

Deposition of diffraction images to be discussed at the Open Meeting of the Commission on Biological Macromolecules of the IUCr in Osaka

Edward Baker,^a Zbigniew Dauter,^b Mitchell Guss^c and Howard Einspahr^d

^aSchool of Biological Sciences, University of Auckland, Private Bag 92-019, Auckland, New Zealand, ^bMCL, National Cancer Institute, Argonne National Laboratory, Biosciences Division, Bldg 202, Room Q142, Argonne, IL 60439, USA, ^cSchool of Molecular and Microbial Biosciences, University of Sydney, NSW 2006, Australia, and ^dPO Box 6395, Lawrenceville, NJ 08648-0395, USA

Over the years the macromolecular crystallography community has been proactive in advocating the deposition of both the results of X-ray structural analyses, in the form of refined coordinates, and of the experimental data from which those results were obtained. This advocacy has resulted in guidelines, formalized by the International Union of Crystallography (IUCr), which shape practice throughout the community.

Today, all journals which publish macromolecular structures require that the refined coordinates be deposited in the Protein Data Bank (PDB). This is also demanded by most funding agencies. All crystallographers (and other structural biologists) treat this obligation as a natural way of making their results available to the community. Deposition of coordinates is nowadays easy through the online systems provided by the PDB. However, it was not always so easy and we, the undersigned, remember times when files were sent to Brookhaven by post in the form of 12 inch magnetic tape reels. The PDB deposition software is constantly improving, validation of deposited models becomes more automatic and elaborate, file formats are more unified and PDB customers have a wide choice of different programs for perusing and analyzing all kinds of information stored in more than 45 000 structural models.

However, the structural models expressed in the form of atomic coordinates (as well as displacement parameters and occupancies) do not constitute the primary data of crystallographic diffraction experiments (or NMR measurements). Community sentiment and the IUCr guidelines have continued to evolve, and for some years the guidelines (and *Acta Crystallographica* practice) required that the structure factor amplitudes used for model refinement be deposited, albeit that a delay of their release of up to six months was permitted. Today some 90% of deposited structures are accompanied by structure-factor amplitudes but, unfortunately, the requirement is not as strictly monitored as for coordinates. We humbly admit that, even in our own journals, several publications have slipped through without deposited experimental data. The deposition of structure-factor amplitudes is now required by some funding bodies, for example in structural genomics projects, and the IUCr Commission on Biological Macromolecules and the NIH directors are working in coordination with the wwPDB to recommend and implement guidelines for the uniform and obligatory deposition of structure factors, to be released on the date of publication of results in scientific journals.

In fact, the structure-factor amplitudes do not represent the actual raw data of a diffraction experiment. Diffraction images are nowadays almost always recorded with two-dimensional detectors, chiefly CCDs and imaging plates, in the form of digitized computer files, and the structure-factor amplitudes result from interpretation of diffraction patterns and integration of reflection intensities by elaborate data-processing programs. Every year the process of data collection and reduction becomes easier and quicker, as the appropriate hardware and software get more elaborate and automatic. With the help of new software pipelines it is sometimes possible to obtain an atomic model of a crystal structure of a macromolecule within minutes after the end of a synchrotron data collection session. Unfortunately, errors can be made and complete reliance on automatic processing of diffraction images can, in difficult or atypical cases, lead to misin-

terpretations, to difficulties in later stages of structure analysis, and even to wrong results. To be able to reinterpret more complicated cases one should return not to the integrated intensities, but to the original set of diffraction images.

A few years ago it was not realistically possible to store thousands of sets of images, each extending perhaps to several gigabytes. With the enormously fast progress in computer technology, however, this task becomes more feasible and in few years may be as easy as the deposition of coordinates is today. It is still too early to demand that images be deposited, but voluntary actions would be a good beginning. In fact, our Australian colleagues have started such an initiative already (<http://www.tardis.edu.au>). The data bank of images may be used not only for potential reinterpretation of structures, but also for developing new data processing programs and for teaching purposes.

The idea of depositing diffraction images is already in the air. We believe that, sooner rather than later, deposition of images will be treated as naturally and routinely as deposition of coordinates is treated now. To us, it is a question then of not whether, but when. This subject will be discussed in an Open Commission Meeting of the Commission on Biological Macromolecules of the IUCr at the Congress in Osaka. This is an opportunity for all members of the community to express their views, as in the earlier discussions on data deposition. Although at present there are certainly many difficulties of a technical and organizational nature, we would like to prompt the readers of *Acta Crystallographica Section D* and *Section F* to express their opinions about this idea. We will be glad to collect the readers' opinions and pass them on for consideration by the Commission.