Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Deposition and release of macromolecular structural data

Edward N. Baker[a]*† and Wolfram Saenger[b]‡

[a]School of Biological Sciences, University of Auckland, Auckland, New Zealand, and [b]Institut fur Kristallographie, Freie Universitat Berlin, Takustrasse 6, 14195 Berlin, Germany

† President, International Union of Crystallography.
‡ Chair, International Union of Crystallography Commission on Biological Macromolecules.

Correspondence e-mail:
ted.baker@auckland.ac.nz

1. Background

There has been much recent discussion on the issue of the deposition and release of macromolecular structural data [see editorials in *Nature*, **391**, 617 (1998) and *Nature Structural Biology*, **5**, 83–84 and 165–166 (1998) and the letter from Wlodawer *et al.* in *Science*, 16 January 1998]. These editorials and letters rightly point out the critical value that macromolecular structural data now have for biology and medicine and the need to make these data available to the wider scientific community.

The crystallographic community has in fact been at the forefront of efforts to ensure that these data should be deposited, for the general good. Some ten years ago, extensive consultation by the International Union of Crystallography (IUCr), through its Commission on Biological Macromolecules, led to a set of guidelines that were intended to encourage deposition of structural data (both atomic coordinates and structure-factor amplitudes) with the Protein Data Bank (PDB). This action has been manifestly successful. Prior to that time, few journals made any requirements and deposition was largely voluntary. Now almost all journals require that at least the coordinates are deposited, and there is widespread acceptance that this should be so.

The original IUCr guidelines included a provision that release of the data might be delayed for a period after deposition (coordinates for 1 year, structure-factor amplitudes for 4 years). This was essentially a compromise, acknowledging the risky long-term nature of the research and its often ‘unfinished’ nature at the time of initial publication. It is this temporary ‘hold’ provision that has been the focus of current attention. The issue is much wider than this, however, involving deposition practice as well as the question of release. The Commission on Biological Macromolecules, on behalf of the IUCr, has therefore engaged in extensive discussion aimed at reviewing the original guidelines. Below, we summarize some of the issues, and set out proposals for the future.

2. Issues to consider

(i) There is wide agreement that the most important issue is that the data should actually be deposited. Some journals still do not require deposition at all. Others do not check that deposition has in fact occurred. Many have deposition criteria only for coordinates and not for structure-factor amplitudes. It is of the utmost importance that strict deposition criteria should be observed, and that this should apply to both the coordinates and the diffraction data, or structure-factor amplitudes. In fact no true validation of published structural results is possible without the structure factors. In contrast, the coordinates provide an elaboration of the structural results rather than a true validation.

(ii) Two ethical principles argue that the deposited data should be released immediately on publication. The first is proper publication practice, which requires that the data on which a publication is based should be available for validation and extension. The second is the public good. This applies with force to publicly funded research, but becomes more complex for academic–industrial collaborations or in-house industrial research.

(iii) Some difficult questions remain. The position of industry-based, or industry-associated, research is one of these. Although this sector is still relatively small, it is significant and growing. If removal of the ‘hold’ provision results in these groups either not publishing, or publishing in journals that do not require deposition, then everyone is the loser. A complete database, albeit with some data unavailable for a short period, is arguably of more value to science than one which is incomplete.

(iv) The question of lower resolution structures, which may be of considerable biological importance, but for which detailed atomic coordinates are hardly appropriate for publication, also has to be considered. Finally, although it is true that many protein structures are now determined far more rapidly than previously, more difficult and less tractable proteins are now being tackled and these still frequently require many years of effort.

(v) There is significant support for the removal of the ‘hold’ provision on coordinates, as judged by the recent *Nature Structural Biology* poll. Of significance, however, is that 33% were opposed, and it is clear from this and from our own consultations that the macromolecular crystallography community is still evenly split. Attitudes are changing, and will change, but it is important that the culture of deposition fostered by our current guidelines is not endangered.

(vi) We recognize that other parties are involved in these issues, including biologists and modellers, as well as funding bodies and journals. It is entirely appropriate that funding bodies should place their own requirements on the researchers they support. Likewise journals are free to set their own policy, and we welcome the importance that is being given to revisiting this issue. The IUCr, for its part, must speak on behalf of the crystallographic community, but also act in the best interests of international science.

3. Recommendations

3.1. Deposition

All publications that describe macromolecular three-dimensional structures at the level of individual atomic positions must be accompanied by deposition of both the atomic coordinates and the structure-factor amplitudes in the appropriate database (PDB or NDB). In the case of low-resolution structures for which only a chain trace is reported a set of Cα positions and structure-factor amplitudes may be sufficient.

3.2. Publication of ID codes

An article should not finally be accepted for publication by a journal until the relevant PDB entry codes have been provided. These entry codes should be quoted in both the text and the abstract (so that they will be picked up by abstracting services). Journals are also urged to check that the appropriate data have actually been deposited, as part of the process of final acceptance.

3.3. Release

Authors are urged to release the atomic parameters and structure-factor amplitudes immediately after the publication date. This should be the normal practice. They can, however, request a delay of 6 months in the release of the atomic parameter data and 1 year in the release of the structure-factor amplitudes. This provision will be formally reviewed at the time of the IUCr Congress in Glasgow, during August 1999.

3.4. Other aspects

The IUCr recognizes the principle that the results of publicly funded research should be publicly available. While this is a matter for the funding bodies in different countries, the IUCr encourages its members to put pressure on funding bodies in the public domain to insist on deposition of atomic parameters and structure-factor amplitudes and to require release at time of publication. The IUCr also urges its members to use their influence to ensure that all journals that publish articles on protein three-dimensional structure do require the deposition of both atomic parameters and structure-factor amplitudes.