

To save crystallization data

Bernard Lorber

Département 'Mécanismes et Macromolécules de la Synthèse Protéique', Laboratoire de Synthèse Protéique et Cristallogénèse, UPR 9002, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 Rue René Descartes, 67084 Strasbourg cedex, France

Correspondence e-mail:
b.lorber@ibmc.u-strasbg.fr

Received 27 September 2000

Accepted 16 January 2001

Note from the Section Editor: the points raised in this paper will be discussed by the IUCr Commission on Journals. We consider that submission of the crystallization data to the relevant database should be strongly recommended.

An eighth international conference in Florida, two practical workshops sponsored by the European Molecular Biology Organization in Germany and France, and an increasing number of multidisciplinary articles confirm that the fine art of crystallizing biological macromolecules is a field of research *per se*. On one hand, crystallogenesis investigations combining biological with physical chemical and mathematical approaches on model proteins such as hen egg-white lysozyme gradually shed light on the mechanisms of crystal nucleation and growth. On the other, data accumulate on thousands of macromolecules of which some are extensively studied and others with a potential pharmaceutical interest are investigated with the sole purpose of identifying those promising for crystallographic analysis. Surprisingly, in both cases the experimental conditions that produce exploitable crystals are often neglected with regard to the three-dimensional structures that are derived from them. Once an average sample composition is reported that is supposed to yield the best crystal form, all technical details about its crucial preparation and its failures remain buried in inaccessible laboratory notebooks. The belief that good quality crystals are just a bridge to crystallographic structures and the lack of understanding of how crystals which are obtained today can be the key to preparing those of tomorrow may explain this situation. Together with the facts that our current knowledge suffers from decades of incomplete reports and that printed space is limited in publications, this should stimulate us to implement a rational method to save the hard-earned crystallization recipes in a standardized way before they are lost forever. Collecting all data in a single repository would open the possibility of processing them and distributing them throughout the scientific community. A number of academic and industrial laboratories have already set up in-house databases and the idea seems worth pursuing.

Which publicly accessible database will be able to deal with the avalanche of information anticipated from automated high-throughput crystallization experiments performed in the frame of structural genomics and proteomics projects or from large-scale assays to be performed aboard the International Space

Station? Presently, the Biological Macromolecule Crystallization Database (<http://www.bmcd.nist.gov:8080/bmcd/bmcd.html>), which includes the NASA Protein Crystal Growth Archive, is one candidate. This 'basis for a crystallization strategy' compiles from the literature over 3500 crystal entries for more than 2500 macromolecules (proteins, nucleic acids, complexes and viruses). To expand, it urges crystal growers to contribute their results. The Protein Data Bank (<http://www.rcsb.org/pdb/>), which successfully centralizes the atomic coordinates of more than 14 000 biological macromolecules, is another candidate. It will soon record details on protein expression, purification, characterization and crystallization (Berman *et al.*, 2000).

Thus, analogously to the systematic deposition of atomic coordinates prior to publication, the deposition of crystallization data can become a common practice. The accepted database should not only contain optimal conditions but also the ranges of pH, temperature, macromolecule and precipitant concentrations *etc.* where crystals grow, as well as morphology changes and diffraction properties. Links to related sequence, three-dimensional structure and publication databases will save search time. For crystal growers, the benefit will be a broader view of the crystallogenesis of a variety of structurally distinct molecules and assemblies possessing characteristics that are far from those of the present-day models. The search for crystallization conditions of functionally similar macromolecules having a close solution behavior will increasingly rely more on statistical methods than on an empirical hunt with sparse matrices (Hennessy *et al.*, 2000). As a novel tool in structural biology, the crystallization database will also be highly relevant to macromolecular and biomaterials researchers designing structures and their interactions.

References

- Berman, H. M., Bhat, T. N., Bourne, P. E., Feng, Z., Gilliland, G., Weissig, H. & Westbrook, J. (2000). *Nature Struct. Biol. Struct. Genomics Suppl.*, pp. 957–959.
- Hennessy, D., Buchanan, B., Subramanian, D., Wolosz, P. A. & Rosenberg, J. M. (2000). *Acta Cryst. D* **56**(7), 817–827.