

Of crystals, structure factors and diffraction images

Luca Jovine,* Ekaterina Morgunova and Rudolf Ladenstein

Center for Structural Biochemistry, Department of Biosciences and Nutrition, Karolinska Institutet, Hälsovägen 7, SE-141 57 Huddinge, Sweden. Correspondence e-mail: luca.jovine@ki.se

It is suggested that it would be useful if raw X-ray diffraction images could be included in data depositions with the Protein Data Bank.

Recent comments have stressed the importance of depositing with the Protein Data Bank (PDB) experimental data in the form of structure factors, in addition to atomic model coordinates, whenever publications describing crystallographic structures are submitted (Jones & Kleywegt, 2007; Joosten & Vriend, 2007). We suggest that investigators should also be allowed to deposit raw X-ray diffraction images with the PDB, because these constitute the actual experimental evidence from which structure factors are computationally derived. In light of the current low cost of data storage (< 1\$ per GB), ongoing efforts towards the development of an optimal deposition format (Hammersley *et al.*, 2005) and the availability of highly efficient lossless image compression algorithms (Witten *et al.*, 1999), we believe that such depositions would be preferable for several reasons. Firstly, conversion from images to structure factors depends on a number of parameters. Some of these (X-ray wavelength, crystal-detector distance, oscillation range and detector type) are determined by the experimental setup and should also be specified when depositing raw data. Others – such as crystal space group, resolution cutoff and data range used for scaling – are ultimately the result of decisions made by crystallographers. While in most instances these choices are straightforward, it is nevertheless impossible to reevaluate the data unless the original diffraction images are available. This is particularly relevant if, for example, a space group of too high a symmetry has been assigned because of undetected crystal twinning; although careful analysis of structure factors can reveal such cases, users do not have the option of reprocessing the data in the correct space group. Secondly, general availability of diffraction images would undoubtedly be beneficial for further development of data reduction software, on which all other steps of crystal structure determination depend. For instance, current data processing programs do not make use of diffuse scattering information, which can be evident in diffraction images but is lost upon conversion into structure factors (Fig. 1) (Glover *et al.*, 1991; Wall *et al.*, 1997). Thirdly, and most importantly, a centralized archive of raw X-ray data would ensure that this experimental evidence is preserved, to the long-term advantage of both depositors and users. For the same reason, the PDB could consider collaborating with nonprofit organizations such as Addgene (<http://www.addgene.org/>), with the aim of creating a physical repository of expression constructs successfully used for structure determination. Macromolecular models consist of a simple list of spatial coordinates; as crystallographers, we have the responsibility to ensure that the often hard-won experimental evidence

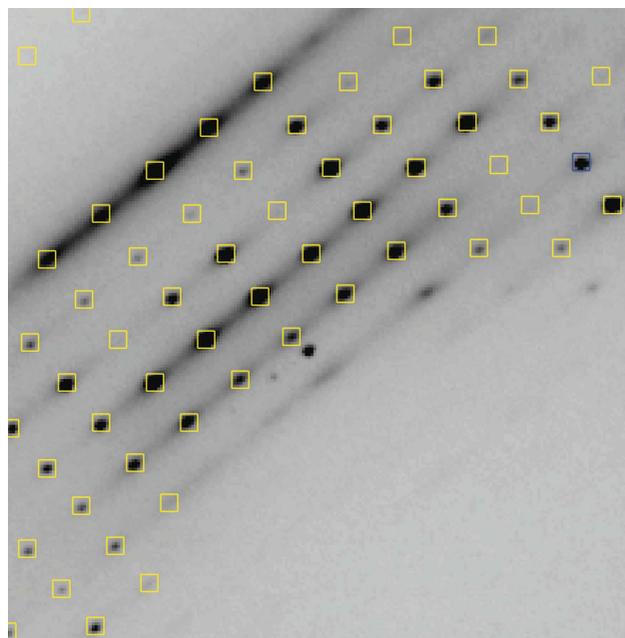


Figure 1
 Example of strong X-ray diffuse scattering from a trigonal crystal of 4.5S RNA domain IV (PDB code 1duh). Although the significant intensity of the regions between Bragg peaks (boxes) contains information about crystal disorder, currently this is not taken into account by standard data processing software.

leading to these numbers is maintained, so that future generations will be able to reproduce as well as improve on our findings.

References

- Glover, I. D., Harris, G. W., Helliwell, J. R. & Moss, D. S. (1991). *Acta Cryst.* **B47**, 960–968.
- Hammersley, A. P., Bernstein, H. J. & Westbrook, J. D. (2005). *Image CIF Dictionary (imgCIF) and Crystallographic Binary File Dictionary (CBF)*. Version 1.3.2. ftp://ftp.iucr.org/cifdics/cif_img_1.3.2.dic.pdf.
- Jones, T. A. & Kleywegt, G. J. (2007). *Science*, **317**, 194–195.
- Joosten, R. P. & Vriend, G. (2007). *Science*, **317**, 195–196.
- Wall, M. E., Clarage, J. B. & Phillips, G. N. (1997). *Structure*, **5**, 1599–1612.
- Witten, I. H., Moffat, A. & Bell, T. C. (1999). *Managing Gigabytes: Compressing and Indexing Documents and Images*, 2nd ed. San Francisco: Morgan Kaufmann.