

Expectation bias and information content

Zbigniew Dauter,^{a*} Manfred S. Weiss,^{b*} Howard Einspahr^{c*} and Edward N. Baker^{d*}

^aArgonne National Laboratory, Biosciences Division, Bldg 202, Room Q142, Argonne, IL 60439, USA, ^bHelmholtz-Zentrum Berlin für Materialien und Energie, Macromolecular Crystallography (HZB-MX), Albert-Einstein-Strasse 15, D-12489 Berlin, Germany, ^cPO Box 6483, Lawrenceville, NJ 08648-0483, USA, and ^dSchool of Biological Sciences, University of Auckland, Private Bag 92-019, Auckland, New Zealand

It has long been realised that the interpretation of macromolecular structures from electron-density maps involves some degree of subjectivity (Brändén & Jones, 1990). This is a direct consequence of factors such as limited data resolution or inherent disorder, in solvent regions, in segments of a macromolecule, or in a bound ligand. Whereas the well behaved parts of the structure may often be confidently interpreted, and often refined completely automatically without human intervention, in practice there are usually some regions that require very careful inspection and interpretation by the experienced scientist. Often such difficult-to-interpret features are simultaneously the most important in the context of chemistry and biology, as with, for example, biologically active ligands interacting with enzymes or receptors.

For quite some time now, it has been the rule that publication of crystal structures must be accompanied by the deposition of corresponding coordinates and structure-factor amplitudes in the Protein Data Bank (PDB). Of course, the PDB is limited to curation of deposited structure data and does not correct or re-interpret the results submitted by the authors. However, with all data being available to the community, anyone is now in a position to generate the relevant electron-density maps and interpret the results independently. Unfortunately, as evidenced in the two associated articles by Pozharski *et al.* (2013) and Weichenberger *et al.* (2013) in the current issues of *Acta Crystallographica, Section D* and *Acta Crystallographica, Section F*, respectively, re-interpretation sometimes unearths serious problems. In the present case, the focus is on protein structures with small-molecule ligands, where the ligands are seriously misinterpreted.

There may be various reasons for such misinterpretations, including lack of experience, inadequate supervision or apparent wishful thinking, that is, too strong a belief in expectations or importance. Whatever the cause, such misinterpretations present false results and may impair subsequent research based on them.

It is our view that proper validation or re-interpretation of structures deposited in the PDB, either by actions such as those presented in the above-mentioned papers, or by the activity of initiatives such as the PDB_REDO group (Joosten, Salzemann *et al.*, 2009; Joosten, Womack *et al.*, 2009; Joosten *et al.*, 2012), are highly beneficial for the community of structural biologists. They not only uncover problem structures, but also keep us mindful of potential pitfalls in our own work and that of others. Such efforts should be applauded as a great service to us all.

References

- Brändén, C.-I. & Jones, T. A. (1990). *Nature (London)*, **343**, 687–689.
Joosten, R. P., Joosten, K., Murshudov, G. N. & Perrakis, A. (2012). *Acta Cryst.* **D68**, 484–496.
Joosten, R. P., Salzemann, J. *et al.* (2009). *J. Appl. Cryst.* **42**, 376–384.
Joosten, R. P., Womack, T., Vriend, G. & Bricogne, G. (2009). *Acta Cryst.* **D65**, 176–185.
Pozharski, E., Weichenberger, C. X. & Rupp, B. (2013). *Acta Cryst.* **D69**, 150–167.
Weichenberger, C. X., Pozharski, E. & Rupp, B. (2013). *Acta Cryst.* **F69**, doi:10.1107/S1744309112044387.