

Numerology

Zbigniew Dauter^a and
Ted Baker^b

^aMCL, National Cancer Institute, Argonne
National Laboratory, Biosciences Division, Bldg
202, Room Q142, Argonne, IL 60439, USA,
and ^bSchool of Biological Sciences, University of
Auckland, Private Bag 92-019, Auckland, New
Zealand

It is obvious that the numerical results of all scientific activity should be presented faithfully, with precision corresponding to the best possible estimate of accuracy of various measured or calculated values. Certainly the same applies to macromolecular crystallography as well. However, the readers and editors of *Acta Crystallographica Section D* (as well as other journals), and users of the Protein Data Bank, are sometimes perplexed by the effects of 'numerology', which sometimes (frankly, too often) sneak into otherwise valuable scientific reports.

Several more or less typical examples of numerology are common in the description of crystal structures and the underlying data. Reporting cell dimensions such as 123.456 Å, with precision far exceeding that of the used X-ray wavelength or crystal-to-detector distance, or giving a cell volume with seven-digit precision are not uncommon. The r.m.s.d. values of bond lengths in refined protein models are quoted with precision that is too low, e.g. 0.01 Å, or with too optimistic precision, such as 0.004631 Å in the two (independently refined by the same authors!) structures in the PDB. As Alex Wlodawer (2007) pointed out, the latter value is given with precision exceeding the diameter of an electron. The other examples of such 'wishful thinking' are reporting atomic *B* factors with three or more (in one case six) decimal digits, putative hydrogen bonds with three or more decimal digits, or the accessible protein surface or other parameters difficult even to define accurately, with too optimistic precision.

In the majority of cases, such unrealistic numerical values result from uncritical copying from a calculator or a program output. Some programs indeed report certain numerical results with too many digits and we should urge the programmers to pay more attention to the format of the numerical values printed out by their software. Cases of such numerology are often spotted and corrected by journal editors or reviewers, but of course the ultimate responsibility for reporting results accurately, appropriately and faithfully remains in the hands of the authors.

In the past, macromolecular crystallography was practiced mostly by researchers with a rigorous background in physical sciences. Today methods have become much improved and programs highly automatic and easier to use, thus structure solution and refinement usually takes days, not months. Protein crystallography has matured and is (rightly) used as a tool by biologists, who may not have a good feeling for the strict meaning of some of the reported numerical values. In contrast to small-molecule crystallography, macromolecular crystallography does not usually allow a strictly statistically valid estimation of the true accuracy of the resulting parameters to be obtained. However, in most cases the examples of numerology result not from the lack of knowledge, but from the lack of critical thought in quoting numerical results.

We ask the authors submitting their results to our (and other) journals to be vigilant in reporting the numerical values in their manuscripts, in data submitted to the Protein Data Bank, and in other presentations.

References

Wlodawer, A. (2007). *Acta Cryst.* **D63**, 421–423.