Letters

Crystallographic Citations

Daniel E. Koshland, Jr. (Editorial, 5 Aug., p. 637), draws attention to obligations of scientists in assuring the correctness of research publications, including the axiom that all data needed to verify the reported results should be available to the reader. He quite rightly identifies some of the special problems associated with results from modern science that are too voluminous to print in journal pages. The need to deposit coordinates for x-ray structures of macromolecules is a timely example, and Koshland uses this to illustrate his point. I wish to draw attention to an associated scientific obligation to cite the primary sources of all information used in conducting studies that are reported in research papers. In particular, I am disturbed by a trend in structural surveys toward citations only to the Brookhaven Protein Data Bank (*1*) and not to the original literature.

Information stored in the Protein Data Bank proves to be very useful in several kinds of studies, as Barbara Jasny reports (Research News, 6 May, p. 722). Important among these applications are structural comparisons designed to elucidate general principles of protein conformation, folding, and evolution. Unfortunately, several significant articles of this kind published recently in *Science* (*2*) and elsewhere (*3*) do not cite primary references for the coordinates that were used. I raise these examples not to discredit the findings or these particular authors but to show pervasiveness. This practice has several regrettable consequences: the contributions of the original investigators go unrecognized; readers are forced to access the magnetic database, rather than library holdings, in order to check on the provenance of data; and sometimes the specific identity of coordinate sets is left ambiguous. Moreover, in some instances, "discoveries" based on a straight survey of the database may have antecedent descriptions in the original literature that go overlooked.

This situation with respect to crystallographic citations is somewhat understandable in light of confusion from past deposition practice. Initially, deposits into the Protein Data Bank were viewed as discretionary rather than obligatory. No doubt, data were withheld by some out of sloth or selfishness, as suggested by Koshland; but a concern for accuracy motivated many to hold back until satisfactory refinement. In any case, many early deposits are not strictly associated with particular publications. Some ambivalence toward deposition still persists in the community.

During recent committee deliberations on deposition policy, it became clear that such policy should be founded on the principle that a scientific publication ought to report both the observations and the results of the investigation. In this view, the data (diffraction measurements) and primary results (atomic coordinates) from a crystallographic study are essential components of a structural report, and these should be deposited into the Protein Data Bank as supplements to the publication. It follows that users of these supplementary materials are obliged to cite the associated publication as well as the repository.

Adherence to these standards of scholarship in citing the literature can be accommodated with little perturbation in most cases. Only on the order of 20 to 40 citations would have been required for each of the studies noted above (*2, 3*), and comparable numbers were readily handled in similar studies (*4*). Even in the event of studies based on multitudinous publications, one can hold to principle and deposit references as supplementary material. Murray-Rust and Glusker (*5*) deposited 46 pages of references to hundreds of structures from the Cambridge Crystallographic Data File that were analyzed for O–H-X hydrogen bonding interactions.

Spurred by the example and the proddings of respected leaders, the crystallographic community is now markedly more forthcoming with depositions than it was even a short time ago. The supportive editorial policy of *Science* and other journals, as well as the vigilance of reviewers, are very much appreciated in this regard. I trust that users of these deposits will see that citation to the original literature is an associated responsibility.

WAYNE A. HENDRICKSON
Howard Hughes Medical Institute,
Department of Biochemistry
and Molecular Biophysics,
College of Physicians and Surgeons of
Columbia University,
New York, NY 10032

REFERENCES

1. F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977).
2. L. G. Presta and G. D. Rose, *Science* **240**, 1632 (1988); J. S. Richardson and D. C. Richardson, *ibid.*, p. 1648.
3. S. Rackovsky and D. A. Goldstein, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 777 (1988); J. W. Ponder and F. M. Richards, *J. Mol. Biol.* **193**, 775 (1987); E. J. Milner-White, *ibid.* **199**, 503 (1988); D. J. Barlow and J. M. Thornton, *ibid.* **201**, 601 (1988).
4. S. Miller, A. M. Lesk, J. Janin, C. Chothia, *Nature* **328**, 834 (1987); N. Thanki, J. M. Thornton, J. M. Goodfellow, *J. Mol. Biol.* **202**, 637 (1988).
5. P. Murray-Rust and J. P. Glusker, *J. Am. Chem. Soc.* **106**, 1018 (1984).

Response: Nolo contendere.

Of course, one can think of extreme circumstances in which coordinates are available but the relevant citation is not obvious, or studies in which very large numbers of citations require special treatment—but these are quibbles.

I fully support both parts of Hendrickson's proposal: responsible citation of the primary literature and timely deposition of x-ray coordinates.

GEORGE D. ROSE
Department of Biological Chemistry,
Hershey Medical Center,
Pennsylvania State University,
Hershey, PA 17033

Response: Although listed among the "badies" in Hendrickson's letter, I would like to strongly support the points he makes.

Independent but related efforts are now under way by groups of independent investigators and by the International Union of Crystallography to regularize the deposition of x-ray data and models at, or close to, the time of publication. If these efforts result in the acceptance of reasonably uniform deposition requirements by the relevant scientific journals, then the problem addressed by Hendrickson will be the principal outstanding issue. Proper credit must be given on a regular basis to those who have invested the enormous amount of time and effort that goes into the solution of a macromolecular structure.

There are several problems with citation.

1) Reports of studies employing files from the Protein Data Bank should identify the specific files used. Journals and their referees should demand this in future manuscripts.

2) The original publication to which reference should be made is usually clear in the Protein Data Bank file, but not always. The original authors should supply this information. In the absence of a regular journal article, the reference might, in some form, be to the Protein Data Bank file itself. The Advisory Board to the Data Bank is an appropriate group to suggest the possible forms of reference.

3) In some studies the large number of references used presents a logistical problem. Up to 50 or so references can be handled by the current standard procedures. The potential expansion of the Data Bank may expand this number to the point of impracticality. Hendrickson's suggestion that this material be deposited as supplementary material is reasonable, but it has one drawback. The intense competition in the current scientific scene has unfortunately reinforced the paper and citation counting that so frequently plays an important role in