Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Black sheep among the flock of protein structures

Zbigniew Dauter[a] **and Edward N. Baker**[b]

[a]National Cancer Institute, Argonne National Laboratory, Biosciences Division, Bldg 202, Room Q142, Argonne, IL 60439, USA, and [b]School of Biological Sciences, University of Auckland, Private Bag 92-019, Auckland, New Zealand

The recent announcement made by the University of Alabama at Birmingham (USA) that a number of crystal structures produced in the laboratory of Dr H. M. Krishna Murthy will have to be removed from the Protein Data Bank and retracted from the literature, spread a shockwave among the macromolecular crystallography community. These structures were published over a period of seven years (1999–2006), and included such important proteins as dengue virus protease (1bep, 1df9, 2qid), complement component proteins (2hr0, 1g40, 1g44), vaccinia complement proteins (1rid, 1y8e), apolipoproteins (1i6l, 2ou1, 2a01), and Taq DNA polymerase (1cmw, 1bgx). In the past there have been cases in which structures determined by X-ray crystallography have had to be retracted because of errors in data interpretation or in the programs utilized in structure solution. These are understandable, as mistakes do occur. This time, however, it appears that the retracted structures were deliberately fabricated and there is no evidence that any experimental data were actually collected.

This is of course very bad for the credibility of X-ray crystallography. It has always been accepted that diffraction methods, due to the intrinsic interdependence between direct and reciprocal space, give what are arguably the most faithful and difficult-to-manipulate results of all approaches used in structural chemistry and biology. However, the current case shows that, in spite of the existence of many validation methods, it has proved possible to smuggle fraudulent structures through the process of manuscript refereeing and PDB deposition. One good point, however, is that flags were raised by vigilant members of the macromolecular crystallography community, who detected some definite abnormalities in the information stored for the incriminated structures in the PDB.

In a way, this sad story has been possible because of the accelerated progress of protein crystallography. The stream of new protein X-ray structures is now so rapid that it was possible for a few black sheep to sneak through the net of shepherd dogs (referees) and join the flock (PDB). The current situation suggests that more strict validation of X-ray models is required prior to publication, and that the validation data should be made available to the referees and others who evaluate new structures. It is also clear that submission to the PDB of structure-factor amplitudes must be enforced (half of the structures in question were deposited without structure factors, in violation of the rules of the journals and the funding agencies). Efforts are in progress towards these goals on the part of both the PDB and the IUCr. In this context, the current initiatives to store the raw diffraction images, which would give the opportunity for ultimate validation of the structure analyses, have special importance.