

## Giving credit where credit is due

### To the editor:

Structural genomics<sup>1,2</sup> efforts have now been underway in the United States for five years under the auspices of the Protein Structure Initiative (PSI). The first phase of the PSI is nearing completion, with the second one to follow. A number of centers in other countries are also engaged in similar work. Although the emphasis on solving structures quickly with little or no time to analyze them in any detail did not receive great enthusiasm from many old-fashioned structural biologists (including this writer), it is no longer possible to ignore the contribution of structural genomics to modern structural biology. Many structures deposited in the Protein Data Bank (PDB)<sup>3</sup> are the result of structural genomics efforts<sup>4</sup> (1,540 by a recent count). The US rules require deposition of the PSI-derived structures within six weeks of completion (to be shortened to four weeks in the second phase of PSI), and the worldwide rules allow only a six-month delay. Usually, their release by the PDB follows almost immediately, leaving little time for publication. This raises two questions: what rules should apply to the use of structural genomics results by others, and how should these structures be referenced in order to appropriately credit the scientists who solved them?

The functions of proteins solved by structural genomics are often not known. However, with the continuing development of theoretical 'structure to function' methods<sup>5-7</sup> and with structure-inspired experimental work aimed at explaining the biological significance of such novel proteins, groups other than the original depositors will be able to derive important information from the deposited structures. It is clear that such efforts should not only be allowed but encouraged, as long as proper credit is given to those who originally solved the structures. What is less clear is how best to reference structural genomics-derived structures.

Although PDB codes are unique and permanent identifiers of coordinate sets, they are not considered publications and thus they cannot be cited in journal articles as normal references. However, accessing a record in the RCSB PDB

(<http://www.rcsb.org/pdb/>)<sup>8</sup> yields much more information than what can be found solely in the coordinate files themselves. Through a series of links, one can view the structure, see details of ligand interactions, learn about the protein's structural classification and its structural neighbors, and access other types of information. Together with details of expression, crystallization and data collection listed in the PDB coordinate file, these data are not much less than what can be found in a published structural communication. If more methodology is included, it should be possible to use a deposited structure's PDB website as a preliminary reference. All that is needed is a digital object identifier understood by reference software. Ideally, such a 'publication' should not preclude subsequent submission of a manuscript describing the structural work to a scientific journal, but it should suffice as a proper reference in the meantime.

Another problem that needs to be discussed relates to publication ethics and is not limited to referencing the output of structural genomics. It is not uncommon for two or more laboratories to publish almost simultaneously structures that are the same or highly similar. Quite often the authors simply pretend that their competition does not exist and fail to reference work performed by competing laboratories. How far should that go?

Again, I believe that the release of PDB coordinates (as opposed to their submission, often kept anonymous) should be the guiding principle and that such released coordinates should be properly cited. All authors of structural reports should feel obligated to check the PDB before finalizing their papers for journal submission and publication and should give proper credit where due. It would help if the PDB could include a list of structures that are similar either in sequence or fold as part of the evaluation returned to the author of a structure submission. The author would be free to decide whether or not to use this information, but at least it would be available. Obviously, we cannot expect either journals or the PDB to play a policing role, but clear instructions to authors and reviewers outlining the obligation to cite

structural genomics results or data from competitors would go a long way toward resolving the problem. I wonder whether the structural biology community would agree.

Alexander Wlodawer

Macromolecular Crystallography Laboratory,  
National Cancer Institute, Frederick, Maryland  
21702, USA. e-mail: wlodawer@ncifcrf.gov

1. Vitkup, D., Melamud, E., Moulton, J. & Sander, C. *Nat. Struct. Biol.* **8**, 559–566 (2001).
2. O'Toole, N., Grabowski, M., Otwinowski, Z., Minor, W. & Cygler, M. *Proteins* **56**, 201–210 (2004).
3. Berman, H.M., Henrick, K. and Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980 (2003).
4. Todd, A.E., Marsden, R.L., Thornton, J.M. & Orengo, C.A. *J. Mol. Biol.* **348**, 1235–1260 (2005).
5. Orengo, C.A., Todd, A.E. & Thornton, J.M. *Curr. Opin. Struct. Biol.* **9**, 374–382 (1999).
6. Wolfson, H.J. et al. *Curr. Protein Pept. Sci.* **6**, 171–183 (2005).
7. Zhang, C. & Kim, S.H. *Curr. Opin. Chem. Biol.* **7**, 28–32 (2003).
8. Berman, H.M., et al. *Nucleic Acids Res.* **28**, 235–242 (2000).

### Helen M. Berman responds:

In order to integrate the growing number of unpublished structure entries in the PDB with the larger body of scientific literature, the PDB will assign a document object identifier (DOI) to each PDB entry.

The PDB has recently contracted with the CrossRef organization to obtain DOIs for all of the entries in the current archive and for future entries. CrossRef provides a service to the community of scholarly publishers linking DOIs to their corresponding electronic publications. The DOI has therefore become a universal electronic identifier for scientific literature. With a DOI assigned to each PDB entry, structures deposited in the PDB can be referenced and accessed in the same manner as other electronic publications.

Helen M. Berman

Director, RCSB Protein Data Bank, Department of  
Chemistry and Chemical Biology, Rutgers, The State  
University of New Jersey, Piscataway, New Jersey  
08855, USA. e-mail: berman@rcsb.rutgers.edu