

nature Structural biology

molecular form & function

may 2003
vol. 10 no. 5

A database for 'em

In recent years, the question of open access to published scientific information has been hotly debated, which seems curious because publishing, after all, involves communicating relevant information to readers. As intensive projects have produced exquisitely detailed macromolecular structures and vast libraries of nucleic acid sequence data, the traditional scientific publishing model has been complicated slightly. Every large data set will contain errors, for example, and scientists may prefer not to make such data sets available to others until they are confident that errors have been minimized to an acceptable level or eliminated altogether. Alternatively, investigators who have invested time and effort to produce a given data set might wish to restrict access to it, so they can retain priority in making future discoveries based on the data, which could, of course, have attendant commercial implications.

Common sense has prevailed in the world of structural biology, and it is now accepted that scientists who publish macromolecular structures obtained by X-ray crystallography will deposit the coordinates for their atomic models in public databases, principally the protein data bank (PDB). *Nature Structural Biology* and its older sibling *Nature* require that supporting data from crystallographic studies are also deposited, which should make the information more useful to interested parties. Atomic coordinates obtained using NMR are deposited in PDB too, as are those gleaned from electron microscopic (EM) work. But EM has yielded rather less in terms of data that are both portable and usable. This is partly because the information generated by EM is different from that produced by other techniques, taking the form of 'envelopes of electron densities', which may, although worthwhile in their own right, be difficult to interpret completely or unambiguously.

With a view to establishing a systematic repository for EM data, EM database (EMDB, previously known as 3D-EM; <http://www.ebi.ac.uk/msd-srv/emdep/helpGeneral.html>) has now been established under the auspices of the European Bioinformatics Institute (EBI) near Cambridge in the UK. The goal of this European Union-funded project is to provide a per-

manent archive for data on macromolecular structures that have been studied by EM. EMDB is part of a larger project known as MSD (macromolecular structure database), which should allow the efficient dovetailing of information on a given macromolecule or assemblage, which has been harvested by different techniques. In the future, it is likely that EMDB will evolve so as to be more international in flavor, perhaps with several portals permitting access to a single database along the PDB model.

This development promises to be valuable for structural, molecular and cell biologists alike. Not only will EM data be stored reliably for future use (together with detailed information on experimental parameters), but the data itself should become much more useful. Some of the value will derive simply from the establishment of a standard format for deposited data, which should make the information readily accessible and usable without the need for specialized software. Comparing data from independent EM studies on related molecules will be facilitated, which should speed progress in deciphering the function of complex and dynamic entities like ribosomes or ion channels. And, viewing the PDB pejoratively for a moment as a child's construction kit of macromolecular fragments, the project should provide impetus to the challenging task of interpreting EM structures using partial information derived from crystallographic and/or NMR work.

At the time of writing, EMDB is still in its infancy, containing ~40 EM structures (although some of these maps are subject to hold periods of one, two or four years, which the curators currently allow). But more than a hundred EM structures are now being published in the journals in a typical year. Without EMDB, these data would not be archived for future general use. So the size and usefulness of the database are likely to increase dramatically. *Nature Structural Biology* is strongly supportive of the general principle that scientific data should be professionally maintained and freely accessible, and so its editors will from now on encourage scientists to deposit their work in EMDB when papers describing EM structures are published in the journal. We hope to be reporting on the project's success in the future.