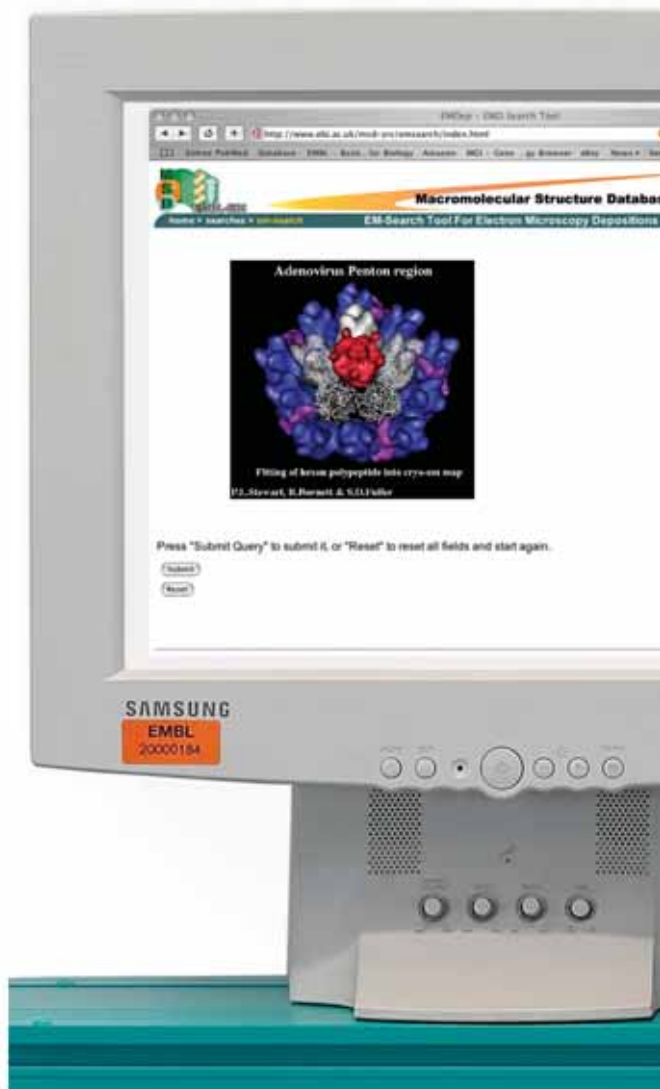


# When large, fuzzy objects enter the database

The EMSD expands to incorporate electron microscope images

If you want to see snapshots of protein machines, like those Patrick Aloy and Rob Russell have been studying, you can now look them up on the Internet. Among the many other types of biological data that it provides, the European Bioinformatics Institute has now added





**Left:** Kim Henrick. **Above:** the EMSD

images obtained by cryo-electron microscopy (cryo-EM) to one of its databases.

Until now, visitors to the European Macromolecular Structures Database (EMSD) have been able to obtain detailed data about the physical structure of proteins. Usually that information comes from X-ray crystallography experiments, where the goal is to achieve the highest possible resolution. In the best cases, the structure will reveal the position of single atoms within a molecule or a complex. That permits detailed insights into molecules: how they bind to each other, the chemical properties that control their activity, and other aspects of their functions.

But getting information at this level of detail usually requires obtaining proteins in crystal form, which is often difficult or impossible with large molecules and complexes. With methods like cryo-EM, researchers can sometimes obtain lower-resolution pictures that still provide useful information: they may show the overall shape

of a molecular machine, for example. That may allow scientists to fit bits and pieces of more detailed structures into maps.

Other scientists would profit from these images if they could get to them easily. This led Kim Henrick and the EMSD team to propose a project that would add lower-resolution structures to the database. They received initial funding from the European Union to do so. Now scientists can add images obtained through EM and view those submitted by others by going through the project website.

“Anyone who has obtained three-dimensional images of molecules using cryo-EM can use our new web deposition system,” says Richard Newman, who has been working on the project with Kim’s team. “That involves more than simply uploading their images. To make the data most useful, we have to capture information about what’s in a structure, how the experiments were designed and carried out. The scientist who created the image knows all of that, and the submission process is set up to capture his or her expertise. That’s done in an interactive way thanks to submission tools designed by Mohammed Tagari.”

Cryo-EM information is different than atomic-resolution data on proteins. Richard says you can think of an image as a transparent box, filled with small blocks, arranged in the shape of the protein complex or the object. The size of each block represents the pixel size – the maximum resolution of the picture. That will be lower than most structures obtained through X-rays – where scientists aim to achieve one-Ångstrom resolution, the diameter of a hydrogen atom. The very best electron microscope pictures have a pixel size of a few Ångstroms.

The resource opened to submissions in 2002 in a pilot phase, and researchers submitted data on viruses, ribosomes and chaperonins. The database currently holds about 70 structures. “The work will continue to expand and be maintained through funding from the European Commission, under a new network of excellence,” Kim says. “Its focus is specifically three-dimensional electron microscopy. That will allow us to bring in more depositions, especially from Europe.”

The more data that is submitted, the more useful the database will be, Richard says. The journals *Nature* and *Nature Structural and Molecular Biology* have already agreed that authors of papers need to put their structures into the database as part of the paper submission process. This step was a key factor in turning other EMBL databases – of genes and proteins, for example – into key resources for the scientific community. The EBI expects other journals to follow suit soon.