

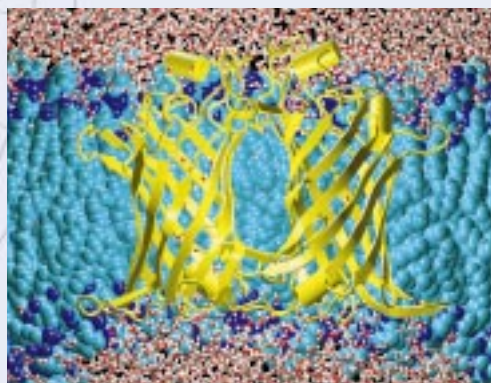
Stimulating simulated biology

The aim of the BioSimGrid project is to make the results of large-scale computer simulations of biomolecules more accessible to the biological community. Such simulations of the motions of proteins are a key component in understanding how the structure of a protein is related to its dynamic function.

A major impediment to making such simulations available to biologists has been the absence of a database of simulation results. Using the power of the Grid the developers hope to solve these problems, thus enabling them to take a comparative, i.e. genuinely biological, approach to analysis of simulation data.

In an ideal world, all simulation data would be available to all interested parties. However, at present simulation data reside in the "home" laboratory and are not accessible to other research groups. One solution to this problem would be to deposit all simulation results in a centralised database, but in reality the vast quantities of data mean that rapid access would become impossible. The Grid provides the opportunity to draw together distributed collections of simulation data in disparate formats, whilst maintaining a centrally accessible federated database.

The developers have selected an initial application that stretches their methodology and also allows them to explore an important biological question. They are comparing simulations of two enzymes: (i) acetylcholinesterase (AChE), a key enzyme of the nervous system; and (ii) OMPLA, a bacterial enzyme involved in pathogenesis. Structural data show that these two enzymes have similar active sites (a triad of amino acid sidechains



involved in their catalytic mechanisms). Otherwise, the structures of the two proteins are completely unrelated. The researchers are analysing simulations to compare the patterns of catalytic sidechain dynamics in these two distantly related enzymes, so as to understand the relationship between sidechain mobility and catalytic mechanism.

But, there is a catch. The AChE simulations were performed in Andrew McCammon's group at the University of California, San Diego; the OMPLA simulations in Mark Sansom's group in Oxford. Normally, such data would never "meet" and would reside in the separate laboratories. Furthermore, the simulations were run using different programs and protocols and the data are in very different formats. Rather than re-run one (or both) of the simulations (which would consume many days of costly supercomputer time), the researchers are using BioSimGrid to make the comparison. This apparently simple test case is non-trivial. It involves the researchers working closely with the simulation code developers to generate methods for automatic data deposition and for generation of ontogenies for simulation data. However, by bringing together a team of simulation and computer scientists



BioSimGrid can simulate proteins in motion

they are making good progress. One of the first successes will be to have a definitive comparison of the active sites dynamics of AChE and OMPLA.

This provides a microcosm of the many comparisons that will become possible once BioSimGrid is fully operational. Biomolecular simulation groups from around the UK will be able to deposit their simulation data for wide-ranging comparative analyses that so far have been impossible. This will propel biomolecular simulation studies into the post-genomic era.

www.biosimgrid.org/

BioSimGrid Project Leader: Mark Sansom,
University of Oxford