ProteinShop Gives Researchers a Hands-On Tool for Manipulating, Visualizing Protein Structures

The Human Genome Project and other biological research efforts are creating an avalanche of new data about the chemical makeup and genetic codes of living organisms. But in order to make sense of this raw data, researchers need software tools which let them explore and model data in a more intuitive fashion.

With this in mind, researchers at Lawrence Berkeley National Laboratory and the University of California, Davis, have developed ProteinShop, a visualization and modeling program which allows researchers to manipulate protein structures with pinpoint control, guided in large part by their own biological and experimental instincts.

Biologists have spent the last half century trying to unravel the "protein folding problem," which refers to the way chains of amino acids physically fold themselves into three-dimensional proteins. This final shape, which resembles a crumpled ribbon or piece of origami, is what determines how the protein functions and translates genetic information. Understanding and modeling this geometrically complex formation is no easy matter.

ProteinShop takes a given sequence of amino acids and uses visualization guides to help generate predictions about the secondary structures, identifying alpha helices and flat beta strands, and the coil regions that bind them. Once secondary structures are in place, researchers can twist and turn these pre-configurations until they come up with a number of possible tertiary structure conformations. In turn, these are fed into a computationally intensive optimization procedure that tries to find the final, three-dimensional protein structure.

Most importantly, ProteinShop allows users to add human knowledge and intuition to the protein structure prediction process, thus bypassing bad configurations that would otherwise be fruitless for optimization. This saves compute cycles and accelerates the entire process, so that more and larger problems can be attempted.

ProteinShop was put to the test in the Fifth Critical Assessment of Techniques for Structure Prediction (CASP5) competition held in 2002. The team included Elizabeth Eskow, Richard Byrd, and Bobby Schnabel, scientists from the University of Colorado, and Berkeley Lab's Teresa Head-Gordon and Silvia Crivelli. While many of the competing teams used information on known proteins stored in the Protein Data Bank, this team came up with secondary and tertiary protein structures starting with nothing more than amino acid sequences decoded from genes. The Colorado-Berkeley team developed a global optimization method that takes several weeks to converge using a supercomputer at the National Energy Research Scientific Computing Center (NERSC). This distributed memory IBM SP supercomputer, named "Seaborg," is a 6,656-processor system with a peak speed of 10 teraflop/s.

In the end, the team predicted 20 new or difficult protein folds ranging from 53 to 417 amino acids in length.

Aside from raw computing power, another way ProteinShop may speed up the discovery process is using concepts from robotics and animation, which allow researchers to re-configure structures without breaking them up. These “inverse kinematics” robotics concepts enable all the angles in the structure to move as jointed segments – in effect, simulating the way joints in our bodies interrelate. Whereas incorporating all the likely alpha helical regions and beta sheets used to consume days of computational time, this predictive process is now done in a matter of hours.

A fundamental concept in protein research is that every 3D structure of a protein is the one in which the free energy is at its lowest. In order to find the final structure of a protein, then, researchers must find this “global energy minimum” formation. Taking this into account, the program allows users to experiment among various potential conformations while simultaneously monitoring the energy profile of the protein.

Currently, the program designers are working to make ProteinShop more applicable and adaptable to different protein folding methodologies. If users could manipulate structures from a biological point of view, and then put them back in the queue for more optimization, the process of experimentation and discovery in protein research could be greatly enhanced. The group is also investigating the use of stereoscopic rendering and three-dimensional input devices to remove the limitations of a two-dimensional interface.

Clearly, protein-folding research will have far-reaching ramifications. It could lead to new insights about diseases ranging from Alzheimer’s to Cystic fibrosis, which scientists believe are caused by protein folding gone wrong. A better understanding of protein structures could also lead to the engineering of altogether new proteins, and shed light on how drugs bind proteins to alter their structure and function.

Above all, ProteinShop is an important tool that will help scientists unravel one of the most challenging problems that theoretical and computational chemistry has to offer.

The ProteinShop development was led by Silvia Crivelli of the Visualization Group in the Berkeley Lab’s Computational Research Division, and also includes Wes Bethel of Berkeley Lab, and UC Davis's Nelson Max, Bernd Hamann, and Oliver Kreylos. The group’s recent paper, "Interactive Protein Manipulation," won the Best Application Award at the IEEE Visualization 2003 Conference.

ProteinShop is now available for download from http://proteinshop.lbl.gov. While there is no cost to academic users, registration is required.