COMPUTATIONAL CHALLENGES

HIGH-PERFORMANCE COMPUTERS

YOKOHAMA, Japan—A number of problems in today's biomedical computing environment present significant computational challenges. The most obvious of these is the inability of conventional sequential computers to handle the computational rates required to deal with applications such as determining the three-dimensional structure of viruses from electron micrographs. The answer to such problems may well be found in the realm of high-performance parallel computing hardware and software.

Such was the thesis presented by Robert Martino of the division of computer research and technology (DCRT) at the National Institutes of Health (NIH, Bethesda, MD) at the Supercomputing Japan 92 conference held recently in Yokohama. While presenting his paper—entitled "Determining Biological Structure and Function Using Scalable Parallel Architectures"—Martino said that DCRT is "developing parallel algorithms for a number of biomedical applications that can benefit from computational speedup." These applications include image processing of electron micrographs, protein and nucleic acid sequence analysis, protein-folding prediction, and nuclear magnetic resonance spectroscopy.

Martino presented a number of cases illustrating the speed advantages of DCRT's system, an Intel Supercomputer Systems Division iPSC/860 parallel computer. The iPSC/860 is known as a multiple instruction stream/multiple data stream distributed memory system, with 128 nodes. Each node has its own processor and 16 megabytes of memory. Each processor can execute its own program and communicate with other nodes over a high-speed data pathway. The DCRT system also has an Intel concurrent I/O file system for fast-access mass storage.

In one example, Martino discussed the computationally intensive calculation of the solvent-accessible surface area of a protein. Such a calculation is key to determining the three-dimensional structure of a protein—so as to predict its function—when only its amino-acid sequence is known (the so-called protein-folding problem). At DCRT, Martino and his colleagues employed an approximation method to estimate a protein's solvent-accessible surface area. A protein-folding problem typically requires at least 200,000 time-consuming area calculations. Such a task is often beyond the limits of conventional computers. On the Intel iPSC/860, however, the computation was straightforward. Using a 64-processor configuration, the surface-area calculation for a protein with a 333-amino-acid sequence was performed over 30 times faster than with an IBM 3090 mainframe and more than 75 times faster than with a Convex C240 superminicomputer.

Martino pointed out that in implementing a protein database search algorithm—which is used by NIH's Human Genome Project for protein sequence analysis—a 64-processor iPSC/860 performed a similarity search of five protein sequences with a total length of 2,391 residues in four minutes. This compares to the 23 hours taken by a Sun workstation, using a sequential version of the search program. The parallel search ran 345 times faster than the sequential search.—**Stuart M. Dambrot**