

[\[Link to publication\]](#)

News Briefs from the ISMB Conference Floor

By Vivien Marx

Biosof and IBM Collaborate on Proteomics Hardware/Software Platform

Biosof, a Columbia University spin-out devoted to drug discovery, agricultural science, and homeland security, is collaborating with IBM to provide a server platform for PredictProtein, the brainchild of Columbia University researcher Burkhard Rost.

IBM and Columbia are also exploring the option of porting PredictProtein to IBM's Cell Broadband Engine.

The main goal of the project is to allow users to run PredictProtein jobs on IBM high-performance computing hardware via a virtual private network.

IBM's Janis Landry-Lane, program manager in the company's deep computing group, told BioInform that the university had set up a web portal for users to use PredictProtein that was running on a shared Linux cluster. PredictProtein has approximately 3,000 users a month.

The scientists essentially "wanted to be able to work on the algorithms and not be bothered with IT infrastructure," she said, noting that IBM is working with Biosof on the hardware environment as well as specific demands of some clients.

Landry-Lane said that IBM offered the university the opportunity to benchmark their application on different platforms and to develop it for maximum performance within its financial constraints. IBM will do testing on various Linux options, she said.

In order to address the concerns of commercial users such as pharmaceutical firms who might be wary of using the service over the Internet, she said the package will run on IBM's Deep Computing Capacity on Demand, the firm's virtual private network that can host services to run applications.

This plan also eliminates the need for Columbia to run a virtual private network while maintaining the hardware and software system, including the need to pay for power and cooling, she said. IBM will be providing power, space, cooling, and IT services to support Bi-Sof.

She added that the partners are also exploring whether IBM's Cell Broadband Engine can be used to accelerate components of the PredictProtein algorithm.

The Cell Broadband Engine, jointly developed by Sony, Toshiba and IBM, is a multiprocessor with high-performance features. It combines a two-way simultaneous multithreading PowerPC processor core with eight DSP-like SIMD processing units that can handle compute-intensive operations and a high bandwidth-memory subsystem.

It was initially developed for Sony's Playstation 3 game console, but it is being explored for other uses, such as bioinformatics. These processors are at the core of Roadrunner, the IBM machine that recently broke the petaflop barrier and tops the list of the world fastest supercomputers.

IBM scientists Kathy Tzeng and her colleagues have been exploring the use of computationally intensive applications with this processor for such applications as hidden Markov model-based protein profile searches. She and her team have modified HMMer, an implementation of an HMM-based protein profile search for the Cell Broadband Engine architecture.

For the Columbia collaboration, Tzeng is looking at the most computationally intensive and time-consuming functions of PredictProtein and exploring putting them on the Cell, said Landry-Lane.

"In an IBM Blade Center you can insert Cell blades and Intel or AMD blades as a single system," so some of the jobs can be routed to Intel or AMD blades while other could be routed to Cell. "Some of the exploratory work we are looking at for Biosof is, 'Can the Cell processor be used for parts of their algorithm to be accelerated?'" she said.

Preliminary studies have shown "considerable speed-ups" over other chips, Landry-Lane explained.

Rost told BioInform that he released the first ProteinProtein server 16 years ago at the European Molecular Biology Laboratory, "and wondered then about what we might need in 10 years." In that time frame he said, hundreds of thousands of users have used the tool.

PredictProtein offers algorithms to help to predict protein structure and function and to study such aspects as binding and active sites, subcellular localization, sequence motifs, fold recognition, secondary structure, and solvent accessibility.