

Toolset teams computers to design drugs

By Ted Smalley Bowen, Technology Research News

Computational grids provide the raw material for assembling temporary, virtual computers from sometimes far-flung resources connected to the Internet or private networks. They came about because researchers often require processing power, storage, and bandwidth far beyond the scope of their own systems.

This type of distributed computing, which can also include scientific instruments, makes the means to tackle complex applications available on an ad hoc basis, and allows researchers to draw on widely-dispersed stores of information.

The molecular modeling programs used to design drugs are especially data-hungry and computationally intensive applications. Designing a drug involves screening massive databases of molecules to identify pairs that can be combined, and figuring out the best way to combine them to achieve a certain affect. The molecules could be enzymes, protein receptors, DNA, or the drugs designed to act on them.

During this molecular docking process, researchers try to match the generally small molecules of prospective drugs with the larger biological molecules they are designed to affect, such as proteins or DNA. These searches can entail sifting through millions of files that contain three-dimensional representations of the molecules.

A group of researchers in Australia has put together a set of software tools to perform molecular docking over a computational grid. The tools tap into remote databases of chemical structures in order to carry out the molecular matching process.

Grid computing software finds and accesses resources from networked computers that can be physically located almost anywhere. It coordinates scheduling and security among systems that may be running different operating systems, to combine, for example, the processing capabilities of half a dozen Unix servers and a supercomputer with databases stored in a collection of disk drives connected to yet another computer.

The researchers adapted a molecular docking program to work on a grid configuration by having it run several copies of a molecular matching program on different systems or portions of systems. The software performed many computations at once on different subsets of the data, then combined the results. This type of parallel processing, also known as a parameter sweep, enabled the grid application to work through the matching process more quickly.

The complexity of each molecule record and the scale of the database searches involved in molecular docking put such applications beyond the reach of most labs' conventional computing resources, according to Rajkumar Buyya, a research scientist at Monash University in Australia. "Screening each compound, depending on structural complexity, can take hours on a standard PC, which means screening all compounds in a single database can take years."

Even on a supercomputer, "large-scale exploration is still limited by the availability of processing power," he said.

Using a computational grid, however, researchers could feed extensive computing jobs to a coordinated mix of PCs, workstations, multiprocessor systems and supercomputers, in order to crunch the numbers simultaneously.

A drug design problem that requires screening 180,000 compounds at three hours each would take a single PC about 61 years to process, and would tie-up a typical 64-node supercomputer for about a year, according to Buyya. "The problem can be solved with a large scale grid of hundreds of supercomputers in a day," he said.

To run the docking application on a computational grid, the researchers developed a program to index chemical databases, and software for accessing the chemical databases.

To speed the scheme, the researchers replicated the chemical database so that more requests for database information could be processed at once. To further speed the process, the researchers wrote a database server program that allowed computers to field more than one database query at a time.

The researcher's scheme compensates for the uneven bandwidth, processing speeds, and available resources among grid-linked systems by mapping the location of files and selecting the optimal computer to query, according to Buyya. "The data broker assists in the discovery and selection of a suitable source... depending on... availability, network proximity, load, and the access price," he said.

Because the performance of database applications suffers over network connections, the researchers generated indices for each chemical database, including references to each record's size.

January 16, 2002

[Page One](#)

[Morphing DNA makes motor](#)

[Toolset teams computers to design drugs](#)

[Atom clouds ease quantum computing](#)

[Web pages cluster by content type](#)

[Quantum effect alters device motion](#)

News:

[Research News Roundup](#)

[Research Watch blog](#)

Features:

[View from the High Ground Q&A](#)

[How It Works](#)

RSS Feeds:

[News XML](#) | [Blog XML](#) | [Books XML](#)

[What's New Log](#)

Ad links:

[Buy an ad link](#)

Advertisements:

HTS for
Gq GPCR

With IP-One,
don't miss
your hits
anymore



Read TRN's Making the Future reports:

- DNA Technology
- Robotics
- Internet Structure
- Internet Applications
- Pattern Recognition
- Nanomechanics/ Nanoelectronics
- Quantum Computing
- Self-Assembly
- Computer Displays
- Power Sources
- Security
- Biochips
- Alternative Chips
- Computer Interfaces
- Data Storage
- Quantum Cryptography

Get your road maps to the future now.



This allowed each computer to respond to queries by first checking the index file for the record's size and location and then accessing the record directly from the database file, rather than sequentially sifting through the database, said Buyya.

The application requirements and the tools used to meet them are specific to molecular docking, but similar software would speed compute-intensive tasks like high-energy physics calculations and risk analysis, according to Buyya.

The researchers tested the scheduling portion of their scheme on the World Wide Grid test-bed of systems in Australia, Japan and the US, and successfully estimated the time and cost required to run the applications in configurations optimized for speed and for budget, Buyya said.

Using the test bed, they screened files of 200 candidate molecules for docking with the target enzyme endothelin-converting enzyme (ECE), which is associated with low blood pressure.

The researchers' use of grid computing tools to automate molecular docking is "an excellent application of grid computing," said Julie Mitchell, an assistant principal research scientist at the San Diego Supercomputer Center. Features like "deadline- and budget-constrained scheduling should make the software very attractive to pharmaceutical companies" and to companies interested in such computationally demanding applications as risk analysis, scientific visualization and complex modeling said Mitchell. "There's nothing specific to molecular biology in their tools, and I imagine they could be applied quite readily in other areas."

The researchers also handled the process management aspects of adapting the applications to grids well, she added.

"The [researchers'] approach is obviously the way to go for those type of applications on the Computational Grid," said Henri Casanova, a research scientist in the computer science and engineering department of the University of California at San Diego. "The notion of providing remote access to small portions of domain-specific databases is clearly a good idea and fits the molecular docking applications," he said.

The economic concepts underlying the scheduling and costing of grid applications application are still immature, Casanova added. "The results concerning application execution are based on a Grid economy model and policies that are not yet in place. There are only vague notions of "Grid credit unit" in the community and the authors of the paper assume some arbitrary charging scheme for their experiments. This is an interesting avenue of research, but...there is very little in terms of Grid economy that is in place at the moment," he said.

The data access and computation techniques are technically ready to be used in practical applications today, according to Buyya.

Buyya's research colleagues were Jon Giddy, and David Abramson of Monash University in Australia and Kim Branson of the Walter and Eliza Hall Institute, in Australia. The research was funded by the Australian Cooperative Research Center for Enterprise Distributed Systems Technology (EDST), Monash University, the Walter and Eliza Hall Institute of Medical Research, the IEEE Computer Society, and Advanced Micro Devices Corp.

Timeline: Now

Funding: Corporate; Government

TRN Categories: Distributed Computing; Applied Computing; Supercomputing

Story Type: News

Related Elements: Technical paper, "The Virtual Laboratory: Enabling On-Demand Drug Design with the World Wide Grid," posted on the computer research repository (CoRR) at <http://xxx.lanl.gov/abs/cs.DC/0111047>.

Advertisements:

HTS for Gq GPCR

With IP-One, don't miss your hits anymore

Ad links: [Clear History](#)

[Buy an ad link](#)

[Home](#) [Archive](#) [Resources](#) [Feeds](#) [Offline Publications](#) [Glossary](#)

[TRN Finder](#) [Research Dir.](#) [Events Dir.](#) [Researchers](#) [Bookshelf](#)

[Contribute](#) [Under Development](#) [T-shirts etc.](#) [Classifieds](#)

[Forum](#) [Comments](#) [Feedback](#) [About TRN](#)

TRN [Newswire and Headline Feeds](#) for Web sites