

Medical University of South Carolina Applies Bioinformatics Theory with MathWorks Tools

Advances in molecular biology are producing unprecedented amounts of data. The data contains a wealth of information about the structure and dynamics of genes, proteins, and metabolites, but to extract and interpret that information, researchers must combine emerging theories with computational methods and mathematical models.

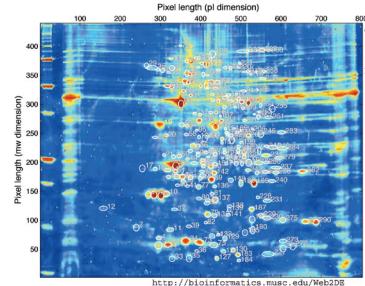
Researchers in the Department of Biostatistics, Bioinformatics, and Epidemiology at the Medical University of South Carolina (MUSC) use MathWorks tools to incorporate genomic and metabolic data into systems models for testing hypotheses. The university also uses MathWorks tools to support a number of the department's research initiatives in genomics, proteomics, and bioinformatics, such as systems biology, metabolism modeling and multivariate statistics.

"The drive for algorithm identification is rooted in biology, and so our biologists must develop fast-prototype applications," says Dr. Jonas Almeida, associate professor of Bioinformatics at MUSC. "MATLAB® provides a user-friendly environment that enables us go from theory to application quickly and painlessly."

THE CHALLENGE

The department's research team sought a robust mathematical modeling environment for solving systems of ordinary differential equations, image processing, statistical analysis, optimization, and sequence alignment.

The environment would need to integrate with their existing Web technologies, open-source software, and data from the National Center for Biotechnology Information (NCBI).



Two-dimensional gel analysis.

Because life scientists often lack experience with mathematical modeling software, the environment would also need to be easy to learn and use.

THE SOLUTION

After considering a number of commercial and open-source mathematical modeling tools, the group chose MATLAB for its ease of use, interoperability, industry acceptance, and rich modeling and computational capabilities.

Research teams within the group use MATLAB to develop applications for genomic and proteomic analysis such as biomarker identification, two-dimensional gel analysis, and artificial neural networks. They make these packages available to other groups and the scientific community via the Web using the MATLAB Web Server.

MATLAB and the Statistics and Optimization toolboxes provide the foundation for much of the group's work. "MATLAB is at the core of our ability to function. And everything we do uses the Statistics Toolbox. We also use the Optimization Toolbox for numerical decoupling by tracing state variables with neural networks in conjunction with genetic algorithms," says Almeida.

THE CHALLENGE

To advance computational biology research by enabling students and biologists to apply theory through mathematical modeling

THE SOLUTION

Use MathWorks tools to analyze experimental data, develop algorithms, and use existing open-source technologies

THE RESULTS

- Prestigious research grant
- High student productivity
- Shorter computation times

“MATLAB enables young biologists to learn enough programming and math without being afraid of the code. They can write in MATLAB as if it were English.”

Dr. Jonas Almeida,

Medical University of South Carolina

The group uses the Bioinformatics Toolbox to simplify sequence alignment using Needleman-Wunsch and Smith-Waterman algorithms. The toolbox also enables them to normalize, visualize, and import microarrays, including data from NCBI's Gene Expression Omnibus. They also use the SOAP client in MATLAB to interact with local and public data on the Internet.

MUSC uses the Wavelet and Image Processing toolboxes to denoise and identify clusters of proteins in two-dimensional gel samples.

In the area of biochemical system theory, researchers use the Symbolic Math Toolbox to numerically decouple and recast systems of nonlinear differential equations.

They also use MATLAB to make system calls to other open-source technologies, including the PostgreSQL database and the statistical package *R*, and parse the results.

“There is a strong bias in the bioinformatics community for open-source tools, but we really believe MATLAB is comparable to an open-source tool because the code we develop has an open architecture, so anyone can see the source code,” says Almeida. “I can't ever remember one of my reviewed papers not getting accepted because I used MATLAB.”

MUSC researchers plan to continue using MathWorks tools to develop applications for genomic, transcriptomic, and proteomic analysis.

APPLICATION AREAS

- Biotechnology, pharmaceutical, and medical
- Statistics and data analysis
- Image processing

PRODUCTS USED

- MATLAB
- MATLAB Web Server
- Statistics Toolbox
- Optimization Toolbox
- Bioinformatics Toolbox
- Wavelet Toolbox
- Image Processing Toolbox
- Symbolic Math Toolbox

THE RESULTS

- **Prestigious research grant.** MathWorks tools helped the department win grants, including one of only 15 Bioinformatics Training grants from the National Library of Medicine and one of only 10 National Heart, Lung, and Blood Institute Proteomic Center awards. “Graphics play a big role in winning grants,” says Almeida. “A color picture of a gel showing clusters of proteomic spots on the side makes a big impact. With MathWorks tools, we can produce that in minutes.”
- **High student productivity.** “Biology students can learn dynamic programming very easily with MATLAB,” Almeida explains. “We basically have one semester to train them, and with MATLAB that is enough.”
- **Shorter computation times.** Using MATLAB and a technique for decoupling systems of differential equations, the group accelerated the process of determining pathway structure from metabolic or proteomic time-series data. On a single machine using MATLAB, they completed the process in fewer than 15 minutes. Approaches, which use heavy parallelization on clusters of several hundred processors, take hours to produce results.

To learn more about the Medical University of South Carolina, visit bioinformatics.musc.edu or proteomics.musc.edu

www.mathworks.com