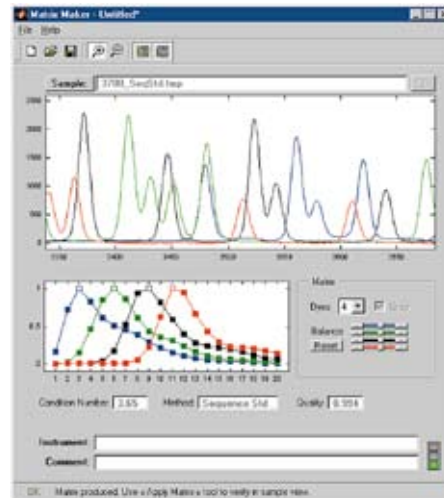


Applied Biosystems Develops a Crucial DNA Sequencing Algorithm in MATLAB

Genome sequencing reveals the genetic make-up of an organism by reading off the sequence of the four DNA bases (chemical compounds). This sequence encodes all the information necessary for the life and functioning of the organism.

The recent advances in human genome sequencing, including the International Human Genome Project, have been made possible by sophisticated new instruments, such as the ABI PRISM® series from Applied Biosystems Group. The ABI PRISM® 3700 DNA Analyzer is the first instrument developed specifically for large-scale projects such as human genome sequencing. It contains proprietary software developed with MATLAB® and MATLAB Compiler™.



A portion of the DNA dye-label spectral profile, which allows the researcher to read the sequence of bases in a selected strand of DNA.

THE CHALLENGE

To develop a robust yet flexible spectral calibration algorithm to be included in a high-throughput DNA analysis instrument

THE SOLUTION

Use MATLAB to test ideas and code a prototype, and then use the MATLAB Compiler tools to convert the algorithm to C/C++ code and test it at a component level

THE RESULTS

- Research options extended
- Calibration errors reliably detected
- Development completed in 2-3 weeks

THE CHALLENGE

Jim Labrenz, senior software developer at Applied Biosystems, had to develop a new spectral calibration algorithm for the 3700 DNA Analyzer. This algorithm would automatically compute the DNA dye-label spectral profiles, as “seen” by the optics of each sequencing instrument, in each capillary.

The 3700 DNA Analyzer uses capillary electrophoresis, a technology in which DNA fragments differing in length by one base are injected into a capillary and separated by size as they migrate through a polymer matrix under the influence of an electric field. Before injection, chemical reactions are performed on a target strand of DNA to create a set of fragments called a size ladder. Incrementally larger fragments ending in a particular base in the sequence can then be labeled with a specific dye. As the fragments

emerge from the capillary, the dye labels are excited by a laser and emit a fluorescent signature that is recorded by a detector.

The fluorescent emissions are recorded in four or more intensity signals, each corresponding to light collected in a particular wavelength range. Because the four dyes have overlapping color spectra, these signals must be deconvolved (separated out) to produce a four-channel chromatogram where each “color” indicates the presence or absence of a corresponding base in the target strand of DNA.

To ensure that the spectral deconvolution produces an accurate representation of the DNA sequence, the dyes’ spectral profiles must be predetermined for each capillary, on each 3700 DNA Analyzer.

“ Having one integrated package is a big advantage. Using MATLAB and MATLAB Compiler reduced my development time by a factor of 4 or 5. ”

Jim Labrenz, Applied Biosystems

Labrenz needed a calibration algorithm that could analyze dye mixtures in the raw-data signals and determine the spectral profiles automatically. This algorithm had to be robust and flexible enough to handle any future machine or dye modifications and to work with various instruments, including the older ABI PRISM® 310 and 377 models.

THE SOLUTION

MATLAB allowed Labrenz to reuse C/C++ code from other projects directly within MATLAB and to start experimenting almost immediately. He could then easily code a prototype to test the spectral calibration algorithm.

After optimizing and verifying his prototype, Labrenz used MATLAB Compiler to convert his application to C/C++ code. “Having one integrated package is a big advantage,” says Labrenz. “Using MATLAB and MATLAB Compiler reduced my development time by a factor of 4 or 5.”

Having generated portions of his MATLAB code as a MEX file, he used MATLAB Compiler’s MEX wrapper function to integrate his MATLAB code with other C/C++ code and test it at a component level. He generated output from the MATLAB command line, checked for coding errors, and then integrated the corrected code with the code that he was compiling. “Without MATLAB Compiler, the algorithm would have taken longer to implement and would have been hard to test,” says Labrenz.

MATLAB Compiler enabled Labrenz to build some code as a Windows NT DLL and to add APIs to call the calibration routines from a Java application.

Labrenz completed a user interface for version 3.7 of Applied Biosystems’ Sequencing

Analysis software. This interface drives the algorithms of the spectral calibration matrix and lets the researcher edit and view the results of the matrix. It was developed entirely in MATLAB and compiled with MATLAB Compiler.

“MATLAB has always been ideal as an algorithm prototyping tool,” Labrenz concludes, “but MATLAB Compiler adds a whole new dimension, allowing rapid delivery of sophisticated solutions.”

THE RESULTS

▪ **Research options extended.** The spectral calibration algorithm in the 3700 DNA Analyzer can be run as a standalone application or launched from another application. Researchers can calibrate directly from a data sample and do not require a special chemical kit.

▪ **Calibration errors reliably detected.** “When you edit the matrix, it applies that calibration back to raw data,” explains Labrenz. “A trained eye can then detect a bad matrix or calibration. This saves researchers time and expense.”

▪ **Development completed in 2–3 weeks.** MATLAB was essential to this time frame. “What speeded up the development was having everything there,” says Labrenz. “It was fast!”

To learn more about Applied Biosystems, visit www.appliedbiosystems.com

APPLICATION AREAS

- Algorithm development
- Biomedical engineering

PRODUCTS USED

- MATLAB®
- MATLAB Compiler™

www.mathworks.com



© 2008 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

9853v01 11/08