Proteomics LIMS: A caBIG™ Project, Year 1

Olga Tchuvatkina, M.S., Liat Shimoni, Ph.D., Michael F. Ochs, Ph.D. and Thomas Moloshok, Ph.D.

Bioinformatics Facility, Fox Chase Cancer Center, Philadelphia, PA 19111

Abstract:

protLIMS is a web-based proteomics laboratory information management system. In February, we released version 1 of protLIMS, completed to the year one goal: The web-interface provides for recording of the biological material, protein mixture preparation, 2D-PAGE, gel image files and spot and plug mapping. Associated files may be uploaded and retrieved through the web-interface to the file system. In year two, protLIMS will be extended to accommodate mass spectrometric analyses and protein identification.

Description:

To meet the needs of proteomics researchers, the Bioinformatics Facility at Fox Chase Cancer Center (FCCC) has initiated development of a proteomics laboratory information management system (protLIMS) under the National Cancer Institute's (NCI's) Cancer Biomedical Informatics Grid (caBIG™) initiative. This is a multi-year, open-source, development project in the Integrative Cancer Research working group. This poster reports the project status after completion of year one and the issues involved with caBIG™ development projects.

The central focus of proteomics is on proteins whether, they are being studied *in situ* or *in vitro*, singly or *en masse*, or in mixtures or in various levels of purification. protLIMS is being developed to target the needs of *in vitro* analyses, but with enough flexibility to allow for variability in workflow and analytical procedures.

The system is comprises an N-tier architecture with an Enterprise Java Beans business tier coupled to a SQL-compliant database backend and a web-based client interface. While experimental details are stored in the database, associated files, such as images and files exported by analytical equipment, are stored in a dedicated space in the server file system. Locations of these files are stored in the database for later access.

Initial workflows and procedures being accommodated are those used by the FCCC Proteomics Facility and by the Proteomics Facility of our caBIG "Adopter Site," the H. Lee Moffitt Cancer Center. These workflows begin with biological samples, the extract the proteins, separate the proteins by 2-dimensional polyacrylamide gel electrophoresis (2D-PAGE), and analyze and identify the resolved proteins by mass spectrometry. The two proteomics facilities take different approaches to this generalized workflow: protLIMS accommodates both workflows. In addition, protLIMS is customizable to allow accommodation of other workflows. Being an open-source software project, customization is possible by alteration of the program code itself when necessary.

The goal of year one was to develop protLIMS to an endpoint of gel-spot localization and excision. Each spot on the gel indicates the presence of a protein or protein isoform (multiple proteins or isoforms, if not fully resolved). Once excised from the gel, the gel "plugs" would be processed for further analysis. Capturing of the data and metadata for this later analysis is the goal of year two.

In February, FCCC released version 1 of protLIMS, the prototype, completed to year one goals. In accordance with the caBIG initiative, and in association with the Adopter site participants, system requirements and specifications were defined and deposited in the caBIG repository in the form of textual and UML documentation. In addition, common data elements (CDEs) were defined and incorporated into NCI's Cancer Data Standards Repository (caDSR) in collaboration with the caBIG Vocabularies and CDE working group. These CDEs will be reused in additional caBIG projects, beginning the establishment of links between different platforms, just as protLIMS incorporated available CDEs.

Availability:

ProtLIMS, version 1, and supporting documentation is available at:

http://bioinformatics.fccc.edu/software/caBIG/protLIMS/protlims.shtml