

197a Plenary: High Throughput Proteomics Informatics Challenges

Gordon A. Anderson, Ken J. Auberry, Gary R. Kiebel, Mathew E. Monroe, Elena S. Peterson, Eric F. Strittmatter, and Nikola Tolic

Proteomic analysis of biological samples produces large volumes of data from different mass spectrometer (MS) technologies. These datasets allow the identification of peptides and proteins as well as allowing quantitation of peptide and protein abundances. This research often requires hundreds to thousands of separate mass spectrometer experiments. These experiments include a liquid chromatographic (LC) separation step coupled to both MS and tandem MS experiments. Data analysis tools are then used to perform database searches to identify peptides from tandem MS dataset, while other tools interpret and extract detected masses from MS datasets, and assign peptide identifications to those detected masses using additional information from LC elution times. These complex multi-stage analyses require tracking of experimental conditions and sample pedigree. Additionally, quality control analysis is performed at several stages of the processing to insure instrument performance and sample preparation quality. In order to manage this large volume of data and metadata the Proteomics Research Information Storage and Management system (PRISM) was developed. This system performs the required data management tasks in addition to automation of the data process pipeline. PRISM converts the data produced by multiple disparate mass spectrometers into information about the proteins that are present in biological samples. It has been in operation for over three years in support of proteomics research at Pacific Northwest National Laboratory (PNNL), and has successfully managed information relating to thousands of experiments on behalf of our researchers. PRISM is tightly integrated into the laboratory processes and instrumentation and is highly automated in order to support high-throughput proteomics investigation while minimizing the time demand on the scientific and technical staff.

Conducting proteomics research at any significant level of throughput using mass spectrometry requires automated information management, as the volume of data is too large and the processing rates required are too rapid to be managed manually. PRISM was designed with flexibility and scalability in mind to readily accommodate future needs and expanded functionality, and has grown and evolved to meet increasingly sophisticated research needs.