Chapter XI
Management and Analysis of Mass Spectrometry Proteomics Data on the Grid

Mario Cannataro
University Magna Graecia of Catanzaro, Italy

Pietro Hiram Guzzi
University Magna Graecia of Catanzaro, Italy

Giuseppe Tradigo
University Magna Graecia of Catanzaro, Italy

Pierangelo Veltri
University Magna Graecia of Catanzaro, Italy

ABSTRACT

Recent advances in high throughput technologies analysing biological samples enabled the researchers to collect a huge amount of data. In particular, mass spectrometry-based proteomics uses the mass spectrometry to investigate proteins expressed in an organism or a cell. The manual inspection of spectra is unfeasible, so the need to introduce a set of algorithms, tools and platforms to manage and analyze them arises. Computational Proteomics regards the computational methods for analyzing spectra data in qualitative (i.e. peptide/protein identification in tandem mass spectrometry), and quantitative proteomics (i.e. protein expression in samples), as well as in biomarker discovery (i.e. the identification of a molecular signature of a disease directly from spectra). This chapter presents main standards, tools, and technologies for building scalable, reusable, and portable applications in this field. The chapter surveys available solutions for computational proteomics and includes a deep description of MS-Analyzer, a Grid-based software platform for the integrated management and analysis of spectra data. MS-Analyzer provides efficient spectra management through a specialized spectra database, and supports the semantic composition of pre-processing and data mining services to analyze spectra on the Grid.
INTRODUCTION

Proteomics is about the study of the proteins expressed in an organism or a cell. Computational Proteomics regards the computational methods, algorithms, databases, and methodologies used to manage, analyze and interpret the data produced in proteomics experiments (Cannataro, 2008).

The broad application of proteomics and the increasing resolution offered by technological platforms, especially in Mass Spectrometry-based high-throughput proteomics, make the analysis of proteomics experiments difficult and error prone without efficient algorithms and easy-to-use tools.

Mass Spectrometry-based proteomics (Aebersold and Mann, 2003) requires computational methods for analyzing data in qualitative (e.g. peptide/protein identification in tandem mass spectrometry) and in quantitative proteomics (e.g. protein expression in samples). Other important issues regard the way how spectra are cleaned, pre-processed, organized, and stored in an efficient way. The data heterogeneity introduced by several available platforms, as well as the need to conduct repeatable biomedical experiments on large populations (e.g. for outcome research), demand for suitable standards for the representation, storage, transmission, and sharing of proteomics data among different research centers.

Main research fields of computational proteomics are: (i) qualitative proteomics, i.e. the identification of proteins expressed in a cell (ii) quantitative proteomics, the determination of the abundance of protein expressed, and (iii) biomarker discovery, e.g. finding most discriminant features among spectra coming from two different populations (e.g. healthy samples versus diseased ones).

In summary, Computational Proteomics requires a combination of (i) recognized standards for data representation and exchange, (ii) efficient algorithms and tools constituting the building blocks of proteomics studies, (iii) computer science technologies enabling an easy composition of tools, sharing of data, and repeatability of experiments, (iv) large storage repositories for storing and sharing the ever increasing large amount of proteomics data, (v) high-performance implementations of data pre-processing, analysis and visualization algorithms.

After presenting requirements of Computational Proteomics applications, with a special focus on Mass Spectrometry-based Proteomics, the chapter describes main standards, tools, and technologies for building scalable, reusable, and portable applications in this field. Relevant distributed and Grid-based Computational Proteomics platforms are surveyed, including a description of MS-Analyzer, a Grid-based software platform for the integrated management and analysis of spectra data. MS-Analyzer provides efficient spectra management through a specialized spectra database, and supports the semantic composition of spectra pre-processing and data mining services to analyze spectra on the Grid.

BACKGROUND

Recent developments in high-throughput technologies have enabled the study of the cell on a large scale. The resulting scenario is characterized by the birth of new disciplines leading to the start of a new era, the so-called omics era. The suffix omics identifies a discipline that analyzes the relation of various biological compounds in their environment. For instance, proteomics is the study of the proteins expressed in a cell (Aebersold and Mann, 2003).

Mass Spectrometry (MS) has become a core technology in proteomics. The application of techniques based on mass spectrometry to analyze samples has a big impact in understanding both qualitative and quantitative distribution of proteins in a sample and finally to individuate potential biomarkers (Listgarten et al., 2005; Petricoin et al., 2002; Morris et al., 2005) .
Related Content

EEG Based Detection of Alcoholics: A Selective Review
[www.igi-global.com/article/eeg-based-detection-alcoholics/73694?camid=4v1a](www.igi-global.com/article/eeg-based-detection-alcoholics/73694?camid=4v1a)

Tools and Considerations to Develop the Blueprint for the Next Generation of Clinical Care Technology

Immersive Image Mining in Cardiology
[www.igi-global.com/chapter/immersive-image-mining-cardiology/26271?camid=4v1a](www.igi-global.com/chapter/immersive-image-mining-cardiology/26271?camid=4v1a)

Classification of Brain MR Images Using Corpus Callosum Shape Measurements
[www.igi-global.com/article/classification-of-brain-mr-images-using-corpus-callosum-shape-measurements/138227?camid=4v1a](www.igi-global.com/article/classification-of-brain-mr-images-using-corpus-callosum-shape-measurements/138227?camid=4v1a)