
Preface

With the rapid development of proteomic technologies in life sciences and in clinical applications, many bioinformatics methodologies, databases, and software tools have been developed to support comparative proteomics study. This volume aims to highlight the current status, challenges, open problems, and future trends in developing bioinformatics tools and resources for comparative proteomics research and to serve as a definitive source of reference providing both the breadth and depth needed on the subject of *Bioinformatics for Comparative Proteomics*.

The volume is structured to introduce three major areas of research methods: (1) basic bioinformatics frameworks related to comparative proteomics, (2) bioinformatics databases and tools for proteomics data analysis, and (3) integrated bioinformatics systems and approaches for studying comparative proteomics in the systems biology context.

Part I (Bioinformatics Framework for Comparative Proteomics) consists of seven chapters:

Chapter 1 presents a comprehensive review (with categorization and description) of major protein bioinformatics databases and resources that are relevant to comparative proteomics research.

Chapter 2 provides a practical guide to the comparative proteomics community for exploiting the knowledge captured from and the services provided in UniProt databases.

Chapter 3 introduces the InterPro protein classification system for automatic protein annotation and reviews the signature methods used in the InterPro database.

Chapter 4 introduces the Reactome Knowledgebase that provides an integrated view of the molecular details of human biological processes.

Chapter 5 introduces eFIP (extraction of Functional Impact of Phosphorylation), a Web-based text mining system that can aid scientists in quickly finding abstracts from literature related to the phosphorylation (including site and kinase), interactions, and functional aspects of a given protein.

Chapter 6 presents a tutorial for the Protein Ontology (PRO) Web resources to help researchers in their proteomic studies by providing key information about protein diversity in terms of evolutionary-related protein classes based on full-length sequence conservation and the various protein forms that arise from a gene along with the specific functional annotation.

Chapter 7 describes a method for the annotation of functional residues within experimentally uncharacterized proteins using position-specific site annotation rules derived from structural and experimental information.

Part II (Proteomic Bioinformatics) consists of ten chapters:

Chapter 8 describes how the detailed understanding of information value of mass spectrometry-based proteomics data can be elucidated by performing simulations using synthetic data.

Chapter 9 describes the concepts, prerequisites, and methods required to analyze a shotgun proteomics data set using a tandem mass spectrometry search engine.

Chapter 10 presents computational methods for quantification and comparison of peptides by label-free LC-MS analysis, including data preprocessing, multivariate statistical methods, and detection of differential protein expression.

Chapter 11 proposes an alternative to MS/MS spectrum identification by combining the uninterpreted MS/MS spectra from overlapping peptides and then determining the consensus identifications for sets of aligned MS/MS spectra.

Chapter 12 describes the Trans-Proteomic Pipeline, a freely available open-source software suite that provides uniform analysis of LC-MS/MS data from raw data to quantified sample proteins.

Chapter 13 provides an overview of a set of open-source software tools and steps involved in ELISA microarray data analysis.

Chapter 14 presents the state of the art on the Proteomics Databases and Repositories.

Chapter 15 is a brief guide to preparing both large- and small-scale protein interaction data for publication.

Chapter 16 demonstrates a new graphical user interface tool called PRIDE Converter, which greatly simplifies the submission of MS data to PRIDE database for submitted proteomics manuscripts.

Chapter 17 presents a method for describing a protein's posttranslational modifications by integrating the top-down and bottom-up MS data using the Protein Inference Engine.

Chapter 18 describes an integrated top-down and bottom-up approach facilitated by concurrent liquid chromatography-mass spectrometry analysis and fraction collection for comprehensive high-throughput intact protein profiling.

Part III (Comparative Proteomics in Systems Biology) consists of four chapters:

Chapter 19 gives an overview of the content and usage of the PhosphoPep database, which supports systems biology signaling research by providing interactive interrogation of MS-derived phosphorylation data from four different organisms.

Chapter 20 describes "omics" data integration to map a list of identified proteins to a common representation of the protein and uses the related structural, functional, genetic, and disease information for functional categorization and pathway mapping.

Chapter 21 describes a knowledge-based approach relying on existing metabolic pathway information and a direct data-driven approach for a metabolic pathway-centric integration of proteomics and metabolomics data.

Chapter 22 provides a detailed description of a method used to study temporal changes in the endoplasmic reticulum (ER) proteome of fibroblast cells exposed to ER stress agents (tunicamycin and thapsigargin).

This volume targets the readers who wish to learn about state-of-the-art bioinformatics databases and tools, novel computational methods and future trends in proteomics data analysis, and comparative proteomics in systems biology. The audience may range from graduate students embarking upon a research project, to practicing biologists working on proteomics and systems biology research, and to bioinformaticians developing advanced databases, analysis tools, and integrative systems. With its interdisciplinary nature, this volume is expected to find a broad audience in biotechnology and pharmaceutical companies and in various academic departments in biological and medical sciences (such as biochemistry, molecular biology, protein chemistry, and genomics) and computational sciences and engineering (such as bioinformatics and computational biology, computer science, and biomedical engineering).

We thank all the authors and coauthors who had contributed to this volume. We thank our series editor, Dr. John M. Walker, for reviewing all the chapter manuscripts and providing constructive comments. We also thank Dr. Winona C. Barker from Georgetown University for reviewing the manuscripts. We thank Dr. Qinghua Wu for proof reading the book draft. Finally, we would like to extend our thanks to David C. Casey and Anne Meagher of Springer US, Jeya Ruby and Ravi Amina of SPi for their help in the compilation of this book.

Newark, DE, USA

Cathy H. Wu and Chuming Chen