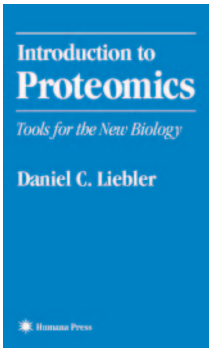

Introduction to Proteomics
Edited by Daniel C Liebler
Humana Press, Totowa, NJ, USA
(2002) 210 pp.

Introduction to proteomics: tools for the new biology

'...excellent overview of this field in an easy-to-read, succinct format.'

Gary Hardiman

Expert Rev. Proteomics 1(1), 9–10 (2004)

In the past decade, we have witnessed a technological revolution that culminated in the sequencing of the human genome and the dawning of the postgenomic era. The terms proteomics and proteome were coined by Marc Williams and colleagues in the early 1990s, and have since been adopted by the research community at large. The proteome refers to the full complement of proteins encoded by the genome. Biochemical studies of protein activity have historically focused on the analyses of single molecular species. The rapid pace of discovery of new gene products, via large-scale genomic and proteomic initiatives, has necessitated the development of alternative strategies to evaluate protein function. The challenge in recent years has been to develop high-throughput approaches to facilitate systematic protein analysis in biological samples, map functional interactions between proteins on a global scale, and place them in a biological context.

Daniel C Liebler's '*Introduction to Proteomics: Tools for the New Biology*' (2002) reviews the important tools and applications of proteomics, providing an excellent overview of this field in an easy-to-read, succinct format. The target audience is primarily newcomers to the field, graduate students embarking upon a research project, and practicing biologists refocusing their studies on the proteome. The author clarifies the differences between traditional low-throughput protein chemistry and the relatively new field of proteomics. Both are overlapping, complementary and involve protein identification. Proteomics, however, is grander, concerning itself with multiprotein systems where the focus is on the interplay of numerous distinct proteins, orchestrating their roles as part of larger systems or networks. Protein chemistry simply studies individual proteins, one protein or multisubunit protein complex at a time, focusing on their complete sequence analysis, modeling studies and structure determination.

Every organism possesses one genome and multiple proteomes. Cells express genes, which encode proteins with unique, cell-specific functions, in addition to proteins that carry out generic, yet essential functions. Approximately 200 different post-translational modifications have been reported, encompassing a wide variety of reversible and irreversible chemical reactions, which influence the diversity, affinity, function, cellular abundance and transport of proteins. The proteome comprises all of these forms, which makes its analysis incredibly complex, as any protein may be present in a cell, in any form, at any given time. Proteomes are constantly changing in response to environmental stimuli, chemicals and drug treatments, as well as growth and disease processes. Many of these changes hold considerable interest for researchers seeking to comprehend complicated disease pathologies, such as cancer. As the messenger RNA abundance in a cell often correlates poorly with the amount of protein synthesized, the proteome, rather than the transcriptome, holds promise for the identification of novel targets for therapeutic drugs.

The book is divided into three sections. The first considers the subject matter itself and describes its role in biomedical research. The second part introduces various proteomics tools and explains how they function. The final section explains how these tools are

 ISSN 1478-9450

applied in protein expression profiling, protein–protein interaction studies and the mapping of protein modifications. Despite the complexity of these topics, the chapters are all very straightforward, with illustrations to provide additional help to those unfamiliar with the subject matter. In the first section, the author highlights the fact that it is the proteome, rather than the genome, which confers biological complexity. Comparative genomics revealed that the fruit fly, *Drosophila melanogaster*, contains fewer numbers of genes and a smaller core proteome than the nematode worm, *Caenorhabditis elegans*, even though it is a more complex organism. Regulation of the genes and functions of the protein products may account for the greater complexity of the fly. This implies that the intricacy of the human genome lies in the diversity of the proteomes rather than in its actual size.

The proteome in any organism is a collection of between 30 and 80% of the possible gene products, many of which are expressed at relatively low levels. Regardless of absolute expression levels, most proteins exist in multiple forms owing to post-translational modifications. This fact, as the author notes, poses the greatest challenge to analytical proteomics, the essence of which is the conversion of proteins to peptides, followed by amino acid sequence determination and identification. In a given proteomic analysis, a typical human cell may contain approximately 20,000 different expressed proteins, with an average size of 50 kDa and contain average numbers of lysine and arginine residues, each of which would yield 30 peptides upon tryptic digestion. Thus, one cell could generate six million tryptic peptides, which would pose a formidable technical challenge to proteome analysis. Protein separation and digestion techniques are therefore needed to facilitate analytical proteomics. The author describes how proteins are extracted from biological samples and explains the principles behind techniques, such as 1- and 2D sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), preparative isoelectric focusing and high-performance liquid chromatography. Tandem LC techniques, such as multidimensional protein identification technique (MudPIT), are explained in detail. Subsequent chapters cover the working of matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) and electrospray ionization tandem mass spectrometry (ESI-MS/MS) instruments. As the author notes, however, the most flexible comprehensive strategy for proteome characterization is a combination of methods, or a hybrid approach. Proteins are first separated as intact species either by preparative isoelectric focusing, preparative 1D SDS-PAGE or high-performance liquid chromatography and the fractions obtained are subjected to high-performance liquid chromatography separations prior to introduction into the mass spectrometer.

Analysis of intact proteins, as the author states, is not realistic at the present time. It is very difficult to obtain mass measurements for large hydrophobic proteins, as the sensitivity is inferior to that for peptide mass measurements and peptide MS/MS analyses. Peptide fragments of six to 20 amino acids in length are ideal for mass spectrometry analysis and subsequent identification.

Despite excellent sections on analytical proteomics, one shortcoming of this book is the sparse amount of detail afforded to protein arrays in the final chapter. Protein arrays represent novel approaches to bridge the information gap between genomics and proteomics and offer the possibility of comprehensive analysis of proteomes, since they have the capacity to detect thousands of specific proteins in parallel. Many challenges exist in the generation of protein arrays and a chapter outlining these difficulties would have been welcome. The disadvantage of a publication of this nature is that it is not the best source of detailed information on any particular method. However, the book provides an excellent overview of the many facets of the proteome. The author succeeds in presenting the information in a manner that familiarizes the inexperienced reader with the important concepts and tools of proteomics.

Affiliation

- *Gary Hardiman*, PhD
  Department of Medicine, University of California, San Diego, La Jolla, CA 92093-0349, USA
  Tel.: +1 858 822 3792
  Fax: +1 858 822 3021
  ghardiman@ucsd.edu