
Foreword

A remarkable development in the post-genome era is the re-emergence of proteomics as a new discipline with roots in old-fashioned chemistry and biochemistry, but with new branches in genomics and informatics. The appeal of proteomics stems from the fact that proteins are the most functional component encoded for in the genome and thus represent a direct path to functionality. Proteomics emphasizes the global profiling of cells, tissues, and biological fluids, but there is a long road from applying various proteomics tools to the discovery, for example, of proteins that have clinical utility as disease markers or as therapeutic targets. Given the complexity of various cell and tissue proteomes and the challenges of identifying proteins of particular interest, informatics is central to all aspects of proteomics. However, protein informatics is still in its early stages, as is the entire field of proteomics.

Although collections of protein sequences have preceded genomic sequence databases by more than two decades, there is a substantial need for protein databases as basic protein information resources. There is a need for implementing algorithms, statistical methods, and computer applications that facilitate pattern recognition and biomarker discovery by integrating data from multiple sources. This book, which is dedicated to protein informatics, is intended to serve as a valuable resource for people interested in protein analysis, particularly in the context of biomedical studies. An expert group of authors has been assembled with proteomics informatics-related expertise that is highly valuable in guiding proteomic studies, particularly since currently the analysis of proteomics data is rather informal and largely dependent on the idiosyncrasies of the analyst.

Several chapters address the need for infrastructures for proteomic research and cover the status of public protein databases and interfaces. The creation of a national virtual knowledge environment and information management systems for proteomic research is timely and clearly addressed. Issues surrounding data standardization and integration are very well presented. They are captured in a chapter that describes ongoing initiatives within the Human Proteome Organization (HUPO). A major strength of the book is in the detailed review and discussion of applications of statistical and bioinformatic tools to data analysis and data mining. Much concern at the present time surrounds the analysis of proteomics data by mass spectrometry for a variety of applications. The book shines in its presentation in several chapters of various approaches and issues surrounding mass spectrometry data analysis.

Although the field of proteomics and related informatics is highly evolving, this book captures not only the current state-of-the-art but also presents a vision for where the field is heading. As a result, the contributions of the book and its component chapters will have long-lasting value.

Sam Hanash, M.D.
Fred Hutchinson Cancer Center
Seattle, Washington