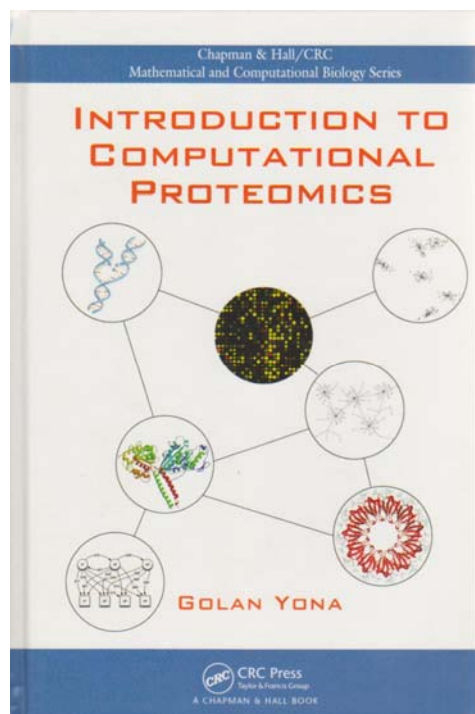


GOLAN YONA

INTRODUCTION TO COMPUTATIONAL PROTEOMICS



Chapman & Hall/CRC
Taylor and Francis Group
ISBN: 978-1-58488-555-9
Hardcover
767 pages
2011

Introduction to Computational Proteomics introduces the field of computational biology through a focused approach that tackles the different steps and problems involved with protein analysis, classification, and meta-organization. The book starts with the analysis of individual entities and works its way through the analysis of more complex entities, from protein families to interactions, cellular pathways, and gene networks.

The first part of the book presents methods for identifying the building blocks of the protein space, such as motifs and domains. It also describes algorithms for assessing similarity between proteins based on sequence and structure analysis as well as mathematical models, such as hidden Markov models and support vector machines, that are used to represent protein families and classify new instances.

The second part covers methods that investigate higher order structure in the protein space through the application of unsupervised learning algorithms, such as clustering and embedding. The book also explores the broader context of proteins. It discusses methods for analyzing gene expression data, predicting protein-protein interactions, elucidating cellular pathways, and reconstructing gene networks.

This book provides a coherent and thorough introduction to proteome analysis. It offers rigorous, formal descriptions, along with detailed algorithmic solutions and models. Each chapter includes problem sets from courses taught by the author at Cornell University and the Technion. Software downloads, data sets, and other material are available at biozon.org.

Table of Contents

Chapter 1.	The Basics	1
1.	What is Computational Proteomics?	3
2.	Basic Notions in Molecular Biology	9
3.	Sequence Comparison	23
4.	Multiple Sequence Alignment, Profiles, and Partial Order Graphs	105
5.	Motif Discovery	155
6.	Markov Models of Protein Families	183
7.	Classifiers and Kernels	235
8.	Protein Structure Analysis	291
9.	Protein Domains	345



Chapter 2. Putting All the Pieces Together	395
10. Clustering and Classification	397
11. Embedding Algorithms and Vectorial Representations	459
12. Analysis of Gene Expression Data	505
13. Protein-protein Interactions	553
14. Cellular Pathways	615
15. Learning Gene Networks with Bayesian Networks	649
References	687
Conference Abbreviations	735
Acronyms	737
Index	739