BRIEFINGS IN BIOINFORMATICS. VOL 9. NO 1. 91–93
Advance Access publication October 26, 2007
doi:10.1093/bib/bbm049

Book Review

Bioinformatics—From Genomes to Therapies.

Edited by Thomas Lengauer, Wiley-VCH Verlag GmbH & Co KGaA, Weinheim; 2007; ISBN: 978-3-527-312788; Hardcover; 3 Volume Set; 1814 pp.; $625.00.

'Bioinformatics—From Genomes to Therapies', edited by Thomas Lengauer, is big collection of contributions that includes the work of 91 authors arranged in 45 chapters, 11 parts and three volumes of 1814 pages. As an expanded sequel of the author's previous publication, 'Bioinformatics—From Genomes to Drugs', the book includes diverse topics in bioinformatics focusing on the challenges in understanding, diagnosing and curing of diseases.

Remembering aforementioned focus on the design, development and therapies of drugs for diseases, the three volumes are organized to help understand the building blocks of sequences and structures (Volume 1), the inner workings of molecular interactions (Volume 2) and the integrated picture of molecular function (Volume 3), respectively. Each volume in turn consists of several parts, each of which consists of several chapters under the same umbrella of a specific topic. References are provided at the end of each chapter giving both beginners and more experienced researchers additional information on the chapter. In addition, there is a global index that includes an extensive list of definitions, concepts and methods in bioinformatics addressed throughout the book. To sum up, Volume 1 is composed of 15 chapters in four parts: Introduction (Part 1), Sequencing Genomes (Part 2), Sequence Analysis (Part 3) and Molecular Structure Prediction (Part 4). Volume 2 is composed of 13 chapters in three parts: Analysis of Molecular Interactions (Part 5), Molecular Networks (Part 6) and Analysis of Expression Data (Part 7). Volume 3 is composed of 17 chapters in four parts: Protein Function Prediction (Part 8), Comparative Genomics and Evolution of Genomes (Part 9), Basic Bioinformatics Technologies (Part 10) and Outlook (Part 11).

The book begins with Part 1 consisting of an introductory chapter (Chapter 1) that explains the molecular basics of a disease and molecular approach to its cure (including protein targeting, genomics and proteomics, information on genes and proteins, drug development, therapy optimization). Chapter 1 also contains a nice summary on the organization of the book, from which this review is written.

Part 2 consists of a single chapter (Chapter 2) that discusses the assembly strategies for genome sequencing with algorithmic issues followed by examples of some of the existing assemblers.

Part 3 comprises Chapters 3–8 on sequence analysis. Chapter 3 includes the pairwise and multiple sequence alignment techniques and the sequence search in databases. Chapter 4 explains the phylogeny reconstruction approaches for ancestral inference by the tree reconstruction from the leaves, finding the optimal tree and then generating split networks from the trees. Chapters 5–7 start with introductory material followed by common methodologies for finding the protein-coding genes, for analyzing the regulatory regions and for finding the repeats in genomes, respectively. Chapter 8 ends Part 3 touching on the analysis of genome rearrangements with various methods (e.g. distance-based, maximum parsimony, maximum likelihood).

Part 4 deals with the molecular structure prediction in Chapter 9 through 15. Chapter 9 explains the hard task of three-dimensional protein structure prediction and describes simplified features of the task including the secondary structures, transmembrane regions, solvent accessibility, inter-residue contacts, flexible and intrinsically disordered regions and protein domains. Chapter 10 covers the homology modeling in biology and medicine (i.e. structure prediction based on the sequence alignment with template proteins) provided with modeling methods and results, followed by Chapter 11 that deals with the protein fold recognition task based on distant homologs, using a variety of

computational models (e.g. Hidden Markov Models, Support Vector Machines). Chapter 12 explores the *De Novo* structure prediction method (as a method that does not rely on homology between the query and the prototype sequences), and its application in systems biology. Chapter 13 is an introductory chapter on structural genomics that attempts to predict protein structures over whole proteomes and combines the experimental as well as homology-modeling based approaches. Chapters 14 and 15 conclude Part 4 by discussing the secondary and tertiary structure prediction of RNA, respectively.

In Part 5, molecular interactions are discussed in Chapter 16 through 19, focusing on the interactions between a drug and its target protein for drug therapy. Chapter 16 explores the docking and scoring methods for drug design, explaining how to dock ligands into protein-binding sites and how to assemble new ligands inside the binding site. Chapter 17 covers a variety of approaches to protein–protein and protein–DNA docking (e.g. Correlation, Monte Carlo techniques). Chapters 18 and 19 discuss lead identification and optimization (for drug discovery) using virtual screening and combination of relevant features (or biological variables), with the aid of data mining and automated decision-making models.

Part 6 covers molecular networks in four chapters, in order to understand the interactions between diseases and their therapies. Chapter 20 exhibits various approaches for modeling and simulating metabolic networks statically as well as dynamically, considering the red bold cell metabolism as an example. Chapter 21 introduces the motivation and requirements of gene regulatory networks and discusses methods for the inference of the networks. Chapter 22 explains cell signaling networks and the methods for their analysis (e.g. Boolean, Bayesian networks). Chapter 23 concludes Part 6 by bringing about a medical application, the dynamics of interaction between the viruses and the host cells.

Part 7 comprises Chapter 24 through 28, focusing on the analysis of expression data. Chapter 24 contains an overview on the technologies and applications in DNA microarrays. Chapter 25 explains the low-level details of processes in micro-array experiments. Chapters 26 and 27 touch on the classification of patients and genes, respectively, and exhibit a wide spectrum of classification and clustering algorithms that have been developed and used commonly in machine learning and data mining research communities. Chapter 28 introduces a relatively new area, *Proteomics*, beyond the genome analysis to deal with the diverse and dynamic characteristics of proteins.

Part 8 discusses the prediction of protein functions in eight chapters. Chapter 29 introduces diverse ontologies in molecular biology that need to be clarified prior to the prediction. Chapter 30 explains techniques for inferring protein functions from sequences and describes representative databases, as well. Chapter 31 explores various approaches to the analysis of protein interaction networks, based on different types of data (e.g. sequences, structures, textual information). Chapters 32 and 33 again focus on the inference of protein functions, based on genomic context (i.e. genomes, genes, gene arrangements) and protein structures, respectively. Chapter 34 deals with text mining on protein functions such as information retrieval and extraction, question answering and natural language generation. Chapter 35 discusses integration of different data such as features, classifiers and networks for protein function prediction. Chapter 36 finally discusses methods for predicting druggability of proteins in drug design considering the suitability of their shapes in binding.

Part 9 comprises Chapter 37 through 41 and discusses the comparative genomics and the evolution of genomes, focusing on the analysis of relationships and differences between genomes. Chapter 37 provides the definition, motivation and technologies of comparative genomics in a bid to figure out what we can learn from different genomes of individuals of the same species. Chapter 38 deals with the association studies of diseases to understand the genetic differences among people on the susceptibility of the disease, and provides related statistical methods. Chapter 39 starts with introductory material on pharmacogenetics and pharmacogenomics for personalized drugs, and provides associated data and tools. Chapter 40 focuses on the evolution of drug resistance in HIV, providing introductory material as well as techniques, resources and issues in the domain. Chapter 41 is another chapter focusing on a specific task, analysis of the evolution of infectious bacteria.

Part 10 addresses some of the basic technologies in bioinformatics in three chapters. Chapter 42 touches on the issue of integrating diverse biological databases and explains the data models, integration methods and implementation technologies. Chapter 43 handles the data visualization, showing many methods for different types of data. Chapter 44 concludes

Part 10 by delivering the issues and technologies of distributed computing in bioinformatics.

The last part (Part 11) is an outlook in bioinformatics. It includes a concluding chapter (Chapter 45) that summarizes the field and enumerates future research areas in bioinformatics and computational biology.

With the diverse but integrated contents provided for the molecular analysis of diseases and their therapies, the book can be a valuable asset to the researchers from the computer science, mathematics, statistics and biological sciences, in performing interdisciplinary research in bioinformatics in general and in pharmaceutics and molecular medicine in particular. Especially for computer scientists, the book provides relevant knowledge and addresses research issues in the biological domain, which makes it possible for them to accomplish significant achievements based on their computational and algorithmic backgrounds in computer science.

Jihoon Yang
Department of Computer Science, Sogang University,
1 Shinsoo-Dong, Mapo-Ku, Seoul 121-742, Korea
E-mail: yangjh@sogang.ac.kr