Advanced Bioinformatics and Computational Biology

Contributors: Katarzyna Hupert-Kocurek, Jon M. Kaguni, et al.

About the Book

Bioinformatics is the application of computer technology to the management of biological information. Computers are used to gather, store, analyze and integrate biological and genetic information which can then be applied to gene-based drug discovery and development. The need for Bioinformatics capabilities has been precipitated by the explosion of publicly available genomic information resulting from the Human Genome Project. Computational biology has soared from being an auxiliary discipline to being a crucial element for progress in practically all aspects of the biological sciences. The field of bioinformatics and computational biology arose due to the need to apply techniques from computer science, statistics, informatics, and applied mathematics to solve biological problems. Scientists have been trying to study biology at a molecular level using techniques derived from biochemistry, biophysics, and genetics. Progress has greatly accelerated with the discovery of fast and inexpensive automated DNA sequencing techniques. As the genomes of more and more organisms are sequenced and assembled, scientists are discovering many useful facts by tracing the evolution of organisms by measuring changes in their DNA, rather than through physical characteristics alone. This has led to rapid growth in the related fields of phylogenetics, the study of evolutionary relatedness among various groups of organisms, and comparative genomics.

Advanced Bioinformatics and Computational Biology attempts to connect different sub-fields of bioinformatics, as well as those that examine the history of bioinformatics and computational biology. This book presents major developments and trends in bioinformatics and its applications. It is a compilation of original research articles as well as review articles in all areas of bioinformatics and computational biology. The discipline of bioinformatics and computational biology sits at the interface between biology, the quantitative sciences (such as mathematics, statistics, computer science) and engineering, and spans levels of investigation from single genes to systems. It considers significant computational biology advances of the last decade, and reflect on some key challenges ahead. In addition, since computational biology has become closely tied to experimental research, progress is not purely computational; it is tied to experiment. And that's as it should be. Ten years ago, computational biology was not entirely trusted by experimental biologists. By contrast, today computational biology is integrated in the community. It's easier for computational biologists to collaborate across disciplines. Laboratory scientists have a better understanding of the merit of computational models for hypothesis generation as well as the need to iterate between modeling and laboratory testing.

KORO PRESS
www.korospress.com

ISBN 978-1-78569-333-5
9 781785 693335 00000

Kocurek,
Kaguni

Advanced Bioinformatics and Computational Biology

Advanced Bioinformatics and Computational Biology

Contributors

Katarzyna Hupert-Kocurek, Jon M. Kaguni, et al.

KOROS PRESS LIMITED

Advanced Bioinformatics and Computational Biology

Contributors: Katarzyna Hupert-Kocurek, Jon M. Kaguni, et al.

Published by Koros Press Limited

www.korospress.com

United Kingdom

Copyright 2017

The information in this book has been obtained from highly regarded resources. The copyrights for individual articles remain with the authors, as indicated. All chapters are distributed under the terms of the Creative Commons Attribution License, which permit unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Notice

Contributors, whose names have been given on the book cover, are not associated with the Publisher. The editors and the Publisher have attempted to trace the copyright holders of all material reproduced in this publication and apologise to copyright holders if permission has not been obtained. If any copyright holder has not been acknowledged, please write to us so we may rectify.

Reasonable efforts have been made to publish reliable data. The views articulated in the chapters are those of the individual contributors, and not necessarily those of the editors or the Publisher. Editors and/or the Publisher are not responsible for the accuracy of the information in the published chapters or consequences from their use. The Publisher accepts no responsibility for any damage or grievance to individual(s) or property arising out of the use of any material(s), instruction(s), methods or thoughts in the book.

Advanced Bioinformatics and Computational Biology

ISBN: 978-1-78569-333-5

British Library Cataloguing in Publication Data
A CIP record for this book is available from the British Library

Printed in the United Kingdom

此为试读，需要完整PDF请访问：www.ertongbook.com

Advanced Bioinformatics and Computational Biology

List of Contributors

Katarzyna Hupert-Kocurek
University of Silesia, Poland

Jon M. Kaguni
Michigan State University, United States of America

Kuo-Yuan Hwa
Institute of Organic & Polymeric MaterialsDepartment of Molecular Science and Engineering, Taiwan, ROC
Centre for Biomedical IndustriesNational Taipei University of Technology, Taipei, , Taiwan, ROC

Wan-Man Lin
Institute of Organic & Polymeric MaterialsDepartment of Molecular Science and Engineering, Taiwan, ROC
Centre for Biomedical IndustriesNational Taipei University of Technology, Taipei, , Taiwan, ROC

Boopathi Subramani
Institute of Organic & Polymeric MaterialsDepartment of Molecular Science and Engineering, Taiwan, ROC
Centre for Biomedical IndustriesNational Taipei University of Technology, Taipei, , Taiwan, ROC

Rosa M. Vitale
Istituto di Chimica Biomolecolare, CNR, Pozzuoli , Italy

Giovanni Renzone
Laboratorio di Proteomica e Spettrometria di Massa, ISPAAM, CNR, Naples, Italy

Andrea Scaloni
Laboratorio di Proteomica e Spettrometria di Massa, ISPAAM, CNR, Naples, Italy

Pietro Amodeo
Istituto di Chimica Biomolecolare, CNR, Pozzuoli , Italy

Mauricio Salcedo
Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, , México DF

Sergio Juarez-Mendez
Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, , México DF

Vanessa Villegas-Ruiz
Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, , México DF

Hugo Arreola
Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, , México DF

Oscar Perez
Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, Laboratorio de Oncología Experimental, Instituto Nacional de Pediatría, SS, , México

Guillermo Gómez
Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, Centro Nacional de Clínica de Displasias, Unidad de Oncología, Hospital Servicio de Oncología, Hospital General de México, SS. , México DF

Edgar Roman-Bassaure
Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, General de México, SS., México DF

Pablo Romero
Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, , México DF

Raúl Peralta
Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, Laboratorio de Oncología Genómica, Unidad de Investigación en Enfermedades Oncológicas, Hospital de Oncología, CMN-SXXI, IMSS, , México DF

Ryusuke Sawada
Nagoya University, Japan

Shigeki Mitaku
Nagoya University, Japan

Chia-Han Chu
National Tsing Hua University, , Taiwan

Chun Yuan Lin
Chang Gung University, , Taiwan

Cheng-Wen Chang
National Tsing Hua University, , Taiwan

Chihan Lee
Chipboud Technology Corporation, , Taiwan

Chuan Yi Tang
Providence University, Taiwan

Kaiser Jamil
Head, Genetics Department, Bhagwan Mahavir Medical Research Centre, Mahavir Marg, Hyderabad, & Dean, School of Life Sciences (SLS), Jawaharlal Nehru Institute of Advanced Studies, (JNIAS), India

M. Sabeena
Research Associate, Centre for Biotechnology and Bioinformatics- (CBB), Jawaharlal Nehru Institute of Advanced Studies, (JNIAS), Secunderabad, India

David Piedra
IBMB-CSIC, Spain

Marco d'Abramo
CNIO, Spain

Xavier de la Cruz
IBMB-CSIC, Spain
ICREA, Spain

Vikrant Nain
Gautam Buddha University, Greater Noida, India

Shakti Sahi
Gautam Buddha University, Greater Noida, India

Polumetla Ananda Kumar
National Research Centre on Plant Biotechnology, New Delhi, India

Lígia R. Rodriguesand
IBB – Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Campus de Gualtar, Braga, Portugal

Leon D. Kluskens
IBB – Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Campus de Gualtar, Braga, Portugal

Aoife Lowery
Department of Surgery, National University of Ireland Galway, , Ireland
John Van Geest Cancer Research Centre, School of Science & Technology, Nottingham Trent University, Nottingham, UK

Christophe Lemetre
Department of Surgery, National University of Ireland Galway, , Ireland
John Van Geest Cancer Research Centre, School of Science & Technology, Nottingham Trent University, Nottingham, UK

Graham Ball
Department of Surgery, National University of Ireland Galway, , Ireland
John Van Geest Cancer Research Centre, School of Science & Technology, Nottingham Trent University, Nottingham, UK

Michael Kerin
Department of Surgery, National University of Ireland Galway, , Ireland
John Van Geest Cancer Research Centre, School of Science & Technology, Nottingham Trent University, Nottingham, UK

此为试读,需要完整PDF请访问: www.ertongbook.com

Preface

Bioinformatics is the application of computer technology to the management of biological information. Computers are used to gather, store, analyze and integrate biological and genetic information which can then be applied to gene-based drug discovery and development. The need for Bioinformatics capabilities has been precipitated by the explosion of publicly available genomic information resulting from the Human Genome Project. Computational biology has soared from being an auxiliary discipline to being a crucial element for progress in practically all aspects of the biological sciences. The field of bioinformatics and computational biology arose due to the need to apply techniques from computer science, statistics, informatics, and applied mathematics to solve biological problems. Scientists have been trying to study biology at a molecular level using techniques derived from biochemistry, biophysics, and genetics. Progress has greatly accelerated with the discovery of fast and inexpensive automated DNA sequencing techniques. As the genomes of more and more organisms are sequenced and assembled, scientists are discovering many useful facts by tracing the evolution of organisms by measuring changes in their DNA, rather than through physical characteristics alone. This has led to rapid growth in the related fields of phylogenetics, the study of evolutionary relatedness among various groups of organisms, and comparative genomics.

Advanced Bioinformatics and Computational Biology attempts to connect different sub-fields of bioinformatics, as well as those that examine the history of bioinformatics and computational biology. This book presents major developments and trends in bioinformatics and its applications. It is a compilation of original research articles as well as review articles in all areas of bioinformatics and computational biology. The discipline of bioinformatics and computational biology sits at the interface between biology, the quantitative sciences (such as mathematics, statistics, computer science) and engineering, and spans levels of investigation from single genes to systems. It considers significant computational biology advances of the last decade, and reflect on some key challenges ahead. In addition, since computational biology has become closely tied to experimental research, progress is not purely computational; it is tied to experiment. And that's as it should be. Ten years ago, computational biology was not entirely trusted by experimental biologists. By contrast, today computational biology is integrated in the community. It's easier for computational biologists to collaborate across disciplines. Laboratory scientists have a better understanding of the merit of computational models for hypothesis generation as well as the need to iterate between modeling and laboratory testing.

Contents

List of Contributors ..... *vii*

Preface ..... *xi*

Chapter 1 **Understanding Protein Function - The Disparity Between Bioinformatics and Molecular Methods** **1**
Katarzyna Hupert-Kocurek And Jon M. Kaguni

Chapter 2 **Introduction to Biological Databases** **15**

Chapter 3 **In Silico Analysis of Golgi Glycosyltransferases: A Case Study on the Large-Like Protein Family** **27**
Kuo-Yuan Hwa, Wan-Man Lin and Boopathi Subramani

Chapter 4 **Concepts of Bioinformatics** **41**

Chapter 5 **Computational Methods in Mass Spectrometry-Based Protein 3d Studies** **55**
Rosa M. Vitale, Giovanni Renzone, Andrea Scaloni and Pietro Amodeo

Chapter 6 **Pairwise Sequence Alignment** **73**

Chapter 7 **Functional Analysis of the Cervical Carcinoma Transcriptome: Networks and New Genes Associated to Cancer** **85**
Mauricio Salcedo, Sergio Juarez-Mendez, Vanessa Villegas-Ruiz, Hugo Arreola, Oscar Perez, Guillermo Gómez, Edgar Roman-Bassaure, Pablo Romero And Raúl Peralta

Chapter 8 **Number Distribution of Transmembrane Helices in Prokaryote Genomes** **101**
Ryusuke Sawada and Shigeki Mitaku

Chapter 9 **Classifying Tim Barrel Protein Domain Structure by an Alignment Approach Using Best Hit Strategy and Psi-Blast** **107**
Chia-Han Chu, Chun Yuan Lin, Cheng-Wen Chang, Chihan Lee and Chuan Yi Tang

Chapter 10 **Identification of Functional Diversity in the Enolase Superfamily Proteins** **129**
Kaiser Jamil and M. Sabeena

Chapter 11 **Contributions of Structure Comparison Methods to the Protein Structure Prediction Field** **141**
David Piedra, Marco D'abramo and Xavier De La Cruz

Chapter 12 **In Silico Identification of Regulatory Elements in Promoters** **153**
Vikrant Nain, Shakti Sahi and Polumetla Ananda Kumar

Chapter 13 **Synthetic Biology & Bioinformatics Prospects in the Cancer Arena** **169**
Lígia R. Rodriguesand and Leon D. Kluskens

Chapter 14 **Microarray Technology - Expression Profiling of Mrna and Microrna in Breast Cancer** **189**
Aoife Lowery, Christophe Lemetre, Graham Ball and Michael Kerin

Citations **215**

Index **217**

1

UNDERSTANDING PROTEIN FUNCTION - THE DISPARITY BETWEEN BIOINFORMATICS AND MOLECULAR METHODS

Katarzyna Hupert-Kocurek[1] and Jon M. Kaguni[2]

[1] University of Silesia, Poland

[2] Michigan State University, United States of America

INTRODUCTION

Bioinformatics has its origins in the development of DNA sequencing methods by Alan Maxam and Walter Gilbert (Maxam and Gilbert, 1977), and by Frederick Sanger and coworkers (Sanger et al., 1977). By entirely different approaches, the first genomes determined at the nucleotide sequence level were that of bacteriophage ϕX174, and the recombinant plasmid named pBR322 composed of about 5,400 (Sanger et al., 1977), or 4,400 base pairs (Sutcliffe, 1979), respectively. In contrast, two articles that appeared in February 2001 reported on the preliminary DNA sequence of the human genome, which corresponds to 3 billion nucleotides of DNA sequence information (Lander et al., 2001; Venter et al., 2001). Only two years later, the GenBank sequence database contained more than 29.3 billion nucleotide bases in greater than 23 million sequences. With the development of new technologies, experts predict that the cost to sequence an individual's DNA will be about $1000. This reduction in cost suggests that efforts in the area of comparative genomics will increase substantially, leading to an enormous database that vastly exceeds the existing one.

By way of comparative genomics approaches, computational methods have led to the identification of homologous genes shared among species, and their classification into superfamilies based on amino acid sequence similarity. In combination with their evolutionary relatedness, superfamily members have been clustered into clades. In addition, high throughput sequencing of small RNAs and bioinformatics analyses have contributed to the identification of regions between genes that can code small RNAs (siRNA, microRNA, and long noncoding RNA), which act during the development of an organism to modulate gene expression at the post-transcriptional level (Fire et al., 1998; Hamilton and Baulcombe, 1999) reviewed in Elbashir et al., 2001; Ghildiyal and Zamore, 2009; Christensen et al., 2010). An emerging area is functional genomics whereby gene function is deduced using large-scale methods by identifying the involvement of specific genes in metabolic pathways. More recently, phenotype microarray methods have been used to correlate the functions of genes of microbes with cell phenotypes under a variety of growth conditions (Bochner, 2009).

These methods contrast with the traditional approach of mapping a gene via the phenotype of a mutation, and deducing the function of the gene product based on its biochemical analysis in concert with physiological studies. Such studies have been performed to confirm the functional importance of conserved residues shared by superfamily members, and also to determine the role of specific residues for a given protein. In comparison, comparative genomics methods are unable to distinguish if a nonconserved amino acid among superfamily members is functionally important, or simply reflects sequence divergence due to the absence of selection during evolution. Without functional information, it is not possible to determine if a nonconserved amino acid is important.

BIOINFORMATICS ANALYSIS OF AAA+ PROTEINS

On the basis of bioinformatics analysis, P-loop nucleotide hydrolases compose a very large group of proteins that use an amino acid motif named the phosphate binding loop (P-loop) to hydrolyze the phosphate ester bonds of nucleotides. A positively charged group in the side chain of an amino acid (often lysine) in the P-loop promotes nucleotide hydrolysis by interacting with the phosphate of the bound nucleotide. Additional bioinformatics analysis of this group of proteins led to a category of nucleotidases containing the Walker A and B motifs, as well as additional motifs shared by the AAA (ATPases Associated with diverse cellular Activities) superfamily (Beyer, 1997; Swaffield and Purugganan, 1997). These diverse activities include protein unfolding and degradation, vesicle transport and membrane fusion, transcription and DNA replication. The additional motifs of the AAA superfamily differentiate its members from the larger set of P-loop nucleotidases. Neuwald *et al.*, and Iyer *et al.* then integrated structural information with bioinformatics analysis to classify members of the AAA+ superfamily into clades (Neuwald et al., 1999; Iyer et al., 2004). These clades are the clamp loader clade, the DnaA/CDC6/ORC clade, the classical AAA clade, the HslU/ClpX/Lon/ClpAB-C clade, and the Helix-2 insert clade. The last two clades have been organized into the Pre-sensor 1 hairpin superclade.

Members of the superfamily of AAA+ ATPases carry a nucleotide-binding pocket called the AAA+ domain that ranges from 200 to 250 amino acids, which is formed by an αβα-type Rossmann fold followed by several α helices (Figure 1) (Lupas and Martin, 2002; Iyer et al., 2004; Hanson and Whiteheart, 2005). Such proteins often assemble into ring-shaped or helical oligomers (Davey et al., 2002 ; Iyer et al., 2004; Erzberger and Berger, 2006). Using the nomenclature of Iyer *et al.*, the Rossmann fold is formed by a β sheet of parallel strands arranged in a β5-β1-β4-β3-β2 series. Its structure resembles a wedge. An α helix preceding the β1 strand and a loop that is situated across the face of the β sheet is a distinguishing feature of the AAA+ superfamily. Another characteristic is the position of several α helices positioned above the wide end of the wedge. The P-loop or the Walker A motif (GX_4GKT/S where X is any amino acid) is located between the β1 strand and the following α helix. The Walker B motif (ϕϕϕϕϕDE where ϕ is a hydrophobic amino acid) coordinates a magnesium ion complexed with the nucleoside triphosphate via the conserved aspartate residue. The conserved glutamate is thought to interact with a water molecule to make it a better nucleophile for nucleotide hydrolysis.

AAA+ proteins also share conserved motifs named the Sensor 1, Box VII, and Sensor 2 motifs that coordinate ATP hydrolysis with a change in conformation (Figure 1) (Lupas and Martin, 2002; Iyer et al., 2004; Hanson and Whiteheart, 2005). Relative to the primary amino acid sequence, these motifs are on the C-terminal side of the Walker B motif. The Sensor 1 motif contains a polar amino acid at the end of the β4 strand. On the basis of the X-ray crystal structure of N-ethylmaleimide-sensitive factor, an ATPase involved in intracellular vesicle fusion (Beyer, 1997; Swaffield and Purugganan, 1997), this amino acid together with the acidic residue in the Walker B motif interacts with and aligns the activated water molecule during nucleotide hydrolysis. The Box VII motif, which is also called the SRH (Second Region of Homology) motif, contains an arginine named the arginine finger by its analogous function with the corresponding arginine of GTPase activator proteins that interacts with GTP bound to a small G protein partner to promote GTP hydrolysis. The crystal structures of several AAA+ proteins have shown that the Box VII motif in an individual molecule is located some distance away from the nucleotide binding pocket. In AAA+ proteins that assemble into ring-shaped or helical oligomers, the Box VII motif of one protomer directs an arginine residue responsible for interaction with the γ phosphate of ATP toward the ATP binding pocket of the neighboring molecule. It is proposed that this interaction or lack thereof coordinates ATP hydrolysis with a conformational change. The Sensor 2 motif, which resides in one of the α helices that follow the Rossmann fold, also contains a conserved arginine. For proteins whose structures contain the bound nucleoside triphosphate or a nucleotide analogue, this amino acid interacts with the γ phosphate of the nucleotide. As reviewed by Ogura (Ogura et al., 2004), this residue is involved in ATP binding or its hydrolysis in some but not all AAA+ proteins. Like the arginine finger residue, this arginine is thought to coordinate a change in protein conformation with nucleotide hydrolysis.

此为试读，需要完整PDF请访问：www.ertongbook.com