

Computational Biology

Editors-in-Chief

Andreas Dress
University of Bielefeld (Germany)

Martin Vingron
Max Planck Institute for Molecular Genetics (Germany)

Editorial Board

Gene Myers, Janelia Farm Research Campus, Howard Hughes Medical Institute (USA)
Robert Giegerich, University of Bielefeld (Germany)
Walter Fitch, University of California, Irvine (USA)
Pavel A. Pevzner, University of California, San Diego (USA)

Advisory Board

Gordon Crippen, University of Michigan (USA)
Joe Felsenstein, University of Washington (USA)
Dan Gusfield, University of California, Davis (USA)
Sorin Istrail, Brown University, Providence (USA)
Samuel Karlin, Stanford University (USA)
Thomas Lengauer, Max Planck Institut Informatik (Germany)
Marcella McClure, Montana State University (USA)
Martin Nowak, Harvard University (USA)
David Sankoff, University of Ottawa (Canada)
Ron Shamir, Tel Aviv University (Israel)
Mike Steel, University of Canterbury (New Zealand)
Gary Stormo, Washington University Medical School (USA)
Simon Tavaré, University of Southern California (USA)
Tandy Warnow, University of Texas, Austin (USA)

The *Computational Biology* series publishes the very latest, high-quality research devoted to specific issues in computer-assisted analysis of biological data. The main emphasis is on current scientific developments and innovative techniques in computational biology (bioinformatics), bringing to light methods from mathematics, statistics and computer science that directly address biological problems currently under investigation.

The series offers publications that present the state-of-the-art regarding the problems in question; show computational biology/bioinformatics methods at work; and finally discuss anticipated demands regarding developments in future methodology. Titles can range from focused monographs, to undergraduate and graduate textbooks, and professional text/reference works.

Author guidelines: springer.com > Authors > Author Guidelines

For other titles published in this series, go to <http://www.springer.com/series/5769>

Masao Nagasaki • Ayumu Saito • Atsushi Doi •
Hiroshi Matsuno • Satoru Miyano

Foundations of Systems Biology

Using Cell Illustrator[®] and Pathway Databases

Dr. Masao Nagasaki
Dr. Ayumu Saito
Prof. Satoru Miyano
University of Tokyo
Inst. Medical Science
Human Genome Center
4-6-1 Shirokanedai
Tokyo
Minato-ku
108-8639 Japan
miyano@ims.u-tokyo.ac.jp

Dr. Atsushi Doi
Institute of System LSI Design Industry
Fukuoka R & D Center
3-8-34 Momochihama
Fukuoka
Office 608, Sawara-ku
814-0001 Japan

Prof. Hiroshi Matsuno
Yamaguchi University
Graduate School of Science &
Engineering
Yamaguchi
753-8512 Japan

Computational Biology Series ISSN 1568-2684

ISBN: 978-1-84882-022-7

e-ISBN: 978-1-84882-023-4

DOI: 10.1007/978-1-84882-023-4

Translated by Satoru Miyano, Masao Nagasaki and Ayumu Saito

© Springer-Verlag London Limited 2009

© 2007 Atsushi Doi, Masao Nagasaki, Ayumu Saito, Hiroshi Matsuno, Satoru Miyano

Shisutemu seibutugaku ga wakarū! Seruirasutore-ta wo tsukatte miyōu

ISBN: 978-4-320-05658-9 was originally published in Japanese language by Kyoritsu Shuppan Co., Ltd., Tokyo, Japan in 2007. This translation is published by arrangement with Kyoritsu Shuppan Co., Ltd., Tokyo, Japan.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from Kyoritsu Shuppan Co., Ltd.

Cell Illustrator is the property of Tokyo University and is distributed worldwide by BIOBASE GmbH.

TRANSPATH is a registered trademark of BIOBASE GmbH, Halchtersche Strasse 33, Wolfenbüttel 38304 Germany.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2009922124

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

The use of registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Cover design: KünkelLopka GmbH, Heidelberg

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

Today, as hundreds of genomes have been sequenced and thousands of proteins and more than ten thousand metabolites have been identified, navigating safely through this wealth of information without getting completely lost has become crucial for research in, and teaching of, molecular biology.

Consequently, a considerable number of tools have been developed and put on the market in the last two decades that describe the multitude of potential/putative interactions between genes, proteins, metabolites, and other biologically relevant compounds in terms of metabolic, genetic, signaling, and other networks, their aim being to support all sorts of explorations through bio-data bases currently called *Systems Biology*.

As a result, navigating safely through this wealth of information-processing tools has become equally crucial for successful work in molecular biology.

To help perform such navigation tasks successfully, this book starts by providing an extremely useful overview of existing tools for finding (or designing) and investigating metabolic, genetic, signaling, and other network databases, addressing also user-relevant practical questions like

- Is the database viewable through a web browser?
- Is there a licensing fee?
- What is the data type (metabolic, gene regulatory, signaling, etc.)?
- Is the database developed/maintained by a curator or a computer?
- Is there any software for editing pathways?
- Is it possible to simulate the pathway?

It then goes on to introduce a specific such tool, that is, the fabulous “Cell Illustrator 3.0” tool developed by the authors. The book explains in great detail how this tool can be used for creating, analyzing, and simulating models explicating and testing our current understanding of basic biological processes. They pertain, for example, to

- the organization and control of metabolic networks and metabolic flux analysis,
- the regulation of gene transcription, processing, and translation, or

— the processing of information via signaling pathways.

The book deals with such topics by providing a fascinating array of detailed examples. Thus, it can serve as a perfect introduction to contemporary cell biology for anybody who wants to quickly gain insight into the most important and topical directions of research in this field. In particular, the book provides invaluable help for anybody who wants to learn more about why and how the current big bio-data bases can be used to develop and support Systems Biology research.

Therefore, any biology student can, and actually should, just work through these examples on his own screen to quickly gain important and solid expertise and become a valuable and well-informed member of the continuously growing Systems Biology research community.

The authors Masao Nagasaki, Ayumu Saito, Atsushi Doi, Hiroshi Matsuno, and Satoru Miyano have been working at the forefront of *in silico*-based biology for quite a few years, and are highly respected in the community.

I am therefore very happy to have their book appear in this series, and I congratulate the publishers for the very good work they have done in dealing with the challenging task of appropriately editing such a strongly digitally-oriented manuscript.

Prof. Dr. Andreas Dress

Director

Department of Combinatorics and Geometry (DCG)

CAS-MPG Partner Institute for Computational Biology (PICB)

Shanghai Institutes for Biological Sciences (SIBS)

Chinese Academy of Sciences (CAS)

June 2008

Preface

It has been said that “Systems Biology” is an important postgenomic challenge in biology to understand “life as systems”. That being said, what does it mean? What can be done with signaling pathways, metabolic pathways, and gene regulatory networks using computers? For those with similar concerns or questions, this should be the first book you consult for an understanding of *Systems Biology*.

The definition of *Systems Biology* varies from scientist to scientist. Some of you may have skimmed books or scientific papers with “Systems Biology” in the title and seen alien terms such as “robustness analysis”, “stochastic differential equations”, or “bifurcation analysis” fly by. Some may have felt that this is similar to lining up toy soldiers called differential equations and making them march. Those of you who have felt that way are the intended audience of this book.

Biological organisms consist of many molecules, such as proteins, which fulfill their functions and interact with others. One of the ways to understand this system is to construct the system in parts on a computer and analyze. Beneath the current attentions to Systems Biology is the compilation of large amounts of genomic data and biological knowledge on the parts that compose everything from bacteria to human beings. Since the basic mechanisms of these parts have been considerably well defined, it is now time to understand how the interactions between these parts create the high degree of complexity in biological systems.

On one hand, man-made systems such as electrical circuits and machinery can be made over and over once there are parts and blueprints, since the system is known from the beginning. On the other hand, organisms are made by nature and evolution, and there is a large gap between gathering the parts and understanding the system. Modeling and simulation are necessary technologies to close this gap. In order to understand this system, it needs to be modeled with a high-level language including mathematics and entered into a computer for computation. We should say a goodbye to messy (in Japanese, we say “Gochagocha”) printed diagrams with arrows and circles of various shapes with narrations. This is the point of entry of “Cell Illustrator”, which is a software tool for biological pathway modeling and simulation.

Reading the book and using Cell Illustrator bundled in the CD-ROM should make it possible to create highly complex pathways and simulations. There is no need for

prior knowledge in differential equations or programming. The prerequisites are interest in biology, ability to operate a cell phone (or equivalent), and mathematical ability of a standard middle school student or better.

Using Cell Illustrator, reading the book, and finishing the exercises—answers are provided—should make you realize how easy this can be “(^ o ^) v”. Although pathway drawing does not require any mathematical or programming skills, drawing pathways may require some artistic sense. In addition, just by drawing pathways using Cell Illustrator, pathway knowledge will become better organized, and the reader should feel a sense of accomplishment. The columns interspersed in the book are addendums and digressions; they can be skimmed at the reader’s discretion.

This book is designed and structured to be used for a semester-long course text at the undergraduate level or can be used as a part of graduate courses. Chapter 1 describes a minimum biological knowledge and Chapters 2 and 3 explain some of the important pathway databases and software tools together with their related concepts. Chapter 4 describes the detailed first steps and elements for modeling pathways with Cell Illustrator. The reader may find that graphical pictures representing biological entities and processes help understanding the elements of pathways. Chapter 5 will guide the reader to model three kinds of pathways in a step-by-step manner as exercises. Chapter 6 discusses the computational functionalities required for Systems Biology. This book is an English translation of the original Japanese version published by Kyoritsu Shuppan Co., Ltd. With this edition, the data on software and database versions are updated and Chapter 6 is enhanced with some new topics.

We are grateful to many people. First and foremost, we would like to thank the current and former members of the Cell System Markup Language Project: Emi Ikeda, Euna Jeong, Kaname Kojima, Chen Li, Hiroko Nishihata, Kazuyuki Numata, Yayoi Sekiya, Yoshinori Tamada, Kazuko Ueno of Human Genome Center; Kanji Hioka, Yuto Ikegami, Hironori Kitakaze, Yoshimasa Miwa, Daichi Saihara, Tomoaki Yamamoto of Yamaguchi University.

Andreas Dress should be specially acknowledged for the foreword of this book. For this English version, we were encouraged by Holger Karas and Edgar Wingender of BIOBASE and Wayne Wheeler of Springer U.K. as well as Koichi Nobusawa and Yumiko Kita of Kyoritsu Shuppan Co., Ltd. for the original Japanese version. Special thanks go to Jocelyne Bruand of UCSC and Tatsunori Hashimoto of Harvard University for helping this translation, and to Seiya Imoto, Rui Yamaguchi, Teppei Shimamura, André Fujita, Yosuke Hatanaka, Eric Perrier, Jin Hwan Do, and Takashi Yamamoto for their tremendous supports for Cell Illustrator.

Tokyo,
June 2008

Masao Nagasaki
Ayumu Saito
Atsushi Doi
Hiroshi Matsuno
Satoru Miyano

Contents

Foreword	v
Preface	vii
1 Introduction	1
1.1 Intracellular Events	1
1.1.1 Transcription, Translation, and Regulation	1
1.1.2 Signaling Pathways and Proteins	3
1.1.3 Metabolism and Genes	3
1.2 Intracellular Reactions and Pathways	3
2 Pathway Databases	5
2.1 Major Pathway Databases	5
2.1.1 KEGG	6
2.1.2 BioCyc	8
2.1.3 Ingenuity Pathways Knowledge Base	8
2.1.4 TRANSPATH	8
2.1.5 ResNet	9
2.1.6 Signal Transduction Knowledge Environment (STKE): Database of Cell Signaling	9
2.1.7 Reactome	11
2.1.8 Metabolome.jp	12
2.1.9 Summary and Conclusion	12
2.2 Software for Pathway Display	13
2.2.1 Ingenuity Pathway Analysis (IPA)	13
2.2.2 Pathway Builder	14
2.2.3 Pathway Studio	14
2.2.4 Connections Maps	14
2.2.5 Cytoscape	14
2.3 File Formats for Pathways	15
2.3.1 Gene Ontology	15

2.3.2	PSI MI	16
2.3.3	CellML	16
2.3.4	SBML	16
2.3.5	BioPAX	16
2.3.6	CSML/CSO	17
3	Pathway Simulation Software	19
3.1	Simulation Software Backend	19
3.1.1	Architecture: Deterministic, Probabilistic, or Hybrid?	20
3.1.2	Methods of Pathway Modeling	20
3.2	Major Simulation Software Tools	21
3.2.1	Gepasi/COPASI	21
3.2.2	Virtual Cell	21
3.2.3	Systems Biology Workbench (SBW), Cell Designer, JDesigner	21
3.2.4	Dizzy	22
3.2.5	E-Cell	22
3.2.6	Cell Illustrator	22
3.2.7	Summary	24
4	Starting Cell Illustrator	25
4.1	Installing Cell Illustrator	25
4.1.1	Operating Systems and Hardware Requirements	25
4.1.2	Cell Illustrator Lineup	26
4.1.3	Installing and Running Cell Illustrator	26
4.1.4	License Install	28
4.2	Basic Concepts in Cell Illustrator	28
4.2.1	Basic Concepts	28
4.2.2	Entity	28
4.2.3	Process	30
4.2.4	Connector	33
4.2.5	Rules for Connecting Elements	34
4.2.6	Icons for Elements	35
4.3	Editing a Model on Cell Illustrator	36
4.3.1	Adding Elements	36
4.3.2	Model Editing and Canvas Controls	39
4.4	Simulating Models	41
4.4.1	Simulation Settings	41
4.4.2	Graph Settings	41
4.4.3	Executing Simulation	43
4.5	Simulation Parameters and Rules	44
4.5.1	Creating a Model with Discrete Entity and Process	44
4.5.2	Creating a Model with Continuous Entity and Process	49
4.5.3	Concepts of Discrete and Continuous	51
4.6	Pathway Modeling Using Illustrated Elements	52

4.7	Creating Pathway Models Using Cell Illustrator	55
4.7.1	Degradation	55
4.7.2	Translocation	57
4.7.3	Transcription	60
4.7.4	Binding	62
4.7.5	Dissociation	64
4.7.6	Inhibition	66
4.7.7	Phosphorylation by Enzyme Reaction	68
4.8	Conclusion	73
5	Pathway Modeling and Simulation	75
5.1	Modeling Signaling Pathway	75
5.1.1	Main Players: Ligand and Receptor	75
5.1.2	Modeling EGFR Signaling with EGF Stimulation	76
5.2	Modeling Metabolic Pathways	87
5.2.1	Chemical Equations and Pathway Representations	87
5.2.2	Michaelis-Menten Kinetics and Cell Illustrator Pathway Representation	88
5.2.3	Creating Glycolysis Pathway Model	89
5.2.4	Simulation of Glycolysis Pathway	101
5.2.5	Improving the Model	101
5.3	Modeling Gene Regulatory Networks	106
5.3.1	Biological Clocks and Circadian Rhythms	106
5.3.2	Gene Regulatory Network for Circadian Rhythms in Mice	107
5.3.3	Modeling Circadian Rhythms in Mice	108
5.3.4	Creating Hypothesis by Simulation	119
5.4	Summary	124
6	Computational Platform for Systems Biology	127
6.1	Gene Network of Yeast	127
6.2	Computational Analysis of Gene Network	128
6.2.1	Displaying Gene Network	128
6.2.2	Layout of Gene Networks	130
6.2.3	Pathway Search Function	131
6.2.4	Extracting Subnetworks	132
6.2.5	Comparing Two Subnetworks	133
6.3	Further Functionalities for Systems Biology	136
6.3.1	Languages for Pathways: CSML 3.0 and CSO	136
6.3.2	SaaS Technology	137
6.3.3	Pathway Parameter Search	138
6.3.4	Much Faster Simulation	138
6.3.5	Exporting Pathway Models to Programming Languages	138
6.3.6	Pathway Layout Algorithms	139
6.3.7	Pathway Database Management System	141
6.3.8	More Visually: Automatic Generation of Icons	142

Bibliographic Notes	145
Index	151