BIOLOGICAL MODELING AND SIMULATION

A Survey of Practical Models, Algorithms, and Numerical Methods

Russell Schwartz

The MIT Press
Cambridge, Massachusetts
London, England

© 2008 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Times New Roman and Syntax on 3B2 by Asco Typesetters, Hong Kong. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Schwartz, Russell.
Biological modeling and simulation : a survey of practical models, algorithms, and numerical methods / Russell Schwartz.
p. cm. — (Computational molecular biology)
Includes bibliographical references and index.
ISBN 978-0-262-19584-3 (hardcover : alk. paper) 1. Biology—Simulation methods. 2. Biology—Mathematical models. I. Title.
QH323.5.S364 2008
570.1′1—dc22 2008005539

10 9 8 7 6 5 4 3 2 1

Preface

This text arose from a class on biological modeling I have been teaching annually at Carnegie Mellon University since 2004. I created the class to fill what I saw as a gap in the available computational biology teaching materials. There are many excellent sources from which one can learn about successful approaches that have been developed for various core problems in computational biology (e.g., building phylogenies, implementing molecular simulations, or inferring DNA binding motifs). What seems to me to have been missing, though, is material to prepare aspiring computational biologists to solve the next problem, the one that no one has studied yet. Too often, computational biology courses assume that if a student is well prepared in biology and in computer science, then he or she can figure out how to apply the one to the other. In my experience, however, a computational biologist who wants to be prepared for a broad range of unexpected problems needs a great deal of specialized knowledge that is not part of the standard curriculum of either discipline. The material included here reflects my attempt to prepare my students for the sorts of unanticipated problems a computational biology researcher is like to encounter by collecting in one place a set of broadly useful models and methods one would ordinarily find scattered across many classes in several disciplines.

Meeting this challenge—preparing students for solving a wide array of problems without knowing what those problems will be—requires some compromises. Many potentially useful tools had to be omitted, and none could be covered in as much depth as I might have liked so that I could put together a "bag of tricks" that is likely to serve the aspiring researcher well on a broad class of biological problems. I have for the most part chosen techniques that have proved useful in diverse biological modeling contexts in the past. In a few cases, I have selected methods that are not yet widely used in biological modeling but that I believe have great potential. For every topic, I have tried to focus on what the practitioner needs to know in order to use these techniques effectively, sacrificing theoretical depth to accommodate greater breadth. This approach will surely grate on some readers, and indeed I feel that this material is best treated not as a way to master any particular techniques, but rather

as a set of possible starting points for use in the modeling problems one encounters. My goal is that a reader who learns the material in this text will be able to make at least a first attempt at solving nearly any computational problem he or she will encounter in biology, and will have a good idea where to go to learn more if that first attempt proves inadequate.

This text is designed for readers who already have some familiarity with computational and biological topics. It assumes an introductory knowledge of algorithms and their analysis. Portions of the text also assume knowledge of calculus, linear algebra, and probability at the introductory undergraduate level. Furthermore, though the text teaches computational methods, its goal is to help readers solve biological problems. The reader should therefore be prepared to encounter many toy examples and a few extended case studies showing how the methods covered here have been applied to various real problems in biology. Readers are therefore likely to need a general knowledge of biology at the level of at least an undergraduate introductory survey course. When I teach this material, a key part of the learning experience consists of exercises in which students are presented with biological problems and are expected to formulate, and often implement, models using the techniques covered here. While one need not necessarily use the text in that way, it is written for readers capable of writing their own computer code.

I would like to thank the many people who have made this work possible. Sorin Istrail, one of my mentors in this field, provided very helpful encouragement for this project, as did my editors at the MIT Press, Bob Prior and Katherine Almeida. Mor Harchol-Balter provided valuable advice on clarifying my presentation of continuous-time Markov models. And I am grateful to my many teachers throughout the years in whose classes I picked up bits and pieces of the material of this text. I had the mixed blessing of having realized I wanted to be a computational biologist as a student in the days before computational biology classes were widespread. Many of the topics here are pieced together from subjects I found useful in inventing my own computational biology curriculum with the advice of my graduate mentor, Bonnie Berger. Most important in preparing this work have been the students in my class, who have provided much helpful criticism as this material evolved from handwritten lecture notes to typeset handouts, and finally to its present form. Though all of my students deserve some thanks, the following have been particularly helpful in offering corrections and criticism on various editions of this work and suggesting new topics that made their way into the final version: Byoungkoo Lee, Srinath Sridhar, Tiequan Zhang, Arvind Ramanathan, and Warren Ruder.

This material is based upon work supported by the National Science Foundation under glant no. 0346981. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.