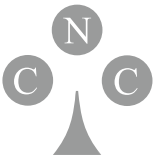
Natural Computing Series



For further volumes:
http://www.springer.com/series/4190

Nataša Jonoska
Masahico Saito
Editors

Discrete and Topological Models in Molecular Biology

Springer

Editors
Nataša Jonoska
Masahico Saito
Department of Mathematics & Statistics
University of South Florida
Tampa FL, Florida
USA

Series Editors
G. Rozenberg (Managing Editor)

Th. Bäck, J.N. Kok, H.P. Spaink
Leiden Center for Natural Computing
Leiden University
Leiden, The Netherlands

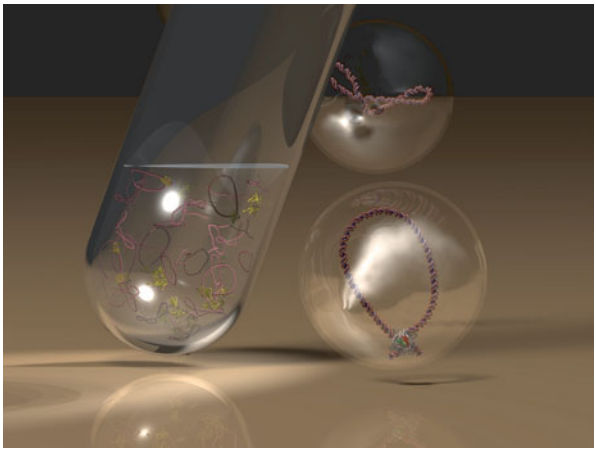
A.E. Eiben
Vrije Universiteit Amsterdam
The Netherlands

ISSN 1619-7127 Natural Computing Series
ISBN 978-3-642-40192-3 ISBN 978-3-642-40193-0 (eBook)
DOI 10.1007/978-3-642-40193-0
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013957136

Printed on acid-free paper

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



The Cre-loxP recombination system has been used as a tool to characterize DNA tertiary structure, for example in difference-topology experiments that are based on tangle analysis. Difference topology depends on the formation of a looped recombinase-DNA intermediate; however, kinetic details of the Crerecombination pathway have not been sufficiently characterized for the recombination system to be used quantitatively. By analyzing the synapsis steps in Cre recombination, Shoura and Levene showed that the free energy of DNA-loop formation can be directly measured using a novel fluorescence resonance energy transfer (FRET)-based reporter technique. The featured method uses special DNAs bearing loxP recombination sites that contain fluorophore tags (in red and green) at specific nucleotide positions. Addition of Cre protein (yellow) to the reaction generates a looped synaptic complex and recombination products (linear and circular DNA). Innovative applications of their technique include superhelical, knotted, or catenated DNA substrates in vitro and in living cells. See the chapter by Shoura and Levene for more details. (Conceptual design and artwork is by Massa Shoura and Udayana Ranatunga. Supercoiled DNA PDB are from MD simulations by Sarah Harris. Other DNA structures were made using VMD and the GraphiteLifeExplorer software. Final image was rendered using PovRay.)

Preface

Commonly used models in mathematical biology involve dynamical systems, differential equations, and statistics. These fields often study the general behavior and dynamics of biological systems either at the population level or in terms of cell-to-cell interactions, tissue development, or organ function, but they are less often used in the study of biomolecular processes such as genetics, biomolecular structures and interactions.

With the explosion of research in molecular biology, and in particular the enormous experimental data generated in the last couple of decades, new mathematical tools are being developed using graph theory, algebra, combinatorics, discrete stochastic processes, and topology. These new methods allow us to "zoom-in" to the cell to better understand spatial macromolecular arrangements and molecular interactions within the cell, or a portion of the cell.

With this volume, we wish to introduce several aspects of these contemporary approaches in mathematical (molecular) biology that contain a variety of models, covering a wide spectrum of problems in molecular biology.

The chapter authors are experts in their own fields and have diverse scientific background ranging from biology to biophysics, physics, computer science, and mathematics. The collection of their experiences gives different perspectives on sometimes similar biological problems and, we hope, will help in understanding the mathematical tools as well as the biological process.

The book is divided into five parts devoted to general biological themes, while the mathematical methods introduced in each theme differ dramatically.

The first part of the book concentrates on data analysis, including genetic data and data related to brain activities. The chapters by Franco and Angeleska et al. deal with short nucleotide segments. While Franco describes methods from formal language theory to learn about the functionality of genetic segments, Angeleska et al. survey methods, mainly based on graph theory, for parsing and annotating sequencing data for genome assembly. The next two chapters, by Carbone and Bonizzoni et al., describe methods to understand gene or protein (co)evolutions through analysis of distances in phylogenetic trees (Carbone) or graph theoretical methods (Bonizzoni et al.). The last two chapters deal with different types of data

analyses. Based on MRI scans, Daley describes a method to construct a graph for the brain neural network through threshold functions, while Nanda and Sazdanović show how to use algebraic topology to analyze a variety of data types, including neural connections in the brain.

The second part of the book deals with biomolecular spatial arrangements. Chapters in this part use techniques from combinatorics and graph theory as well as purely algebraic methods using groups of symmetries. This segment starts with problems about RNA secondary structures (Heitch and Poznanović); continues with rigid and flexible regions in a protein ternary structure (Fox and Streinu), supramolecular assembly of viral capsids with clusters of proteins (Sitharam), and transitions of dodecahedral and icosahedral symmetries in viral capsid expansions (Cermelli et al.); and concludes with three-dimensional synthetic DNA structures (Ellis-Monaghan et al.).

DNA rearrangements have been observed on both developmental and evolutionary scales, and some of the most extensive shuffling of genetic material has been observed in certain species of single-cell organisms (ciliates). Goldman et al. start the third part of the book with a brief survey of the biological process, while the next two chapters cover mathematical methods that describe the rearrangement process using matrix algebras (Brijder and Hoogeboom) and topological aspects of graphs (Dolzhenko and Valencia).

The fourth part on spatial embeddings of biomolecules starts with an introduction to DNA topology by Darcy et al., providing basic biological and topological background of the subject. Buck's exposition on methods from knot theory capturing enzymatic actions that control topological embeddings of DNA is followed by Baker's mathematical development of these methods. This part of the book ends with a chapter by Ishihara et al. applying the described methods to a specific experimentally observed biological process.

The fifth and last part of the book deals with the kinetics and dynamics of molecular interactions. It starts with analyzing looping of DNA through enzyme kinetics (Shoura and Levene) and moves into reaction networks using the quasi-steady-state assumption by differential and matrix equations (Pantea et al.) and a survey of algebraic methods in systems biology (Laubenbacher et al.). This part, and the book, concludes with chapters by Savageau and Lomnitz, who use dynamical systems to study phenotype development, and Rejniak, who shows a computational model that can capture spatial tissue development including mutant morphologies.

We hope this volume will be suitable as a reference book for researchers in mathematics and theoretical computer science who are interested in modeling molecular and biological phenomena using discrete methods as well as for biologists looking for available mathematical tools in discrete models. It may also serve as a guide and supplement for a graduate course in mathematical biology or bioinformatics, to introduce discrete aspects of mathematical biology. Many chapters end with open problems that can serve as the basis for research developments.

We wish to thank all the contributing authors for their work in producing these chapters. Every contribution was reviewed by at least two researchers, whose valuable comments and suggestions helped in composing the final product. We

express our deep gratitude to them for their time and effort. Many of the contributors were also participants of a workshop on Discrete and Topological Models in Molecular Biology held at the University of South Florida in March 2012. The workshop received generous support from the National Science Foundation through the grant DMS-1157242. This allowed for a very successful workshop and provided opportunities for the research community to meet and exchange ideas that spurred the development of this volume. Finally, we wish to acknowledge support for this project, in part, by the NSF grant DMS-0900671 and CCF-1117254.

Tampa, FL, USA
June 2013

Nataša Jonoska
Masahico Saito

Contents