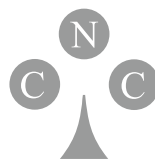


# Natural Computing Series



Series Editors: G. Rozenberg

Th. Bäck A.E. Eiben J.N. Kok H.P. Spaink

Leiden Center for Natural Computing

---

Advisory Board: S. Amari G. Brassard K.A. De Jong C.C.A.M. Gielen  
T. Head L. Kari L. Landweber T. Martinez Z. Michalewicz M.C. Mozer  
E. Oja G. Păun J. Reif H. Rubin A. Salomaa M. Schoenauer  
H.-P. Schwefel C. Torras D. Whitley E. Winfree J.M. Zurada

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. Spaijk  
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

© Springer-Verlag Berlin Heidelberg 2014

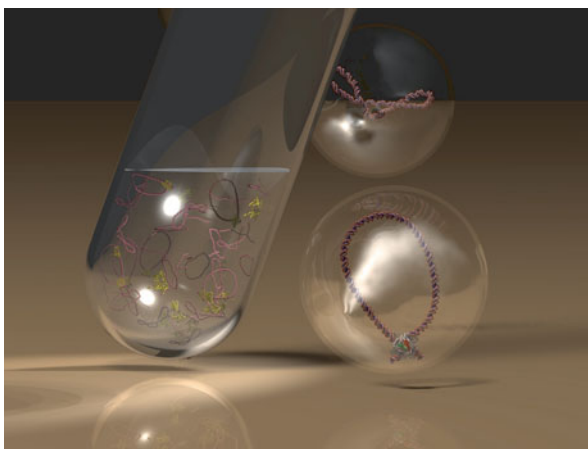
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

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

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://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 Cre recombination 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 Hoozeboom) 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

## Part I Discrete and Graph-Theoretic Models for Data Analysis

<b>Perspectives in Computational Genome Analysis</b> .....	3
Giuditta Franco	
<b>The Sequence Reconstruction Problem</b> .....	23
Angela Angeleska, Sabrina Kleessen, and Zoran Nikoloski	
<b>Extracting Coevolving Characters from a Tree of Species</b> .....	45
Alessandra Carbone	
<b>When and How the Perfect Phylogeny Model Explains Evolution</b> .....	67
Paola Bonizzoni, Anna Paola Carrieri, Gianluca Della Vedova, Riccardo Dondi, and Teresa M. Przytycka	
<b>An Invitation to the Study of Brain Networks, with Some Statistical Analysis of Thresholding Techniques</b> .....	85
Mark Daley	
<b>Simplicial Models and Topological Inference in Biological Systems</b> .....	109
Vidit Nanda and Radmila Sazdanović	

## Part II Molecular Arrangements and Structures

<b>Combinatorial Insights into RNA Secondary Structure</b> .....	145
Christine Heitsch and Svetlana Poznanović	
<b>Redundant and Critical Noncovalent Interactions in Protein Rigid Cluster Analysis</b> .....	167
Naomi Fox and Ileana Streinu	
<b>Modeling Autonomous Supramolecular Assembly</b> .....	197
Meera Sitharam	

<b>The Role of Symmetry in Conformational Changes of Viral Capsids: A Mathematical Approach .....</b>	<b>217</b>
Paolo Cermelli, Giuliana Indelicato, and Reidun Twarock	
<b>Minimal Tile and Bond-Edge Types for Self-Assembling DNA Graphs .....</b>	<b>241</b>
Joanna Ellis-Monaghan, Greta Pangborn, Laura Beaudin, David Miller, Nick Bruno, and Akie Hashimoto	
<b>Part III Gene Rearrangements</b>	
<b>Programmed Genome Processing in Ciliates .....</b>	<b>273</b>
Aaron David Goldman, Elizabeth M. Stein, John R. Bracht, and Laura F. Landweber	
<b>The Algebra of Gene Assembly in Ciliates.....</b>	<b>289</b>
Robert Brijder and Hendrik Jan Hooeboom	
<b>Invariants of Graphs Modeling Nucleotide Rearrangements .....</b>	<b>309</b>
Egor Dolzhenko and Karin Valencia	
<b>Part IV Topological Models and Spatial DNA Embeddings</b>	
<b>Introduction to DNA Topology .....</b>	<b>327</b>
Isabel K. Darcy, Stephen D. Levene, and Robert G. Scharein	
<b>Reactions Mediated by Topoisomerases and Other Enzymes: Modelling Localised DNA Transformations .....</b>	<b>347</b>
Dorothy Buck	
<b>Site-Specific Recombination on Unknot and Unlink Substrates Producing Two-Bridge Links.....</b>	<b>363</b>
Kenneth L. Baker	
<b>Site-Specific Recombination Modeled as a Band Surgery: Applications to Xer Recombination .....</b>	<b>387</b>
Kai Ishihara, Koya Shimokawa, and Mariel Vazquez	
<b>Part V Dynamics and Kinetics of Molecular Interactions</b>	
<b>Understanding DNA Looping Through Cre-Recombination Kinetics .....</b>	<b>405</b>
Massa J. Shoura and Stephen D. Levene	
<b>The QSSA in Chemical Kinetics: As Taught and as Practiced .....</b>	<b>419</b>
Casian Pantea, Ankur Gupta, James B. Rawlings, and Gheorghe Craciun	
<b>Algebraic Models and Their Use in Systems Biology .....</b>	<b>443</b>
Reinhard Laubenbacher, Franziska Hinkelmann, David Murrugarra, and Alan Veliz-Cuba	

**Deconstructing Complex Nonlinear Models in System Design Space..... 475**  
Michael A. Savageau and Jason G. Lomnitz

**IBCell Morphocharts: A Computational Model for Linking Cell  
Molecular Activity with Emerging Tissue Morphology ..... 507**  
Katarzyna A. Rejniak