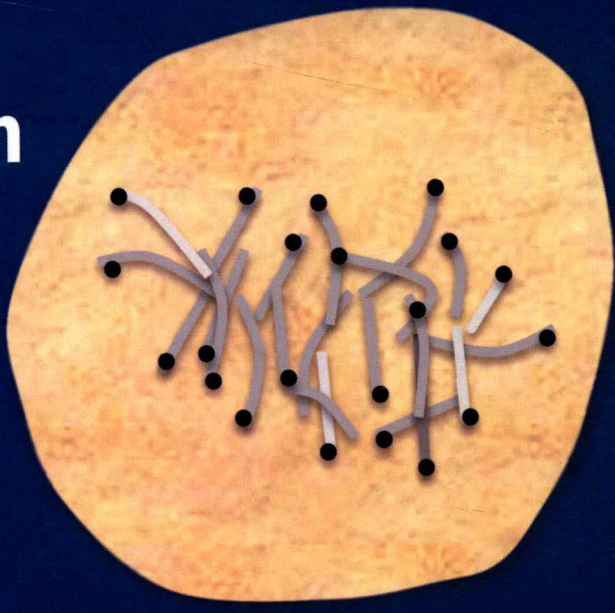
Chapman & Hall/CRC
Mathematical and Computational Biology Series

CHROMATIN

Structure
Dynamics
Regulation



Ralf Blossey

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Chapman & Hall/CRC Mathematical and Computational Biology Series

CHROMATIN

Structure
Dynamics
Regulation

Ralf Blossey

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper

International Standard Book Number-13: 978-1-4987-2938-3 (Hardback)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

CHROMATIN

Structure
Dynamics
Regulation

CHAPMAN & HALL/CRC
Mathematical and Computational Biology Series

Aims and scope:
This series aims to capture new developments and summarize what is known over the entire spectrum of mathematical and computational biology and medicine. It seeks to encourage the integration of mathematical, statistical, and computational methods into biology by publishing a broad range of textbooks, reference works, and handbooks. The titles included in the series are meant to appeal to students, researchers, and professionals in the mathematical, statistical and computational sciences, fundamental biology and bioengineering, as well as interdisciplinary researchers involved in the field. The inclusion of concrete examples and applications, and programming techniques and examples, is highly encouraged.

Series Editors

Proposals for the series should be submitted to one of the series editors above or directly to:
CRC Press, Taylor & Francis Group
3 Park Square, Milton Park
Abingdon, Oxfordshire OX14 4RN
UK

Published Titles

An Introduction to Systems Biology: Design Principles of Biological Circuits
Uri Alon

Glycome Informatics: Methods and Applications
Kiyoko F. Aoki-Kinoshita

Computational Systems Biology of Cancer
Emmanuel Barillot, Laurence Calzone, Philippe Hupé, Jean-Philippe Vert, and Andrei Zinovyev

Python for Bioinformatics
Sebastian Bassi

Quantitative Biology: From Molecular to Cellular Systems
Sebastian Bassi

Methods in Medical Informatics: Fundamentals of Healthcare Programming in Perl, Python, and Ruby
Jules J. Berman

Chromatin: Structure, Dynamics, Regulation
Ralf Blossey

Computational Biology: A Statistical Mechanics Perspective
Ralf Blossey

Game-Theoretical Models in Biology
Mark Broom and Jan Rychtář

Computational and Visualization Techniques for Structural Bioinformatics Using Chimera
Forbes J. Burkowski

Structural Bioinformatics: An Algorithmic Approach
Forbes J. Burkowski

Spatial Ecology
Stephen Cantrell, Chris Cosner, and Shigui Ruan

Cell Mechanics: From Single Scale-Based Models to Multiscale Modeling
Arnaud Chauvière, Luigi Preziosi, and Claude Verdier

Bayesian Phylogenetics: Methods, Algorithms, and Applications
Ming-Hui Chen, Lynn Kuo, and Paul O. Lewis

Statistical Methods for QTL Mapping
Zehua Chen

An Introduction to Physical Oncology: How Mechanistic Mathematical Modeling Can Improve Cancer Therapy Outcomes
Vittorio Cristini, Eugene J. Koay, and Zhihui Wang

Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems
Qiang Cui and Ivet Bahar

Kinetic Modelling in Systems Biology
Oleg Demin and Igor Goryanin

Data Analysis Tools for DNA Microarrays
Sorin Draghici

Statistics and Data Analysis for Microarrays Using R and Bioconductor, Second Edition
Sorin Drăghici

Computational Neuroscience: A Comprehensive Approach
Jianfeng Feng

Biological Sequence Analysis Using the SeqAn C++ Library
Andreas Gogol-Döring and Knut Reinert

Gene Expression Studies Using Affymetrix Microarrays
Hinrich Göhlmann and Willem Talloen

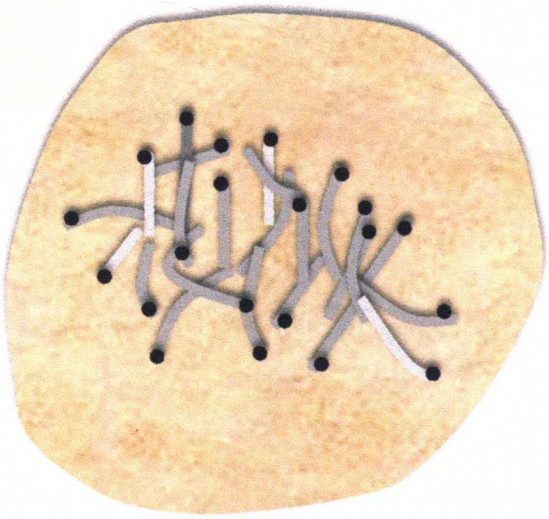
Handbook of Hidden Markov Models in Bioinformatics
Martin Gollery

Meta-analysis and Combining Information in Genetics and Genomics
Rudy Guerra and Darlene R. Goldstein

Differential Equations and Mathematical Biology, Second Edition
D.S. Jones, M.J. Plank, and B.D. Sleeman

Knowledge Discovery in Proteomics
Igor Jurisica and Dennis Wigle

Introduction to Proteins: Structure, Function, and Motion
Amit Kessel and Nir Ben-Tal



Theorie ist bei mir immer nachträglich.

Max Frisch

Preface

Chromatin, the genome-containing complex in the nucleus, is the central player in eukaryote biology. It is a huge macromolecular complex made up from DNA, histone proteins, and many other protein and RNA components. It contains the basic genetic information coded in DNA, but also all machinery to read it. In this analogy drawn with information technology, chromatin is both the software saved on the disk (DNA), but also all the electronics that go along with it to let the software run.

Understanding chromatin is a complicated task and meanwhile scientists with different backgrounds are active in trying to disentangle its properties. Biologists of different breeds have been joined by physicists and computational biologists. The purpose of this book is to provide a quick entry for a newcomer to the field who comes to the topic with a more "computational" background. I have attempted to present what is currently established about chromatin as far as its basic structure and involvement in gene regulation is concerned. Very much about chromatin is, however, although already qualitatively and quantitatively described, not yet really understood, and I came across a lot of debated topics when reading the literature. I have therefore avoided tracing these discussions here. Another difficulty for a newcomer to the field is that, common to biology, many chromatin-related topics are discussed in specific genetic or even medical contexts, which to the uninitiated hides the forest for the trees. This book is more about some of the trees. On the other side, developing a however "general" theory of chromatin is still a far way out. I try to reflect this in the motto I put at the beginning of the book, a quote from Swiss-German writer Max Frisch, which translates into English as *For me, theory always comes later.* In the case of chromatin, for all of us this is the case.

This book comes along with its subject dissected into three of its aspects, chromatin structure, dynamics and regulation. These topics are very much intertwined such that there is no natural or obvious way to separate them altogether. In the chapters of this book, the reader will therefore encounter these aspects in a sometimes mixed way. The material covered in the book is organized in six chapters, ranging from basic structural aspects to the dynamic and regulatory topics on different time and spatial scales.

Chromatin structure in this text is understood to be first of all the structure of the *chromatin fiber*, i.e., DNA and its organization by the histone proteins. Likewise, the notion of chromatin dynamics refers to the dynamic properties of this fiber and of the role the nucleosomes play, as well as the topoisomerases as one class of specific enzymes acting on DNA. A second set of chromatin-related enzymes, the so-called chromatin remodeling complexes, appear on stage in a chapter on histones and histone-acting enzymes. As this is a topic I worked on myself, it receives special attention here. Finally, this book interprets the notion of regulation in terms of the key biological function of chromatin, i.e., the switching on or off of *genes*, either temporarily or permanently.

This book results from several years of thinking about chromatin – mainly about its regulation – and the discussions I had with many colleagues. They are all thanked for the insights they shared with me, in particular my collaborators Guillaume Brysbaert, Fabrizio Cleri, Ana Maria Florescu, Marc Lensink, Helmut Schiessel, Raghav Singh and Yves Vandecan. I am also very grateful to numerous colleagues I met throughout the years and whose insights influenced my own path through this field. I thank particularly Alain Arneodo, Arndt Benecke (who introduced me to the topic), Bradley Cairns, Andrew Flaus, Ulrich Gerland, Jörg Langowski, Christophe Lavelle, Geeta Narlikar, Tom Owen-Hughes, Franklin Pugh, Karsten Rippe, Cédric Vaillant and Jean-Marc Victor and many others.

In the course of my interest in the field of chromatin I also enjoyed a number of discussions with Jon Widom. Jon's insights have set essential marks in the field; he is deeply missed by his colleagues for the unique combination of the profoundness of his scientific vision and his human side. I dedicate this book to his memory.

Ralf Blossey
Lille, May 2017

Contents

List of Figures

List of Tables

CHAPTER 1

DNA and the nucleosome

DNA structure

In our age of genomics, the biology of the cell is dominated by a DNA-centered view. In the early days of cellular biology this was not the case, as observations were limited to light microscopy. In the course of his studies of the transformations a eukaryotic cell undergoes – what we nowadays call the cell cycle – the biologist Walther Flemming observed the dynamical changes of the material in the nucleus of a cell and made drawings of the different stages of the cell cycle – like the sketch shown in the beginning of the book, which is inspired by the original drawings which can be found in the reprint (Flemming, 1965). Flemming called the complex structure in the nucleus *chromatin*, as it could be dyed and made visible in this way. In the course of the cell cycle, the nuclear material undergoes numerous visible shape changes, and, in particular in metaphase, condenses into the well-known form of X-shaped chromosomes.

In contrast to Flemming we now know that DNA is the substrate of these structures, and in order for it to undergo the structural changes a great number of different, very specific proteins intervene. In this book we try to understand the key features of this material as they are currently known. We will do this from the bottom up, hence start with DNA, and in the course of the chapters we will move up from the level of DNA to the chromatin complex in the cell.

The DNA molecule is a *nucleic acid* built from two single strands of nucleotides. Each of the strands is composed of a nitrogen-containing *nucleobase*, either cytosine (C) and thymine (T), the pyrimidines, and guanine (G) and adenine (A), the purines, and a monosaccharide sugar called deoxyribose and a phosphate group. The nucleobases form hydrogen bonds between corresponding Watson–Crick pairs, two for AT and three for CG, thereby linking the

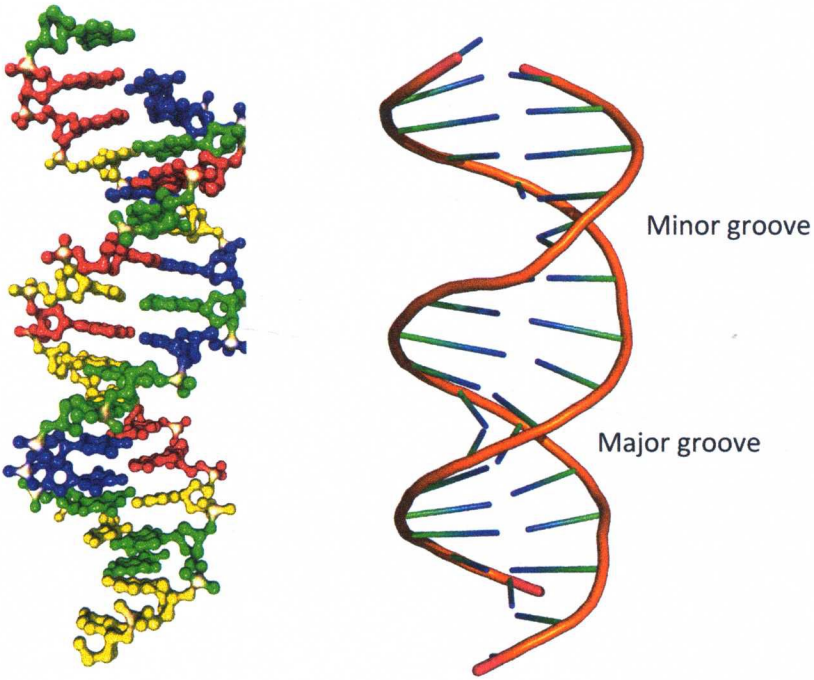


FIGURE 1.1 The molecular structure of DNA. Left: hyperball representation with color code as red = adenine, blue = thymine, green = guanine, yellow = cytosine. Realized with UnityMol 0.9.6. Right: cartoon representation of the same molecule, created with Pymol 1.7. Data taken from PDB entry 1HJC. Courtesy G. Brysbaert.

two strands into the double helix. The nucleobases are stacked on each other by covalent bonds between the sugar of one nucleotide and the phosphate of the next, which leads to an alternating sugar-phosphate backbone. Figure 1.1 displays an exemplary structure of the DNA double helix.

Both DNA strands wind around each other every 10.4 base pairs (bp), giving rise to the double-helical structure of DNA with its associated minor and major grooves, which are also visible in Figure 1.1. The DNA molecule can appear in several conformations which differ in the stacking of the base pairs; these are called the A-, B- and Z forms, whereby the A-form is a more compact structure than B-DNA, and the Z-form is left-winding. The B-DNA structure is the biologically most relevant one, and also the one originally observed in X-ray crystallography. Figure 1.2 shows a sketch of the original X-ray graph from