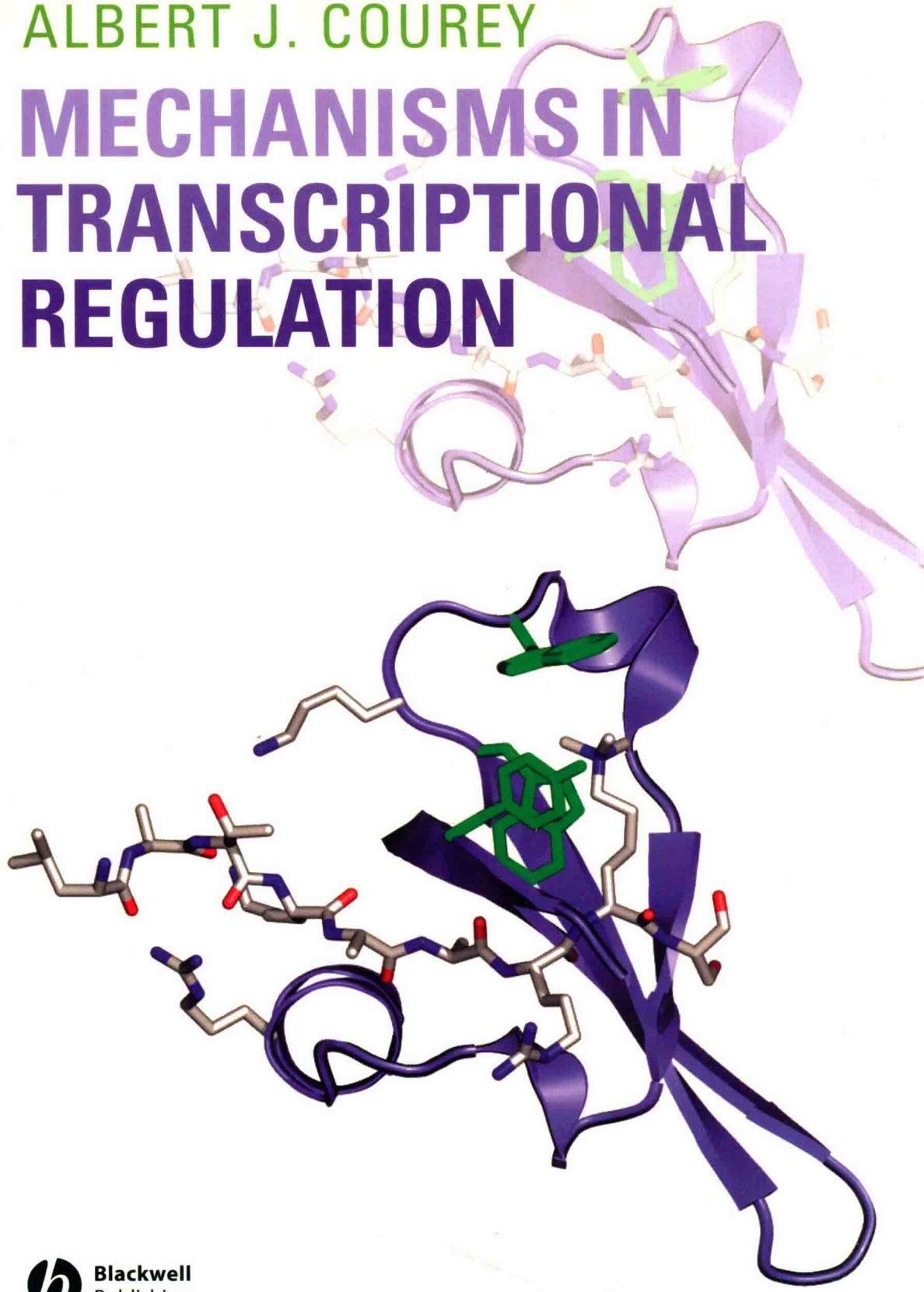
ALBERT J. COUREY

# MECHANISMS IN TRANSCRIPTIONAL REGULATION



Blackwell Publishing

# Mechanisms in Transcriptional Regulation

**Albert J. Courey**

University of California, Los Angeles

**Blackwell**
Publishing

BLACKWELL PUBLISHING
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford OX4 2DQ, UK
550 Swanston Street, Carlton, Victoria 3053, Australia



First published 2008 by Blackwell Publishing Ltd

1 2008

*Library of Congress Cataloging-in-Publication Data*

Courey, Albert J.
Mechanisms in transcriptional regulation / Albert J. Courey.
p.; cm.
Includes bibliographical references and index.
ISBN-13: 978-1-4051-0370-1 (pbk. : alk.paper)
ISBN-10: 1-4051-0370-1 (pbk. : alk.paper)
1. Genetic transcription – Regulation. 2. Transcription factors. I. Title.
[DNLM: 1. Eukaryotic Cells – physiology. 2. Transcription, Genetic. 3. Bacteria – genetics. QH450.2 C859m 2008]

QH450.2C68 2008
572.8'845 – dc22 2007038526

A catalogue record for this title is available from the British Library.

Set in 10/12.5 Meridien
by Graphicraft Limited, Hong Kong
Printed and bound in Singapore
by Fabulous Printers Pte Ltd

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy, and which has been manufactured from pulp processed using acid-free and elementary chlorine-free practices. Furthermore, the publisher ensures that the text paper and cover board used have met acceptable environmental accreditation standards.

For further information on
Blackwell Publishing, visit our website:
www.blackwellpublishing.com

*Cover illustration*: The Polycomb chromodomain (Protein Data Bank id 1PDQ) complexed to a portion of the histone H3 N-terminal tail containing trimethyl-lysine 27. This structure was first described in: Fischle, W., Wang, Y., Jacobs, S. A., Allis, C. D. and Khorasanizadeh, S. (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev*, **17**, 1870–1881.

# Mechanisms in Transcriptional Regulation

*Dedicated to Jody, Sibyl, and Ethan for their unconditional love and support*

# Preface

The impetus for this book is a course that I teach at UCLA – "Mechanisms of Eukaryotic Transcriptional Regulation". While this course focuses on eukaryotic transcription, almost all of the students take it just after completing a course entitled "Mechanisms of Prokaryotic Transcriptional Regulation" taught by my colleague Jay Gralla. In Jay's course, the students are introduced to a number of ideas in bacterial transcription that greatly aid an understanding of eukaryotic transcription. Thus, although this book focuses on eukaryotes, I have included key ideas and examples from bacterial transcription.

Like my course, this book is intended for both upper division undergraduates and graduate students in the molecular life sciences. It should also be useful to more senior scientists who find that their research has, either by accident or by design, entered the realm of transcription, and who therefore require a basic introduction to the field.

This book is by no means a comprehensive account of transcription, since a complete treatment of this large and rapidly expanding field would require multiple volumes. Instead, it discusses a few topics that are essential to an appreciation of the field and illustrates these topics with a number of carefully selected examples. These topics include the workings of the basal transcriptional machinery (Chapters 2 and 3), mechanisms of activation (Chapter 4), the role of chromatin in eukaryotic transcriptional control including mechanisms of epigenetic regulation (Chapters 5 and 6), and mechanisms of combinatorial control (Chapter 7).

While this book emphasizes unifying ideas, it is only through an appreciation of the details that underlie these ideas that one gains a deep understanding of them. Therefore, I have attempted to include enough examples to help readers feel at home with the concepts and the experimental basis behind them. My goal is to equip readers with the ability to approach the primary scientific literature with a critical mind.

While most of the conclusions presented in this book are broadly accepted, transcription is a rapidly evolving field being actively pursued in thousands of research labs around the world. In an effort to capture some of the current excitement in the field, I have chosen to cover certain topics that are not yet completely settled. Rather than attempting

to present all sides of every issue, I have, in some cases, chosen to take a particular stance that I believe is supported by the currently available evidence. Thus, some of the conclusions presented in this book will no doubt require future revision.

To increase accessibility, each chapter includes essential background information set off from the main text. These boxes introduce relevant experimental approaches (e.g., protein chromatography, the use of chemical probes, chromatin immunoprecipitation assays, genetic suppression) and relevant topics in biology and biochemistry (e.g., protein structure visualization, cooperativity, *Drosophila* segmentation). In addition, a few essential terms that are not defined in the main text are explained in the margins, and a comprehensive list of definitions is provided in the form of a glossary at the end of the book (terms highlighted in color in the text).

Since it is just as important to understand what we do not know as it is to understand what we do know, explicit discussions of some of the most important open questions is presented in side boxes. Furthermore, each chapter also includes a set of problems, many of which bring up complex unresolved issues and are intended to stimulate discussion. Possible answers to these problems are provided at the end of the book.

Finally, each chapter contains a list of suggested further reading grouped and ordered to relate to the way that the material is presented in the text. These are not meant to be comprehensive bibliographies, but instead include a few articles mainly from the primary research literature that have influenced my thinking and that have withstood the test of time. In generating these reading lists, I have left out many equally important and worthy papers, and for this, I apologize to my colleagues in the scientific community.

Albert Courey

# Acknowledgments

I would like to acknowledge my colleagues at UCLA, including the members of my research group and the Department of Chemistry and Biochemistry, for their support as I devoted long hours to this project. I especially thank Jay Gralla, with whom I began co-teaching the Mechanisms of Transcriptional Regulation course series at UCLA 17 years ago. It is thanks to Jay, whose research delves into mechanisms of both bacterial and eukaryotic transcription, that I came to realize that bacteria and eukaryotes are not such different beasts after all, and that an appreciation of one is essential to an understanding of the other. The comments, questions, and responses from our many students over the years have helped me to improve my presentation of the beautiful intricacies of transcriptional control.

My affiliation with the UCLA Interdepartmental Graduate Program in Gene Regulation has played a central role in shaping my thinking about the subject of this book. The weekly journal club associated with this program is a forum for penetrating discussion of the latest developments in the field of transcription. My frequent exchanges with such colleagues as Arnold Berk, Michael Carey, Michael Grunstein, Siavash Kurdistani, and Stephen Smale have been extremely stimulating.

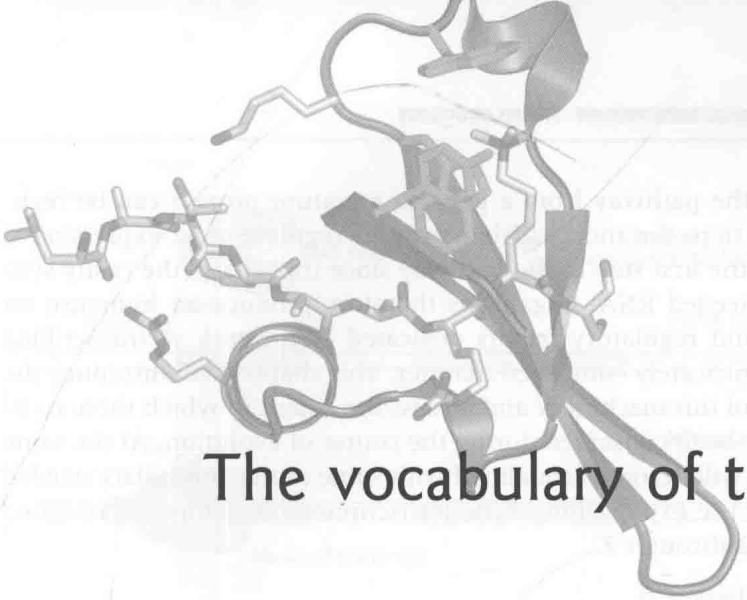
This book has benefited greatly from the insightful comments of James Kadonaga (UC San Diego), Songtao Jia (Columbia University), and Stephen Small (New York University). Their thought-provoking observations and questions have been invaluable in my efforts to improve the readability and focus of the text.

I am also grateful to Blackwell Publishing for shepherding this book through to completion: Nancy Whilton visited me year after year before finally convincing me to take on the project; Elizabeth Frank and Haze Humbert showed great patience when I missed deadlines; Steve Weaver and Karen Chambers continued to support the project after those who initiated it at Blackwell moved on to other things; Jane Andrew paid great attention to detail during the proofreading process. Special thanks are also due to Chris Lear and his colleagues at J&L Composition for their patience in repeatedly revising the illustrations.

Finally, my deepest thanks go to James C. Wang (Harvard University) and Robert Tjian (UC Berkeley) for teaching me that one can only appreciate the big picture after a thorough consideration of the details.

# Full contents

# List of boxes



# 1 The vocabulary of transcription

**Key concepts**

- Mechanisms of transcription and its regulation are conserved across the domains of life
- The basal machinery, which catalyzes transcription, consists of core RNA polymerases and accessory factors
- The basal machinery is controlled by a regulatory machinery consisting of activators, repressors, coactivators, and corepressors

**Outline**



## 1.1 INTRODUCTION

Proteins control nearly everything that happens in living organisms – they decode genetic information; they control cell shape, cell movement, and segregation of the chromosomes during cell division; they determine the localization of every cellular constituent; and catalyze every essential metabolic reaction. So if we are to understand how organisms work, we need to understand the processes that control when and where each protein is synthesized.

The production of a protein is a multistep process beginning with the synthesis of RNA (transcription), the processing of this RNA, and its transport to the ribosomes where it serves as a template for the synthesis of polypeptides (translation). These polypeptides must then be correctly modified and folded to form mature proteins.

Although the pathway from a gene to a mature protein can be regulated at any step, the most efficient way to regulate gene expression is to modulate the first step in the pathway since this avoids the costly synthesis of unneeded RNA. Organisms therefore produce an immense set of catalytic and regulatory factors dedicated to the task of transcribing genes in an intricately controlled manner. This chapter will introduce the components of this machinery and discuss the extent to which these components have been conserved during the course of evolution. At the same time, readers will become acquainted with some of the vocabulary needed to appreciate the exploration of the transcriptional machinery presented in Chapters 2 through 7.

## 1.2 THE VOCABULARY OF TRANSCRIPTION

Readers of this book should have had some previous exposure to most of the terms given in bold face in this section, and to basic ideas about proteins and nucleic acids through a college level course in biochemistry or molecular biology. Other important terms defined in this section, to which readers may or may not have previous exposure, will be given in italics.

### 1.2.1 RNA biogenesis

**Transcription**, the first step in *RNA biogenesis*, is the DNA template-dependent synthesis of RNA, a process that is catalyzed by **RNA polymerases**. Special DNA sequences termed **promoters**, which are found near the beginning of each transcribed region, direct these enzymes to initiate transcription. *Transcriptional initiation* is followed by *transcriptional elongation*, during which polymerase moves along the DNA, catalyzing the template-directed joining of nucleotides via **phosphoester linkages** to form a full-length *primary transcript* that is complementary in sequence to one of the DNA strands (the "template strand"). Elongation is, in turn, followed by *transcriptional termination*, in which the polymerase encounters a termination signal triggering the release of both the transcript and the DNA from the polymerase.

The job of most RNA molecules is to direct protein synthesis by serving as a template (**mRNA**) or as components of a catalytic machine for decoding this template (**tRNA** and **rRNA**). The production of RNA molecules ready to participate in protein synthesis requires numerous modifications to the primary transcripts. rRNA and tRNA modification occurs in all organisms, while mRNA modification (mRNA processing) is largely restricted to eukaryotes. Eukaryotic mRNA processing includes 5′ *capping* to produce the mature 5′ end of the transcript, *splicing* to remove introns, and *cleavage and polyadenylation* to produce the mature 3′ end. These processing events begin before synthesis of the primary transcript is complete (Figure 1.1).

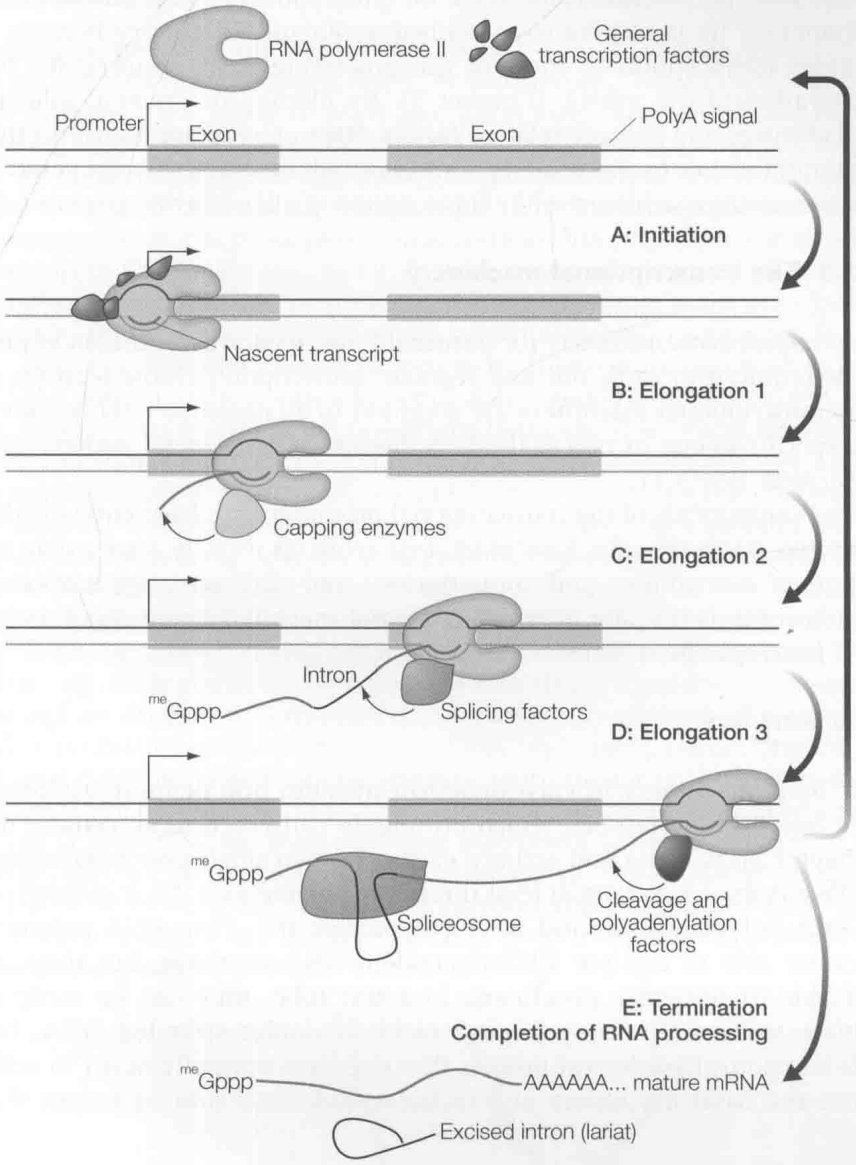



**Figure 1.1** *RNA polymerase II is an mRNA factory.* RNA polymerase II (Pol II) interacts with a myriad of transcription factors and RNA processing factors to coordinate the process of mRNA biogenesis. (A) Promoter recognition and transcriptional initiation by Pol II requires general transcription factors. These proteins form a complex bound to the promoter, which then opens up the DNA exposing the template strand. Although it is not shown, after Pol II leaves the promoter and begins elongation, some of the basal factors remain behind at the promoter, others remain bound to the elongating Pol II, and others are released back into solution. (B–D) Early during the elongation phase, Pol II binds the enzymes that catalyze capping, which then direct the formation of a $^{me}$Gppp cap in which 7-methylguanosine is attached to the 5′ end of the transcript via a 5′ to 5′ triphosphate linkage.

The RNA polymerase responsible for eukaryotic mRNA synthesis (RNA polymerase II) is sometimes described as an mRNA factory because it couples transcription to most of the processing steps required for the maturation of the mRNA (Chapter 3). By binding to capping, splicing, and cleavage and polyadenylation factors, RNA polymerase II ensures that appropriate sites in the primary transcript will be efficiently delivered to the processing machinery while transcription is still occurring (Figure 1.1).

### 1.2.2 The transcriptional machinery

The *transcriptional machinery* (or transcriptional apparatus) is the set of proteins required to carry out and regulate transcription. Some features of the transcriptional machinery are common to all organisms, while others are specific to one or two of the *three domains of life* (*bacteria*, *archaea*, and *eukaryotes*; Box 1.1).

The components of the transcriptional machinery are frequently divided into two categories: the *basal machinery*, which is directly responsible for promoter recognition and transcription, and the *regulatory machinery*, which controls the rate at which the basal machinery carries out its job in a gene-specific manner.

#### *Basal machinery*

The basal machinery in each organism includes one or more multisubunit *core RNA polymerases*, which are highly conserved in all cellular life (Chapter 2). Bacteria and archaea each contain a single core polymerase, while eukaryotes contain at least three core polymerases (RNA polymerase I, RNA polymerase II, and RNA polymerase III). Core RNA polymerases are able to catalyze DNA-dependent RNA synthesis, but they are not able to recognize promoters. In a test tube, they can be made to initiate transcription at gaps and nicks in double-stranded DNA, but this is a non-physiological process that does not normally occur in cells. Thus, the basal machinery also includes additional protein factors that

**Figure 1.1** (*continued*) As elongation proceeds, Pol II binds splicing factors, which then trigger the formation at each intron of a spliceosome (a large ribonucleoprotein complex that catalyzes splicing). Pol II also binds the factors that catalyze cleavage of the transcript at the PolyA signal and the addition of several hundred adenylate residues to the newly formed 3′ end (polyadenylation). (E) After Pol II passes the PolyA signal, termination occurs releasing the transcript and DNA template from Pol II. The splicing and formation of the mature 3′ end may not be completed until after termination. The mature mRNA contains the 5′ cap and 3′ PolyA tail, which help to stabilize the RNA and are required for efficient export and translation of the mRNA. In the process of splicing together the exons, the introns are excised from the transcript and released as lariats.