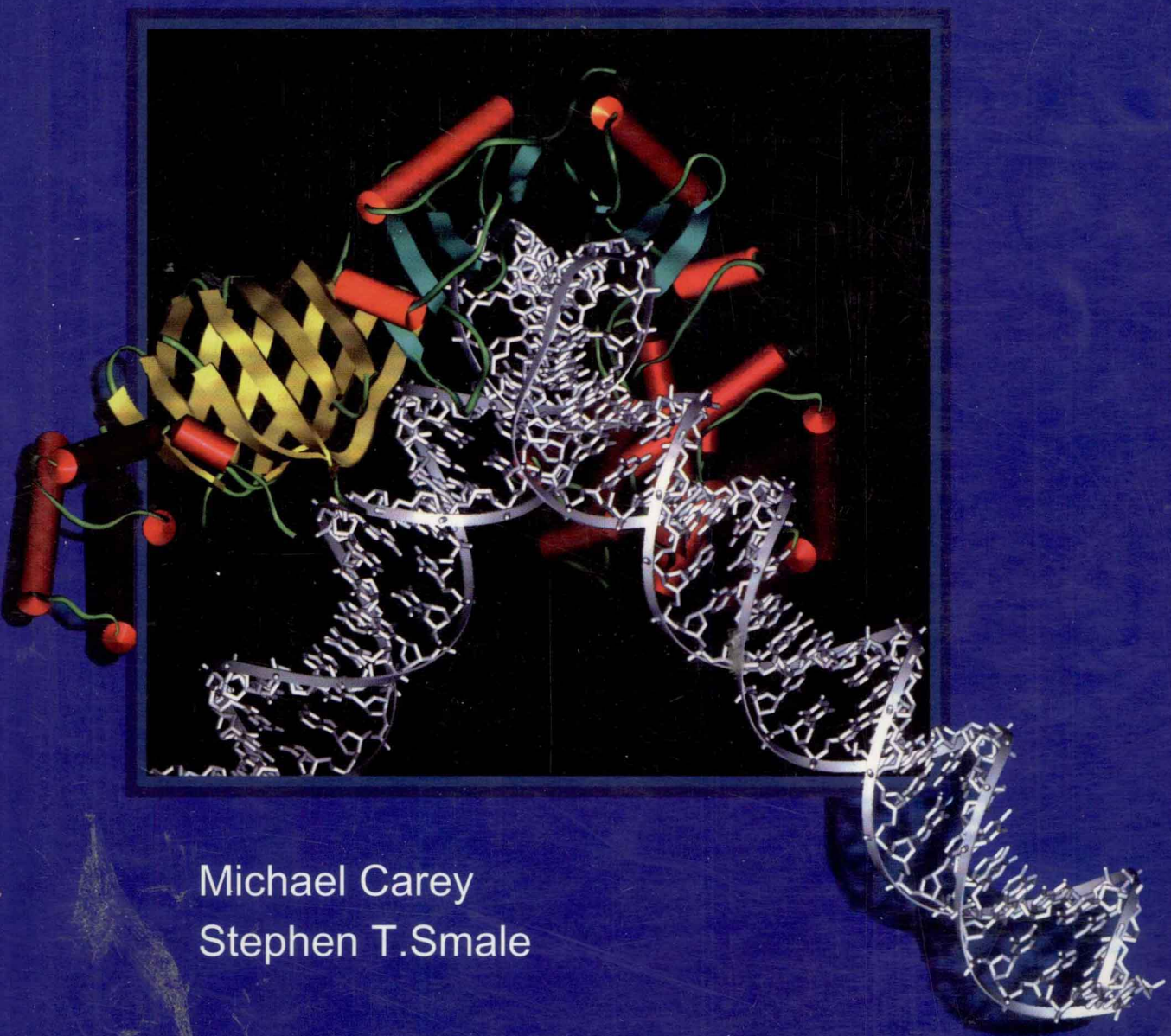
真核生物转录调控

Transcriptional Regulation in Eukaryotes 概念，策略和方法

Concepts, Strategies, and techniques



Michael Carey
Stephen T.Smale

清华大学出版社
http://www.tup.tsinghua.edu.cn
冷泉港实验室出版社

Transcriptional Regulation in Eukaryotes

Concepts, Strategies, and Techniques

Michael Carey
University of California, Los Angeles

Stephen T. Smale
Howard Hughes Medical Institute and University of California, Los Angeles

清华大学出版社

冷泉港实验室出版社

（京）新登字 158 号

Transcriptional Regulation in Eukaryotes

Concepts, Strategies, and Techniques

Michael Carey
Stephen T. Smale

北京市版权局著作权合同登记号： 图字 01-2001-4299

书　名：真核生物转录调控　概念，策略和方法
作　者：Michael Carey
　　Stephen T. Smale
出版者：清华大学出版社（北京清华大学学研大厦，邮编 100084）
　　http:// www.tup.tsinghua.edu.cn
印刷者：清华大学印刷厂
发行者：新华书店总店北京发行所
开　本：850×1168 1/16　印张：42
版　次：2002 年 1 月第 1 版　2002 年 1 月第 1 次印刷
书　号：ISBN 7-302-05073-2/Q・17
印　数：0001～4000
定　价：68.00 元

Preface

Since the advent of recombinant DNA technology three decades ago, thousands of eukaryotic genes have been isolated. The differential expression of these genes is critical for both normal cellular processes and abnormal processes associated with disease. To understand these processes, a growing number of investigators from diverse fields of biology have begun to study the molecular mechanisms regulating gene transcription. Furthermore, the genome projects under way throughout the world have led to the identification of the entire gene complements of *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and numerous archaeal and eubacterial organisms. Within the next few years, the approximately 100,000 genes within the human genome will have been identified. After this goal is realized, the need to dissect mammalian transcriptional control regions and regulatory mechanisms rigorously will increase dramatically.

Despite the global interest in elucidating mechanisms of transcriptional regulation, a comprehensive source of strategic, conceptual, and technical information has not been available for those entering the field for the first time. Although protocols for numerous techniques have been published, the strategic decisions necessary to carry out a step-by-step analysis have not been outlined. This deficiency became apparent to us while we were serving as instructors for the Eukaryotic Gene Expression course held each summer at Cold Spring Harbor Laboratory. This laboratory course was designed for physician-scientists interested in understanding the regulation of a specific disease-related gene, Ph.D. scientists trained in other fields who became interested in the regulatory mechanisms for a gene involved in a particular biological process, and graduate students or postdoctoral fellows who were initiating transcriptional regulation projects. This book is targeted toward this same diverse group of scientists who have developed an interest in transcriptional regulation.

In writing this book, we have focused on issues that the average investigator faces when undertaking a transcriptional regulation analysis, and we have outlined recommended strategies for completing the analysis. One risk of describing a prescribed step-by-step approach is that it may suppress creativity and may not be applicable to all regulatory scenarios. To the contrary, our hope is that our recommendations will enhance creativity by allowing it to evolve from an informed perspective.

We thank the many participants in the Eukaryotic Gene Expression Course from 1994 through 1998 for providing the inspiration and motivation for this book. We also acknowledge our colleagues at UCLA, the members of our laboratories, and our co-instructors for the Eukaryotic Gene Expression course, including Marc Learned, Ken Burtis, Grace Gill, David Gilmour, and Jim Goodrich, for many valuable discussions. We are deeply indebted to a number of colleagues for specific contributions and reading of sections, including Doug Black, Mike Haykinson, Leila Hebshi, Reid Johnson, Ranjan Sen, and Amy

Weinmann. We are particularly grateful to our editor Judy Cuddihy and the book's reviewers, Grace Gill, Bill Tansey, and Steve Hahn, whose generous contribution of time and ideas made the undertaking intellectually rewarding and personally enjoyable. The book was greatly improved by the work of Birgit Woelker and Maryliz Dickerson at Cold Spring Harbor Laboratory Press, as well as Jan Argentine, Pat Barker, and Denise Weiss. Finally, we acknowledge Cold Spring Harbor Laboratory Press Director John Inglis, whose encouragement was essential for the completion of this novel project.

M.C. and S.T.S.

Overview

The goal of this book is to provide a detailed description of the approaches to be employed and issues to be considered when undertaking a molecular analysis of the transcriptional regulatory mechanisms for a newly isolated gene, or a biochemical analysis of a new transcription factor. Our emphasis is on mammalian transcription, which is complicated by the combinatorial nature of regulation and the lack of facile genetics. We refer periodically to studies in yeast, *Drosophila*, and other organisms where more tractable genetic approaches have led to a detailed understanding of particular mechanistic issues. The topics covered in the book extend from the determination of whether a gene is in fact regulated at the level of transcription initiation to advanced strategies for characterizing the biochemical mechanism underlying its combinatorial regulation by activators. Although numerous specialized and detailed techniques are included, the unique characteristics of this book are its strategic and conceptual emphasis on analysis of individual genes and the transcription factors that regulate them.

Chapter 1 reviews the current state of the RNA polymerase II transcription field. This chapter provides an investigator entering the field with a comprehensive introduction into areas of active research and the types of regulatory strategies that will be confronted. We have defined the general properties of known regulatory regions (i.e., enhancers, promoters, silencers), components of the transcriptional machinery (mediator components and the general transcription factors), activators, and repressors. Select review articles and online information sources are included for the novice interested in additional details on the various topics. Emphasis is placed on the role of macromolecular complexes in regulation.

Chapters 2–9 were conceived as a step-by-step guide for an investigator who wants to pursue the regulatory mechanisms for a new gene that has been identified. Chapter 2 presents general strategic issues to consider before the analysis is initiated. First and foremost is a discussion of the goals of the analysis. This topic was included because it has become apparent that many investigators enter the transcription field with unrealistic expectations. Presumably, these expectations arise because a preliminary analysis of a control region, using basic reporter assays and electrophoretic mobility shift assays, is relatively straightforward. To the contrary, a substantial amount of effort is usually required to make meaningful progress toward an understanding of a gene's regulatory mechanisms. Chapter 2 also contains a discussion of the feasibility of achieving the goals. The feasibility is largely dependent on the availability of particular tools, including appropriate cell lines for functional and biochemical studies, and an appropriate functional assay. The chapter concludes with a discussion of whether to begin the analysis by studying the promoter or, alternatively, distant control regions, with a brief description of the initial steps required for each starting point. In this book, the phrase "distant control regions" is used in reference to any control region that is distinct from the promoter, such as enhancers, locus control regions, and silencers.

One issue that will become apparent in Chapter 2 and in all subsequent chapters is that specific protocols are not included for many of the methods described. Instead, references

are given to standard methods manuals, in particular Sambrook et al. (1989) and Ausubel et al. (1994). The intention was to avoid duplication of the valuable information provided in pre-existing manuals and to instead focus on strategic advice. Although the book could have been written without any protocols, since they all can be found in the literature, we chose to include selected protocols for three reasons. First, some of the protocols were chosen because we felt that the reader would benefit from a detailed explanation of the specific steps and history of the methodology, information generally not found in other manuals. Second, in some instances we felt it necessary to provide the reader with a sense of the mechanics of a technique while reading the book. Finally, several protocols were included because of their special nature (e.g., permanganate footprinting, TFIID binding studies) and the fact that no general source exists for such methods.

Chapter 3 continues the step-by-step guide by describing how to determine the mode of regulation for a new gene. At the outset, this chapter emphasizes the fact that the regulation of a biochemical activity does not necessarily mean that the gene encoding the protein is subject to regulation. Alternative possibilities are the regulation of protein synthesis or degradation, or posttranslational regulation of the biochemical activity itself. Furthermore, if the gene is found to be regulated, it is not necessarily regulated at the level of transcription initiation. Rather, it may be regulated at the level of transcription elongation, mRNA stability, pre-mRNA splicing, polyadenylation, or mRNA transport. Because regulation at the level of transcription initiation is most difficult to distinguish from regulation of mRNA stability and transcription elongation, the basic principles of these latter modes of regulation are discussed. Furthermore, strategies for distinguishing between the various modes of regulation are presented, along with a detailed protocol for one important technique, the nuclear run-on.

As stated above, one critical decision discussed in Chapter 2 is whether to begin an analysis of transcriptional regulation by studying the promoter or, alternatively, the distant control regions for the gene. If the investigator opts to study the promoter, the approaches detailed in Chapters 4 and 5 should be followed if the gene is found to be regulated at the level of transcription initiation. Chapter 4 describes methods for determining the location of the transcription start site, an essential first step in every promoter analysis. Four methods for start-site mapping are described, including the primer extension, RNase protection, S1 nuclease, and RACE methods. The advantages and limitations of each method are discussed, and detailed protocols are included for the first three.

Chapter 5 considers the development of a functional assay for a promoter; in other words, the development of an assay that can be used to identify, by mutagenesis (see Chapter 7), the individual control elements required for promoter activity. Transient and stable transfection assays are discussed in detail, including an overview of transfection procedures, reporter genes, vectors, and assays, and the initial design and interpretation of experiments. Alternative functional assays, including in vitro transcription and transgenic assays, are also briefly mentioned, along with their advantages and disadvantages. Chapter 5 is the first of several chapters where the text becomes strongly focused toward a discussion of transcriptional activation, with very little discussion of transcriptional repression. The intention was not to minimize the importance of repression mechanisms for transcriptional regulation; however, a discussion of each point from the perspective of both activation and repression would have been unmanageable. In most cases, it therefore is left to the reader to determine how the principles discussed can be applied to a repression analysis.

If an investigator chooses to pursue distant control regions instead of, or in addition to, the promoter, Chapters 5 and 6 are designed to follow Chapter 3. Chapter 5, as described

above, contains basic information regarding the design of functional assays. This information is applicable to both promoters and distant regions. Chapter 6 describes approaches for identifying distal control regions, including the recommended starting point of performing DNase I hypersensitivity experiments. Chapter 6 also describes special strategies not discussed in Chapter 5 for developing functional assays to analyze distant control regions.

After a functional assay is developed for a promoter (Chapters 4 and 5) or distant control region (Chapters 5 and 6), the next step is to dissect the individual DNA elements constituting the region. These procedures, which usually involve a systematic mutant analysis, are described in Chapter 7. This chapter stresses the benefits of a mutant analysis, but also describes other strategies that may lead to the identification of important DNA elements within a control region.

After the DNA elements are identified, the proteins that bind to them must be identified and their genes cloned. These procedures are described in Chapter 8, beginning with the development of EMSA and DNase I footprinting assays for use with crude nuclear extracts. These assays are discussed in greater detail in Chapter 13 from the perspective of an analysis of a pure recombinant protein. An attempt was made to minimize the duplication of information between these two chapters. However, to maintain the logical progression of the book, some redundancy was unavoidable. Various strategies that can be used to clone the gene encoding a DNA-binding protein are then described, including protein purification, the yeast one-hybrid screen, in vitro expression library screen, mammalian expression cloning, degenerate PCR, and database approaches.

Chapter 9 completes the step-by-step outline of the characterization of a new gene by focusing on a crucial issue: After a factor that binds an important DNA element in vitro is identified, how can one determine whether that factor is indeed responsible for the function of the control element in vivo? Although no experiment is available that can provide conclusive evidence that the protein is functionally relevant, twelve experimental strategies are described that can be used to test the hypothesis. As with all science, the strength of the hypothesis will correspond to the number of rigorous tests to which it has been subjected.

The analysis of a control region, using the strategies described in Chapters 2–9, relies on the use of artificial assays, such as transfection assays and in vitro DNA-binding assays. To complement these approaches, it can be helpful to study the properties of the endogenous control region within its natural environment. Chapter 10 describes experimental strategies for such a characterization, beginning with genomic footprinting and in vivo crosslinking/immunoprecipitation strategies for visualizing specific protein–DNA interactions at the endogenous locus. Chromatin structure is also known to be an important contributor to gene regulation and is best studied in the context of the endogenous locus. Therefore, strategies are included for determining nucleosome positioning and remodeling. Strategies for analyzing DNA methylation status and subnuclear localization of a gene are also briefly discussed.

From a biochemical point of view, an understanding of the mechanism of gene regulation involves recreating regulated transcription in vitro and delineating the precise protein–protein and protein–DNA interactions involved in the process. Chapters 11–15 describe approaches for recreating and studying gene regulation in vitro using purified and reconstituted biochemical systems.

The initial starting point in a biochemical analysis of any regulatory protein is to synthesize the protein and its derivatives in recombinant form. Chapter 11 provides a list of approaches for expressing proteins, and guides the investigator through the strategic and technical decisions encountered in choosing an appropriate system for diverse applica-

tions. The chapter outlines the fundamentals of using *E. coli* to generate small regulatory molecules (e.g., DNA-binding domains of activators and repressors) and baculovirus and retroviral systems to generate multi-protein complexes.

Typically, as an investigator proceeds through different stages of an analysis, it becomes imperative to delineate the protein domains engaged in interactions with other regulatory proteins and with the transcriptional machinery. This information is essential for completing a biochemical analysis of mechanism. The approach employed to gain such insights is termed "structure–function" analysis. This is not a trivial task, and the approach and decision-making are often based on the particular type of regulatory protein being studied. Chapter 12 discusses structure–function analysis from several perspectives. Approaches for studying protein interactions are described briefly to permit the investigator to design specific assays for analyzing the relevant domains. Simple deletion analysis is discussed as a means to delineate how different regions of a regulatory protein contribute to different aspects of DNA binding and transcriptional regulation. This discussion serves as a springboard to more advanced approaches, including domain swapping, a straightforward means to ascribe precise functions to portions of proteins. Most importantly, however, a molecular understanding of transcription is often derived from knowledge of the specific amino acid residues mediating the relevant contacts. Particular emphasis is placed on guiding the investigator through different conceptual approaches to generating site-directed mutants, how such mutants are modeled, and case studies in which mutagenesis is compared with the results of crystal structures. Finally, the chapter discusses the exciting and emerging concept that structural information can be employed to generate novel "altered specificity" genetic systems for analyzing transcriptional mechanisms.

DNA recognition by combinations of proteins is the major contributor to the cell and developmental specificity of a transcriptional response. The mechanisms employed by proteins to bind a promoter or enhancer, both alone and cooperatively with other proteins, are key areas of study in the transcription field. As new transcription factors are identified from the genome project, even more focus will be placed on understanding DNA-binding cooperativity and combinatorial interactions. Chapter 13 describes the fundamentals of equilibrium binding. It introduces the concepts of DNA recognition, describes the chemistry of DNA–protein interactions to the novice, and finally, discusses how chemical and nuclease probes can be employed to generate detailed models for DNA binding. Furthermore, the chapter outlines case studies where models derived from chemical probing are compared with the results of crystal structures of DNA–protein co-complexes. Finally, but most importantly, the chapter provides a basic introduction to the concept and study of nucleoprotein complexes called enhanceosomes, an emerging area of research that underlies the combinatorial action of transcription factors.

Ultimately, the investigator may wish to understand the detailed biochemical steps affected by activators. This goal involves two undertakings: First, development of a robust in vitro transcription system that recreates the regulatory phenomenon in vitro and, second, design of mechanistic experiments with highly specialized reagents including purified transcription factors and chromatin templates. Chapter 14 guides the investigator through the logistical decisions and reagents necessary to design the appropriate reporter templates and to develop active transcription systems. The chapter discusses how in vitro transcription reactions are measured and optimized, including G-less cassettes and primer extension, while expanding on the nuances of in vitro systems presented originally in Chapter 8. Descriptions of the available methods for generating reconstituted systems with crude or pure general factors and Pol II and the development of systems for analyzing chromatin templates are also presented.

Once activators are shown to stimulate transcription in vitro, the investigator may wish to further pursue the biochemical mechanism of activated transcription using purified transcription reagents. This is a rapidly evolving area in terms of both new concepts and specialized reagents. Chapter 15 presents a historical overview of how different methods were originally applied for understanding basal and activated transcription. The chapter then outlines numerous strategies employed to study specific steps in activated transcription using crude and pure reagents. These include approaches for analyzing transcription complex assembly including sarkosyl sensitivity, the immobilized template approach, permanganate probing, and others. The emphasis is on assay development and data interpretation. The chapter also attempts to provide an up-to-date tabulation of sources for specialized reagents including systems for expressing and purifying recombinant transcription factors and multi-component complexes such as the human holoenzyme, chromatin remodeling machines, human mediator, and TFIID.

REFERENCES

Ausubel F.M., Brent R.E., Kingston E., Moore D.D., Seidman J.G., Smith J.A., and Struhl K. 1994. *Current protocols in molecular biology*. John Wiley and Sons, New York.

Sambrook J., Fritsch E.F., and Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Abbreviations and Acronyms

In addition to standard abbreviations for metric measurements (e.g., ml) and chemical symbols (e.g., HCl), the abbreviations and acronyms below are used throughout this manual.

A, adenine
AcPNV, *Autographa californica* polyhedrosis virus
AdMLP, adenovirus major late promoter
AMV, avian myeloblastosis virus
AR, androgen receptor
ARC, activator-recruited co-factor
ARS, autonomous replication sequence
AOX1, alcohol oxidase
ARE, AU-rich response element
ATP, adenosine triphosphate
att site, attachment site

BAC, bacterial artificial chromosome
BEAF, boundary element-associated factor
bHLH, basic helix-loop-helix
BrdU, bromodeoxyuridine
BRE, TFIIB recognition element
BSA, bovine serum albumin
bZIP, basic leucine zipper

C, cytosine
CAP, catabolite activator protein
CAT, chloramphenicol acetyltransferase
CBP, CREB-binding protein
cDNA, complementary DNA
C/EBP, CCAAT enhancer-binding protein
CHD, chromodomain SWI/SNF-like helicase/ATPase domain and DNA-binding domain
CITE, cap-independent translational enhancers
CMV, cytomegalovirus
CREB, cAMP receptor element binding protein
cRNA, complementary RNA
cs, cold sensitive
CTD, carboxy-terminal domain
CTP, cytosine triphosphate

DAN, deadenylating nuclease
dATP, deoxyadenosine triphosphate
dCTP, deoxycytidine triphosphate
DEPC, diethyl pyrocarbonate
dGTP, deoxyguanosine triphosphate
DHFR, dihydrofolate reductase
DMP, dimethyl pimelidate dihydrochloride
DMS, dimethyl sulfate
DMSO, dimethyl sulfoxide
DPE, downstream core promoter element
DR, direct repeat
DRB, 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole
DTT, dithiothreitol

ECMV, encephalomyocarditis virus
EBV, Epstein-Barr virus
EDTA, ethylenediaminetetraacetic acid
EKLF, erythroid Kruppel-like factor
EMCV, encephalomyocarditis virus
EMSA, electrophoretic mobility shift assay (gel shift)
ES, embryonic stem (cells)
EST, expressed sequence tag
ETL, early-to-late promoter
***Exo*III**, exonuclease III

FACS, fluorescence-activated cell sorting
FISH, fluorescence in situ hybridization
FBS, fetal bovine serum

β-gal, β-galactosidase
G, guanine
GFP, green fluorescent protein
GTFs, general transcription factors
gpt, guanine phosphoribosyltransferase
GST, glutathione-*S*-transferase
GTP, guanosine triphosphate

H_2O_2, hydrogen peroxide
HA, hemagglutinin
HAT, histone acetyltransferase
HCF, host cell factor
HEBS, HEPES-buffered saline
HEPES, *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid
HisD, histidinol dehydrogenase
HIV, human immunodeficiency virus
HIV-1, human immunodeficiency virus type 1
HKLM, heat-killed *Listeria monocytogenes*
HLH, helix-loop-helix
HMBA, hexamethylene bisacetamide

HMG, high mobility group
HMK, heart muscle kinase
hpi, hours post induction
HPLC, high-performance liquid chromatography
HS, hypersensitive
hsp70, heat shock protein 70
HSTF, heat shock transcription factor
HSV, herpes simplex virus
HSV-1, herpes simplex virus type 1
HSV-TK, herpes simplex virus thymidine kinase

IFN-β, interferon-β
Ig, immunoglobulin
IgM, immunoglobulin heavy-chain protein
IL-2, interleukin-2
IL-12, interleukin-12
Inr, initiator elements
int, integrase
IPTG, isopropyl-β-D-thiogalactoside
IRE, iron-responsive element
IRP, iron-regulating protein
ISWI, imitation SWI

LCR, locus control region
LIS, lithium diiodosalicylate
LM-PCR, ligation-mediated PCR
LPS, lipopolysaccharide
LTR, long terminal repeat

M, molar
MAR, matrix attachment region
MBP, maltose binding protein
MEL, mouse erythroleukemia (cells)
MMLV, Moloney murine leukemia virus
MMTV, mouse mammary tumor virus
MNase, micrococcal nuclease
moi, multiplicity of infection
MOPS, 3-(*N*-morpholino) propanesulfonic acid
MPE, methidium propyl EDTA
mRNA, messenger RNA
MTX, methotrexate

NAT, negative activator of transcription
NER, nucleotide excision repair
neo, aminoglycoside phosphotransferase
NHP, nonhistone proteins
Ni-NTA, nickel-nitriloacetic acid
NMR, nuclear magnetic resonance
NP-40, Nonidet P-40

NTP(s), nucleotide triphosphate(s)
NURF, nucleosome remodeling factor
O_L, leftward operator
O_R, rightward operator
OH-radical, hydroxyl-radical
ONPG, O-nitrophenyl-β-D-galactopyranoside
OP-Cu, Cu-phenanthroline
ORC, origin recognition complex
ori, origin of replication

P_R, promoter in rightward direction
P_{RM}, promoter for repressor maintenance
PAGE, polyacrylamide gel electrophoresis
PAN, poly(A) nuclease
PBS, phosphate-buffered saline
pc, positive control
PCR, polymerase chain reaction
PCV, packed cell volume
PEG, polyethylene glycol
PEI, polyethylenimine
PIC, preinitiation complex
PIPES, piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)
PMSF, phenylmethylsulfonyl fluoride
PNK, polynucleotide kinase
PPAR-γ, peroxisome proliferator-activated receptor-γ

RACE, rapid amplification of cDNA ends
RAR, retinoic acid receptor
Rb, retinoblastoma
rbs, ribosome binding site
RDA, representative difference analysis
RNA, ribonucleic acid
RNAP, RNA polymerase
RRE, Rev-responsive element
RSV, Rous sarcoma virus
RT, reverse transcriptase
RT-PCR, reverse transcription polymerase chain reaction

SAAB, selected and amplified binding site analysis
SAGA, SPT-ADA-GCN acetyltransferase
SAR, scaffold-associated region
SCAP, SREBP cleavage-activating protein
SDS, sodium dodecyl sulfate
SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEAP, secreted alkaline phosphatase
SMCC, SRB MED co-activator complex
sn, small nuclear
SPR, surface plasmon resonance

SRB, suppressor of RNA polymerase B
SREBP-1, sterol response element binding protein
SSC, standard saline citrate
Su(Hw), suppressor of hairy wing

Tac, Trp-Lac (promoter)
TAE, Tris-acetate-EDTA
TAF, TBP-associated factor
TAg, T antigen
TBE, Tris-borate-EDTA
TBP, TATA-box binding protein
TCR-α, T-cell receptor-α
TdT, terminal transferase
TE, Tris/EDTA buffer
TES, *N*-Tris[hydroxymethyl]methyl-2-amino ethane sulfonic acid
Tet, tetracycline
TetR, Tet repressor
TFII, transcription factor for Pol II
TICS, TAF_{II}- and initiator-dependent cofactors
TK, thymidine kinase
TLC, thin-layer chromatography
TR, thryoid receptor
TRAP, TR-associated protein
TRF, TFIID-related factor
tRNA, transfer RNA
TRRD, transcription regulatory region database
TSA, trichostatin A

U, uracil
UAS, upstream activating sequence
UAS_G, galactose upstream activating sequence
USA, upstream stimulatory activity
UTL, untranslated leader
UTP, uridine triphosphate

VAF, virus-inducible transcription activator complex
VHL, von Hippel-Lindau
VSV, vesicular stomatitis virus

WCE, whole cell extract

Xis, excision protein

YAC, yeast artificial chromosome

Contents