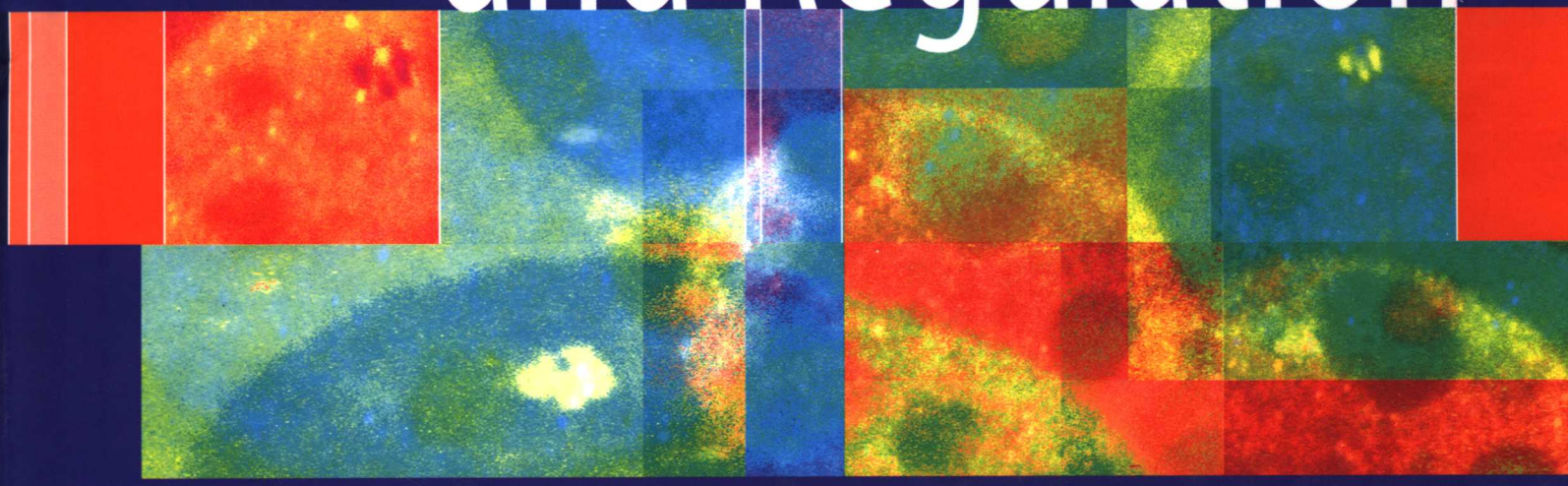
Jun Ma *Editor*

Gene Expression and Regulation



高等教育出版社
Higher Education Press

Jun MA

Division of Developmental Biology
Cincinnati Children's Hospital Research Foundation
University of Cincinnati College of Medicine
3333 Burnet Avenue
Cincinnati, OH 45229
USA

E-mail: *Jun.ma@cchmc.org*

图书在版编目（CIP）数据

基因的表达与调控=Gene Expression and Regulation/马骏主编.
—北京：高等教育出版社，2006
ISBN 7-04-017675-0

Ⅰ.基… Ⅱ.马… Ⅲ.基因表达—研究—英文 Ⅳ.Q753

中国版本图书馆 CIP 数据核字（2005）第 140172 号

图字：01-2006-0366

ISBN 7-04-017675-0

Printed in P. R. China

About the Editor

Jun Ma is an Associate Professor at the Division of Developmental Biology, Cincinnati Children's Hospital Research Foundation and University of Cincinnati College of Medicine. He graduated from Peking University in 1982, majoring in Biology. He did his graduate work with Mark Ptashne at the Department of Biochemistry and Molecular Biology in Harvard University, and was a Junior Fellow at the Harvard Society of Fellows between 1989–1992. He spent the summer of 1988 in the laboratory of Christiane Nüsslein-Volhard at the Max-Planck-Institute for Developmental Biology in Tübingen to collaborate with Wolfgang Driever. He joined the faculty of the University of Cincinnati College of Medicine in 1992 and has remained there since. Currently he also has a collaborative base at the Institute of Biophysics of the Chinese Academy of Sciences in Beijing. His earlier work on the yeast activator GAL4 helped pave the way to the development of the yeast two-hybrid system. His current research focuses on the mechanisms of transcription control and development in *Drosophila*.

Preface

All genes must be expressed to exhibit their biological activities. How genes are expressed and regulated is a central question in molecular biology and our knowledge in this area has been expanding enormously in recent years. The complexity of gene regulation is compounded by the fact that gene activities reach every corner of biology. Transcription is universally the first step toward expressing a gene. It is a highly regulated process. Understanding the molecular mechanisms of transcription regulation is of fundamental importance. For protein-coding genes, post-transcriptional steps, including pre-mRNA processing, mRNA transport and translation, can also play important roles in regulating gene expression. To contain the scope of this book, we will focus primarily on RNA polymerase II transcription and regulation. We will explore not only the biochemical basis of transcription but also the biological consequences of, and biological influences on gene transcription.

The book is composed of 35 individual review articles written by authorities in the field. The chapters are organized into five sections: The History, The Machinery, The Regulators, The Genome, and Special Topics. The History section contains one chapter, written by James Goodrich and Robert Tjian, who provide an excellent historical perspective and overview of the transcription process. The Machinery section has six chapters that cover essential topics on the transcriptional apparatus, general cofactors, chromatin structure, and core promoter structure. The Regulators section has thirteen chapters. While the first two of them investigate the mechanisms of transcriptional activation and repression, the remaining eleven chapters discuss in depth selected gene-specific transcription factors that play critical roles in a variety of biological processes, including STATs, Smads, NFκB, nuclear receptors, NFAT, Rb, p53, HIV Tat, ATFs, c-Jun and Hox proteins. The Genome section contains six chapters that examine topics relevant to transcription regulation and genome behavior, including chromatin boundaries, heterochromatin, DNA methylation, genomic analysis, genomic integrity, and cell death. Finally, the Special Topics section contains nine chapters that investigate such important issues as pre-mRNA splicing, DNA supercoiling, microRNA, transcription factor dynamics, role of actin in transcription, gene therapy, and transcription regulation in bacteria, plants and developmental signaling.

When Higher Education Press invited me to write a textbook for their Current Scientific Frontiers book series two years ago, I did not think I had the time needed to tackle such a big project. Instead, I made a proposal—endorsed quickly by HEP—to explore the possibility of editing a book (resembling a textbook style) on the topic of gene expression and regulation, with individual review articles written by experts in the field. Without the enthusiastic support and generous commitment from the contributors, this project would have never even started. I am deeply indebted to all of them. Every chapter in this book is a scholarly work reflecting numerous hours of intense efforts of the contributors. I would like to express my special thanks to Cheng-Ming Chiang for generously contributing two excellent chapters, a few contributors for kindly agreeing to write on relatively short notice, and Gordon Hager for providing the cover photo and design suggestions. I would also like to thank HEP for their flexibility and trust in this project, and the HEP and Springer editorial and design teams, in particular Li Shen at HEP, for their excellent work. Finally, I would like to thank Bingxiang Li at HEP for the countless email communications and her hard work—at every step along the way—that made this book a reality.

Jun Ma
Cincinnati, USA
November 18, 2005

Contents

Section V Special Topics

Section I

The History

Chapter 01

Transcription: The Never Ending Story

James A. Goodrich[1] and Robert Tjian[2]

[1]*Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309*
[2]*Molecular and Cell Biology Department, University of California, Berkeley, Berkeley, CA 94720*

Key Words: transcription, promoter, activator, coactivator, general factors, chromatin, RNA polymerase II

Summary

After more than 30 years of intense and sustained activity, the field of transcriptional control in eukaryotes continues to deliver unexpected and revealing montages of the remarkably complex yet elegant consequences of evolution. Transcription research started from humble beginnings with the isolation of 3 distinct RNA polymerases. This was followed by a rich period of mapping promoters, enhancers and the isolation of the first sequence specific DNA binding regulatory factors. These studies in turn led to the unraveling of the multi-subunit pre-initiation apparatus culminating with the modern era of co-activators and chromatin remodeling complexes. Throughout this opus of biochemical discovery we have witnessed a beautiful convergence of *in vitro* biochemical tour-de-force combined with the power of molecular genetics and cell biology. In this short preamble, we offer a brief and very likely incomplete history of the maturing of eukaryotic transcription and its prospects for the future.

Fumbling in the Dark: Hoping for Simplicity

Emboldened by the inspiring successes of pioneering work in the biochemistry of DNA replication and bacterial phage transcription, early workers struggling with animal and human gene regulation followed suit by isolating not one but three distinct enzymes: RNA polymerase I, II and III each dedicated to the synthesis of rRNA, mRNA, and tRNA/5sRNA respectively (Krebs and Chambon, 1976; Sklar *et al.*, 1975). However, due to the lack of promoter specific DNA templates or the ability to obtain sufficient quantities of "cloned" DNA, the ability of these 3 distinct enzymes to discriminate between the different classes of genes remained obscure. Nevertheless, the chromatographic separation and *in vitro* biochemical assays for detecting the RNA polymerases opened the first doors to the future development of high fidelity promoter specific and eventually activator regulated transcription in cell free systems.

Because, eukaryotic RNA polymerases behaved in a rather promiscuous and DNA template independent fashion *in vitro*, there was a brief period, (after the discovery of heterogeneous nuclear RNA) in which it was popular to posit that, unlike bacterial transcription which is temporally regulated by cascades of σ-factors, eukaryotic transcription may be "unregulated". Instead, one imagined that post transcriptional RNA processing (i.e. splicing, poly A addition, capping, etc.) would largely determine the population of mRNA's destined for gene product expression. Although this "random transcription" model fit with some early data regarding the apparent lack of promoter DNA selectivity *in vitro* of eukaryotic RNA polymerases, it soon became clear from studies of mammalian viruses (SV40, Adeno 2) that at the very least, specific DNA sequences that lie near transcription start sites (i.e. TATA elements and GC boxes) played some role in determining elements of the eukaryotic "promoter" (Fig.1.1) (Myers *et al.*, 1981; Rio *et al.*, 1980; Tjian, 1978).

As is often the case with biology in general but especially in the study of eukaryotic transcriptional regulation, we invariably opted for simplicity and hoped that a well defined −35/−10 like element such as the

Corresponding Author: Robert Tjian, Tel: (510) 642-0884, FAX: (510) 643-9547, E-mail:jmlim@Berkeley.edu

popular TATA box of Ad2 would suffice to designate the necessary cis-regulatory information of a promoter (Corden *et al.*, 1980; Hu and Manley, 1981). This rather minimalist view was, however, decisively toppled when both *in vitro* and cell based assays were developed that revealed the existence of important upstream distal as well as proximal DNA sequences in eukaryotic promoters (Banerji *et al.*, 1981; Benoist and Chambon, 1981; Fromm and Berg, 1983; Gidoni *et al.*, 1985; Myers and Tjian, 1980; Picard and Schaffner, 1984). With the emergence of cloned promoter sequences and DNA template dependent *in vitro* transcription reactions measured by run-off and primer-extension assays the combinatorial nature of multiple cis-control elements of eukaryotic gene regulatory units became firmly embedded (Mitchell and Tjian, 1989). Add to these *in vitro* assays the advent of transient transfection assays and microinjection in animal cells that revealed the existence of "orientation and distance independent" enhancer elements and we began, for the first time, to get a glimpse of the complex regulatory network of gene transcription that would follow in succeeding decades (McKnight and Tjian, 1986; McKnight, 1982; Picard and Schaffner, 1984; Treisman *et al.*, 1983). To this day, the precise mechanisms mediating "long distance" enhancer or silencer functions remain largely obscure despite many plausible models including DNA looping, scanning etc.

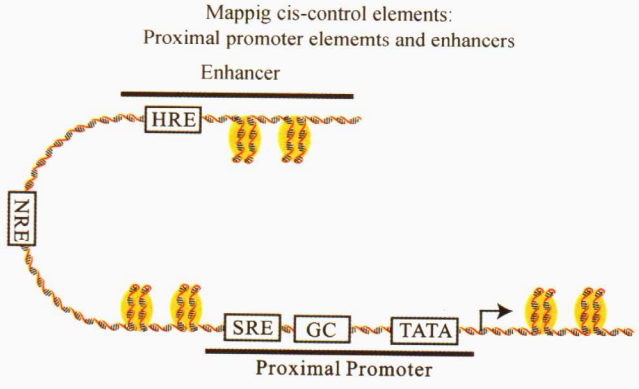


Fig.1.1 Cis-control elements in RNA polymerase II promoters can be located near the transcription start site or at great distances away. Some of the cis-control elements identified by early studies are shown. Abbreviations: TATA, TATA box; GC, GC box; SRE, sterol response element; NRE, nuclear hormone receptor response element; HRE, heat shock response element.

A Shaft of Light: Sequence Specific Transcription Factors

After a flurry of intense promoter bashing experiments with all manner of DNA templates, cell-types and gene systems, we were confronted with the daunting task of determining what was actually recognizing and keying off these composite arrays of cis-control DNA sequences to govern gene specific transcription. One important step along this pathway of discovery was the rapid deployment of various elegant *in vitro* mutagenesis techniques such as linker scanning clustered point mutations and deletions (McKnight *et al.*, 1981; McKnight and Kingsbury, 1982; Myers and Tjian, 1980). At the same time, powerful new biochemical assays such as DNase I footprint protection were being developed (Galas and Schmitz, 1978). Perhaps the single most influential strategy for those of us attempting to dissect the molecular identity of transcriptional regulatory factors was the promoter selective *in vitro* transcription assay (Manley *et al.*, 1980; Rio *et al.*, 1980; Weil *et al.*, 1979; Wu, 1978). This "bucket biochemistry" approach allowed us to use cloned DNA fragments containing well mapped and carefully defined promoters to drive accurate and factor dependent transcription by partially purified RNA polymerases. The tacit assumption in establishing such *in vitro* promoter dependent assays was that purified eukaryotic RNA polymerase II was necessary but not sufficient to direct accurate initiation of transcription. We therefore assumed that one or more additional transcription factors (whose identity and mode of action had remained unknown) was needed in order to instruct or otherwise impart upon RNA pol II the ability to discriminate one promoter from another. Indeed, since no such cellular factors in eukaryotes had yet been identified or isolated in 1980, we had little clue as to the biochemical properties of such factors (i.e. were these factors proteins, nucleic acids, carbohydrate, etc.?). The closest candidate at that time was the SV40 T-ag, a viral encoded protein that displayed many of the hallmarks of a bona fide promoter recognition factor (Rio *et al.*, 1980; Tjian, 1978). Also, whether they would directly bind RNA polymerase á la σ-factors or they would behave more like CAP in the lac operon system and bind DNA in a sequence specific manner was a big question.

Indeed, one of the unappreciated and hidden advantages of using fairly crude nuclear extracts (i.e. from Hela cells or *Drosophila* embryos) to carry out systematic biochemical "complementation" tests *in vitro* allowed us the freedom to be unbiased and simply search for whatever molecules stimulated transcription

of one promoter but not another. Using this approach, factors such as Sp1 were first identified as functional transcriptional activators for RNA pol II that could discriminate, for example, between the SV40 and AdML promoters (Fig.1.2) (Carthew *et al.*, 1985; Dynan and Tjian, 1983a; Dynan and Tjian, 1983b; Sawadogo and Roeder, 1985). Similar biochemical fractionation and *in vitro* assays led to the isolation of TFIIIA for Pol III and UBF for Pol I (Engelke *et al.*, 1980; Learned *et al.*, 1985; Learned *et al.*, 1986; Pelham and Brown, 1980; Wu, 1978). However, it did not take long given the availability of various discriminating DNA binding assays available at that time to determine that these transcription factors were indeed sequence specific DNA binding "activators". And thus, there was a nice alignment of cis-regulatory elements and DNA binding transcription factors. We anticipate that a similar biochemical dissection and reconstitution of *in vitro* transcription reactions that are responsive to distal enhancers, tethering elements, silencers and boundary elements are still needed to fill-in our gaps of knowledge vis-à-vis the molecular players and mechanisms that govern "long distance" regulation so prevalent in metazoan organisms.

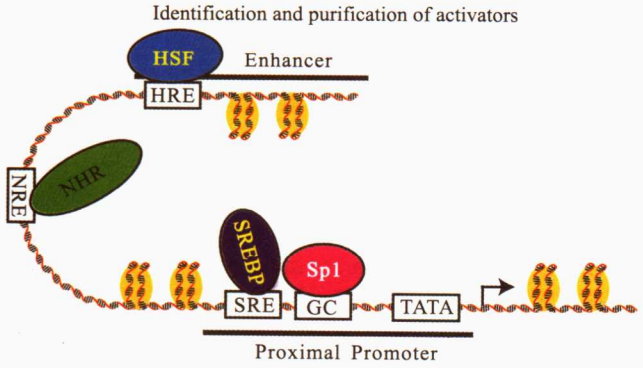


Fig.1.2 Trans-acting factors bind to RNA polymerase II promoters. Abbreviations: Sp1, specificity protein 1; SREBP, sterol response element binding protein; NHR, nuclear hormone receptor; HSF, heat shock factor.

A New Era of Transcription Biochemistry Arrives: Clone, Sequence, Express & Reconstitute

The next big hurdle was to actually purify, clone, and characterize these seemingly powerful transcriptional activators. As often happens in emerging fields, advances in concepts and techniques must go hand in hand. For the transcription field, the development of sequence specific DNA affinity chromatography and a host of affiliated techniques revolutionized our capacity to detect, purify and clone the genes encoding sequence-specific transcription factors (Briggs *et al.*, 1986; Jones *et al.*, 1985; Kadonaga *et al.*, 1987; Kadonaga and Tjian, 1986). Once the genes encoding the first few bona fide transcriptional activators (and repressors) such as Sp1, TFIIIA, CTF, AP1, GCN4, Gal4, GR, and HSF were characterized — a flood of paradigm shifting concepts emerged (Berg, 1988; Bohmann *et al.*, 1987; Courey *et al.*, 1989; Kadonaga *et al.*, 1987; Kadonaga *et al.*, 1988; Mermod *et al.*, 1989; Miller *et al.*, 1985; Mitchell *et al.*, 1987; Triezenberg *et al.*, 1988; Turner and Tjian, 1989). For instance, the remarkably modular nature of transcriptional activators was revealed (Ma and Ptashne, 1987a; Ma and Ptashne, 1987b). The subsequent cloning and sequencing of transcription factors rapidly advanced our ability to recognize DNA binding motifs (i.e. Zn finger, B-HLH, homeodomains, etc.) dimerization domains (LZ, histone folds) activation domains (gln-rich, acidic, etc.) and regulatory/ligand binding domains (AF2).

Initially, as a result of the pioneering work on transcription factor structures derived from studies of the λ-repressor and other phage and bacterial transcription factors (Anderson *et al.*, 1985; Wharton *et al.*, 1984), there was a tendency to assume that all transcription factors would utilize a helix-turn-helix DNA binding domain and an "acidic" activation domain. However, the structure/ function analysis of eukaryotic transcription factors such as Sp1, TFIIIA, steroid receptors, Jun/Fos AP1, C/EBP, CTF etc. quickly dispelled the over-simplified notion that there were only one or two motifs for DNA binding and transcription activation (Gill and Ptashne, 1988). Indeed, it became clear that in eukaryotes and especially metazoan organisms, the repertoire of structural domains that had evolved to accommodate transcriptional specificity was astoundingly diverse and elaborate.

One of the most impressive accomplishments during this rich middle period (1985 — 1995) of transcription research was not only the rapid identification, cloning and characterization of hundreds of sequence specific transcription factors, but also a quantum leap in our understanding of the relationship between function and structure — particularly with regards to DNA binding motifs (Pabo and Sauer, 1992). The high resolution X ray structures of countless DNA binding domains were solved and this rich body of information continues to provide a basis for rapid genome wide functional analysis of novel gene products. The discovery of thousands of different transcriptional activators (repressors) and their pivotal role in complex biological processes such as anterior-posterior and dorsal-ventral patterning in metazoans firmly cemented

the importance of this vast family of proteins. Indeed, after the first few different metazoan genomes were determined, it became apparent that between 5%~10% of the coding capacity of eukaryotes is devoted to encoding such transcriptional regulators. These findings provided another inexorable clue to the essential, universal, and yet diverse nature of transcriptional control mechanisms. However, despite this exponential growth in knowledge about transcription factors, not everything was rosy or well understood about transcriptional regulation. Indeed, although DNA binding motifs and their structures had proven to be highly informative with respect to structure/function relationships, a similar understanding of activation domains was sorely lacking and largely remains so even today.

Mix and Match: Combinatorial Control, Modularity, and Enhanceosomes

As activators and genes were being characterized in greater detail, it became apparent that the simple paradigm of a single activator or single repressor controlling transcription of a gene, as was the case in some bacterial and even yeast systems, did not apply in higher eukaryotes. The regulatory regions of mammalian and *Drosophila* genes, enhancers and silencers, contain binding sites for many transcriptional regulators. An enhancer might bind 10 or more DNA binding factors, including many different activators as well as multiple copies of a single activator. This complexity was further amplified by the presence of large activator families (Homeo-box, FOXO, AP1 etc.) in which individual members had similar DNA binding specificities, but distinct activation domains and presumably different functions (Mitchell and Tjian, 1989). Combinatorial control and the notion of cis-regulatory networks help explain observations indicating that it is the precise complement of activators and repressors present at a promoter that gives rise to gene specific activation in a spatial and temporally regulated pattern (DeFranco and Yamamoto, 1986; Diamond *et al.*, 1990). Cooperativity in DNA binding and synergy in transcriptional activation further contribute to an uncanny level of control over gene transcription. Our understanding of enhancers and activators was substantially advanced with the detailed characterization of the interferon-β and T-cell receptor α enhancers, where the correct function of the enhancers requires not only the presence of the appropriate array of transcriptional activator proteins but also the association of architectural proteins, and the proper spatial orientation of all of these factors dictates the ultimate outcome (Giese *et al.*, 1992; Giese *et al.*, 1995; Thanos and Maniatis, 1995).

Unimagined complexity: The General Transcription Factors, PIC formation, and Promoter-Specific Transcription

From early *in vitro* studies it was realized that while core RNA polymerase II was capable of synthesizing an RNA product, it required additional factors to initiate transcription at specific promoters (Weil *et al.*, 1979). The general concept of dissociable and essential transcription factors had been firmly established in bacteria, where core RNA polymerase required a sigma subunit for promoter-specific transcription. While this paradigm provided a useful framework for studying eukaryotic transcription, the requirement for a single sigma-like subunit was quickly dispelled in eukaryotic transcription systems. Employing biochemical fractionation and promoter specific DNA templates to drive *in vitro* transcription reactions, an unexpectedly large number of critical accessory factors were painstakingly teased out and characterized, initially as crude fractions eluted from columns (Matsui *et al.*, 1980). Of course, like any good biochemist, once you have an assay, next you want to purify the critical activity, characterize its biochemical properties, and identify the gene encoding the factor. After many hundreds of researcher years, all of the general factors and their genes from human, *Drosophila* and yeast eventually were isolated (Aso *et al.*, 1992; DeJong and Roeder, 1993; Eisenmann *et al.*, 1989; Finkelstein *et al.*, 1992; Fischer *et al.*, 1992; Ha *et al.*, 1991; Hahn *et al.*, 1989a; Hahn *et al.*, 1989b; Hoey *et al.*, 1990; Horikoshi *et al.*, 1989; Kao *et al.*, 1990; Ma *et al.*, 1993; Peterson *et al.*, 1991; Peterson *et al.*, 1990; Schaeffer *et al.*, 1993; Shiekhattar *et al.*, 1995; Sopta *et al.*, 1989; Yokomori *et al.*, 1993). Thus, the general or basal transcription factors TFII-A, -B, -D (TBP), -E, -F, and -H were identified (Fig.1.3).

The general transcription factors were unlike the sequence specific activators in that most of them showed little or no propensity to bind DNA in a sequence dependent manner, but instead associated with RNA polymerase II and participated in complex ways towards the assembly of the pre-initiation complex (PIC). Among this large clan of general transcription factors—one that stood out early on was the fraction originally designated TFIID which revealed a weak tendency to bind TATA elements (Reinberg *et al.*, 1987; Sawadogo and Roeder, 1985). Attempts to purify and

characterize this activity proved to be particularly intransigent. After many attempts and failures on the part of several labs, through a combination of persistent biochemistry and fortuitous genetics—the all important TATA binding protein (TBP) was isolated and cloned in the late 1980s (Hahn *et al.*, 1989b; Hoey *et al.*, 1990; Horikoshi *et al.*, 1989; Kao *et al.*, 1990; Peterson *et al.*, 1990). TBP, the central subunit of the TFIID complex, itself has the ability to bind specifically to TATA box elements found in many, but by no means most RNA polymerase II promoters. The surprising observation that the single subunit TBP could replace the crude TFIID function in directing preinitiation complex assembly and basal transcription *in vitro* enabled biochemical experiments to establish an order of assembly for the preinitiation complex—TBP, TFIIA, TFIIB, TFIIF/RNA polymerase II, TFIIE, and TFIIH (Buratowski *et al.*, 1989; Buratowski *et al.*, 1988; Flores *et al.*, 1992). In later *in vitro* experiments, RNA polymerase II was also found in larger complexes containing some of the general transcription factors and provided an alternative mode of preinitiation complex assembly (Koleske and Young, 1994), in which a RNA polymerase II complex is recruited to TFIID and TFIIA pre-assembled on promoter DNA.

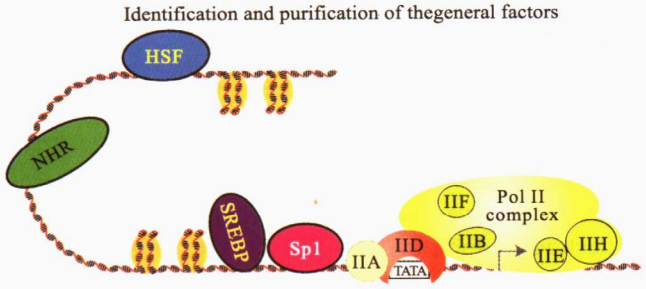


Fig.1.3 The RNA polymerase II general transcription machinery loads into preinitiation complexes encompassing the start site of transcription.

Many surprises surfaced during studies of the general transcription factors. Some of these factors functioned in multiple stages of the transcription reaction—TFIIF is required for initiation and stimulates elongation (Flores *et al.*, 1989; Saltzman and Weinmann, 1989). Others had enzymatic activity—a subunit of TFIIH is a kinase (Feaver *et al.*, 1994; Lu *et al.*, 1992; Roy *et al.*, 1994; Serizawa *et al.*, 1995), two others are helicases (Schaeffer *et al.*, 1994; Schaeffer *et al.*, 1993), and the largest subunit of TFIID has kinase (Dikstein *et al.*, 1996a), acetyltransferase (Mizzen *et al.*, 1996), and ubiquitin-activating/conjugating (Pham and Sauer, 2000) activities. While not detected initially, some of the general transcription factors are now known to bind core promoter DNA with sequence specificity—TFIIB binds the BRE (Lagrange *et al.*, 1998) and subunits of TFIID other than TBP bind the initiator and DPE (Burke and Kadonaga, 1997; Kaufmann and Smale, 1994; Verrijzer *et al.*, 1994). These observations lead us to wonder what other functions will be discovered in future studies of the general transcription factors.

As the general transcription factors were discovered and some of their functions revealed, an intense interest mounted in uncovering the three dimensional structures of these critical factors. TBP was found to be "saddle shaped", with the underside making intricate contacts with and dramatically bending the TATA DNA, while the upper surface presented itself for numerous interactions with other proteins (Kim *et al.*, 1993a; Kim *et al.*, 1993b; Nikolov *et al.*, 1992). This was followed by crystallography studies that unveiled the molecular architecture of complexes contain TFIIB and TFIIA along with TBP and DNA (Geiger *et al.*, 1996; Nikolov *et al.*, 1995; Tan *et al.*, 1996). Larger complexes, such as TFIID and TFIIH have been envisioned using electron microscopy, which revealed the overall shape of these massive entities (Andel *et al.*, 1999; Brand *et al.*, 1999; Chang and Kornberg, 2000; Schultz *et al.*, 2000). The RNA polymerase II enzyme itself was the focus of a major structural effort that began with EM and ultimately surrendered to X-ray crystallography (Cramer *et al.*, 2000; Darst *et al.*, 1991; Gnatt *et al.*, 2001). The amazing structures that have so far resulted from this endeavor provide an unimaginably intricate view of the polymerase alone and bound to other molecules. The value of these structures in opening new lines of research is incredible and these initial glimpses leave us longing for high resolution structures of larger and more elaborate transcription complexes.

The Paradox of Transcriptional Activation: Insufficiency of the General Machinery

With studies rapidly progressing on transcriptional activators and the general transcription machinery a number of labs began to puzzle over the nagging question: How does a sequence specific transcription factor such as Sp1 actually promote the initiation of transcription by RNA pol II? This simple question would eventually lead to a most elaborate and unexpected molecular landscape that today dominates our thinking about how specific mechanisms of transcriptional control are executed in temporally and spatially restricted programs of gene expression. By

1989, we thought all of the molecular components necessary to form an active PIC were in hand. Thus, it seemed a simple matter to reconstitute *in vitro* transcription with purified TBP, RNA Pol II, TFII-A, -B, -E, -F, and -H. Indeed, this constellation of factors isolated either from yeast, *Drosophila* or HeLa cells efficiently produced accurately initiated transcription on any number of well defined promoters. However, to our consternation and frustration, when attempts were made to reconstitute "regulated" transcription using factors such as Sp1, VP16, NR, etc—there was no response to activators in these *in vitro* reactions—contrary to what was expected from *in vivo* studies (Fig.1.4). Thus began a new chapter in the transcription story—the hunt for the elusive co-activators, which can be categorized as the 3rd class of transcription factors, the other two being the sequence specific DNA binding factors and the general factors. Again, relying on our old standby strategy of *in vitro* biochemical complementation as well as genetics in yeast using model activators such as Sp1 or Gal4, the first evidence for a new class of transcription factors emerged — loosely named co-activators and mediators (Berger *et al.*, 1990; Kelleher III *et al.*, 1990; Pugh and Tjian, 1990). Members of this new class of factors were not required for accurate transcription initiation but provided a key function in transcriptional activation, and perhaps a link between DNA binding activators and the core transcription machinery. After much additional research, we now know that many co-activators (and co-repressors) exist, with two groups of proteins playing critical roles in activation of many if not all genes: the TAF subunits of TFIID and subunits of the Mediator complexes.

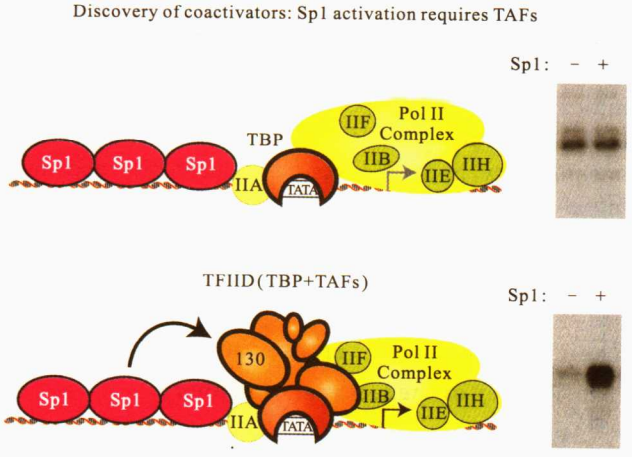
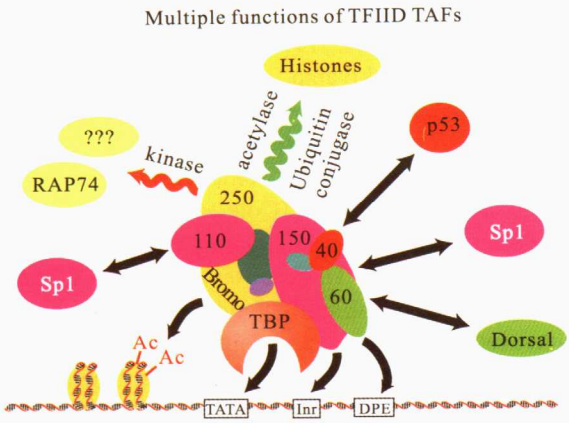


Fig.1.4　The TAF subunits of TFIID are required for transcriptional activation *in vitro*, serving as coactivators. The upper panel shows the lack of activation by Sp1 in transcription reactions reconstituted with TBP in place of TFIID. The lower panel shows Sp1 activation under identical conditions, with the exception that the holo-TFIID complex containing TBP and TAFs has replaced the single subunit TBP. The $TAF_{II}130$ (hTAF4) subunit of TFIID serves as a coactivator via interaction with Sp1.

The TFIID TAF Saga: Pride and Prejudice

The first well characterized group of co-activators turned out to be subunits associated with TBP and were thus called TAFs (TBP-associated factors) (Dynlacht *et al.*, 1991; Pugh and Tjian, 1990). Although the TAF's were originally discovered in *Drosophila* and human systems, eventually it was revealed that these subunits of the TFIID complex are, in fact, universal in eukaryotes and largely conserved from yeast to man. Thus, the paradox surrounding TFIID and its relation to TBP was finally resolved: TFIID is actually composed of TBP, a subunit essential for basal transcription, while the cluster of tightly associated TAF subunits are necessary for the co-activator function of TFIID (Fig.1.5). After a great deal of structure/function analysis *in vitro* and *in vivo*, we now know that the TAF/TBP complex actually participates in several distinct aspects of transcription including recognition of composite core promoter elements (i.e. INR, TATA, DPE) by TBP and several of the TAF subunits (Burke and Kadonaga, 1997; Hahn *et al.*, 1989b; Hoey *et al.*, 1990; Horikoshi *et al.*, 1989; Kao *et al.*, 1990; Kaufmann and Smale, 1994; Peterson *et al.*, 1990; Verrijzer *et al.*, 1994). Another important co-activator function involves direct or indirect targeting of TAFs by select activation domains (Chen *et al.*, 1994; Goodrich *et al.*, 1993; Hoey *et al.*, 1993). TAFs are not limited to TFIID, but are also found in other complexes (SAGA, STAGA, and TFTC) that function in regulated transcription (Grant *et al.*, 1998; Martinez *et al.*, 1998; Wieczorek *et al.*, 1998). Moreover, some of the TAFs carry out various enzymatic functions including protein phosphorylation, acetylation, and ubiquitination (Dikstein *et al.*, 1996a; Mizzen *et al.*, 1996; Pham and Sauer, 2000). Most intriguingly, one of the TAF's bears bromo-domains that are responsible for binding and discriminating between acetylated and non-acetylated histones in the context of chromatin (Jacobson *et al.*, 2000). Thus, it appears that co-activators such as TFIID/TAFs participate in numerous functions that may serve to integrate regulatory signals from DNA bound activators (repressors) and thus help potentiate transcription activation and control.



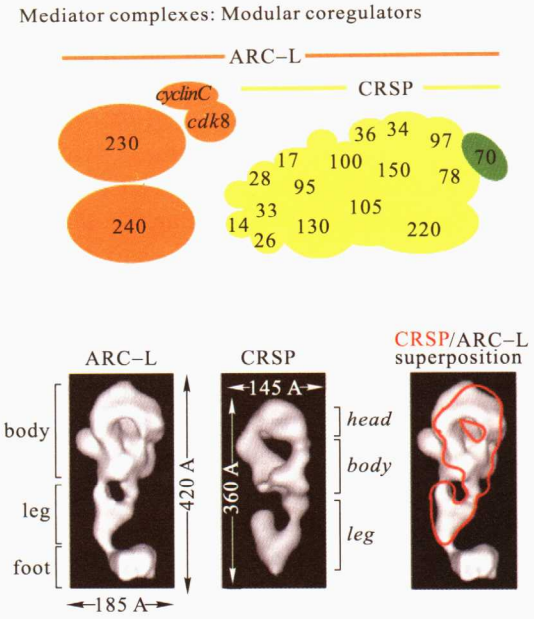


Fig.1.5 The TFIID complex plays multiple roles in the initiation and regulation of RNA polymerase II transcription. The names of the TAFs shown are the apparent molecular weights of the subunits of *Drosophila* TFIID: $TAF_{II}250$ (dTAF1), $TAF_{II}150$ (dTAF2), $TAF_{II}110$ (dTAF4), $TAF_{II}60$ (dTAF6), $TAF_{II}40$ (dTAF9).

Co-activators Abound: Mediator, CBP/p300, OCA, and Others

Of course, the TAFs and TFIID turned out to be merely the tip of the iceberg when it comes to co-activators. Using a combination of biochemistry and genetics a large number of co-factors, mediators, and co-regulators soon emerged. Among them were the yeast mediator and a series of mammalian coactivator complexes isolated in multiple labs and named CRISP, TRAP, DRIP, etc (Boyer *et al.*, 1999; Fondell *et al.*, 1996; Kim *et al.*, 1994b; Naar *et al.*, 1998; Rachez *et al.*, 1999; Ryu *et al.*, 1999; Sun *et al.*, 1998). Upon further purification and identification of the subunits, all of these complexes were found to be related and are now generally referred to as the Mediator. As a co-activator complex, the Mediator is not required for basal transcription *in vitro* and has not been found to bind DNA directly. Instead, it is thought to be recruited to promoters via interaction with promoter bound transcriptional activators where it facilitates the binding of RNA polymerase II. This class of co-activators also is able to directly bind to the CTD of RNA pol II and thus further integrate complex mechanisms of transcriptional control (Kim *et al.*, 1994a). A large and diverse group of activators have been found to bind the mediator complex, and EM studies revealed that the binding of activators can grossly alter the conformation of the co-activator complexes (Fig.1.6) (Taatjes *et al.*, 2002; Taatjes *et al.*, 2004).

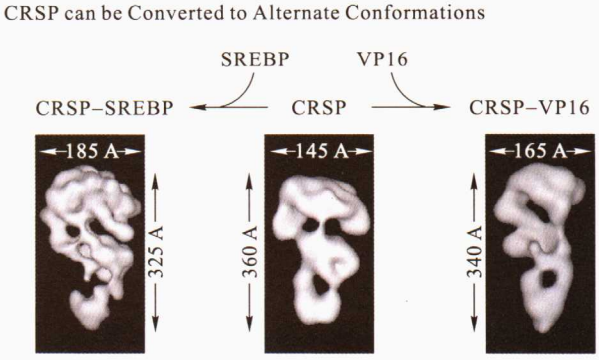


Fig.1.6 Multisubunit and Modular Co-regulatory Complexes. EM analysis revealed that ARC and CRSP, the mammalian counterpart of the yeast mediator are structurally related (yellow subunits are common) but distinct (orange and green subunits are unique) co-factors that display dramatically different functional properties. The larger ARC complex is inactive while the smaller CRSP complex is highly potent as a co-activator *in vitro*. Remarkably, the 3D structure of CRSP can undergo dramatic conformational changes dependent on the activator bound to target subunits within the CRSP assembly. Thus, the 3D structure of the unliganded, VP16-bound and SREBP-bound CRSP complexes display distinct structures as determined by negative stain EM and single particle reconstruction.

Although it may seem that the TAFs and the Mediator, which are ubiquitous transcriptional co-activators, would be sufficient for activating all genes, eukaryotic transcription once again proved to me more elaborate than imagined. Many other coactivators have now been identified. CBP, which was first identified as

a co-activator for phosphorylated CREB, and p300 are two highly related proteins that are now known to function in transcriptional activation at many genes (Chrivia *et al.*, 1993; Eckner *et al.*, 1994; Kwok *et al.*, 1994). The mechanism by which these two factors, as well as others (e.g. GCN5), co-activated transcription was partly illuminated by the finding that these proteins harbored histone acetyltransferase activity (Bannister and Kouzarides, 1996; Ogryzko *et al.*, 1996). Activator-specific, cell type-specific, and developmentally regulated co-activators soon followed: for example, OCA-B co-activates Oct transcription (Luo and Roeder, 1995), $TAF_{II}105$ (TAF4b) is found in a cell type specific version of TFIID in B cells (Dikstein *et al.*, 1996b; Freiman *et al.*, 2002), and multiple testis-specific TAF isoforms have been found to function in spermatid development (Hiller *et al.*, 2004; Hiller *et al.*, 2001). Clearly, when it comes to transcriptional regulation in eukaryotes, complexity is the dominant theme. Only time and considerably more research will reveal how vast the co-activator universe is and the diverse spectrum of mechanisms they use to potentiate transcriptional activation.

Paving the Way: Remodeling Nucleosomes at Promoters

While many labs were focusing considerable effort on identifying and characterizing the transcriptional machinery, a few bold researchers had the foresight to ask how activators, co-activators, and the general machinery could possibly overcome the repressive effects of nucleosomes and higher order chromatin structures present in eukaryotic nuclei. Inroads in this area came from the integration of complementary findings from experiments in yeast, *Drosophila*, and human, which showed that nucleosomes could be remodeled (Cote *et al.*, 1994; Kwon *et al.*, 1994; Pazin *et al.*, 1994; Tsukiyama *et al.*, 1994). The yeast SWI/SNF complex, subunits of which had been discovered in genetic screens, turned out to be an ATP-dependent chromatin remodeling complex and effector of transcription (Cote *et al.*, 1994). Other complexes that could assemble chromatin were found to have similar activity. These observations led to the idea that activators capable of binding native chromatin might recruit remodeling complexes to promoters, thereby opening the chromatin and allowing access to other transcriptional activators, co-activators, and the general transcription machinery. This proved to be the case (Neely *et al.*, 1999; Yudkovsky *et al.*, 1999), and the role of chromatin structure and its modulation was brought to the forefront of transcription research (Fig.1.7).

A Missing Link: Histone Modifications

For years, it had been known that in cells histones were differentially modified with acetyl, methyl, ubiquitin, and other post-translationally added groups. Dogma had it that histones in euchromatin, which was transcriptionally active, were hyper-acetylated, while histones in transcriptionally silenced heterochromatic regions were hypo-acetylated. Theories abounded to explain the correlation between histone acetylation and transcriptional competence, but for the most part, the transcription community paid little attention to these theories. This all changed with the identification of a nuclear histone acetyltransferase purified from *Tetrahymena* (Brownell *et al.*, 1996). Surprisingly, the *Tetrahymena* HAT had high sequence similarity to a known yeast co-activator, Gcn5p. Instantaneously, the

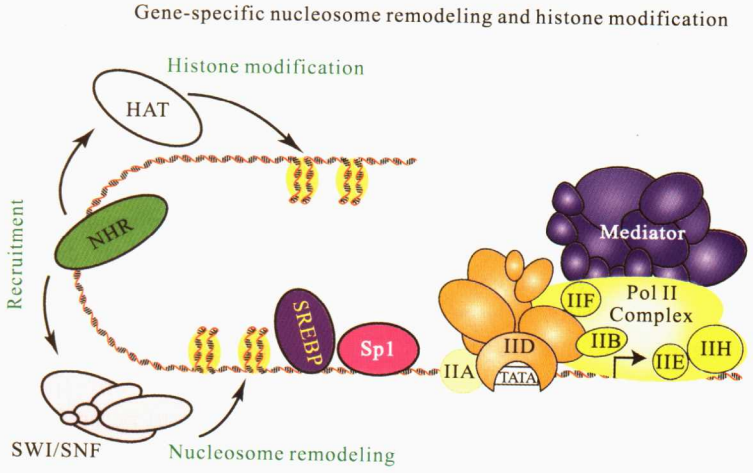


Fig.1.7 Nucleosome remodeling and histone modifying complexes are recruited to promoters via interactions with activators.

collective eyes of the transcription community opened to the possibility that many transcription factors might bear HAT activity. When the dust settled, multiple previously identified co-activators were found to be HATs, and ultimately it was realized that histones were not the only substrates of these acetyltransferases; indeed, activators themselves could be acetylated. With the subsequent discovery of deacetylases (Taunton *et al.*, 1996), acetylation was added to phosphorylation as a reversible post-translational modification used by intracellular signaling pathways to regulate gene expression. Ultimately, enzymes placing other modifications on histones (e.g. methylation, phosphorylation, and ubiquitination) were identified and characterized, and in some cases also found to be co-activators or co-repressors of transcription. Moreover, these enzymes can be recruited to promoters by gene specific activators and repressors to control levels of transcription.

The number of possible combinations of covalent modifications on the eight histones in any single nucleosome was dumbfounding. What was the function of all of these histone modifications? A seductive idea was posited: perhaps, specific patterns of post-translational modifications on the core histones in nucleosomes in individual promoters or regions of the genome help set the levels of transcription from those genes (Jenuwein and Allis, 2001; Strahl and Allis, 2000). For example, activation correlates with acetylation of specific lysines, while repression is observed upon acetylation or methylation of other lysines. Thus was born the Histone Code Hypothesis (Jenuwein and Allis, 2001; Strahl and Allis, 2000). While the putative histone "code" is far from understood, or even the notion of a true code accepted, it is clear that modification of histones adds another level of dynamic encoded information to the static DNA sequence present in a genome.

Escaped and On the Run: Regulation of Post-initiation Steps of Transcription

During the time that activators, co-activators, and general factors were being discovered and their roles in forming preinitiation complexes were initially characterized, some of the same labs and others embarked on understanding the mechanism and regulation of the RNA synthesis steps of the RNA polymerase II reaction. RNA synthesis is not simply the monotonous creation of phosphodiester bonds, but instead is a phase of the reaction rich in regulation (Fig.1.8). TFIIF and the TFIIH helicase function during promoter escape (Chang *et al.*, 1993; Goodrich and Tjian, 1994). The TFIIH kinase phosphorylates the CTD of RNA polymerase II as the enzyme leaves the promoter (Lu *et al.*, 1992). At the HSP70 promoter polymerase pauses after synthesis of a short (~20 nt) RNA, and is poised to fire the moment heat shock is sensed (via the Heat Shock Factor) (Gilmour and Lis, 1986; Rougvie and Lis, 1988). P-TEFb and DSIF/NELF have opposing effects on elongation (Marshall and Price, 1995; Wada *et al.*, 1998; Yamaguchi *et al.*, 1999). Elongation factors were discovered, including TFIIF, TFIIS, Elongin, etc, and indeed, the overall rate of elongation can be controlled globally and in a gene specific fashion (Aso *et al.*, 1995; Reinberg and Roeder, 1987). HIV TAT, regulates the transcription reaction by binding a TAR element in the nascent RNA, which is reminiscent of bacterial phage factors that control transcriptional termination by binding the RNA transcript (Kao *et al.*, 1987). RNA itself has recently appeared in the transcriptional regulatory picture, as a number of small noncoding RNAs have been found to control the RNA polymerase II transcription reaction via association with transcription factors and RNA polymerase II (Allen *et al.*, 2004; Espinoza *et al.*, 2004; Kwek *et al.*, 2002; Nguyen *et al.*, 2001; Yang *et al.*, 2001). It seems that evolution has taken advantage of many different regulatory mechanisms beyond simply controlling the formation of preinitiation complexes, and we have only begun to appreciate and understand the multiple layers of regulation that can come into play.

Keeping the End in Sight: Coupling RNA Processing to Transcription

As transcription factors were identified and characterized using biochemical and genetic approaches, individual pieces of data began to support the notion that the transcriptional apparatus in eukaryotic cells is tightly coupled to the RNA processing machinery, and moreover that the transcription reaction itself can be influenced by factors that add the 5′ Cap, splice the RNA, process the 3′ end of the transcript, and transport the mature transcript out of the nucleus (Fig.1.8) (Cho *et al.*, 1997; Dantonel *et al.*, 1997; Fong and Zhou, 2001; Hirose *et al.*, 1999; McCracken *et al.*, 1997a; McCracken *et al.*, 1997b; Strasser *et al.*, 2002). In hindsight, the coupling between transcription and RNA processing is logical, however, observations of splicing factors influencing transcription, and indications that RNA processing factors are recruited via interaction with the Pol II CTD were surprising, and the implications profound. We now envision that the nucleus contains mRNA synthesis/processing machines,