

A Unified Theory of Gene Expression

Review

George Orphanides¹ and Danny Reinberg^{2,3}

¹Syngenta Central Toxicology Laboratory
Alderley Park
Cheshire SK10 4TJ
United Kingdom

²Howard Hughes Medical Institute
Division of Nucleic Acids Enzymology
Department of Biochemistry
University of Medicine and Dentistry of New Jersey
Robert Wood Johnson Medical School
Piscataway, New Jersey 08854

The human genome has been called “the blueprint for life.” This master plan is realized through the process of gene expression. Recent progress has revealed that many of the steps in the pathway from gene sequence to active protein are connected, suggesting a unified theory of gene expression.

Introduction

Few scientific events justify a press conference involving the President of the United States of America and the director of its National Institutes of Health. The completion of the sequencing of the human genome was one such event and heralded the dawn of a new era in biology and medicine (Davies, 2001). However, the identification of the DNA sequence of every human gene is of limited value without a description of the function and regulation of the gene products. Metazoans consist of hundreds of different cell types, each designed to perform a specific role that contributes to the overall functioning of the organism. Every one of these cells contains the same 35,000–50,000 genes. The remarkable diversity in cell specialization is achieved through the tightly controlled expression and regulation of a precise subset of these genes in each cell lineage. Furthermore, the response of a cell to physiological and environmental cues requires tight regulation of these gene products. The reviews in this issue of *Cell* address the mechanisms used to manufacture and regulate a cell's complement of proteins—processes that bring the genome to life.

Figure 1 illustrates the process of gene expression in its broadest sense, from the activation of transcriptional regulators to the synthesis of a functional protein. The expression level of most genes is regulated by transcription factors that bind to DNA regulatory sequences situated upstream of the site at which transcription is initiated. The fact that more than 5% of our genes are predicted to encode transcription factors underscores the importance of this protein family in biology (Tupler et al., 2001). The activities of these proteins are controlled by a diverse array of regulatory pathways. For example, transcription factors that regulate genes involved in cell cycle progression are controlled by cell cycle signals, while factors that modulate the expression

of metabolic enzymes are often regulated directly by metabolites. Once activated, transcription factors bind to gene regulatory elements and, through interactions with other components of the transcription machinery, promote access to DNA and facilitate the recruitment of the RNA polymerase enzymes to the transcriptional start site. Three RNA polymerases function in eukaryotes (RNAP I, II, and III). Transcription of protein-coding genes is catalyzed by RNAP II. In this review, we will focus on the expression of genes transcribed by RNAP II, although many of the basic principles apply to the other polymerases.

Soon after RNAP II initiates transcription, the nascent RNA is modified by the addition of a “cap” structure at its 5' end. This cap serves initially to protect the new transcript from attack by nucleases and later serves as a binding site for proteins involved in export of the mature mRNA into the cytoplasm and its translation into protein (reviewed by Proudfoot et al., 2002 [this issue of *Cell*]). The capping process appears to coordinate early transcriptional events by regulating the transition between transcription “initiation,” during which RNAP II begins RNA synthesis, and transcription “elongation,” in which the polymerase moves 5' to 3' along the gene sequence and extends the transcript (see below). The elongation phase of transcription is subject to regulation by a family of “elongation factors” (Uptain et al., 1997). Coding sequences in the gene (exons) are often interrupted by long noncoding sequences (introns), which are removed by pre-mRNA splicing. Upon reaching the end of a gene, RNAP II stops transcription (“termination”), the newly synthesized RNA is cleaved (“cleavage”) and a polyadenosine [poly(A)] tail is added to the 3' end of the transcript (“polyadenylation;” all reviewed by Proudfoot et al., 2002 [this issue of *Cell*]).

The processes by which information is transferred from DNA to RNA (transcription) and from RNA to protein (translation) are physically separated in eukaryotes by a membrane that surrounds the nucleus; transcription occurs in the nucleus, whereas translation is a cytoplasmic event (however, see below). Therefore, processed mRNAs must be transported from the nucleus to the cytoplasm before translation can occur. Protein-lined pores in the nuclear membrane allow the bidirectional transport of macromolecules between nucleus and cytoplasm. As reviewed by Reed and Hurt (2002 [this issue of *Cell*]), export of mRNA is mediated by factors that bind to mRNA molecules in the nucleus and direct them into the cytoplasm through interactions with proteins that line the nuclear pores. Translation of mRNA into protein takes place on large ribonucleoprotein complexes called ribosomes (reviewed by Ramakrishnan, 2002 [this issue of *Cell*]) and is mechanistically analogous to transcription. It begins with the location of the start codon by translational initiation factors in conjunction with subunits of the ribosome and involves elongation and termination phases (reviewed by Dever, 2002 [this issue of *Cell*]). The nascent polypeptide chain then undergoes folding (reviewed by Fersht and Daggett, 2002 [this issue of *Cell*]) and often posttranslational

³Correspondence: reinbedf@umdnj.edu

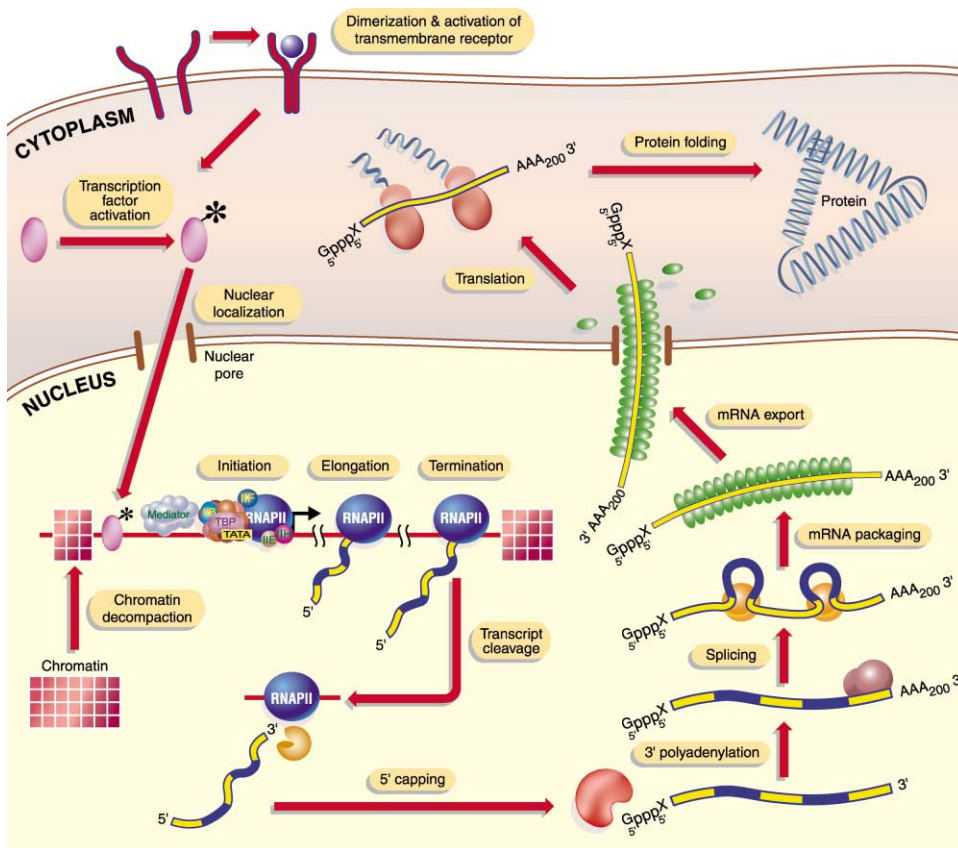


Figure 1. A Traditional View of Gene Expression

The different steps in the pathway from gene to protein have traditionally been viewed as independent events, with each going to completion before the next begins (see text for details).

chemical modification to generate the final active protein.

A Unified Theory of Gene Expression

The complexity of each of the steps in the pathway from gene to protein has required that they be studied in isolation, and most of our knowledge in this area has been generated using classical biochemistry. Using this approach, the biochemist first obtains experimental conditions in which the process of interest (e.g., transcription or translation) can be reconstituted *in vitro* in a cell-free extract. The protein machineries involved are then purified from the protein extract, allowing the process to be recapitulated using purified proteins and allowing the role of each player to be analyzed mechanistically. While this type of approach has been very useful, it forces the scientist to take a reductionist view. Each step in the pathway (e.g., transcription, pre-mRNA processing, and translation) is studied separately, often with little thought being given to the connections between steps. Consequently, the different steps have traditionally been viewed as discrete, unconnected events. As a testimony to this, most biochemical texts deal with these phases independently, with a chapter devoted to each. This separation of events is also most compatible with human thought process, which can most easily visualize complex processes as a linear series of events,

with each step going to completion before the next begins.

In recent years, the way in which we view gene expression has changed significantly, and decade-old observations suggesting that consecutive steps in the pathway are interdependent or are influenced by one another have taken on new meaning. A growing number of genetic studies have revealed functional links between the protein factors that carry out the different steps in the gene expression pathway. Similarly, conventional biochemical approaches and large-scale mapping of protein-protein interaction networks have uncovered physical interactions between the various machineries. In combination, these studies suggest that each stage is a subdivision of a continuous process, with each phase physically and functionally connected to the next. It has now been demonstrated, for example, that the transcriptional apparatus plays an active role in recruiting the machinery that caps and processes the nascent RNA transcript (Shatkin and Manley, 2000, also see Proudfoot et al., 2002 [this issue of *Cell*]), and that pre-mRNA splicing promotes transcription elongation (Fong and Zhou, 2001) and is required for efficient export of the resulting mRNA into the cytoplasm (Reed and Hurt, 2002 [this issue of *Cell*]). The temporal separation of each step has also been questioned; pre-mRNA splicing and packaging of the mRNA for export occurs even as the nascent

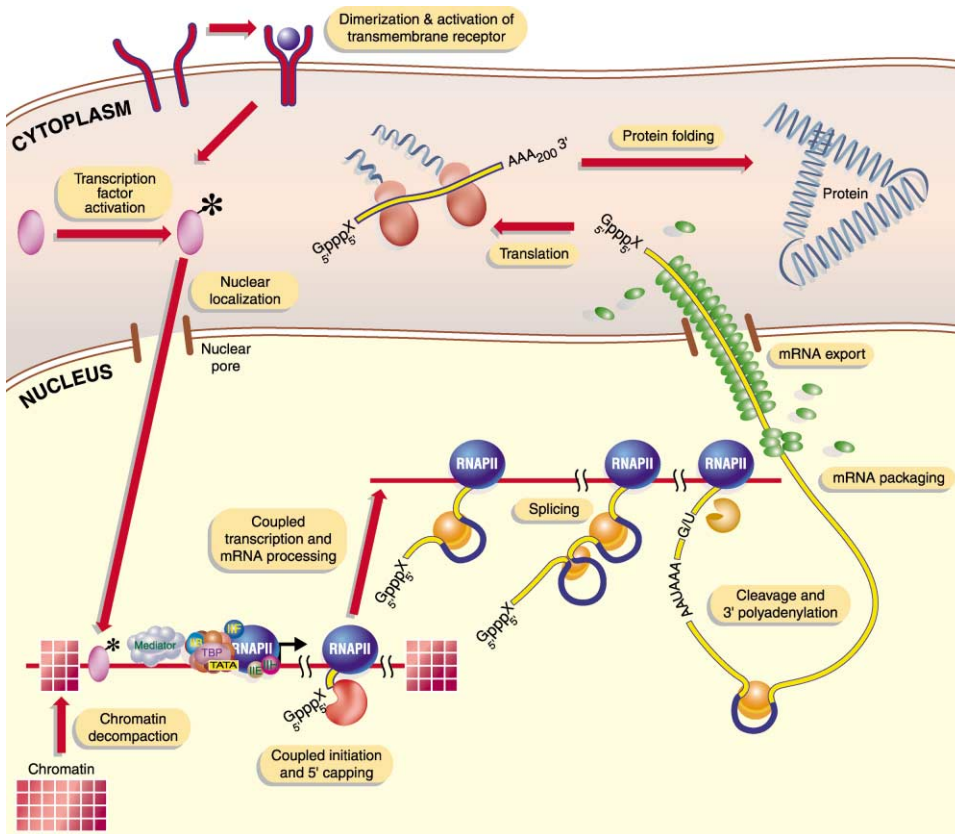


Figure 2. A Contemporary View of Gene Expression

Recent findings suggest that each step regulating gene expression (from transcription to translation) is a subdivision of a continuous process. In this contemporary view of gene expression, each stage is physically and functionally connected to the next, ensuring that there is efficient transfer between manipulations and that no individual step is omitted (see text for details).

transcript is spooling off the transcribing RNAP II. The picture that is emerging is one in which most steps are physically and functionally connected—conveyor belt-style—ensuring efficient transfer from one manipulation to the next (Figure 2). This organization of events may also introduce a series of quality control mechanisms, as it ensures that no individual step is omitted.

The results of a large body of work have revealed at least three general principles. (1) The protein factors responsible for each individual step in the pathway from gene to protein are functionally, and sometimes physically, connected. (2) Regulation of the pathway is controlled at multiple stages. (3) No general rules exist describing how the pathway is regulated. Different classes of gene are regulated at different stages.

In this review, we focus on novel paradigms that describe the functioning and regulation of the gene expression pathway and the connections that exist between the constituent steps of this process.

The Role of Chromatin Structure in Gene Expression: Not Just Packaging

The DNA in our cells is not naked, but packaged into a highly organized and compact nucleoprotein structure known as chromatin. The basic organizational unit of chromatin is the nucleosome, which consists of 146 bp of DNA wrapped almost twice around a protein core

containing two copies each of four histone proteins: H2A, H2B, H3, and H4 (Luger et al., 1997). These small, positively charged proteins show remarkable conservation among eukaryotes and are the protein building blocks of our chromosomes. Further compaction of our genes is achieved via poorly defined levels of higher-order nucleosome folding.

Once thought of as being a static organizational framework for DNA, it is now apparent that chromatin plays a pivotal role in regulating gene transcription by marshalling access of the transcriptional apparatus to genes (reviewed by Narlikar et al., 2002 [this issue of *Cell*]). However, not all chromatin is equal. Untranscribed regions of the genome are packaged into highly condensed “heterochromatin,” while transcribed genes are present in more accessible “euchromatin” (reviewed by Richards and Elgin, 2002 [this issue of *Cell*]). Each cell type packages its genes into a unique pattern of heterochromatin and euchromatin, and this pattern is maintained after cell division. The pattern of packaging into these alternative chromatin states determines which genes will be active in a newly divided cell, thus ensuring that the unique characteristics of each cell lineage are transferred from generation to generation. To activate gene expression, transcriptional activator proteins must, therefore, contend with inaccessible and repressive chromatin structures. As we discuss below,

they do this by nucleating events that lead to increased DNA accessibility.

Regulating the Regulators

Higher eukaryotes have developed sophisticated mechanisms for controlling the rate of gene transcription. The end point of many signal transduction cascades is the activation of transcriptional regulator proteins that bind to short sequence motifs found in the promoter and enhancer regions of genes. The regulatory sequences of most eukaryotic genes contain binding sites for multiple transcription factors, allowing each gene to respond to multiple signaling pathways and facilitating the fine-tuning of transcript levels (Lefstin and Yamamoto, 1998; McKenna and O'Malley, 2002 [this issue of *Cell*]). The activities of many transcription factors are context dependent and can be modulated by other regulators bound nearby. Thus, a single activated transcription factor can induce transcription of one gene while repressing that of another. This combinatorial and context-dependent regulation of transcription allows metazoans to respond to a surprisingly diverse array of stimuli using the same factors. Transcriptional control is a simpler affair in prokaryotes, where metabolically related genes are coregulated in common transcription units (operons) by a single transcriptional activator or repressor.

Considering the diversity of physiological signals that regulate gene expression, it is not surprising that the activities of transcription regulators are subject to multiple modes of regulation. A common theme in their regulation is the transport of a protein between nuclear and cytoplasmic compartments. This occurs through "nuclear pores," specialized gateways that span the nuclear membrane and control the passage of macromolecules. A family of transport factors that recognize short amino acid motifs found in proteins to be transported mediates movement of proteins through these pores. Two types of transport motif exist: nuclear export signals (NES) are found in proteins transported from the nucleus, while nuclear localization signals (NLS) label a protein for nuclear import (reviewed by Carmo-Fonseca, 2002 [this issue of *Cell*]). The beauty of this system lies in the ease at which protein transport can be controlled. Simply masking an NLS or NES by covalent modification or through the binding of a transport inhibitor prevents its recognition by the transport machinery.

To respond to changes in stimulus, a cell must be able to inactivate a transcriptional activator as quickly as it is induced. A number of recent reports have linked the ubiquitin protease system to both transcription factor activation and degradation. In yeast, ubiquitination potentiates the activity of the archetypal VP16 activation domain and also targets it for destruction (Salghetti et al., 2001). Similarly, in mammalian cells, the activity of the 26S proteasome is required for the transcriptional activity of the estrogen receptor and its degradation in the presence of ligand (Lonard et al., 2000). The tight coupling of transcription factor activation and degradation may operate to restrict prolonged transcription factor activity. Thus, ubiquitination may "license" transcription factors by linking their activity to their inactivation (Salghetti et al., 2001). The rapid turnover of promoter bound activated transcription factors resets the signaling pathway and allows the cell to continuously monitor

its environment. If signaling is prolonged, a newly activated protein will replace the degraded transcription factor. In the absence of signal, the degraded factor is not replaced and transcription ceases. This is consistent with observations that the mammalian glucocorticoid receptor (McNally et al., 2000) and the yeast Swi5 protein (Cosma et al., 2001) bind only transiently to their respective promoters. The role of the ubiquitin ligase system may extend to other phases of transcription. The largest subunit of RNAP II is ubiquitinated during transcription in vitro (Huibregtse et al., 1997; Mitsui and Sharp, 1999). This ubiquitination is induced by UV- or cisplatin-induced DNA damage in vivo (Bregman et al., 1996; Beaudenon et al., 1999) and by cisplatin-induced DNA damage in vitro (Lee et al., 2002). Furthermore, the yeast 19S proteasome is required for efficient transcription elongation by RNAP II and interacts physically and genetically with subunits of a known elongation factor, and mutations in its subunits lead to defects in elongation (Ferdous et al., 2001). Taken together, these findings raise the possibility that the entire transcriptional process is closely linked to the cellular processes that degrade proteins, allowing the rapid termination of transcription at multiple stages in response to general or gene-specific signals.

In addition to protein ubiquitination, a number of other posttranslational modifications play important roles in regulating transcription factor activity. Protein phosphorylation is the best-studied modification, and this may be due to the ease with which protein kinases can be identified by homology searching and the convenience with which phosphorylated proteins can be detected. However, transcription factors are subject to many other modifications, including acetylation on lysine residues and methylation on arginine and lysine residues (Zhang and Reinberg, 2001). Many of the enzymes that catalyze these modifications have been identified only recently. The p53 tumor suppressor protein, which responds to stress signals and coordinates a wide variety of cellular processes, was among the first transcription factors shown to be acetylated with functional consequences. Recent studies suggest that acetylation of several lysine residues at the C terminus of p53 by the p300 protein regulates its transcriptional activity by modulating interactions with coactivator and repressor proteins (Barlev et al., 2001; Prives and Manley, 2001).

How Do Transcription Factors Regulate Gene Expression?

Transcriptional activator proteins must bind to and decompact repressive chromatin structures to induce transcription (Narlikar et al., 2002 [this issue of *Cell*]). The way in which they do this is gradually becoming clear. To elicit their effects on gene expression, activators require the cooperation of a diverse family of coregulator proteins (McKenna and O'Malley, 2002 [this issue of *Cell*]). The function of these ancillary proteins was obscure until it was found that many were subunits of protein complexes that alter chromatin structure, or were themselves chromatin-modifying enzymes (see below). Thus, the recruitment of coactivators by DNA bound transcription factors leads to local chromatin decompaction and allows access of RNAP II and the general transcription machinery to the promoter.

Ordered DNA Binding and Chromatin Remodeling at Promoters—Insights into a “Chicken and Egg” Scenario

The requirement for chromatin decompaction for transcription factor binding to DNA appears to create a “chicken and egg” scenario: these factors induce local chromatin remodeling, but their interaction with DNA requires prior chromatin decompaction. How, then, do transcription factors access the DNA in the first place? The answer probably lies in the fact that some transcription activators can bind to their DNA recognition sequences even when they are packaged into nucleosomes. For example, the glucocorticoid receptor (GR) binds to a short DNA element and makes contacts with DNA in the major groove on one side of the double helix. Consequently, GR is able to bind its cognate sequence in a nucleosome. By contrast, the nuclear factor 1 (NF1) activator, which binds to a longer DNA sequence and completely surrounds the double helix, is unable to bind to nucleosomal DNA. Furthermore, access of a DNA binding protein to nucleosomal DNA will depend on the precise location of its binding site on the surface of the nucleosome (Urnov and Wolffe, 2001). The biological significance of this functional division between factors is best illustrated by studies on the induction of the MMTV gene (see Di Croce et al., 1999; Fletcher et al., 2000, and references therein). The promoter of this gene contains five GR binding sites and two NF1 sites. GR is able to bind to at least two of these sites when the promoter is packaged into nucleosomes. Promoter bound GR then recruits chromatin-modifying activities that increase DNA accessibility and facilitate the binding of NF1, thereby promoting synergy between the activation functions of the two activators.

Coregulator Proteins: Master Regulators of Transcriptional Networks?

Coregulators are recruited to promoters by sequence-specific DNA binding transcription factors and are required for the regulation of gene expression (Hampsey and Reinberg, 1999; Naar et al., 2001; McKenna and O’Malley, 2002 [this issue of *Cell*]). The role of coregulators has been studied most extensively for transcription that is regulated by the nuclear hormone receptor superfamily of ligand-activated transcription factors (McKenna and O’Malley, 2002 [this issue of *Cell*]). Two types of nuclear receptor coregulators exist: coactivators, recruited by ligand bound nuclear receptors, and corepressors, recruited by unliganded or antagonist bound receptors. These coregulators play pivotal roles in generating the promoter- and tissue-specific responses characteristic of nuclear receptor action. Another important transcriptional coregulator is the mediator, a multi-subunit complex that can associate with RNAP II and is required for the regulation of many different gene families (Hampsey and Reinberg, 1999). The modular nature of the mediator allows it to integrate both positive and negative regulatory signals (see Woychik and Hampsey, 2002 [this issue of *Cell*]).

The discovery that many transcriptional coregulators are enzymes that modulate chromatin structure underlines the importance of DNA packaging in gene expression. Coregulators that act on chromatin can be divided into two general classes: ATP-dependent nucleosome remodeling complexes and activities that catalyze post-translational modification of histones. ATP-dependent

chromatin remodeling complexes facilitate access of DNA binding proteins to DNA by repositioning nucleosomes at the promoter or by inducing conformational changes in nucleosomes (Narlikar et al., 2002 [this issue of *Cell*]). Four classes of histone modifiers have been implicated in transcriptional regulation. These are the histone acetyltransferases (HATs), the histone deacetylases (HDACs), the histone methyltransferases (HMTs), and the histone kinases (Narlikar et al., 2002 [this issue of *Cell*]). Histone acetylation was the first modification shown to correlate with transcriptional competence and is likely to be the event that initiates the breakdown of chromatin structure (Struhl, 1998). Recruitment of HATs and HMTs to promoters by activators results in the acetylation and methylation, respectively, of residues located in the N-terminal tails of histones and is crucial for the activation of many classes of gene (reviewed by Struhl, 1998; Turner, 2000; Zhang and Reinberg 2001; Jenuwein and Allis, 2001; Roth et al., 2001). Conversely, recruitment of HDACs by transcriptional repressors leads to deacetylation of the histone tails and is required for repression.

A Gene Switch Controlled by Arginine Methylation

Recent studies have shown that one particular coactivator, CARM1, is at the center of a sensitive molecular switch that regulates the decision to express either genes under the control of nuclear receptors or those regulated by the CREB transcription factor (Xu et al., 2001; Nishioka and Reinberg, 2001). CARM1 is a coactivator with arginine HMT activity toward the tail of histone H3. In synergy with the HAT activity of the P300/CBP coactivator, this activity promotes the induction of genes regulated by nuclear receptors. P300/CBP is present at limiting concentrations in many cell types and is an obligatory coactivator for a number of transcriptional pathways. A recent study (Xu et al., 2001) demonstrates that CARM1 also methylates an arginine residue in a domain of P300/CBP required for the interaction with CREB. Crucially, methylation of P300/CBP by CARM1 disrupts its interaction with CREB, thus inactivating the transcriptional activity of CREB. In this way, CARM1 simultaneously functions as a coactivator for nuclear receptor-mediated transcription and a corepressor for CREB-mediated transcription. It is likely that molecular switches of this kind operate throughout the gene regulatory network of a cell.

How Do Histone Tail Modifications Regulate Transcription?

The relationship between histone tail modification and gene expression is complex: the transcriptional consequence of any individual histone tail modification is influenced by other modifications on the same tail (Strahl and Allis, 2000; Turner, 2000). This was first realized in studies that examined phosphorylation and acetylation occurring on the histone H3 tail (Lo et al., 2000; Cheung et al., 2000; Thomson et al., 2001). However, recent studies have found a similar interplay in histone tail methylation. Two types of HMTs are involved in regulating transcription: those that target arginine residues and those that modify lysines. Arginine methylation of histones has thus far only been associated with the activation of transcription, whereas the effects of lysine methylation are context-dependent. Methylation of lysine 9 of histone H3 by SUV39 is associated with repression (Richards and Elgin, 2002 [this issue of *Cell*]). This modifica-

tion appears to define the heterochromatic state and is found on the histones that package a diverse range of untranscribed chromosomal regions, from the silenced β -globin genes (Litt et al., 2001) to the transcriptionally inactive X chromosome (Heard et al., 2001).

By contrast, methylation of lysine 4 by the Set1 complex in yeast (Roguev et al., 2001) or Set9 in humans (Nishioka et al., 2002) leads to transcriptional activation. Importantly, however, Set1-mediated methylation of lysine 4 of histone H3 within the ribosomal RNA locus leads to repression of transcription by RNAP II (Briggs et al., 2001). Moreover, it appears that methylation at lysine 9 can be converted from a repressive signal to an activating signal by methylation at lysines 4 or 27 on the same tail. It is not clear whether histone methylation is reversible, as activities capable of demethylating histones have not been identified. One possibility is that the repressing or activating effects of methylation are neutralized by other histone tail modifications. Alternatively, methylated histones may be targeted for degradation, or the tails may be clipped by proteases (Allis et al., 1980; Lin et al., 1991). Given the abundance of modifications that occur on the tails of all four histones, understanding the functional significance of each combination of modifications will be a major challenge.

The "Histone Code" Hypothesis

The way in which histone modifications result in the reorganization of chromatin structure is currently the subject of intense research. All four histone tails contain multiple targets for covalent modification, and various combinations of these modifications have been observed. This has led to the proposal that the modification state of the histone N termini make up a "histone code" read by proteins that modulate transitions between the different chromatin states (Strahl and Allis, 2000; Turner, 2000). This hypothesis predicts that histone modifications create binding sites for accessory proteins. This prediction is beginning to be realized: bromodomains are found in many proteins that modulate chromatin structure and are used by some proteins for recognition of histone tails acetylated at specific lysines (Jeanmougin et al., 1997; Doerks et al., 2001). Furthermore, heterochromatin protein 1 (HP1) uses its chromodomain to recognize the histone H3 lysine 9 modification found in heterochromatin (Richards and Elgin, 2002 [this issue of *Cell*]). It is likely that bromodomains and chromodomains have evolved to recognize histone tails carrying specific modifications, although it is important to point out that not all proteins containing these domains possess this function. The molecular interactions that dictate the specificity of binding between these domains and modified histone tails will become apparent once the atomic structures of a sufficient number of complexes have been determined. However, it is likely that conserved residues in these domains function to define the overall three-dimensional structure of the modules, while nonconserved residues determine specificity.

How Is Chromatin Disruption Propagated into Coding Regions?

Decompaction of chromatin at the promoter is not sufficient for efficient transcription. RNAP II often needs to traverse thousands of base pairs of compacted chromatin downstream of the promoter. How, then, is the decompaction of chromatin spread from the promoter area to the remainder of the gene to be transcribed? Recent

observations suggest that disruption of chromatin structure downstream of the promoter is coupled to transcription elongation (reviewed in Orphanides and Reinberg, 2000). Two protein factors have been implicated in this process. The first is the chromatin-specific transcription elongation factor, FACT, which facilitates RNAP II elongation through nucleosomes and plays a role in elongation *in vivo* (Orphanides et al., 1998, 1999). In a reconstituted transcription system employing chromatin templates, FACT facilitates RNAP II elongation by binding to nucleosomes and promoting the dissociation of histones H2A and H2B. Although FACT can facilitate elongation through chromatin independently, there is evidence that *in vivo* it also recruits histone acetyltransferases (HATs) to regions downstream of the promoter (reviewed in Orphanides and Reinberg, 2000). The second complex implicated in disrupting chromatin downstream of the promoter is the elongator, originally isolated as a component of elongating RNAP II (Otero et al., 1999) and recently shown to promote transcription through chromatin (Kim et al., 2002). It is likely that the disruption of chromatin structure in transcribed regions downstream of the promoter involves factors that track with RNAP II during transcription and other factors that help the polymerase to destabilize nucleosome structure.

Coordinating Transcription and Pre-mRNA Processing: A Pivotal Role for RNAP II

The engine at the heart of the transcriptional apparatus is RNAP II, whose 12 subunits are remarkably conserved throughout eukaryotes. Heroic crystallographic efforts using yeast RNAP II have given us the first detailed insight, at the atomic level, of the molecular mechanism used by eukaryotic RNA polymerases (Cramer et al., 2001; Gnatt et al., 2001; Woychik and Hampsey, 2002 [this issue of *Cell*]). Unlike the prokaryotic enzymes, eukaryotic RNA polymerases cannot recognize the promoters of their target genes and instead rely on a series of accessory factors known as the general transcription factors (GTFs; Orphanides et al., 1996; Roeder, 1996; Woychik and Hampsey, 2002 [this issue of *Cell*]). These protein factors recognize the conserved "TATA" box and "initiator" sequences present in most protein-coding genes and recruit RNAP II to the start site of transcription. The GTFs and RNAP II were identified as independent, chromatographically distinct factors. However, the purification of preassembled complexes, from yeast and human cells, containing GTFs, RNAP II, and other regulatory factors, popularized the view that RNAP II and its accessory factors are recruited to the promoter as a preformed complex, in the form of a "holoenzyme," as is the case in bacteria. This notion, however, has recently been questioned by reports demonstrating that certain holoenzyme components are recruited to some promoters independently of RNAP II (Cosma et al., 2001; Bhoite et al., 2001). Importantly, the relative abundance of the yeast GTFs is more compatible with a step-wise assembly model (Borggreve et al., 2001).

Pre-mRNA Processing Occurs Cotranscriptionally: The Role of the Carboxy-Terminal Domain of the Largest Subunit of RNAP II

A plethora of evidence suggests that pre-mRNA processing reactions occur during, and not after, transcription (Shatkin and Manley, 2000; Hirose and Manley, 2000; Proudfoot et al., 2002 [this issue of *Cell*]). Capping

of the 5' end of the RNA is the first of these modifications and occurs as soon as the 5' end of the nascent transcript becomes accessible, usually after transcription of 20–30 nt of RNA. Curiously, it has long been known that the polymerase pauses transcription at around this point (Woychik and Hampsey, 2002 [this issue of *Ce/I*]). Therefore, it is possible that *in vivo* this pausing occurs to allow time for 5' capping, and that RNAP II will not continue until this protective modification has been added. This would be analogous to the “checkpoints” that operate during the cell cycle to ensure that each phase of the cycle is complete before the next begins.

The key player in the coupling of pre-mRNA transcription and processing is a unique and unusual domain present at the C terminus of the largest subunit of RNAP II known as the “CTD” (carboxy-terminal domain). In mammals, this domain consists of 52 repeats of the consensus heptapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser and appears to be unstructured in crystallographic studies (Cramer et al., 2001). The initiation of transcription by RNAP II is a multistep process involving separation of the DNA strands at the initiation site (“promoter melting”), formation of the first phosphodiester bond of the transcript, and disruption of the interactions between RNAP II and the promoter (“promoter clearance”). The transition from initiation to elongation is accompanied by massive phosphorylation of the CTD of RNAP II. The precise function of the CTD has been elusive, although it was speculated almost a decade ago that it constitutes a binding site for other protein factors (Greenleaf, 1993). It now appears that the CTD is a platform for the ordered assembly of the different families of pre-mRNA processing machinery (Figure 3). Consistent with this suggestion, the CTD is required for efficient pre-mRNA processing and interacts specifically with all classes of processing factor (Shatkin and Manley, 2000; Hirose and Manley, 2000; Proudfoot et al., 2002 [this issue of *Ce/I*]). Therefore, it is likely that phosphorylation of the CTD during transcription coordinates the recruitment of pre-mRNA capping, splicing, and 3' processing factors at different stages in the synthesis of the nascent transcript. Experiments utilizing a series of truncated CTD peptides have revealed that the ability of the CTD to stimulate pre-mRNA cleavage in a reconstituted 3' processing reaction is not dependent on any particular region of the domain (Ryan et al., 2002). Instead, the cleavage-stimulation activity of the CTD is largely dependent on the number of repeats present. This observation is consistent with a scaffolding role for the CTD in pre-mRNA cleavage and perhaps other pre-mRNA processing reactions.

Any model implicating the CTD in coordinating the association of protein machineries with the pre-mRNA predicts the existence of a mechanism by which the CTD can “signal” to the machinery the phase of transcription it is engaged in and, therefore, the status of its transcript. Two observations provide circumstantial support for this mechanism. First, the phosphorylation pattern of the CTD changes during transcription, with phosphorylation of serine 5 in the CTD motif occurring between transcription initiation and promoter clearance and modification of serine 2 being found only when the polymerase is associated with the coding region (Komarnitsky et al., 2000). Second, different pre-mRNA processing factors recognize distinct regions of the CTD

(Fong and Bentley, 2001). Thus, a “CTD code” may exist, analogous to the histone code described above. Changes in CTD phosphorylation triggered upon completion of each stage of transcription, or after each mRNA processing reaction, may create docking sites for enzymes that catalyze the next processing step. These docking sites would act synergistically with sequences in the nascent RNA to recruit the processing apparatus.

The coordinating role played by the RNAP II CTD in RNA processing may also ensure that the reactions occur in the correct order and that the transitions between the reactions are efficient. Furthermore, connecting the pre-mRNA processing steps in this way introduces a series of quality control mechanisms to ensure that no step is omitted. It is tempting to speculate that the language of the CTD code also includes a word for “RNA degradation,” to be used by RNAP II when it is associated with a defective transcript in order to recruit factors that degrade the RNA. Signals on the CTD may also be used when RNAP II arrests during transcription. Indeed, transcriptional arrest induced by DNA damage results in ubiquitination of RNAP II on its largest subunit (see above), and also inhibits the activities of the 3' processing machinery (Kleiman and Manley, 2001).

A Checkpoint Model for the Coupling of 5' Pre-mRNA Capping and Early Transcription Initiation

Identification of the proteins that couple transcription and pre-mRNA processing facilitates a detailed analysis of the molecular mechanism involved. This has been best studied for the coupling of 5' RNA capping with early elongation, allowing construction of a model involving interplay among RNAP II, factors that regulate elongation near the promoter, kinases that phosphorylate the CTD, and the capping enzymes. As illustrated in Figure 4, the model begins with the binding of the DSIF factor (Wada et al., 1998a, 1998b) to RNAP II shortly after initiation or during formation of the transcription complex at the promoter. DSIF then recruits NELF (Yamaguchi et al., 1999), which arrests transcription. The cdk7 subunit of the initiation factor TFIIF phosphorylates the CTD of RNAP II on serine 5 of its heptapeptide motif between initiation and arrest (Woychik and Hampsey, 2002 [this issue of *Ce/I*]). The paused RNAP II is then joined by the capping enzymes through interactions with the serine 5-modified CTD and DSIF (Wada et al., 1998b; Wen and Shatkin, 1999; Kim et al., 2002). Following the addition of a cap to the 5' end of the nascent RNA, the P-TEFb CTD kinase binds to RNAP II and phosphorylates serine 2 (Cho et al., 2001) and DSIF (Ivanov et al., 2000; Kim and Sharp, 2001). This neutralizes the repressive action of NELF and allows the polymerase to resume elongation.

Protein-protein interactions reported between the factors involved in this coupling process are generally consistent with the model presented above. However, it is not clear how P-TEFb is recruited. P-TEFb may be recruited to RNAP II at the promoter, and its kinase activity may be inactive until the cap has been added. Indeed, the observations that two transcriptional activator proteins interact with P-TEFb suggest that this complex may be recruited to the initiation complex at the promoter (Kanazawa et al., 2000; Barbaric et al., 2001). Alternatively, P-TEFb may join the paused RNAP II complex after capping. The mechanism of P-TEFb recruitment and release

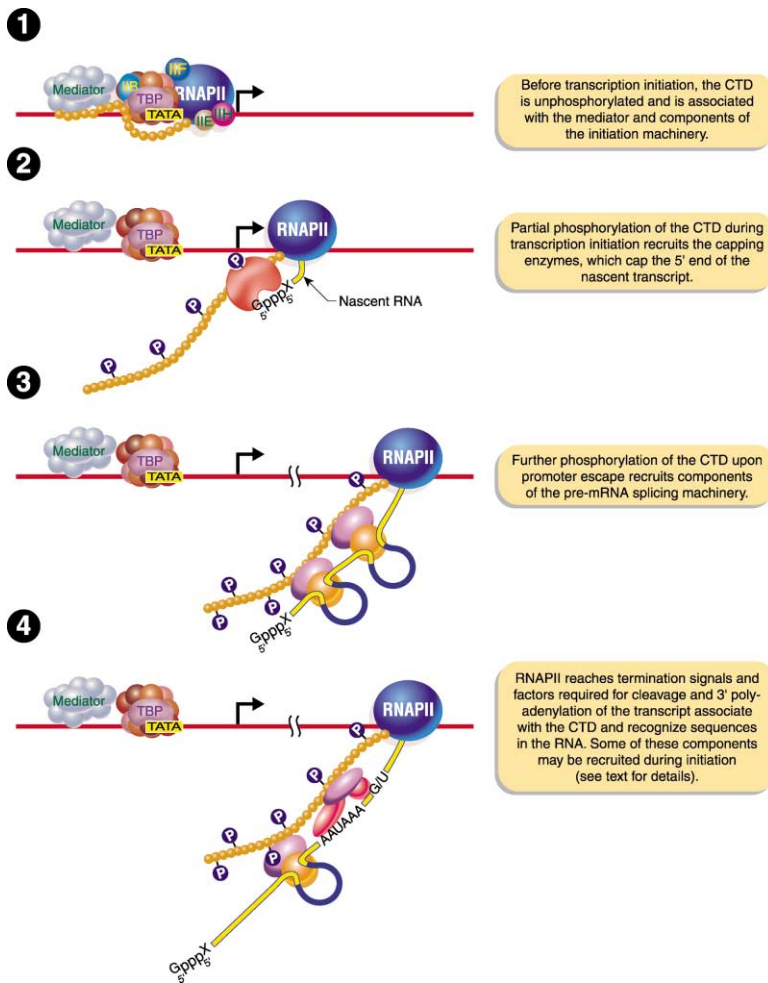


Figure 3. The C-Terminal Domain (CTD) of RNA Polymerase II Coordinates Transcription and Pre-mRNA Processing

The CTD consists of 52 repeats of the consensus heptapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser and serves as a platform for the ordered assembly of the factors responsible for transcription, pre-mRNA 5' capping, splicing, and 3' processing at different stages in the synthesis of the nascent transcript.

from RNAP II pausing is best understood for HIV transcription. Here, the TAT viral transactivator recognizes a stem-loop structure that forms on the nascent RNA shortly after initiation and recruits P-TEFb to the stalled RNAP complex to facilitate release from arrest. Cellular transcripts, however, do not form stem-loop structures recognized by TAT and must therefore recruit P-TEFb by alternative means. On these genes, it is possible that the capping enzyme complex recruits the P-TEFb kinase. Thus, the capping enzyme has two roles in the coupling process: to cap the 5' end of the RNA and to recruit P-TEFb. The latter role reverses RNAP II stalling. In this way, a checkpoint mechanism operates, ensuring that the polymerase does not extend uncapped transcripts. It is likely that protein-protein interactions of this kind, cocordinated by protein phosphorylation, operate in the coupling of other steps in the gene expression pathway.

Coupling 3' Transcript Processing with Transcription, Splicing, and 5' Capping

Upon reaching the end of a gene, the nascent RNA is cleaved, a polyadenosine tail of approximately 200 nt is added at the 3' end of the transcript, and RNAP II terminates transcription (Proudfoot et al., 2002 [this issue of *Cell*]). Convincing evidence exists for the coupling of pre-mRNA splicing and 3' polyadenylation with the

5' capping process. Biochemical analyses have revealed that an intact 5' cap is required for efficient splicing and polyadenylation (Shatkin and Manley, 2000; Hirose and Manley, 2000; Proudfoot et al., 2002 [this issue of *Cell*]). The way in which these processes are coupled has not been defined, but the mechanism is likely to involve an interaction of the cap binding protein complex (CBC), bound at the 5' end of the RNA, with components of the splicing and 3' end-processing machinery. The interaction between the CBC and the splicing machinery is only required for excision of the first intron, and a likely possibility is that the coupling between capping and 3' processing is important for processing intronless mRNA precursors.

A number of independent observations point to a functional interaction between the first and last phases of gene transcription: initiation and termination/3' end formation. Experiments with vaccinia virus demonstrated that the enzyme that adds the cap structure at the 5' end of the nascent transcript during initiation also mediates transcription termination (Shuman et al., 1987). More recently, three studies analyzing RNAP II transcription have identified physical and functional interactions between proteins involved in initiation that act at the promoter and subunits of the cleavage-polyadenylation specificity factor (CPSF) and cleavage stimulation factor

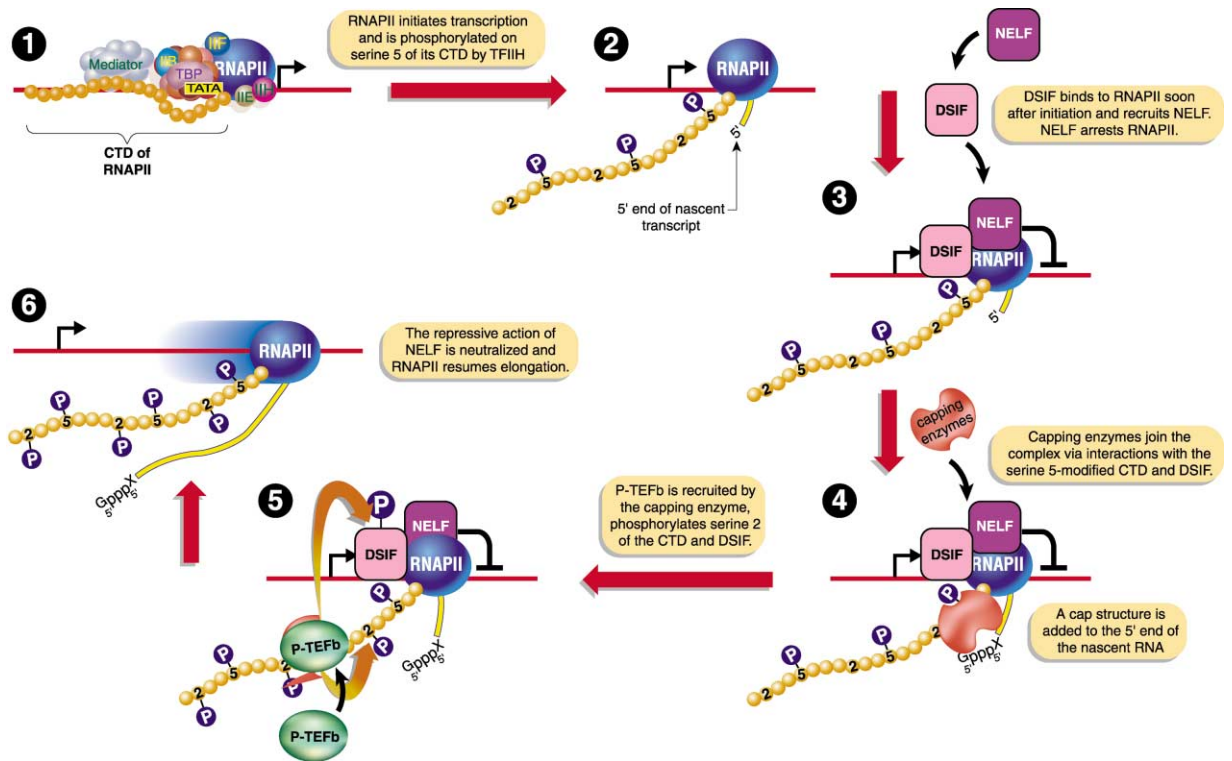


Figure 4. A "Checkpoint" Model for the Coupling of 5' Pre-mRNA Capping and Early Transcription Initiation

Capping of the nascent transcript occurs as soon as the 5' end of the RNA becomes accessible, usually after transcription of 20–30 nt. RNAP II has long been known to temporarily stop transcription at this point, although the functional significance of this pausing has been obscure. Recent observations suggest that RNAP II pauses to allow time for 5' capping to occur and that this pausing is part of checkpoint mechanism that ensures that uncapped transcripts are not extended. In the model presented here, a series of specific protein-protein interactions and phosphorylation events involving components of the initiation and 5' capping machineries function to coordinate this checkpoint.

(CstF), required for 3' processing. First, the CPSF complex is associated in nuclear extracts with the GTF TFIID, which binds to the TATA element and other sequences in the vicinity of the transcription start site (Dantonel et al., 1997). Second, a yeast two-hybrid interaction screen for proteins that bind to the 64 kDa subunit of CstF (CstF-64) identified a protein named PC4 (Calvo and Manley, 2001), which is a well-studied transcriptional coactivator (Ge and Roeder, 1994). Furthermore, Rna15 and Sub1, the yeast homologs of CstF-64 and PC4, interact genetically and are associated in vivo (Calvo and Manley, 2001). The final observation indicating a coupling between transcription initiation and 3' processing involves the initiation factor TFIIB. In genetic experiments, the gene encoding TFIIB exhibits allele-specific interactions with the *Ssu72* gene (Sun and Hampsey, 1996), which was recently found to be a component of the cleavage and polyadenylation machinery (Gavin et al., 2002). Intriguingly, TFIIB shows the same allele-specific interaction with Sub1 (Wu et al., 1999), thereby linking TFIIB, Sub1, and *Ssu72*. The multitude of protein-protein interactions defined between components of the initiation and termination apparatus raises the possibility that the two processes are physically connected in the nucleus. This could facilitate efficient reinitiation of transcription by RNAP II complexes that have terminated and might be analogous to the circularization of mRNAs during eukaryotic protein translation. A similar mecha-

nism has been proposed for the mechanism of termination and reinitiation by RNAP III (Dieci and Sentenac, 1996).

Getting Transcripts into the Cytoplasm: Coupling mRNA Export with Transcription and Pre-mRNA Processing

Once an mRNA has been fully processed, it must be transported to the site of protein translation in the cytoplasm. mRNA transport utilizes the same nuclear pore channels used to transport proteins to and from the nucleus (Reed and Hurt, 2002 [this issue of *Cell*]). The molecular mechanisms of bulk mRNA transport are not fully understood. However, it is assumed that mRNA transport proteins recognize and bind to a conserved element found in processed transcripts and target them to nuclear pores. The best candidate for a factor that carries out this role is the Mex67/Mtr2 complex (reviewed by Reed and Hurt, 2002 [this issue of *Cell*]). This factor is essential for RNA export in yeast and interacts with both mRNA and nucleoporin proteins that line the nuclear pore. Furthermore, this heterodimer has been observed to shuttle between the nucleus and cytoplasm. Splicing of pre-mRNAs has been shown to promote their export, suggesting that the two processes may be coupled (Luo and Reed, 1999). The proteins that couple splicing and transport are beginning to be identified. Significantly, one of these coupling proteins, Aly, is only

recruited to mRNAs during splicing. This may function as a quality control mechanism that prevents export of mRNAs that are not fully processed. Although unproven, it is likely that this coupling is facilitated by direct interactions between mRNA splicing and export factors. Furthermore, the finding that members of the SR family of pre-mRNA splicing factors are involved in mRNA transport (Huang and Steitz, 2001) and in regulating mRNA stability (Lemaire et al., 2002) adds further complexity to the coupling of these events.

Protein Synthesis: Solely a Cytoplasmic Event?

The processes of transcription and pre-mRNA processing described above function to convert the genetic code into a form that can be read by the protein translation apparatus. Translation is a cytoplasmic event that takes place on large ribonucleoprotein complexes called ribosomes. As reviewed by Ramakrishnan, 2002 [this issue of *Cell*], our understanding of the mechanism of translation has recently been revolutionized by the visualization of the ribosome structure at atomic resolution.

In bacteria, translation can occur as the nascent transcript emerges from the RNA polymerase. Eukaryotic cells contain a nuclear membrane that separates the cell into nuclear and cytoplasmic compartments. Therefore, it is assumed that transcription and translation are spatially separated events in eukaryotes. This view has recently been challenged. In the mid-1970s, Littauer and colleagues reported the existence of ribosome-like structures in the nucleus (Gozes et al., 1977). The significance of this observation was unclear until last year, when observations made by Cook and coworkers suggested that a fraction of a cell's protein synthesis occurs in the nucleus and may depend on concurrent transcription of the translated mRNA (Iborra et al., 2001). If correct, this observation has implications for regulation of transcription in eukaryotes. In bacteria, transcription of certain genes encoding enzymes that synthesize amino acids and pyrimidines are subject to feedback regulation mediated by the ribosome, a process known as attenuation. When amino acid levels are limiting, translation is slowed and the ribosome pauses, leading to stabilization of a stem-loop structure in the transcriptionally engaged mRNA that prematurely terminates transcription (Yanofsky, 1988). Nuclear translation may allow similar mechanisms to operate in eukaryotes, although a mechanism integrating translation of nascent mRNA precursors and splicing is not without difficulties.

Regulation of Gene Expression Occurs at Multiple Stages

Regulating a rate-limiting step is an efficient way to control the overall rate of a multistep process. However, there is a limit to the level of regulation that can be achieved by controlling a single step; there is no point in increasing the rate of one step if another soon becomes rate limiting. Therefore, eukaryotes have developed methods to regulate the expression of their proteins at multiple levels in a coordinated fashion. The rich diversity of mechanisms that exist to regulate protein expression and activity is likely to be a result of the clear selective advantage conferred by the ability to finely tune the activities of proteins in a cell. Regulation can

be gene-specific or general. For example, specific classes of genes can be controlled at the transcriptional level by regulating a transcription factor, while certain stimuli (e.g., hormones) can also regulate protein synthesis to increase the overall translational capacity of a cell. In regulatory terms, each gene is an individual: there are no general rules that govern how genes are controlled. The mechanisms used largely depend on the level of regulation required for proper gene function and are selected through evolution. Genes whose expression must be rapidly and tightly controlled tend to be quickly transcribed and translated, and their mRNAs and proteins have short half-lives. By contrast, genes that play general or "housekeeping" roles are often induced slowly and have long half-lives. A detailed discussion of all mechanisms of gene regulation is beyond the scope of this review. However, a few general principles are discussed below.

Regulating Protein Activity

The simplest way of regulating the activity of a protein is by changing the number of protein molecules in the cell. This is generally achieved by regulating gene transcription, mRNA translation, and the rate of mRNA and protein turnover. A more rapid increase in protein activity can be achieved by keeping a protein in an inactive state until activated by posttranslational modification. For example, activation of a kinase by protein phosphorylation is a hallmark of intracellular signaling cascades (reviewed by Hunter, 2000)

Generating Protein Diversity

In addition to controlling the activity of their protein complement, eukaryotes can alter the properties of their proteins by using different combinations of coding exons, a process known as alternative splicing. Importantly, however, although alternative splicing of pre-mRNA molecules allows multiple isoforms of a protein to be made from a single gene, the process is unidirectional and thus ensures that the genetic information encoded in the DNA is maintained. RNA editing, while rare in metazoans, can produce further diversity by altering the coding sequence of the mRNA itself (Gerber and Keller, 2001; also see Madison-Antenucci, 2002 [this issue of *Cell*]). Add to this the additional diversity in protein function contributed by posttranslational chemical modification of proteins, and the theoretical number of different protein activities far exceeds the number of genes in the genome.

Regulation by Compartmentalization: Controlling the Local Concentration of a Protein

The overall activity of a protein in a cell is dictated by the concentration of active molecules. Therefore, increasing the local concentration of a protein by localizing it to a particular cell compartment is an effective way to regulate activity. In this way, proteins are sent to the sites where they are needed. Short stretches of amino acids recognized by transport factors direct the cellular localization of proteins (see above; Carmo-Fonseca, 2002 [this issue of *Cell*]). Similarly, nucleotide sequences found in the untranslated regions of mRNAs are recognized by factors that transport and anchor them to specific cellular structures (reviewed by Kloc et al., 2002 [this issue of *Cell*]). The localization of mRNAs ensures that proteins are translated (and therefore accumulate) at the appropriate cellular location. The mechanisms of

mRNA and protein transport and localization are just beginning to be elucidated, and it is likely that many proteins are regulated in this manner. The way in which a given gene is regulated is encoded in its sequence. Different DNA sequences present in promoter regions, untranslated sequences, and translated sequences dictate the rate of transcription, pre-mRNA processing and splicing, mRNA export, and protein synthesis. Additional sequences govern the localization and stability of mRNAs and proteins. Therefore, definition of these different classes of regulatory sequences will, in theory, allow the modes of regulation applied to any given gene be discerned from its sequence.

Perspectives and Future Directions

We are entering a period that will revolutionize the study of gene expression. We now know enough about the individual steps in the gene expression pathway to begin to understand the connections between them and the processes that regulate them. As is often the case, advances in technology have driven scientific breakthroughs. For example, developments in the application of fluorescent technologies in cell biology now allow the interactions between biological macromolecules to be studied in real time and in living cells. The analysis of proteins in their natural environment has already provided some surprising insights, such as the rapid recycling of promoter bound transcription factors (McNally et al., 2000). Analysis of the structural organization of this natural environment—the cell—is also likely to yield some important breakthroughs. The traditional view of a cell as a bag of molecules separated into compartments by membranes is beginning to be challenged. For example, the results of recent experiments in which transcription has been studied in intact nuclei have been taken to suggest that eukaryotic transcription occurs in a defined number of transcription factories (or “transcriptosomes”) containing factors necessary for transcription and pre-mRNA processing (reviewed by Cook, 1999). This view has not been universally accepted but is consistent with the abundance of protein-protein interactions made between components of the transcription and processing apparatus. We are also beginning to understand the workings of the gene expression machinery at the molecular level. Crystallographic analysis of large, multisubunit complexes is becoming more routine, and recent studies have revealed the structures of two of the central protein engines in gene expression, the ribosome (reviewed by Ramakrishnan, 2002 [this issue of *Cell*]) and RNAP II (reviewed by Woychik and Hampsey, 2002 [this issue of *Cell*]), at atomic resolution.

The sequencing of the genomes of a number of organisms has led to a significant philosophical shift in the way in which experiments are conducted. Technologies that permit global transcript profiling force the experimentalist to take a holistic view and to consider biological pathways and processes that would otherwise be ignored. Genome sequences provide a richness of information not limited to the coding sequences of the genes themselves. Gene regulatory sequences hold the key to understanding how genes are regulated by programmed and environmental signals. Global analysis of the binding of transcriptional regulators to these sequences can

provide multitiered information on the transcriptional networks that control cellular processes (Iyer et al., 2001; Ren et al., 2000; Simon et al., 2001).

As we move into the “postgenomic” era, the emphasis will be placed on identifying the functions of our gene products and the way in which they are regulated. This discipline has often been referred to as “functional genomics,” although the techniques involved are often those used in “classical biochemistry.” Understanding the regulatory networks that govern normal cellular processes will in turn lead to an appreciation of how these mechanisms go wrong in the disease state. Furthermore, comprehending the mechanisms that regulate specific gene families will facilitate the rational design of pharmaceutical and agrochemical molecules with enhanced selectivity.

Acknowledgments

We thank Drs. Michael Hampsey, Jim Manley, and Aaron Shatkin for providing helpful comments on the manuscript. We apologize to colleagues whose work we did not cite due to space limitations. Work in the laboratory of D.R. is supported by the National Institutes of Health and the Howard Hughes Medical Institute.

References

- Allis, C.D., Bowen, J.K., Abraham, G.N., Glover, C.V., and Gorovsky, M.A. (1980). Proteolytic processing of histone H3 in chromatin: a physiologically regulated event in *Tetrahymena* micronuclei. *Cell* 20, 55–64.
- Barbaric, M., Nissen, R.M., Kanazawa, S., Jabrane-Ferrat, N., and Peterlin, B.M. (2001). NF- κ B binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Mol. Cell* 8, 327–337.
- Barlev, N.A., Liu, L., Chehab, N.H., Mansfield, K., Harris, K.G., Halazonetis, T.D., and Berger, S.L. (2001). Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases. *Mol. Cell* 8, 1243–1254.
- Beaudenon, S.L., Huacani, M.R., Wang, G., McDonnell, D.P., and Huibregtse, J.M. (1999). Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 19, 6972–6979.
- Bhoite, L.T., Yu, Y., and Stillman, D.J. (2001). The Swi5 activator recruits the Mediator complex to the HO promoter without RNA polymerase II. *Genes Dev.* 15, 2457–2469.
- Borggreve, T., Davis, R., Bareket-Samish, A., and Kornberg, R.D. (2001). Quantitation of the RNA polymerase II transcription machinery in yeast. *J. Biol. Chem.* 276, 47150–47153.
- Bregman, D.B., Halaban, R., van Gool, A.J., Henning, K.A., Friedberg, E.C., and Warren, S.L. (1996). UV-induced ubiquitination of RNA polymerase II: a novel modification deficient in Cockayne syndrome cells. *Proc. Natl. Acad. Sci. USA* 93, 11586–11590.
- Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F., and Allis, C.D. (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev.* 15, 3286–3295.
- Calvo, O., and Manley, J.L. (2001). Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination. *Mol. Cell* 7, 1013–1023.
- Carmo-Fonseca, M. (2002). The contribution of nuclear compartmentalization to gene regulation. *Cell* 108, this issue, 513–521.
- Cheung, P., Tanner, K.G., Cheung, W.L., Sassone-Corsi, P., Denu, J.M., and Allis, C.D. (2000). Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. *Mol. Cell* 5, 905–915.
- Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J., and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at

- Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev.* 15, 3319–3329.
- Cook, P.R. (1999). The organization of replication and transcription. *Science* 284, 1790–1795.
- Cosma, M.P., Panizza, S., and Nasmyth, K. (2001). Cdk1 triggers association of RNA polymerase to cell cycle promoters only after recruitment of the mediator by SBF. *Mol. Cell* 7, 1213–1220.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* 292, 1863–1876.
- Dantonel, J.C., Murthy, K.G., Manley, J.L., and Tora, L. (1997). Transcription factor TFIID recruits factor GPSF for formation of 3' end of mRNA. *Nature* 389, 399–402.
- Davies, K. (2001). After the genome: DNA and human disease. *Cell* 104, 465–467.
- Dever, T.E. (2002). Gene-specific regulation by general translation factors. *Cell* 108, this issue, 545–556.
- Di Croce, L., Koop, R., Venditti, P., Westphal, H.M., Nightingale, K.P., Corona, D.F., Becker, P.B., and Beato, M. (1999). Two-step synergism between the progesterone receptor and the DNA-binding domain of nuclear factor 1 on MMTV minichromosomes. *Mol. Cell* 4, 45–54.
- Dieci, G., and Sentenac, A. (1996). Facilitated recycling pathway for RNA polymerase III. *Cell* 84, 245–252.
- Doerks, T., Copley, R., and Bork, P. (2001). DDT—a novel domain in different transcription and chromosome remodeling factors. *Trends Biochem. Sci.* 26, 145–146.
- Ferdous, A., Gonzalez, F., Sun, L., Kodadek, T., and Johnston, S.A. (2001). The 19S regulatory particle of the proteasome is required for efficient transcription elongation by RNA polymerase II. *Mol. Cell* 7, 981–991.
- Fersht, A.R., and Daggett, V. (2002). Protein folding and unfolding at atomic resolution. *Cell* 108, this issue, 573–582.
- Fletcher, T.M., Ryu, B.W., Baumann, C.T., Warren, B.S., Fragoso, G., John, S., and Hager, G.L. (2000). Structure and dynamic properties of a glucocorticoid receptor-induced chromatin transition. *Mol. Cell Biol.* 20, 6466–6475.
- Fong, N., and Bentley, D.L. (2001). Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev.* 15, 1783–1795.
- Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature* 414, 929–933.
- Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147.
- Ge, H., and Roeder, R.G. (1994). Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. *Cell* 78, 513–523.
- Gerber, A.P., and Keller, W. (2001). RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem. Sci.* 26, 376–384.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292, 1876–1882.
- Gozes, I., Walker, M.D., Kaye, A.M., and Littauer, U.Z. (1977). Synthesis of tubulin and actin by neuronal and glial nuclear preparations from developing rat brain. *J. Biol. Chem.* 252, 1819–1825.
- Greenleaf, A.L. (1993). Positive patches and negative noodles: linking RNA processing to transcription? *Trends Biochem. Sci.* 18, 117–119.
- Hampsey, M., and Reinberg, D. (1999). RNA polymerase II as a control panel for multiple coactivator complexes. *Curr. Opin. Genet. Dev.* 9, 132–139.
- Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C.D., and Specator, D.L. (2001). Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* 107, 727–738.
- Hirose, Y., and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. *Genes Dev.* 14, 1415–1429.
- Huang, Y., and Steitz, J.A. (2001). Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol. Cell* 7, 899–905.
- Huibregtse, J.M., Yang, J.C., and Beaudenon, S.L. (1997). The large subunit of RNA polymerase II is a substrate of the Rsp5 ubiquitin-protein ligase. *Proc. Natl. Acad. Sci. USA* 94, 3656–3661.
- Hunter, T. (2000). Signaling—2000 and beyond. *Cell* 100, 113–127.
- Iborra, F.J., Jackson, D.A., and Cook, P.R. (2001). Coupled transcription and translation within nuclei of mammalian cells. *Science* 293, 1139–1142.
- Ivanov, D., Kwak, Y.T., Guo, J., and Gaynor, R.B. (2000). Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Mol. Cell Biol.* 20, 2970–2983.
- Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409, 533–538.
- Jeanmougin, F., Wurtz, J.M., Le Douarin, B., Chambon, P., and Losson, R. (1997). The bromodomain revisited. *Trends Biochem. Sci.* 22, 151–153.
- Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. *Science* 293, 1074–1080.
- Kanazawa, S., Okamoto, T., and Peterlin, B.M. (2000). Ta⁺ completes with CIITA for the binding to P-TEFb and blocks the expression MHC class II genes in HIV infection. *Immunity* 12, 61–70.
- Kim, J.B., and Sharp, P.A. (2001). Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. *J. Biol. Chem.* 276, 12317–12323.
- Kim, J.-H., Lane, W.S., and Reinberg, D. (2002). Human elongator facilitates RNA polymerase II transcription through chromatin. *Proc. Natl. Acad. Sci. USA*, in press.
- Kleiman, F.E., and Manley, J.L. (2001). The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. *Cell* 104, 743–753.
- Kloc, M., Zearfoss, N.R., and Etkin, L.D. (2002). Mechanisms of subcellular mRNA localization. *Cell* 108, this issue, 533–544.
- Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* 14, 2452–2460.
- Lee, K.-B., Wong, D., Lippard, S.J., and Sharp, P.A. (2002). Transcription-coupled and DNA damage-dependent ubiquitination or RNA polymerase II in vitro. *Mol. Cell Biol.*, in press.
- Lefstin, J.A., and Yamamoto, K.R. (1998). Allosteric effects of DNA on transcriptional regulators. *Nature* 392, 885–888.
- Lemaire, R., Prasad, J., Kashima, T., Gustafson, J., Manley, J.L., and Lafyatis, R. (2002). Stability of a PKC δ -1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins. *Genes Dev.*, in press.
- Lin, R., Cook, R.G., and Allis, C.D. (1991). Proteolytic removal of core histone amino termini and dephosphorylation of histone H1 correlate with the formation of condensed chromatin and transcriptional silencing during *Tetrahymena* macronuclear development. *Genes Dev.* 5, 1601–1610.
- Litt, M.D., Simpson, M., Gaszner, M., Allis, C.D., and Felsenfeld, G. (2001). Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. *Science* 293, 2453–2455.
- Lo, W.S., Trievel, R.C., Rojas, J.R., Duggan, L., Hsu, J.Y., Allis, C.D., Marmorstein, R., and Berger, S.L. (2000). Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. *Mol. Cell* 5, 917–926.
- Lonard, D.M., Nawaz, Z., Smith, C.L., and O'Malley, B.W. (2000). The 26S proteasome is required for estrogen receptor-alpha and coactivator turnover and for efficient estrogen receptor-alpha transactivation. *Mol. Cell* 5, 939–948.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Rich-

- mond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- Luo, M.J., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. *Proc. Natl. Acad. Sci. USA* 96, 14937–14942.
- Madison-Antenucci, S., Grams, J., and Hajduk, S.L. (2002). Editing machines: the complexities of trypanosome RNA Editing. *Cell* 108, this issue, 435–438.
- McKenna, N.J., and O'Malley, B.W. (2002). Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* 108, this issue, 465–474.
- McNally, J.G., Muller, W.G., Walker, D., Wolford, R., and Hager, G.L. (2000). The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. *Science* 287, 1262–1265.
- Mitsui, A., and Sharp, P.A. (1999). Ubiquitination of RNA polymerase II large subunit signaled by phosphorylation of carboxyl-terminal domain. *Proc. Natl. Acad. Sci. USA* 96, 6054–6059.
- Naar, A.M., Lemon, B.D., and Tjian, R. (2001). Transcriptional coactivator complexes. *Annu. Rev. Biochem.* 70, 475–501.
- Narlikar, G.J., Fan, H.-Y., and Kingston, R.E. (2002). Cooperation between complexes that regulate chromatin structure and transcription. *Cell* 108, this issue, 475–487.
- Nishioka, K., and Reinberg, D. (2001). Transcription. Switching partners in a regulatory tango. *Science* 294, 2497–2498.
- Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, D.C., Tempst, P., and Reinberg, D. (2002). Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes Dev.*, in press.
- Orphanides, G., and Reinberg, D. (2000). RNA polymerase II elongation through chromatin. *Nature* 407, 471–475.
- Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev.* 10, 2657–2683.
- Orphanides, G., LeRoy, G., Chang, C.H., Luse, D.S., and Reinberg, D. (1998). FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* 92, 105–116.
- Orphanides, G., Wu, W.H., Lane, W.S., Hampsey, M., and Reinberg, D. (1999). The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* 400, 284–288.
- Otero, G., Fellows, J., Li, Y., de Bizemont, T., Dirac, A.M., Gustafsson, C.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (1999). Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol. Cell* 3, 109–118.
- Prives, C., and Manley, J.L. (2001). Why is p53 acetylated? *Cell* 107, 815–818.
- Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. *Cell* 108, this issue, 501–512.
- Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell* 108, this issue, 557–572.
- Reed, R., and Hurt, E. (2002). A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* 108, this issue, 523–531.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309.
- Richards, E.J., and Elgin, S.C.R. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* 108, this issue, 489–500.
- Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* 21, 327–335.
- Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., and Stewart, A.F. (2001). The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J.* 20, 7137–7148.
- Roth, S.Y., Denu, J.M., and Allis, C.D. (2001). Histone acetyltransferases. *Annu. Rev. Biochem.* 70, 81–120.
- Ryan, K., Murthy, K.G.K., Kaneko, S., and Manley, J.L. (2002). Requirements of the RNA polymerase II C-terminal domain for reconstituting pre-mRNA 3' cleavage. *Mol. Cell. Biol.* 22, in press.
- Salghetti, S.E., Caudy, A.A., Chenoweth, J.G., and Tansey, W.P. (2001). Regulation of transcriptional activation domain function by ubiquitin. *Science* 293, 1651–1653.
- Shatkin, A.J., and Manley, J.L. (2000). The ends of the affair: capping and polyadenylation. *Nat. Struct. Biol.* 7, 838–842.
- Shuman, S., Broyles, S.S., and Moss, B. (1987). Purification and characterization of a transcription termination factor from vaccinia virus. *J. Biol. Chem.* 262, 12372–12380.
- Simon, I., Barnett, J., Hannett, N., Harbison, C.T., Rinaldi, N.J., Volkert, T.L., Wyrick, J.J., Zeitlinger, J., Gifford, D.K., Jaakkola, T.S., and Young, R.A. (2001). Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* 106, 697–708.
- Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* 403, 41–45.
- Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* 12, 599–606.
- Sun, Z.W., and Hampsey, M. (1996). Synthetic enhancement of a TFIIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo. *Mol. Cell. Biol.* 16, 1557–1566.
- Thomson, S., Clayton, A.L., and Mahadevan, L.C. (2001). Independent dynamic regulation of histone phosphorylation and acetylation during immediate-early gene induction. *Mol. Cell* 8, 1231–1241.
- Tupler, R., Perini, G., and Green, M.R. (2001). Expressing the human genome. *Nature* 409, 832–833.
- Turner, B.M. (2000). Histone acetylation and an epigenetic code. *Bioessays* 22, 836–845.
- Uptain, S.M., Kane, C.M., and Chamberlin, M.J. (1997). Basic mechanisms of transcript elongation and its regulation. *Annu. Rev. Biochem.* 66, 117–172.
- Urnov, F.D., and Wolffe, A.P. (2001). A necessary good: nuclear hormone receptors and their chromatin templates. *Mol. Endocrinol.* 15, 1–16.
- Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., et al. (1998a). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* 12, 343–356.
- Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., and Handa, H. (1998b). Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *EMBO J.* 17, 7395–7403.
- Wen, Y., and Shatkin, A.J. (1999). Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev.* 13, 1774–1779.
- Woychick, N.A., and Hampsey, M. (2002). Simplifying the RNA polymerase II transcription complex: when structure meets function. *Cell* 108, this issue, 453–463.
- Wu, W.H., Pinto, I., Chen, B.S., and Hampsey, M. (1999). Mutational analysis of yeast TFIIIB. A functional relationship between Ssu72 and Sub1/Tsp1 defined by allele-specific interactions with TFIIIB. *Genetics* 153, 643–652.
- Xu, W., Chen, H., Du, K., Asahara, H., Tini, M., Emerson, B.M., Montminy, M., and Evans, R.M. (2001). A transcriptional switch mediated by cofactor methylation. *Science* 294, 2507–2511.
- Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., and Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* 97, 41–51.
- Yanofsky, C. (1988). Transcription attenuation. *J. Biol. Chem.* 263, 609–612.
- Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* 15, 2343–2360.