a0005

2.04 Control of Gene Expression

B J Clark, University of Louisville, Louisville, KY, USA

Abbreviations

ACTR	activator of retinoic acid receptor
AIB-1	amplified in breast cancer-1
ARNT	aryl hydrocarbon receptor nuclear translocator protein
bHLH	basic helix–loop–helix
bZIP	basic leucine zipper
Carm1	coactivator-associated arginine methyltransferase 1
CBP	cap-binding protein
CBP	CREB-binding protein coactivator
ChIP	chromatin immunoprecipitation
CPSF	cleavage-polyadenylation specificity factor
CREB	cAmp-responsive element binding protein
CstF	cleavage-stimulating factor
CTD	carboxy-terminal domain
DNMT	DNA cytosine methyltransferase
DPE	downstream promoter element
dsRNA	double-stranded RNA
EJC	exon-junction complex
Erf	elongation release factor
ESE	exon-splicing enhancer
HAT	histone acetyltransferase
HP1	heterochromatin protein 1
HTH	helix–turn–helix
KID	kinase-interacting domain
LSD1	lysine-specific histone demethylase 1
MBD	methyl-binding domain
MeCP2	methyl CpG-binding protein 2
mRNP	messenger ribonucleoprotein particle
NCoA	nuclear receptor coactivator
N-CoR	nuclear receptor corepressor
NMD	nonsense-mediated decay
NuRD	nucleosome remodeling and histone deacetylation
PABP	poly(A)-binding protein
PAP	poly(A) polymerase
P/CAF	p300/CBP-associated factor
p/CIP	p300/CBP-interacting protein
PIC	preinitiation complex
PKC	protein kinase C
PRMT	protein arginine methyltransferase
PTC	premature termination codon
RISC	RNA-induced silencing complex

RNAi	RNA interference	**SRC**	steroid receptor coactivator
RNAPII	RNA polymerase II	**TAF**	TATA-binding protein and associated factor
SMRT	silencing mediator of retinoid and thyroid receptor	**TBP**	TATA-binding protein
snRNP	small nuclear ribonucleoprotein particle	**TSS**	transcription start site
snRNPs	small nuclear ribonucleoprotein particles	**Upf**	upstream processing factor
		XRE	xenobiotic responsive element

s0005

2.04.1 Introduction

p0005 Proper control of gene expression is fundamental for the correct spatio-temporal expression of a functional protein in response to developmental or environmental signals. Both transcriptional and post-transcriptional mechanisms contribute to the control of gene expression and both of these regulatory arms are susceptible to disruption. Therefore, this chapter will review the current models that guide our understanding of transcriptional and posttranscriptional mechanisms in the study of gene expression.

p0010 Gene transcription is controlled at multiple levels including recruitment of transcription factors and coactivator proteins, chromatin remodeling, and assembly of the general transcriptional machinery and RNA polymerase II (RNAPII) for transcription initiation. Chromatin remodeling involves a dynamic interchange between open (transcriptionally active) and closed (transcriptionally repressed or silenced) chromatin structures and is controlled by enzymatic modification of the DNA and/or histone proteins, such as DNA methylation and histone acetylation, respectively. This type of regulation is epigenetic, in that gene expression is not dictated by information encoded within the gene. Epigenetic mechanisms can be inheritable traits; therefore, the impact of chemical disruption of epigenetic control points in gene regulation may have long-lasting effects. Recruitment of transcription factors and coactivator proteins involves protein–DNA interactions between transcription factors and target gene promoter elements and enhancer sequences as well as protein–protein interactions between coactivator proteins. This step is dependent upon the accessibility of the promoter region to the transcription factors and therefore is dependent upon chromatin remodeling. Factors that influence gene expression at this level include the strength of protein–DNA interactions, function of the coactivator proteins, and cell-specific expression of the transcription factors, all of which can be regulated by posttranslational modification of the transcription factors and coactivator proteins. Thus, xenobiotics that affect signal transduction pathways can alter factor modification and function, modulating either DNA binding or enzymatic activity, thereby indirectly altering gene expression.

p0015 While transcription is a complex and critical control point for gene expression, this process produces a primary RNA transcript, or pre-mRNA, that must be processed to generate a functional mRNA. This requires posttranscriptional processing events, which include modifications of the transcript ends by the addition of 7-methylguanosine nucleotide to the 5′-end generating the 5′-cap structure, addition of a polyadenosine tail to the 3′-end, and removal of introns to generate a contiguous protein-encoding sequence. As will be discussed, these events are actually cotranscriptional processes, but represent unique points that are potentially susceptible to disruption. Additional posttranscriptional control points for gene regulation include nuclear export of messenger ribonucleoprotein particle (mRNP, a complex of mRNA and proteins) and mRNA stability. It is intuitive that mRNA half-life, the transient time the mRNA is present in the cytoplasm and available for translation, can be an important control point for gene expression, yet this is a relatively underexplored area of investigation compared to transcriptional control.

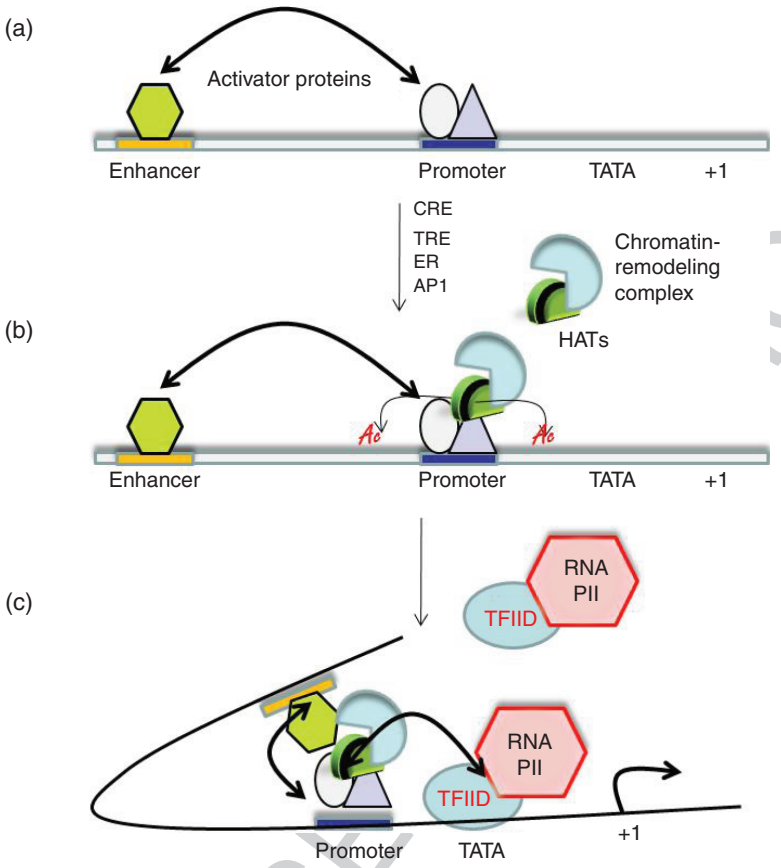
s0010

2.04.2 Transcriptional Control of Gene Expression

s0015

2.04.2.1 Gene Structure and Transcription Initiation

p0020 A gene that is a target for transcription by RNAPII has distinctive functional elements that define or localize the transcription start site (TSS). The TSS is designated as the first nucleotide transcribed into pre-mRNA and is denoted numerically by +1 with the functional elements defined by their relative



f0005 **Figure 1** Gene promoter regulatory elements. A schematic representation of the relative positions of the core promoter (TATA), responsive or regulatory elements (promoter), and enhancer regions within the 5′-flanking region of a gene. (a) Binding of activator proteins to their respective regulatory elements. Examples provided: CRE, cAMP-responsive element; TRE, phorbol ester (TPA)-responsive element; ER, estrogen-responsive element; and AP1, activator protein-1-responsive element. (b) Recruitment of coactivator proteins of the histone acetyltransferase class (HAT) and the ATPase–helicase chromatin-remodeling class promotes histone acetylation and nucleosome repositioning, making the core promoter accessible to RNAPII (c).

position to TSS. The promoter region of a gene is typically within 250 bp upstream of the TSS and contains sequence-specific *cis*-acting DNA elements, or regulatory elements, that are recognized and bound by *trans*-acting DNA-binding proteins (transcription factors) required to initiate transcription (**Figure 1a**). First, a core promoter element is found within 40–50 bp of the TSS and marks the minimal region required for RNAPII to initiate transcription. There are two major classes of core promoter sequences for RNAPII binding: an AT-rich DNA sequence containing a consensus TATAAAA element called the TATA box is found in ~25% of mammalian genes, and CG-rich regions referred to as a TATA-less promoter are found in the remaining 75% of the genes. In general, TATA box elements are found in highly regulated, cell-specifically expressed genes while CG-rich regions serve as promoters for many housekeeping or commonly expressed genes. Recruitment of RNAPII to the core promoter and proper TSS of the gene is controlled by the assembly of the general transcription complex, which together with RNAPII forms the preinitiation complex (PIC).

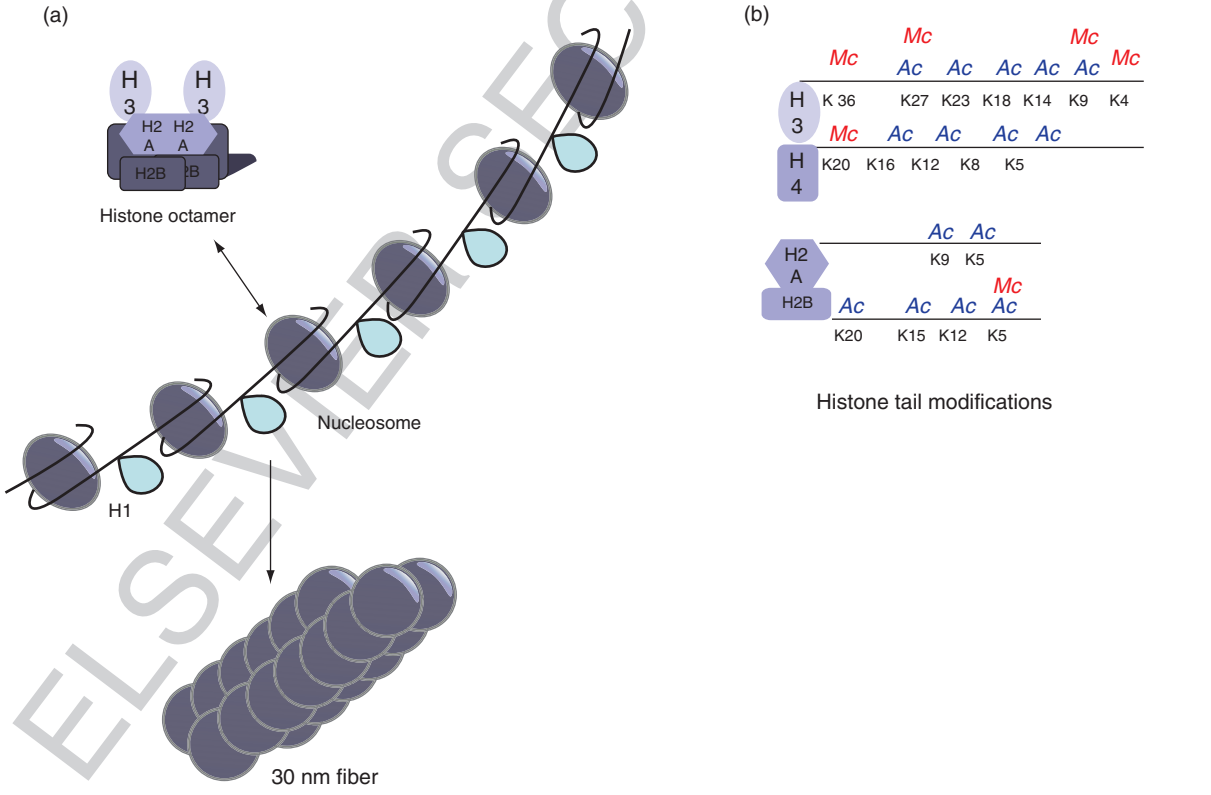
The general transcription complex is composed of p0025
basal transcription factors, denoted by TFII for their association with RNAPII (Thomas and Chiang 2006). The activity of RNAPII is dependent upon these factors to provide functions for (1) site-specific DNA binding at the TATA box (TFIID), (2) stabilizing RNAPII DNA binding and unwinding the DNA (TFIIA, TFIIB, TFIIE, TFIIF, and TFIIH), and (3) phosphorylating the carboxy-terminal domain (CTD) of the large subunit of RNAPII (TFIIH, see Section 2.04.3.1). The composition of the general transcription

complex may be cell specific but common core components include TFIID, TFIIA, TFIIB, and TFIIH. With the exception of TFIIB, each of these TFII components are multisubunit complexes. As an example, the TFIID component contains TBP (TATA-binding protein) and associated factors (TAFs). For TATA-less genes, accurate transcription is mediated by a TFIID-associated TAF binding to a downstream promoter element (DPE) located +28 to +32 relative to the TSS.

p0030 The binding of the general transcription complex and RNAPII to the core promoter is enhanced by the assembly of activator and coactivator proteins at regulatory elements within the promoter, thus the default for most genes is low expression. Additional *cis*-acting DNA regulatory elements, termed enhancers, may lie upstream or downstream of the TSS and, as the name implies, enhance promoter activity. Similar to promoter elements, enhancers bind cognate activator proteins, yet they are distinguished by their ability to function in a position- and orientation-independent manner indicating that they can function from great distances (**Figure 1b**). To summarize transcription initiation, gene regulatory region(s) can be composed of multiple regulatory elements separated by great distances that bind several activator proteins and recruit coactivator proteins that modify chromatin structure leading to enhanced access of the core promoter to general transcription complex and RNAPII to initiate transcription (**Figures 1a–1c** and **2**).

2.04.2.1.1 Activator and coactivator/corepressor proteins s0020

Activator proteins bind to their respective sequence- p0035 specific DNA regulatory elements either constitutively to maintain steady-state levels of expression or in a developmental or regulated manner to activate gene transcription in response to external stimuli such as hormonal or chemical signaling (**Table 1**). The DNA regulatory elements typically contain a two-fold symmetry and the activator



f0010 **Figure 2** Nucleosome structure and histone tail modifications. (a) Schematic representations for the histone octamer composition and nucleosome structure. (b) The lysine residues known to be modified on vertebrate histones by acetylation (Ac) and methylation (Me) are indicated. H3K9 can be dimethylated and H3K27 trimethylated in heterochromatin while these sites are targets for acetylation by HATs during gene activation and H3K4 and H3K27 are methylated during the transcription cycle.

t0005 **Table 1** Classes of transcriptional activators

Target gene expression	*T factor*	*Core RE sequence*	*RE*	*Ligand/STP*	*DBM*
Constitutive	SP1	GGGGCGGGG	Sp1		Zn finger
	c/EBP	CCAAT	CAAT box		bZIP
Developmental	GATA	GATA	GATA		Zn finger
	HOX	TAAT/ATTA	Homeobox		bHTH
	MyoD	CANNTG	E-box		bHLH
Regulated					
Ligand-dependent					
Steroid receptor superfamily			HRE		
Homodimers	ER	AGGTCA-N(3)-TGACCT	ERE	Estradiol, xenoestrogens	Zn finger
	GR	AGAACA-N(3)-TGTTCT	GRE	Glucocorticoids	Zn finger
	PR	AGAACA-N(3)-TGTTCT	PRE	Progestins	Zn finger
	MR	AGAACA-N(3)-TGTTCT	MRE	Mineralocorticoids	Zn finger
RXR heterodimers		AGGTCA-N(x)-DR, IR, ER	NR half-site		
	PPAR	IR-1, DR-4, DR-1	PPARE	PUFA	Zn finger
	LXR	IR-1, DR-4, DR-1	LXRE	Oxysterols	Zn finger
	FXR	IR-1, DR-4, DR-1	FXRE	Bile acids	Zn finger
	CAR	DR-4	CARE	Androstanes, drugs	Zn finger
	PXR	DR-3, ER-6	PXRE	Steroids, drugs	Zn finger
	VDR	DR-3	VDRE	Vitamin D	Zn finger
Xenobiotic responsive	AhR	TNGCGTG	XRE, DRE	TCDD, BP	bHLH
Cellular signaling	CREB/ ATF	TGAC/GTCA	CRE	cAMP-PKA	bZIP
	AP1	TGAG/CTCA	TRE	PKC, cAMP-PKA	bZIP
	HIF1a	[AG]CGTG	ARE	Hypoxia	bHLH

Shown are examples of transcription factors (T Factor) commonly associated with constitutive, developmental, or regulated gene expression. The core regulatory element (RE) sequence is provided where N stands for any nucleotide and the symmetry indicated by DR for direct repeat, IR for inverted repeat, and ER everted repeat. For example, steroid receptor ER homodimers bind to an inverted repeat of AGGTCA separated by 3 nt while the LXR:RXR heterodimer can bind to the sequence AGGTCA as an inverted repeat separated by 1 nt (IR-1), or as a direct repeat separated by either 4 nt (DR-4) or 1 nt (DR-1). The responsive elements (RE) are typically named for the ligand or activator protein followed by the RE designation such as ERE for estrogen-responsive element, CRE for cAMP-responsive element, or LXRE for liver X receptor (LXR) response element. The classical ligands are listed for the regulated T factors of the steroid receptor superfamily with ligand binding promoting dimerization and DNA binding. Many T factors can be regulated by posttranslational modification, and the signal transduction pathway(s) (STP) that controls gene expression via corresponding T factor modification is listed for the cellular signaling T factor. DNA-binding motifs (DBMs) are provided to indicate potential interaction between factors with common motifs. Abbreviations: SP1, specificity protein 1; c/EBP, CAAT enhancer-binding protein; GATA, binds GATA sequence; HOX, homeodomain proteins; MyoD, myogenesis D; ER, estrogen receptor; GR, glucocorticoid receptor; PR, progesterone receptor; MR, mineralocorticoid receptor; PPAR, peroxisomal proliferator receptor; LXR, liver X receptor; CAR, constitutive androstane receptor; PXR pregnane X receptor; VDR, vitamin D receptor; RXR, retinoid X receptor; AhR, aryl-hydrocarbon receptor; CREB/ATF, cAMP-responsive element-binding protein/activating transcription factor; AP1, activator protein-1 commonly composed of c-Jun:c-fos heterodimers or c-Jun heterodimers; HIF1a, hypoxia-inducible factor 1.

proteins bind as homo- or heterodimers. Common DNA- and protein-interacting motifs identified within the activator proteins include the helix–turn–helix (HTH), basic helix–loop–helix (bHLH), zinc finger (Zn finger), and basic leucine zipper (bZIP) motifs. The HTH motif was identified in the bacterial Lac repressor protein and is found in eukaryotic homeodomain proteins that are involved in development. The bHLH motif is distinguished by a conserved region of 40–50 amino acids that folds into two amphipathic helices separated by a loop that generates a protein–protein interaction domain flanked by a stretch of basic residues (basic region) that bind to major groove of the DNA. The bHLH proteins bind to E-box elements as either homo- or heterodimers. A well-studied example of a bHLH protein in gene regulation in toxicology is the aryl-hydrocarbon receptor (AhR), which heterodimerizes with ARNT protein and binds to the xenobiotic responsive element (XRE) of target genes in response to exogenous ligands (**Table 1**). bZIP proteins have a conserved region aligned by the spacing of leucine resides at every seventh amino acid that folds into an α-helix containing a hydrophobic face with evenly

spaced leucines. The interdigitation of the leucines between these α-helices (i.e., a leucine zipper) promotes protein dimerization. A basic region is adjacent to each helix and binds the half-site of the response element; therefore, dimerization is necessary for function (DNA binding). bZIP motifs are commonly found in activators that are regulated in response to cellular signaling, such as the c-fos and c-jun proteins of the AP1 transcription family or the cAMP-responsive element-binding protein (CREB), which are activated by protein kinase C (PKC) and protein kinase A (PKA) signaling pathways, respectively. Another common DNA-binding motif is the Zn finger; the coordination of Zn ions by specifically spaced cysteine (Cys4) or cysteine and histidine (Cys_2His_2) residues helps to form adjacent finger-like structures that bind the major groove of DNA and promote protein dimerization. The Zn finger Cys_2His_2 motif is shared by members of the steroid hormone/nuclear receptor superfamily where ligand binding enhances homo- or heterodimer formation and DNA binding (**Table 1**).

p0040 Coactivator and corepressor proteins do not bind DNA directly but are recruited to the promoters via protein interactions with the activator proteins (reviewed in Aoyagi and Archer 2008; Lonard and O'Malley 2007; Rosenfeld *et al.* 2006). Conserved protein-interacting motifs include the LXXLL motif for nuclear receptor–coactivator interactions and kinase-interacting domain (KID) for CREB/ATF family members and coactivators (Rosenfeld *et al.* 2006; Savkur and Burris 2004). Protein–protein interactions can be dependent upon posttranslational modification(s), such as phosphorylation of CREB on Ser133 by PKA or Ca^{2+} signaling, which results in increased association with CREB-binding protein (CBP/p300) (Plevin *et al.* 2005; Savkur and Burris 2004; Xu *et al.* 1999, 2007). Co-activators classically have intrinsic histone acetyltransferase activity (HAT) that catalyze the transfer of the acetylgroup from Coenzyme A to the epsilon-amino group of a lysine residue within the amino terminal region of one of the core histone proteins (H2A, H2B, H3 and H4) forming an amide bond. Members of the HAT family include CBP/p300 (CREB-binding protein), P/CAF (p300/CBP-associated factor), and the p160 family. The p160 family of coactivators interact with nuclear hormone receptors and include SRC-1/NCOA-1 (steroid receptor coactivator-1/nuclear receptor coactivator-1), SRC-2/NCOA-2, and SRC-3 (P/CIP, AIB-1, NCOA-3, ACTR). Thus, the recruitment of HAT coactivators to the promoter results in local histone acetylation and activates gene transcription (discussed in Section 2.04.2.2.3). Additional histone-modifying enzymes have been recently shown to contribute to coactivator function. Specifically, coactivator-associated arginine methyltransferase 1 (Carm1) and protein arginine methyltransferase (PRMT) have been identified as p160 coactivator-interacting proteins that methylate histone tails and further enhance transcription (Wysocka *et al.* 2006).

p0045 Conversely, a gene may be actively silenced by the constitutive presence of a corepressor complex on the promoter. Corepressor proteins, like coactivators, are associated with DNA-bound activator proteins via protein–protein interactions thereby localizing the corepressor to specific promoter regions of the gene (Rosenfeld *et al.* 2006). The repressor protein then recruits HDAC proteins, enzymes with intrinsic histone deacetylase activity that help maintain histones in the deacetylated state and promote gene silencing. mSin3 and NuRD (nucleosome remodeling and histone deacetylation) are two common repressors utilized by multiple activator proteins to recruit HDAC1 to the promoter. Nuclear receptors also can utilize corepressors SMRT (silencing mediator of retinoid and thyroid receptors) and N-CoR (nuclear receptor corepressor) to recruit HDACs either directly or via Sin3A complex formation (Downes *et al.* 2000; Kao *et al.* 2000).

2.04.2.2 Chromatin Structure

s0025

2.04.2.2.1 Nucleosome structure

s0030

p0050 In eukaryotic cells, the total 2 m length of DNA must be condensed to fit within the $\sim 20\,\mu m^3$ space of the nucleus. This is achieved by assembly of highly ordered DNA–protein interactions that form the nucleosome, the structural unit of chromatin (Tremethick 2007). The nucleosome is composed of a protein particle containing histone proteins H2A, H2B, H3, and H4 arranged in two sets of H3/H4 and H2A/H2B heterodimers. These form an octomer core with $\sim$146 bp of DNA that wraps around the octomer (**Figure 2a**). The histone–DNA interactions are facilitated by the positively charged amino-terminal portion of the histone proteins that extend from the core and interact with the negatively charged phosphate backbone of DNA. Each nucleosome is separated by a linker region of DNA $\sim$20–100 bp in length that is bound by histone H1. H1 promotes coiling of nucleosomes into the classical 30 nm fiber of DNA observed within cells (Bednar *et al.* 1998).

The significance of nucleosome structure in condensed chromatin in terms of gene regulation is that the positioning of DNA within the nucleosome dictates accessibility of a gene to the general transcriptional machinery and RNAPII. The accessibility of a gene is further influenced by epigenetic factors, such as DNA methylation and posttranslational histone modifications that close or open chromatin structure leading to inactive or active transcription, respectively (Latham and Dent 2007).

s0035 #### 2.04.2.2.2 DNA methylation

p0055 DNA methylation occurs at C5 of the cytosine base within the context of CpG dinucleotide repeat. The presence of CpG dinucleotides within the genome is relatively rare, that is, at only 5–10% of the predicted frequency for this dinucleotide. However, ~70–80% of CpGs are methylated with methylated DNA associated with inactive chromatin (Beck and Rakyan 2008). These DNA methylation patterns are proposed to be stable, heritable traits that contribute to imprinting and X-chromosome inactivation (Cohen and Lee 2002). One mechanism proposed for gene silencing by DNA methylation is mediated by the family of methyl-CpG-binding proteins that share a common methyl-binding domain (MBD). Methyl CpG-binding protein 2 (MeCP2) is an example of an MBD protein that binds methylated DNA and promotes complex formation with repressor protein Sin3A and histone deacetylases HDAC1 or HDAC2. Thus, methylated DNA may trigger formation of corepressor complexes leading to histone deacetylation that favors inactive heterochromatin structure (Clouaire and Stancheva 2008).

p0060 Many gene promoter regions are GC rich with an apparent bias for the CpG dinucleotide repeats within these regions. In some cases, CpG islands, long stretches of 500–5000 bp that are enriched in CpG dinucleotide repeats, are found in gene promoter regions. Paradoxically, however, the CpG islands in these regions are typically not methylated and hypermethylation of these sites will repress transcription. The concept that aberrant promoter methylation may lead to gene inactivation is supported by the finding that in some cancers the loss of tumor suppressor gene expression has been correlated with overexpression of DNA methyltransferases (DNMTs), the enzymes responsible for cytosine methylation and promoter hypermethylation (Esteller 2008). A classical example is retinoblastoma protein promoter hypermethylation in retinoblastoma (Cohen *et al.* 2008). However, cancer cells in general display an overall decrease in DNA methylation, for example, global hypomethylation, and in some cases this affects promoter regions of normally suppressed oncogenes. Thus, a general disruption of DNA methylation patterns can have a significant impact on gene expression.

p0065 The regulation of DNA methyltransferase expression and function is a potential target for controlling gene expression (Gius *et al.* 2005). In eukaryotes, DNMTs function in either establishing methylation patterns during embryogenesis, for example, the *de novo* methyltransferases DNMT3A and DNMT3b, or maintaining the established methylation patterns during DNA replication. DNMT1 is considered as the maintenance methyltransferase but this function may be performed by other family members as well. DNMT inhibition by arsenic, a heavy metal considered as a major environmental contaminant, was shown to be correlated with promoter hypomethylation and reactivation of tumor suppressor genes, such as p16INK4a, in liver cancer cell lines (Cui *et al.* 2006). Limited prenatal exposure to arsenite resulted in decreased ERα promoter methylation in liver from adult male mice, leading to increased estrogen-mediated cell proliferation and liver cancer (Waalkes *et al.* 2004), supporting the idea that prenatal methylation patterning persists into adulthood and indicating that alterations to DNA methylation during the prenatal periods susceptible to epigenetic disruption may impact adult function. If the methylation pattern is disrupted in germ cells, the effects can be perpetuated through several generations. This transgenerational potential of such an alteration in DNA methylation was demonstrated by studies on vinclozolin, an antiandrogenic fungicide that disrupted testis function in adult rats after exposure *in utero* during the critical embryonic period for testis differentiation (Anway *et al.* 2005; Uzumcu *et al.* 2004). Strikingly, prenatal vinclozolin exposure altered germ cell DNA methylation patterns in male offspring and these methylation patterns were still present after four generations along with the testicular disorders (Anway *et al.* 2005; Chang *et al.* 2006).

p0070 To gain greater insight into the function(s) of DNA methylation in gene regulation, genome-wide methylation patterns, termed 'methylomes,' are being established (Beck and Rakyan 2008; Schones and Zhao 2008; Shen *et al.* 2007; Zhang *et al.* 2006). These studies have identified different classes of CpG islands based on CpG densities and surrounding *cis*-acting DNA sequences that may influence DNA methylation patterns in a gene- and

tissue-specific manner (Bock *et al.* 2007; Shen *et al.* 2007). For example, the *Arabidopsis* methylome confirmed that CpG regions in inactive heterochromatin were highly methylated while gene promoter regions were typically not methylated (Zhang *et al.* 2006). However, methylation of sparse CpG regions within the transcribed regions of genes was also detected and correlated with constitutive transcriptional activity where the highly methylated genes had greater transcriptional activity compared to genes that lacked methylation (Zhang *et al.* 2006). Global DNA methylation analysis of CpG dense promoter regions in normal peripheral blood identified a small subset of genes that were highly methylated and suppressed in somatic cells yet hypomethylated and expressed in testicular germ cells (Shen *et al.* 2007). Thus, DNA promoter methylation may represent a broader mechanism for gene suppression in normal somatic cells than previously appreciated.

s0040 ***2.04.2.2.3 Histone modifications***

p0075 Histone proteins are covalently modified via acetylation, methylation, phosphorylation, ubiquitylation, ADP-ribosylation, and sumoylation by recruitment of nonhistone proteins such as the coactivator HATs, methylases, and kinases to a localized region of the gene. Thus, as stated above, chromatin structure can be controlled locally by specific cofactor association with activator proteins of the gene. A major question is how multiple histone modifications are linked to function and the histone code hypothesis proposes that it is the combination or sequence of modifications that determine an active or repressed state for gene expression (Aoyagi and Archer 2008; Latham and Dent 2007; Rosenfeld *et al.* 2006; Shahbazian and Grunstein 2007; Turner 2007).

p0080 Histone acetylation has been well characterized and specific lysine residues in the N-terminal extensions of the H2A, H2B, H3, and H4 proteins have been identified as acetylation targets (**Figure 2b**). Correlating the specific acetylation patterns with subsequent protein interactions and nucleosome structure modification(s) and gene activation is complex and may be cell- and/or gene-specific in some cases (Kimura *et al.* 2005; Shahbazian and Grunstein 2007). In general, active transcription correlates with acetylated histones H3 and H4. One putative mechanism for histone acetylation resulting in a more open chromatin structure is the interaction of positively charged lysine residues with the DNA phosphate backbone providing stability to the nucleosome structure, and introduction of the acetyl groups decreases the interaction between the DNA and the histones. A second consequence of histone acetylation is specific acetylated residues provide binding sites for bromodomain-containing proteins (Hassan *et al.* 2007; Kanno *et al.* 2004). These proteins are commonly found in chromatin-remodeling complexes and HATs (Rosenfeld *et al.* 2006). For example, acetylation of histone H4 at K12 has been shown to bind Brd2, a component of the SWI/SNF chromatin-remodeling complex, leading to an open chromatin structure and active transcription (Martens and Winston 2003) (see Section 2.04.2.2.4 and Chapter 2.20).

p0085 Histone methylation is found within both heterochromatin and euchromatin regions of the genome with mono-, di-, and trimethylated modifications of specific lysine residues of H3 and H4. The general pattern for histone methylation in heterochromatin is di- and trimethylation of lysine residues 9 and 27 of histone H3 (H3K9me2 and H3K27me3) while in euchromatin monomethylation of H3K9 and H3K27 along with H3K4me, H4K20me, and H2BK5me is present (**Figure 2b**). The di- and trimethylated species of H3K9 and H3K27 may contribute to heterochromatin structure indirectly as these modified histones are bound by the heterochromatin protein 1 (HP1), which promotes closed chromatin structure (Schones and Zhao 2008). However, histone H3 becomes methylated at K4 and K36 at the 5′-end and throughout the coding region, respectively, during transcription. These modifications appear to help recruit coactivator proteins, leading to histone H3 and H4 acetylation at the promoter, thereby contributing to enhanced transcription. Conversely, histone H3K4 demethylation by lysine-specific histone demethylase 1 (LSD1) appears to contribute to transcriptional repression of some target genes (Forneris *et al.* 2008). LSD1 was recently identified as part of a corepressor complex containing HDAC proteins as well. Thus, the pattern(s) of posttranslational modification of histone tails appears to provide unique binding sites for coactivator proteins and subsequent chromatin remodeling. However, posttranslation modification of the histone tails appears to be a dynamic process with the interchange of HAT coactivator for HDACs corepressor complexes and/or histone methylases for demethylases modulating expression. Recent advances using chromatin immunoprecipitation (ChIP) assays have monitored the temporal association of activator and coactivator/corepressor proteins with gene promoter regions and support a model for

cycling of transcription factors on and off the promoter with specific periodicity (Aoyagi and Archer 2008; Metivier *et al.* 2003, 2008; Rosenfeld *et al.* 2006). Thus, the complexity is increased by the potential of cyclic-dependent recruitment of coactivator versus corepressor complexes.

s0045 #### 2.04.2.2.4 Chromatin remodeling

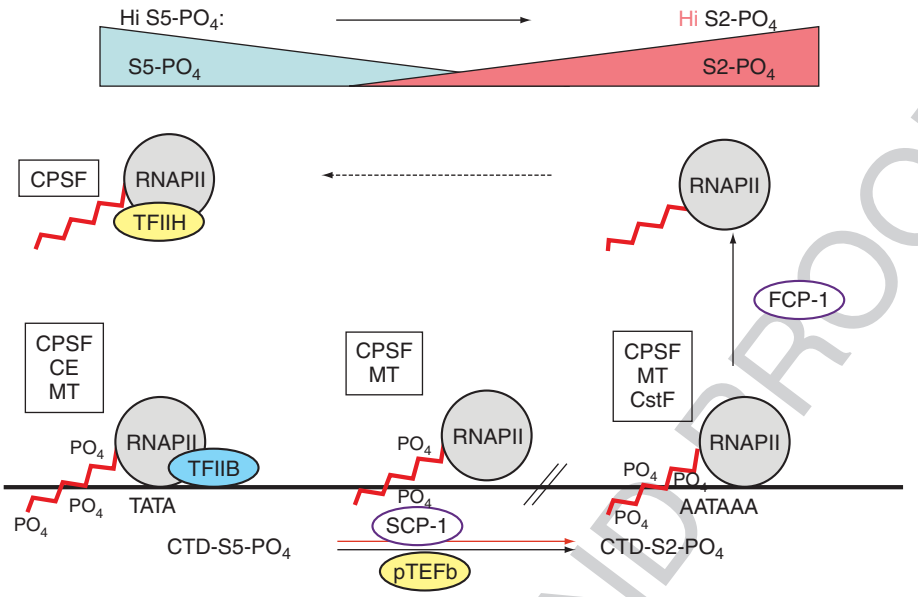
p0090 The model for chromatin remodeling in gene activation states that activator proteins function to increase the rate of transcription initiation by facilitating RNAPII binding to the core promoter (**Figure 3**). One mechanism is by recruitment of coactivator proteins that modify the chromatin structure around the promoter to make this region of the gene more accessible to RNAPII. As described above, one class of coactivator proteins has intrinsic HAT activity, with the yeast Gcn5p and Esa1p/mof/Tip60, and mammalian GCN5/PCAF, CBP/p300, and p160 family of proteins being the best characterized of this class. A second class of coactivator proteins belongs to the family of DNA-dependent ATPases, typified by the yeast SWI/SNF complex, that bind as part of a large multiprotein complex to the promoter and alter chromatin structure via ATP-dependent helicase activity resulting in repositioning of nucleosomes (see Chapter 2.20). The HATs and chromatin-remodeling complexes may work independently or together to promote gene activation. The precise order of events for assembly of an active promoter complex is not well defined and most likely is promoter- and cell-specific. However, in the simplest terms (see **Figure 1**), the model for transcription initiation starts with an activator protein binding directly to its cognate sequence-specific DNA element and recruiting coactivator proteins that locally remodel chromatin structure. Recruitment of HATs results in acetylation of histone H3 and H4 tails that contributes to loosening of the DNA–histone interactions. Acetylated histones also provide binding sites for bromodomain-containing proteins and recruitment of ATP-dependent chromatin-remodeling complex that promotes nucleosome sliding. The result is an open structure around the core promoter that is accessible to the general transcription complex and RNAPII. Activator proteins also have been shown to interact directly with the general transcription factors and RNAPII to enhance DNA binding and transcription initiation, increasing the complexity for transcription initiation (see Gangaraju and Bartholomew 2007; Hager *et al.* 2004; Korzus *et al.* 1998; Kumaran *et al.* 2008; Lonard and O'Malley 2007; Mellor 2006; Nunez *et al.* 2008; Rosenfeld *et al.* 2006).

p0095 Current studies indicate that the concept of open versus condensed chromatin determining active versus inactive transcription may not be applicable to all genes. Specifically, active genes have been localized in regions of condensed chromatin and, conversely, inactive genes in open chromatin regions (Gilbert 2004). Thus, histone and DNA modifications may serve additional roles other than chromatin remodeling for transcription initiation (Spector 2004). Another unresolved question is where active transcription occurs. The above model implies that RNAPII is recruited to the DNA at promoter regions. However, a challenge to this model comes from studies that monitored active transcription in real time and showed the process was localized to discrete subnuclear regions termed transcription factories (Jackson 1998). Active genes colocalized to the factory site (reviewed in Schneider and Grosschedl 2007), including genes separated by great distances (Kumaran *et al.* 2008; Nunez *et al.* 2008; Osborne *et al.* 2004, 2007), suggesting the DNA is dynamic and moving to sites of RNAPII localization for initiation of transcription. The composition of the transcription factory and elucidating the mechanisms that control gene recruitment to this site are important unresolved issues that need to be addressed.

2.04.3 Control of Gene Expression by Pre-mRNA Processing

s0050

p0100 To generate a functional protein-encoding mRNA, the primary RNA transcript produced by RNAPII must be modified at the 5′- and 3′-ends to enhance stability and the noncoding intronic sequences must be removed by splicing. Processing of pre-mRNA transcripts refers to these posttranscriptional modifications and each processing event represents a potential control point in gene expression. However, over the past decade, it has become clear that all the processing events are enhanced during ongoing transcription (Aguilera 2005; Bentley 2005; Hagiwara and Nojima 2007; Kornblihtt *et al.* 2004). Thus, pre-mRNA-processing events are cotranscriptional processes. In particular, the processing factors have been shown to be recruited to the carboxy-terminal domain (CTD) tail of the large subunit of RNAPII and this recruitment is dependent upon the CTD phosphorylation status and RNAPII transient down the gene.



CTD PO4	Enzyme		Function
Serine 5	TFIIH/Cdk7	Kin28	Kinase
	Scp1	Ssu72	Phosphatase
Serine 2	P-TEFb/Cdk9/TFIIH	Ctk1	Kinase
	Fcp1		Phosphatase
5'CAP	CE (capping enzyme)	CET-1	Triphosphatase
		CEG-1	Guanylyltransferase
	MT (methyltransferase)	Abd1	Methyltransferase
Polyadenylation	CPSF (cleavage-polyadenylation specificity factor)	Yhh1	Binds AAUAAA
	CstF (cleavge-stimulating factor)	Rna14, Rna15	Binds GU-rich region
	CF-I, CF-II (cleavage factors)	Pcf11	Endonucleases
	PAP (polyadenylation polymerase)	Pap1	Polyadenylation

f0015 **Figure 3** Cotranscriptional processing of pre-mRNA. The dynamic phosphorylation status of the CTD tail of RNAPII is depicted as a shift from high serine 5 phosphorylation at the 5′-end of the gene to high serine 2 phosphorylation at the 3′-end of the gene. The CTD tail is shown as a jagged line extending from RNAPII. The kinases and phosphatases (circles) and processing factors associated with either serine 5 or serine 2 CTD phosphorylation are indicated (box above the CTD tail). The enzymes and functions are provided in the table below the figure.

s0055 ### 2.04.3.1 RNA Polymerase II CTD Phosphorylation

p0105 RNAPII is composed of 12 subunits that function to synthesize RNA from a DNA template, and therefore is classified as a DNA-dependent RNA polymerase. The carboxy-terminus of the large subunit of RNAPII (Rbp1) has a unique structure composed of heptad repeats of the sequence Y-S2-P-T-S5-P-S termed the carboxy-terminal domain. In higher eukaryotes, this heptad is repeated 52 times and serines 2 and 5 are targets of kinase activity and go through dynamic phosphorylation cycles (Buratowski 2003). Serine 5 is phosphorylated (S5-PO_4) by a cyclin-dependent kinase, Cdk7, associated with TFIIH and this modification is required for the release of RNAPII from the promoter and a switch to transcription elongation (**Figure 3**). During the elongation phase, serine 2 is phosphorylated by kinases associated with either TFIIH or elongation factors (pTEF) and serine 5 is dephosphorylated by phosphatases, such as Scp1, recruited to the CTD. Serine 2-specific phosphatases dephosphorylate the CTD and RNAPII returns to a hypophosphorylated state that is set for another round of transcription initiation. ChIP studies support the model that serine 5 phosphorylation and associated kinases are present at the promoter region and 5′-end of the gene and decline at the 3′-end of the gene while serine 2 phosphorylation and associated kinases increase later during the elongation cycle (**Figure 3**) (Aguilera 2005; Ahn *et al.* 2004; Hirose and Ohkuma 2007; Komarnitsky *et al.* 2000). Similarly, pre-mRNA processing factors were shown to be associated preferentially with either the serine 5 or serine 2 phosphorylation status of the CTD, indicating a link between transcription elongation and processing, supporting that the CTD tail of Rbp1 recruits processing enzymes specific for the nascent or pre-mRNA transcripts (Kim *et al.* 2004; Lima 2005; McCracken *et al.* 1997a,b; Millhouse and Manley 2005; Proudfoot *et al.* 2002). Deletion of the CTD tail results in decreased efficiency or loss of all processing steps.

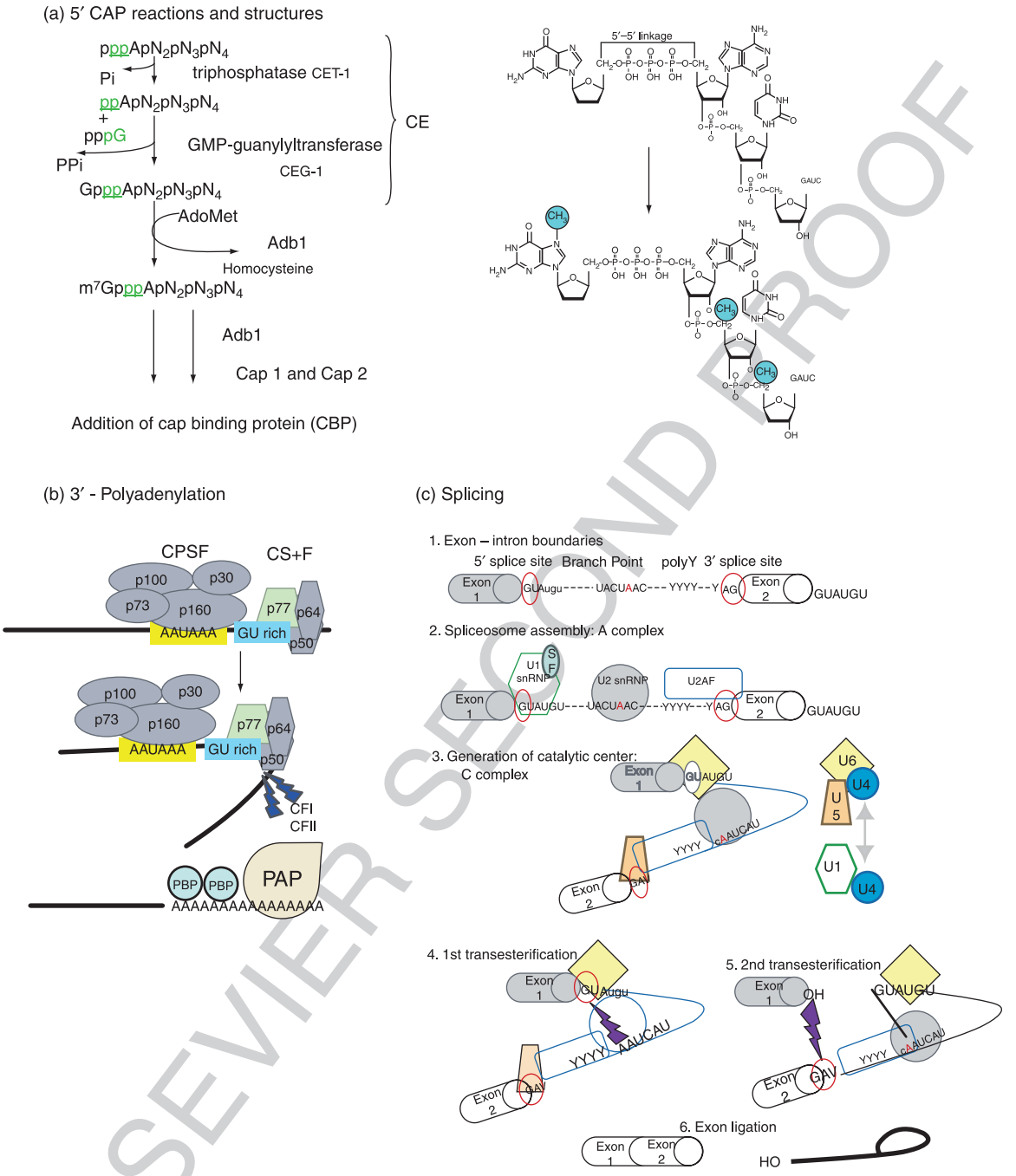
s0060 ### 2.04.3.2 5′-pre-mRNA Processing

p0110 Addition of the 7-methyl guanine nucleotide to the 5′-end of the nascent pre-mRNA transcript is characteristic of RNAPII-dependent transcripts and this modification is referred as the 5′-cap. Guanine is attached by a 5′–5′ triphosphate linkage to the first nucleotide of the nascent transcript, which is usually a purine (**Figure 4a**). In brief, RNA triphosphatase catalyzes the cleavage of the 5′-γ-phosphate group followed by RNA guanylyltransferase activity that uses GTP hydrolysis to form GTP–RNA transcript cap structure (**Figure 4a**). In higher eukaryotes, the capping enzyme (CE) is a bifunctional enzyme containing both the triphosphatase and guanylyltransferase activities while in yeast these activities are encoded by separate genes: CET-1 (triphosphatase) and CEG-1 (guanylyltransferase) (**Figure 3**). Lastly, the N7 of guanine is methylated by RNA guanine 7-methyltransferase. Additional methylation on C2 of the ribose backbone within the first two nucleotides can also be present (**Figure 4a**). Protein–protein interactions between guanylyltransferase (CE) and the RNAPII CTD tail occur only when serine 5 is phosphorylated. ChIP analysis supports that CE association with the S5-PO_4 CTD and at the promoter and 5′-end of reporter genes with loss of CE association with the CTD tail correlated with loss of serine 5 phosphorylation (**Figure 3**) (Komarnitsky *et al.* 2000).

2.04.3.3 3′-pre-mRNA Processing s0065

p0115 Mature mRNAs contain a stretch of adenosines at the 3′-end of the transcript referred to as the poly(A) tail. 3′-end processing is the addition of the poly(A) tail and involves two steps: endonucleolytic cleavage followed by polyadenylation. Cleavage of the pre-mRNA requires recognition of an AAUAAA sequence near the 3′-end of the nascent transcript. This sequence is conserved and is located ~10–30 nt upstream of a GU-rich sequence. Cleavage-polyadenylation specificity factor (CPSF) binds as a tetramer to the AAUAAA element and binding of CPSF promotes binding of a second heterotrimeric protein complex, the cleavage-stimulating factor (CstF), to the GU-rich sequence (**Figures 3** and **4b**). CPSF and CstF facilitate structural changes in the transcript that promote access to cleavage factors I and II (CFI and CFII) that purportedly have endonuclease activity that cleaves the RNA providing the 3′-end for subsequent polyadenylation. Poly(A) polymerase (PAP) catalyzes the processive addition of adenines to achieve a final length of ~200–300 nt. PAP function is enhanced by binding of poly(A)-binding protein (PABP) to poly(A) stretches of ~10 nt in length, which helps to stabilize the PAP interaction with DNA.

p0120 The CPSF complex is associated with TFIID and binds to Rbp1 CTD tail independent of



f0020 **Figure 4** Pre-mRNA-processing mechanisms. (a) 5′-end cap modification. (b) 3′-Polyadenylation. (c) Spliceosome assembly. Complex A is formed by U1 and U2 binding to their respective 5′-splice site and branch point elements along with associated factors (1 and 2). The heterotrimer U4/U6/U5 incorporates into the spliceosome with U5 binding at the 5′-splice site and U6 binds U2 to form the B1 complex (not shown). Subsequent rearrangement of the RNA:RNA interactions results in U1 release and U6 interaction with the 5′-splice site to form the B2 complex (not shown). The B1 and B2 complexes are intermediate stages in the formation of the catalytic C1 complex, defined by U5 binding to the 3′-splice site, release of U4, and specific U2:U6 RNA:RNA interactions that form the catalytic center for the first transesterification reaction (3 and 4). U2:U6:U5 complex remains bound to the intron lariat and the second transesterification occurs (5), releasing the lariat and ligating the upstream and downstream exons (6).

phosphorylation status; therefore, it appears that this 3′-processing factor is present at promoter and constitutively associated with RNAPII (McCracken *et al.* 1997b). CFI and CFII, however, are recruited to the CTD tail only when serine 2 is phosphorylated and thus this represents a later event in transcriptional elongation (**Figure 3**) (Ahn *et al.* 2004).

s0070 ### 2.04.3.4 Splicing: The Process of Intron Removal

p0125 Protein-encoding sequences within a gene are localized within multiple discrete regions termed exons and are interrupted by noncoding intervening regions termed introns (Sharp 2005). In higher eukaryotes, the exonic sequences are relatively short 100–200 bp regions while introns typically span 1.0–2.0 kb in length. Splicing is the process of intron removal and exon ligation to generate a contiguous sequence encoding a full-length functional protein. Splicing is a complex process with sequential steps used to explain the formation of the spliceosome, a megadalton ribonucleoprotein particle that catalyzes the splicing reaction(s). The spliceosome is composed of six small nuclear ribonucleoprotein particles (snRNPs called snurps) and auxiliary proteins. Each snRNP contains 1 small RNA (U1–U6) and ~20 associated proteins and snRNPs are named for the RNA component, such as U1snRNP, U2snRNP, U3snRNP, U4snRNP, U5snRNP, and U6snRNP. Important auxiliary proteins of the spliceosome belong to the SR protein family that is characterized by an arginine–serine-rich C-terminal domain that participates in protein–protein interactions and an N-terminal RNA-binding domain (Hertel and Graveley 2005; Sanford *et al.* 2005). This family of proteins are proposed to act as bridging proteins that help define the splice sites and exonic sequences as will be discussed below.

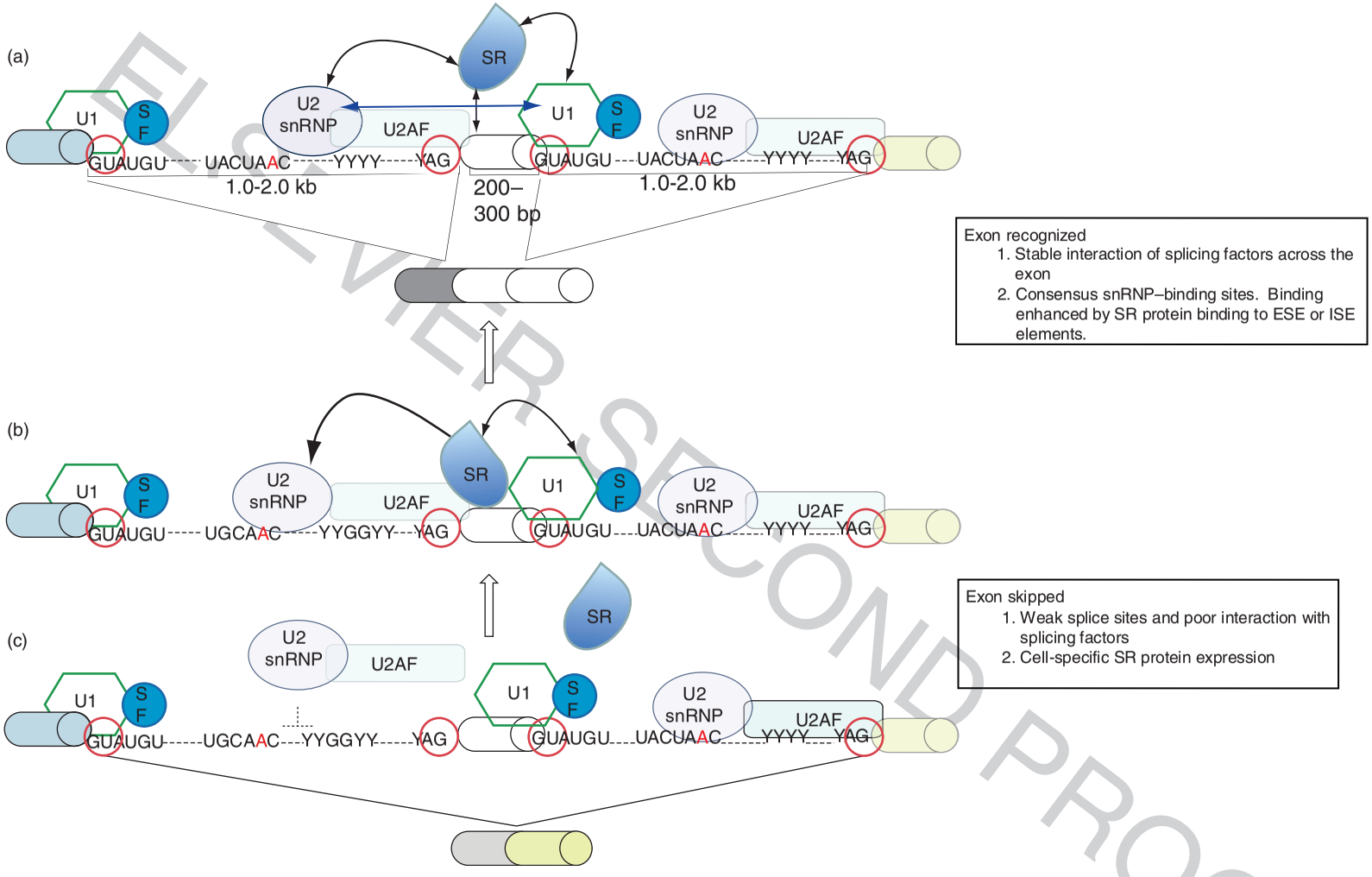
p0130 The first step in splicing is recognition of conserved sequence-specific *cis*-acting elements by the snRNPs. These include the splice sites, branch site, and polypyrimidine tract elements (**Figure 4c**). Splice sites define the intron boundaries and follow the GU–AG rule with conserved GU and AG dinucleotides at the 5′- and 3′-ends of the intron, respectively. The branch site element has the sequence 5′- UACUAAC-3′, located 18–40 nt upstream of the 3′- AG splice site, and contains the nucleophile (underlined A) for the first catalytic step of splicing. U1snRNP binds to the 5′- splice site and binding is enhanced by the presence of SR proteins. U2snRNP binds to the branch site and the polypyrimidine track lies between the branch site and the 3′-splice site and binding of auxiliary proteins to this track enhances U2snRNP binding. One view of splicing is to consider that the snRNP RNAs drive complementary base pairing with the target RNA to generate a catalytic core for the splicing reaction (Valadkhan 2007). In brief, the U4–U5–U6 snRNPs form a heterotrimer stabilized by RNA:RNA base pairing between U4 and U6 RNA. Through mutually exclusive binding, the snRNP RNA:RNA base pairing is exchanged for snRNP RNA:intron RNA base pairing. U6 RNA displaces U1 at the 5′-splice site and U6 RNA:U2RNA interactions displace U6 RNA:U4RNA interactions (**Figure 4c**). These RNA rearrangements are catalyzed by the ATP-dependent RNA helicase activity of a DExD box protein family member that is associated with the U5 snRNP (Bleichert and Baserga 2007; Rocak and Linder 2004). The U6:U2 RNA:RNA interactions generate the catalytic core that promotes two sequential transesterification reactions that remove the intravening RNA sequence as a lariat structure and ligation of upstream exon and downstream exons (Madhani and Guthrie 1992).

p0135 In higher eukaryotic cells where the typical protein-encoding genes are comprised of multiple exons of relatively short length separated by introns 10–20 times their size, an exon definition model has been proposed, which states that the spliceosome scans for closely spaced splice sites (Berget 1995). Therefore, the upstream 3′-splice site and downstream 5′-splice site that span an internal exon define the exonic sequence (**Figure 5**). In this model, the SR proteins serve important roles as bridging factors via binding to short degenerate sequences, termed exon-splicing enhancers (ESEs), and protein–protein interactions to establish a link between the upstream 3′-U2snRNP and the downstream 5′-U1snRNP splice sites (Blencowe 2006; Wang and Burge 2008).

2.04.3.5 Alternative Splicing

s0075

p0140 Alternative splicing is a process by which different combinations of exons are joined together and result in the production of multiple forms of mRNA from a single pre-mRNA (Blencowe 2006). The completion of the human genome project revealed a discord between the genome and the proteome with too few genes identified to account for the number of proteins (Birney *et al.* 2007; Pennisi 2005). Thus, alternative splicing has gained renewed interest as



f0025 **Figure 5** The exon definition model and alternative splicing. Binding of the U1snRNP and U2snRNP to the 5′-splice site and branch point splice site is indicated. (a) The interaction between the upstream U2snRNP and the downstream U1snRNP that flank the internal exon (open cylinder) is indicated by the double-headed arrow. Binding of SR protein to an ESE will help stabilize the interaction and promote exon recognition resulting in inclusion of the internal exon. (b and c) Nonconsensus branch point sequence can result in weak U2snRNP interaction and loss of the interaction between the upstream U2snRNP and the downstream U1snRNP that flank the internal exon. (c) No exon definition results in loss of exon in spliced transcript. (b) The presence of an SR protein can enhance U2snRNP binding and reassemble the interaction between the upstream U2snRNP and the downstream U1snRNP providing exon definition. In this model (b and c), the internal exon is alternatively spliced based on cell-specific or regulated expression of SR proteins.

an important mechanism in the control of gene expression. There is a predicted average of 2–3 transcripts per gene with many transcripts expressed ubiquitously while alternative forms are expressed by developmental or cell-specific mechanisms. From the exon definition model perspective, deviations from consensus sequences and cell type-specific expression of SR proteins will influence whether an alternative exon is recognized (defined) or skipped (spliced out of the final transcript) (Blencowe 2006; Wang and Burge 2008). Second, elements within either the exonic or intronic sequences may be recognized by SR proteins and either enhance or suppress splicing. These elements are referred to as exon or intron splicing enhancers or silencers, respectively (**Figure 5**) (Wang and Burge 2008). The generalized concept for alternative splicing is that weak splice sites, defined as a sequence that deviates from the consensus, do not stably bind the snRNP resulting in exon skipping. However, the presence of SR protein can stabilize snRNP binding and promote exon recognition resulting in retaining the exon in the mRNA. Models for cell-specific alternative splicing propose that the expression levels of SR proteins influence splicing: the same target exon will be skipped in one cell type that has low or absence SR protein expression while retained in another cell type by the presence of the SR protein.

p0145 The basis for certain diseases has been linked to mutations within splice sites or within the exonic sequences affecting the function of ESE elements (Stoilov *et al.* 2002). Potential adverse functional consequences of aberrant alternative splicing include decreased protein expression if alternative exon contains an in-frame stop codon, or production of dominant negative form of a protein if a functional domain is spliced out (Blencowe 2006).

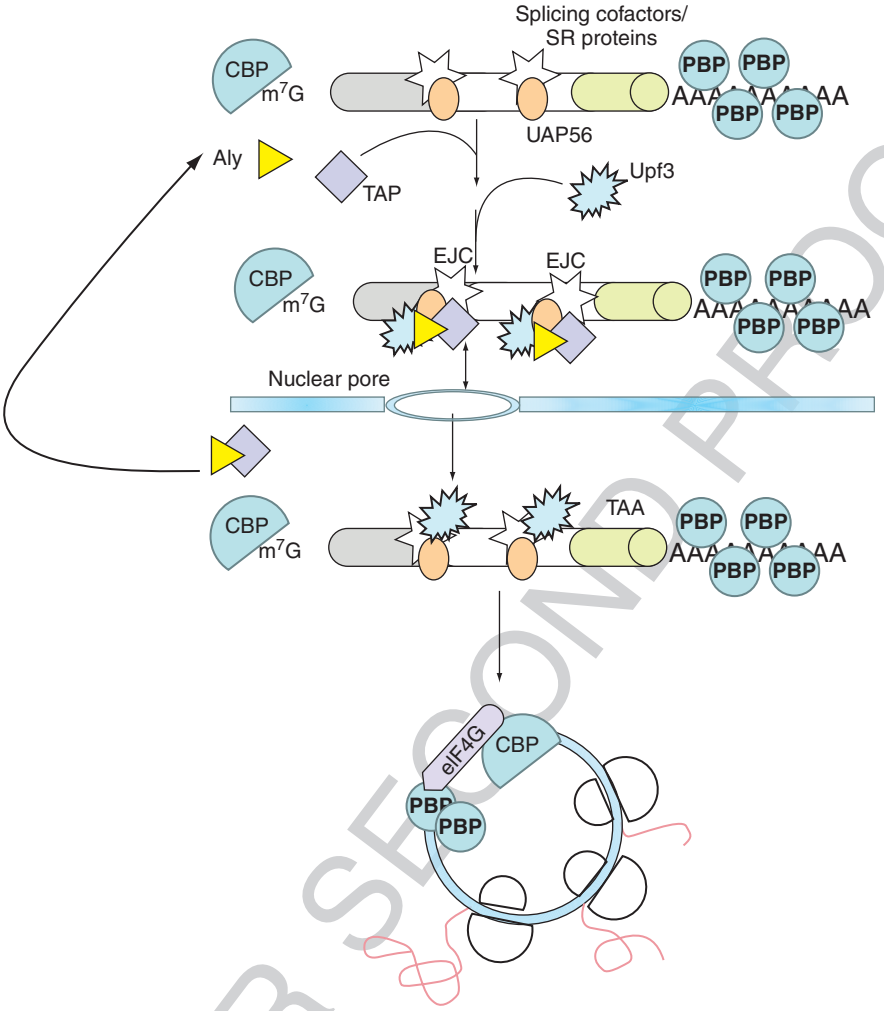
s0080 2.04.4 PostTranscriptional Control of Gene Expression

p0150 Pre-mRNA processing is critical for mRNA stability. The 5′-cap containing the 7-methylguanosine in the 5′–5′ linkage is recognized and bound by the cap-binding protein (CBP) and protects the 5′-end of the mRNA from exonuclease digestion. Similarly, the binding of PABP to the poly(A) tail protects the transcript from 3′-end exonuclease or exosome degradation (Houseley *et al.* 2006). Finally, splicing is a recorded event with the spliced transcript marked by proteins that form a complex at the mRNA exon–exon boundaries, termed the exon-junction complex (EJC) (Le Hir *et al.* 2000). EJC proteins play important roles in signaling nuclear export and mRNA stability in the cytoplasm (Le Hir *et al.* 2001; Reed and Hurt 2002). Thus, the processed mRNA is complexed with proteins at the 5′-end (CBP), the 3′-end (PABP), and EJCs forming an export competent mRNP particle (**Figure 6**). These nuclear processing events are linked to the cytoplasmic fate of mRNAs.

p0155 The spliceosome factor Aly is an important component of the EJC that serves as a bridging factor between splicing and nuclear export and mRNA surveillance, an mRNA quality control mechanism in the cytoplasm (**Figure 6**). Specifically, Aly recruits nuclear export (TAP) and mRNA surveillance factors (Upf3) to the EJC that are required for subsequent export and mRNA surveillance in the cytoplasm. Once the mRNP complex is in the cytoplasm, the nuclear export components are recycled back to the nucleus and additional mRNA surveillance components (e.g., Upf3) are recruited to the EJC.

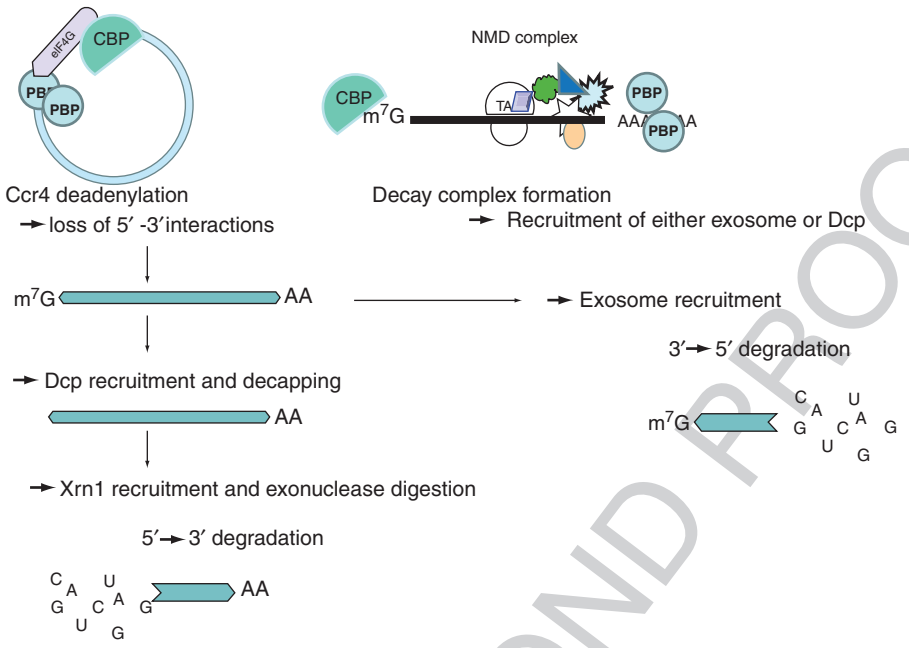
s0085 2.04.4.1 mRNA Stability

p0160 The final step in gene expression is translation of the mRNA in the cytoplasm. This chapter will focus only on the role of the initial round of translation as a quality control check point for mRNA (Chang *et al.* 2007; Doma and Parker 2007; Shyu *et al.* 2008). The current model for a stable, translation-competent mRNA depicts protein–protein interactions between the 5′- and 3′-ends generating a circular transcript (**Figure 6**). This model is dependent upon proper termination, which requires an initial round of translation for the ribosome to read through the mRNA transcript and remove all EJC. The consequence of the loss of all EJCs is reorganization of mRNP to promote protein–protein interactions between the 5′- and 3′-ends that stabilize the mRNA in a functionally active form. Each transcript has an inherent mRNA half-life with mRNA turnover mediated by two potential degradation pathways based on susceptibility of the mRNA ends. One pathway is that processive loss of the poly(A) tail by the action of the deadenylase Ccr4 disrupts the 5′–3′ protein–protein interactions and exposes the 5′-end to the decapping enzyme Dcp1. Loss of the cap renders the 5′-end susceptible to the Xrn1 exonuclease that degrades the mRNA in a 5′ to 3′ direction. Alternatively, the shortened poly(A) tail can serve as a site for recruitment of the exosome and degradation from the 3′- to 5′-end of the transcript.



f0030 **Figure 6** Nuclear–cytoplasmic connection with mRNA processing. The mRNA is 'marked' with nuclear proteins that are derived from nuclear pre-mRNA processing. 5′- and 3′-end processing is marked with cap-binding protein (CBP) and poly(A)-binding protein (PBP). Splicing is marked by the exon-junction complex (EJC) composed of protein from the spliceosome and recruited factors. UAP56 is a U2snRNP-associated factor that binds Aly. Aly recruits upstream processing factor 3 (Upf3) and nuclear export factor TAP. An initial round of translation removes all EJC and the stable, actively translated mRNA is depicted as a circle with 5′–3′ protein–protein interactions and translating ribosomes (open half circles).

p0165 The ability of the cell to recognize and degrade aberrant transcripts is necessary to ensure proper gene expression. Transcripts that contain nonsense mutations, processing errors, or extended 3′-UTR typically lead to improper translation termination and subsequent loss of structural rearrangement of the mRNA into a stable, translation-competent transcript. mRNA surveillance is a major mechanism used to target an aberrant mRNA containing a premature termination codon (PTC) for degradation by the nonsense-mediated decay (NMD) pathway (Chang *et al.* 2007; Doma and Parker 2007). The ribosome pauses at a termination codon and recruits elongation release factors (Erfs) that (1) promote translation termination and (2) form a surveillance complex to check for additional termination codons. As stated above, all EJCs will be removed when termination occurs at the proper stop codon. However, assembly of the surveillance complex at a PTC will result in interaction with the downstream EJC and trigger NMD (**Figure 7**). Genetic screens identified upstream processing factors (Upfs) as components required to trigger mRNA surveillance and NMD. Erf proteins of the ribosome surveillance



f0035 **Figure 7** mRNA degradation. The stable mRNA transcript is depicted as a closed circle with protein–protein interactions between the translation factor eIF4G that binds the cap-binding protein (CBP) and the poly(A)-binding protein(s) (PBP). The actions of the deadenylase enzyme (Ccr4), decapping enzyme (Dcp), and 5′-exonuclease (Xrn1) are indicated. Both the 5′- and 3′-ends are susceptible to degradation by Dcp → Xrn1 pathway and the exosome pathway, respectively. Nonsense-mediated decay (NMD) complex includes the EJC and Upf3 as shown in **Figure 6**. Upf1 (cloud shape) and Upf2 (triangle) are recruited to the EJC and interact with the elongation release factor (Erf, parallelogram) to form the decay complex. The decay complex signals for either Dcp or exosome recruitment for mRNA degradation.

complex and Upf3 work together to recruit Upf2 and Upf1 to form a decay complex at the PTC–EJC junction (**Figure 7**). Upf1 is an RNA-dependent, ATP-dependent helicase that is required to destabilize the mRNA structure leading to deadenylation, decapping and exonuclease degradation, and/or exosome degradation as outlined for 'normal' decay processes.

s0090 ### 2.04.4.2 Translation Repression by RNA Interference

p0170 RNA interference (RNAi) is a posttranscriptional gene silencing pathway that is triggered by small cytoplasmic double-stranded RNAs (dsRNAs) 20–30 nt in length. The cytoplasmic dsRNAs can be from exogenous origin (siRNA) or genome-encoded (microRNAs (miRNAs)). miRNAs serve important biological roles in development, metabolism, apoptosis, and differentiation processes. miRNAs are discussed in greater detail in Chapter 2.23 and therefore will only be highlighted herein for their function in translation repression and mRNA degradation.

p0175 In brief, miRNA-mediated gene silencing is dependent upon specific recognition of target mRNAs through complementary base pairing within the 3′-untranslated region that then triggers the target mRNA for either cleavage and degradation or translation inhibition (Filipowicz *et al.* 2008). Both nuclear and cytoplasmic steps are required for generation of miRNA. First, miRNAs are encoded in the genome and transcribed by an RNAPII-dependent mechanism that produces a longer primary miRNA transcript (pri-RNA) that is ~70 nt long and adopts a stem–loop structure. Since miRNAs are under the control of RNAPII transcription, their expression is controlled by developmental, cell-specific, and regulated mechanisms. Thus, the expression pattern for miRNAs is another potential regulatory point in terms of control of gene expression. In the nucleus, the pri-RNA is processed by the endonuclease, Drosha, a class II RNase III enzyme, to form a shorter pre-miRNA that still retains the stem–loop structure and is exported to the cytoplasm by Exportin. In the cytoplasm, the pre-mRNA is cleaved by Dicer, a class III RNase III enzyme, to produce an miRNA, a short

linear dsRNA. Dicer generates short 21–23 nt dsRNA containing 3′ overhangs of 2–3 nt in length and a 5′ phosphate group, characteristic of RNase III enzymes. The miRNA is incorporated into an RNA-induced silencing complex (RISC) that aligns the antisense miRNA with the target mRNA and associates with stress granules (or P bodies in yeast). For targets with 100% complementarity to miRNA, the target mRNA is cut by the action of argonaute 2, a protein of the RISC, and another RNase III enzyme (Peters and Meister 2007). The cleaved mRNA now has susceptible 5′- and 3′-ends for exonuclease or exosome digestion. Targets with <100% sequence complementarity to miRNA undergo translation arrest.

s0095

2.04.5 Conclusions

p0180 As stated in the introduction, proper control of gene expression is fundamental for correct spatio-temporal expression patterns of a protein. This chapter has reviewed major regulatory points at the transcriptional and posttranscriptional levels that must all function properly to ensure that the protein-encoding genes are accurately expressed. The relevance to toxicology research is that each point is susceptible to disruption by chemical insults and elucidating the molecular mechanisms of chemical disruption at the transcriptional and posttranscriptional levels is an active and important research focus for developmental and reproductive toxicology. The early work focused on xenobiotics that activate the AhR:ARNT signaling pathway(s) as the molecular mechanism responsible for inducing expression of the drug-metabolizing enzymes of the cytochrome P450 systems for clearance of the xenobiotic. Currently, there is great interest in understanding the impact of endocrine disruptors on disease. Endocrine disruptor is a generic term used to represent an environmental contaminant that promotes altered gene expression by (1) mimicking the action of endogenous steroid hormones and aberrantly activating or inhibiting the steroid hormone/nuclear receptor transcriptional activators, (2) promoting cross-talk between the AhR and the steroid hormone/nuclear receptor pathways, and (3) altering epigenetic DNA modifications. Potential gene targets of transcriptional disruption include mRNA processing factors and miRNA expression. Therefore, both direct and indirect mechanisms for altered gene expression need to be considered when evaluating the potential impact of an environmental contaminant on disrupting spatio-temporal expression patterns of proteins.

References

Aguilera, A. *Curr. Opin. Cell Biol.* **2005**, *17* (3), 242–250. b0005
Ahn, S. H.; Kim, M.; Buratowski, S. *Mol. Cell* **2004**, *13* (1), 67–76. b0010
Anway, M. D.; Cupp, A. S.; Uzumcu, M.; Skinner, M. K. *Science* **2005**, *308* (5727), 1466–1469. b0015
Aoyagi, S.; Archer, T. K. *Mol. Cell. Endocrinol.* **2008**, *280* (1–2), 1–5. b0020
Beck, S.; Rakyan, V. K. *Trends Genet.* **2008**, *24* (5), 231–237. b0025
Bednar, J.; Horowitz, R. A.; Grigoryev, S. A.; Carruthers, L. M.; Hansen, J. C.; Koster, A. J.; Woodcock, C. L. *Proc. Natl. Acad. Sci. USA* **1998**, *95* (24), 14173–14178. b0030
Bentley, D. L. *Curr. Opin. Cell Biol.* **2005**, *17* (3), 251–256. b0035
Berget, S. M. *J. Biol. Chem.* **1995**, *270* (6), 2411–2414. b0040
Birney, E.; Stamatoyannopoulos, J. A.; Dutta, A.; Guigo, R.; Gingeras, T. R.; Margulies, E. H.; Weng, Z.; Snyder, M.; Dermitzakis, E. T.; Thurman, R. E., *et al. Nature* **2007**, *447* (7146), 799–816. b0045
Bleichert, F.; Baserga, S. J. *Mol. Cell* **2007**, *27* (3), 339–352. b0050
Blencowe, B. J. *Cell* **2006**, *126* (1), 37–47. b0055
Bock, C.; Walter, J.; Paulsen, M.; Lengauer, T. *PLoS Comput. Biol.* **2007**, *3* (6), e110. b0060
Buratowski, S. *Nat. Struct. Biol.* **2003**, *10* (9), 679–680. b0065
Chang, H. S.; Anway, M. D.; Rekow, S. S.; Skinner, M. K. *Endocrinology* **2006**, *147* (12), 5524–5541. b0070
Chang, Y. F.; Imam, J. S.; Wilkinson, M. F. *Annu. Rev. Biochem.* **2007**, *76*, 51–74. b0075
Clouaire, T.; Stancheva, I. *Cell. Mol. Life Sci.* **2008**, *65* (10), 1509–1522. b0080
Cohen, D. E.; Lee, J. T. *Curr. Opin. Genet. Dev.* **2002**, *12* (2), 219–224. b0085
Cohen, Y.; Merhavi-Shoham, E.; Avraham, R. B.; Frenkel, S.; Pe'er, J.; Goldenberg-Cohen, N. *Exp. Eye Res.* **2008**, *86* (2), 201–206. b0090
Cui, X.; Wakai, T.; Shirai, Y.; Yokoyama, N.; Hatakeyama, K.; Hirano, S. *Hum. Pathol.* **2006**, *37* (3), 298–311. b0095
Doma, M. K.; Parker, R. *Cell* **2007**, *131* (4), 660–668. b0100
Downes, M.; Ordentlich, P.; Kao, H. Y.; Alvarez, J. G.; Evans, R. M. *Proc. Natl. Acad. Sci. USA* **2000**, *97* (19), 10330–10335. b0105
Esteller, M. *N. Engl. J. Med.* **2008**, *358* (11), 1148–1159. b0110
Filipowicz, W.; Bhattacharyya, S. N.; Sonenberg, N. *Nat. Rev. Genet.* **2008**, *9* (2), 102–114. b0115
Forneris, F.; Binda, C.; Battaglioli, E.; Mattevi, A. *Trends Biochem. Sci.* **2008**, *33* (4), 181–189. b0120
Gangaraju, V. K.; Bartholomew, B. *Mutat. Res.* **2007**, *618* (1–2), 3–17. b0125
Gilbert, N.; Boyle, S.; Fiegler, H.; Woodfine, K.; Carter, N. P.; Bickmore, W. A. *Cell* **2004**, *118* (5), 555–566. b9000
Gius, D.; Bradbury, C. M.; Sun, L.; Awwad, R. T.; Huang, L.; Smart, D. D.; Bisht, K. S.; Ho, A. S.; Nguyen, P. *Cancer* **2005**, *104* (9), 1789–1793. b0130
Hager, G. L.; Nagaich, A. K.; Johnson, T. A.; Walker, D. A.; John, S. *Biochim. Biophys. Acta* **2004**, *1677* (1–3), 46–51. b0135
Hagiwara, M.; Nojima, T. *J. Biochem. (Tokyo)* **2007**, *142* (1), 11–15. b0140
Hassan, A. H.; Awad, S.; Al-Natour, Z.; Othman, S.; Mustafa, F.; Rizvi, T. A. *Biochem. J.* **2007**, *402* (1), 125–133. b0145
Hertel, K. J.; Graveley, B. R. *Trends Biochem. Sci.* **2005**, *30* (3), 115–118. b0150
Hirose, Y.; Ohkuma, Y. *J. Biochem. (Tokyo)* **2007**, *141* (5), 601–608. b0155
Houseley, J.; LaCava, J.; Tollervey, D. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (7), 529–539. b0160

b9005 Jackson, D. A.; Iborra, F. J.; Manders, E. M. M.; Cook, P. R. *Mol. Biol. Cell* **1998**, *9* (6), 1523–1536.

b0165 Kanno, T.; Kanno, Y.; Siegel, R. M.; Jang, M. K.; Lenardo, M. J.; Ozato, K. *Mol. Cell* **2004**, *13* (1), 33–43.

b0170 Kao, H. Y.; Downes, M.; Ordentlich, P.; Evans, R. M. *Genes Dev.* **2000**, *14* (1), 55–66.

b0175 Kim, H. J.; Jeong, S. H.; Heo, J. H.; Jeong, S. J.; Kim, S. T.; Youn, H. D.; Han, J. W.; Lee, H. W.; Cho, E. J. *Mol. Cell Biol.* **2004**, *24* (14), 6184–6193.

b0180 Kimura, A.; Matsubara, K.; Horikoshi, M. *J. Biochem. (Tokyo)* **2005**, *138* (6), 647–662.

b0185 Komarnitsky, P.; Cho, E. J.; Buratowski, S. *Genes Dev.* **2000**, *14* (19), 2452–2460.

b0190 Kornblihtt, A. R.; de la Mata, M.; Fededa, J. P.; Munoz, M. J.; Nogues, G. *RNA* **2004**, *10* (10), 1489–1498.

b0195 Korzus, E.; Torchia, J.; Rose, D. W.; Xu, L.; Kurokawa, R.; Mclnerney, E. M.; Mullen, T.-M.; Glass, C. K.; Rosenfeld, M. G. *Science* **1998**, *279* (5351), 703–707.

b0200 Kumaran, R. I.; Thakar, R.; Spector, D. L. *Cell* **2008**, *132* (6), 929–934.

b0205 Latham, J. A.; Dent, S. Y. R. *Nat. Struct. Mol. Biol.* **2007**, *14* (11), 1017–1024.

b0210 Le Hir, H.; Gatfield, D.; Izaurralde, E.; Moore, M. J. *Embo. J.* **2001**, *20* (17), 4987–4997.

b0215 Le Hir, H.; Izaurralde, E.; Maquat, L. E.; Moore, M. J. *Embo. J.* **2000**, *19* (24), 6860–6869.

b0220 Lima, C. D. *Nat. Struct. Mol. Biol.* **2005**, *12* (2), 102–103.

b0225 Lonard, D. M.; O'Malley, B. W. *Mol. Cell* **2007**, *27* (5), 691–700.

b0230 Madhani, H. D.; Guthrie, C. *Cell* **1992**, *71* (5), 803–817.

b0235 Martens, J. A.; Winston, F. *Curr. Opin. Genet. Dev.* **2003**, *13* (2), 136–142.

b0240 McCracken, S.; Fong, N.; Rosonina, E.; Yankulov, K.; Brothers, G.; Siderovski, D.; Hessel, A.; Foster, S.; Shuman, S.; Bentley, D. L. *Genes Dev.* **1997a**, *11* (24), 3306–3318.

b0245 McCracken, S.; Fong, N.; Yankulov, K.; Ballantyne, S.; Pan, G.; Greenblatt, J.; Patterson, S. D.; Wickens, M.; Bentley, D. L. *Nature* **1997b**, *385* (6614), 357–361.

b0250 Mellor, J. *Trends Genet.* **2006**, *22* (6), 320–329.

b0255 Metivier, R.; Gallais, R.; Tiffoche, C.; Le Peron, C.; Jurkowska, R. Z.; Carmouche, R. P.; Ibberson, D.; Barath, P.; Demay, F.; Reid, G., *et al. Nature* **2008**, *452* (7183), 45–50.

b0260 Metivier, R.; Penot, G.; Hubner, M. R.; Reid, G.; Brand, H.; Kos, M.; Gannon, F. *Cell* **2003**, *115* (6), 751–763.

b0265 Millhouse, S.; Manley, J. L. *Mol. Cell. Biol.* **2005**, *25* (2), 533–544.

b0270 Nunez, E.; Kwon, Y. S.; Hutt, K. R.; Hu, Q.; Cardamone, M. D.; Ohgi, K. A.; Garcia-Bassets, I.; Rose, D. W.; Glass, C. K.; Rosenfeld, M. G., *et al. Cell* **2008**, *132* (6), 996–1010.

b0275 Osborne, C. S.; Chakalova, L.; Brown, K. E.; Carter, D.; Horton, A.; Debrand, E.; Goyenechea, B.; Mitchell, J. A.; Lopes, S.; Reik, W., *et al. Nat. Genet.* **2004**, *36* (10), 1065–1071.

Osborne, C. S.; Chakalova, L.; Mitchell, J. A.; Horton, A.; Wood, A. L.; Bolland, D. J.; Corcoran, A. E.; Fraser, P. *PLoS Biol.* **2007**, *5* (8), e192. b0280

Pennisi, E. *Science* **2005**, *309* (5731), 80. b0285

Peters, L.; Meister, G. *Mol. Cell* **2007**, *26* (5), 611–623. b0290

Plevin, M. J.; Mills, M. M.; Ikura, M. *Trends Biochem. Sci.* **2005**, *30* (2), 66–69. b0295

Proudfoot, N. J.; Furger, A.; Dye, M. J. *Cell* **2002**, *108* (4), 501–512. b0300

Reed, R.; Hurt, E. *Cell* **2002**, *108* (4), 523–531. b0305

Rocak, S.; Linder, P. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (3), 232–241. b0310

Rosenfeld, M. G.; Lunyak, V. V.; Glass, C. K. *Genes Dev.* **2006**, *20* (11), 1405–1428. b0315

Sanford, J. R.; Ellis, J.; Caceres, J. F. *Biochem. Soc. Trans.* **2005**, *33* (Pt. 3), 443–446. b0320

Savkur, R. S.; Burris, T. P. *J. Pept. Res.* **2004**, *63* (3), 207–212. b0325

Schneider, R.; Grosschedl, R. *Genes Dev.* **2007**, *21* (23), 3027–3043. b0330

Schones, D. E.; Zhao, K. *Nat. Rev. Genet.* **2008**, *9* (3), 179–191. b0335

Shahbazian, M. D.; Grunstein, M. *Annu. Rev. Biochem.* **2007**, *76*, 75–100. b0340

Sharp, P. A. *Trends Biochem. Sci.* **2005**, *30* (6), 279–281. b0345

Shen, L.; Kondo, Y.; Guo, Y.; Zhang, J.; Zhang, L.; Ahmed, S.; Shu, J.; Chen, X.; Waterland, R. A.; Issa, J. P. *PLoS Genet.* **2007**, *3* (10), 2023–2036. b0350

Shyu, A. B.; Wilkinson, M. F.; van Hoof, A. *EMBO J.* **2008**, *27* (3), 471–481. b0355

Spector, D. L. *Annu. Rev. Biochem.* **2003**, *72*, 573–608. b9010

Stoilov, P.; Meshorer, E.; Gencheva, M.; Glick, D.; Soreq, H.; Stamm, S. *DNA Cell Biol.* **2002**, *21* (11), 803–818. b0360

Thomas, M. C.; Chiang, C. M. *Crit. Rev. Biochem. Mol. Biol.* **2006**, *41* (3), 105–178. b0365

Tremethick, D. J. *Cell* **2007**, *128* (4), 651–654. b0370

Turner, B. M. *Nat. Cell Biol.* **2007**, *9* (1), 2–6. b0375

Uzumcu, M.; Suzuki, H.; Skinner, M. K. *Reprod. Toxicol.* **2004**, *18* (6), 765–774. b0380

Valadkhan, S. *Biol. Chem.* **2007**, *388* (7), 693–697. b0385

Waalkes, M. P.; Liu, J.; Chen, H.; Xie, Y.; Achanzar, W. E.; Zhou, Y. S.; Cheng, M. L.; Diwan, B. A. *J. Natl. Cancer Inst.* **2004**, *96* (6), 466–474. b0390

Wang, Z.; Burge, C. B. *RNA* **2008**, *14* (5), 802–813. b0395

Wysocka, J.; Allis, C. D.; Coonrod, S. *Front. Biosci.* **2006**, *11*, 344–355. b0400

Xu, L.; Glass, C. K.; Rosenfeld, M. G. *Curr. Opin. Genet. Dev.* **1999**, *9* (2), 140–147. b0405

Xu, W.; Kasper, L. H.; Lerach, S.; Jeevan, T.; Brindle, P. K. *Embo. J.* **2007**, *26* (12), 2890–2903. b0410

Zhang, X.; Yazaki, J.; Sundaresan, A.; Cokus, S.; Chan, S. W.; Chen, H.; Henderson, I. R.; Shinn, P.; Pellegrini, M.; Jacobsen, S. E., *et al. Cell* **2006**, *126* (6), 1189–1201. b0415

Author's Contact Information

Dr. Barbara J. Clark
Department of Biochemistry and Molecular Biology
University of Louisville
HSC A Room 502, 319 Abraham Flexner Way
Louisville, KY 40292
USA
e-mail: bjclark@louisville.edu

Keywords: epigenetic DNA modification; gene regulation; RNA processing; transcription

Abstract

This chapter will review the current models that guide our understanding of transcriptional and posttranscriptional mechanisms that control gene expression. Gene transcription is regulated at multiple levels and this chapter will focus on recruitment of transcription factors and coactivator proteins to the promoter, chromatin structure and remodeling, and assembly of the general transcriptional machinery and RNA polymerase II on the promoter for transcription initiation. Posttranscriptional mechanisms for primary RNA processing and mRNA stability will also be highlighted. RNA processing requires proper splicing of a primary transcript and modification of the 5′- and 3′-ends to generate a mature mRNA and the focus will be on the interdependence of these RNA-processing events with ongoing transcription. To ensure that the protein-encoding genes are accurately expressed, all regulatory processes must function properly and disruption at any point will alter gene expression patterns. For example, it is established that exposure to select environmental contaminants can impact the function of transcriptional activators and/or alter epigenetic DNA modifications, which subsequently result in aberrant gene expression. Therefore, understanding the fundamental molecular mechanisms of transcriptional and posttranscriptional regulation of gene expression is important for evaluating the potential impact of an environmental contaminant on disrupting spatio-temporal gene expression patterns.