

2

Multiple Transcript Initiation as a Mechanism for Regulating Gene Expression

Robert E. Farrell, Jr. and Carole L. Bassett

2.1. Nuclear Gene Transcription – An Overview

Transcription is the intermediary process that copies a DNA-encoded gene into a form which is either functional in its own right (stable RNAs, such as ribosomal or transfer RNAs) or can be decoded by the translational machinery into a functional protein. Transcripts destined for translation are called messenger RNAs (mRNAs), since they act as go-betweens from DNA to protein. Although RNAs are transcribed as single-stranded molecules, most can assume complicated secondary and tertiary structures that are critical for proper functioning. As a result, each mRNA contains not only the sequence information required to synthesize a protein, but also structural components that can regulate mRNA localization, stability, and translation efficiency. Thus the initiation of transcription occupies a preeminent place in the regulation of gene expression.

Since transcription is compartmentalized within a membrane-bound nucleus in eukaryotic cells, it is both spatially and temporally separated from translation. In both plant and animal cells, the stability of a typical mRNA transcript is significantly greater than that of a prokaryotic transcript which is often being both translated and degraded simultaneously. The enhanced stability of eukaryotic transcripts has facilitated their isolation and manipulation using molecular techniques.

RNA is produced from different chromosomal loci at different rates. It is obvious that transcript abundance reflects differences in the rate of RNA synthesis, stability of the transcript, and/or differences in the rate of RNA degradation. In this chapter we shall emphasize aspects of RNA synthesis with a focus on transcription initiation, particularly as it relates to multiple transcript start sites (TSS). mRNA transport and turnover are discussed in detail in Chapter 6.

The first level of eukaryotic nuclear gene regulation lies in the structure and organization of the chromosomal DNA itself. For most of the first half of the 20th century it was generally accepted that the interphase nucleus was unorganized, with the DNA and attendant proteins relatively uniformly

distributed. The discovery that DNA was “packaged” into discrete structures termed nucleosomes was the first step towards understanding the role of DNA in regulating transcription. Nucleosomes consist of an association of DNA wrapped around a group of eight histone proteins which, when viewed under an electron microscope, resemble “beads on a string”. Prior to the interaction between chromatin and early initiators of transcription, this three-dimensional architecture prevents RNA polymerase and associated transcription factors from interacting with the promoter, a stretch of sequence upstream of the coding region of an affiliated gene that determines its expression. Access to a given gene is provided by local unwinding which was thought to occur via an unwinding enzyme. However, a recent report documents an unexpected aspect of the unwinding of DNA prior to transcription. Ju *et al.* (2006) demonstrate that both strands of DNA are broken prior to active transcription and that specific repair enzymes are recruited to reverse the damage after transcription is initiated. It appears that transcription initiation is a much more “violent” process than was originally thought.

Methylation of the cytosine in CpG doublets is yet another way in which chromosomal structure affects gene expression, since methylated promoter sequences very often interfere with transcription factor binding (as in transcriptional silencing). Both plants (Antequera and Bird, 1988) and animals (Lander *et al.*, 2001; Venter *et al.*, 2001) possess CG islands consisting of multiple repetitions of the CpG dinucleotide which are found near or within promoters. The majority of CG clusters are methylated and widespread (Pradhan and Adams, 1995). Higher plants methylate CpN with a distinct preference for CpG. In addition plants can also methylate cytosines in CNG, with a preference for CAG and CTG triplets. For example, two different methyltransferases have been isolated from *Pisum sativum*, one of which is specific for CpG and the other of which is specific for CAG and CTG (Pradhan and Adams, 1995). Methylation has also been linked to posttranscriptional gene silencing in plants (discussed in Chapter 5).

Another level of organization affecting gene expression is the subnuclear compartmentation of genes, transcripts, and mRNA binding proteins (see Misteli, 2000). It is well known that the nuclear periphery contains large amounts of heterochromatin that is for the most part transcriptionally inactive. For example, the yeast mating-type locus is silenced by segregation to the nuclear periphery where Sir (silent information regulator) proteins mediate silencing of the locus (Cockell and Gasser, 1999). Another familiar subnuclear compartment is that of the nucleolus, the active site of rRNA transcription. Other nuclear compartments include Cajal bodies frequently associated with transcriptionally active histone loci (Schul *et al.*, 1999) and “speckles” (Spector, 1993) formed by splice factor complexes.

Coding regions of the DNA contain genes which are comprised of exons (regions that encode proteins or parts of proteins) and introns (intervening sequences that are generally noncoding). Introns are distributed throughout plant and animal genes and are variable in number, length, and nucleotide content. A salient feature of introns is the presence of highly conserved di-

nucleotide pairs that define the intron boundaries: introns typically begin with GU and end with AG. By recognizing these dinucleotides and adjacent or more distant motifs, the spliceosomal machinery excises introns from heterogeneous nuclear (hn) RNA transcripts during transcription in a very precise manner. It is noteworthy that several human disease states are a direct result of splicing errors (Cooper and Mattox, 1997). Transgenic expression of animal genes in plants suggests that plant splicing machinery recognizes somewhat different motifs, since animal genes are frequently cut into nonfunctional fragments because of this. It is well known that reporter genes, like green fluorescent protein derived from animal sources, have to be modified to remove bases recognized as splicing features by the typical plant spliceosome (Haseloff et al., 1997). A small number of introns contain open reading frames (ORFs, *i.e.*, encoded peptides or proteins), and, in a few cases, promoters and other regulatory sequences have been found in introns, especially those at the 5'-most region of a gene (discussed in Section 2.2.2.). Alternative processing of introns, including retention of introns within the transcript, contributes to differential gene expression by creating multiple proteins from a single transcript (discussed in detail in Chapter 3).

Genes in plant cells, as in animal cells, are transcribed by enzymes known as DNA-dependent RNA polymerases. In eukaryotic organisms RNA polymerase II (RNA pol II) is the enzyme that transcribes mRNAs. RNA pol II in both plants and animals is a very large enzyme consisting of a multi-subunit core with numerous transient interacting polypeptides, including transcription factors and their accessory proteins; it is the interaction of RNA pol II with other proteins and factors that modulates the activity of genes, often conferring inducibility/suppression in response to external and internal signals, as well as tissue and temporal specificity. In mammals and plants, a new RNA polymerase (RNA polymerase IV) has recently been described (Onodera *et al.*, 2005; Kanno *et al.*, 2005; Kravchenko *et al.*, 2005). In mammals this polypeptide, single-polypeptide nuclear RNA polymerase IV (spRNAP-IV), is derived from a nuclear-encoded mitochondrial RNA polymerase by alternative splicing which removes the mitochondrial transit sequence. As a result, spRNAP-IV is relocated to the nucleus where it regulates a subset of nuclear-encoded genes. The plant RNA pol IV appears to be quite different from spRNAP-IV both in sequence and function. It is associated with the DNA methylation aspects of gene silencing (Onodera *et al.*, 2005; Kanno *et al.*, 2005; see also Chapter 5), and one of the subunits, DDR3, is unique to plants (Kanno *et al.*, 2005).

2.1.1. *Initiation of Transcription: Transcription Factors and Promoter Elements*

Initiation of transcription requires at minimum a promoter and one or more transcription factors to direct RNA pol II to the appropriate gene. Promoters are usually arranged in a *cis* configuration upstream of the translation start codon (Fig. 2.1). All of the essential regulatory components of the promoter

AAGGTATCCAACCAGATTCCGGTACTACTCTCTCCCGTAATTAGATTCAAAGATTGAGGCAGGATTTCGATCC
 TTTTTCGCGCGCGCGCCCGGAATAAAATCACTACCATGAGTCGATCGAACGTTACGGGCACCCATTTCGAG
 GGCCTATTTTCAGCCCGCCAGCGCAGGGAATATCCTGACGTGTGATCATGCCATAATGCATCCATTATAA
 AAGAGGAGCCGATTAGTAAGTACCGACTACTGTCCGATACCATGAAGTGATCAATCGATACGGGTACCAT
 CGGACTATATAAAGCTACCCGTATTAGCATAATTGACCGGTACCTCAGTCCTTCTAGGCTATGATTCTAAC
 CCTAGAACTACATGACGTACCTACCGGTACGGAATTCATTCCGGTACGAATTGGAGCATGGACTTACTTGG

FIGURE 2.1. A typical promoter region of a eukaryotic gene is illustrated. A CpG island is underlined. The TATA element is outlined by a solid box and an upstream transcription factor binding site element (potential myb transcription factor binding site) is outlined by a solid oval. The initiator element is boxed by a dotted line, and the major transcription start site 37 nts from the center of the TATA element is indicated by a right-facing solid arrow above the 5'-most nucleotide. Minor TSSs are indicated by right-facing dotted arrows. Upstream AUGs are overlined and the translation start site of the encoded polypeptide is shown in gray type. Note that the upstream AUGs are not in "good" context (explained in Section 2.2.4.1.).

are typically located within about 300 bp 5' from the TSS. The core promoter is the shortest DNA template sequence that can direct transcription. The primary *cis* element for RNA pol II recruitment is the TATA box which consists of the core sequence, TATAA, and is centered around -30 to -60 bases from the TSS. This element can bind both RNA polymerases II and III to initiate transcription (Mattaj *et al.*, 1988; Lobo and Hernandez, 1989; Mitchell *et al.*, 1992). Other *cis*-elements upstream of the TATA box influence the directionality of transcription, and it is thought that they determine which polymerase dominates by preferentially recruiting one type over the other (Huang *et al.*, 1996). Additional elements up- and downstream of the TATA box modulate activity in a sequence-specific manner by binding transcription factors with their associated accessory factors in response to environmental and cellular cues.

TATA boxes appear to be required to initiate activated mRNA transcription from simple promoters in plants; however, in more complex promoters, the role of TATA elements may be more supportive (Pan *et al.*, 2000). Similarly, the TATA box is not an absolute requirement for transcription initiation in eukaryotes, as a variety of genes utilize TATA-less promoters (Burke and Kadonaga, 1996; Lagrange *et al.*, 1998). An additional motif, Py₂CAPy₅, known as an initiator sequence (Inr) is able to promote transcription autonomously or in conjunction with a TATA box or other element (Smale and Baltimore, 1989). Although some plant promoters have Inr-like sequences that appear to function with the TATA element, most plant promoters do not have consensus Inr sequences. The precise geometry of these elements in conjunction with the tertiary and quaternary structure of the pol II complex plus accessory factors will determine the transcription start site.

The position of the TSS determines the sequence and length of the 5' leader, also known as the 5' untranslated region (5' UTR). Features in the 5' leader sequence of the mRNA preceding the coding region are known to affect the efficiency of translation (Pain, 1996). This is particularly true of secondary structures, like hairpins and loops, that can positively or negatively affect translation efficiency depending on their location relative to the translation start codon. These features are discussed in more detail in Section 2.2.4. Despite these differences among individual mRNAs, there are a few generalizations that can be made regarding plant leader sequences. For example, dicot leader sequences tend to be more AU-rich than monocot leaders (Joshi, 1987). For most plant mRNAs the average length of the 5' leader is around 70 bases, although there are some exceptions (see Section 2.2.4.1., below).

2.1.2. *Transcription of Cytoplasmic Genomes*

The genomes of mitochondria and chloroplasts are covalently closed circles of double-stranded DNA and are typically present in multiple copies. There is considerable variation in the size of these organellar genomes. In plants, mtDNA can be as large as 570 kbp, which is substantially larger than mammalian mtDNA (about 17 kbp).

Human mtDNA contains 37 genes which are encoded on both strands, with some of the genes actually overlapping completely. In mammals there are no introns found in mtDNA (Anderson *et al.*, 1981; Lewin, 2004), though introns are observed in the mtDNA of fungi and plants (reviewed in Saldanha *et al.*, 1993). In plants and lower eukaryotes, mitochondria and chloroplast structural RNAs and protein-encoding genes are associated with introns referred to as Group I and II introns. Group I introns can also be found in a few nuclear genes, but they differ from nuclear-encoded introns primarily by the fact that they are self-splicing and require no spliceosomal machinery for processing. On the other hand, Group II introns which appear to be associated exclusively with mtDNA (plants and fungi) and cpDNA (plants) are spliced *in vivo* in a manner very similar to that observed with nuclear introns.

2.1.3. *Organellar vs Cytoplasmic mRNAs*

The mechanics of transcription and translation within the mitochondrion and chloroplast are highly conserved. Extranuclear genes residing in the mtDNA or the cpDNA are transcribed, and later translated, in that same organelle. When transcribed, mammalian mitochondrial genomes produce a single large transcript; the smaller individual RNAs are then liberated by the extensive posttranscriptional processing that occurs in the organelle. Similarly, cpRNAs have also been found to be polycistronic, likely related to their postulated blue-green algal ancestry.

Mitochondrial mRNAs characteristically lack the 5' cap that is associated with virtually all mRNAs in the eukaryotic cytoplasm, have small or non-existent leader sequences, and exhibit a start codon close the 5' end. Mitochondrial mRNAs have a poly(A) tail, but it is usually much shorter compared to the formidable poly(A) tail commonly found on cytoplasmic messages. In mammals, the addition of the poly(A) tail creates the stop codon for translation termination of the majority of mitochondrial mRNAs (Ojala *et al.*, 1981). Chloroplast mRNAs also lack a 5' cap. Although chloroplast mRNAs are also known to possess poly(A) tails, the effect of polyadenylation on these organellar transcripts is destabilizing (Lisitsky *et al.*, 1997; Gagliardi and Leaver, 1999; reviewed by Hayes *et al.*, 1999; Schuster *et al.*, 1999), rather than stabilizing, as is seen with cytoplasmic polyadenylated mRNAs (Manley and Proudfoot, 1994).

2.2. The Origins of Multiple Transcripts

A recent analysis of alternative transcript initiation using a bioinformatic approach to compare six different species was reported (Nagakasi *et al.*, 2005). Using the National Center for Biotechnology Information (NCBI) UniGene set for *Arabidopsis*, these authors found a relatively small number (435) of loci associated with alternative transcript initiation. This number is probably a significant underestimate given the fact that full length cDNAs do not necessarily represent the primary (capped) transcript and considering the fact that some genes in the *Arabidopsis* genome may be incorrectly annotated (see, for example, Meyers *et al.*, 2002; also reviewed in Pennisi, 1999 and Mathé *et al.*, 2002). However, the study does support the experimental evidence that alternative transcript initiation is one form of gene regulation that has been evolutionarily conserved, since it is seen in both animals and plants.

2.2.1. Multiple Promoters

In vertebrate mitochondrial DNAs, a single promoter governs transcription from each strand, and multiple products arise as a consequence of extensive processing and stability differences (reviewed in Tracy and Stern, 1995; Holec *et al.*, 2006). In contrast, in plants and fungi multiple transcription initiation is commonly observed in mitochondria (also reviewed in Tracy and Stern, 1995; Binder *et al.*, 1996), and has recently been verified for several different plant species (Lupold *et al.*, 1999; Kühn and Binder, 2002; Kühn *et al.*, 2005). Similarly, plant chloroplast-encoded genes with multiple promoters are not uncommon (Nakazono *et al.*, 1995; Vera *et al.*, 1996).

Nuclear genes with multiple promoters have been reported in animals (Weitzel *et al.*, 2000; Weitzel *et al.*, 2001). For example, several studies have reported multiple promoters for a nuclear-encoded rat mitochondrial glycerol-

3-phosphate dehydrogenase (G3PD). The presence of three different first exons associated with G3PD transcripts suggested the possibility of different promoters, and subsequent studies indicating a high level of expression in certain tissues and a strong response to thyroid hormone supported the prospect of different regulatory mechanisms (Gong *et al.*, 1998). A more detailed analysis of the expression of G3PD indicated that it was indeed governed by three promoters (Weitzel *et al.*, 2000), one of which was responsive to thyroid and steroid hormones, and the other two of which being tissue-specific did not respond to these hormones (Weitzel *et al.*, 2001).

A recent, interesting example of how different levels of transcript regulation act together to modulate gene expression is illustrated by the mouse cationic amino acid transporter 2 (*mCAT2*) locus. Two RNAs are transcribed from the locus, the smaller of which (*mCAT2* mRNA) is transported to the cytoplasm and the larger (8 kb) of which (*cat2* transcribed nuclear RNA or *CTN-RNA*) is retained in the nucleus. Each RNA is transcribed from a different promoter. Both RNAs are polyadenylated, but the *CTN-RNA* is edited (adenosine-to-inosine) in the 3' UTR which contributes to its retention in the nucleus. *mCAT2* mRNA is alternatively spliced in a tissue-specific manner, giving rise to a high-affinity protein (in stomach, brain and muscle) and a low-affinity protein translated in the cytoplasm of mouse liver cells (Closs *et al.*, 1993). However, under stress, the *CTN-RNA* transcript is cleaved at its 3' UTR, producing an additional *mCAT2* mRNA with a unique 5' UTR which is also translated in the cytoplasm of liver cells (Prasanth *et al.*, 2005). In this manner, production of *CTN-RNA* acts as a reservoir for the transporter protein in liver cells, so that under stress conditions, a relatively large supply of *mCAT2* RNA is rapidly transported into the cytoplasm and made available for translation.

Actual examples of multiple promoters regulating expression of a single nuclear gene in plants have not been described as yet. An early report on the regulation of the nuclear-encoded chloroplast ribosomal protein, L21, suggested alternative promoters that were differentially regulated in roots and leaves (Lagrange *et al.*, 1993). However, analysis of transcript initiation indicates that the two transcripts arise from a single apparently TATA-less promoter and are initiated only 43 nt apart from each other. This is consistent with numerous reports of single promoters with multiple TSSs, rather than multiple promoters *per se*.

2.2.2. Transcription Start Sites in Introns

An interesting example of an intron containing a promoter in animals is the case of the chicken type III collagen gene. Two transcripts are detected from this gene, one that initiates upstream of the primary ATG and one that initiates inside of intron 23 (Zhang *et al.*, 1997). The upstream-initiating transcript encodes the canonical collagen protein, whereas the intron-initiating transcript encodes a protein of unknown function (Cohen *et al.*, 2002). Each

transcript and protein are expressed in a tissue-specific manner. However, the biological significance of these observations is not yet clear.

Alternative promoters in the introns of plant nuclear genes are almost exclusively associated with the first or second intron. In such cases the resulting mRNA lacks the first exon or may encode a truncated protein (Sheen, 1991; Tamaoki *et al.*, 1995; Bai *et al.*, 1999). In the case of the maize pyruvate Pi-dikinase (PPDK) gene, two forms of the enzyme were known to exist. One form is associated with the cyclic C₄ photosynthetic pathway and is located in mesophyll chloroplasts. The second form is found in plants that use the standard C₃ type of photosynthesis and is associated with the cytoplasmic reactions comprising the anapleurotic pathway to regenerate phosphoenolpyruvate. Both enzymes are derived from a single gene (Sheen, 1991). The C₄ form of the enzyme is encoded by a transcript that initiates upstream of the first exon. Removal of intron 1 during processing results in a transcript that, when translated, yields a PPDK isoform with a transit peptide directing it to the chloroplast. In contrast, a cytoplasmic version of PPDK originates from a transcript which initiates *via* a promoter located within the 3' end of intron 1. The resulting product is located in the cytoplasm. In addition to governing changes in intracellular location, the two PPDK promoters also contribute to the observed differences in abundance reported for PPDK in C₄ vs. C₃ plants.

2.2.3. *Multiple TATA Boxes in a Single Promoter*

Improved methods have allowed us to identify more precisely the actual TSSs for a large number of genes. As mentioned previously, in many cases there is a primary TSS and several nearby secondary TSSs responding to a single TATA box (Wray *et al.*, 2003). Although reports of multiple TATA boxes within the same promoter are comparatively rare, a number of examples of multiple TATA boxes with multiple TSSs have been described in animals (for example, see Fra *et al.*, 2000). One recent example is the case of the prolactin-releasing peptide (PrRP) gene which was shown to produce transcripts in rat from three different TSSs (Yamada *et al.*, 2001) in conjunction with multiple TATA boxes within the same promoter region.

Examples of multiple TATA boxes initiating transcripts at different TSSs are not as common in plants; however, some interesting examples have been recently reported. RNA ligase-mediated 5'-rapid amplification of cDNA ends (RLM-5' RACE) was used to demonstrate that a leucine-rich repeat receptor kinase gene (*inrpk1*) in morning glory (*Ipomoea [Pharbitis] nil*) produced transcripts from at least nine different TSSs using three different consensus and non-consensus TATA boxes (Bassett *et al.*, 2004). One of the more remarkable features of this promoter is the observation that one of the TATA boxes is selected in a tissue-specific manner. Similarly, the *PhyA* genes from both *Arabidopsis* and tobacco have been shown to initiate transcription at different TATA elements in the promoter (Dehesh *et al.*, 1994; Adams *et al.*, 1995).

In tomato, two different TATA motifs direct the synthesis of long and short mRNAs from the phenylalanine ammonia-lyase gene (*PAL5*), in which the TATA elements appear to be differentially responsive to various stresses (Lee *et al.*, 1994). Also found in tomato is the Ca^{++} -ATPase gene (*LCAI*) promoter (Navarro-Avino *et al.*, 1999) in which one mRNA has a 5' leader sequence that is more than 900 bp longer than the shorter transcript. Each transcript appears to be derived from two different TATA boxes, although the functional significance of this observation is not clear at present. Recent analysis of *LCAI* response to auxin and abscisic acid indicates that *cis*-sequences corresponding to hormone response elements are located approximately 200 bases upstream of the distal TSS (Navarro-Avino and Bennett, 2005), suggesting that the significance of dual TATA boxes may lie in differences in the upstream *cis*-elements influencing them, *i.e.*, response to different hormones or environmental factors. Based on all the existing reports, however, this is a very unusual occurrence.

2.2.4. How Alternative TSSs Influence Gene Expression

2.2.4.1 Upstream AUGs in Different 5' Transcript Leader Sequences

Once a mature mRNA reaches the cytoplasm, it is very likely that it will be translated, but this is not guaranteed. Transcripts compete for a limited number of ribosomes, and it is clear that some mRNAs are translated with greater efficiency than others (for a review, see Bag, 1991). The prevailing hypothesis of the mechanics of initiation of eukaryotic translation is widely known as the ribosome scanning model (Kozak, 1978; 1991b). According to this model, the first event associated with the process of translation in eukaryotic cells is recognition of the 5' methylated cap (7-methylguanosine joined 5'→5' to the first transcribed nucleotide) that is associated with virtually all eukaryotic cytoplasmic mRNAs. The scanning model postulates that the smaller subunit of the eukaryotic ribosome, *i.e.*, the 40S subunit with attached initiation factors, binds to the mRNA proximal to the cap, and then migrates linearly downstream along the mRNA until the first AUG codon in "good context" is encountered. Migration in this manner is instrumental for melting any secondary structure associated with the transcript, especially around the 5' untranslated leader. Although the first AUG codon encountered will usually function as the initiation codon, simply being the first AUG codon after the 5' end of the transcript does not guarantee that it will serve as the initiation codon. "Good context", which is the key to successful recognition of the appropriate translation start site, refers to the sequence of nucleotides surrounding the initiator AUG and is critical to the selection of the appropriate AUG for translation initiation. The "good context" consensus sequence associated with translation initiation in higher animals is GCCGCCPu-CCAUGG, and the importance of this sequence has been demonstrated by site-directed mutagenesis (Kozak, 1986; Kozak, 1987b; Kozak 1997).

Higher plants, in contrast show a slightly different context consensus sequence of caA(A/C)aAUGGCg; the sequences for dicots and monocots is aaA(A/C)aAUGGCu and c(a/c)Pu(A/C)cAUGGCG, respectively (Joshi *et al.*, 1997).

One of the most important developments in understanding the control of protein synthesis was the discovery of AUG codons in the 5' leader sequence of many mRNA species. These so-called upstream AUGs (uAUGs) along with often-associated up-stream open reading frames (uORFs) are regarded as *cis*-acting elements in the mRNA itself. While something of a rarity when considering transcripts as a whole, uAUGs are very common in regulatory genes like kinases and transcription factors, including several oncogenes (Kozak, 1987a; 1991a; Morris, 1995). When multiple AUG codons are contained within a 5' leader sequence, initiation at downstream AUG codons is favored when any of the following occur (Kozak, 1989): (1) the first AUG codon occurs fewer than 10 nt downstream from the 5' cap; (2) the first AUG codon is in an unfavorable context for initiation; (3) reinitiation occurs following translation of a minor protein associated with the upstream AUG, *i.e.*, inframe stop codons specify termination of translation. In instances when the first AUG is not selected as a functional translation initiation codon for any reason, the phenomenon is known as "leaky scanning".

Following translation of a uORF, it is possible for the ribosome to continue scanning the mRNA and reinitiate translation downstream at the next initiation codon in good context. While in some cases it does not appear that the uORF has a negative impact on the efficiency of translation downstream (Cao and Geballe, 1995; Mize *et al.*, 1998), there are other examples in which the peptide produced by uORF translation may cause the ribosome to stall, thereby preventing the loading of additional ribosomes onto the transcript *i.e.*, polysome formation. It has further been suggested that the uORF protein may play its role in the regulation of gene expression by destabilizing the transcript (Ruiz-Echevarria *et al.*, 1996; 2000). Other examples of gene regulation by uORFs include the mammalian gene *AdoMetDC*, a key enzyme in the biosynthesis of spermidine and spermine (Mize *et al.*, 1998) and, in *Neurospora*, the small subunit of carbamoyl phosphate synthetase, which is encoded by the *arg-2* gene. When the level of arginine is elevated in the growth medium, the uORF causes ribosomes to arrest on the leader sequence, while at low concentrations the uORF codon is ignored *via* leaky scanning, thereby facilitating translation of the major protein (Wang and Sachs, 1997).

Because upstream ORFs are most often associated with regulatory genes whose mRNAs are typically not very abundant, it is thought that the uORFs act as an additional "brake" on expression, since translation of the downstream ORF is usually reduced. A potential example of this in plants was reported for *inrpk1* from morning glory (Bassett *et al.*, 2004). As previously mentioned, multiple transcripts were observed to initiate from three TATA boxes. Transcripts from the first two TATA boxes were found predominantly

in aerial tissues, whereas transcripts originating from the 3'-most TATA box were found predominantly in roots. Interestingly, six uAUGs were identified in the longer transcripts; the two 5'-most uAUGs were in "good" Kozak context and potentially encoded a peptide of 73 amino acids. The root-specific transcript initiated further downstream and therefore lacked the two distal uAUGs. Because the root-specific transcripts contained the four proximal non-contextual uAUGs, the prediction would be that the protein would be rare in aerial tissues, but comparably abundant in roots. This is consistent with previous indications of greater RNA abundance in roots and root-specific processing of a cryptic intron from *inrpk1* (Bassett *et al.*, 2000).

2.2.4.2. Avoidance of Secondary Structures

As seen in the previous section, uAUGs can dramatically influence translation efficiency. Other features in the 5' untranslated leader can also have profound effects on translation, including the formation of secondary structures upstream of the start codon (Meijer and Thomas, 2002). In instances where an mRNA forms secondary structure involving the 5' leader sequence, ribosomal scanning of the mRNA can be so severely limited that translation may not be possible. This phenomenon has been demonstrated in both plant and animal models (Kozak, 1991b; Kumar and Dinesh-Miller, 1993; Futterer and Hohn, 1996; Pain, 1996; for review, see Kozak, 1999) and in some plant viral mRNAs (Ryabova *et al.*, 2000; Gowda *et al.*, 2003). The degree of translation inhibition is dependent upon the thermodynamic stability of the hairpin and its position. It is possible that translational control of this nature may well have evolved in order to prevent the over expression of a particular protein. While moderately-to-very stable hairpins upstream of an AUG codon can efficiently suppress translation, a weak (−19 kcal/mol) hairpin 14 nts downstream of an AUG codon can actually enhance the initiation of translation (Kozak, 1990). One can speculate that the downstream hairpin causes strategic pausing of the ribosomal complex directly over the initiation AUG, thereby enhancing translation initiation. Thus placement relative to an AUG codon appears to be a critical factor in determining how secondary structures affect translation initiation efficiency.

An example of translational repression by plant mRNA secondary structure is represented by the maize (*Zea mays*) *Lc* gene (Wang and Wessler, 2001), which encodes a transcriptional activator involved in the anthocyanin biosynthetic pathway. The 5' untranslated leader is unusually long and contains a uORF which itself restricts translation initiation at the usual AUG (Damiani and Wessler, 1993). However, an additional level of translational control is achieved by the location of a potential RNA hairpin 18 nt after the 5' end of the leader sequence of the *Lc* transcript. Disruption of the secondary structure through mutation and deletion of paired bases enhances translation of the uORF and the downstream gene both *in vitro* and *in vivo* (Wang and Wessler, 2001). Thus the putative hairpin operates to repress downstream

translation initiation and does so in a manner that is independent of the uORF's effect on translation. In fact the two structures appear to act in an additive manner on the main AUG codon. The proximity of the hairpin to the 5' end of the transcript suggests that the hairpin may impede the ability of ribosomes to load efficiently onto the mRNA. Despite numerous examples of secondary structure effects on translation efficiency, there are relatively few examples in the literature where multiple TSSs have been reported in which one transcript avoids a region of secondary structure found in the longer 5' untranslated leader, *e.g.*, Myers *et al.* (1998) and Myers *et al.* (2004), although this is probably a more common phenomenon than has been reported so far. No reports of a similar mechanism in plants with multiple TSSs have been published to date.

2.2.4.3. Changes in the Location of Gene Product

In plants multiple transcripts originating from a single promoter can give rise to polypeptides potentially located in different intracellular compartments. For example, Cheng *et al.*, (1994) mapped TSSs for *Atrbp33*, a gene encoding an RNA-binding protein (RBP) from *Arabidopsis*. The transcripts fell into two major groups corresponding to a 1.2 kb and a 1.0 kb transcript. The longer transcript is predicted to encode a typical chloroplast RBP containing a transit peptide, an acidic region and two RNA binding domains. The shorter transcript begins about 200 bases 3' of the longer transcript TSS and therefore lacks the first AUG. Translation initiation at the 3'-most AUG would yield an amino-terminally-truncated protein containing only the second RBP which would likely be cytoplasmic. A similar case was observed for the *Arabidopsis* threonyl-tRNA and valyl-tRNA synthetase genes (Souciet *et al.*, 1999). Two transcripts for each gene were identified, one of which initiated upstream of the first AUG and one of which initiated between the first and second AUGs. The long transcripts translated mitochondrial forms of the enzymes, while the shorter transcripts translated cytoplasmic versions. Interestingly, the mitochondrial forms were translated at reduced levels compared to the cytosolic forms because of differences in the contexts of the AUGs. In carrot the dihydrofolate reductase-thymidylate synthase gene promoter contains two consensus TATA elements that direct synthesis of two transcripts (Luo *et al.*, 1997). The longer transcript has an upstream AUG (in frame with the main AUG) which encodes a polypeptide with a transit sequence that directs the enzyme to the plastid.

More recently, a similar example was found involving the dehydrin 2 gene (*PpDhn2*) in peach (Wisniewski *et al.*, 2006). This gene initiates from a consensus TATA box approximately 100 bp upstream of the translation start site in bark tissue sampled in either July or December (Fig. 2.2). Two predominant TSSs were identified, one 28 bases downstream from the TATA box and the other 32 bases further downstream. However, in mature fruit, initiation occurs an additional 125 bp upstream of the conventional ATG from an

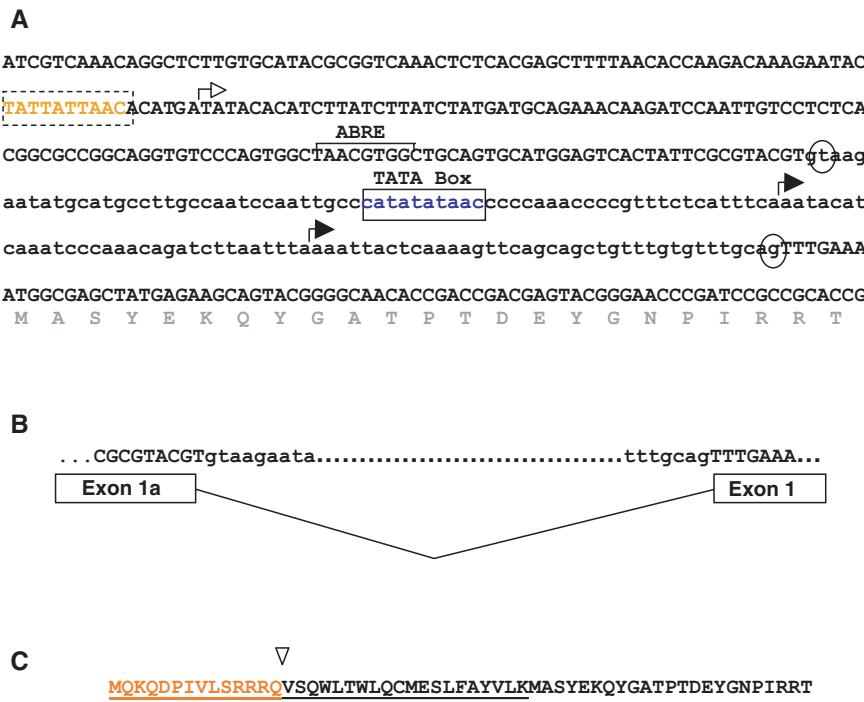


FIGURE 2.2. The promoter region of a peach dehydrin gene is illustrated. A. Sequence of the 490 bases upstream of the canonical translation start site is shown. The 5' leader intron is shown in lower case lettering, and the TATA element (in blue) that regulates expression in bark is outlined by a solid box. Two transcription start sites identified in bark tissues are indicated by the two right-facing black arrows. The dotted box outlines a possible non-consensus TATA element (in orange), and the TSS corresponding to fruit transcripts is shown by the right-facing open arrow. Intron junctions are enclosed in ovals. An abscisic acid response element (ABRE) is enclosed in horizontal brackets. B. Intron junction sequences and removal of the intron are illustrated. C. Putative polypeptide generated by the removal of the 5' leader intron. Additional amino acids at the N-terminus are underlined. The inverted arrow marks a potential cleavage site for a putative transit peptide (in orange) directing the protein into mitochondria. After Bassett *et al.*, in press. (See Color Plates)

unidentified element which could be a nearby nonconsensus TATA box (Bassett *et al.*, in press; Fig. 2.2A). The resulting fruit-specific transcript contains an upstream AUG in good translational context, followed by a typical intron, which interestingly contains the consensus TATA box used in bark tissues. Splicing out the intron places the upstream AUG in frame with the conventional translation start of the protein and would add 34 amino acids to the N-terminus (Fig. 2.2B and C). Although bioinformatics programs

predict that the new protein could be translocated to mitochondria, there is as yet no direct evidence to support this possibility.

2.3. Bicistronic mRNAs

2.3.1. *Monocistronic vs. Polycistronic mRNA*

Since much of the early information about transcription regulation came from studies of bacteria and viruses, it was first assumed that eukaryotic transcription would be similar, if not identical. In bacteria and archaea, genes performing different steps in the synthesis of a specific compound are often clustered into operons which direct synthesis of large transcripts containing more than one coding sequence.

Translation of individual coding regions in a polycistronic transcript occurs essentially independently, such that several polypeptides are made simultaneously. This arrangement insures that groups of genes encoding proteins in the same pathways are coordinately regulated. Observations that large mRNAs could also be found in animals suggested that, like their prokaryote counterparts, eukaryotes might also synthesize polycistronic mRNAs. The finding that most of these large RNAs were not polycistronic at all, but contained introns or multiple poly(A) additions to the 3' UTR, eventually convinced most researchers that eukaryotes only synthesized monocistronic messages. This conclusion was supported by observations that translation initiation in eukaryotes follows a "ribosomal scanning" type mechanism which generally precludes translation of AUGs downstream of the first AUG in "good" context (discussed above).

Many eukaryotic viruses, both plant and animal, synthesize polycistronic mRNA and, given the ribosomal scanning hypothesis, it would seem to be an inefficient process to insure their survival. However, many of these viruses have evolved mechanisms for bypassing the restrictions in eukaryotic translation initiation (Kozak, 2001). One mechanism uses sequence-specific sites within the polycistronic mRNAs to begin translating internal exons. These sites, called IRESs (Internal Ribosome Entry Site) allow the virus to translate virtually all of the proteins encoded in the polycistronic mRNA using host machinery. Although there have been some reports of similar sequences present in eukaryotic mRNAs, very few have been rigorously tested, and it is uncertain whether they actually function as IRESs or whether a portion of the initial transcripts originate from other TSSs within upstream introns, thus appearing to function like IRES (reviewed in Kozak, 2002 and Kozak, 2003). Another mechanism used by eukaryotic viruses is that of translating polycistronic mRNAs into a single, polyprotein which is subsequently cleaved into individual proteins. Polyproteins are found naturally in yeast, for example in the mating factor α pheromone (Kurjan and Herskowitz, 1982), and animal hormones such as vasopressin and oxytocin (Richter, 1983) among others. In plants the best example of a polyprotein comes from studies of the

copied-type retrotransposons which encode a polyprotein consisting of a protease, integrase, reverse transcriptase and RNase H (Boeke and Corres, 1989). Another polyprotein paradigm is found with the mitochondrial ribosomal small subunit protein, RPS14. In some plants, *e.g.*, broad bean and pea, the gene is located in the mitochondrial genome (Wahleithner and Wolstenholme, 1988; Hoffmann *et al.*, 1999), but in others like *Arabidopsis*, maize and rice the gene is encoded in the nucleus (Figueroa *et al.*, 1999; Kubo *et al.*, 1999). In rice and maize, *rps14* has integrated between two exons of succinate dehydrogenase subunit B (SDHB), and two alternatively spliced transcripts are generated from this locus. One of the transcripts encodes a full length SDHB, while the other encodes a chimeric protein consisting of the RPS14 polypeptide fused to the C-terminus of the SDHB pre-protein, creating a nearly full-length SDHB which is missing a small portion of its C-terminal end (Figueroa *et al.*, 1999; Kubo *et al.*, 1999). The polyprotein is processed into a functional RPS14 protein in mitochondria (Figueroa *et al.*, 2000).

By definition a polycistronic transcript contains more than one coding sequence or more than one functional RNA, but there are several ways to render multiple products from a single RNA (Fig. 2.3). For example, mRNAs which are alternatively processed into two or more derivative polypeptides could be said in the strictest sense to be dicistronic. In plants a good example is represented by the chloroplastic ascorbate peroxidase locus (*ApxII*) in spinach (Ishikawa *et al.*, 1997). Two isoforms of chloroplastic ascorbate peroxidase exist, one encoding a stromal isoform and one encoding a thylakoid enzyme. Both isoforms are derived from a single 8.5 kb transcript by a mechanism of alternative splicing known as exon skipping. In this case the stromal enzyme contains exons 11 and 12 (predominately noncoding), while exon 12 is deleted from the thylakoid isoform by an alternative 3' splice site, thus joining exons 11 and 13. The alternative splice event by deleting exon 12 with its associated poly(A) site allows a second polyadenylation site at the end of exon 13 to become functional. The biological significance of this example lies in the altered location of the two isoforms within the chloroplast.

Bacteria and viruses commonly utilize overlapping polypeptides translated from a single RNA in different frames (also known as re-coding) as a mechanism for compressing genetic information (Normark *et al.*, 1983). Although this type of mechanism technically requires a bicistronic mRNA, it is (very rarely) seen in eukaryotic nuclear-encoded mRNAs (Dillon, 1987). For example, in human fibroblasts the anti-enzyme (antizyme) for ornithine decarboxylase, an enzyme involved in polyamine metabolism, is encoded by a single mRNA containing two overlapping reading frames. A very precise frameshift is required to move the ribosome from the initiator AUG of the first open reading frame (ORF) to the second ORF (Rom and Kahana, 1994). The success of this action depends on the frameshift occurring exactly one base before the stop codon of the first ORF, and also involves the six codons and a pseudoknot that follow the stop codon (Ivanov *et al.*, 2000). It is not known at this time whether or not plant polyamine metabolism utilizes a similar mechanism.

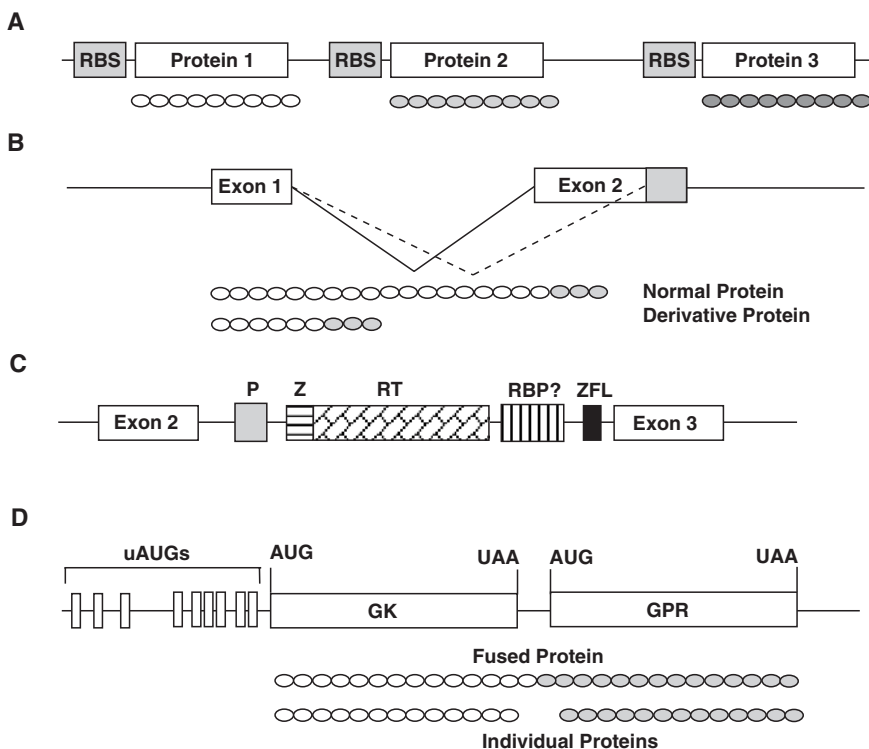


FIGURE 2.3. Different types of polycistronic mRNAs are illustrated. Exons: boxes; introns and intergenic regions: solid and dotted lines; protein products are shown below the transcripts. RBS: ribosomal binding site; AUG: initiator codon; UAA: translation stop codon. A. A typical prokaryotic polycistronic mRNA from an operon encoding three proteins is shown. B. Alternative acceptor splice sites in the mRNA generate two related polypeptides. C. An mRNA encoding ORFs within the yeast (*S. cerevisiae*) Intron 2 [Group II organellar intron] of the *coxI* gene. White boxes: exons of the primary protein; shaded/hatched boxes: intron ORFs. P: protease-like region; Z: non-long-terminal repeat retroelement; RT: reverse transcriptase domain; RBP: putative RNA intron binding domain; ZFL: zinc finger-like region. D. The organization of the putative bicistronic Δ^1 -pyrroline-5-carboxylate synthetase mRNA of tomato. GK: γ -glutamyl kinase; GPR: γ -glutamyl phosphate reductase; uAUG: upstream AUG.

This contrasts with a more common form of gene overlap where RNAs are transcribed from both strands of the same gene, although the overlapping region is usually short (< 300 b). A good example of this type of overlap is represented by the OTC and AUL1 genes in *Arabidopsis* (Quesada *et al.*, 1999).

Another type of multicistronic mRNA is represented by genes that encode a second polypeptide within an intron which gives rise to a polypeptide

different from that of the normally processed transcript. The most common examples of this type of bicistronic mRNA are found in the self-splicing introns of bacteria and eukaryote organelles, both mitochondria and chloroplast. Group I introns encode proteins exhibiting site-specific endoDNase activity (Lambowitz and Belfort, 1993). Polypeptides encoded within the Group II introns contain domains homologous to reverse transcriptase and Zn-finger proteins (Mohr *et al.*, 1993). Both types of polypeptides appear to be involved with intron movement and/or transfer. Except for the previously mentioned *rps14* gene, no documentation of such polypeptides encoded by introns in other genes transcribed by RNA pol II has been reported.

Small nucleolar RNAs (snoRNAs) are a class of noncoding RNAs believed to be of ancient origin with distinct cellular functions, including 2'-O-methylation of various RNAs, processing of rRNA transcripts, and synthesis of telomeric DNA (reviewed in Kiss, 2002). In vertebrates and yeast, the majority of snoRNAs are encoded individually in introns from which they are processed by exonucleolytic trimming of linearized snoRNA-containing intron lariats (Kiss and Filipowicz 1995; Maxwell and Fournier 1995). A few non-intron located snoRNAs are present in vertebrates, and they are transcribed from their own promoters using RNA pol II. In contrast, most of the plant snoRNAs are clustered and are transcribed into single polycistronic RNAs (reviewed in Brown *et al.*, 2003). The rare individual (non-clustered) plant snoRNAs, *e.g.*, U3, are transcribed by Pol III (Kiss *et al.*, 1991), but it is not yet known which RNA polymerase transcribes the clustered snoRNAs. Polycistronic snoRNA transcripts are processed by a mechanism that is independent of splicing (Leader *et al.*, 1997; Leader *et al.*, 1999; Brown *et al.*, 2001; Kruska *et al.*, 2003). A novel type of organization of these genes in rice (*Oryza sativa* L.) has been recently reported (Liang *et al.*, 2002), where four of the six identified snoRNA clusters were found in introns. This observation appears to be unique to plants, and, like their non-intron counterparts, these clusters were also transcribed as polycistronic RNAs.

2.3.2. Classical Bicistronic mRNA(s) in Plants

2.3.2.1. A prokaryote-like Δ^1 -pyrroline-5-carboxylate synthase in tomato

Perhaps the earliest report of a bicistronic mRNA in plants was the discovery of a locus in tomato (*tomPro1*) encoding a Δ^1 -pyrroline-5-carboxylate synthase (P5CS) that was synthesized as a single mRNA separately encoding the γ -glutamyl kinase (GK) and γ -glutamyl phosphate reductase (GPR) activities associated with P5CS, normally a bifunctional enzyme catalyzing the first step in proline synthesis (García-Ríos *et al.*, 1997). The P5CS enzyme in other plants such as *Arabidopsis* and *Vigna aconitifolia* is a hybrid enzyme with GK activity located at the N-terminal end of the polypeptide and GPR activity at the carboxy-terminus (Hu *et al.*, 1992; Savouré *et al.*, 1995; Yoshida *et al.*, 1995). Unfortunately, only antibody to the GK activity

was available in the tomato study, and the resulting polypeptide observed in Western blots was somewhat larger than predicted. It is, therefore, not known whether two separate polypeptides are synthesized *in vivo* and only the GK protein detected, being somewhat larger due to posttranslational modifications, or whether only a single, hybrid molecule is synthesized (Fig. 2.3C). These observations are complicated by subsequent studies in tomato that uncovered a second locus, *tomPro2*, which, like *Arabidopsis* and *V. aconitifolia* encodes a hybrid enzyme (Fujita *et al.*, 1998). The significance and relationship between the two tomato loci are not clear, and the possibility that the *tomPro1* transcript is processed to yield two separate RNAs or has an alternative TSS has not been rigorously excluded. If the *tomPro1* locus encodes a single hybrid polypeptide, it must arise from a mechanism like translational frame-shifting (a termination codon is present at the end of the GK coding region just 5 bases upstream of the ATG translation initiation codon of the GDR coding region); on the other hand, if two separate polypeptides are synthesized, then a ribosomal re-initiation mechanism must be operating.

The origin of *tomPro1* is also of interest, since it appears to have more homology to bacterial *proA* loci, than to plant *P5CS* genes. If, as the authors speculate, *tomPro1* originated by horizontal gene transfer from a bacterium, then fusion of the GK-GDR activities seen in the monocistronic *tomPro2* and other plant homologs of the hybrid *P5CS* gene must have occurred before the monocot-dicot split. It would then appear that the origin of *tomPro1* is a fairly recent event possibly occurring before or right after the divergence of the Solanaceae. The timing of the origin of *tomPro1* will depend on whether or not homologs can be detected in other dicot species. BLAST analysis of the *tomPro1* nucleotide sequence against the plant EST and *Arabidopsis* genome databases with the expected value set to 10 did not reveal any related genes (Bassett, personal observation), suggesting that *tomPro1* may be a feature restricted to the Solanaceae.

2.3.2.2. Multiple Bicistronic mRNAs in the *Arabidopsis* Genome

A recent genome-wide survey of *Arabidopsis* revealed some unexpected features associated with the cytochrome P450 monooxygenase (P450) gene family transcript structure (Thimmapuram *et al.*, 2005). Of the 272 individual P450 genes, transcripts from six of these genes contained more than one gene locus. Some of these loci represented tandem P450 gene family members, while others represented known or hypothetical loci lying 5' of a P450 gene. In some cases the proteins are predicted to be translated into individual polypeptides; however, in at least two cases, conceptual translation of the bicistronic transcript predicts fusion proteins, one containing a P450 protein fused to an O-methyl transferase, and one containing two P450 proteins joined to create duplicate heme-binding domains (summarized in Table 2.1.).

TABLE 2.1. Summary of *Arabidopsis* P450 bicistronic transcripts.

Gene loci	Monocistronic transcripts	Predicted protein	Bicistronic transcript ID	Predicted product
At3g26310	BT_011754	CYP71B35	AY139766 [retains 2 nd intron of At3g26310]	Truncated CYP71B35 and a full length CYP71B34 protein
At3g26300	NM_113537 ^a	CYP71B34		
At3g20080	AY65358		AY064016	Products not likely ^b
	BT001066	CYP705A15	AY090446	Full length CYP705A15
	AY064146			
At3g20083	No ^c	CYP705A16		
At3g53130	AY091083		AF367289	Full length CYP97C1 and OMT
	AY424805	CYP97C1		
At3g53140	AY089164	O-methyl Transferase [OMT]		
	AY133618			
At5g57260	BT004038	CYP71B10	NM_125108 ^a	Fusion protein containing entire CYP71B10 joined to the predicted At5g57250 protein
At5g57250	No	Hypothetical		
At4g59480	RAFL15-02-O16 ^a	CYP96A9	NM_120108 ^a	Fusion protein merging CYP96A9 with CYP96A10
At4g59490	No	CYP96A10		
At4g20240	No	CYP71A27	NM_118143 ^a	Fusion of CYP71A27 with CYP71A28
		CYP71A28		

^a Provisional REFSEQ.

^b Products are predicted to be truncated at aminoterminal end due to long 5' UTRs and unpredicted splicing events.

^c No = no individual full length transcript identified to date.

Taken from Thimmapuram *et al.* (2005)

Alignment of over 13,000 full-length cDNA clones with the *Arabidopsis* genome further identified 58 loci (almost 0.4% of the annotated genome) having transcripts containing more than one gene. Nearly three-quarters of these transcripts originate in cDNA pools from stressed plants.

2.4. Conclusion

Transcript initiation, mRNA biogenesis, and the many aspects of translational regulation of gene expression are far more complex than anyone believed in recent years. The ongoing sequencing of the genomes of various plants and animals has revealed insufficient numbers of genes needed to encode the observed proteomes in these same organisms. The rapidly growing knowledge base related to transcriptional, posttranscriptional, translational, and posttranslational regulation of gene expression has demonstrated that the observed differences between known genes and their cognate proteins is attributable, at least in part, to multiple TSSs, alternative transcript processing, differences in cytoplasmic/nuclear stability, and functional uORFs. As refinements in transcription and translation assays are made, the resulting increase in sensitivity and resolution will reveal further details pertaining to these critical biochemical events and help us better understand the relationship between gene expression and cellular complexity. Perhaps the greatest need for biologists in the near future is to develop mechanisms for collating, processing and extracting usable information from the massive accumulation of data associated with current and future genomic projects. It is possible that further developments in the creation of artificial intelligence will lead to ways in which all this information can be managed. Similarly, creating extensions of our own intelligence through implanted microchips or nano-processors may be a better alternative.

Acknowledgements. The authors would like to acknowledge the dedicated technicians and students who have provided the technical expertise which drives research forward. Without their tireless assistance, little in the way of scientific progress would be made. We would also like to acknowledge the many helpful comments and suggestions for improving this chapter made by Dr. Christopher Dardick and Dr. Ann Callahan. Results of original research presented here were supported by the USDA, Agricultural Research Service, CRIS 1931-21220-0014-00D.

References

- Adam, E., Kozma-Bognar, L., Dallmann, G., and Nagy, F., 1995, Transcription of tobacco phytochrome-A genes initiates at multiple start sites and requires multiple cis-acting regulatory elements, *Plant Mol. Biol.* **29**:983-993.
- Anderson, S., Bankier, A.T., Barrell, B.G., deBruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R., and Young, I.G., 1981, Sequence and organization of the human mitochondrial genome, *Nature* **290**:457-465.
- Antequera, F., and Bird, A.P., 1988, Unmethylated CpG islands associated with genes in higher plant DNA, *EMBO J.* **7**:2295-2299.
- Bag, J., 1991, mRNA and mRNP, in: *Translation in Eukaryotes* pp.71-79. CRC Press, Ft. Lauderdale.
- Bai, X., Peirson, B.N., Dong, F., Xue, C., and Makaroff, C.A., 1999, Isolation and characterization of SYN1, a RAD21-like gene essential for meiosis in *Arabidopsis*, *Plant Cell* **11**:417-430.
- Bassett, C.L., Nickerson, M.L., Cohen, R.A., and Rajeevan, M.S., 2000, Alternative transcript initiation and novel post-transcriptional processing of a leucine-rich repeat receptor-like protein kinase gene that responds to short-day photoperiodic floral induction in morning glory (*Ipomoea nil*), *Plant Mol. Biol.* **43**:43-58.
- Bassett, C.L., Nickerson, M.L., Farrell Jr., R.E., and Harrison, M., 2004, Multiple transcripts of a gene for a leucine-rich repeat receptor kinase from morning glory (*Ipomoea nil*) originate from different TATA boxes in a tissue-specific manner, *Mol. Gen. Genomics* **271**:752-760.
- Binder, S., Marchfelder, A., and Brennicke, A., 1996, Regulation of gene expression in plant mitochondria, *Plant Mol. Biol.* **32**:303-314.
- Boeke, J.D., and Corres, V.G., 1989, transcription and reverse transcription of retrotransposons. *Ann. Rev. Microbiol.* **43**:403-434.
- Brown, J.W.S., Clark, G.P., Leader, D.J., Simpson, C.G., and Lowe, T., 2001, Multiple snoRNA gene clusters from *Arabidopsis*, *RNA* **7**:1817-1832.
- Brown, J.W.S., Echeverria, M., and Qu, L.-H., 2003, Plant snoRNAs: functional evolution and new modes of gene expression, *Trends in Plant Sci.* **8**:42-49.
- Burke, T.W., and Kadonaga, J.T., 1996, *Drosophila* TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters, *Genes Dev.* **10**:711-724.
- Cao, J.H., and Geballe, A.P., 1995, Translational inhibition by a human cytomegalovirus upstream open reading frame despite inefficient utilization of its AUG codon, *J. Virol.* **69**:1030-1036.
- Cheng, S.H., Cline, K., and DeLisle, A.J., 1994, An *Arabidopsis* chloroplast RNA-binding protein gene encodes multiple mRNAs with different 5' ends, *Plant Physiol.* **106**:303-311.
- Closs, E.I., Lyons, C.R., Kelly, C., and Cunningham, J.M., 1993, Characterization of the third member of the mCAT family of cationic amino acid transporters, *J. Biol. Chem.* **268**:20796-20800.
- Cockell, M., and Gasser, S.M., 1999, Nuclear compartments and gene regulation. *Curr. Opin. Genet. Dev.* **9**:199-205.
- Cohen, A.J., Lakshmi, T.R., Niu, Z., Trindade, J., Billings, P.C., and Adams, S.L., 2002, A novel noncollagenous protein encoded by an alternative transcript of the chick type III collagen gene is expressed in cartilage, bone and muscle, *Mech. Dev.* **114**:177-180.

- Cooper, T.A., and Mattox, W., 1997, The regulation of splice-site selection and its role in human disease. *Am. J. Hum. Genet.* **61**:259-266.
- Damiani, Jr., R.D., and Wessler, S.R., 1993, An upstream open reading frame represses expression of *Lc*, a member of the *R/B* family of maize transcriptional activators. *Proc. Natl. Acad. Sci. USA* **90**:8244-8248.
- Dehesh, K., Franci, C., Sharrock, R.A., Somers, D.E., Welsch, J.A., and Quail, P.H., 1994, The *Arabidopsis* phytochrome A gene has multiple transcription start sites and a promoter sequence motif homologous to the repressor element of monocot phytochrome A genes. *Photochem. Photobiol.* **59**:379-384.
- Dillon, L.S., 1987, *The Gene. Its Structure, Function, and Evolution*. Plenum Press, New York, pp. 896.
- Dinesh-Kumar, S.P., and Miller, W.A., 1993, Control of start codon choice on a plant viral RNA encoding overlapping genes. *Plant Cell* **5**:679-692.
- Farrell, Jr., R.E., 2005, *RNA Methodologies: A Laboratory Guide for Isolation and Characterization*, (3/e). Elsevier Academic Press, San Diego, CA, pp. 1-767.
- Figueroa, P., Holuigue, L., Araya, A., and Jordana, X., 2000, The nuclear-encoded SDH2-RPS14 precursor is proteolytically processed between SDH2 and RPS14 to generate maize mitochondrial RPS14. *Biochem. Biophys. Res. Commun.* **271**:380-385.
- Figueroa, P., Gomez, I., Holuigue, L., Araya, A., and Jordana, X., 1999, Transfer of rps14 from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron-sulphur subunit of succinate dehydrogenase and expression by alternative splicing. *Plant J.* **18**:601-609.
- Fra, A.M., Pasqualetto, E., Mancini, M., and Sitia, R., 2000, Genomic organization and transcriptional analysis of the human genes encoding caveolin-1 and caveolin-2. *Gene* **243**:75-83.
- Fujita, T., Maggio, A., García-Ríos, M., Bressan, R.A., and Csonka, L.N., 1998, Comparative analysis of the regulation of expression and structures of two evolutionarily divergent genes for Δ^1 -pyrroline-5-carboxylate synthetase from tomato. *Plant Physiol.* **118**:661-674.
- Futterer, J., and Hohn, T., 1996, Translation in plants -rules and exceptions. *Plant Mol. Biol.* **32**:159-189.
- Gagliardi, D. and Leaver, C.J., 1999, Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. *EMBO J.* **18**:3757-3766.
- García-Ríos, M., Fujita, T., LaRosa, P.C., Locy, R.D., Clithero, J.M., Bressan, R.A., and Csonka, L.N., 1997, Cloning of a polycistronic cDNA from tomato encoding γ -glutamyl kinase and γ -glutamyl phosphate reductase. *Proc. Natl. Acad. Sci. USA* **94**:8249-8254.
- Gong, D.-W., Bi, S., Weintraub, B.D., and Reitman, M., 1998, Rat mitochondrial glycerol-3-phosphate dehydrogenase gene: multiple promoters, high levels in brown adipose tissue, and tissue-specific regulation by thyroid hormone. *DNA Cell Biol.* **17**:301-309.
- Gowda, S., Satyanarayana, T., Ayllon, M.A., Moreno, P., Flores, R., and Dawson, W.O., 2003, The conserved structures of the 5' nontranslated region of *Citrus tristeza virus* are involved in replication and virion assembly. *Virology* **317**:50-64.
- Haseloff, J., Siemering, K.R., Prasher, D.C., and Hodge, S., 1997, removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. *Proc. Natl. Acad. Sci. USA* **94**:2122-2127.

- Hayes, R., Kudla, J., and Gruissem, W., 1999, Degradation of chloroplast mRNA: the role of polyadenylation, *Trends Biochem. Sci.* **24**:199-202.
- Hoffmann, M., Dombrowski, S., Guha, C., and Binder, S., 1999, Cotranscription of the *rp15-rps14-cob* gene cluster in pea mitochondria. *Mol. Gen. Genet.* **261**:537-545.
- Holec, S., Lange, H., Kuhn, K., Alioua, M., Börner, T., and Gagliardi, D., 2006, Relaxed transcription in *Arabidopsis* mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase, *Mol. Cell. Biol.* **26**:2869-2876.
- Hu, C.A., Delauney, A.J., and Verma, D.P.S., 1992, A bifunctional enzyme (Δ^1 -pyrroline-5-carboxylate synthetase) catalyzes the first two steps in proline biosynthesis in plants, *Proc. Natl. Acad. Sci. USA* **89**:9354-9358.
- Huang, W., Wong, J.M., and Bateman, E., 1996, TATA elements direct bi-directional transcription by RNA polymerases II and III, *Nucl. Acids Res.* **24**:1158-1163.
- Ishikawa, T., Yoshimura, K., Tamoi, M., Takeda, T., and Shigeoka, S., 1997, Alternative mRNA splicing of 3'-terminal exons generates ascorbate peroxidase isoenzymes in spinach (*Spinacia oleracea*) chloroplasts, *Biochem. J.* **328**:795-800.
- Ivanov, I.P., Matsufuji, S., Murakami, Y., gesteland, R.F., and Atkins, J.F., 2000, Conservation of polyamine regulation by translational frameshifting from yeast to mammals. *EMBO J.* **19**:1907-1917.
- Joshi, C.P., 1987, An inspection of the domain between putative TATA box and translation start site in 79 plant genes, *Nucl. Acids Res.* **15**:6643-6652.
- Joshi, C.P., Zhou, H., Huang, X., and Chiang, V.L., 1997, Context sequences of translation initiation codon in plants. *Plant Molecular Biol.* **35**:993-1001.
- Ju, B.-G., Lunyak, V.V., Perissi, V., Garcia-Bassets, I., Rose, D.W., Glass, C.K., and Rosenfeld, M.G., 2006, Estrogen-initiated gene transcription involves enzymatically driven DNA cleavage, repair and local reconfiguration of chromatin. *Science* **312**:1798-1802.
- Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J.M., 2005, Atypical RNA polymerase subunits required for RNA-directed DNA methylation, *Nat Genet* **37**:761-765.
- Kiss, T., 2002, Small nucleolar RNAs: An abundant group of noncoding RNAs with diverse cellular functions, *Cell* **109**:145-148.
- Kiss, T., and Filipowicz, W., 1995, Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns, *Genes Dev.* **9**:1411-1424.
- Kiss, T., Marshallsay, C., and Filipowicz, W., 1991, Alteration of the RNA polymerase specificity of U3 snRNA genes during evolution and in vitro, *Cell* **65**:517-526.
- Kozak, M., 1978, How do eukaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**:1109-1123.
- Kozak, M., 1986, Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes, *Cell* **44**:283-292.
- Kozak, M., 1987a, An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs, *Nucl. Acids Res.* **15**, 8125-8148.
- Kozak, M., 1987b, At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells, *J. Mol. Biol.* **196**:947-950.
- Kozak, M., 1989, The scanning model for translation: an update, *J. Cell Biol.* **108**: 229-241.
- Kozak, M., 1990, Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes, *Proc. Natl. Acad. Sci. USA* **87**:8301-8305.

- Kozak, M., 1991a, An analysis of vertebrate mRNA sequences: intimations of translational control, *J. Cell Biol.* **115**: 887-903.
- Kozak, M., 1991b, Structural features in eukaryotic mRNAs that modulate initiation of translation, *J. Biol. Chem.* **266**:19867-19870.
- Kozak, M., 1997, Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6, *EMBO J.* **16**:2482-2492.
- Kozak, M., 1999, Initiation of translation in prokaryotes and eukaryotes, *Gene* **234**:187-208.
- Kozak M., 2001, Constraints on reinitiation of translation in mammals, *Nucl. Acids Res.* **29**:5226-5232.
- Kozak, M., 2002, Pushing the limits of the scanning mechanism for initiation of translation, *Gene* **299**:1-34.
- Kozak, M., 2003, Alternative ways to think about mRNA sequences and proteins that appear to promote internal initiation of translation, *Gene* **318**:1-23.
- Kravchenko, J.E., Rogozin, I.B., Koonin, E.V., and Chumakov, P.M., 2005, Transcription of mammalian messenger RNAs by a nuclear RNA polymerase of mitochondrial origin, *Nature* **436**:735-739.
- Kruszka, K., Barneche, F., Guyot, R., Ailhas, J., Meneau, I., Schiffer, S., Marchfelder, A., and Echeverria, M., 2003, Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z, *EMBO J.* **22**:621-632.
- Kubo, N., Harada, K., Hirai, A., and Kadowaki, K., 1999, A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: common use of the same mitochondrial targeting signal for different proteins. *Proc. Natl. Acad. Sci. USA* **96**:9207-9211.
- Kühn, K., Weihe, A., and Börner, T., 2005, Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*, *Nucl. Acids Res.* **33**:337-346.
- Kühn, J. and Binder, S., 2002, RT-PCR analysis of 5' to 3'-end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria, *Nucl. Acids Res.* **30**:439-446.
- Kurjan, J. and Herskowitz, I., 1982, Structure of a yeast pheromone gene (MF α): A putative α -factor precursor contains four tandem copies of mature α -factor, *Cell* **30**:933-943.
- Lagrange, T., Franzetti, B., Axelos, M., Mache, R., and Lerbs-Mache, S., 1993, Structure and expression of the nuclear gene coding for the chloroplast ribosomal protein L21: Developmental regulation of a housekeeping gene by alternative promoters, *Molec. Cell. Biol.* **13**:2614-2622.
- Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H., 1998, New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB, *Genes Dev.* **12**:34-44.
- Lambowitz, A.M., and Belfort, M., 1993, Introns as mobile genetic elements, *Ann. Rev. Biochem.* **62**:587-622.
- Lander, E.S., *et al.*, 2001, Initial sequencing and analysis of the human genome, *Nature* **409**:860-92.
- Leader, D.J., Clark, G.P., Watters, J., Beven, A.F., Shaw, P.J., and Brown, J.W., 1997, Clusters of multiple different small nucleolar RNA genes in plants are expressed as and processed from polycistronic pre-sno-RNAs, *EMBO J.* **16**:5742-5751.
- Leader, D.J., Clark, G.P., Watters, J., Beven, A.F., Shaw, P.J., and Brown, J.W., 1999, Splicing-independent processing of plant box C/D and box H/ACA small nucleolar RNAs, *Plant Mol. Biol.* **39**:1091-1100.

- Lee, S.W., Heinz, R., Robb, J., and Nazar, R.N., 1994, Differential utilization of alternative initiation sites in a plant defense gene responding to environmental stimuli, *Eur. J. Biochem.* **226**:109-114.
- Lewin, B., 2004, *Genes VIII*, Pearson Education, Upper Saddle River, NJ, pp. 617.
- Liang, D., Zhou, H., Zhang, P., Chen, Y.Q., Chen, X., Chen, C.L., and Qu, L.H., 2002, A novel gene organization: intronic snoRNA gene clusters from *Oryza sativa*, *Nuc. Acids Res.* **30**:3262-3272.
- Lisitsky, I., Klaff, P., and Schuster, G., 1997, Blocking polyadenylation of mRNA in the chloroplast inhibits its degradation. *Plant J.* **12**:1173-1178.
- Lobo, S.M., and Hernandez, N., 1989, A 7 bp mutation converts a human RNA polymerase II snRNA promoter into an RNA polymerase III promoter, *Cell* **58**:55-67.
- Luo, M., Orsi, R., Patrucco, E., Pancaldi, S., and Cella, R., 1997, Multiple transcription start sites of the carrot dihydrofolate reductase-thymidylate synthase gene, and sub-cellular localization of the bifunctional protein, *Plant Mol. Biol.* **33**:709-722.
- Lupold, D.S., Caoile, A.G., and Stern, D.B., 1999, The maize mitochondrial *cox2* gene has five promoters in two genomic regions, including a complex promoter consisting of seven overlapping units, *J. Biol. Chem.* **274**:3897-3903.
- Manley, J.L., and Proudfoot, N.J., 1994, RNA 3' ends - Formation and function, *Genes Dev.* **8**:259-264.
- Mathé, C., Sagot, M.-F., Schiex, T., and Rouzé, P., 2002, Current methods of gene prediction, their strengths and weaknesses, *Nucl. Acids Res.* **30**:4103-4117.
- Mattaj, I.W., Dathan, N.A., Parry, H.D., Carbon, P., and Krol, A., 1988, Changing the RNA polymerase specificity of U snRNA gene promoters, *Cell* **55**:435-442.
- Maxwell, E.S., and Fournier, M.J., 1995, The small nucleolar RNAs, *Annu. Rev. Biochem.* **35**: 897-934.
- Meijer, H.A., and Thomas, A.A., 2002, Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA, *Biochem J.* **367**:1-11.
- Meyers, B.C., Morgante, M., and Michelmore, R.W., 2002, TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes, *Plant J.* **32**:77-92.
- Mitchell, M.T., Hobson, G.M., and Benfield, P.A., 1992, TATA box-mediated polymerase III transcription in vitro, *J. Biol. Chem.* **267**:1995-2005.
- Misteli, T., 2000, Cell biology of transcription and pre-mRNA splicing: nuclear architecture meets nuclear function. *J. Cell Sci.* **113**:1841-1849.
- Mize, G.J., Ruan, H.J., Low, J.J., and Morris, D.R., 1998, The inhibitory upstream open reading frame from mammalian S-adenosylmethionine decarboxylase mRNA has a strict sequence specificity in critical positions. *J. Biol. Chem.* **273**, 32500-32505.
- Mohr, G., Perlman, P.S., and Lambowitz, A.M., 1993, Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function, *Nucl. Acids Res.* **21**:4991-4997.
- Morris, D.R., 1995, Growth control of translation in mammalian cells, *Prog. Nucl. Acid Res. Mol. Biol.* **51**:339-363.
- Myers, S.J., Peters, J., Huang, Y., Comer, M.B., Barthel, F. and Dingledine, R., 1998, Transcriptional regulation of the GluR2 gene: neural-specific expression, multiple promoters, and regulatory elements, *J. Neurosci.* **18**:6723-6739.
- Myers, S.J., Huang, Y., Genetta, T., and Dingledine, R., 2004, Inhibition of glutamate receptor 2 translation by a polymorphic repeat sequence in the 5'-untranslated leaders, *J. Neurosci.* **24**:3489-3499.

- Nagasaki, H., Arita, M., Nishizawa, T., Suwa, M., and Gotoh, O., 2005, Species-specific variation of alternative splicing and transcriptional initiation in six eukaryotes, *Gene* **364**:53-62.
- Nakazono, M., Itadani, H., Wakasugi, T., Tsutsumi, N., Sugiura, M., and Hirai, A., 1995, The *rps3-rpl16-nad3-rps12* gene cluster in rice mitochondrial DNA is transcribed from alternative promoters, *Curr. Genet.* **27**:184-189.
- Navarro-Avino, J.P., and Bennett, A.B., 2005, Role of a Ca^{2+} -ATPase induced by ABA and IAA in the generation of specific Ca^{2+} signals, *Biochem. Biophys. Res. Commun.* **329**:406-415.
- Navarro-Avino, J.P., Hentzen, A.E., and Bennett, A.B., 1999, Alternative transcription initiation sites generate two *LCA1* Ca^{++} -ATPase mRNA transcripts in tomato roots, *Plant Mol. Biol.* **40**:133-140.
- Normark, S., Bergstrom, S., Edlund, T., Grundstrom, T., Jaurin, B., Lindberg, F.P., and Olsson, O., 1983, Overlapping genes, *Ann. Rev. Genet.* **17**:499-525.
- Ojala, D., Montoya, J., and Attardi, G., 1981, tRNA punctuation model of RNA processing in human mitochondria, *Nature* **290**:470-474.
- Onodera, Y., J.R. Haag, T. Ream, P.C. Nunes, O. Pontes, and C.S. Pikaard. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**:613-622
- Pain, V., 1996, Initiation of protein synthesis in eukaryotic cells, *Eur. J. Biochem.* **236**:747-771.
- Pan, S., Czarnecka-Verner, E., and Gurley, W.B., 2000, Role of the TATA binding protein-transcription factor IIB interaction in supporting basal and activated transcription in plant cells, *Plant Cell* **12**:125-135.
- Pennisi, E., 1999, Keeping genome databases clean and up to date. *Science* **286**:447-450.
- Pradhan, S. and Adams, R.L.P., 1995, Distinct CG and CNG DNA methyltransferases in *Pisum sativum*, *Plant J.* **7**:471-481.
- Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L., 2005, Regulating gene expression through RNA nuclear retention, *Cell* **123**:249-263.
- Quesada, V., Ponce, M.R., and Micol, J.L., 1999, OTC and AUL1, two convergent and overlapping genes in the nuclear genome of *Arabidopsis thaliana*, *FEBS Lett.* **461**:101-106.
- Richter, D., 1983, Vasopressin and oxytocin are expressed as polyproteins, *Trends Biochem. Sci.* **8**:278-291.
- Rom, E., and Kahana, C., 1994, Polyamines regulate the expression of ornithine decarboxylase antizyme in vitro by inducing ribosomal frame-shifting. *Proc. Natl. Acad. Sci. USA* **91**:3959-3963.
- Ruiz-Echevarria, M.J., Czaplinski, K., and Peltz, S.W., 1996, Making sense of non-sense in yeast, *Trends Biochem. Sci.* **21**:433-438.
- Ruiz-Echevarria, M., and Peltz, S.W., 2000, The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames, *Cell* **101**:741-751.
- Ryabova, L.A., Pooggin, M.M., Dominguez, D.I., and Hohn, T., 2000, Continuous and discontinuous ribosome scanning on the cauliflower mosaic virus 35S RNA leader is controlled by short open reading frames, *J. Biol. Chem.* **275**:37278-37284.
- Saldanha, R., Mohr, G., Belfort, M., and Lambowitz, A.M., 1993, Group I and group II introns, *Fed. Amer. Soc. Exp. Biol. (FASEB) J.* **7**:15-24.

- Savouré, A., Jaoua, S., Hua, X.-J., Ardiles, W., Van Montagu, M., and Vergruggen, N., 1995, Isolation, characterization, and chromosomal location of a gene encoding the Δ^1 -pyrroline-5-carboxylate synthetase in *Arabidopsis thaliana*, *FEBS Lett.* **372**:13-19.
- Schul, W., van Der Kraan, I., Matera, A.G., van Driel, R., and de Jong, L., 1999, Nuclear domains enriched in RNA 3'-processing factors associate with coiled bodies and histone genes in a cell cycle-dependent manner. *Mol. Biol. Cell* **10**:3815-3824.
- Schuster, G., Lisitsky, I., and Klaff, P., 1999, Update on chloroplast molecular biology: polyadenylation and degradation of mRNA in the chloroplast, *Plant Physiol.* **120**:937-944.
- Sheen, J., 1991, Molecular mechanism underlying the differential expression of maize pyruvate, orthophosphate dikinase genes, *Plant Cell* **3**:225-245.
- Smale, S.T., and Baltimore, D., 1989, The "initiator" as a transcription control element, *Cell* **57**:103-113.
- Souciet, G., Menand, B., Ovesna, J., Cosset, A., Dietrich, A., and Wintz, H., 1999, Characterization of two bifunctional *Arabidopsis thaliana* genes coding for mitochondrial and cytosolic forms of valyl-tRNA synthetase and threonyl-tRNA synthetase by alternative use of two in-frame AUGs, *Eur. J. Biochem.* **266**:848-854.
- Spector, D.L., 1993, Macromolecular domains within the cell nucleus. *Ann. Rev. Cell Biol.* **9**:265-315.
- Tamaoki, M., Tsugawa, H., Minami, E., Kayano, T., Yamamoto, N., Kano-Murakami, Y., and Matsuoka, M., 1995, Alternative RNA products form a rice homeobox gene. *Plant J.* **7**:927-938.
- Thimmapuram, J., Duan, H., Liu, L., and Schuler, M.A., 2005, Bicistronic and fused monocistronic transcripts are derived from adjacent loci in the *Arabidopsis* genome, *RNA* **11**:128-138.
- Tracy, R.L., and Stern, D.B., 1995, Mitochondrial transcription initiation: promoter structures and RNA polymerases, *Curr. Genet.* **28**:205-216.
- Venter, J.C., *et al.*, 2001, The sequence of the human genome, *Science* **291**:1304-1351.
- Vera, A., Hirose, T., and Sugiura, M., 1996, A ribosomal protein gene (*rpl32*) from tobacco chloroplast DNA is transcribed from alternative promoters: similarities in promoter region organization in plastid housekeeping genes, *Mol. Gen. Genet.* **251**:518-525.
- Wahleithner, J.A., and Wolstenholme, D.R., 1988, A sequence encoding a maturase-related protein in a group II intron of a plant mitochondrial nad1 gene. *Nucl. Acids Res.* **16**:6897-6913.
- Wang, L., and Wessler, S.R., 2001, Role of mRNA secondary structure in translational repression of the maize transcriptional activator *Lc*, *Plant Physiol.* **125**:1380-1387.
- Wang, Z., and Sachs, M.S., 1997, Ribosome stalling is responsible for arginine-specific translational attenuation in *Neurospora crassa*, *Mol. Cell. Biol.* **17**:4904-4913.
- Weitzel, J.M., Grott, S., Radtke, C., Kutz, S., and Steitz, H.J., 2000, Multiple promoters direct the tissue-specific expression of rat mitochondrial glycerol-3-phosphate dehydrogenase, *Biol. Chem.* **381**:611-614.
- Wetzel, J.M., Kutz, S., Radtke, C., Grott, S., and Steitz, H.J., 2001, Hormonal regulation of multiple promoters of the rat mitochondrial glycerol-3-phosphate dehydrogenase gene, *Eur. J. Biochem.* **268**:4095-4103.
- Wisniewski, M.E., Bassett, C.L., Renaut, J., Farrell, Jr., R.E., Tworowski, T., and Artlip, T.S., 2006, Differential regulation of two dehydrin genes from peach (*Prunus*

- persica* L. Batsch) by photoperiod, low temperature, and water deficit, *Tree Physiol.* **26**:575-584.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A., 2003, The evolution of transcriptional regulation in eukaryotes, *Mol. Biol. Evol.* **20**:1377-1419.
- Yamada, M., Ozawa, A., Ishii, S., Shibusawa, N., Hashida, T., Ishizuka, T., Hosoya, T., Monden, T., Satoh, T., and Mori, M., 2001, Isolation and characterization of the rat prolactin-releasing peptide gene: multiple TATA boxes in the promoter region, *Biochem. Biophys. Res. Comm.* **281**:53-56
- Yoshiba, Y., Kiyosue, T., Katagiri, T., Ueda, H., Mizoguchi, T., Yamaguchi-Shinozaki, K., Wada, K., Harada, Y., and Shinozaki, K., 1995, Correlation between the induction of a gene for Δ^1 -pyrroline-5-carboxylate synthetase and the accumulation of proline in *Arabidopsis thaliana* under osmotic stress, *Plant J.* **7**:751-760.
- Zhang, Y., Niu, Z., Cohen, A.J., Nah, H.-D., and Adams, S.L., 1997, The chick type III collagen gene contains two promoters that are preferentially expressed in different cell types and are separated over 20 kb by DNA containing 23 exons, *Nucl. Acids Res.* **25**:2470-2477.

Regulation of Gene Expression in Plants
The Role of Transcript Structure and Processing
Bassett, C.L. (Ed.)
2007, XVI, 196 p., Hardcover
ISBN: 978-0-387-35449-1