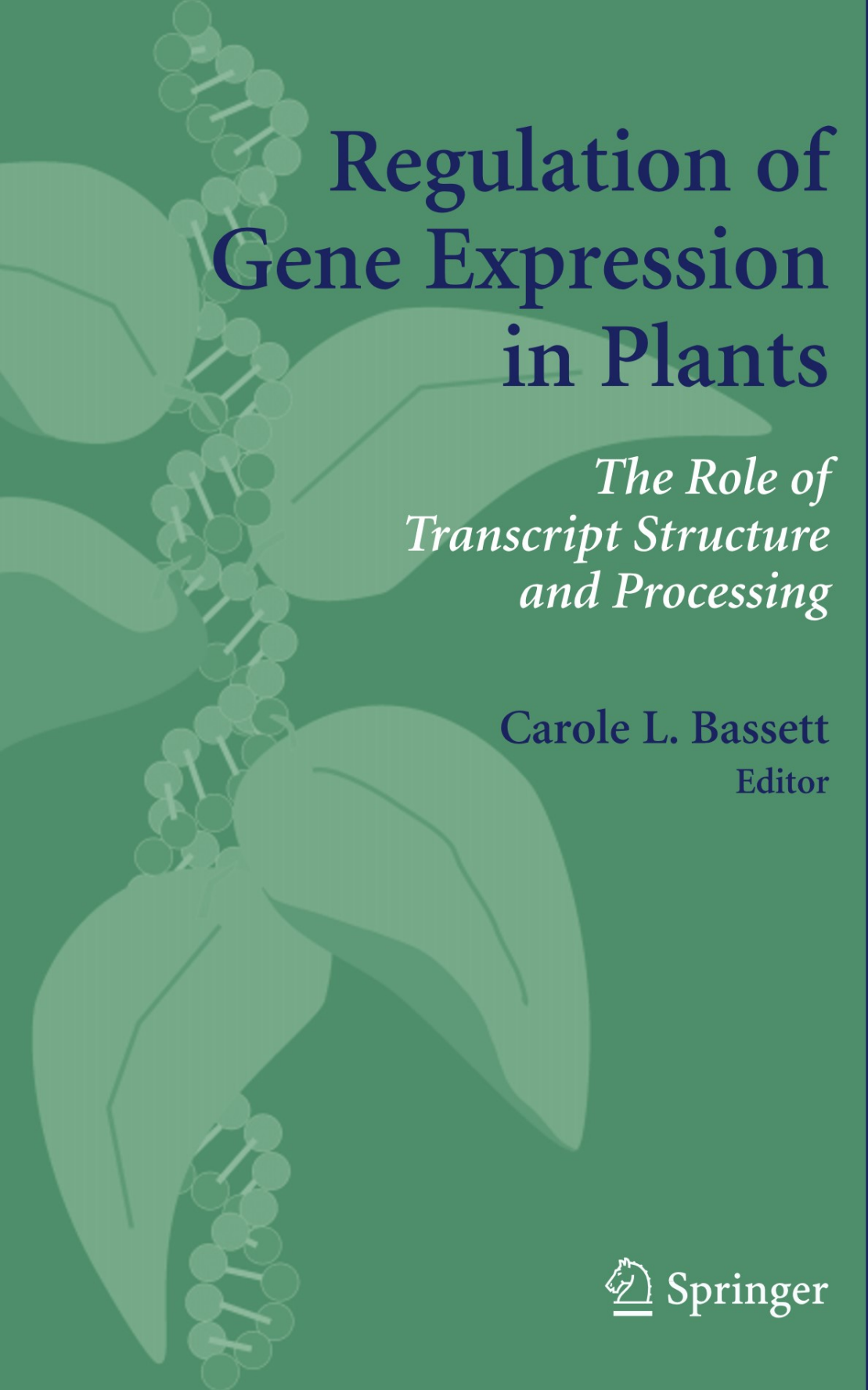


Regulation of Gene Expression in Plants

The Role of Transcript Structure and Processing

Carole L. Bassett
Editor

Springer

Regulation of Gene Expression in Plants

Regulation of Gene Expression in Plants

The Role of Transcript Structure and Processing

Edited by

Carole L. Bassett
United States Department of Agriculture
Kearneysville, WV, USA

Springer

Carole L. Bassett
USDA - ARS
2217 Wiltshire Road
Kearneysville, WV 25430
USA

Library of Congress Control Number: 2006937882

ISBN-10: 0-387-35449-2 e-ISBN-10: 0-387-35640-1
ISBN-13: 978-0-387-35449-1 e-ISBN-13: 978-0-387-35640-2

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

Dedication

The contributors would like to dedicate this book to their families and friends who have so patiently ignored the long hours and frustrations that are a natural part of scientific research and supported us intellectually and emotionally in our efforts. We hope they have also shared with us in the satisfaction of achievement and the joy of discovery that occur all too infrequently in the realm of science. The editor would also like to dedicate this book to her father, Dr. Boude Bowman Leavel (1919-2003), who inspired her interest in science and nurtured her love of learning. He was a dedicated physician and as close to a perfect husband and father as will ever be found.

Foreword

Regulation of gene expression in eukaryotes is complex and occurs on many levels, often overlapping each other. In a hierarchical sense, the chromatin structure itself is the first level influencing gene expression through methylation of select bases, posttranslational modification of histones, and alterations in scaffolding. The second level includes transcription and posttranscriptional processing, including export and turnover of mRNAs. It is with this level of gene regulation that the current book is primarily focused. The third level of regulation involves translation and posttranslational events. As we shall see in the following chapters, it is sometimes difficult to separate these processes from each other. Of necessity we have narrowed our emphasis on transcript expression to the structure and processing of mRNAs and have opted to minimize the mechanics of transcription initiation, elongation, and termination. This is not to say that such processes are not important, only that they have been amply covered in numerous recent reviews.

Preface

In almost all of the areas of gene expression control except one, plant research has lagged considerably behind studies in yeast, insects and vertebrates. Advances in animal gene expression control have also benefited plant research, as we continue to find that much of the machinery and mechanisms controlling gene expression have been preserved in all eukaryotes. However, there are some interesting differences in gene structure and regulation between plants and animals. First, vertebrate genes can be quite large, often spanning tens of thousands of base pairs and usually separated by numerous large introns, whereas plant genes tend to be much smaller (averaging between 1–2 kb (kilobases) with fewer and smaller introns. Second, as we shall see in the following chapters, plant transcripts retain introns more often than do animal transcripts (30% of all genes in the model plant, *Arabidopsis,* compared to 10% in humans). Unfortunately, at this time we have only a few plant models for gene regulation, and only *Arabidopsis*, rice and poplar have been fully sequenced. Since *Arabidopsis* was the first plant genome to be fully sequenced, most of our information has come from studies of its transcriptome, and it is not known to what extent it truly represents other members of the plant kingdom. In addition, as our knowledge of factors influencing gene expression increases, so too, does our recognition that the current annotations in the gene databases will need to be updated to reflect new information as it appears in the literature. Finally, compared to animals, plants have evolved different signaling mechanisms, partly because plant hormones do not exactly function as do those in animals, and partly because plants must cope with environmental changes differently than animals, since they cannot physically escape their environments except possibly through reproduction. Therefore, plants have evolved very complex, interacting signaling pathways in response to developmental signals and biotic/abiotic stresses. All of these observations ultimately reflect some of the differences plants display in regulating gene expression compared to yeasts, insects and vertebrates.

Although we have touched upon some of the differences between animal and plant control of gene expression here, it is equally important to recognize and appreciate the common mechanisms they share. For example, the basic

transcriptional machinery *via* DNA-dependent RNA Polymerase II is virtually the same in plants and animals. Furthermore, some transcription factors, like *myb* and *myc* factors are similar in structure and function in both plants and animals. Plants and animals contain introns separating the coding regions of most genes and again, they utilize similar machinery to process the introns and form mature mRNAs. Since translation in all eukaryotes is basically the same, we see more similarities between plants and animals in this process, than differences. These mechanisms relate to mRNA structure, including sequences in the 5′ leader region and those in the 3′ untranslated region which influence the efficiency and selectivity of translation. It is hoped that the following chapters will expose the reader to some of the most recent, novel and fascinating examples of transcriptional and posttranscriptional control of gene expression in plants and, where appropriate, provide comparison to notable examples of animal gene regulation.

Contents

Contributors

Carole L. Bassett
USDA-ARS, Appalachian Fruit Research Station,
2217 Wiltshire Road, Kearneysville, WV 25430, USA

Robert E. Farrell, Jr.
Biology Department, Pennsylvania State University,
208 Mueller Lab, University Park, PA 16802-5301, USA

Jean-Michel Hily
Génomique dévelopment du Pouvoir pathogène, INRA-Université Bordeaux,
Segalen/Bordeaux, BP 81, 33883 Villenave d'Ornon, France

Arthur G. Hunt
Department of Plant and Soil Sciences, University of Kentucky,
Plant Sciences Building, 1405 Veterans Drive, Lexington, KY 40546-0312, USA

Zongrang Liu
USDA-ARS, Appalachian Fruit Research Station,
2217 Wiltshire Road, Kearneysville, WV 25430, USA

Eliezer S. Louzada
Department of Agronomy and Resource Sciences, Texas A & M University,
312 N. International Blvd., Weslaco, TX 78596, USA

1
The Regulation of Gene Expression in Plants and Animals

Robert E. Farrell, Jr.

1.1. Overview of Eukaryotic Transcription

The control of gene expression in all cells involves an elaborate and dynamic interplay among what might best be described as regulatory molecules. These molecules include RNA polymerases, myriad transcription factors, the DNA template, the RNA produced by transcription, and the protein produced by translation with its attendant processing. To interfere with or in some way modify any of these critical elements is to potentially cause a profound change in the phenotypic manifestation of one or more genes or gene relays. It is clear that transcription is a consequence of a series of well-orchestrated, ordered events. Contemporary methods that examine this aspect of gene expression have revealed the dynamic nature of this biochemical process.

The examination of gene expression often revolves around quantifying the abundance of a particular transcript. This assessment may be performed using state-of-the-art methods such as high throughput microarrays and real-time PCR, or by the use of the time-honored methods of Northern analysis, nuclease protection, or semi-quantitative PCR (described in Farrell, 2005). While all of these methods provide information about the transcriptional activity of genes, they measure steady-state levels only, i.e., the final accumulation of RNA in the cell at the moment of lysis; these methods fail to take into account the stability of the RNA, as well as the rate of transcription at the specific loci under investigation. In order to describe more completely the nature of gene expression modulation, it is incumbent upon the investigator to examine the level of the corresponding protein(s) so as to determine whether an alteration of transcript levels, protein levels, or both are in some way associated with a phenotypic change.

1.1.1. Regulation of Gene Expression

The modulation of transcript and protein levels is a nonstatic, on-going process that governs all cell and tissue function. Experiments performed using adenovirus and SV40 models were the first to elucidate some of the

biochemical events associated with the synthesis of mRNA. Current research has demonstrated that the secondary and tertiary structures formed by RNA and numerous RNA binding proteins influence the posttranscriptional fate of the transcript. While the number of potential gene regulatory control points is virtually infinite, in the *broadest* sense one might consider the following four major headings:

1. Transcriptional regulation
2. Posttranscriptional regulation
3. Translational regulation
4. Posttranslational regulation

Eukaryotic RNA molecules are synthesized by transcription in the nucleus, in mitochondria, or in chloroplasts. This is transcriptional-level regulation: a gene either is or is not transcribed. The same subcellular compartment where transcripts are synthesized is also the location where they are processed to maturity. Maturation, stability, and export of mRNAs represent posttranscriptional levels of gene regulation. With respect to nuclear transcripts, mature mRNA molecules may be exported into the cytoplasm, usually lacking the intron sequences that were part of heterogeneous nuclear RNA (hnRNA). Nucleocytoplasmic transport is an extremely discriminating mechanism, involving movement to and then through nuclear pore complexes (for reviews see Piñol-Roma and Dreyfuss, 1993; Maquat, 1997; Chapter 6). It is very likely that mature mRNA will become translated, but this is not always the case. The chemical stability of mRNA is itself an important posttranscriptional regulatory control point, as is the formation of nontranslatable mRNA:protein complexes previously known as "informosomes" (Preobrazhensky and Spirin, 1978; Bag, 1991). Translation initiation is another major control point affecting gene expression. mRNAs that bind eukaryotic initiation factors (eIFs) with low affinity typically do not synthesize as much polypeptide product as mRNAs with a higher binding affinity. This level of control occurs at the translational level and includes the elongation and termination of polypeptide synthesis. If translation of a particular mRNA does occur, then the translated polypeptide itself is subject to extensive posttranslational modifications. The magnitude and diverse nature of these posttranslational modifications is only now coming to light with the advent of the emerging field of proteomics. Some of the common posttranslational modifications include, but are not limited to, proteolytic cleavage (with or without ubiquitination), acylation, methylation, prenylation, carboxylation, glycosylation, phosphorylation, neddylation, acetylation, sumoylation, and hydroxylation.

The biochemical status of a cell or tissue sample can be qualitatively and quantitatively defined as a function of mRNA complexity. For example, one may investigate whether modulation of gene expression is regulated transcriptionally or by a posttranscriptional event(s). Thus, insightful consideration

must be given to the method of cellular disruption and RNA purification because of the numerous sub-populations of transcripts in the cell (for protocols, see Farrell, 2005). In order to paint a more complete picture of the methods by which one or more genes of experimental interest are being regulated in a model system, the abundance of the corresponding protein(s), and modulation thereof, should also be scrutinized.

1.1.2. Nature of Transcription

Transcription is that process by which the information residing in a double-stranded DNA molecule (dsDNA) is transferred to a nascent RNA molecule. "Transcription" is so-named because the code contained within the sequence of the DNA is rewritten in the same language, namely, the language of nucleic acids. RNA polynucleotides are assembled at a rate of about 40 nucleotides (nt)/sec in eukaryotic and in prokaryotic cells. Translation, in contrast, is that process by which a protein or polypeptide is synthesized from the information encoded in the RNA; it is so-named because the information contained within the RNA is rewritten in an entirely different language, namely, the language of proteins. Proteins are assembled from amino acid monomers at a rate of about 5–8 amino acids (aa)/sec in eukaryotes and about twice as fast in bacteria. The mechanics of the processes of transcription and translation ensure the requisite colinearity and accuracy of information transfer from the DNA to the resulting protein which, in turn, ensures proper biochemical function. To do otherwise would constitute a mutation, with potentially serious consequences for both the cell and the organism.

In eukaryotic cells, transcription is compartmentalized within membrane-bound organelles and, in the case of the nucleus, transcription is both spatially and temporally separated from translation. In both plant and animal cells, the half-life of a typical mRNA transcript is significantly greater than the half-life of mRNA molecules in bacteria; prokaryotic transcripts are often being both translated and degraded from their 5′ ends even while transcription is still on-going at the 3′ ends of the "nascent" transcripts. The enhanced stability of eukaryotic transcripts clearly facilitates transport across the nuclear membrane into the cytoplasm where mRNAs may have an opportunity to engage the protein translation apparatus.

RNA is produced from different loci at different rates. The quantity of a particular RNA species in a cell is referred to as its abundance. It is common to classify genes based on the abundance category into which their transcripts fall and to describe measured changes in the expression of a gene at the RNA level as a change in the relative abundance of that transcript. This terminology is used to describe (1) a change in the quantity of a particular transcript relative to another transcript, such as a housekeeping gene, or (2) a change in the quantity of a transcript after experimental manipulation relative to an untreated control.

1.1.2.1. Important Plant and Animal Models for Studying Transcription

Plants: The genomes of *Arabidopsis* (*Arabidopsis thaliana*), rice (*Oryza sativa*), and poplar (*Populus trichocarpa*) have been completely sequenced. Projects to sequence tomato (*Lycopersicon esculentum*), *Medicago truncatula* (a forage legume) and lotus (*Lotus japonica*) and many other plants are currently underway. Because *Arabidopsis* was the first plant whose genome was completely sequenced, it has served as a model plant for gene organization and structure, as well as for transcript processing and regulation.

Models for inheritance and gene mapping studies have historically been maize (*Zea mays*) and tomato. More recently wheat and barley, as well as soybean and alfalfa have been the focus of trait mapping, especially QTL (quantitative trait loci) identification of disease and stress resistance. Genetic studies in plants are often complicated by the fact that some of the more important crops are polyploids or aneuploids. Common plant models for the study of non-Mendelian inheritance include shoot variegation in the four o'clock *Mirabilis jalapa*, cytoplasmic male sterility in corn, and uniparental inheritance of mating type in the alga, *Chlamydomonas reinhardtii*.

Animals: Much of the pioneering research in the area of transcription and transcriptional control of gene expression involved the use of such unassuming animals as *Drosophila*, sea urchins, *Caenorhabditis elegans* (nematode), rat, mouse, and just as importantly, animal cell culture models, such as the CHO-K1 (Chinese Hamster Ovary) and NIH 3T3 (swiss mouse fibroblast) cell lines.

1.1.2.2. Genome Organization Affects Nuclear Gene Transcription

The eukaryotic genome is both large and quite sophisticated. If the DNA from a typical eukaryotic cell were stretched end to end it would reach several feet in length. Organization of the DNA is critically important for packaging into a microscopic nucleus, which must be done in such a way that specific DNA sequences (genes) and their regulatory elements (promoters) can be periodically accessed in order to be transcribed. Supercoiling is the key.

The first level of eukaryotic nuclear DNA organization involves the formation of nucleosome particles. Nucleosomes consist of an association of 146 bp of core DNA wrapped around a group of eight histone proteins, i.e., two molecules each of the histones H2A, H2B, H3, and H4. The orderly assemblage of nucleosomes is often described as beads on a string when viewed by electron microscopy. A fifth histone protein, H1, is located on the outside of the nucleosome particle and can be readily removed from the complex under conditions of low ionic strength.

Prior to the initial interaction of the chromatin with the early initiators of transcription, this three-dimensional nucleosome architecture prevents RNA polymerase and transcription factor association with the promoter. The local unwinding of tightly packaged DNA is a first step needed to provide transcription factors the opportunity to interact with the regulatory elements governing expression of a particular gene. Beyond the formation of

nucleosomes, subsequent levels of coiling ultimately lead to the characteristic formation of chromatin which, at the onset of mitosis, further condenses into recognizable chromosomes.

1.1.2.3. Gene Organization Affects Nuclear Gene Transcription

The typical eukaryotic genome is well-known for its excessive amount of non-coding, intergenic spacers, as well as the fact that the structural portions of genes are divided into coding regions, known as exons, and intervening, noncoding regions known as introns. While the precise origin of interrupted genes is anyone's guess, one of the more rational attempts to explain this phenomenon is that introns are remnants of genome invasion attempts by primordial prokaryotes. The intron sequences that are observed distributed throughout plant and animal genes are variable in number, length, and nucleotide content, and infrequently are found to contain open reading frames (ORFs). The only predictable feature of introns is the presence of highly conserved dinucleotide pairs that define the intron boundaries: introns begin with GU and end with AG. By recognizing these dinucleotides, and the adjacent bases that are conserved to a lesser extent, introns are excised from hnRNA transcripts during and shortly after transcription. A considerable amount of the typical eukaryotic genome does not appear to encode anything, and these intergenic regions are often impressive in size.

Most animal cells are diploid with the exception of gametes. As such, each cell has at least two copies of each gene, and there are many genes that exist in multiple copies. The genome is also subject to change, not only by random mutation, but also through the orderly duplication of specific genes in response to external stimuli. For example, the drug methotrexate, which is commonly used to treat patients with either cancer or arthritis, is known to induce amplification of the dihydrofolate reductase (DHFR) gene (Schimke, 1981). Another example of the nonstatic nature of the animal genome occurs in lymphocytes, in which the immunoglobulin genes are spliced in order to facilitate the production of a customized antibody in response to exposure to a previously unknown antigen (reviewed by Lewin, 2004).

In plants, the parameters governing genome stability are similar to those in animals, though it is not unusual to find triploid, tetraploid, and octaploid tissues and organisms in the plant kingdom. In addition, plants seem better able to tolerate aneuploidy than mammals, and some hybrid progeny from interspecies crosses in the plant kingdom are fertile, whereas this rarely happens in higher animals.

1.1.2.4. RNA Polymerases I, II, III, and IV

Genes in plant cells, as in animal cells, are transcribed by enzymes known as DNA-dependent RNA polymerases. In eukaryotic organisms there are three nuclear RNA polymerases, designated RNA polymerase I, RNA polymerase II, and RNA polymerase III. These enzymes, which are composed of

multiple subunits and are among the largest and most complex enzymes in the cell, each transcribe a different class of genes. This approach to transcription is markedly different compared to prokaryotic cells where one type of RNA polymerase is responsible for the transcription of all genes. Each eukaryotic RNA polymerase is functional only in the presence of DNA, which acts as a template. These enzymes also exhibit an absolute cofactor requirement for Mg^{++} and are dependent upon the presence of myriad transcription factors to initiate and support transcription.

RNA polymerase I is responsible for the transcription of the ribosomal genes, and is therefore ultimately responsible for the production of 28S ribosomal RNA (rRNA), 18S rRNA, and 5.8S rRNA. RNA polymerase II transcribes genes that encode proteins, first producing hnRNA which can mature into mRNA molecules that can support translation. RNA polymerase III transcribes the transfer RNA (tRNA) genes and a small ribosomal rRNA gene that produces the 5S rRNA. RNA polymerases are also responsible for the synthesis of small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), and small cytoplasmic RNA (scRNA). In mammals, a fourth RNA polymerase (single-polypeptide nuclear RNA polymerase IV [spRNAP-IV]) has recently been described (Kravchenko *et al*., 2005). This polypeptide is derived from a nuclear-encoded mitochondrial RNA polymerase by alternative splicing which results in a truncated polypeptide lacking 262 amino acids near the amino terminus, including the mitochondrial transit sequence. spRNAP-IV is located in the nucleus and regulates a subset of nuclear-encoded genes.

Similarly, a novel polymerase, RNA polymerase IV, or simply Pol IV, has been discovered in plants (Onodera *et al*., 2005). Pol IV does not appear to be necessary for viability, nor does its activity overlap with the known functions of RNA polymerases I, II, and III; in contrast, it is used by the cell to promote methylation-associated higher-order heterochromatin formation (Kanno *et al*., 2005; Chapter 5).

As with all nucleic acid polymerases, RNA polynucleotides are assembled $5' \rightarrow 3'$, and in this case, in a DNA-dependent manner. Mutations notwithstanding, the products of transcription are identical to the base sequence found on the DNA coding strand, with the exception of the substitution of uracil in RNA in place of the thymine found in DNA. Thus, the newly transcribed RNA is likewise complementary to the DNA template strand upon which it was synthesized directly. Curiously, more than 70% of all transcribed RNA is subsequently degraded in the nucleus, never maturing into corresponding cytoplasmic species (Soeiro *et al*., 1968; Jackson *et al*., 2000). While this may seem like an inordinate waste of cellular resources, it is nonetheless a well-documented mode of posttranscriptional regulation of gene expression.

1.1.2.5. Phage-Type RNA Polymerases: The RpoT Genes

Angiosperms possess a small, unique group of nuclear-encoded RNA polymerases that resemble the well-characterized RNA polymerases associated with the T7 and T3 bacteriophages. These genes, known as the RpoT

polymerase family, are believed to have arisen by duplication from an ancestral gene encoding the mitochondrial RNA polymerase. At least some of the genes in this family are found in both monocots and dicots. The RpoT genes were initially identified and characterized in *Arabidopsis thaliana* (Hedtke *et al*., 1997), *Cheopodium album* (Weihe *et al*., 1997), *Physcomitella patens* (Richter *et al*., 2002), maize (Young *et al*., 1998; Chang *et al*., 1999), and wheat (Ikeda and Gray, 1999). In *Arabidopsis*, the three members of this family that have been found so far are known as RpoT1, RpoT2, and RpoT3. RpoT1 is responsible for the transcription of mtDNA (mitochondrial DNA) while RpoT3 transcribes cpDNA (chloroplast DNA). Particularly interesting is RpoT2, the N-terminal amino acid composition of which directs RpoT2 into chloroplasts and mitochondria, suggesting a dual transcriptional role for this enzyme involving two different genomes (Hedtke *et al*., 2000). In moss, the RpoT2 mRNA contains two inframe AUG codons near the 5′ terminus (Richter *et al*., 2002). When translation initiation occurs at the second AUG codon, the resulting enzyme is directed into the mitochondria. Translation initiation from the first AUG codon, in contrast, produces a chloroplast-targeted translation product.

1.1.3. Transcription Factors and Promoter Elements

Initiation of the process of transcription is a complex event, involving far more components than were appreciated even as recently as five years ago. The specific DNA sequences required for the binding of transcription-associated proteins and enzymes are known generically as regulatory elements, and these include motifs commonly known as the TATA box, the CAAT box, enhancers, and the GC-rich elements. The highly variegated proteins that bind to the DNA and to each other to facilitate the initiation of transcription are collectively referred to as transcription factors. Those DNA sequences that are spatially associated with a particular genetic locus are known as *cis*-acting elements, while the various RNA polymerases and all of the transcription factors that are encoded at other genetic loci, are referred to as *trans*-acting factors. For example, in plants, the *cis*-acting motif, (T)(T)TGAC(C/T), is known as a W box, to which members of the WRKY superfamily of transcription factors bind in response to wounding and certain other stresses (Cheong *et al*., 2002). Before an RNA polymerase can engage a transcription template, a pre-initiation complex consisting of several different transcription factors binds to the DNA promoter itself, and this precedes the binding of RNA polymerase (reviewed by Lewin, 2004). An overview of transcription and associated processes is illustrated in Figure 1.1.

The transcription factors associated with genes that encode proteins, as opposed to genes which encode tRNA and rRNA, appear to be highly conserved among eukaryotes, as is RNA polymerase II, which transcribes these genes. Basal transcription factors associated with genes that encode mRNA are generically referred to as $TF_{II}X$, with X representing a specific

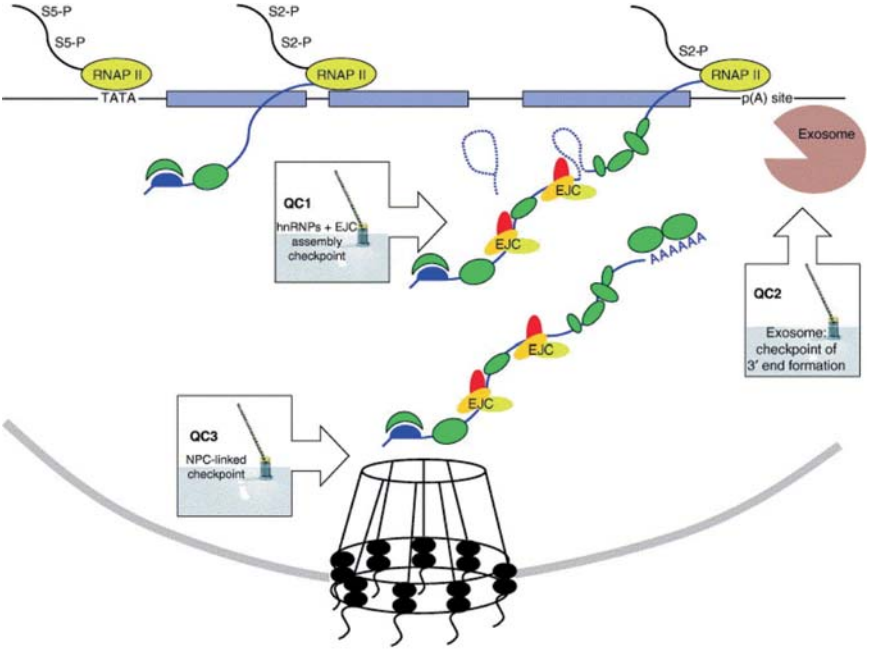


FIGURE 1.1. Key features and quality control checkpoints of the 'mRNA factory'. In the CTD (C-terminal domain) of RNAP II, the changing phosphorylation of repeat residues Serine 5 (S5) and Serine 2 (S2) during transcription regulates the dynamic interactions of the CTD with enzymes of mRNA maturation (capping, splicing and polyadenylation factors, not shown). The 5′ cap is added early in transcription, whereas splicing of introns usually occurs cotranscriptionally but can happen later as well. Splicing leads to a deposition of the exon junction complex (EJC) upstream of the splice junctions. Numerous hnRNPs and other RNA binding proteins (represented as green symbols of different shapes and sizes), including cap binding protein eIF4E and poly(A) binding protein, associate with maturing transcripts; the complement of these factors is likely to be distinct for different mRNAs. The appropriate assembly of hnRNP proteins and completion of 3′ end formation is monitored at quality control checkpoints QC1 and QC2, respectively; messages improperly matured at these stages are subject to retention at the transcription site and degradation by the exosome complex. The nuclear pore complex-linked checkpoint, QC3, ensures that only spliced transcripts are exported from the nucleus and causes mRNA to be kept at the site of transcription when the nuclear-pore complex-linked export step is blocked. Reproduced with permission, from Belostotsky and Rose, 2005, Trends in Plant Science, 10(7), 347–353, © Elsevier Ltd. (*See Color Plates*)

transcription factor, *e.g.*, $TF_{II}D$. Table 1.1 lists several representative classes of transcription factors that have been identified in plants and in animals.

Many promoters exhibit a DNA sequence known as a TATA box, so named because of the highly conserved sequence, TATAA, or a close variant

TABLE 1.1. Classes of eukaryotic transcription factors[1].

	Transcription factor categories			
Super family	Examples of major classes	Observed in plants	Observed in animals	Transcriptional factor examples
Basic domains	Leucine zippers	Yes	Yes	c-jun (human) c-fos (mouse) CPRF-3 (parsley)
	Helix-loop-helix	Yes	Yes	Id2 (mouse) Lc (maize) NUC-1 (*Neurospora*)
	NF-1	No	Yes	CTF-3 (human) NF-1C1 (pig) NF-1A1 (chick)
Zinc-coordinating DNA binding Domains	Cys4 zinc finger (nuclear receptor type)	No	Yes	T3R-α (rat) AR (human) CF1 (fruit fly)
	Cys2-His2 zinc finger domain	Yes	Yes	$TF_{III}A$ (ubiquitous TF WT1 (human) GAL4 (yeast)
	Diverse Cys4 zinc fingers	Yes	Yes	GATA-1 (frog) AREA/NIT-2 (*Neurospora*)
Helix-turn-helix	Homeo domain	Yes	Yes	HOXC6 (zebra fish) Knox3 (barley) KN1 (maize)
	Heat shock factors	Yes	Yes	HSF1 (human) HSF3 (chick) HSF24 (tomato)
	Tryptophan clusters	Yes	Yes	c-myb (human) GL1 (*Arabidopsis*) MYB.pH3 (petunia)
β-Scaffold factors	STAT	No	Yes	STAT2 (human) STAT4 (mouse) STAT5A (sheep)
	p53	No	Yes	p53 (human) p53as (mouse)
	MADS Box	Yes	Yes	GLO (tobacco) AG (*Arabidopsis*) ZEM1 (maize)
	TATA-binding proteins	Yes	Yes	TBP (human) TBP (*Arabidopsis*) TBP (yeast)
Other types of transcription factors	Copper fist proteins (fungal)	No	No	ACE1 (yeast) AMT1 (yeast)
	HMGI(Y)	No	Yes	HMG Y (human) HMGI-C (mouse)
	Pocket domain	No	Yes	Rb (mink) CBP (mouse) p107 (human)

(Continued)

TABLE 1.1. Classes of eukaryotic transcription factors[1]. — Cont'd

Super family	Transcription factor categories: Examples of major classes	Observed in plants	Observed in animals	Transcriptional factor examples
	E1A-like factors (adenovirus)	No	No	E1A 12S (adenovirus) E1A 13S (adenovirus)
	AP2/EREBP-related factors	Yes	No	AP2 (*Arabidopsis*) EREBP-1 (tobacco) ARF1 (*Arabidopsis*)

[1] Some of this information was derived from www.gene-regulation.com/pub/databases/transfac/cl.html.
Visit this site for a more comprehensive listing of known TFs.

thereof. Most often a TATAA box is situated at –25, meaning 25 nts upstream from the transcription start site (TSS). The TATA box is considered a fundamental part of transcript initiation, although transcription can also be initiated from TATA-less promoters. In animals this occurs when the transcription factor IIB recognition element ($TF_{II}B$; Lagrange et al, 1998) and downstream promoter element (DPE; Burke and Kadonaga, 1996) are present in place of the TATA box element. An additional motif, known as an initiator sequence (Inr) is able to promote transcription autonomously or in conjunction with a TATA or other element (Smale and Baltimore, 1989). The precise geometry of these elements will determine the TSS. In *Arabidopsis*, the TATA box is the primary regulatory element required to initiate transcription in less than one-third of promoters analyzed (assuming that full length cDNAs represented the primary transcript; Molina and Grotewold 2005). This means fully two-thirds of the putative *Arabidopsis* promoters are without TATA boxes. It is noteworthy that the DPE and Inr sequences mentioned above in TATA-less animal promoters were not over-represented in the TATA-less *Arabidopsis* promoters, suggesting that there are fundamental architectural differences between plant and animal promoters.

Tissue-specific expression of particular genes is fundamental to proper function at the tissue level and beyond. As such, identification of the functional TATA box and its associated regulatory element in a particular tissue provides a greater understanding of gene regulation in that tissue. Currently available software is particularly useful for the *in silico* identification of potentially functional TATA boxes, especially when the region upstream of the TSS is AT-rich. While no single software product can predict TATA box functionality with absolute confidence, merging the results generated through the use of several different gene analysis programs allows the investigator to make a more informed judgment. For example, the presence of three operational TATA boxes was recently reported in the promoter region of the

inrpk1 gene of morning glory using this basic strategy (Bassett *et al*., 2004). This gene encodes a leucine-rich repeat receptor kinase, and using 5′ RLM-RACE with RNA from roots, first leaves, and cotyledons, the selection of TATA box from within the promoter was confirmed to be tissue-specific. In addition, through the use of different primer combinations to include, or exclude, a particular TSS, it was demonstrated that no fewer than nine TSSs were utilized in a highly tissue-specific manner. While the occurrence of multiple promoters for a specific gene is common in animal models, it appears to be considerably rarer in plants.

The overall efficiency of a promoter can be markedly increased by coming under the influence of an enhancer sequence, which can be located a short distance upstream or downstream of the promoter or, occasionally, ten or more kilobases (kb) away from the promoter. These interesting DNA sequences consist of elements that are similar to those found within the promoters themselves and appear to function most often in *cis*. It is widely believed that enhancer sequences may exert their influence by increasing the local concentration of transcription activators. Silencers, which might best be regarded as negative enhancers, can also modulate gene expression by repressing transcription from a distance of up to 2.5 kb away.

1.1.4. Chromosomal Structure Influences Gene Expression

That a number of diverse elements influence the initiation and progression of transcription is undisputed. This includes the structure of the chromatin itself. The formation of the modified base methylcytosine is commonly associated with CG doublets (often written as CpG) within genomic DNA, and methylated promoter sequences very often interfere with transcription factor binding. CG islands consist of multiple repetitions of this CG dinucleotide and are found among all vertebrates near the promoters of genes that are expressed constitutively; CG islands occur with a frequency of 29,000 in the human genome (Lander *et al*., 2001; Venter *et al*., 2001). As with vertebrates, angiosperm genomes likewise exhibit CG islands (Antequera and Bird, 1988), and more recent research has demonstrated that a majority of CG clusters are methylated and widespread in plants (Pradhan and Adams, 1995). Higher plants also display cytosine methylation associated with CNG[1] and CWG[2] trinucleotides, and two different methyltransferases have been isolated from *Pisum sativum* which are independently responsible for the methylation of CNG and CWG clusters (Pradhan and Adams, 1995). While two different enzymes suggest different regulatory roles for CNG and CWG methylation, the precise function(s) of these methylation patterns in plants

[1] N = A, C, G, or T
[2] W = A or T

remains somewhat obscure. It is clear, however, that promoter methylation correlates well with transcription suppression.

Transcription necessitates a local, transient disruption of the chromatin in order for RNA polymerase and all requisite transcription factors to engage the promoter associated with the gene. Perhaps the most fundamental events leading to the onset of transcription are the covalent posttranslational modifications of the histone proteins that permit a local relaxation of supercoiled chromatin. Interestingly, the sites of these modifications tend to be localized close to the N-termini of the histones. Histone modifications, particularly of H3 and H4, include acetylation of lysine, methylation of lysine, histidine, and arginine, and phosphorylation of histidine and serine. In particular the methylation of lysine 9 in histone 3 (H3K9me) appears to be a key event (Nakayama *et al*., 2001; Tuzon *et al*., 2004). As suggested above, these modifications are dynamic, and the exact degree of posttranslational modification of these organizing proteins regulates the transcriptional status of the associated chromatin. In general, histone methylation correlates with gene silencing, while acetylation tends to correlate with chromatin activation, though exceptions abound. While protein phosphorylation and dephosphorylation are certainly associated with progression through the cell cycle, the significance of histone modification in this manner remains unclear.

1.1.5. Extranuclear Transcriptionally Active Compartments: Mitochondria and Chloroplasts

Mitochondria are oblong-shaped, double-membrane-bound organelles found in the cytoplasm of all aerobic eukaryotic cells. Often referred to as the powerhouses of the cell, mitochondria play a key role in biochemical processes other than cellular respiration, one noteworthy example of which is apoptosis (Loeffler and Kroemer, 2000). Chloroplasts are round-to-oval organelles found only in green plants and photosynthetic protists. Like the mitochondria, chloroplasts are also double-membrane-bound. It is well established that both of these common eukaryotic organelles possess a circular, double-stranded DNA genome that is inherited in a non-Mendelian manner, and each individual organelle contains multiple copies of its genome. In order to avoid confusion with the nuclear genome, the mitochondrial chromosome is designated mtDNA and the chloroplast genome is designated cpDNA (and less frequently, ctDNA). Transcription, therefore, occurs in two compartments in animal cells, namely the nucleus and the mitochondrion, while in plants, a third transcriptionally active compartment is the chloroplast.

In plants, mtDNA can be as large as 570 kb, which is substantially larger than mammalian mtDNA (about 17 kb). Because the GC content of mtDNA is often substantially different from that of genomic DNA, it is possible to separate these two DNA types by isopycnic centrifugation in CsCl. Alternatively, the entire organelle can be isolated intact in a step gradient. The primary difference between mtDNA from plants, animals, and fungi is

that plant and fungal mitochondrial genomes contain numerous noncoding regions, while mitochondrial genomes in animals demonstrate an economy of sequences, nearly all of which appear to encode specific products. In general, the mitochondrial genes encode many of the enzymes associated with the oxidative reactions of cellular respiration, including pyruvate dehydrogenase, electron transport-associated enzymes, the enzymes needed to run the Kreb's cycle, and those required for the oxidation of fatty acids.

Genes encoded in mtDNA are found on both strands and, in addition to the protein-encoding genes, rRNA and tRNA genes are also found in mtDNA. The identification of genes in several mtDNA genomes was computer aided, in which mitochondrial sequences were searched for start codons and nonsense (stop) codons in the same reading frame. Such a region is known as an open reading frame (ORF), and searching for ORFs is a fundamental bioinformatics strategy. Gene products produced by expression of mtDNA remain in the mitochondria, though ancillary proteins such as ribosomal proteins are encoded in nuclear DNA and then imported into the mitochondria to complete the assembly of these organelles. Following transcription, pre-mRNA transcripts are subject to processing that includes shortening and the addition of the commonly observed poly(A) tail to mRNAs at the 3′ end; newly transcribed tRNAs are modified by the addition of the trinucleotide motif CAA to the 3′ terminus. In contrast to cytoplasmic mRNA, mitochondrial mRNAs are not capped at the 5′ end, have small or nonexistent leader sequences, and exhibit a start codon close to the 5′ end. With respect to protein synthesis, it appears that the nuclear genetic code governs mitochondrial translation in plants. In animals and other eukaryotes, for example, there are multiple noteworthy differences in the genetic code. The unique character of gene expression (both transcription and translation) in this eukaryotic organelle has been likened to gene expression in prokaryotes.

cpDNA is similar in organization and content to mtDNA. Chloroplast genomes often possess more than one hundred genes and range in size from 120 – 200 kb, depending on the species. This is many more genes than are ordinarily encoded in mtDNA, and the chloroplast chromosome is much larger. As is the case with mtDNA, the genes encoded in cpDNA are transcribed and then translated entirely within the organelle, though additional proteins need to be transported into the chloroplast from the cytoplasm to support chloroplast-localized transcription and translation. Curiously, there is a similar requirement for the importation of certain proteins needed to support the photosynthetic pathway. Like nuclear genes, chloroplast genes are frequently interrupted. Introns found in chloroplast tRNA genes are generally localized to the anticodon loop, while introns associated with protein-encoding genes resemble mitochondrial introns.

Through bioinformatics applications involving cpDNA more than 100 ORFs have been identified, most of which correlate nicely with known chloroplast proteins. The most noteworthy example of a chloroplast-encoded protein is the large subunit (LSU) of ribulose-1,5-biphosphate carboxylase/oxygenase

(RUBISCO), which is the first of several enzymes needed to support photosynthesis. Because RUBISCO constitutes at least half of all of the protein in green plant tissue, it is widely acclaimed as the most abundant protein in the world. The protein itself exhibits quarternary structure, consisting of eight small, nuclear-encoded polypeptides and eight larger, chloroplast-encoded polypeptides. Hence, RUBISCO is an excellent paradigm for cooperation and coordination of gene expression between two subcellular compartments.

1.1.6. Types of Nuclear Transcripts Produced

The major types of transcripts that are produced in a eukaryotic cell are rRNA, tRNA, and hnRNA. rRNA forms the scaffolding of the ribosome; ribosomes are non-membrane-bound organelles consisting of rRNA and approximately 80 ribosome-specific RNA binding proteins. tRNA molecules function much like shuttles in that they transport amino acids from the cytoplasm to the aminoacyl site of the ribosome in order to support translation. There are only four major types of rRNA and about 20 common tRNAs, and these do not change as a function of cell state. Other minor transcripts are confined to the nucleus and have a role in the mechanics of hnRNA splicing.

hnRNA, by comparison, is remarkably diverse, and it is this type of transcript that will mature into mRNA after a number of processing steps described below. Cells respond to external and internal stimuli by altering the complexity and membership of mRNA in the cytoplasm. Not only are specific genes up- and downregulated, but alternative processing of transcripts from a single genetic locus has been observed, as have changes in spatial and temporal patterns of gene expression.

1.1.6.1. Processing, Structure, and Stability of Nuclear Transcripts

The primary products of transcription require a fair degree of processing and maturation before they will become functional. This is true of the products of all three RNA polymerases. RNA polymerase I produces an initial large 45S rRNA, containing the 28S, 18S and 5.8S rRNAs; these are liberated from this common transcript following several splicing events. Subsequent interaction with rRNA binding proteins and the independently transcribed 5S rRNA will ultimately support the assembly of a functional ribosome. tRNA genes, when transcribed, render long RNA molecules that contain concatenated tRNAs that are released from the primary transcript via post-transcriptional splicing. The transcription products of RNA polymerases I and III are rather unremarkable, and do not generally change as a function of cell state. In sharp contrast, the products of RNA polymerase II activity, namely mRNA and its unspliced precursor hnRNA, are remarkably variegated. This diversity is not unexpected because mRNAs are translated into proteins.

The first modification of mRNA transcripts is the capping of hnRNA, an event which occurs as soon as the 5′ end of the nascent transcript moves away from the RNA polymerase. The cap consists of an unusual inverted 5′ – 5′ linkage between the first transcribed nucleotide and the 7-methylguanosine cap which is then added to it enzymatically. The function of the cap is two-fold in nature: not only does the cap protect the 5′ end of cytoplasmic mRNAs from nuclease degradation, it also serves as an identification signal by which ribosomes are able to distinguish mRNA from non-mRNA transcripts.

The second major type of hnRNA modification is the addition of a tract of adenosine residues, the so-called poly(A) tail, by the action of nuclear enzymes known as poly(A) polymerases; transcripts so-endowed are known collectively as poly(A)$^+$ RNA. Most cytoplasmic eukaryotic mRNAs display this structure, though there are some naturally occurring non-poly(A) mRNA species, including the very abundant histone mRNAs. The poly(A) tail is generally bound by poly(A) binding protein (PABP) and the length of the tail is usually indicative of how long it will remain stable in the cytoplasm.

The addition of the poly(A) tail is a precisely orchestrated series of events. The required splicing near the 3′ end of the hnRNA, as well as the efficiency of polyadenylation itself, are both dependent on the highly conserved hexanucleotide motif AAUAAA, i.e., the polyadenylation signal, and to a lesser extent upon a GU-rich region located downstream from the polyadenylation signal. The sequence AAUAAA is not a universal polyadenylation signal, however, since subtly altered versions such as AUUAAA and AAUAUA have been observed to direct the placement of a poly(A) tail (Jung *et al*., 1980; Ahmed *et al*., 1982; Tosi *et al*., 1981). Furthermore, many plant mRNAs do not have such a motif, and AAUAAA elements located in upstream regions of the mRNA are ignored by the plant polyadenylation machinery. The cleavage of hnRNA is mediated, at least in part, by small nuclear ribonucleoproteins (snRNPs), resulting in a new –OH group that functions as the substrate for poly(A) polymerase (Birnstiel *et al*., 1985; Berget and Robberson, 1986; Black and Steitz, 1986, Hashimoto and Steitz, 1986). See Chapter 4 for further details.

The third type of processing that is observed in the eukaryotic nucleus is the systematic removal of introns and the joining together of exons. Splicing is mediated by snRNAs and is associated with several protein complexes (discussed in Chapter 3). Because introns characteristically begin with a GU dinucleotide and end with an AG dinucleotide, one may predict the splicing pattern for each transcript. Further, eukaryotic mRNAs, particularly in animal cells, are typically monocistronic, meaning that each mRNA encodes a single polypeptide. This is in contrast to prokaryotic transcripts which are often polycistronic, usually encoding protein products related to a common metabolic pathway (*e.g*., the *lac* operon).

Nuclear RNA is transported through membranes from one compartment to another in eukaryotic cells (reviewed in Chapter 6). In order to facilitate efficient transmembrane relocation, transcripts form ribonucleoprotein

(RNP) particles by interaction with RNA binding proteins. Nuclear egress in this manner is probably related to the unfavorable thermodynamic transport of mRNA with a high negative charge density across the hydrophobic interior of nuclear membranes. Some RNA molecules are transported into the cytoplasm and then into the mitochondria; for example, some tRNAs that are not encoded in mtDNA but yet are required to support the synthesis of mitochondrial proteins fall into this category.

mRNA has also been shown to travel from one cell to another in plants via the phloem. This was originally demonstrated in tomato. Briefly, wild type tomato plants were grafted onto plants exhibiting the *Me* phenotype, a dominant mutation that affects leaf morphology. mRNA from the mutant rootstock was able to travel to the grafted wild type plants and alter their morphology (Kim *et al.*, 2001).

RNA stability in the cytoplasm has a profound impact on the levels of specific proteins in the cell, as well as homeostasis in general. Plant and animal mRNAs both exhibit at least two features (5′ cap and poly(A)$^+$ tail) that maximize the likelihood that mRNAs will survive long enough in the cytoplasm to have an opportunity to support the synthesis of a required protein. At the same time, all transcripts have a natural turnover rate: this ensures that an mRNA does not persist as an indefinite template for translation, since overproduction of a normal protein can lead to grave consequences for the cell and for the organism.

1.1.6.2. Monocot vs. Dicot mRNAs

The region of a transcript just downstream from the 5′ cap and prior to the initiation codon is known as the leader sequence or the 5′ untranslated region (UTR). The length and base composition of this mRNA region is variable from one gene to the next. In plants, there are significant differences between monocot and dicot mRNA leader sequences, including differences in GC content, length, and context of the start codon used to initiate translation. This is important because the features of the RNA that precede the initiation codon are known to affect the efficiency of translation (Pain, 1996). In general, mRNAs with longer leader sequences tend to be translated at higher rates than short-leader sequence mRNAs (Kozak, 1991b; Futterer and Hohn, 1996), though this is not a hard-and-fast rule. Dicot mRNA leader sequences also tend to be AU-rich (Joshi, 1987), and many range from 10–120 nt. In contrast the leader sequences associated with monocot mRNAs are C-rich, ranging from about 40–120 nt long.

1.2. Translation of Nuclear Transcripts

Translational control of gene expression is a fundamental regulatory process in the cell. All cell function is ultimately governed by the types of proteins that are synthesized and the abundance of each; therefore failure to produce a

protein when needed or the overproduction of protein are events likely to alter the physiology of the cell. The mechanics of translation in the cytoplasm are detailed in Section 1.2.1. in this chapter. The abundance of a transcript, its half-life in the cytoplasm, and the degree to which secondary structures form within the transcript are all important aspects of the regulation of gene expression at the translational level. It is also known that a particular species of mRNA can be transiently prevented from associating with the translation apparatus in response to certain external or internal cues. This is one means of suppressing the production of a biochemically significant protein.

Proteins are synthesized in the cytoplasm either by membrane-bound ribosomes, in which the nascent peptide is directed into the endomembrane system, or by free ribosomes (non-membrane-bound). The latter usually results in protein localization to another subcellular compartment, such as the nucleus, chloroplasts, or mitochondria; this accounts for the majority of the protein found in these organelles. Entry into a chloroplast or a mitochondrion is mediated by sequences at the amino terminus of the protein. The "entry code" for mitochondrial-targeted proteins is 12 or 13 amino acids that form an amphipathic helix, while the transit sequences of chloroplast-targeted proteins contain 25 or more amino acids and have a net positive charge. These N-terminal sequences appear to be recognized by organelle-bound surface receptors.

Translation can be divided into three stages. Stage 1 is initiation and involves the formation of an initiation complex at the mRNA 5′ leader (see Section 1.2.1., below). The 40S ribosomal subunit is primarily responsible for recruiting the mRNA and associated initiation factors. Recent evidence indicates that poly(A) binding protein enhances initiation of translation by binding a component of eIF4F, a multi-protein complex required for association of the 43S pre-initiation complex, thus forming a bridge between the initiation factors and the mRNA cap (Tarun and Sachs, 1996; Kahvejian *et al.*, 2005). Following successful initiation, the ribosomes begin and continue elongation of the polypeptide until a stop codon (UGA, UAA, UAG) is encountered. The 60S ribosomal subunit directs the elongation process and subsequent termination of translation. At this point, class 1 and class 2 release factors are required to dissociate the tRNA from the carboxy terminal amino acid of the nascent peptide and then, in turn, the ribosome from the class 1 release factor, respectively. These events occur while the last charged tRNA and the nascent protein are within the peptidyl (P) site of the ribosome. Termination of translation is another point where gene expression can be affected, as for example, in nonsense-mediated decay (discussed in Chapter 6). Figure 1.2. provides a summary of eukaryotic translation.

1.2.1. mRNA Sequence and Structure Affect Translation

Translation can be envisioned as a collection of dissimilar mechanisms acting concertedly to initiate and sustain protein synthesis. The prevailing understanding of the mechanics of the initiation of translation is widely

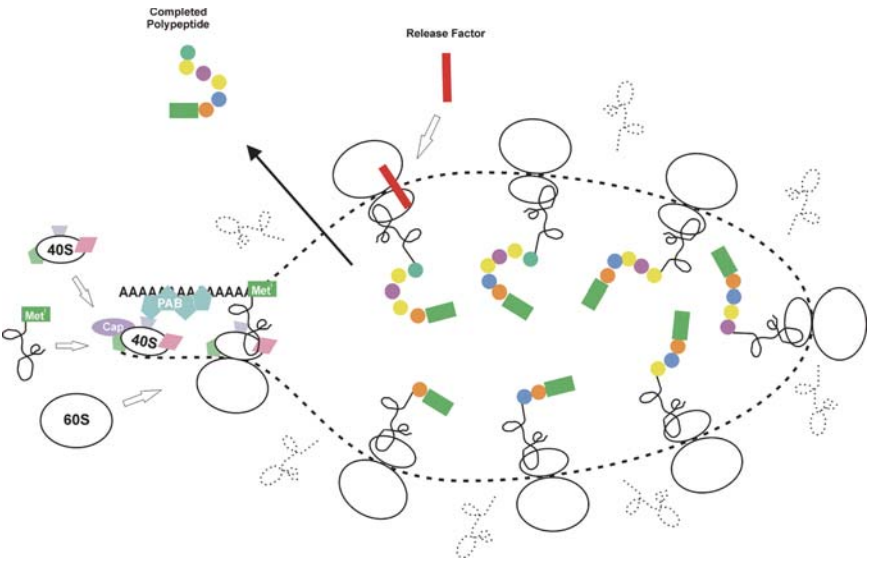


FIGURE 1.2. Summary of Eukaryotic Translation. The small ribosomal subunit (40S) associates with initiation factors (shown as colored shapes) and binds to the mRNA (dotted line) via the cap (red circle). This initiation complex then recruits the 60S ribosomal subunit to begin translation. A bridge formed between eIF4G and the poly (A) binding protein (PAB, purple pentagons) is thought to facilitate translation initiation. Various elongation factors modulate the rate of elongation of the nascent polypeptide. Different charged tRNAs (shown with the charged amino acid as a colored or gray rectangle) alternately bind the A (acceptor) and P (peptidyl) sites to position the correct amino acid specified by the triplet codon and to form the covalent peptide linkage. When the stop codon is encountered, release factors dissociate the completed polypeptide from the ribosomes, and the ribosomal subunits and accessory factors are recycled. (*See Color Plates*)

known as the scanning model (Kozak, 1978; 1991b). According to this model, the first event associated with the process of translation in eukaryotic cells is the 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 initiating AUG codon is encountered. Migration in this manner is instrumental for melting any secondary structure associated with the transcript. As a general rule, the first AUG codon encountered will function as the initiation codon, thereby directing the start of protein synthesis. This is true for more than 90% of the known eukaryotic mRNAs. It is abundantly clear, however, that being the first

AUG codon, meaning proximity to the 5′ end of the transcript, is not sufficient justification to function as the initiation codon.

1.2.1.1. Importance of AUG Codons in "Good Context"

The context of an AUG codon is fundamental to translation initiation control. "Good context", which is the key to successful translation, refers to the most favorable sequence combination(s) of the nucleotides surrounding the AUG. In vertebrate mRNA, the critical locations are adenine three bases upstream (–3) of the AUG and guanine four bases downstream (+4), keeping in mind that the A in AUG is considered nt +1. While other bases in the vicinity may have a secondary role in determining whether an AUG triplet will support translation initiation, they are of significantly less importance. 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".

The consensus "good context" sequence that has emerged associated with higher animal genes is GCCGCCPuCC*AUG*G, and the importance of this sequence has been demonstrated by site-directed mutagenesis (Kozak, 1986; Kozak, 1987a, b). Higher plants, in contrast show a slightly different context consensus sequence of caA(A/C)a*AUG*GCg; the sequences for dicots and monocots are aaA(A/C)a*AUG*GCu and c(a/c)Pu(A/C)c*AUG*GCG, respectively (Joshi *et al*., 1997).

1.2.1.2. Non-Canonical Initiation Codons

AUG is overwhelmingly the most commonly used translation initiation codon in eukaryotes, though other non-AUG codons have been demonstrated to function in this capacity, particularly in animal and viral mRNA (Hann, 1994; Huang et al., 1993). Sometimes, however, non-AUG initiation can result in a modified protein, such as murine FGF-3 (Acland *et al*., 1990) and human FGF-2 (Bugler *et al*., 1991), the altered properties of which can then direct the translation products to different subcellular locations. What is especially interesting is that these non-AUG start codons differ in sequence from AUG by a single base, though not always in the same position. Noteworthy examples include GUG, ACG, and CUG. In spite of the codon modification, a charged tRNA carrying methionine is still able to base-pair with this codon, albeit weakly, thereby positioning methionine at the amino terminus of the nascent polypeptide. How this is tolerated can only be explained by the "good context" of these non-AUG codons (Kozak, 1989a, b; Kozak, 1997).

AGAMOUS (AG) is another example of the use of a non-AUG initiator codon (Reichmann et al., 1999). AG is a floral homeotic gene in *Arabidopsis*. Many homeotic genes, including AG, are regulated at the transcriptional level, in this case by the negative-acting regulator APETALA2 (AP2) and the positive-acting regulator LEAFY (Busch *et al*., 1999). Translation of AG mRNA has been shown to initiate at an ACG codon (Riechman *et al*., 1999)

and is believed to do so because the codon ($ACG_{100–102}$) is situated in good context. Further sequence analysis of the transcript also revealed the potential for secondary structure formation that could enhance the interaction of the 40S ribosomal unit with upstream sequences. These data demonstrated that non-AUG codons are quite capable of functioning as the only translation initiation codon for certain plant mRNAs.

1.2.1.3. Upstream ORFs and the Ribosome Scanning Hypothesis

The AUG codon closest to the 5′ end usually functions as the initiation codon. Initiation at downstream AUG codons is favored, however, when any of the following occur (Kozak, 1989a): (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 an upstream AUG. Condition (3) presupposes that the ribosome has already encountered a stop codon in the same reading frame as the first AUG and upstream of the second AUG.

One of the most important developments in understanding the control of protein synthesis was the discovery of multiple AUG codons in the 5′ leader sequence of many mRNA species. These so-called upstream AUGs (uAUGs) along with often-associated upstream 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 proliferation- and differentiation-associated genes, including several oncogenes (Kozak, 1987a; 1991a; Morris, 1995). The positional definition of a uAUG is not always clear, however, given the fact that transcripts are often subjected to alternative splicing in response to experimental challenge, the resulting altered TSSs may change the ranking of AUGs found within the transcript. It appears that the parameters governing uORF translation initiation apply equally to ORF translation downstream.

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; Damiani and Wessler, 1993; Wang and Wessler, 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 a role in the regulation of gene expression by destabilizing the transcript (Ruiz-Echevarria *et al*., 1996; Ruiz-Echevarria and Peltz, 2000; Wiese *et al*., 2005). Other examples of gene regulation by uORFs include the mammalian gene *AdoMetDC*, a key enzyme in the biosynthesis of spermidine and spermine and, in N*eurospora*, 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). Sucrose sensing in *Arabidopsis* appears to be controlled by expression of a uORF which is required to translationally repress bZIP-type transcription factors (Wiese *et al*., 2005).

1.2.2. Non-Canonical Initiation of Translation

As mentioned in Section 1.2.1.2., the use of non-AUG translation initiation codons, although somewhat unusual among higher eukaryotes, can influence gene expression. Another form of non-canonical translation initiation involves ribosome binding at internal sites on the mRNA. Many eukaryotic viruses, both plant and animal, synthesize polycistronic mRNAs, and given the ribosomal scanning hypothesis, it would seem to be an inefficient process to ensure the survival of these entities. However, many of these viruses have evolved a mechanism for getting around the restrictions in eukaryotic translation initiation by using sequence-specific sites within their mRNAs to begin translating internal exons. These sites, called IRESs (Internal Ribosome Entry Sites) allow the virus to translate virtually all of the proteins encoded in the polycistronic mRNA using the host's 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 2003).

1.2.3. Role of Secondary mRNA Structure on Translational Control

Translation in the cell results when a ribosome successfully engages a transcript, finds an initiation codon in good context, and then travels unencumbered until it reaches a stop codon. The variety and abundance of mRNAs in the cell is impressive, though mRNA abundance does not always show a direct correlation with protein abundance. Transcripts are competing for a limited number of ribosomes, and it is clear that some mRNAs are translated with greater efficiency than others.

In instances where an mRNA forms secondary structure(s) involving the 5′ leader sequence, the 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 (Pain, 1996; Kozak, 1991b; Dinesh-Kumar and Miller, 1993; Futterer and Hohn, 1996; for review, see Kozak, 1999). 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 efficiently suppress translation, a –19 kcal/mol hairpin 14 nts downstream of an AUG codon can enhance the initiation of translation (Kozak, 1990). One might well speculate that the downstream hairpin causes strategic pausing of the ribosomal complex directly over AUG, thereby enhancing translation initiation.

One of the better characterized paradigms of translational repression by mRNA secondary structure is the maize (*Zea mays) Lc* gene (Wang and Wessler, 2001), the product of which is involved in the anthocyanin biosynthetic pathway. A potential RNA hairpin within the 5′ leader sequence of the *Lc* transcript has been shown to repress translation *in vitro*. Very convincing experiments have demonstrated that mutation and deletions that reduce the stability of the hairpin result in an increase in the amount of protein encoded downstream (Wang and Wessler, 2001), and proximity of the hairpin to the 5′ end of a transcript may impact the ability of ribosomes to load onto the mRNA.

Perhaps more intriguing are the recent reports that the 3′ UTR which, by definition, is found downstream from the coding region of a transcript, may have a profound role in the regulation of the initiation of translation. It is widely known that most, if not all, eukaryotic mRNAs assume a circular or closed-loop configuration which brings the 3′ end of the transcript (and any proteins bound to it) into the proximity of the 5′ cap and the 5′-UTR. This phenomenon is observed in eukaryotic cells and eukaryotic viruses (Tarun and Sachs, 1996; Wells, *et al*., 1998; Mazumder *et al*., 2003, Allen *et al*., 1999). It now appears that the association of proteins and/or the formation of secondary structures within the 3′-UTR is able to influence the efficiency of translation. This mechanism of translation initiation control appears to be widespread in animals (Allard *et al*., 2005; Wax *et al*., 2005), plants (Browning, 1996), and among plant viruses, such as Barley yellow dwarf virus (Guo *et al*., 2001).

1.3. Mainstream Molecular Techniques to Study RNA as a Parameter of Gene Expression

That transcript levels change in cells in response to internal signals and external cues, including biotic and abiotic stress, is undisputed. One must realize, however, that an observed modulation of transcript abundance may be the result of an increase in the rate of transcription from a particular locus (hnRNA synthesis) or a change in the rate of the subsequent processing, nucleocytoplasmic transport, or stability of a transcript, i.e. mRNA, following the completion of its synthesis. Unfortunately, it is all too commonplace to find reports in which total cellular RNA (nuclear RNA mixed with cytoplasmic RNA) is extracted from the cell for assay: analysis of total RNA does not allow one to distinguish the individual contributions made by nuclear

and cytoplasmic RNA species. To do so requires that cells be disrupted gently so that RNA can be isolated from the cytoplasm independently of the RNA found in the nucleus; this approach is particularly useful for determining if a gene is being regulated transcriptionally or at any one of several posttranscriptional levels. While the clever design of primers that span an intron may provide somewhat clearer information about the level at which gene regulation is occurring, it is not a completely accurate or unambiguous approach, and requires knowledge of the sequence being investigated. It is also recognized that this type of analysis may not be possible in all experimental settings, and the suggestion of examining various subpopulations of RNA is proffered only to place the interpretation of data from various forms of transcriptional assays in "good context".

1.3.1. Non-PCR Methods: Northern Analysis, Nuclease Protection, and Nuclear Runoff Assay

Among the traditional methods for transcript assay are the Northern analysis (Alwine *et al*., 1977, 1979) and the S1- and ribonuclease protection assays (RPA) (Berk and Sharp, 1977; Casey and Davidson, 1977; Vogt, 1980; Melton *et al*., 1984; Quarless and Heinrich, 1986). In each of these methods, target RNA from one or more samples is hybridized to a labeled DNA or RNA probe. At the end of the hybridization the goals are to localize and quantify the probe. These methods measure steady-state RNA, i.e. the final accumulation of transcripts in the cell, with no regard for the rate at which these transcripts were synthesized or undergoing degradation. Rate issues can be addressed by performing the somewhat cumbersome nuclear runoff assay.

In the case of Northern analysis, RNA is electrophoresed under denaturing conditions, most often in the presence of formaldehyde and formamide, such that all like-species of RNA comigrate. The electrophoresed RNA is then transferred to a solid support such as a nylon filter to which it is permanently crosslinked by UV irradiation. The filter paper is incubated under moderate-to-high stringency conditions in the presence of a nucleic acid probe, during which time the probe will hybridize to any complementary RNA molecules on the nylon filter, and will do so in proportion to the abundance of that particular transcript. Finally, a piece of x-ray film placed in proximity to the filter is used to record either radioactive decay in the case of isotopically-labelled probes or the emission of visible light in the case of detection by chemiluminescence. Northern analysis is an example of mixed-phased hybridization because one of the two strands participating in hybrid formation is fixed in place on the surface of the filter (the RNA target) while the other strand (the probe) is free-floating. While few would dispute the reputation that Northern analysis has for lack of sensitivity, the key advantage of using this technique is that one is able to determine the full length size of the transcript because the RNA is electrophoresed following isolation from the biological source and without any prior manipulation. In contrast, PCR

typically amplifies only a section of a (cDNA) template i.e. that which is framed by the two primers.

The S1 assay and the RPA were developed in order to address the serious lack of sensitivity associated with Northern analysis. Succinctly, sample RNA is mixed directly with and hybridized to a radiolabelled probe under high-stringency conditions. In this approach, both probe and target are free-floating (solution hybridization), resulting in complete saturation of all target molecules. In contrast, the requisite act of crosslinking target transcripts onto a filter paper for Northern analysis renders some of those transcript molecules incapable of hybridization, thereby greatly compromising sensitivity. Following solution hybridization, S1 nuclease or a ribonuclease cocktail is used to degrade any non-hybridized excess probe as well as any nontarget transcripts that did not participate in hybridization. These assays are termed "protection assays" because any hybrid molecules that form between target (mRNA or cDNA) and probe are resistant or "protected" from post-hybridization nuclease degradation. The result of the enzymatic digestion of all remaining single-stranded molecules is a collection of double-stranded hybrid molecules. In addition, the 5′ and 3′ ends of the hybridized target and probe molecules will also be removed by the nuclease, thereby trimming the protected fragment down to the size of the probe or, often, a bit less. The protected fragment(s) is then electrophoresed on a polyacrylamide gel which is subsequently dried down so that autoradiography can be performed. As in the case of x-ray film exposure in the Northern analysis, the signal recorded on the film is directly proportional to the amount of a particular mRNA target within a sample. Compared to Northern analysis, however, one may expect an enormous boost in sensitivity because there is no filter paper or fixation of molecules involved.

A third approach for the assay of gene expression, albeit a method far less popular than Northern analysis or nuclease protection, is the nuclear runoff assay. This method involves the isolation of intact nuclei from cells grown in culture or from tissue samples followed by radiolabelling nascent transcripts. Labeling is accomplished by incubating the nuclei in a ribonucleotide cocktail that includes ^{32}P-labelled UTP. The incorporation of radiolabel occurs in proportion to the rate at which transcription is occurring. This technique has several advantages: (1) the natural geometry of the chromatin is preserved, (2) the mechanics of the assay do not support the initiation of new transcripts, but only the elongation of transcripts that were initiated prior to cellular disruption, and (3) because all transcripts will incorporate radiolabel, one may assay the transcriptional behavior of multiple genes simultaneously. The nuclear runoff assay, while time-consuming and technically "messy", provides information about the relative rate of transcription. Nuclei from several biological sources can be prepared in advance and stored at -80° for up to one year, so as to minimize the number of times that the assay needs to be run.

1.3.2. PCR-Based Methods: 5′ RACE (Rapid Amplification of cDNA Ends)

The polymerase chain reaction[3] (PCR) offers unparalleled sensitivity in the assay of gene expression, as well as in other applications that require the amplification of small starting masses of either genomic DNA or cDNA. While PCR is not without its shortcomings, the methodology constitutes a rapid enzymatic amplification of DNA sequences and does so with high fidelity. With minimal target sequence information available, one may mass produce exact or nearly exact copies of starting material in as few as two to three hours. Many permutations of PCR have been developed to answer very specific questions. Among the more interesting, contemporary questions with respect to gene expression is the possibility of identifying alternative transcription start sites coming from a single locus.

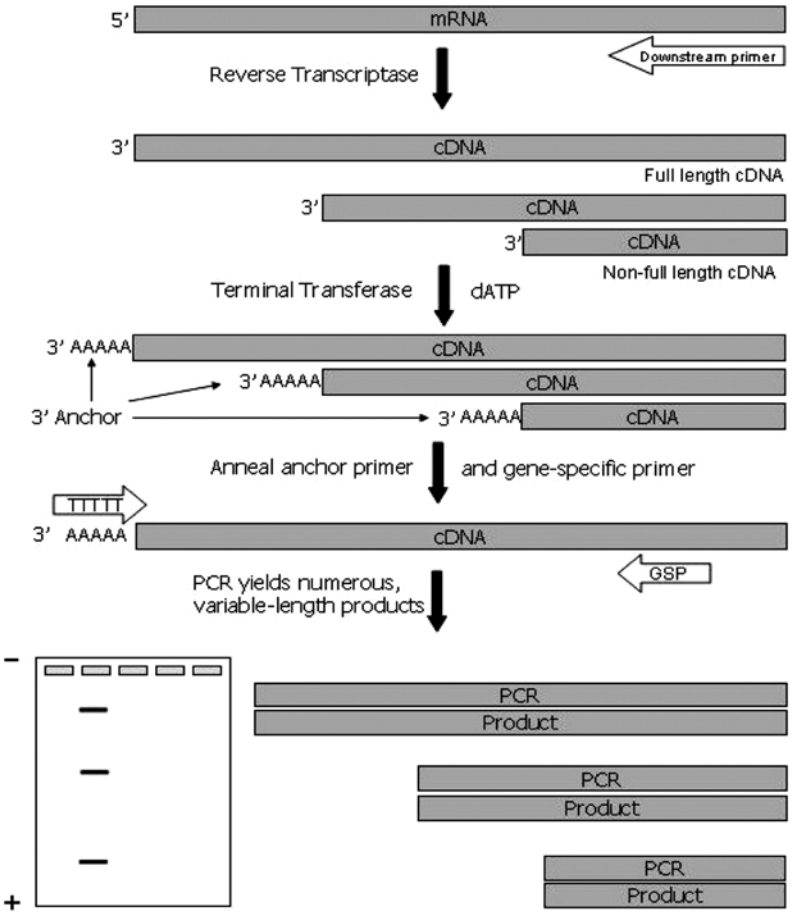
A method known as 5′ RACE (rapid amplification of cDNA ends) was originally developed to facilitate the cloning of sequences found close to the 5′-most end of a transcript by first converting the transcript to cDNA and then performing a typical PCR amplification. Originally known as anchor PCR and one-sided PCR, 5′ RACE has been expanded to identify precisely the first transcribed nucleotide from a locus and concomitantly identify transcripts that result from one or alternative transcription start sites.

1.3.2.1. Classical Race

The original approach to performing RACE was a glorified form of a very old technique in molecular biology known as homopolymeric tailing. In this approach, mRNA is converted into first-strand cDNA using any of a variety of reverse transcriptases and priming strategies. The 3′ end of the newly synthesized, single-stranded cDNA provides the required -OH group for another enzyme known as terminal deoxynucleotidyl transferase, or simply "terminal transferase". This enzyme is an unusual polymerase in that it is able to add nucleotides in a non-template-dependent manner to protruding 3′ ends of both single- and double-stranded cDNA. Thus, in the presence of a single dNTP, for example dATP, a small tract of adenine-containing nucleotides, four to eight bases long, will be added to the template. This newly added tract now functions as an anchor, to which an anchor primer can be designed, *e.g.* poly dT, and which can support PCR when used in conjunction with a gene specific primer (Figure 1.3).

---

[3] The PCR process is covered by patents currently owned by Roche Molecular Systems, Inc. and F. Hoffmann-La Roche Ltd. Use of PCR requires a limited license under their patents. PCR must also be performed in a licensed thermal cycler. For specific details, contact the Director of Licensing, Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404.



There are significant disadvantages associated with this traditional form of RACE and, in the context of a discussion of alternative transcription start sites, the most severe problem is the fact that *any* first-strand cDNA, full length or not, is a candidate for homopolymeric tailing, resulting in a mixed population of tailed cDNAs as a consequence of a non-robust reverse transcriptase reaction. This has the unhappy consequence of obscuring the presence of any cDNAs that do, in fact, reflect an alternative TSS, as they may be mixed among

a number of false positives. While rationale primer design and improvements in commercially available reverse transcriptase formulations lessened some potential sources of error, the entire approach remained flawed because both the anchor and the anchor primer were either 100% G+C or 100% A+T. This creates a severe thermodynamic imbalance between the two primers needed to support the PCR component of this technique because the T_m of the gene-specific primer would differ greatly from the T_m of the anchor primer.

1.3.2.2. RLM-Race

The fundamental limitations of classical 5′ RACE led to the development of a completely new approach known as RNA ligase-mediated RACE, or simply RLM-RACE. As with classical RACE, the goal of RLM-RACE is to map precisely the 5′ end of a transcript and detect multiple transcription start sites, if present. Elimination of the homopolymeric tailing and inclusion of a mid T_m-range adapter (the anchor) have dramatically improved the precision of this assay. The entire process, described below, is outlined in Figure 1.4.

Prior to reverse transcription, the RNA sample is incubated in the presence of calf intestinal phosphatase (CIP) which will remove the 5′ phosphate from any un-capped transcripts, non-full-length transcripts, and non-mRNAs. This is important because mRNAs that have been uncapped or experienced any degree of degradation at or near the 5′ end could be included in the final PCR amplification. Therefore this first step ensures that only primary, full-length mRNA transcripts will be supported in the subsequent steps. Next the RNA sample is incubated with the enzyme, tobacco acid pyrophosphatase (TAP), which has the unusual ability to remove the two 5′-most phosphates that constitute the 5′-cap found on the vast majority of eukaryotic mRNAs. This renders uncapped, full-length mRNAs with a single exposed 5′-phosphate group. At this point the RNA sample consists of a mixed population of RNAs with and without 5′ phosphate groups. In the next step, an RNA oligonucleotide adapter is joined only to full-length mRNAs by the action of T4 RNA ligase because (1) all non-mRNA molecules and non-full-length mRNAs are missing a 5′ phosphate required for ligation, and (2) previously capped, full-length mRNAs have been uncapped, leaving a 5′ phosphate substrate that is required to complete the ligation to the RNA adapter: the mRNA provides the 5′ phosphate and the RNA adapter provides the 3′-OH substrate. Finally, the RNA is converted into cDNA using a standard reverse transcriptase reaction, and cDNAs corresponding to the locus of interest are then amplified by standard PCR methods. The adapter primer can only anneal to complementary adapter anchor sequences, which were added exclusively to full-length RNA. The resulting PCR products mirror the transcription start site(s), and the gene-specific downstream primer can be moved closer to or farther away from the adapter primer to not only change the expected size of the product but also to test for the functionality of different promoter elements (Bassett *et al.*, 2000, 2004).

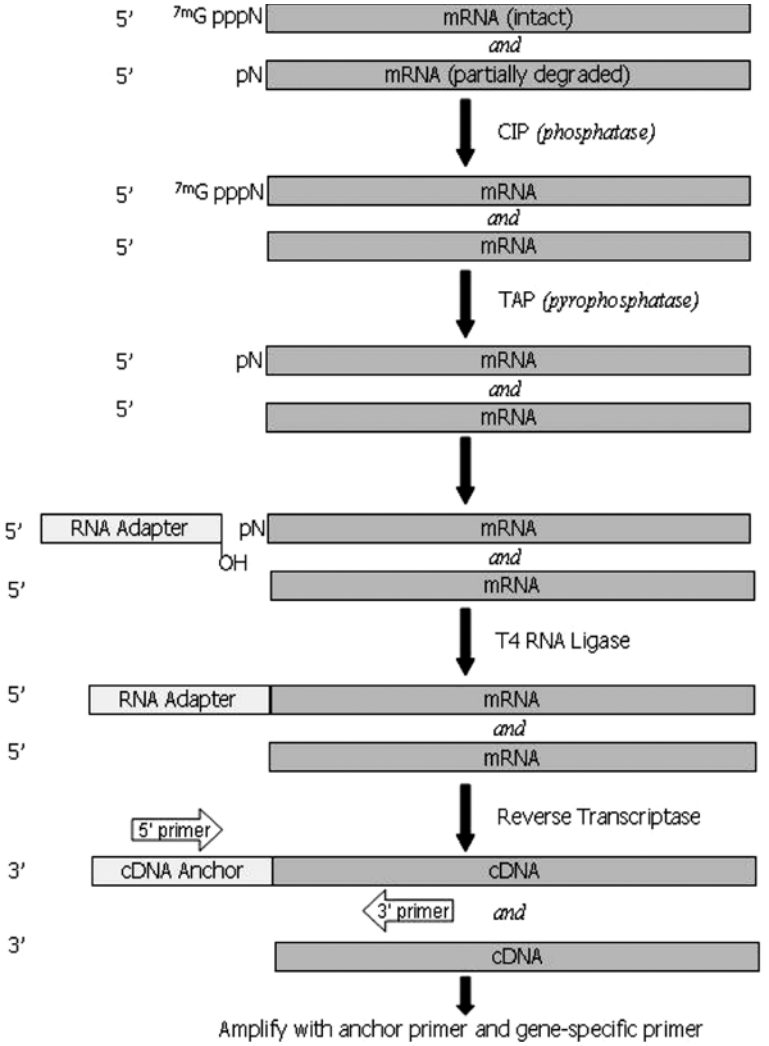


FIGURE 1.4. RLM-RACE. The ligation of an RNA adapter to full-length, uncapped mRNA ensures the synthesis of cDNA molecules that represent the 5′ end of the transcript. This approach is an accurate, highly reproducible method for mapping one or more transcription start sites associated with a particular species of mRNA. From *RNA Methodologies* (2005), with permission of Elsevier Academic Press.

1.3.3. In Silico Tools

The marriage of information technology and biotechnology, made possible by high speed computers with remarkably fast processors and impressive memory capacity, is widely known as bioinformatics. The increase in DNA

and protein sequencing automation is responsible for the vast amount of new information that is added daily to databases world-wide. Because of the enormity of information that is already in these databases, software tools have been developed to facilitate data-mining, and these "search and rescue" tools undergo continual refinements. In some laboratories, scientists with expertise in this area have written their own software to perform laboratory-specific analyses.

The place to begin the assessment of or to search for sequence data is the National Center for Biotechnology Information (NCBI; Bethesda, Maryland) web site, located at www.ncbi.nlm.nih.gov. Many web sites that are related to bio-informatics have direct links to the NCBI web site (Table 1.2.). Once there, the user may select from any of a number of tools for performing the desired analysis. Two other popular sites, the European Molecular Biology Laboratory (EMBL; Heidelberg, Germany), and the DNA Data Bank of Japan (DDBJ; Mishima, Japan) coordinate their efforts with NCBI so that investigators can have access to these massive quantities of data 24 hours per day from virtually any computer terminal in the world that has internet access. Other specialized databases are also readily available by accessing The Institute for Genome Research (TIGR) or Swiss-Prot, which is an annotated protein sequence database.

Sequencing databases are huge, and various computer programs have been written that simplify the tasks of searching for and identifying DNA or protein sequences of interest. The best known of these is Basic Local Alignment Search Tool software, more commonly known as BLAST®. This versatile software tool, which can be accessed at the NCBI web site, searches databases for nucleic acid or amino acid sequences that match the input sequence, and the output shows the sequence alignment. Another favorite software tool is GRAIL (http://compbio.ornl.gov/public/tools), an analysis tool that identifies entire genes, exons, and other characteristics of input DNA.

The most straightforward method for performing a search is to select "nucleotide-nucleotide BLAST" (blastn). Users are asked to enter the desired

TABLE 1.2. Select bioinformatics web sites.

Formal name	Common name	URL
National Center for Biotechnology Information	NCBI	www.ncbi.nlm.nih.gov
The Institute for Genomic Research	TIGR	www.tigr.org
The *Arabidopsis* Genome Resource	TAIR	www.arabidopsis.org
European Molecular Biology Laboratory	EMBL	www.embl.org
DNA Data Bank of Japan	DDBJ	www.ddbj.nig.ac.jp
Swiss-Prot Protein Knowledgebase	Swiss-Prot	http://us.expasy.org/sprot
The Proteome Society	Proteome Society	www.proteome.org
Human Proteome Organization	HUPO	www.hupo.org
Codon Usage Database	Codon Tables	www.kazusa.or.jp/codon
Microarray Gene Expression Data Society	MGED	www.mged.org

search string (the nucleotide sequence being investigated; lower case letters only) which can be done manually or by cutting and pasting. The user then has the option of searching the entire unknown sequence or portions thereof. It is best for new users to perform a BLAST search using the nonredundant database "nr" which will perform the most comprehensive search possible. Depending on a number of variables, the search may require from several seconds to hours for completion. The results of each search will include two scores assigned to each sequence: first the *S score* (bit score), which is a statistical value associated with the number of matches retrieved from the data base, and then the second score, the *E value* (expect value), which is the number of matches with the same S score that one would expect to find by chance. Thus, for any BLAST search, the higher the S score and the lower the E value (ideally much lower than 1.0), the better the match.

1.3.4. In Vitro Translation and Western Analysis

As described above, regulation of gene expression may occur at the translational level and/or at the posttranslational level. In the case of the former, the extent to which biologically competent and chemically stable mRNAs become engaged by the protein translation apparatus determines the extent of synthesis of the primary amino acid sequence. In the case of the latter, phenotype alterations at both the cell and organism levels result from changes in the abundance of functional proteins, meaning proteins that have experienced a set of appropriate posttranslational modifications that occur after the protein has been released from the ribosome and which modulate its activity. It is well known that alterations in the cellular biochemistry or in the extracellular milieu can have a profound effect on the functionality of cytoplasmic transcripts. In the context of the identification and characterization of alternative transcripts and alternative transcription start sites, it is ultimately of great importance to determine whether the transcripts under investigation can direct the synthesis of a functional protein.

The methods used for the study of protein synthesis *in vitro* fall into one of two major categories. First, mRNAs can be used to direct the synthesis of an encoded protein in a cell-free system, such as the wheat germ extract or rabbit reticulocyte systems. Second, transcription and translation can be coupled by cloning cDNA or genomic DNA into a specialized vector which is then used to direct the synthesis of the encoded transcript first and then the corresponding protein. A third method for studying translation *in vitro* is the far less frequently used and far more technically challenging method of mRNA microinjection. This is a highly specialized technique which, for most investigations, would not be expected to render any data that could not be gleaned using either of the first two approaches.

With respect to protein synthesis in the cell, a commonly used method for measuring levels of a specific protein is Western analysis. In this approach, proteins are isolated from the biological source and then electrophoresed on

a polyacrylamide gel. At the conclusion of the electrophoresis period, the gel can either be stained directly with Coomassie blue or silver-stained, or the protein can be transferred from the gel onto a filter membrane. The transfer process, known as electroblotting, uses an electrical current to drive the protein from the gel and onto an adjacent sheet of nylon, nitrocellulose, or polyvinylidine difluoride (PVDF). This is analogous to the blotting of RNA in Northern analysis and the blotting of DNA in Southern analysis (Southern, 1975) though nucleic transfer ordinarily does not require an electrical current. The entire sample, now on the surface of the filter, can then be stained with Ponceau S and a duplicate filter probed with an antibody to assay the abundance and size of a particular protein. This method is also known as immunoblotting.

Another, infrequently used method that can provide clearer definition to the concept of mRNA translatability is known as polysome purification. This method determines the extent to which the mRNAs have engaged the translation apparatus. In this approach, one collects the polysome fraction from cells in order to look for the presence of a particular mRNA, indicating that translation of the message is taking place.

1.3.5. Implications for Proteomics

The fluid nature of gene expression is more clearly understood by delving deeply into the nature of the various products of transcription and, ultimately, of translation. The transcriptome is the full set of mRNA molecules produced by a particular cell under a particular set of circumstances. The proteome, in turn, is the full complement of proteins produced by a cell. It is important to note that these are transitive definitions, meaning that as internal and external stimuli occur, both the transcriptome and the proteome are subject to change.

The most remarkable attribute of the proteome is its size and complexity relative to the DNA and RNA sequences that actually encode the proteins. As it turns out, each of the various transcripts can produce a number of protein isoforms. Thus, the 35,000 or so genes in humans may encode 500,000 to 1,000,000 different proteins. Part of the reason for the remarkable diversity of proteins is the extensive nature of posttranslational modifications that occur as newly synthesized polypeptides move through the Golgi apparatus and other components of the endomembrane system.

Proteomics has evolved into a very important discipline for discerning gene function. In contrast to the older methods for the analysis of proteins one at a time, some of the newer proteomics technologies efficiently facilitate rapid, high-through-put studies in a cost-effective manner. Protein informatics is a formidable subdivision of bioinformatics that examines global patterns of protein expression at the cell and tissue levels.

Perhaps the premier tool for examining the proteome is two-dimensional (2-D) gel electrophoresis. In this method proteins are first separated by

isoelectric point (pI), and then by molecular weight in a direction perpendicular to the first dimensional charge separation. The advantage of this method is that it permits the separation of thousands of proteins simultaneously; each protein appears as a unique spot upon visualization. Specialized image informatics software can automate the identification and quantification of unique spots on the gel and provide an electronic means for archiving the image.

Following identification of "spots of interest", unique proteins are recovered from the gel and broken into significantly smaller peptides by the action of proteolytic enzymes. The peptides are identified by mass spectrometry, most often matrix-assisted laser desorption ionization-time of flight (MALDI-TOF, or simply MALDI) analysis, a method that was developed in the 1980s. DNA encodes any and all possible proteins, and so MALDI-derived amino acid sequence information can be used to match the protein with the nucleotide sequence associated with a specific genetic locus. Other procedures are necessary, however, to elucidate function, such as through the use of the yeast two-hybrid system, immunoaffinity capture, nuclear magnetic resonance (NMR), crystallization, capillary electrophoresis, and high performance liquid chromatography (HPLC). Other more traditional systems for proteomics investigations are various cell-free transcription-translation systems, including wheat germ extract, rabbit reticulocyte lysate, canine pancreatic microsomal membranes, and proteases, alluded to in Section 1.3.4.

A state-of-the-art approach to proteome characterization is innovative protein microchip technology, also known as protein microarrays, which are used to detect protein-protein and protein-lipid interactions. Many protein arrays are "themed", allowing the investigator to study specific infectious diseases, cancer, or apoptosis. Peptides, like DNA and RNA oligonucleotides, can also be made-to-order using amino acids with various posttranslational modifications. Customized peptides can be used for epitope mapping in immunology, identification of active peptides for pharmaceutical applications, identification of substrates for and inhibitors of various enzymes, development of novel antibodies for vaccine development, and the characterization of protein-protein and protein-nucleic acid interactions. Posttranslational modifications of proteins and methods for the study thereof have been recently highlighted (Keen and Ashcroft, 1999; Liebler, 2002; Kannicht, 2002; Hamdan and Righetti, 2005). The number of proteomics databases is increasing rapidly, some of which have been included in Table 1.2. The professional organizations related to proteomics are very good starting points for novices in the discipline.

1.4. Summary

The regulation of gene expression in response to internal and environmental cues, occurs at many levels in the plant and animal cell. Myriad transcription factors, particularly those which function *in trans*, control the efficiency with

which the transcription apparatus is assembled and the rate at which transcripts are produced from a genetic locus. While transcription is limited to membrane bound organelles in eukaryotic cells, posttranscriptional processing, transcript export out of the organelle in which synthesis occurred, and the overall efficiency of translation and posttranslational processing are equally important for controlling the amount of functional protein that a cell can produce.

Various classical and novel methods are in use for the study of gene expression at the RNA level and the protein level. The presence of a certain abundance level of transcript implies that a proportionate level of protein may likewise be found in the same cell, though this is not always the case. Northern analysis, quantitative PCR approaches, and various protein assays are reliable tools for the assay of RNA and protein expression, and bioinformatics-based problem solving methods have an increasingly important role in a more holistic portrayal of the cellular biochemistry under a defined set of conditions.

As newer instrumentation is developed that provides previously unimagined levels of sensitivity, more gene sequences and proteins resulting from novel posttranslational modifications will be catalogued in public databases such as GenBank®. Likewise, technological improvements associated with the amplification reaction itself will facilitate identification of specific subsets of genes that respond when a cell is challenged, as well as greater identification efficiency of global changes in gene expression.

References

Acland, P., Dixon, M., Peters G., and Dickson, C., 1990, The subcellular fate of the Int-2 oncoprotein is determined by choice of inititation codon., *Nature* **343**:662-665.

Ahmed, C.M.I., Chanda, R.S., Snow, N.D., and Zain, B.S., 1982, The nucleotide sequence of mRNA for the M_r 19000 glycoprotein from early gene block III of adenovirus 2, *Gene* **20**:339-346.

Allard, P., Yang, Q., Marzluff, W.F., and Clarke, H.J., 2005, The stem-loop binding protein regulates translation of histone mRNA during mammalian oogenesis, *Devel. Biol*. **286**:195-206.

Allen, E., Wang, S., and Miller, S.A., 1999, Barley yellow dwarf virus RNA requires a cap-independent translation sequence because it lacks a 5′ cap, *Virology* **253**:139-144.

Alwine, J.C., Kemp, D.J., and Stark, G.R., 1977, Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes, *Proc. Natl. Acad. Sci*. **74**:5350-5354.

Alwine, J.C., D.J. Kemp, B.A. Parker, J. Reiser, J. Renart, G.R. Stark, and G.M. Wahl. (1979). Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzyloxymethyl paper, *Methods Enzymol*. **68**:220-242.

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*, H. Trachsel, ed., CRC Press, Ft. Lauderdale pp.71-79.

Bassett, C.L., Nickerson, M.L., Cohen, R.A., and Rajeevan, M.S., 2000, Alterantive 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.

Belostotsky, D.A., and Rose, A.B., 2005, Plant gene expression in the age of systems biology: integrating transcriptional and post-transcriptional events, Trends in Plant Sci. **10**:347-353.

Berk, A.J., and Sharp, P.A., 1977, Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids, *Cell* **12**:721-732.

Berget, S.M., and Robberson, B.L., 1986, U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for *in vitro* spicing but not for polyadenylation, *Cell* **46**:691-696.

Birnstiel, M.L., Busslinger, M., and Strub, K.., 1985, Transcription termination and 3′ end processing: the end is in site! *Cell* **41**:349-359.

Black, D.L., and Steitz, J.A., 1986, Pre-mRNA splicing *in vitro* requires intact U4/U6 small nuclear ribonucleoprotein, *Cell* **46**:697-704.

Browning, K.S., 1996, The plant translational apparatus *Plant Mol. Biol*. **32**:107-144.

Bugler, B.F., Amalric, F., and Prats, H., 1991, Alternative initiation of translation determines cytoplasmic or nuclear localization of basic fibroblast growth factor, *Mol. Cell. Biol*. **11**:573-577.

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.

Busch, M.A., Bomblies, K., and Weigel, D., 1999, Activation of a floral homeotic gene is *Arabidopsis, Science* **285**:585-587.

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.

Casey, J., and Davidson, N., 1977, Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentration of formamide, *Nucleic Acids Res*. **4**:1539-1552.

Chang, C., Sheen, J., Bligny, M., Niwa, Y., Lerbs-Mache, S., and Stern, D.B., 1999, Functional analysis of two maize cDNAs encoding T7-like RNA polymerases, *Plant Cell* **11**:911-926.

Cheong, Y.H., Chang, H.S., Gupta, R., Wang, X., Zhu, T., and Luan, S., 2002, Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis, Plant Physiol*. **129**:661-677.

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.767.

Futterer, J., and Hohn, T., 1996, Translation in plants -rules and exceptions, *Plant Mol. Biol*. **32**:159-189.

Guo, L., Allen, E., and Miller, W.A., 2001, Base-pairing between untranslated regions facilitates translation of uncapped, non-polyadenylated viral RNA, *Mol. Cell* **7**:1103-1109.

Hamdan, M., and Righetti, P.G., 2005, *Proteomics Today: Protein Assessment and Biomarkers Using Mass Spectrometry, 2D Electrophoresis, and Microarray Technolog,* John Wiley & Sons, Inc., Hoboken, NJ, pp. 448.

Hann, S.R., 1994, Regulation and function of non-AUG-initiated protooncogenes, *Biochime* **76**:880-886.

Hashimoto, C., and Steitz, J.A., 1986, A small nuclear ribonucleoprotein associates with AAUAAA polyadenylation signal *in vitro*, *Cell* **45**:581-591.

Hedtke, B., Börner, T., and Weihe, A., 1997, Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis, Science* **277**:809-811.

Hedtke, B., Börner, T., and Weihe, A., 2000, One RNA polymerase serving two genomes, *EMBO Reports* **51**:435-440.

Huang, H., Mizukami, Y., Hu, Y., and Ma, H., 1993, Isolation and characterization of the binding sequences for the product of the *Arabidopsis* floral homeotic gene *AGAMOUS, Nucleic Acids Res*. **21**:4769-4776.

Ikeda, T.M., and Gray, M.W., 1999, Identification and characterization of T3/T7 bacteriophage-like RNA polymerase sequences in wheat, *Plant Mol. Biol*. **40**:567-578.

Jackson, D.A., Pombo, A., and Iborra, F., 2000, The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells, *FASEB J*. **14**:242-254.

Joshi, C.P., 1987, An inspection of the domain between putative TATA box and transation start site in 79 plant genes, *Nucleic Acids Res*. **15**:6643-6652.

Joshi, C.P., Zhou, H., Huang, H., and Chiang, V.L., 1997, Context sequences of translation initiation codon in plants, *Plant Molecular Biol*. **35**:993-1001.

Jung, A., Sippel, A.E., Grez, M., and Schutz, G., 1980, Exons encode functional and structural units of chicken lysozyme, *Proc. Natl. Acad. Sci*. **77**:5759-5763.

Kahvejian, A., Svitkin, Y.V., Sukarieh, R., M'Boutchou, M.N., and Sonenberg, N., 2005, Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev*. **19**:104-113.

Kannicht, C., 2002, *Post-Translational Modification of Proteins: Tools for Functional Proteomics*, Humana Press, Totowa, NJ, pp. 336.

Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzkel, M., and Matzke, A.J.M., 2005, Atypical RNA polymerase subunits required for RNA-directed DNA methylation, *Nature Genetics* **37**:761-765.

Keen, J.N., and Ashcroft, AE, 1999, Sequence analysis of expressed proteins, in: *Post-Translational Processing: A Practical Approach*, S.J. Higgins and B.D. Hames, eds, Oxford University Press, New York, NY, pp. 334.

Kim, M., Canio, W., Kessler, S., and Sinha, N., 2001, Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato, *Science* **293**:287-289.

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, *Nucleic 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., 1989a, Context effects and inefficient initiation at non-AUG codons in eukaryotic cell-free translation systems, *Mol. Cell. Biol*. **9**:5073-5080.

Kozak, M., 1989b, The scanning model for translation: an update, *J. Cell Biol*. **108**:229-241.

Kozak, M., 1990, Downstream secondary structure facilitates recognition of inititator codons by eukaryotic ribosomes, *Proc. Natl. Acad. Sci*. **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 the 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., 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.

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.

Lander, E.S., *et al*., 2001, Initial sequencing and analysis of the human genome, *Nature* **409**:860-892.

Lewin, B., 2004, *Genes VIII*, Pearson Education, Upper Saddle River, NJ, pp.1027.

Liebler, D.C., 2002, *Introduction to Proteomics*, Humana Press, Totowa, NJ, pp.194.

Loeffler, M., and Kroemer, G., 2000, The mitochondrion in cell death control: certainties and incognita. *Exp. Cell Res*. **256**:19-26.

Maquat, L.E., 1997, RNA Export from the nucleus, In: *mRNA Metabolism and Post-Transcriptional Gene Regulation*, J.B. Harford and D.R. Morris, eds., Wiley-Liss, New York, NY, pp.107-126.

Mazumder, B., Seshadri, V., and Fox, P.L., 2003, Translational control by the 3′-UTR: the ends specify the means, *Trends Biochem. Sci*. **28**:91-98.

Melton, D.A. Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R., 1984, Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res*. **12**:7035-7056.

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.

Molina, C., and Grotewold, E., 2005, Genome wide analysis of *Arabidopsis* core promoters. *BMC Genomics* **6**:25.

Morris, D.R., 1995, Growth control of translation in mammalian cells, *Prog. Nucleic Acid Res. Mol. Biol*. **51**:339-363.

Nakayama, J., Rice, J.C., Strahl, B. D., Allis, C. D., and Grewal, S. I. S., 2001, Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly, *Science* **292**:110–113.

Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S., 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.

Piñol-Roma, S., and Dreyfus, G., 1993, hnRNP proteins: Localization and transport between the nucleus and the cytoplasm, *Trends Cell Biol*. **3**:151-155.

Pradhan, S., and Adams, R.L.P., 1995, Distinct CG and CNG DNA methyltransferases in *Pisum sativum, Plant J*. **7**:471–481.

Preobrazhensky, A.A., and Spirin, A.S., 1978, Informosomes and their protein components: the present state of knowledge, *Prog. Nucleic Acid Res. Mol. Biol*. **21**:1-38.

Quarless, S.A., and Heinrich, G., 1986, The use of complementary RNA and S1 nuclease for the detection of low abundance mRNA transcripts, *BioTechniques* **4**:434-438.

Richter, U., Kiessling, J., Hedtke, B., Decker, E., Reski, R., Borner, T., and Weihe, A., 2002, Two RpoT genes of *Physcomitrella patens* encode phage-like RNA polymerases with dual targeting to mitochondria and plastids, *Gene* **290**:95-1005.

Riechmann, J.L., Ito, T., and Meyerowitz, E.M., 1999, Non-AUG initiation of *AGAMOUS* mRNA translation in *Arabidopsis thaliana, Mol. Cell. Biol*. **19**:8505-8512.

Ruiz-Echevarria, M.J., Czaplinski, K., and Peltz, S.W., 1996, Making sense of nonsense 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 frame, *Cell* **101**:741-751.

Schimke, R.T., 1981, Chromosomal and extrachromosomal localization of amplified DHFR genes in cultured mammalian cells, *Cold Spring harbor Symp. Quant. Biol*. **45**:785-797.

Smale, S.T., and Baltimore, D., 1989, The “initiator” as a transcription control element, *Cell* **57**:103-113.

Soeiro, R., Vaughan, M.H., Warner, J.R., and Darnell, Jr., J.E., 1968, The turnover of nuclear DNA-like RNA in HELA cells, *J. Biol. Chem*. **39**:112-118.

Southern, E.M., 1975, Detection of specific sequences among DNA fragments separated by gel electrophoresis, *J. Mol. Biol*. **98**:503-517.

Tosi, M., Young, R.A., Hagenbuchle, O., and Schibler, U., 1981, Multiple polyadenylation sites in a mouse α-amylase gene, *Nucleic Acids Research* **9**:2313-2324.

Tarun, S.Z., and Sachs, A.B., 1996, Association of the yeast poly(A) tail binding protein with translation factor eIF-4G, *EMBO J*. **15**:7168-7177.

Tuzon, C.T., Borgstrom, B., Weilguny, D., Egel, R., Cooper, J.P., and Nielsen, O., 2004, The fission yeast heterochromatin protein Rik1 is required for telomere clustering during meiosis, *J. Cell Biol*. **165**:759–765.

Venter, J.C., *et al*., 2001, The sequence of the human genome, *Science* **291**:1304-1351.

Vogt, V.M., 1980, Purification and properties of S1 nuclease from *Aspergillus*. *Methods Enzymol*. **65**:248-255.

Wang, L., and Wessler, S.R., 1998, Inefficient reinitiation is responsible for upstream open reading frame-mediated translational repression of the maize R gene, *Plant Cell* **10**:1733-1745.

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.

Wax, S.D., Nakamura, H., and Anderson, P.J., 2005, The tumor necrosis factor-α AU-rich element inhibits the stable association of the 40S ribsosomal subunit with RNA transcripts, *Biochem. Biophys. Res. Comm*. **333**:1100-1106.

Weihe, A., Hedtke, B., and Börner, T., 1997, Cloning and characterization of a cDNA encoding a bacteriophage-type RNA polymerase from the higher plant *Chenopodium album, Nucleic Acids Res*. **25**:2319-2325.

Wells, S.E., Hillner, P.E., Vale, R.D., and Sachs, A.B., 1998, Circularization of mRNA by eukaryotic initiation factors, *Mol. Cell* **2**:135-140.

Wiese, A., Elzinga, N., Wobbes, B., and Smeekens, S., 2005, Sucrose-induced translational repression of plant bZIP-type transcription factors, *Biochem. Soc. Trans*. **33**:272-275.

Young, D.A., Allen, R.L., Harvey, A.J., and Lonsdale, D.M., 1998, Characterization of a gene encoding a single-subunit bacteriophage-type RNA polymerase from maize which is alternatively spliced, *Mol. Gen. Genet*. **260**:30-37.

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 (*s*ilent *i*nformation *r*egulator) 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

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, Py_2CAPy_5, 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 C_4 photosynthetic pathway and is located in mesophyll chloroplasts. The second form is found in plants that use the standard C_3 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 C_4 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 C_4 *vs*. C_3 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 (*LCA1*) 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 *LCA1* 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 GCCGCCPuCC*AUG*G, 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)a*AUG*GCg; the sequences for dicots and monocots is aaA(A/C)a*AUG*GCu and c(a/c)Pu(A/C)c*AUG*GCG, 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

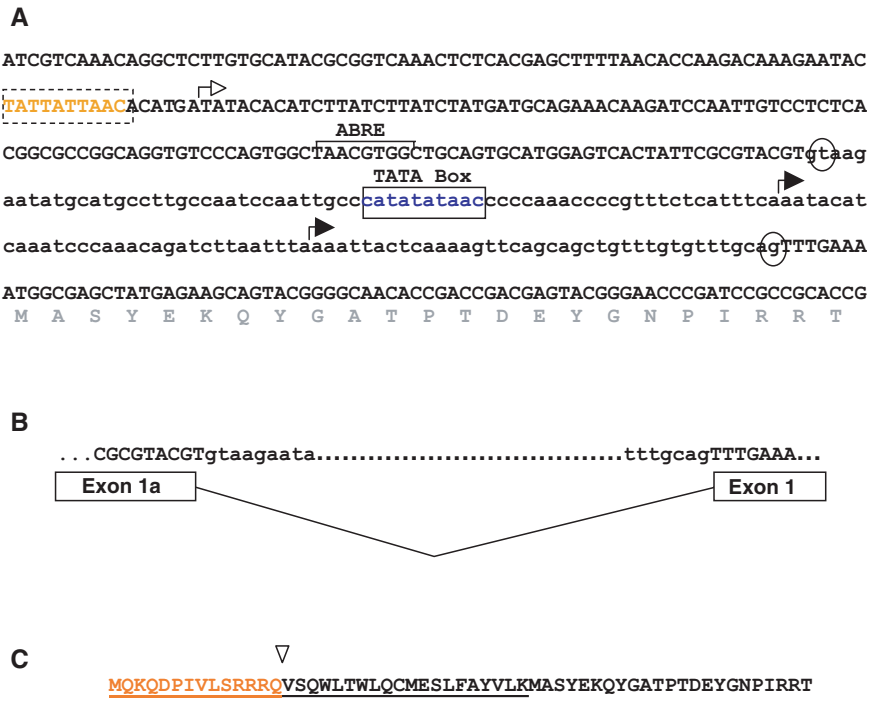


FIGURE 2.2. The promoter region of a peach dehydrin gene is illustrated. A. Sequence of the 490 bases upstream of the canonical translation start site is shown. The 5′ leader intron is shown in lower case lettering, and the TATA element (in blue) that regulates expression in bark is outlined by a solid 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. Moncistronic 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 counterpoints, 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

copia-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 onithine 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.

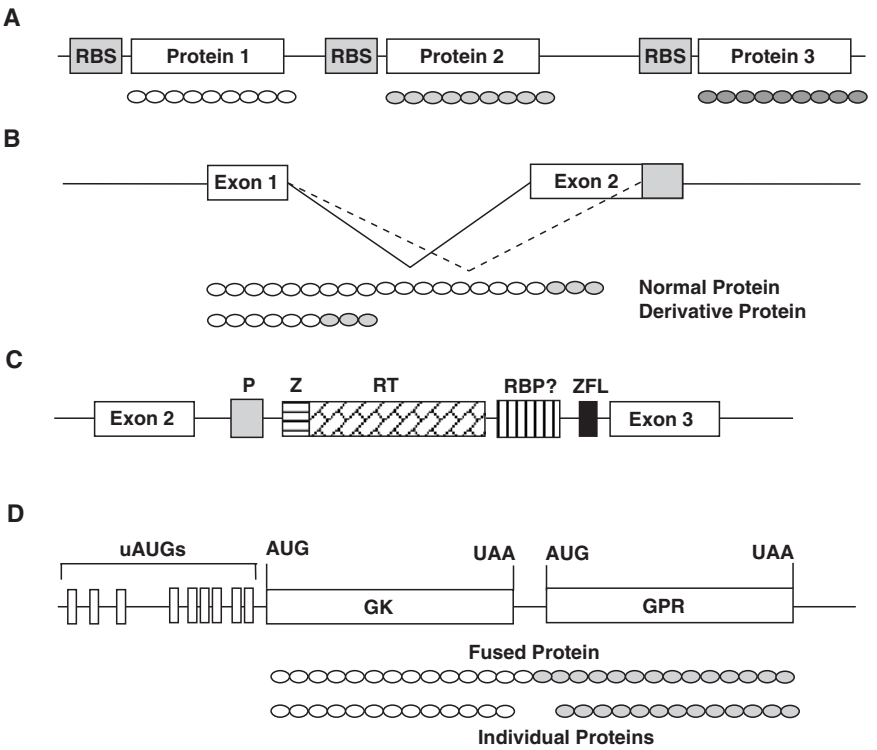


FIGURE 2.3. Different types of polycistronic mRNAs are illustrated. Exons: boxes; introns and intergenic regions: solid and dotted lines; protein products are shown below the transcripts. RBS: ribosomal binding site; AUG: initiator codon; UAA: translation stop codon. A. A typical prokaryotic polycistronic 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; Kruszka *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; Yoshiba *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 2nd 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		
	AY133618	Transferase [OMT]		
At5g57260	BT004038	CYP71B10	NM_125108[a]	Fusion protein containing entire CYP71B10 joined to the predicted At5g57250 protein
At5g57250	No	Hypothetical		
At4g39480	RAFL15-02-O16[a]	CYP96A9	NM_120108[a]	Fusion protein merging CYP96A9 with CYP96A10
At4g39490	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., deBruijin, 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 (MFα): 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 nonsense 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 is 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.

Wetizel, 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., Tworkoski, 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.

3
Alternative Processing as a Mechanism for Regulating Gene Expression

Eliezer S. Louzada

3.1. Introduction

Alternative splicing (AS) is widely accepted to play an important role in the regulation of gene expression in higher organisms and to be responsible for the protein diversity necessary to maintain the complexity of life. The unexpected observation that the number of human genes was not in the hundreds of thousands, but in the range of *ca*. 25,000, created an apparent discrepancy between gene numbers and organismal complexity. It was difficult to explain how such a simple worm like *Caenorhabditis elegans* could have almost 20,000 genes compared to the relatively low number in highly complex humans. Although there are other mechanisms that generate protein diversity, AS is the most important (Grabowski and Black, 2001) because it can modulate transcript expression levels by marking transcripts for degradation *via* nonsense-mediated decay (NMD), and modify the structure of gene products by inserting or deleting protein domains (Stamm *et al*., 2005). Even though, in most cases, a direct link between AS and its function in regulating a specific pathway or process is not known, it is clear that AS is essential for guaranteeing the complexity of an organism and for insuring proper organ functioning, as well. A global analysis of human genes indicates that the majority of AS is involved in transmission, reception and response to cellular signals, with prevalence seen in immune and nervous systems which are involved in processing a large amount of information (Moderek *et al*., 2001). *Dscam*, a gene involved in brain wiring in *Drosophila* and homologous to a human gene, can potentially generate more than 38,000 isoforms (Schumucker *et al*., 2000), indicating the importance of AS in gene regulation.

In plants AS is not as well studied as in mammalian systems; however, recent estimates indicate that 20–25% of *Arabidopsis* and rice genes are alternatively spliced (Wang, 2005). Probably this percentage will increase as further research uncovers additional examples of AS. AS has been reported in potato (Bournay *et al*., 1996), wheat (Mastrangelo *et al*., 2005), tobacco (Dinesh-Kumar and Baker, 2000), apple (Zhang *et al*., 2003), tomato (Snedden and Blumwald, 2000), maize (Grotewold *et al*., 1991), and several other plant

species, indicating that this is also a widespread mechanism for gene regulation in plants. Zhou *et al*., (2003) performed an extensive literature survey and found, at that time, 168 genes were reported to be alternatively spliced in plants across 44 species. The majority of the genes were related to cell growth and/or maintenance, cell communication, and plant development.

In this review we examine the body of knowledge about AS in plants, including mechanisms for its regulation, mode of action and possible functional significance, as well as the relation of AS to stress. Additionally, AS in other organisms will be included for further enlightenment on this important process. The different types of alternative splicing will be referred to in this chapter as AS, unless the meaning is compromised by the abbreviation.

3.2. Regulation of Alternative Splicing

The regulation of AS is a very complex process with multiple control mechanisms requiring the appropriate positioning of sequence elements, binding sites for small nuclear ribonucleoproteins (snRNPs), and the participation of numerous splicing factors. AS is precisely regulated in a tissue-specific and developmental stage-specific manner, and, therefore, changes in the relative levels of alternatively spliced isoforms are expected to affect cellular functions (Nissim-Rafinia and Kerem, 2002).

3.2.1. Splice Site Recognition

The precise processing of a pre-mRNA, with the removal of intron sequences is crucial for the synthesis of native proteins and for the transport of the mature mRNA from the nucleus to the cytoplasm (Müller *et al*., 1998). The cellular machinery responsible for this tightly regulated process is called the supraspliceosome which is composed of four spliceosomes assembled together by pre-mRNA (Sperling *et al*., 1997; Müller *et al*., 1998; Azubel *et al*., 2006). It is suggested that the supraspliceosome functions as a "rolling model" where, in a multi-intronic pre-mRNA after the simultaneous processing of four introns, the RNA rolls, placing a new subset of introns in the correct position to be processed (Azubel *et al*., 2006). This proposed single model supports the hypothesis that in early stages of spliceosome assembly there is a need for communication between splicing factors binding the 3′ and 5′ splicing sites bordering an exon (Robberson *et al*., 1990). The complexity of the splicing mechanism is remarkable, requiring accurate recognition of consensus sequences in the 5′ (AG/GTAAG) and 3′ (TGCAG/G) splice sites. Even though sequences with similarity to the consensus might be found at high frequency in many positions of a pre-mRNA, this alone does not guarantee that a particular sequence will be chosen for splicing (Miriami *et al*., 2002). This in turn suggests that there are mechanisms to prevent the selection of wrong sites and facilitate the selection of the right one.

Intron splicing mechanisms are similar in principle among vertebrates; however, the recognition sequence varies greatly even among closely related organisms. Plants and animals seem to share some similarities, but the mechanism is not identical. Pre-mRNA of the oat phytochrome type 3 and maize bronze locus are accurately spliced with moderate efficiency in a human HeLa cell nuclear extract, while the pre-mRNA of bean phaseolin is not spliced at all (van Santen and Spritz, 1987). Goodall and Filipowicz (1991) reported that a mammalian GC-rich intron is efficiently processed in maize but is completely inactive in *N. plumbaginifolia* protoplasts. Additionally, a human growth hormone (hGH) RNA precursor could not be spliced in sunflower tissue (Barta *et al*., 1986). This difference in the ability of plants to properly process pre-mRNA substrates from animals and the fact that pre-mRNA from some plants, but not from others, can be processed in animal systems indicate that there are differences in the RNA-processing mechanism among organisms. The monocot, maize, is able to recognize and splice many introns that are poorly or not spliced at all in the dicot *Nicotiana plumbaginifolia*, indicating that substantial differences in splicing requirement exist between plants and that monocots appear to have a more relaxed splicing system (Goodall and Filipowicz, 1991). Keith and Chua (1986) reported that a ribulose-1,5-bisphosphate carboxylase small subunit gene (*rbcS*) from wheat is inefficiently processed in tobacco, while a *rbcS* from pea is properly processed. The alcohol dehydrogenase gene (*Adh*-1) from maize is even more poorly processed in tobacco than *rbcS* from wheat. Even among plants from the same taxonomic group, differences in splice mechanisms can be observed. Most monocots and dicots do not require a strict branchpoint like the one in yeast or a highly conserved polypyrimidine tract like that observed with human introns (Goodall and Filipowicz, 1989, 1991; Wiebauer *et al*., 1988; Krämer, 1996). However, a potato invertase gene required a strong branchpoint and a U-rich polypyrimidine tract for inclusion of the 9-nucleotide-long mini-exon in the open reading frame (ORF) (Simpson *et al*., 2002, 2004). The same U-rich sequences (U_{11}) were found to be abundant in both *Arabidopsis* and maize introns, and to play a key role in intron splicing in maize (Ko *et al*., 1968). It was earlier proposed that the precise recognition of 5′ and 3′ splice sites (SSs) in plant pre-mRNA requires an AU rich sequence in the introns (Goodall and Filipowicz, 1989) and a set position of the SS relative to the AU transition point (Lou *et al*., 1993; McCullough *et al*., 1993). This requirement is stricter for dicots than for monocots, where, even though the presence of AU-rich sequences facilitates splicing, it is not an absolute requirement unless the SS sequences strongly deviate from the consensus (Goodall and Filipowicz, 1991). Gniadkowski *et al*., (1996) reported that insertion of U-rich, but not A-rich, segments could activate splicing in a GC-rich synthetic intron regardless of the site of insertion. Furthermore, replacement of one or two Us in the sequence UUUUUAU by A or C had a strong inhibitory effect. The contrast between the U-richness in the intron and the flanking exons determines splicing efficiency, with U-rich sequences being much more effective than A-rich sequences (Ko *et al*., 1968).

Recognition of the 3′ SS in the *Adh*1 gene intron expressed in tobacco is position-dependent in relation to the AU-rich elements. The 3′ (acceptor site) intron boundary (AG/) is defined by multiple AU elements positioned between –66 and –6 nucleotides relative to the 3′ SS. Replacement of U by A residues in these elements activates a cryptic SS upstream in the intron (Lou *et al*., 1993). Moreover, studies using the pea *rbcS3A* intron 1 and the first and second exons indicate that the strength of the U tract immediately preceding the 3′ terminal CAG is an important factor in 3′ SS definition. Replacement of nine U residues by A in the position –4 to –17 upstream from the 3′ SS activates two cryptic exonic SS in the downstream exon. The presence of U-rich sequences upstream of SS in plants, animals and yeast has an important role in 3′ SS selection (Bayton *et al*., 1996).

At the 5′ (donor) splice site, maintenance of complementarity of the 5′ consensus sequence to the U1-small nuclear RNA (U1-snRNA), is an important determinant of 5′ splice site selection (McCullough *et al*., 1993; Zhuang and Weiner, 1986). Moreover, the actual splice site is chosen on the basis of its agreement to the consensus sequence and its position relative to the AU transition point (McCullough *et al*., 1993). The intron boundaries are initially established by recognition of the transition between the AU-moderate exon and AU-rich intron sequences and subsequently defined by their interaction with snRNA and other splicing factors (McCullough and Schuler, 1997). There is also evidence that, as in animals, a strong branchpoint is required for the recognition of the 5′ SS (Simpson *et al*., 1996; Simpson *et al*., 2002; Simpson *et al*., 2004). Simpson *et al*. (1996) suggested that a possible role of the AU-rich sequence is to stabilize the U1-SnRNP-5′ splice site interaction. Undoubtedly, U1 snRNA is closely associated with the 5′ SS selection. The 5′ end of U1 snRNA recognizes the 5′ splice site in the pre-mRNA by complementary base pairing (Ko *et al*., 1998). U1 is part of a variety of small, nuclear RNAs (U1, U2, U4, U5, U6 have been well characterized) and nuclear proteins which are intimately associated with the RNA precursor during the pre-mRNA splicing and spliceosome assembly. There are a multitude of snRNA species in plants with great variation in molecular weight. For example, dicot plants contain multiple U1 snRNAs that are smaller than the ones in monocots, while both groups have multiple U5 snRNAs that are in the same size range. Dicots express only a few U4 snRNAs, while monocots have multiple U4 genes. All plant nuclei contain a single U6 snRNA (Egeland *et al*., 1989). Additionally, monocot and dicot U1 and U2 snRNAs are distinct entities that may associate with specific monocot and dicot snRNP proteins (Musci *et al*., 1992). While it is not yet clear how these snRNAs work on plant pre-mRNA, it is possible that, as in mammals, most of them are involved in the SS recognition and splicing process in plants. In the mammalian system, the 5′ splice site and the branchpoint are partially recognized by base-pairing with U1 and U2 snRNAs, respectively, generating the pre-spliceosome. U4 mediates the entry of U6, together with its three snRNP partners and U5, into the spliceosome to replace U1 at the 5′ SS.

Subsequently, U6 and U2 form a base-pairing interaction that juxtaposes the 5′ SS with the branchpoint, and finally U5, *via* interaction with its loop, promotes the binding of the exons (McConnell and Steitz, 2001).

3.2.2. Factors Affecting Canonical and Alternative Splicing

Canonical and AS are tightly regulated by a multitude of splicing factors (SF) distributed in nuclear domains known as splicing factor compartments or "speckles" (George-Tellez *et al*., 2002; Ali *et al*., 2003; Docquier *et al*., 2004; Lorković and Barta, 2004). In mammalian cells the speckles are highly dynamic structures with splicing factors being rapidly recruited from speckles to the site of transcription after gene activation (Misteli *et al*., 1997). Recruitment of an essential splicing factor, such as a serine/arginine-rich (SR) protein, from nuclear speckles involves the dissociation of the SF from the speckles by phosphorylation of the SR domain. Its subsequent association with the target RNA is mediated by RNA-recognition motifs (RRM) (Misteli *et al*., 1998) present in the protein. Tillemans *et al*. (2005) demonstrated that the SR and RRM domains in *Arabidopsis* SR proteins have a conserved role in targeting the protein to nuclear speckles. Compartmentalization of plant splicing factors in speckles has been documented by several studies (Ali *et al*., 2003; Docquier *et al*., 2004; Lorković and Barta, 2004). Fang *et al*. (2004) demonstrated that protein components from plant speckles rapidly exchange with the nucleoplasmic population, and over time merge, bud, disassemble, or reassemble at a new site; however, all these dynamic events ceased when transcription was blocked, indicating the close relationship of speckles' dynamics with transcription.

In AS, distinct combinatorial sets of generic pre-mRNA splicing factors are recruited to sites of alternatively spliced transcripts and contribute to the splicing outcome. This combinatorial model is valuable because it accounts for the fact that cells can rapidly adjust the splice outcome in response to intracellular and extracellular signals by modulating the binding of existing splicing factors without the need of *de novo* synthesis (Mabon and Misteli, 2005). Shinozaki *et al*. (1999) studied the expression of a tyrosine kinase (*trk*) during neural differentiation in embryonal carcinoma cells and observed that the stoichiometric balance of novel and/or general splicing-associated factors could be associated with AS of *trk* transcripts during differentiation.

Park *et al*. (2004) examined the role of more than 70% of the *Drosophila* RNA-binding-proteins in regulating AS and identified 47 as splicing regulators, 26 of which had not been previously implicated in AS. They found that fluctuation in concentration or activities of core components of the spliceosome, such as U1 snRNP, U2 snRNA auxiliary factor (U2AF), a component of the U2/U6 snRNP, and other proteins, affected AS. Additionally, they observed that AS of different regions in a single pre-mRNA could be regulated by overlapping but distinct sets of proteins.

Wang and Brendel (2004) identified 109 splicing factors and 60 splicing regulators in *Arabidopsis* in addition to many other protein-coding genes related to splicing. Lorković and Barta (2002) previously identified 46 proteins involved in *Arabidopsis* splicing. Around 50% of the splicing-related genes were duplicated in *Arabidopsis* with an even higher percentage for the splicing regulators, suggesting that, although the splicing mechanism is generally conserved among plants, the regulation of splicing may be more variable and flexible (Wang and Brendel, 2004).

SR proteins are by far the most well characterized splicing factors, and heterogeneous nuclear ribonucleoproteins (hnRNPs) and SR protein kinases are the most frequently studied splicing regulators. Splicing regulators are proteins that can either modify splicing factors or compete for their binding sites (Wang and Brendel, 2004). SR proteins are a highly conserved family of essential SFs present in all eukaryotes with the exception of budding yeast (Mabon and Misteli, 2005). They contain one or two RNA-binding domains (RRM domain) and a variable-length arginine/serine-rich domain (RS domain). There are other RS-domain-containing proteins distinct from the SR protein class which are involved in various aspects of pre-mRNA processing (Gravely, 2000). SR proteins are the most versatile SFs with multiple functions reported during the splicing process. For example, one function involves cooperation with U1snRNP to define the functional 5′ interaction with U2AF to promote and/or stabilize complex assembly at the 3′ SS; a second function is bridging the 5′ and 3′ SS to form stable commitment complexes (Fu, 1995). Furthermore, SR proteins are also involved in mRNA transport, stability and translation (Huang, *et al*., 2003; Lin *et al*., 2005).

At least 19 SR proteins are present in *Arabidopsis* (Lorković and Barta, 2002; Kalyna and Barta, 2004) and 20 in rice (Isshiki *et al*., 2006), of which six display plant-specific characteristics. Most of the SR proteins and other splicing-related genes are themselves alternatively spliced. Wang and Brendel (2004) reported that 30.8% of splicing-related genes in *Arabidopsis* were themselves alternatively spliced. Alternative splicing in SR-coding genes was also reported in rice (Isshiki *et al*., 2006), maize (Gao *et al*., 2004; Gupta *et al*., 2005), and *Arabidopsis* (Lazar and Goodman, 2000), suggesting that AS contributes to diversification of splicing factors and to autoregulation of their expression levels. Two SR-coding genes from maize, *zmRSp31A* and *zmRSp31B*, are alternatively spliced, and most of the spliced transcripts are predicted to produce small proteins lacking the RS domain. Curiously, the wild type mRNA, which encodes the SR domain, is present only in the total RNA pool, while some of the smaller AS transcripts are present only in the polysomal fraction, suggesting that some of these spliced transcripts are translated (Gao *et al*., 2004). Each splicing pattern may be unique to each SR protein, as different SFs may stimulate splicing in different regions. *In vivo* assays in rice protoplasts with plant-specific *RSp29* and *RSZp23* show that *RSp29* enhances splicing at the distal site, while *RSZp23* stimulates splicing at the proximal site (Isshiki *et al*., 2006).

The diversification of function of SR proteins may be a result of both their structure (one or two RRM domains plus the RS domain) and their alternative splicing patterns. Variability in amino acid composition in submotifs of the RRM region in the same SR protein family, as observed in rice (Isshiki *et al.*, 2006), might further contribute to diversification and specificity. Recent studies by Liu and Manley (2005) added an extra function for SR proteins beyond their role in splicing. They observed that depletion of the SR protein ASF/SF2 from chicken cells caused the accumulation of high-molecular-weight DNA fragments and the formation of R-loop structures, leading to hypermutation phenotypes. Lack of ASF/SF2 might loosen or destabilize the pre-messenger ribonucleoprotein particles (pre-mRNPs), exposing regions of the RNA to template DNA and forming R loops. Overexpression of RNase H in these cells suppressed the observed genomic instability. The authors suggest that ASF/SF2 prevents R loop formation by binding to nascent pre-mRNA, thus preventing the transcripts from associating with template DNA.

3.2.2.1. Splicing Factors and Phosphorylation

Individual SR proteins, in addition to functioning as essential splicing factors, interact with splicing regulators to control alternative splicing (Tian and Maniatis, 1993). Early studies demonstrated that the association of SR proteins with the splicing regulators Tra and Tra2, products of *transformer* and *transformer 2* genes, were essential to direct the alternative splicing process responsible for sex determination in *Drosophila* (Du *et al.*, 1998). Tra and Tra2 act by recruiting SR proteins to regulatory elements located downstream of a female-specific 3′ splicing site (forming a splicing enhancer complex) in the *doublesex* (*dxs*) pre-mRNA; binding of these proteins promotes female-specific splicing of *dxs* (Tian and Maniatis, 1993). Phosphorylation of the SR protein by SR protein kinase is required for normal *dsx* splicing. Reduction of the activity of the kinase induces the production of male- and female-specific forms of *dsx* mRNA (Du *et al.*, 1998). The state of phosphorylation of the SR protein splicing factor is intimately associated with its function and is closely tied to its ability to coordinate gene expression. Overexpression of two members of the Clk mammalian kinase family, hClk2 and hClk3, causes the redistribution of SR proteins in the nucleus, and may increase the nucleoplasmic concentration of splicing factors, thus changing the balance of factors available to modify SS selection at a particular pre-mRNA (Duncan *et al.*, 1998). Dephosphorylation of SRp38, a conserved SR protein in vertebrates, has a crucial role in HeLa cell survival under heat stress by promoting interaction of SRp38 with the U1 snRNP protein and interfering with 5′-splice site recognition. As a result, alternative splicing is suppressed. SRp38 probably represses AS after heat shock, in part, to prevent the possible accumulation of inaccurately spliced mRNA (Shin *et al.*, 2004). It is clear that changes in the phosphorylation state will interfere with gene

expression *via* SR proteins. Mutation in a gene encoding protein kinase Akt2 in humans is associated with severe insulin resistance and diabetes (George *et al*., 2004). Akt2 directly phosphorylates SRp40 and promotes the inclusion of an exon in a gene encoding the protein kinase CβII (PKCβII), which affects its subcellular localization and substrate specificity. Insulin increases phosphorylation of Akt2 by 3-fold and is an activator of Akt2 and other SR kinases (Patel *et al*., 2001; Patel *et al*., 2005). Mutation of the Akt2 phosphorylation site in SRp40 attenuated PKCβII exon inclusion (Patel *et al*., 2005).

The cyclic phosphorylation/dephosphorylation of SR proteins is vital during mRNA processing and export. Hyperphosphorylated SR proteins are recruited to pre-mRNA where they take part in splicing. After splicing they are dephosphorylated and, as hypophosphorylated forms, they participate in the transport of the mRNP to the cytoplasm. Upon completion of translation they are re-phosphorylated by SR kinase and transported to the nucleus (Huang and Steitz, 2005). Endogenous SR protein kinases are mainly localized in the cytoplasm which is critical for nuclear import of SR proteins. The bipartite SR protein kinase catalytic core is separated by a spacer sequence which guarantees its location in the cytoplasm. Removal of the spacer induces the translocation of the kinase to nuclei, causing aggregation of splicing factors and general inhibition of gene expression due to excessive SR kinase activity (Ding *et al*., 2006).

3.2.2.2. Interactions Define Splicing Outcome

The splice outcome of a particular precursor mRNA is the result of complex protein-protein and protein-RNA interactions, with proteins binding to specific regulatory sequences in the pre-mRNA, followed by recruitment of splicing factors and splicing regulators to the sites of transcription. Alternative splicing is regulated in a combinatorial fashion by the action of auxiliary signals and factors. The auxiliary signals can be classified according to their location in the pre-mRNA (intron or exons) and their function (inducers or repressors), for example as exonic splicing enhancers (ESE, usually purine-rich sequences), exonic splicing silencers (ESS), intronic splicing enhancers (ISE), and intronic splicing silencers (ISS) (Hui and Bindereif, 2005). These signals are sequence elements containing binding sites for splicing factor and regulator proteins (Fig. 3.1). The binding sites are sometimes very peculiar with overlapping sites having positive or negative effects depending on which proteins bind to them. This creates a system of regulation that can respond to a small variation in the relative concentration of a single regulatory factor (Expert-Bezançon *et al*., 2004). Antagonism between splicing factors that recognize the negative and positive elements, in conjunction with the strength of the association of the proteins/signals and the strength of the SS, will affect SS choice (Zhu and Krainer, 2001). Canonical exons usually have strong SSs, eliminating the need of other *cis* elements for

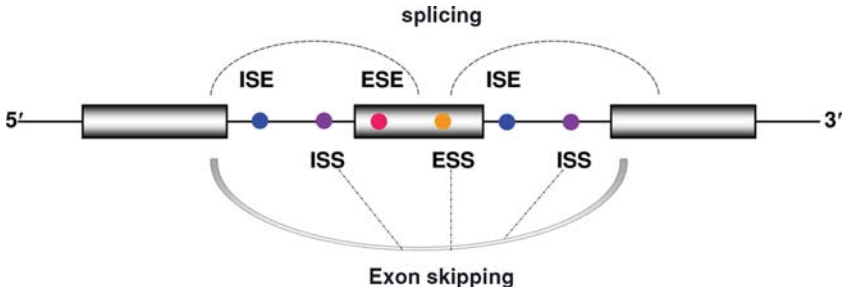


FIGURE 3.1. Enhancer and silencer elements at introns and exons, and splicing consequences. ISE: Intronic splicing enhancer; ESE: Exonic splicing enhancer; ISS: Intronic splicing silencer; ESS: Exonic splicing silencer. (*See Color Plates*)

splicing; however, alternatively spliced exons are frequently weak, leading to exon exclusion. A weak SS on one side of an exon can adversely affect the removal of an intron on the other side; alternatively, a strong SS on one side of a exon can promote the removal of the intron on the other side. In the case of a suboptimal SS, ESEs play an important role in upstream and downstream intron removal (Selvakumar and Helfman, 1999). Additionally, the relative strength of SR protein-mediated association with alternative and canonical exons will also contribute to defining an alternative splicing pattern (Starks *et al.*, 1999). The strength of an ESE depends on the relative activities of the bound SR splicing factor, the number of SR proteins within the ESE complex, and the distance between the enhancer and the intron. Moreover, different SR domains in the SR protein have different potency in activating the 3′ SS from a downstream binding site, depending on the number of alternating arginine and serine residues and, to some degree, on the length of the RS domain (Graveley *et al.*, 1998). Some exons contain multiple ESEs; however, only one ESE complex at a time is able to interact with the spliceosome. The function of multiple ESEs is to increase the probability of interaction between the ESE complex and the splicing machinery (Hertel and Maniatis, 1998). Humphrey *et al.* (1995) studied the effect of a 32-nucleotide ESE on the regulation of alternative splicing of a large exon from a human gene containing two 5′ splicing sites separated by 687 nucleotides. They found that the presence of the ESE was necessary for activation of the internal SS only in the presence of a competing SS and that the ESE had no effect on 5′ SS usage when placed downstream of both SSs. In the absence of the internal SS, the ESE activated a normally silent cryptic SS near the natural internal 5′ SS. This indicates that ESE stimulates upstream 5′ SS selection. When two ESEs, the 32-nucleotide and a similar 30-nucleotide ESE, were positioned between the two competing 5′ SSs, the 32-nucleotide ESE stimulated the upstream 5′ SS, while the smaller ESE activated the

downstream 5′ SS. The difference between the two ESEs responsible for the specificity for one SS or the other was a simple GAG/A repeat at the 3′ end of the ESE. This indicates that small differences in ESE sequences can significantly affect splicing choices (Elrick *et al*., 1998). The ESE/SR protein complex can also function as a barrier to prevent exon skipping and may play a role in guaranteeing the correct 5′→3′ linear order of exons in spliced mRNAs (Ibrahim *et al*., 2005).

hnRNPs are a large family of RNA binding proteins with several members participating in alternative splicing regulation. Although a few members have been described in plants (Lambermon *et al*., 2000; 2002; Lorković *et al*., 2000; Landsberger *et al*., 2002), there have been no reports as yet of the involvement of plant hnRNPs in splicing regulation. While most SR proteins recognize ESEs and promote exon inclusion, some hnRNPs bind to ESSs and block exon recognition (Zhu and Krainer, 2001); however, many other members stimulate AS. In the human CD45 gene, exon 4 usage is suppressed by hnRNP L, and mediated by an ESS (Rothrock *et al*., 2005). Exon 3 of the HIV1 *tat* gene contains several elements with positive and negative effects on splicing, and this juxtaposition of ESE and ESS elements in the same exon seems to be a common phenomenon. hnRNP A1 mediates repression of exon 3 by binding to an ESS. The presence of SF2/ASF and SC35 SR proteins had differential effects. SF2/ASF seems to be able to displace hnRNP A1 and promote the splicing of the exon, while SC35 complemented splicing only after depletion of hnRNP or mutation of the hnRNP A1 site in the ESS (Zhu and Krainer, 2001). The relative concentration of hnRNP A1 and the SF2 SR protein precisely determines which 5′ SS is selected in a pre-mRNA containing duplicated 5′ or 3′ SS: an excess of hnRNP favors distal 5′ SSs, while an excess of SF2 promotes the use of a proximal 5′ SS. This antagonistic activity may play a role in tissue-specific and developmental control of gene expression by AS (Mayeda and Krainer, 1992). The *in vivo* mechanism for the precise removal of large introns from pre-mRNA is not very clear because introns larger than 1kb are not efficiently spliced *in vitro*. Intronic binding sites for hnRNP A/B (A/B binding sites [ABS]) and hnRNP F/H (F binding sites [FBS]) can stimulate *in vitro* pre-mRNA splicing of large introns by bringing the two ends of the intron together *via* association with hnRNP A/B or hnRNP F/H proteins in a looping model. For some pre-mRNAs, the splicing enhancing of the ABS or FBS elements requires their position in both ends of the intron (Martinez-Contreras *et al*., 2006). Several other exonic and intronic sequence elements serve as binding sites for splicing factors and regulators, and many others are still unknown. Short repeats of AC and GT are predicted to function as ISE in fish (Yeo *et al*., 2004). Intronic CA repeats and CA-rich elements can function in the human endothelial nitric oxide synthase (*eNOS*) gene as ISE or ISS depending on their proximity to the 5′ SS. Short distances from the CA-sequence to the alternative 5′ SS correlate with the silencer mode, while longer distances correlate with the enhancer activity. These elements are recognized by hnRNP

and activate 5′ SSs located upstream of the CA repeat (Hui *et al*., 2003; Hui *et al*., 2005). Additional elements reported are groups of intronic G_3s regulating AS in human GH pre-mRNA (McCarthy and Phillips, 1998), intronic UGG repeats coordinating splicing of CD44 alternative exons v8 and v9 (Galiana-Arnoux *et al*., 2005), and the exonic UAGG and the 5′-SS-proximal GGGG motifs functioning cooperatively to silence the brain-region-specific CI exon (Han *et al*., 2005). In rice, all cultivars with more than 18% amylose have the sequence AGGTATA at the leader intron 5′ SS of the gene encoding granule-bound starch synthase (GBSS), while all cultivars with lower amylose content have the sequence AGTTATA. This single nucleotide substitution reduces splicing efficiency, resulting in alternative splicing at multiple sites. The cultivar with multiple isoforms has a substantial increase in mature GBSS transcripts at 18 ˚C compared to 25 or 32 ˚C. The difference in amylose content of the two cultivars at 18 ˚C is much less pronounced than at higher temperatures (Larkin and Park 1999). No association of this sequence with RNA binding proteins has been reported to date.

3.3. Mode of Action of Alternative Splicing

Splicing of introns from precursor mRNA is a vital step for gene expression and involves the recognition of 5′ and 3′ consensus sequences. Considering the limited number of genes in highly complex organisms to carry on the multitude of biochemical tasks and produce the large numbers of proteins needed for life, diversification created by alternative splicing is critical for survival. Alternative splicing can result in skipping of exons, the retention of a whole or part of an intron, the use of alternative 5′ or 3′ splicing sites, and removal of internal fragments of exons, all of which significantly affect gene expression.

Two common types of AS are recognized and two unusual types have also been reported. The most common type of AS is the use of alternative donor or acceptor splice sites (Fig. 3.2). Depending on whether an alternative donor or acceptor site is used, either the 5′-most or 3′-most exon is extended, provided the splice does not change the reading frame. If the reading frame changes, in most cases an in-frame stop codon is encountered and the protein product is truncated at that point. The second most common type of AS involves multiple introns and results in the skipping of one or more exons (Fig. 3.3). In this manner proteins with altered domains or functional regions/motifs can be generated. Of the two unusual modes of AS, intron retention has been most often reported in plants. In general intron retention results in the addition of a few amino acids due to translational read-through at the exon-intron junction. Since for most proteins a stop codon is encountered somewhere within the retained intron, the protein is usually terminated some distance from the 5′ end of the retained intron. The other exceptional type of AS involves the removal of coding sequence as if it were an intron

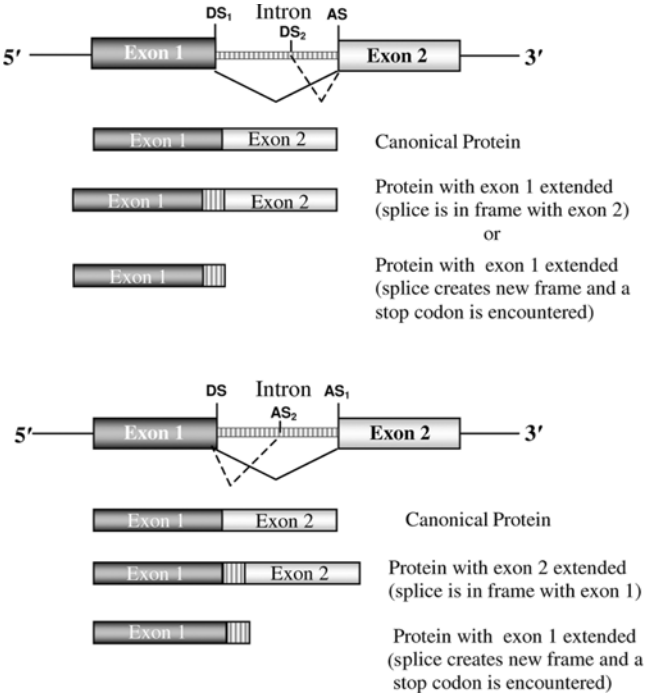


FIGURE 3.2. Illustration of the use of alternative donor and acceptor sites to generate extended or truncated proteins. The mRNAs are shown with the donor and acceptor sites indicated. Exons are shown as rectangles and the intron is shown as a hatched bar. DS1 and DS2: alternative donor sites; AS1 and AS2: alternative acceptor sites. Predicted products are shown below the mRNA. Amino acids added from the intron are shown as a hatched square.

[AU 1] (Fig. 3.4) and, although rarely reported, has been described in both plants and animals. This mode of AS may become more commonly recognized as we develop a better understanding of how splice sites are identified and how they are regulated.

3.3.1. Exon Skipping

The proper recognition of 5′ and 3′ SSs in a pre-mRNA is a crucial step for splicing and for the production of mature translatable mRNA. Two main recognition mechanisms have been proposed: exon definition and intron definition. In the exon definition model, the exons are recognized as a unit as a first step in spliceosome assembly, without the participation of introns. This model requires a combined recognition of 3′ and 5′ SSs by recognition factors and that the SSs are located 300 nucleotides or less apart (Fig. 3.3). If no acceptable 5′ SS is found within 300 nucleotides upstream of the 3′ SS, no

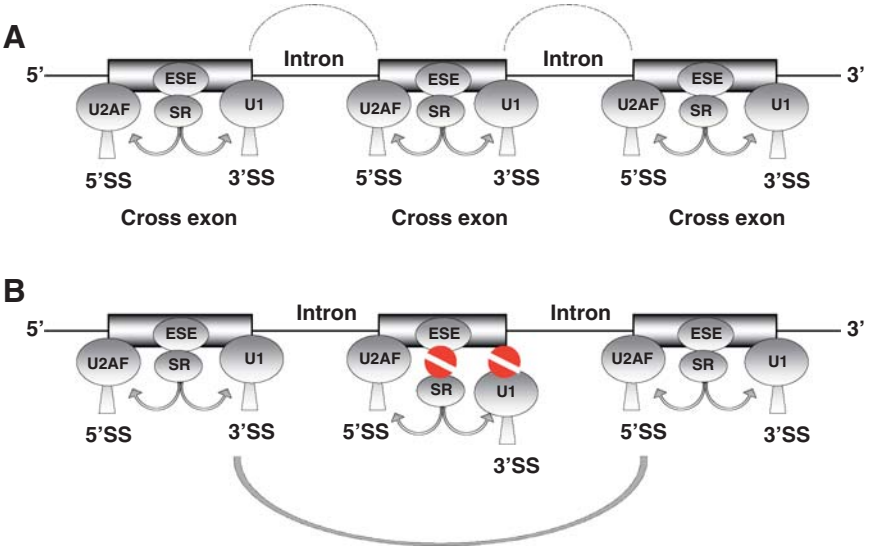


FIGURE 3.3. Mechanism of exon definition during splicing. (A) Exon definition requires the simultaneous recognition of 5′ and 3′ SS across exon. After exon definition, mechanisms at the intron assure that the 3′ end of one exon is joined to the 5′ end of the next exon. (B) Disturbance in the interaction between any components of the splicing machinery (shown in red) can induce exon skipping. (*See Color Plates*)

recognition occurs and the result is exon skipping. This system only applies to internal exons; the first and last exons may be recognized by additional factors (Robberson *et al*., 1990). The majority of vertebrate exons are in the limited size of 300 or less nucleotides (Hawkins, 1998); however, this size requirement creates a problem for the recognition of vertebrate exons that are larger than 300 nucleotides. Chen and Chasin (1994), studied the *in vivo* splicing of several exons, ranging from 50 to 1,500 nucleotides and observed that exons as large as 1,400 nucleotides were spliced. They concluded that exon size is not a limitation for the exon definition mechanism and suggested that factors bound to the two ends of an exon could promote communication across the exon by looping. The intron definition model proposes that the SSs are recognized as pairs across introns, independent of exons, before spliceosome assembly. In this case the intron is the unit of recognition, but the decision to use one system or the other will depend on the relative size of intron and exons. The SSs can be recognized as pairs either across exons or introns, whichever distance is shorter, and a pyrimidine-rich region upstream of the 3′ SS facilitates the exon mode. This system is proposed for small introns, and intron/exon definition could operate in a single pre-mRNA (Telerico and Berget, 1994). Large introns may be spliced by the formation of looping, mediated by hnRNP, which bring the 5′ and 3′ SSs close together (Martinez-Contreras *et al*., 2006). Miriami *et al*. (2003), studied 54 skipped exons and

identified two motifs, both in upstream and downstream introns of skipped exons, which could promote base pairing intra- and inter-introns. They speculated that the exon-skipping decision from 'on' to 'off' could be by shifting from base-pairing inter-intron to intra-intron base pairing. Howe and Ares (1997) identified sequences in introns from yeast pre-mRNA that through complementarity enforced exon inclusion and functioned as an internal intron identity mechanism to promote correct pairing of SSs. Fox-Walsh *et al*. (2005) demonstrated that in *Drosophila* and humans recognition of SSs across introns ceases when the intron sizes are between 200 and 250 nucleotides. Beyond this limit the mechanism switches to recognition across exons. Moreover, the length of the upstream intron is more influential than the downstream intron in inducing alternative splicing. Furthermore, splicing site recognition across introns is more efficient than across exons, stimulating enhanced inclusion of exons with weak SSs.

It is important to realize that exon and intron definitions are basically frameworks where a variety of different factors and sequence elements present in exons and introns work cooperatively to promote recognition of the 5′ and 3′ SSs. In a previous section of this chapter (3.2.2.2.) interactions involving the recognition process and the importance of ESEs, ESSs, ISEs, and ISSs were emphasized. These elements play an important role in splicing of pre-mRNAs containing suboptimal SS. ESEs are recognized by SR proteins and function as a barrier to prevent exon skipping (Ibrahim *et al*., 2005), while ESSs are recognized by certain hnRNPs and block exon recognition (Zhu *et al*., 2001). The presence of ESEs and ESSs together in the same exon may be important for modulation of AS (Asai *et al*., 2003). Additionally, some ESEs can take part in enhancing the second catalytic step of splicing together with SR proteins in a proofreading role to ensure that the correct order of exons is canonically or alternatively spliced (Crew *et al*., 1999). There are fewer reports on ISEs and ISSs; however, they can promote, suppress, or modulate pre-mRNA splicing by associating with different RNA binding proteins (Swanson and Dreyfuss, 1988; Haut and Pintel, 1999; Miyaso *et al*., 2003).

Splice site recognition in plant introns requires AU-rich sequence elements with U residues contributing more than A, and the transition of AU-rich introns to GC-rich exons to function as a SS recognition signal (Bayton *et al*., 1996; Simpson *et al*., 1996; Ko *et al*., 1998). Additionally, in plants U-rich intron elements can function as either a general U-rich signal or as a polypyrimidine tract. Some plants require a branchpoint and a polypyrimidine tract (Simpson *et al*., 1996; Simpson *et al*., 2002; Simpson *et al*., 2004). In some plants, SS selection occurs primarily across introns (intron definition) and secondarily across exons (exon definition), in a combinatorial process (McCullough *et al*., 1996; McCullough and Schuler, 1997; Carle-Urioste *et al*., 1997). The combinatorial model for plant intron recognition establishes that SSs, as well as intron and exon sequences contribute to SS selection and splicing efficiency (Latijnhouwers *et al*., 1999). However, there

is evidence for the operation of exon definition and intron definition mechanisms in plants. The *Arabidopsis* floral organ identity gene *APETALA3* (*AP3*) is required for proper development of petals and stamens in the flower. A flower mutant for the temperature sensitive *ap3-1* allele presents a splicing defect that causes skipping of exon 5 at 23 and 29 ˚C, inducing the petals and stamens to become partially converted to sepals and carpels, respectively. An intronic enhancer has been isolated that suppresses the splicing defect when the plants are grown at 16 ˚C by creating a novel branchpoint sequence that promotes better recognition of exon 5. Exon 5 skipping may be caused by its short length (42 bases) (Yi and Jack, 1998). Small exons (below 50 nucleotides) can induce exon skipping, but the skipping can be reversed by replacing single purines in the polypyrimidine tracts of the upstream intron with pyrimidines (Dominski and Kole, 1991). Exon skipping was also reported in three mutants of the constitutive photomorphogenic locus1 (*COP1*) gene from *Arabidopsis*. A change from G to A in the forth nucleotide upstream from the 3′ SS on intron 5 of the mutant *cop1-1* and G to A in the first nucleotide of intron 6 of *cop1-2* abolished the conserved G within the 5′ consensus and induced skipping of exon 6. In the third mutant, *cop1-8*, a change from G to A in the final nucleotide of intron 10 eliminated the conserved G in the 3′ SS, leading to skipping of exon 11. This suggests that an exon definition mechanism is operating in plants (Simpson *et al*., 1998). On the other hand, intron definition has been reported in maize (Ko *et al*., 1998), tobacco (McCullough *et al*., 1996), potato (Gniadkowski *et al*., 1996; McCullough and Schuler, 1997; Simpson *et al*., 2004), and many other plant species. A synergistic interaction between exon and intron definition was suggested (Carle-Urioste *et al*., 1997).

3.3.2. Intron Retention

Intron retention (IR) is an important phenomenon in humans and plants. In a study comprising more than 21,000 human genes it was observed that almost 15% of them had evidence of IR, with most of the events located in the 3′ untranslated region (3′UTR) (Galante *et al*., 2004). In *Arabidopsis*, more than 30% of AS is reported to be IR, with 17.5% of them appearing as part of coding sequences (CDS) or bridging both CDS and UTRs, while 29% are located at the 5′ or 3′ UTR. Ten percent of the transcripts with IR were related to stress or other stimuli input (Ner-Gaon *et al*., 2004). Low temperature promotes intron retention in wheat (Mastrangelo *et al*., 2005).

In mammals, intron and exon sequences function cooperatively to decide between splicing or non-splicing. Deletion of 115 base pairs from exon 5 of the bovine growth hormone pre-mRNA induces almost complete retention of the up-stream intron (Hampson *et al*., 1989). Insertion of the purine-rich element, GGAAG, which is part of the deleted exon sequence, activates intron splicing. Introduction of several repeats of GGAAG restores splicing to nearly wild type levels. Moreover, mutation of the 5′ SS to conform with

the complementarity needed for U1 snRNA, eliminates the dependence of the intron on the downstream exon (Dirksen *et al*., 1994), indicating that suboptimal 5′ and 3′ SSs are required for IR, and in this case, for splicing to take place there is a need for an ESE (Dirksen *et al*., 1995). Reinforcing the need for interaction between intron/exon, Yue and Akusjärvi (1999) showed that even strong, constitutively active introns are dependent on downstream splicing enhancers. SR proteins binding to ESE or U1 snRNP binding to a downstream 5′ SS can function as splicing-enhancer factors promoting the removal of upstream introns.

In plants, the elevated AU content in introns is well known to be important for their processing (Wiebauer *et al*., 1988). Moreover, U-rich elements, but not A-rich sequences, in introns have a stimulatory effect on splicing, while G residues are strongly inhibitory (Gniadkowski *et al*., 1996). In maize, the presence of U-rich motifs in introns is a key element for their splicing (Ko *et al*., 1998). The presence of U-rich elements in some plant introns can function as a splicing signal or as a polypyrimidine tract (Simpson *et al*., 2004); however, for other plants, rather than a general U- or AU-rich element, the presence of a U_{11} element (*i.e*., polypyrimidine tract) is required in addition to a branchpoint (Simpson *et al*., 1996; Simpson *et al*., 2002). As in other systems, exons and introns function cooperatively to promote intron splicing. Sequences in introns primarily define SS choice in cooperation with weak interactions across exons (McCullough *et al*., 1996). The contrast in base composition between an intron and its flanking exons modulates splicing efficiency. In maize splicing of a GC-rich intron, which still contained a higher percentage of Us than the flanking exons, was increased when GC content was increased in the exons, demonstrating that intron and exons together, cooperate for intron splicing in plants (Carle-Urioste *et al*., 1997).

Intron retention can have profound effects on gene expression, and depending on the location where IR occurs, different results may be observed. Retention at the 5′ UTR may influence tissue specificity, expression levels, and translation efficiency of a gene (Gauss *et al*., 2006), while retention at the 3′ UTR may have drastic effects on mRNA stability (Chan and Yu, 1998; Chen *et al*., 1999). Retention of introns in the open reading frame, *i.e*., ORF, may also result in different effects. In rats, IR in the *Id3* gene, a key regulator of cell proliferation, generates a new functional protein isoform that is absent in normal vessel walls, but is highly expressed following vascular injury and inhibits vascular lesion formation (Forrest *et al*., 2004). Phenyl-ethanolamine *N*-methyltransferase, an enzyme that is differentially regulated in adrenergic neurons in the brain and in adrenal chromaffin cells, is coded by a gene that is alternatively spliced in the prenatal developing rat brainstem, producing two transcripts, one of which retains an intron. The intronless transcript is downregulated postnatally, while the intron-retained transcript is constitutively expressed through adulthood. On the other hand, in the adrenals, only the intronless mRNA is expressed. The intron-retained transcript seems to code for a truncated protein that enzymatically functions to control

expression of the intronlesss transcript (Unsworth *et al*., 1999). Transcripts with a retained intron in the ORF that introduce nonsense stop codons may be a target for nonsense-mediated decay (NMD). Truncated transcripts may also have biological function. Zhang and Gassmann (2003), reported that the function of the *Arabidopsis RPS4* gene, which belongs to the Toll/interleukin-1 receptor(TIR)-nucleotide binding site (NBS)-Leu-rich repeat (LRR) class of disease resistance genes, requires the presence of the full-length and truncated open reading frames. The truncated transcript retains intron 2 or introns 2 and 3. Removal of one or both introns abolishes the function of *RPS4*, even though expression is not affected. The *RPS4* gene confers disease resistance to *Pseudomonas syringae* pv *tomato* strain DC3000.

3.3.3. Cryptic Introns

An unusual class of alternatively processed genes resulting in the creation of two related polypeptides from the same gene was first described for a human fibronectin gene (Vibe-Pedersen *et al*., 1986). A small section (93 bp) of the coding sequence of the IIICS region of the gene was removed, resulting in the deletion of 21 amino acids in-frame from the fibronectin polypeptide. Removal of this "cryptic" intron was not 100% efficient, and both spliced and unspliced forms coexisted in the same cell. The biological function of the deleted region is not apparent as yet. Two recent examples of similar cryptic introns have been reported in plants, both involving protein kinases. In *Arabidopsis* cANP1 encodes a protein with homology to mitogen-activated protein kinase kinase kinases (MAPKKKs) (Nishihama *et al*., 1997). Two transcripts are observed, the shorter one (cANP1S) being derived from the longer by removal of 127 bases of coding sequence. Deletion of this sequence results in a frame shift at the point of removal which gives rise to a truncated kinase with greater activity than the full-length polypeptide. The other example (Bassett *et al*., 2000) involves the removal of approximately 1.5 kbp of sequence encoding leucine-rich repeats from *inrpk1*, a putative leucine-rich repeat receptor protein kinase from morning glory. Two transcripts result from this processing event that differ in acceptor site. Like the case with cANP1S, the shorter *inrpk1b* transcript is frame-shifted and would result in synthesis of a truncated polypeptide. The longer *inrpk1* processed transcript (*inrpk1a*) maintains the original frame and would therefore encode a polypeptide identical to the predicted protein except in the number of leucine-rich repeats which would be reduced from 26 to 5. This type of "cryptic" intron processing is so named because the coding sequences that are removed only resemble introns by their use of similar donor/acceptor recognition sequences; they do not appear to have obvious branching sequences nor are they highly enriched for adenine and/or thymine. Although the few examples illustrated here suggest that cryptic introns occur infrequently, it is likely that such sequences are easily overlooked by most bioinformatics programs. In fact it is hypothesized that this type of processing may be much

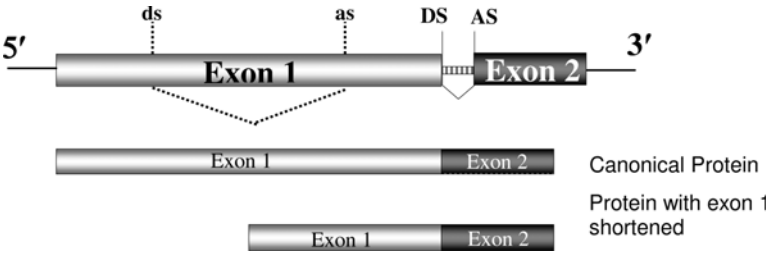


FIGURE 3.4. Cryptic intron processing involves the removal of coding sequence using consensus donor and acceptor sites. Two products can result depending on whether or not removal of the cryptic intron shifts the reading frame. If the frame remains the same, the resulting product is a shortened version of the canonical polypeptide. Cryptic consensus splice sites are shown in lower case letters with a dotted line indicating the portion of the ORF that is removed after splicing.

more common, allowing even greater flexibility in conservation of genetic information and encouraging more subtle levels of control than previously thought possible (Bassett, C.L., personal observation).

3.4. Functional Significance of Alternative Splicing

Alternative Splicing is one of the most important mechanisms for generating protein diversity. A genome wide detection of AS in humans indicates that 74% of AS events occur in the coding sequence, modifying the protein products, while 22% are located at the 5′ UTR and 4% at the 3′ UTR (Moderek *et al.*, 2001). A similar tendency was reported by Nagasaki *et al.* (2005). In rice and *Arabidopsis*, the 5′ UTR is more preferred than the 3′ UTR for AS, except in the case of intron retention, which is mostly located near the 3′ UTR and is the predominant alternative splicing type for plants (Ner-Gaon *et al.*, 2004; Nagasaki *et al.*, 2005). Most of the AS in plants is related to cell growth and/or maintenance, cell communication, development and physiological processes (Zhou *et al.*, 2003). Alternative splicing participates in the regulation of binding properties, intracellular localization, enzymatic activity, stability, and posttranslational modification of many proteins (Stamm et al., 2005).

3.4.1. Nonsense-Mediated mRNA Decay

In general, AS has a strong tendency to preserve the ORF (Moderek *et al.*, 2001; Nagasaki *et al.*, 2005). Preservation of the ORF seems to be a checking mechanism by which the splicing machinery scans the pre-mRNA to reduce the inclusion of pre-mature translation termination codons (PTC)

(Dietz and Kendzoir, 1994; Li *et al*., 2002; Miriami *et al*., 2002; Wachtel *et al*., 2004). Miriami *et al*. (2002) studied 446 protein-coding human genes encompassing 2,311 introns, in which 10,490 latent 5′ SSs were present. The majority of the latent SSs (95.8%), if utilized for splicing, would have introduced PTCs in the resultant mRNA. They suggest a scanning mechanism exists that selects SSs in the pre-mRNA which will maintain the ORF. A latent 5′ SS present in a pre-mRNA from a hamster gene, which conforms better to the consensus sequence than the normal SS (better complementarity to U1, U5, or U6 snRNAs) was not used for splicing in normal growth conditions, but under stress conditions was utilized. The pre-mRNA contains in-frame stop codons upstream of the latent SS (Miriami *et al*., 1994). Elimination of the in-frame stop codons by replacing them with sense codons, promoted the use of the latent SS (Dietz and Kendzoir, 1994).

Besides all the controls to prevent PTCs, still a good percentage of AS results in PTC; however, to prevent the production of truncated polypeptides that could be harmful, a surveillance mechanism called nonsense-mediated mRNA decay (NMD) operates in the cell, degrading mRNPs containing PTCs. Not all mRNPs containing PTCs are degraded by NMD, but only those located more than 50–55 nucleotides upstream of the 3′-most exon-exon junction complex (EJC), measured after splicing (Nagy and Maquat, 1998). The EJC is composed of a group of proteins that is deposited on the pre-mRNA by the splicing machinery and functions as a marker for degradation by NMD (Tange *et al*., 2004). Human EJCs are composed of a series of proteins, such as the up-frameshift, Upfs (hUpf1, hUpf2, hUpf3a, and hUpf3b) proteins involved in pre-mRNA splicing and export, Y14, the mago nashi protein (Magoh), and others (Tange *et al*., 2004; Lejeune and Maquat, 2005). The hUpf proteins assemble a dynamic complex in the nucleus at the mRNA EJC that triggers NMD in the cytoplasm when recognized downstream of a translation termination site (Lykke-Andersen *et al*., 2000). *AtUPF3*, an *Arabidopsis* analog of human Upf proteins, is required for the plant NMD system (Hori and Watanabe, 2005; reviewed in Hilleren and Parker, 1999, Drefuss *et al*., 2002, and Tange *et al*., 2004). The NMD mechanism is, therefore, closely linked to the splicing process. Two mutants of the *waxy* rice mRNA containing a PTC at the same position were differentially degraded by NMD. The fully spliced transcript degraded much faster than the partially spliced, indicating that splicing and NMD are also connected processes in plants (Isshiki *et al*., 2001). Mendell *et al*. (2004), studied the expression of several human genes by microarrays after the suppression of NMD to determine the ones which would be up-regulated, and, therefore, predicted to be substrates for NMD. Among the up-regulated genes, they found transcripts with upstream ORFs in the 5′ UTR, alternatively spliced transcripts with nonsense codons or frameshifts, transcripts containing introns in the 3′ UTR or seleno-cysteine codons, transcripts with ancient transposons, and endogenous retroviruses. All the up-regulated genes had in common the presence of a spliced intron at least 50 nucleotides downstream

of a termination codon. The coupling of AS and NMD provides an efficient mechanism to fine tune gene expression by using the complex AS machinery, while avoiding the potential deleterious consequences associated with the expression of truncated proteins.

NMD is still a mechanism under study, and many nuances are yet to be discovered. Since AS is widespread in basically all organisms, and a large proportion of the alternatively spliced transcripts are candidates for NMD (Green *et al.*, 2003), a question remains: why would the cell produce so many transcripts to later degrade them? In addition to getting rid of "harmful" truncated polypeptides, NMD might be a mechanism to rapidly switch "off" gene expression. Furthermore, NMD usually does not completely down-regulate a transcript. Commonly, 10–30% of the PTC-containing transcripts survive and may produce low levels of physiologically active truncated proteins (Neu-Yilik *et al.*, 2004).

As mentioned previously, the function of the *Arabidopsis RPS4* gene, requires the presence of the full-length and the truncated open reading frames to confer disease resistance to *Pseudomonas syringae* pv *tomato* strain DC3000 (Zhang and Gassmann, 2003). A similar gene from tobacco (*N* gene), which confers resistance to *Tobacco Mosaic Virus* (TMV), produces two transcripts by AS, one short (N_S) and the other long (N_L). The N_S transcript is more prevalent 3h after infection, while N_L, which putatively encodes a truncated protein and lacks 13 of the 14 repeats of the LRR, is prevalent 4–8 h after infection. Only plants containing both forms are completely resistant to TMV (Dinesh-Kumar and Baker, 2000).

3.4.2. Control of Gene Expression

Alternative splicing has the potential to generate a variety of different transcripts from a single gene. Exons can be added or skipped, which could remove a domain or motif or even add a domain. AS at the 3′ or 5′ SS can shorten or increase the length of an exon. Introns can be retained in the 5′ or 3′ UTR modifying the location or stability of the mature mRNA while maintaining the sequence of the cognate protein. Removal of cryptic introns can generate shorter versions of the cognate protein by removing domains or motifs that can affect function. Alternative initiation often changes the protein's N-terminus, while alternative termination can modify the C-terminus of a protein. Furthermore, AS can truncate a protein by inducing a frameshift. Additionally, single or multiple splicing can occur in a single gene, producing a number of different transcripts (Lee and Irizarry, 2003). A study in the human genome identified several genes containing very short segments (< 50 nts) which were products of the retention of introns, and alternative 3′ and 5′ SSs. These segments have secondary structures (non-looping) different from the flanking regions, which change the 3-D structure of the protein, potentially influencing protein function (Wen *et al.*, 2004). Xing *et al.* (2003) studied the effect of AS on transcripts encoding membrane proteins in 1001

human genes. Out of 464 spliced genes encoding transmembrane proteins (TM domain-containing proteins), 188 had their TM domain removed, producing membrane-anchored, ubiquitous, and soluble proteins. The soluble isoforms, in most cases, were localized to a specific tissue.

All these possibilities have great potential to regulate positively or negatively the function of genes in an organism. It is well known that AS is the cause of several diseases in mammals (Cáceres and Kornblihtt, 2002; Green *et al*., 2003; Ghadge *et al*., 2006), but it can also be beneficial (Forrest *et al*., 2004). The power of AS in the regulation of developmental processes is best exemplified by the elaborate controls exercised by the *doublesex* (*dsx)* gene in *Drosophila*. The *dsx* gene produces two differentially spliced transcripts in somatic cells, one specific for males and the other for females. They both encode proteins, which are transcription factors, with DNA binding regions and gender-specific carboxy termini: DSX^M and DSX^F. DSX^M suppresses a group of female-specific genes involved in differentiation, while DSX^F suppress male-specific differentiation genes. Additionally, *dsx* AS is controlled by the products of *transformer* (*tra*), and *transformer-2* (*tra-2*) genes. In males, *tra* is alternatively spliced but produces no functional products; however, in females, *tra* produces a female functional RNA binding protein that associates with the *tra-2* RNA binding protein to direct *dxs* female-specific splicing (MacDougall *et al*., 1995).

Alternative splicing seems to play an important role in disease resistance in plants. Two examples of the involvement of alternatively spliced transcripts in plant disease resistance were provided in previous sections. To further support this idea, in barley, the powdery mildew resistance gene (*Mla 13*) confers resistance to the pathogen, *Blumeria graminis* f. sp. *hordei* (*Bgh*). *Mla 13* expression is dependent on the powdery mildew resistance signaling gene, *Rar1*, and is alternatively spliced to produce five different transcript leader regions, two demonstrating intron retention, and up to three upstream ORFs (uORFs). *Mla 13* transcripts accumulate in much higher levels in incompatible versus compatible interaction with *Bgh*. Moreover, the timing of transcript induction coincides with the attempted penetration of the fungal appressoria in powdery mildew resistant plants. The level of each transcript varied depending on the incubation time with the fungus, indicating that these transcripts show differential response to the recognition stimulus (Halterman *et al*., 2003).

Systemin and its precursor protein, prosystemin, play an important role in the systemic wound response of tomato. Utilization of different 3′ SSs by AS produces *prosysA* and *prosysB*, with *prosysB* accumulating about twice as much as *prosysA* in all tissues that express the prosystemin gene. The transcript *prosysB* differs from *prosysA* by replacement of the residue Arg-57 with Thr-Gly in the non-systemin portion of the protein. Over-expression of *prosysB* in transgenic tomato plants is accompanied by the constitutive expression of defensive genes that are regulated by wounding and systemin. This indicates that the alternatively spliced form is a major prosystemin-encoding transcript in tomato, and active as a signal in the wounding response process (Li and Howe, 2001).

The power of alternative splicing to control gene expression is just beginning to be appreciated in plants. As more studies are carried out, we will continue to document that AS is involved in many different processes that regulate gene expression and guarantee the proper functioning of plant cells. The maize mitochondrial genome does not contain the gene for the ribosomal S14 protein (*rps14*); instead, this gene is integrated into a nuclear gene encoding the iron-sulphur subunit of the respiratory complex II (*Sdh2*). *Sdh2* exon 1 and *rps14* are separated by an intron that is spliced to produce a chimeric *sdh2* (truncated)-*rps14* polypeptide targeted to the mitochondria. In the mitochondria the chimeric protein is subjected to proteolytic digestion and produces a functional *rps14* and a truncated, nonfunctional *sdh2*. Functional *sdh2* is produced by splicing, using the same 5′ SS and a different 3′ SS (Figueroa *et al*., 1999; Figueroa *et al*., 2000). The rice mitochondrial genome does not contain the *sdhB* gene, and has a truncated, nonfunctional *rps14* gene. A single nuclear gene is produced by AS, creating functional *rps14* and *sdhB* genes with mitochondrial signals to localize them in mitochondria (Kubo *et al*., 1999). These examples show how AS in the nucleus can produce two mitochondrial proteins with different functions from the same gene.

Alternative splicing is also a tool that the cell utilizes to promote the regulation of genes in a tissue-specific and developmental manner. A glutamine synthetase gene (GS) from spine dog-fish produces two transcripts by AS, one expressed in liver and the other in neural tissue. The liver isoform contains an additional 95 nucleotides, not present in the neural one, that leads to the formation of a mitochondrial target signal, directing the transcript to the mitochondria. The neural isoform, lacking the 95 bp is maintained in the cytoplasm (Matthews *et al*., 2005). The chloroplast of higher plants produces H_2O_2 scavenger enzymes, ascorbate peroxidases, in two forms, stromal soluble and thylakoid bound. In spinach and tobacco, these enzymes are encoded by the same gene, *ApxII*, which produces four isoforms, one thylakoid bound, and three stromal soluble. The tissue specific splicing is controlled by a splicing enhancer present in exon 12 (Ishikawa *et al*., 1997; Yoshimura *et al*., 2002). All of these examples reinforce the importance of AS as a major point of control of gene expression in living cells.

3.4.3. Alternative Splicing and Stress

The involvement of AS with different kind of stress has been of general interest, and it has been reported to be regulated by major depression in humans (Shimizu *et al*., 1996), by fear (Nijholt *et al*., 2004), subordination (McCobb *et al*., 2003), and psychophysiological stress (Murata *et al*., 2005) in mouse, and mainly by low temperature (Bournay *et al*., 1996) and salt (Shi *et al*., 2002) exposure in plants. There are several reports on low temperature inducing an AS response in plants. Two potato invertase genes, which contain two mini-exons (9 nts long) are correctly spliced under normal conditions, but under cold stress this exon is skipped in leaves and stem. The level of expression of the exon-skipped transcript increases with time of exposure to

cold and is not detected after returning to normal temperature. Wounding does not affect splicing of this gene (Bournay *et al*., 1996). In wheat, exposure to low temperature induces intron retention in two cold up-regulated genes. *7H8* produces four products, the correctly spliced transcript plus three with retained introns, and *6G2* produces the correctly spliced transcript plus a second with intron retention. The cold-dependent splicing is drastically reduced in albino barley, while the wild type barley expresses the additional transcripts as in wheat (Mastrangelo *et al*., 2005). Intron retention promoted by cold stress was also observed in tomatoes (Fung *et al*., in press) and citrus (Robbins and Louzada, 2005). A genome wide analysis in *Arabidopsis*, correlating splicing events with environmental stress shows that AS profiles change with environmental stress, primarily with cold (Lida *et al*., 2004). Besides cold, other stresses that affect alternative splicing in plants include pathogen challenge, water stress, heavy metals, oxidative stress, light, and others (Kazan, 2003). The Bronze2 (*Bz2*) gene encodes a glutathione S transferase in maize, which is involved in anthocyanin biosynthesis. Expression of this gene is induced by cold, ABA, auxin, and heavy metals, particularly cadmium. Treatment of seedlings with cadmium induced a 20-fold increase in the normally spliced transcript, and a 50-fold increase in an intron-containing transcript. Increased expression of the intron-retained transcript does not increase the activity of glutathione S transferase, and may be used to sequester heavy metals (Marrs and Walbot, 1997). Superoxide dismutases are important enzymes involved in protection of the organism against reactive oxygen species. An iron-superoxide dismutase gene (*OsFe-SOD*) from rice produces two isoforms, (*OsFe-SOD* and *OsFe-SODb*). The *OsFe-SODb* transcript retains an intron that cause a frameshift mutation, changing completely the amino acid sequence down-stream of the point of insertion and producing a premature stop codon; however, it still maintains SOD activity. *OsFe-SODb* expression was affected by temperature, with strong inhibition after 4 hours at 4 ˚C (Feng *et al*., 2005). Apetala2/ethylene-responsive factor (AP2/ERF) proteins are AP2 domain-transcription factors and are part of the second largest transcription factor class in plants. A dehydration-responsive factor1 (*HvDRF1*) from barley produces three forms of *HvDRF1* transcripts by AS, and two of them produce AP2 protein. Both transcripts containing the AP2 domain are able to transactivate the expression of a reporter gene driven by a drought inducible promoter, probably mediated by ABA. Expression of *HvDRF1* is up-regulated in vegetative organs by drought, salt and ABA treatment (Xue and Loveridge, 2004).

3.5. Conclusion and Prospectus

As results of new research become available, it is certain that we will continue to see the remarkable contribution of alternative splicing to the regulation of gene expression in plants. The future is wide open, and as we become more knowledgeable about the regulation mechanisms and the regulators

involved in AS in plants we will be able to find potential practical applications for this important process. An area that I believe will became of major interest very soon is the regulation of AS by temperature or other external inducers. The fact that someone can induce a gene to produce two or more different products by changing the environmental temperature or applying certain products may allow us to engineer a plant or suspension cells to produce two kinds of pharmaceutical products, or to have one gene targeting a protein to the roots and another to the fruits, using a non-tissue specific promoter. This is just one example of possible applications among many that surely will come. I recently discovered that a cold responsive gene from citrus produces three additional transcripts when acclimated to low temperature. The transcripts are very unstable at room temperature (unpublished data). Sorting out the biological significance of these observations as they relate to woody plant cold responsiveness will be an exciting challenge for the immediate future.

References

Ali, G.S., Golovkin, M., and Reddy, A.S., 2003, Nuclear localization and *in vivo* dynamics of a plant-specific serine/arginine-rich protein. *Plant J*. **36**:883-893.

Asai, K., Platt, C., and Cochrane, A.,2003, Control of HIV-1 *env* RNA splicing and transport: investigating the role of hnRHP A1 in exon splicing silencer (ESS3a) function, *Virology* **314**:229-242.

Azubel, M., Habib, N., Sperling, R., and Sperling, J., 2006, Native spliceosomes assemble with pre-mRNA to form supraspliceosomes, *J. Mol. Biol*. **356**:955-966.

Barta, A., Sommergruber, K., Thompson, D., Hartmuth, K., Matzke, M.A., and Matzke, A.J.M., 1986, The expression of a nopaline synthase-human growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue, *Plant Mol. Biol*. **6**:347-357.

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.

Bayton, C.E., Potthoff, S.J., McCullough, A.J., and Schuler, M.A., 1996, U-rich tracts enhance 3′ splice site recognition in plant nuclei, *Plant J*. **10**:703-711.

Bournay, A., Hedley, P.E., Maddison, A., Waugh, R., and Marchray, G.C., 1996, Exon skipping induced by cold stress in potato invertase gene script, *Nucl. Acids Res*. **24**:2347-2351.

Cáceres, J.F., and Kornblihtt, A.R., 2002, Alternative splicing: multiple control mechanisms and involvement in human disease, *Trends Genet*. **18**:186-193.

Carle-Urioste, J.C., Brendel, V., and Walbot, V., 1997, A combinatorial role for exon, intron and splice site sequences in splicing in maize, *Plant J*. **11**:1253-1263.

Chan, M., and Yu, S., 1998, The 3′ untranslated region of a rice α-amylase gene functions as a sugar-dependent mRNA stability determinant, *Plant Biol*. **95**:6543-6547.

Chen, I., and Chasin, L.A., 1994, Large exon size does not limit splicing in vivo, *Mol. Cell. Biol*. **14**:2140-2146.

Cheng, W., Taliercio, E.W., and Chourey, P.S., 1999, Sugars modulate an unusual mode of control of the cell-wall invertase gene (*Incw*1) through its 3′ untranslated region in a cell suspension culture of maize, *Plant Biol*. **96**:10512-10517.

Crew, S.L., Liu, H., Mayeda, A., and Krainer, A.R., 1999, Evidence for the function of an exonic enhancer after the first catalytic step of pre-mRNA splicing, *Proc. Natl. Acad. Sci. USA* **96**:10655-10660.

Dietz, H.C. and Kendzoir, Jr, R.J., 1994, Maintenance of an open reading frame as an additional level of scrutiny during splice site selection, *Nat. Genet*. **8**:183-188.

Dinesh-Kumar, S.P., and Baker, B.J.,2000, Alternative spliced N resistance gene transcripts: Their possible role in tobacco mosaic virus resistance, *Proc. Natl. Acad. Sci*. USA **97**:1908-1913.

Ding, J., Zhong, X., Hagopian, J.C., Cruz, M.M., Ghosh, G., Feramisco, J., Adams, J.A. and Fu, X., 2006, Regulated cellular partitioning of SR protein-specific kinases in mammalian cells, *Mol. Biol. Cell* **17**:876-88.

Dirksen, W.P., Hampson, R.K., Sun, Q., and Rottman, F.M., 1994, A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA, *J. Biol. Chem*. **269**:6341-6436.

Dirksen, W.P., Sun, Q., and Rottman, F.M., 1995, Multiple splicing signals control alternative intron retention of bovine growth hormone pre-mRNA, *J. Biol. Chem*. **270**:5346-5352.

Docquier, S., Tillemans, V., Deltour, R., and Motte, P., 2004, Nuclear bodies and compartmentalization of pre-mRNA splicing factors in higher plants, *Chromosoma* **112**:255-266.

Dominski, Z., and Kole, R., 1991, Selection of splice sites in pre-mRNA with short internal exons, *Mol. Cell. Biol*. **11**:6075-6083.

Dreyfuss, G., Kim, V.N., and Kataoka, N., 2002, Messenger-RNA-binding proteins and the messages they carry, *Nature* **3**:195-205.

Du, C., McGuffin, M.E., Dauwalder, B., Rabinow, L., and Mattox, W., 1998, Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in *Drosophila, Mol. Cell* **2**:741-750.

Duncan, P.I., Stojdl, D.F., Marius, R.M., Scheit, K.H., and Bell, J.C., 1998, The Clk2 and Clk3 dual-specificity protein kinases regulate the intranuclear distribution of SR proteins and influence pre-mRNA splicing, *Exp. Cell Res*. **241**:300-308.

Egeland, D.B., Pizanis Stubervant, A., and Schuler, M.A., 1989, Molecular analysis of dicot and monocot small nuclear RNA populations, *Plant Cell* **1**:633-643.

Elrick, L.L., Humphrey, M.B., Cooper, T.A., and Berget, S.M., 1998, A short sequence within two purine-rich enhancers determine 5′ splice site specificity, *Mol. Cell. Biol*. **18**:343-352.

Expert-Bezançon, A., Sureau, A., Durosay, P., Salesses, R., Groeneveld, H., Lecaer, J.P and Marie, J., 2004, hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of β-tropomyosin exon 6B, *J. Biol. Chem*. **279**:38249-38259.

Fang, Y., Hearn, S., and Spector, D.L., 2004, Tissue-specific expression and dynamic organization of SR splicing factors in *Arabidopsis*, *Mol. Biol. Cell* **15**: 2664-2673.

Feng, W., Hogbin, W., Bing, L., and Jinfa, W., 2005, Cloning and characterization of a novel splicing isoform of the iron-superoxide dismutase gene in rice (*Oryza sativa* L.), *Plant Cell Rep*. **24**:734-742.

Figueroa, P., Gómez, 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.

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. Bioph. Res. Comm*. **271**:380-385.

Forrest, S.T., Barringhaus, K.G., Perlegas, D., Hammarskjold, M., and McNamara, C.A., 2004, Intron retention generates a novel Id3 isoform that inhibits vascular lesion formation, *J. Biol. Chem*. **279**:32897-32903.

Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S., Baldi, P.F., and Hertel, K.J., 2005, The architecture of pre-mRNAs affects mechanisms of splice-site pairing, *Proc. Natl. Acad. Sci. USA* **102**:16176-16181.

Fu, X., 1995, The superfamily of arginine/serine-rich splicing factors, *RNA* **1**:663-680.

Fung, R.W.M., Wang, C.Y., Smith, D.L., Gross, K.C., Tao, Y., and Tian M., Characterization of alternative oxidase (AOX) gene expression in response to methyl salicylate and methyl jasmonate pre-treatment and low temperature in tomatoes, *J. Plant Physiol*. In press.

Galante, P.A. F., Sakabe, N.J., Kirschbaum-Slager, N., and De Souza, S.J., 2004, Detection and evaluation of intron retention events in the human transcriptome, *RNA* **10**:757-765.

Galiana-Arnoux, D., Del Gatto-Konczak, F., Gesnel, M., and Breathnach, R., 2005, Intronic UGG repeats coordinate splicing of CD44 alternative exons v8 and v9, Biochem. Bioph. Res. Co. **336**: 667-673.

Gao, H., Gordon-Kamm, W.J., and Lyznik, L.A., 2004, AFS/SF2-like maize pre-mRNA splicing factors affect splice site utilization and theirs transcripts are alternatively spliced, Gene **339**:25-37.

Gauss, K.A., Bunger, P.L., Crawford, M.A., McDermott, B.E., Swearinger, R., Nelson-Overton, L.K., Siemsen, D.W., Kobayashi, S.D., DeLeo, F.R., and Quinn, M.R., 2006, Variants of the 5′-untranslated region of human *NCF2*: expression and translational efficiency, *Gene* **366**:169-179.

George, S., Rochford, J.J., Wolfrum, C., Gray, S.L., Schinner, S.J., Wilson, J.C., Soos, M.A., Murgatroyd, P.R., Williams, R.M., Acerini, C.L., Dunger, D.B., Barford, D., Umpleby, A.M., Wareham, N.J., Davies, H.A., Shafer, A.J., Stoffel, M., O'Rahilly, S., and Barroso, I., 2004, A family with severe insulin resistance and diabetes due to a mutation in AKT2, *Science* **304**:1325-1328.

George-Tellez, R., Segura-Valdez, M.L., Gonzalez-Santos, L., and Jimenez-Garcia, L. F., 2002, Cellular organization of pre-mRNA splicing factors in several tissues: Changes in the uterus by hormone action, *Biol. Cell* **94**:99-108.

Ghadge, G.D., Wang, L., Sharma, K., Monti, A.L., Bindokas, V., Stevens, F.J., and Roos, R.P., 2006, Truncated wild-type SOD1 and FALS-linked mutant SOD1 cause neural cell death in the chick embryo spinal cord, *Neurobiol. Dis*. **21**:194-205.

Gniadkowski, M., Hemmings-Mieszczak, M., Klahre, U., Liu, H. and Filipowicz, W., 1996, Characterisation of intronic uridine-rich sequence elements acting as possible targets for nuclear protein during pre-mRNA splicing in *Nicotiana plumbaginifolia*, *Nucl. Acids Res*. **24**:619-627.

Goodal, G. J., and Filipowicz, W., 1989, The AU-rich sequences present in the introns of plants nuclear pre-mRNAs are required for splicing, *Cell* **58**:473-783.

Goodall, G. J., and Filipowicz, W., 1991, Different effects of nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants, *EMBO J*. **10**:2635-2644.

Grabowski, P.J., and Black D.L., 2001, Alternative RNA splicing in the nervous system, *Prog. Neurobiol*. **65**:289-308.

Graveley, B.R., 2000, Sorting out the complexity of SR protein functions, *RNA* **6**:1197-1211.

Graveley, B.R., Hertel, K.J., and Maniatis, T., 1998, A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers, *EMBO J*. **17**:6747-6756.

Green, R.E., Lewis, B.P., Hillman, R.T., Blanchette, M., Lareau, L. F., Garnett, A.T., Rio, D.C., and Brenner, S.E., 2003, Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and diseased genes, *Bioinformatics* **19**:i118-i121.

Grotewold, E., Athma, P., and Peterson, T., 1991, Alternative spliced products of the maize P gene encode proteins with homology to the DNA-binding domain of *myb*-like transcription factors, *Proc. Natl. Acad. Sci. USA* **88**:4587-4591.

Gupta, S., Wang, B., Stryker, G.A., Zanetti, M.E., and Lal, S.K., 2005, Two novel arginine-serine (SR) proteins in maze are differentially spliced and utilize non-canonical splice sites, *Biochim. Biophys. Acta* **1728**:105-114.

Halterman, D.A., Wei, F., and Wise, R.P., 2003, Powdery mildew-induced *Mla* mRNAs are alternatively spliced and contain multiple upstream open reading frames, *Plant Physiol*. **131**:558-567.

Hampson, R.K., La Follette, L., and Rottman, F. M., 1989, Alternative processing of bovine growth hormone mRNA is influenced by downstream exon sequences, *Mol. Cell. Biol*. **9**:1604-1610.

Han, K., Yeo, G., An, P., Burge, C.B., and Grabowski, P.J., 2005, A combinatorial code for splicing silencing: UAGG and GGGG motifs, *PLoS Biol*. **3**(5):843-860.

Haut, D.D., and Pintel, D.J., 1999, Inclusion of the NS2-specific exon in minute virus of mice mRNA is facilitated by an intronic splicing enhancer that affects definition of the downstream small intron, *Virology* **258**:84-94.

Hawkins, J.D., 1988, A survey on intron and exon lengths, *Nucl. Acids Res*. **16**: 9893-9908.

Hertel , K.J., and Maniatis, T., 1998, The function of multisite splicing enhancers, *Mol. Cell* **1**:449-455.

Hilleren, P., and Parker, R., 1999, Mechanisms of mRNA surveillance in eukaryotes, *Annu. Rev. Genet*. **33**:229-260.

Hori, K., and Watanabe, Y., 2005, UPF3 suppresses aberrant spliced mRNA in *Arabidopsis*, *Plant J*. **43**:530-540.

Howe, K.J., and Ares Jr,M., 1997, Intron self-complementarity enforces exon inclusion in a yeast pre-mRNA, *Proc. Natl. Acad. Sci. USA* **94**:12467-12472.

Humphrey, M.B., Bryan, J., Cooper, T.A., and Berget, S.M., 1995, A 32-nucleotide exon-splicing enhancer regulates the usage of competing 5′ splice sites in a differential internal exon, *Mol. Cell. Biol*. **15**:3979-3988.

Huang, Y., Gattoni, R., Stévenin, J., and Steitz, J.A., 2003, SR splicing factors serve as adapter proteins for TAP-dependent mRNA export, *Mol. Cell* **11**:837-843.

Huang, Y., and Steitz, J.A., 2005, SRprises along a messenger's journey, *Mol. Cell* **17**:613-615.

Hui, J., and Bindereif, A., 2005. Alternative pre-mRNA splicing in the human system: unexpected role of repetitive sequences as regulatory elements, *J. Biol. Chem*. **386**:1265-1271.

Hui, J., Hung, L., Heiner, M., Schreiner, S., Neumüller, N., Reither, G., Haas, S. A., and Bindereif, A., 2005, Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing, *EMBO J*. **24**:1988-1998.

Hui, J., Stangl, K., Lane, W. S., and Bindereif, A., 2003, HnRNP L stimulates splicing of the e*NOS* gene by binding to variable-length CA repeats, *Nat. Struct. Biol*. **10**:33-37.

Ibrahim, E.C., Schaal, T.D., Hertel, K. J., Reed, R. and Maniatis, T, 2005, Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers, *Proc. Natl. Acad. Sci. USA* **102**:5002-5007.

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.

Isshiki, M., Tsumoto, A., and Shimamoto, K., 2006, The serine/arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA, *Plant Cell* **18**:146-158.

Isshiki, M., Yamamoto, Y., Satoh, H., and Shimamoto, K., 2001, Nonsense-mediated decay of mutant *waxy* mRNA in rice, *Plant Physiol*. **125**:1388-1395.

Kalyna, M., and Barta, A., 2004, A plethora of plant serine/arginine-rich proteins: redundancy or evolution of novel gene functions?, *Biochem. Soc. Trans*. **32**:561-564.

Kazan, K., 2003, Alternative splicing and proteome diversity in plants: the tip of the iceberg has just emerged, *Trends Plant Sci*. **8**:468-471.

Keith, B., and Chua, N., 1986, Monocot and dicot pre-mRNAs are processed with different efficiencies in transgenic tobacco, *EMBO J*. **5**:2419-2425.

Ko, C.H., Brendel, V., Taylor, R.D., and Walbot, V., 1998, U-richness is a defining feature of plant introns and may function as an intron recognition signal in maize, Plant Mol. Biol. **36**:573-583.

Krämer, A., 1996, The structure and function of proteins involved in mammalian pre-mRNA splicing, *Annu. Rev. Biochem*. **65**:367-409.

Kubo, N., Harada, K., Hirai, A., and Kodowaki, 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.

Lambermon, M.H.L., Fu, Y., Weiczorek Kirk, D.A., Dupasquier, M., Filipowicz, W., and Lorković, Z., 2002, UBA1 and UBA2, two proteins that interact with UBP1, a multifunctional effector of pre-mRNA maturation in plants, *Mol. Cell. Biol*. **22**:4346-4357.

Lambermon, M.H.L., Simpson, G.G., Wieczorek Kirk, D.A., Hemmings-Mieszczak, M., Klahre, U., and Filipowicz, W., 2000, UBP1, a novel hnRNP-like protein that functions at multiple steps of higher plant nuclear pre-mRNA maturation, *EMBO J*. **19**:1638-1649.

Landsberger, M., Lorković, Z.J., and Oelmüller, R., 2002, Molecular characterization of nucleus-localized RNA-binding proteins from higher plants, *Plant Mol. Biol*. **48**:413-421.

Larkin, P.D., and Park, W.D., 1999, Transcript accumulation and utilization of alternate and non-consensus splice sites in rice granule-bound starch are temperature-sensitive and controlled by a single-nucleotide polymorphism, *Plant Mol. Biol*. **40**:719-727.

Latijnhouwers, M.J., Pairoba, C.F., Brendel, V., Walbot, V., and Carle-Urioste, J.C., 1999, Test of the combinatorial model of intron recognition in a native maize gene, *Plant Mol. Biol*. **41**:637-644.

Lazar, G., and Goodman, H.M., 2000, The *Arabidopsis* splicing factor SR1 is regulated by alternative splicing, *Plant Mol. Biol*. **42**:571-581.

Lee, C.J., and Irizarry, K., 2003, Alternative splicing in the nervous system: and emerging source of diversity and regulation, *Biol. Psychiat*. **54**:771-776.

Lejeune, F., and Maquat, L.E., 2005, Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells, *Curr. Opin. Cell Biol*. **17**:309-315.

Li, B., Wachtel, C., Miriami, E., Yahalom, G., Friedlander, G., Sharon, G., Sperling, R., and Sperling, J., 2002, Stop codons affect 5′ splice site selection by surveillance of splicing, *Proc. Natl. Acad. Sci. USA* **99**:5277-5282.

Li, L., and Howe, G.A., 2001, Alternative splicing of prosystemin pre-mRNA produces two isoforms that are active as signal in the wound response pathway, *Plant Mol. Biol*. **46**:409-419.

Lida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., Konagaya, A., and Shinozaki, K., 2004, Genome-wide analysis of alternative pre-mRNA splicing

in *Arabidopsis thaliana* based on full-length cDNA sequences, *Nucl. Acids Res*. **32**:5096-5103.

Lin, S., Xiao, R., Sun, P., Xu, X., and Fu, X., 2005, Dephosphorylation-dependent sorting of SR splicing factors during mRNP maturation, *Mol. Cell* **20**:413-425.

Liu, X., and Manley, J.L., 2005, Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability, *Cell* **122**:365-378.

Lorković, Z.J., and Barta, A., 2002, Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana*, *Nucl. Acids Res*. **30**:623-635.

Lorković, Z.J., and Barta, A., 2004, Compartmentalization of the splicing machinery in the plant cell nuclei, *Trends Plant Sci*. **9**:565-568.

Lorković, Z.J., Weiczorek Kirk, D.A., Klahre, U., Hemmings-Mieszczak, M., and Filipowicz, W., 2000, RBP45 and RBP47, two oligouridylate-specific hnRNP-like proteins interacting with a poly(A)$^+$ RNA in nuclei of plant cells, *RNA* **6**:1610-1624.

Lou, H., McCullough, A.J., and Schuler, M.A., 1993, 3′Splice site selection in dicot plant nuclei is position dependent, *Mol. Cell. Biol*. **13**:4485-4493.

Lykke-Andersen, J., Shu, M., and Steitz, J.A., 2000, Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon, *Cell* **103**:1121-1131.

Mabon, S.A., and Misteli, T., 2005, Differential recruitment of pre-mRNA splicing factors to alternatively spliced transcripts *in vivo*, *PLoS* Biol. **3**(11):1893-1901.

MacDougall, C., Harbison, D., and Bowness, M., 1995, The developmental consequences of alternate splicing in sex determination and differentiation in *Drosophila*, *Dev. Biol*. **172**:353-376.

Marrs, K.A., and Walbot, V., 1997, Expression and RNA splicing of the maize glutathione s-transferase *Bronze2* gene is regulated by cadmium and other stresses, *Plant Physiol*. **113**:93-102.

Martinez-Contreras, R., Fissette, J., Nasim, F.H., Madden, R., Cordeau, M., and Chabot, B., 2006, Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing, *PLoS Biol*. **4**(2):172-185.

Mastrangelo, A.M., Belloni, S., Barilli, S., Ruperti, B., DiFonzo, N., Stanca, A.M., and Cattivelli, L., 2005, Low temperature promotes intron retention in two *e-cor* genes of durum wheat, *Planta* **221**:705-715.

Matthews, G.D., Gould, R.M., and Vardimon, L., 2005, A single glutamine synthetase gene produces tissue-specific subcellular localization by alternative splicing, *FEBS Lett*. **579**:5527-5534.

Mayeda, A., and Krainer, A.R., 1992, Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2, *Cell* **68**:365-375.

McCarthy, E.M., and Phillips III, J.A., 1998, Characterization of an intron splice enhancer that regulates alternative splicing of human GH pre-mRNA, *Hum. Mol. Genet*. **7**:1491-1496.

McCobb, D.P., Hara, Y., Lai, G., Mahmoud, S.F., and Flügge, G., 2003, Subordination stress alters alternative splicing of the *Slo* gene in tree shrew adrenals, *Horm. Behav*. **43:**180-186.

McConnell, T.S., and Steitz, J.A., 2001, Proximity of the invariant loop of U5 snRNA to the second intron residue during pre-mRNA splicing, *EMBO J*. **20**: 3577-3586.

McCullough, A.J., Lou, H., and Schuler, M.A., 1993, Factors affecting authentic 5′ splice site selection in plant nuclei, *Mol. Cell. Biol*. **13**:1323-1331.

McCullough, A.J., and Schuler, M.A., 1997, Intronic and exonic sequences modulate 5′ splice site selection in plant nuclei, *Nucl. Acids Res*. **25**:1071-1077.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H. C., 2004, Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise, *Nat. Genet*. **36**:1073-1078.

Misteli, T., Caceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F., and Spector, D.L., 1998, Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription *in vivo*, *J. Cell Biol*. **143**:297-307.

Misteli, T., Caceres, J.F., and Spector, D.L., 1997, The dynamics of a pre-mRNA splicing factor in living cells, *Nature* **387**:523-527.

Miriami, E., Motro, U., Sperling, J., and Sperling, R., 2002, Conservation of an open-reading frame as an element affecting 5′ splice site selection, *J. Struct. Biol*. **140**:116-122.

Miriami, E., Sperling, J., and Sperling, R., 1994, Heat shock affects 5′ splice site selection, cleavage and ligation of CAD pre-mRNA in hamster cells, but not its packing in InRNP particles, *Nucl. Acids Res*. **22**:3084-3091.

Moderek, B., Resch, A., Grasso, C., and Lee, C., 2001, Genome-wide detection of alternative splicing in expressed sequences of human genes, *Nucl. Acids Res*. **29**:2850-2859.

Müller, S., Wolpensinger, B., Angenitzki, M., Engel, A., Sperling, J., and Sperling, R., 1998, A supraspliceosome model for large nuclear ribonucleoprotein particles based on mass determination by scanning transmission electron microscopy, *J. Mol. Biol*. **283**:383-394.

Murata, S., Yoshiara, T., Lim, C.R., Sugino, M., Kogure, M., Ohnuki, T., Komurasaki, T., Matsubara, K., 2005, Psychophysiological stress-regulated gene expression in mice, *FEBS Lett*. **579**:2137-2142.

Musci, M.A., Egeland, D.B. and Schuler, M.A., 1992, Molecular comparison of monocot and dicot U1 and U2 snRNAs, *Plant J*. **2**:589-599.

Nagy, E., and Maquat, L.E., 1998, A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance, *Trends in Biochem. Sci*. **23**:198-199.

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.

Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R., and Fluhr, R., 2004, Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*, *Plant J*. **39**:877-885.

Neu-Yilik, G., Gehring, N.H., Hentze, M.W., and Kulozik, A.E., 2004, Nonsense-mediated mRNA decay: from vacuum cleaner to Swiss army knife, *Genome Biol*. **5**:218.1-218.4.

Nijholt, I., Farchi, N., Kye, M., Shoham, S., Verbeure, B., Owen, D., Hochenner, B., Spiess, J., Soreq, H., and Blank, T., 2004, Stress-induced alternative splicing of acetylcholinesterase results in enhanced fear memory and long-term potentiation, *Mol. Psychiatry* **9:**174-183.

Nishihama, R., Banno, H., Kawahara, E., Irie, K., and Machida, Y., 1997, Possible involvement of differential splicing in regulation of the activity of *Arabidopsis* ANP1 that is related to mitogen-activated protein kinase kinase kinases (MAPKKKs), *Plant J*. **12**:39-48.

Nissim-Rafinia, M., and Kerem, B., 2002, Splicing regulation as a potential genetic modifier, *Trends Genet*. **18**:123-127.

Park, J.W., Parisky, K., Celotto, A.M., Reenan, R.A., and Graveley, B.R., 2004, Identification of alternative splicing regulators by RNA interference in *Drosophila*, *Proc. Natl. Acad. Sci. USA* **101:**15974-15979.

Patel, N.A., Chalfant, C.E., Watson, J.E., Wyatt, J.R., Dean, N.M., Eichler, D.C., and Cooper, D.R., 2001, Insulin regulates alternative splicing of protein kinase CβII through a phosphatidylinositol 3-kinase-dependent pathway involving the nuclear serine/arginine-rich splicing factor SRp40, in skeletal muscle cells, *J. Biol. Chem*. **276**:22648-22654.

Patel, N.A., Kaneko, S.,Apostolatos, H.S., Bae, S.S., Watson, J.E., Davidowitz, K., Chappell, D.S., Birnbaum, M.J., Cheng, J.Q., and Cooper, D.R., 2005, Molecular and genetic studies imply Akt-mediated signaling promotes protein kinase CβII alternative splicing *via* phosphorylation of serine /arginine-rich splicing factor SRp40, *J. Biol. Chem*. **280**:14302-14309.

Robberson, B.L., Cote, G.L., and Berget, S. M., 1990, Exon definition may facilitate splice site selection in RNAs with multiple exons, *Mol. Cell. Biol*. **10**:84-94.

Robbins, A.L., and Louzada E. S., 2005, Expression analysis of a cold responsive transcript from trifoliate orange by real-time PCR and RT-PCR, *Plant Cell Rep*. **24**:612-618.

Rothrock, C.R., House, A.E., and Lynch, K.W., 2005, HnRNP L represses exon splicing *via* a regulated exonic splicing silencer, *EMBO J*. **24**:2792-2802.

Sablowski, R.M., and Meyerowitz, E. M., 1998, Temperature-sensitive splicing in the floral homeotic mutant *apetala3-1, Plant Cell*, **10:**1453-1463.

Schumucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E., and Zipursky, S.L., 2000, *Drosophila Dscam* is an axon guidance receptor exhibiting molecular diversity, *Cell* **101**:671-684.

Selvakumar, M., and Helfman, D.M., 1999, Exonic splicing enhancers contribute to the use of both 3′ and 5′ splice site usage of rat β-tropomyosin pre-mRNA, *RNA* **5**:378-394.

Shi, H., Xiong, L., Stevenson, B., Lu, T., and Zhu, J., 2002, The *Arabidopsis salt overly sensitive 4* mutants uncover a critical role for vitamin B6 in plant salt tolerance, *Plant Cell* **14**:575-588.

Shimizu, S., Nomura, K., Ujihara, M., Sakamoto, K., Shibata, H., Suzuki, T., and Demura, H., 1996, An Allele-specific abnormal transcript of the heat shock protein 70 gene in patients with major depression, *Biochem. Biophys. Res. Comm*. **219:** 745-752.

Shin, C., Feng, Y., and Manley, J.L., 2004, Dephosphorylated SRp38 acts as an splicing repressor in response to heat shock, *Nature* **427:**553-558.

Shinozaki, Aahata, K., and Tsukahara, T., 1999, Changes in pre-mRNA splicing factors during neural differentiation in P19 embryonal carcinoma cells, Int. J. Biochem. *Cell Biol*. **31**:1279-1287.

Simpson, C.G., Clark, G., Davidson, D., Smith, P., and Brown, J.W., 1996, Mutation of putative branchpoint consensus sequences in plant introns reduces splicing efficiency, *Plant J*. **9**:369-380.

Simpson, C.G., Jennings, S.N., Clark, G.P., Thow, G., and Brown, J.W.S., 2004, Dual functionality of a plant U-rich intronic sequence element, *Plant J*. **37**:82-91.

Simpson, C.G., McQuade, C., Lyon, J., and Brown, J. W., 1998, Characterization of exon skipping mutants of the *COP*1 gene from *Arabidopsis*, *Plant J*. **15**:125-131.

Simpson, C.G., Thow, G., Clark, G.P., Jennings, S. N., Watters, J.A., and Brown, J.W. S., 2002, Mutational analysis of a plant branchpoint and polypyrimidine tract required for constitutive splicing of a mini-exon, *RNA* **8**:47-56.

Snedden, W.A., and Blumwald, E., 2000, Alternative splicing of a novel diacylglycerol kinase in tomato leads to a calmodulin-binding isoform, *Plant J*. **24**:317-326.

Sperling, R., Koster, A.J., Melamed-Bessudo, C., Rubinstein, A., Angenitzki, M., Berlovitch-Yellin, Z., and Sperling, J., 1997, Three-dimensional image reconstruction of large nuclear RNP (InRNP) particles by automated electron tomography, *J. Mol. Biol*. **267**:570-583.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T.A., and Soreq, H., 2005, Function of alternative splicing, *Gene* **344**:1-20.

Starks, J.M., Cooper, T.A., and Roth, M.B., 1999, The relative strengths of the SR protein-mediated associations of alternative and constitutive exons can influence alternative splicing, *J. Biol. Chem*. **274**:29838-29842.

Swanson, M.S., and Dreyfuss, G., 1988, RNA binding specificity of hnRNP proteins: a subset bind to the 3′ end of introns, *EMBO J*. **7**:3519-3529.

Tange, T., Nott, A., and Moore, M.J., 2004, The ever-increasing complexities of the exon junction complex, *Curr. Opin. Cell Biol*. **16**:279-284 .

Telerico, M., and Berget, S.M., 1994, Intron definition in splicing of small *Drosophila* introns, *Mol. Cell. Biol*. **14**:3434-3445.

Tian, M., and Maniatis, T., 1993, A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA, *Cell* **74**:105-114.

Tillemans, V., Dispa, L., Remacle, C., Collinge, M., and Motte, P., 2005, Functional distribution and dynamics of *Arabidopsis* SR splicing factors in living plant cells, *Plant J*. **41**:567-582.

Unsworth, B.R., Hayman, G.T., Carroll, A., and Lelkes, P.I., 1999, Tissue-specific alternative mRNA splicing of phenylethanolamine N-methyltransferase (PNMT) during development by intron retention, *Int. J. Dev. Neurosci*. **17**:45-55.

van Santen, V.L., and Spritz, R.A., 1987, Splicing of plant pre-mRNAs in animal systems and vice versa, *Gene* **56**:253-265.

Vibe-Pedersen, K., Magnusson, S., and Baralle, F.E., 1986, Donor and acceptor splice signals within an exon of the human fibronectin gene: a new type of differential splicing., *FEBS Let*. **207**:287-291.

Wachtel, C., Li, B., Sperling, J., and Sperling, R., 2004, Stop codon-mediated suppression of splicing is a novel nuclear scanning mechanism not affected by elements of protein synthesis and NMD, *RNA* **10**:1740-1750.

Wang, B., 2005, Genome-wide analysis of splicing related genes and alternative splicing in plants, Iowa State University. PhD dissertation.

Wang, B., and Brendel, V., 2004, The ASRG database: identification and survey of *Arabidopsis thaliana* genes involved in pre-mRNA splicing, *Genome Biol*. **5**:R102.1-R102.22.

Wen, F., Li, F., Xia, H., Lu, X., Zhang, X., and Li, Y., 2004, The impact of very short alternative splicing on protein structures and functions in the human genome, *Trends Genet*. **20**:232-236.

Wiebauer, K., Herrero, J., and Filipowicz, W., 1988, Nuclear pre-mRNA processing in plants: distinct modes of 3′-splice-site selection in plants and animals, *Mol. Cell. Biol*. **8**:2042-2051.

Xing, Y., Xu, Q., and Lee, C., 2003, Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains, *FEBS Lett*. **555**:572-578.

Xue, G., and Loveridge, C. W., 2004, *HvDRF1* is involved in abscisic acid-mediated gene regulation in barley and produces two forms of AP2 transcriptional activators, interacting preferably with a CT-rich element, *Plant J*. **37**:326-339.

Yeo, G., Hoon, S., Venkatesh, B., and Burge, C. B., 2004, Variation in sequence and organization of splicing regulatory elements in vertebrate genes, *Proc. Natl. Acad. Sci. USA* **101**:15700-15705.

Yi, Y., and Jack, T., 1998, An intragenic suppressor of the *Arabidopsis* floral organ identity mutant *apetala* 3-1 functions by suppressing defects in splicing, *Plant Cell* **10**:1465-1477.

Yoshimura, K., Yabuta, Y., Ishikawa, T., and Shigeoka, S., 2002, Identification of a *cis* element for tissue-specific alternative splicing of chloroplast ascorbate peroxidase pre-mRNA in higher plants, *J. Biol. Chem*. **277**:40623-40632.

Yue, B., and Akusjärvi, G., 1999, A downstream splicing enhancer is essential for in vitro pre-mRNA splicing, *FEBS Lett*. **451**:10-14.

Zhuang, Y., and Weiner, A. M., 1986, A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation, *Cell* **46**:827-835.

Zhang, X., and Gassmann, W., 2003, *RPS4*-mediated disease resistance requires the combined presence of *RPS4* transcripts with full-length and truncated open reading frames, *Plant Cell* **15**:2333-2342.

Zhang, Z., Honda, C., Kita, M., Hu, C., Nakayama, M., and Moriguchi, T., 2003, Structure and expression of spermidine synthase genes in apple: two cDNAs are spatially and developmentally regulated through alternative splicing, *Mol. Gen. Genomics* **268**:799-807.

Zhou, Y., Zhou, C., Ye, L., Dong, J., Xu, H., Cai, L., Zhang, L., and Wei, L., 2003, Database and analyses of known alternatively spliced genes in plants, *Genomics* **82**:584-595.

Zhu, J., Mayeda, A. and Krainer, A.R., 2001, Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins, *Mol. Cell* **8**:1351-1361.

4
Messenger RNA 3′-end Formation and the Regulation of Gene Expression

Arthur G. Hunt

4.1. Introduction and An Overview of Polyadenylation

Posttranscriptional control is important in the overall regulation of gene expression in plants. Such control may be manifested through numerous mechanisms such as RNA turnover, transport and/or sequestration, and differential translation. Many of these processes involve, in some manner, the 3′-untranslated region (or UTR) and polyadenylation signal of the gene. It follows that the nature of the 3′-UTR, and choice of polyadenylation site in genes with multiple sites, may play a role in the expression of a gene, with important physiological consequences. Consequently, the processes involved in alternative and regulated RNA polyadenylation, as well as generating 3′-UTR and 3′-end heterogeneity, are of considerable importance in terms of defining the expression profile of the plant genome.

Messenger RNA 3′ end formation and polyadenylation is a ubiquitous feature of gene expression in eukaryotes, and is mediated by an evolutionarily-conserved protein complex of more than 15 subunits (Wahle and Ruegsegger, 1999; Zhao *et al.,* 1999; Proudfoot, 2004). In mammals and yeast, this complex recognizes sequence elements in the 3′ regions of the nascent primary transcript (the polyadenylation signal) and subsequently processes and polyadenylates the RNA. RNA recognition involves a number of RNA-protein interactions and at least three RNA sequence elements. Thus, in mammals, the polyadenylation signal AAUAAA is recognized by the 160 kD subunit of the cleavage and polyadenylation specificity factor (or CPSF160; Murthy and Manley, 1995), the U-rich downstream element by the 64 kD subunit of the cleavage stimulatory factor (or CstF64; MacDonald *et al.,* 1994), and sequences 5′ of AAUAAA by the mammalian cleavage factor I (CFIm; Venkataraman *et al.,* 2005). The human Fip1 protein also has the capacity to bind U-rich RNAs (Kaufmann *et al.,* 2004), and *Drosophila* CPSF30 has a sequence preference for GC-rich RNA (Bai and Tolias, 1998); the relationship of these preferences to the polyadenylation reaction itself is unclear. In yeast, Rna15p (the yeast homolog of CstF64) binds the so-called A-rich positioning element (Gross and Moore, 2001), Hrp1p binds the

efficiency element situated 5′ of the positioning element (Valentini *et al.,* 1999), and any of three different CPSF subunit homologs (Yhh1p, Ydh1p, and Yth1p) binds to sequences near or at the cleavage site (Barabino *et al.,* 2000; Dichtl and Keller, 2001; Dichtl *et al.,* 2002).

Messenger RNA 3′ end formation does not occur apart from other processes; rather, it is coordinated, and possibly occurs in concert, with many other steps in the course of gene expression. Several polyadenylation factor subunits interact with components of the transcription initiation machinery (Calvo and Manley, 2003; Bentley, 2005), and "load" onto the transcribed gene at or near the transcription initiation site (Licatalosi *et al.,* 2002; Venkataraman *et al.,* 2005). The process of mRNA capping has been linked to mRNA 3′ end formation (Flaherty *et al.,* 1997). Other subunits interact with components of the splicing apparatus, and the interplay between splicing and polyadenylation is important for determining (or defining) the 3′-terminal exon in mammalian genes (Niwa *et al.,* 1990; Gunderson *et al.,* 1994; Gunderson *et al.,* 1997; Millevoi *et al.,* 2002). Yet another set of interactions links polyadenylation with transcription termination (Hammell *et al.,* 2002; Buratowski, 2005). Polyadenylation factor subunits also interact with, or play direct roles in, the maturation of cell-cycle-regulated histone mRNAs and snRNAs (Morlando *et al.,* 2002; Nedea *et al.,* 2003; Baillat *et al.,* 2005; Dominski *et al.,* 2005; Kolev and Steitz, 2005). They have been associated with transport of mRNA from the nucleus to the cytoplasm (Brodsky and Silver, 2000; Hammell *et al.,* 2002). Finally, biochemical or genetic associations with DNA repair and chromosome segregation have been reported (Kleiman and Manley, 2001; Wang *et al.,* 2005). These various observations, made in yeast and mammalian systems, reveal both an extensive interconnection between the polyadenylation apparatus and other processes, and a considerable potential for rearrangement and "donation" of parts of the polyadenylation complex to other processes.

Plant genomes possess virtually the entire suite of genes encoding the core polyadenylation apparatus (Belostotsky and Rose, 2005; Q. Q. Li and A. G. Hunt, unpublished observations summarized at http://www.uky.edu/~aghunt00/Table1.html). Many of these have been studied to an extent, and interesting information obtained. The five *Arabidopsis* CPSF subunits exist in a complex in the nucleus (Q. Q. Li, personal communication; Delaney *et al*., 2006). Moreover, one *Arabidopsis* CPSF subunit, AtCPSF100, interacts with at least one poly(A) polymerase (PAP) isoform (Elliott *et al.,* 2003). There appears to be a requirement for tight control of the expression of some of AtCPSF subunit-encoding genes during plant development. For example, plants that carry only one functional allele of the AtCPSF73(II) gene are deficient in female gametogenesis (Xu *et al.,* 2004), while plants that contain an additional copy of the AtCPSF73(I) gene (added as a transgene) are male sterile (Q. Q. Li, personal communication).

Like its mammalian and yeast counterparts, the *Arabidopsis* CPSF30 is an RNA-binding protein (Delaney *et al*., 2006). Interestingly, AtCPSF30 is also

a calmodulin-binding protein, and calmodulin in the presence of calcium inhibits the RNA-binding activity of the protein (Delaney *et al*., 2006). Moreover, mutants in the *Arabidopsis* gene that encodes AtCPSF30 have an oxidative stress-tolerant phenotype (D. F. Falcone, personal communication). These mutants are T-DNA insertions that produce no AtCPSF30, indicating that AtCPSF30 is not absolutely essential for growth. This is unexpected because the yeast counterpart of CPSF30, Yth1p, is an essential protein (Barabino *et al.,* 1997; Barabino *et al.,* 2000). The fact that the *Arabidopsis* CPSF30 gene is not essential suggests that the role of this protein in mRNA 3′ end formation can be fulfilled by other, as yet unidentified proteins. Moreover, the coincidence of the biochemical properties and the phenotype of the mutants implicates calmodulin-regulated RNA binding by AtCPSF30 in the responses of plants to oxidative stress.

The *Arabidopsis* CstF77 and CstF64 proteins also interact, and AtCstF64 is an RNA-binding protein (Yao *et al.,* 2002); both of these properties are also seen in the mammalian and insect polyadenylation complexes (Hatton *et al.,* 2000; Takagaki and Manley, 2000). Curiously, however, the *Arabidopsis* CstF50 homolog does not interact with AtCstF77. In this sense, the plant protein differs from its mammalian counterpart that is part of the heterotrimeric CstF complex and interacts with CstF77 (Takagaki and Manley, 2000). The significance of this difference is not clear. While it may reflect differences in the functioning of CstF50 in polyadenylation, it must be noted that the *Arabidopsis* CstF50 has yet to be linked conceptually with any other polyadenylation factor subunit, and it may be that this protein, while showing a significant degree of sequence similarity to the mammalian CstF50, is actually not associated with mRNA 3′ end formation.

Arabidopsis possesses a Fip1 homolog (AtFip1(V)) that resembles its human counterpart in many ways (Forbes *et al.,* 2006). Thus, the plant protein is an RNA-binding protein and also interacts with PAP. The *Arabidopsis* protein engages in a number of interactions with *Arabidopsis* homologs of CFIm25, CPSF30, CstF77, and PabN, all involving the N-terminus of the *ca* 130 kD protein. The RNA-binding do-main of the protein has a decided preference for poly(G) among the four homopolymers and has a modest but measurable preference for "natural" RNAs that contain a far-upstream element (or FUE; see Section 4.3.1. for a discussion of the elements that make up a plant polyadenylation signal). This suggests that AtFip1(V) may be the subunit responsible for recognizing the FUE.

Plants possesses a probable homolog of the yeast polyadenylation factor subunit, Pfs2p. While the relationship of this protein to other plant polyadenylation factor subunits has not been established, the plant protein (termed FY) has been found to interact with an RNA-binding protein, FCA, that is in turn part of the system that controls the timing of flowering in plants (Simpson *et al.,* 2003). The combination of FY + FCA operates to promote polyadenylation at a novel site in FCA pre-mRNAs, a site that is recognized by FCA itself, thereby establishing a feedback regulatory loop

that yields a balance of full-length and truncated mRNAs. The control of this balance is one contributing factor to the control of flowering time.

Most of the *Arabidopsis* polyadenylation factor subunits (indeed, all of the CPSF and CstF homologs) are encoded by single genes (Q. Q. Li and A. G. Hunt, unpublished observations). In contrast, *Arabidopsis* has four genes whose products are similar to the canonical nuclear PAP of other eukaryotes (Addepalli *et al.,* 2004), and rice possesses six such genes (B. Addepalli *et al*., unpublished results). All four of the *Arabidopsis* PAP genes are expressed, and the recombinant proteins purified from *E. coli* possess poly(A) polymerase activity (Addepalli *et al.,* 2004). Interestingly, all four *Arabidopsis* genes are essential (Meeks, 2005), suggestive of a degree of functional specialization in the products of the four genes. One sort of specialization is apparent from inspection of the amino acid sequences of the four proteins; one of the four (termed PAP(III)) lacks an extended C-terminal domain that, in the other three, includes nuclear localization signals (Addepalli *et al.,* 2004). This is suggestive of a cytoplasmic function for PAP(III). The other three proteins direct GFP exclusively to the nucleus (Meeks, 2005), consistent with the presence of nuclear localization signals in these proteins. Localization studies with PAP(III) were not reported by Meeks, owing to a probable toxicity, even to single cells, of overexpressed protein. As stated above, one of the four PAPs, PAP(II), interacts with AtCPSF100 (Elliott *et al.,* 2003) and another PAP, PAP(IV), interacts with AtFip1(V) (Forbes *et al.,* 2006). Tests of interactions involving other combinations of *Arabidopsis* PAPs have not been reported, and the full scope of interactions between PAPs and other polyadenylation factor subunits is not known. Nonetheless, the studies that have been published reveal an interesting and poorly-understood complexity in the PAP gene family in plants.

Clearly, much remains to be learned about the plant polyadenylation complex. However, the picture that emerges from the studies that have been performed is interesting and has ramifications for the control of gene expression at the level of mRNA 3′ end formation. Twelve *Arabidopsis* polyadenylation factor subunits have been studied to some extent. One of these, AtFip1(V), interacts with five other subunits and with RNA, suggestive of a role as an organizing scaffold in the complex. Two others (AtCPSF30 and AtCstF64) also possess RNA-binding activity. Moreover, AtCPSF160, AtCPSF100, and AtPabN (of which there are three potential isoforms) may also bind RNA, since their mammalian and yeast counterparts do (Wahle, 1991; Murthy and Manley, 1995; Kyburz *et al.,* 2003). In addition, recruitment of the RNA-binding protein FCA by FY, as well as consideration of the nature of the CFIm factor in mammals (which binds RNA and mediates the functioning of elements upstream of AAUAAA in mammals; Venkataraman *et al.,* 2005), raises the possibility that at least two additional RNA-binding activities may participate in 3′ end processing in plants. This plethora of RNA-binding activities lends itself to models involving remodeling of the polyadenylation complex in the course of processing and polyadenylation, or to the existence of different complexes that consist of subsets of the entire

suite of poly-adenylation factor subunits. In either case, possibilities exist for differential action and regulation. Along these lines, four of the subunits (CPSF73(I), CPSF73(II), CPSF30, and FY) have been implicated in regulatory aspects of gene expression in plants (Simpson *et al.,* 2003; Xu *et al.,* 2004; Q. Q. Li, personal communication; D. F. Falcone, personal communication); this reveals that regulation can be effected through components of the 3′ end processing machinery.

4.2. Polymorphism in Polyadenylation Sites in Plants

4.2.1. Regulation via mRNA 3′ end Processing

One means by which regulation may be effected through mRNA 3′ end formation is *via* the use of alternative polyadenylation sites in a single transcription unit. In hypothetical terms, alternative polyadenylation might affect gene expression in various ways. The inherent stability of mRNAs arising from a single transcription unit might depend on the nature of the 3′-UTR, and alternative polyadenylation might be expected to lead to differences in mRNA stability (for example, by including or deleting elements that destabilize RNA). Alternative polyadenylation might also alter the nature of the protein encoded by a gene, due to the inclusion (or not) of different sets of exons in the mature mRNA. The possible scope of the contribution of alternative polyadenylation to the sculpting of the transcriptome is revealed by the observation that as many as half of all mammalian genes may be subject to alternative polyadenylation (Tian *et al.,* 2005; Yan and Marr, 2005).

As a general rule, the population of mRNAs arising from single transcription units in plants is microheterogeneous, with two or more distinct polyadenylation sites typically clustered within tens or hundreds of nucleotides (Hunt, 1994; this phenomenon is illustrated in Fig. 4.1A and termed in this paper as "Class A" alternative polyadenylation). 3′ end microheterogeneity involving as many as 14 different poly(A) sites in a single gene has been reported (Klahre *et al.*, 1995). Hypothetically, mRNAs with different 3′ ends may possess different stabilities, owing to generic nucleotide composition differences and/or the differential presence of specific sequence elements that stabilize or destabilize mRNAs (e.g., RNA 3 in Fig. 4.1A). However, it is not clear that the 3′ end microheterogeneity that is common in plant genes has an impact on RNA stability. While reports of multiple poly(A) sites in plant genes are numerous, the relative stabilities of mRNAs arising from genes that display 3′ end microheterogeneity have not been explicitly examined. It might be expected that the 3′ end microheterogeneity that has been observed is not an absolute determinant of expression levels, since the heterogeneity that is seen could only be detected if the respective mRNAs are stable and abundant enough to be isolated and characterized. However, it is possible that poly(A) sites exist that are not manifest in the set of mRNAs that can be detected in the steady-state population, owing to the inclusion in some RNAs of elements that have destabilizing effects (as illustrated in Fig. 4.1A). Examples

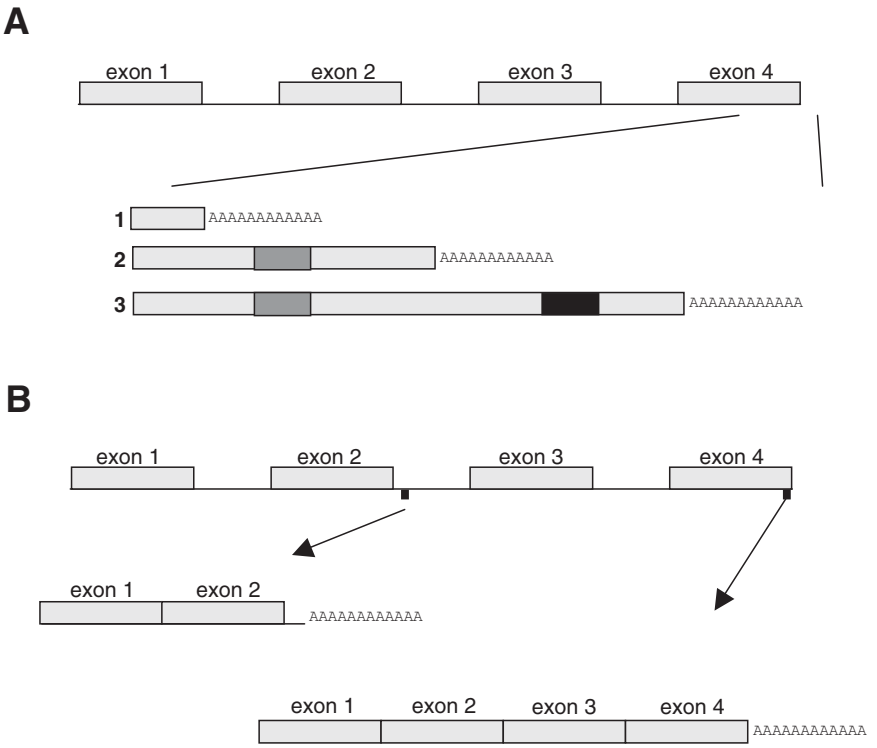


FIGURE 4.1. Themes of alternative polyadenylation in plants. A. 3′ microheterogeneity involving the same 3′-UTR of a gene. A hypothetical gene is shown at the top; exons are depicted with gray boxes, and introns with thin lines. The 3′-extremity is magnified beneath the illustration of the hypothetical gene, showing a typical pattern of microheterogeneity. In this case, 3 possible 3′ ends are shown. RNA 1 is the shortest and, for convenience sake, is a reference. RNA 2 possesses an element (darker gray box) that, as an example, acts as a translation regulatory element. RNA 3 possesses an additional element (black box) that serves as an RNA destabilizing sequence. These are discussed in the text. B. Alternative polyadenylation that results in different exon content of the mRNA. Illustrated is a hypothetical case where a pre-mRNA (top line) can be polyadenylated within an intrn or at the "normal" end of the 3′-terminal exon. Polyadenylation sites are denoted with small black boxes beneath the pre-mRNA. The possibility of alternative polyadenylation sites residing within non-terminal exons is not illustrated, but exists in a formal sense.

of this sort of regulation in plants have not been reported. However, this mode of regulation has been reported in animal systems (Hansen *et al.,* 1996; Miyamoto *et al.,* 1996; Touriol *et al.,* 1999; Caballero *et al.,* 2004); this provides a precedent for the conceptual mode of regulation that is made possible by the extensive 3′ end microheterogeneity that exists in plants.

3′ end microheterogeneity may have other impacts on the process of gene expression. In animals, it has been reported that the translational properties of mRNAs with different poly(A) sites are different (Touriol *et al.,* 2000; Qu

et al., 2002; Simon *et al.,* 2004). Skadsen and Knauer (1995) noted that, in a barley alpha-amylase gene with three different poly(A) sites, one of the three mRNAs was overrepresented in membrane-bound polysomes compared with their overall abundances. This lends credence to the possibility that the translational properties of a set of mRNAs may be affected by their poly(A) site profile, perhaps by differential inclusion of sequences that promote or inhibit translation (illustrated as RNA 2 in Fig. 4.1A).

A more dramatic mode of regulation through 3′ end processing involves the alternative selection of poly(A) sites separated by hundreds or thousands of base pairs along the transcription unit (illustrated in Fig. 4.1B and termed as "Class B"). This form is often accompanied by some type of alternative splicing (*e.g.* intron retention illustrated in Fig. 4.1B). The results of alternative 3′ processing of this sort are different mRNAs that have significantly different protein-coding capacities. For example, the self-incompatibility S locus in *Brassica oleracae* includes two genes (S-locus glycoprotein or SLG and S-receptor kinase or SLK), the transcripts of which have been reported to undergo similar alternative 3′-end processing (Tantikanjana *et al.,* 1993; Giranton *et al.,* 1995). Thus, mRNAs encoded by the so-called SLG gene are processed to give two isoforms differing in size by some 200 nts; the smaller has a 3′ end that lies within an intron that is spliced to produce the larger mRNA (Tantikanjana *et al.,* 1993). An analogous alternative 3′-end processing occurs in the self-incompatibility-associated SLK gene. In this case, the smaller of the two mRNAs contains just the first exon (of seven) and has a 3′ end that lies within the first intron that is spliced to yield the larger mRNA (Giranton *et al.,* 1995). Interestingly, the SLG and SLK genes share substantial nucleotide sequence similarity in their first exons, one that is manifest at the protein level as well. This similarity apparently extends in a functional sense to the first part of the corresponding introns, resulting in the similar patterns of alternative 3′-end processing in the two genes.

In *C. elegans*, alternative polyadenylation of transcripts that encode the so-called DAF-4 receptor kinase (a TGF-beta receptor that is important for larval development) gives rise to truncated mRNAs whose translation products can serve as antagonists of signaling by the full-sized receptor kinase (Gunther and Riddle, 2004). This sort of alternative processing is similar to that seen with the *Brassica* SLK gene involving intron retention leading to the creation of a new polyadenylation site, and suggests parallels for control of signaling in plants. Similarly, a diversity of ethylene receptor-related polypeptides may also be derived by alternative 3′-end processing. Bassett *et al.* (2002) have reported that transcripts encoded by a peach (*Prunus persica*) ethylene receptor (ETR1)-related gene are subject to Class B alternative polyadenylation. The consequence of this is the production of mRNAs that would encode ETR1-related polypeptides that lack the bulk of the C-terminal receiver domain of the receptor. This suggests that, in *P. persica*, a single gene can encode polypeptides analogous to both ETR1 and ERS1 (a related ethylene receptor that lacks the C-terminal receiver domain; Hua *et al.*, 1998). One report of alternative polyadenylation in *Arabidopsis* involves a

gene that encodes a so-called TIR motif-containing protein (Meyers *et al.,* 2002). This instance was identified by massively parallel signature sequence (MPSS) analysis, and involves alternative polyadenylation analogous to that shown in Fig. 4.1B. While the functional significance of these different instances has not been studied in detail, it is tempting to speculate that plants, as animals, utilize alternative polyadenylation to generate a range of receptor-related polypeptides that function together to regulate signaling.

Alternative polyadenylation in plants can involve genes that encode proteins besides receptors. For example, in cotton, alternative polyadenylation gives rise to mRNAs that can encode either a monofunctional lysine-ketoglutarate reductase (LKR) polypeptide or a bifunctional LKR/saccharopine dehydrogenase (SDH) fusion protein (Tang *et al.,* 2002). The monofunctional LKR has a much lower Km for one substrate, lysine, than the bifunctional enzyme, and Tang *et al*. have proposed that this instance of intron retention and alternative polyadenylation is important for enabling efficient flux of lysine catabolism under specific developmental conditions (Tang *et al.,* 2002). In spinach, both the stromal and thylakoid-localized forms of ascorbate peroxidase are derived from alternatively-polyadenylated transcripts from a single gene (Ishikawa *et al.,* 1997) using a form of alternative splicing known as exon skipping (see Chapter 3). The shorter of the two transcripts lacks the 3′-most exon, and with it amino acid sequences that specify a thylakoid-targeting domain, while the longer transcript skips the penultimate exon and instead encodes the 3′-most exon. This is thus another example of the generation of multiple enzyme isoforms from a single gene by a combination of alternative splicing and polyadenylation.

The timing of flowering is also affected by alternative splicing and polyadenylation. Among the regulators of flowering in plants is the RNA-binding protein FCA; this protein is a negative regulator of the flowering suppressor FLC and acts in the so-called autonomous pathway of flowering regulation (Boss *et al.,* 2004). The mRNAs encoded by the FCA gene undergo alternative polyadenylation, and alternative poly(A) site choice yields mRNAs that may encode a full-sized FCA polypeptide or truncated mRNAs that encode smaller, presumably nonfunctional FCA derivatives (Quesada *et al.,* 2003; Simpson *et al.,* 2003). As stated above, FCA binding to its pre-mRNA is integral to the alternative polyadenylation; this mechanism thus establishes a feedback-regulatory loop that helps to control the levels of the FCA.

4.2.2. The Scope of Alternative Polyadenylation in Plants

As indicated in the preceding, alternative poly(A) site choice may play roles in regulated gene expression in plants, and it has the hypothetical capacity to impact gene expression *via* several mechanisms. Consequently, an appreciation of the possible scope and qualitative nature of alternative polyadenylation in plants is needed. Genome-wide surveys have the possibility to identify poly(A) site microheterogeneity on a larger scale, and several studies that address this issue have been published. In a systematic analysis of full-length

cDNAs encoded by *Arabidopsis* chromosome II, Xiao *et al*. (2005) found that 83 of the 187 genes whose derived cDNAs possessed identifiable poly(A) sites encoded RNAs with poly(A) site heterogeneity. Interestingly, while the approach used could have detected alternative poly(A) sites within introns and non-terminal exons, all of the observed microheterogeneity involved Class A-type alternative poly(A) site choice. Most of the heterogeneity was spaced over a region of 10–50 nts of the genomic sequence, but the distance between alternative sites was as great as 543 nts. The number of poly(A) sites in the set of 187 genes ranged from one to five. These observations mirror results obtained in studies of individual plant genes. While analyses of collections of full-length cDNAs may not provide a comprehensive "picture" of the scope of 3′ end microheterogeneity in a transcriptome, the study by Xiao *et al*. provides a lower limit of the percentage (approaching 50%) of *Arabidopsis* genes that may possess multiple polyadenylation sites.

A larger, genome-wide study in *Arabidopsis* suggests that the scope of alternative polyadenylation that leads to changes in the exonic content of derived mRNAs (as in Fig. 4.1B) is substantial (Haas *et al*. 2003). In this study, more than 1300 processing variants were identified amongst a set of some 27,000 annotated protein-coding genes; these data were compiled by analyzing and comparing full-length cDNAs and EST sequences. Among these 1337 genes were 171 with different 3′ ends; this set of genes does not include those that display the sort of 3′ end microheterogeneity (Class A) noted above and in Fig. 4.1A. As is the case with Xiao *et al*., this study does not reveal the complete scope of alternative polyadenylation in plants, as it is limited by the availability of EST and full-length cDNA data, and of the coverage (along specific genes) that EST sequences may provide. However, it indicates that at least 0.5% (roughly) of all *Arabidopsis* genes, and some 10% of those that are alternatively processed, will be subject to Class B alternative polyadenylation.

Similar results have been reported by others (Iida *et al.,* 2004; Nagasaki *et al.,* 2005). Iida *et al*. analyzed some 278,000 cDNA sequences and found, within the 17,130 transcription units that these sequences defined, 458 instances of the use of alternative 3′-terminal exons; this number constituted 21.9% of all instances of alternative RNA processing that were identified, and 3.0% of all transcription units that were represented by more than one sequence in the cDNA collections. Interestingly, there was some bias in the patterns of alternative polyadenylation seen in this study; thus, the alternative 3′ ends seen in libraries prepared from cold-treated plants had a statistically significant tendency to truncate the mRNA, compared with the longer alternative 3′ end. The converse was seen in libraries prepared from siliques and flowers, as well as from stress and hormone-treated plants; in these cases, the bias was against the "shorter" form of the possible transcripts. This study thus suggests that alternative polyadenylation may be regulated globally.

Nagasaki *et al*. (2005) compared alternative transcription initiation and splicing patterns in cDNA collections from six organisms, including *Arabidopsis* and rice. These authors found, in the 15,516 *Arabidopsis* loci that their analysis mapped, 466 instances of alternate poly(A) site usage; this

amounted to 15.2% of the total instances of alternative RNA processing that were identified in the analysis, and 3.0% of all mapped loci. In rice, some 25.4% of all instances of alternative processing involved alternative poly(A) site choice; this corresponded to 3.2% of all mapped loci.

Massively parallel signature sequence studies provide an alternative approach for studying the scope of alternative polyadenylation. This approach has the advantage that it is biased to detect those parts of mRNAs near their 3′ extremities, can sample far more genes in an experiment, and has the potential to detect genes whose expression levels are relatively low. Meyers *et al*. (2004) found MPSS tags within upstream introns and non-terminal exons in approximately 25% of all genes in *Arabidopsis*, suggesting that alternative polyadenylation of the sort that restructures the exonic content of mRNAs is extensive in plants. The scope of this phenomenon is high (25% of all genes in *Arabidopsis*). However, this may be an underestimate; this follows from the realization that a small percentage of the genes identified by Haas *et al*. (2003) as having alternate 3′ ends are not represented in the set of genes possessing sequence tags classified as "Class 5" by Meyers *et al*. ("Class 5" tags are those that map within introns). Given these considerations, it is best to view these approaches as complementary, and to appreciate that these studies likely give underestimates of the true scope of alternative polyadenylation. However, they indicate that alternative polyadenylation is probably a wide-spread feature of gene expression in plants.

Large-scale studies in other plants corroborate the findings obtained with *Arabidopsis*. Thus, in the large-scale sequencing of full-length cDNAs in rice, 4.5% of the expressed genes identified were found to have alternative polyadenylation sites (Kikuchi *et al.,* 2003). In this report, the nature of the variation seen (heterogeneity in 3′-UTRs and terminal exons *vs*. alternative processing involving non-terminal introns and exons) was not detailed. This value is comparable to that obtained in the comparative study of Nagasaki *et al*. (2005) and to those reported for *Arabidopsis* by Iida *et al*. (2004) and Nagasaki *et al*. (2005). However, as discussed in the preceding paragraph, it is likely that this is an underestimate of the scope of alternative polyadenylation, be it Class A or Class B. Regardless, it may be concluded that alternative polyadenylation is a meaningful phenomenon in rice, as well as *Arabidopsis*, and is probably relevant to gene expression in all higher plants.

4.3. Regulation of Polyadenylation in Plants

4.3.1. Recent Developments Regarding the Nature of Polyadenylation Signals in Plants

The widespread occurrence of alternative polyadenylation in plants, and the hypothetical or known instances whereby such events affect gene expression, raise questions regarding the means by which alternative polyadenylation

sites are chosen. Many of these questions are impacted by knowledge of the nature and functioning of polyadenylation signals in plants. Based on studies of several genes in a number of systems (tobacco, maize, wheat), a model of a plant polyadenylation signal was developed (Hunt, 1994; Rothnie, 1996). Briefly, the polyadenylation signal was proposed to be a tripartite unit, consisting of two distinct motifs upstream from the actual site of polyadenylation, as well as the poly(A) site itself (see Fig. 4.2). These motifs, or elements, were termed the "far-upstream element" (abbreviated in this paper as FUE), "near-upstream element" (NUE), and the cleavage/polyadenylation site (CS). The FUE behaved in these experiments as a broad, redundant element whose activity correlated with the occurrence of UG-rich motifs within the respective domains (which could be situated as close as 50 and as far as 150 nts from the poly(A) sites, and may extend over a span of more than 60 nts). The NUE was an A-rich element situated within 30 nts of the poly(A) site; in cases in which the gene possessed an appropriately-situated AAUAAA motif, this motif was found to function as an NUE, leading to the suggestion that the NUE is the functional equivalent of the canonical poly(A) signal AAUAAA. The CS was found to consist of the site of polyadenylation itself, which was noted to correlate with YA sequences embedded within regions of unusually high U content. For two plant genes, the 3′-end microheterogeneity could be attributed to the existence of independent NUEs and CSs clustered in the 3′ region of the genes (MacDonald *et al.,* 1991; Mogen *et al.,* 1992; Li and Hunt, 1995); apparently, many fewer (as few as one) FUEs were able to function in concert with more than one NUE/CS combination to produce the 3′ end profiles that were observed.

Genome-wide analyses of plant 3′ regions corroborate the model that was derived from the earlier studies of a handful of plant genes. Graber *et al.* (1999) analyzed the 3′-UTRs derived from ESTs from rice and *Arabidopsis*. In both organisms, there was a U-richness that extended from the poly(A) site to 150–175 nts upstream from this site. In rice, the relative contributions of the other three bases to the composition of the 3′-UTR were roughly equal. In *Arabidopsis*, there was a greater tendency towards A over G or C. In both organisms, however, A content was dramatically higher between 15 and 30 nts upstream from the poly(A) site. Some 6–8% of the EST sequences contained a match to the motif AAUAAA. However, when a subset (625 of the 4069 total sequences analyzed) were matched with genomic sequences, 14% of the *Arabidopsis* 3′-regions were found to possess AAUAAA. Between 1 and 15 nts upstream from this site, U content was markedly higher. In this study, sequences downstream from the poly(A) were not considered, as most of the plant databases were derived from EST sequences that were not projected onto corresponding genomic DNA sequences.

Loke *et al.* (2005) extended this computational approach to a larger *Arabidopsis* dataset that included sequences 3′, as well as 5′, of the polyadenylation site. In terms of overall base composition, Loke *et al.* reported results identical to those of Graber *et al.* – elevated U content

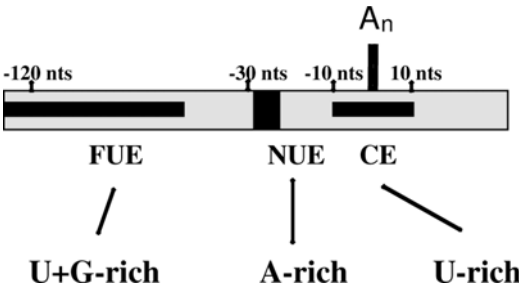


FIGURE 4.2. Structure of a canonical plant polyadenylation signal. The actual cleavage/polyadenylation site is denoted with a thick vertical black line and the notation "A_n". The numbers above the illustration represent characteristic spacings (in nucleotides) of the respective elements. These features are discussed in the text. FUE – far-upstream element; NUE – near-upstream element; CE – cleavage element.

extending up to 175 nts 5′ of the poly(A) site, elevated A content between 15 and 30 nts 5′ of the poly(A) site, and high U content near the poly(A) site. The global trend U>A>G≈C noted by Graber *et al*. was also documented by Loke *et al*. Within the A-rich domain, only 10% of the 8160 sequences analyzed possessed the AAUAAA motif. A number of specific sequence motifs were over-represented in FUEs and NUEs, but no single motif could be assigned a function. The database used by Loke *et al*. included genomic sequences extending 100 nts 3′ of the corresponding poly(A) site; thus, these authors were able to determine that the poly(A) site itself was embedded within what they termed the "cleavage element". This element consists of a region extending from some 10 nts 5′ to 10 nts 3′ of the poly(A) site that had an unusually high U content. Embedded within this U-rich domain is the typical YA dinucleotide that defines the poly(A) site itself. These large-scale genomic analyses mirror, in their results, the outcomes of experimental studies of plant poly(A) signals, and indicate that the model illustrated in Fig. 4.2 is probably applicable to plant genomes as a whole.

4.3.2. Polyadenylation Signals and Alternative 3′ end Processing

From the preceding considerations, a plant polyadenylation signal may be defined as a unit that consists of one FUE combined with a characteristic array of functionally-linked NUEs and CEs. Accordingly, alternative poly(A) site selection may be accomplished by one of two hypothetical mechanisms. The array of 3′ ends defined by a single polyadenylation signal might be controlled, and altered, by changes in the relative efficiencies of the NUEs and/or CEs associated with the signal. Alternatively, different polyadenylation signals (complete combinations of FUE and associated NUEs and CEs) themselves might be selected differentially.

3′ end microheterogeneity is attributable to the presence of multiple NUEs within a single 3′-UTR that act in concert with a shared FUE (MacDonald *et al.,* 1991; Mogen *et al.,* 1992; Li and Hunt, 1995). The expectation follows that differentially-utilized patterns of microheterogeneity should be mediated by factors that recognize the NUE. The factor responsible for this function has not been identified in plants, but may be the counterpart of the mammalian CPSF. This follows from the observations that NUEs, like the mammalian polyadenylation signal AAUAAA, are A-rich (Loke *et al.,* 2005), and that it is the 160 kD subunit of CPSF in mammals that recognizes the AAUAAA element (Murthy and Manley, 1995). However, in yeast, the sub-element that corresponds to the NUE (termed the positioning element) is recognized by Rna15p (Gross and Moore, 2001), the yeast homolog of CstF64. This raises questions as to the identity of the NUE-recognizing factor in plants, and this issue should be considered to be unanswered. Regardless of this, attendant with the expectation that the NUE should be the focus for regulated changes in patterns of 3′ end microheterogeneity is the prospect that different NUEs should have different interactions with the NUE-recognizing factor. These differences may reflect variability in the interaction between the NUE-recognizing subunit and the NUE itself. Alternatively, there may exist different NUE-recognizing subunits amongst the suite of plant polyadenylation factors, each with a somewhat distinctive RNA sequence preference.

Differential selection of polyadenylation signals (as opposed to individual sites associated with a single signal) could, in principle, involve any of the subunits of the polyadenylation apparatus. In animals, alternative polyadenylation of this sort often involves noncanonical polyadenylation sites that, when assayed in isolation, are inherently weaker than canonical (AAUAAA-containing) signals. For example, in the case of the regulated poly(A) site choice that determines immunoglobulin gene expression in B and T cells, elevated levels of a single polyadenylation factor subunit, CstF64, have been correlated with usage of the weaker poly(A) site (Takagaki *et al.,* 1996; Takagaki and Manley, 1998). Elevated levels of CstF64 have also been implicated in alternative poly(A) site choice in lipopolysaccharide-stimulated macrophages (Shell *et al.,* 2005), and a male-specific CstF64 isoform has also been implicated in the functioning of noncanonical poly(A) signals in the male gamete in mammals (Wallace *et al.,* 1999; Wallace *et al.,* 2004). Brown and Gilmartin (2003) noted that elevated levels of the polyadenylation factor CFIm can suppress polyadenylation at AAUAAA-containing signals *in vitro*, while promoting polyadenylation using noncanonical polyadenylation signals. Thus, a mechanism as simple as modulating the absolute levels of a polyadenylation factor or subunit is one means by which alternative polyadenylation is effected in mammals.

Is a similar mechanism operative in plants? The absence in plants of a strongly conserved polyadenylation signal analogous to AAUAAA suggests that plant polyadenylation signals in general may be similar to so-called weak signals in mammals. In this respect, it is possible that regulated alterations in the levels of strategic subsets of plant polyadenylation factor subunits may

alter poly(A) site choice. However, this subject has not been approached experimentally, and the potential mechanisms involved are as many as the numbers of known plant polyadenylation factor subunits.

4.3.3. Involvement of Proteins Apart from Polyadenylation Factor Subunits in 3′ end Processing

An alternative mechanism for alternative polyadenylation is portrayed in the FCA/FY system. In *Arabidopsis*, transcripts encoded by the FCA gene are alternatively polyadenylated, and the choice of polyadenylation site determines whether a full-sized, functional FCA mRNA is produced (Simpson *et al.,* 2003). This poly(A) site choice is in turn regulated by FCA levels. Specifically, full-sized FCA protein has the capacity to bind a sequence located in the third (of 20) intron in the FCA gene. FCA also interacts with FY, a plant homologue of the yeast polyadenylation factor subunit Pfs2p. The combination of interactions of FCA with RNA and FY serves to recruit the polyadenylation apparatus to the third intron, thereby promoting the production of a truncated mRNA. This example thus provides a hypothetical means for alternative polyadenylation in plants, whereby subunits of the polyadenylation complex interact with sequence-specific RNA-binding proteins, leading to polyadenylation at sites (such as introns) that are recognized by these RNA-binding proteins. It also provides a paradigm for linking the polyadenylation apparatus to regulatory signals. In the FCA/FY case, polyadenylation at the "regulated" site is responsive to levels of full-sized FCA protein, levels that in turn are determined by signals that control FCA promoter activity and protein function.

How prevalent might this mode of alternative polyadenylation be in plants? At the present, this question cannot be answered, since the scope of proteins other than polyadenylation factor subunits that interact with the plant polyadenylation apparatus is not known. However, the relatively simple model portrayed by FY/FCA need not be the only mechanism involving proteins other than polyadenylation factor subunits. The *Arabidopsis* homolog of CPSF30 is encoded by a single gene, the transcripts of which are alternatively polyadenylated to yield a small and a large mRNA (Delaney *et al*., 2006; Fig. 4.3). AtCPSF30 itself is encoded by the smaller RNA; purified recombinant AtCPSF30 binds RNA (Delaney *et al*., 2006) and interacts with an *Arabidopsis* Fip1 homolog (Forbes *et al.,* 2006), as well as AtCPSF100 (Q. Q. Li, personal communication). The larger transcript encodes a polypeptide that contains, at its N-terminus, virtually the entire AtCPSF30 protein. The larger C-terminal domain, interestingly, is related to a family of mammalian proteins (typified by the so-called YT521-B protein; (Imai *et al.,* 1998)) that have been associated with pre-mRNA splicing (Stoilov *et al.,* 2002). Thus, an RNA-binding activity that is associated with 3′ end formation also occurs in a potential pre-mRNA splicing factor. This raises the possibility that the balance between splicing and polyadenylation involving introns containing

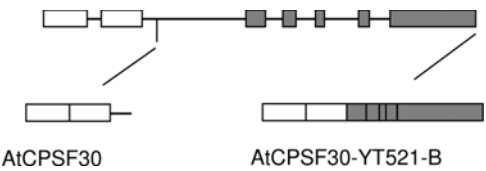


FIGURE 4.3. Structure of the *Arabidopsis* gene that encodes AtCPSF30; exons are depicted with boxes, introns with thin horizontal lines, and the polyadenylation site within the second intron with a thin vertical line. The mRNA products of polyadenylation within the second intron (left) and at the end of the 3′-terminal exon (right) are shown beneath the depiction. AtCPSF30-related exons are shown with white boxes, and YT521-B-related exons with gray boxes. This illustration summarizes results described by Delaney *et al.* (2006).

alternative poly(A) sites might be determined by the nature of the form of CPSF30-containing polypeptides that recognize these introns and bring pre-mRNA splicing into the process of alternative polyadenylation. It should be noted, however, that this possibility is largely hypothetical, as no clear role for YT521-B-related proteins in any process in plants has been established.

4.3.4. Linking Polyadenylation to Environmental and Developmental Cues

If it is to be a regulatory mechanism, alternative polyadenylation must be responsive to developmental and environmental cues. Regulation of poly(A) site choice may be accomplished by changing the levels and activities of subunits of the polyadenylation apparatus itself. The roles of varying CstF64 levels in alternative polyadenylation in mammals provides one precedent for this sort of regulation. It is conceivable that posttranslational modification may play a role; various polyadenylation factors have been shown to be phosphorylated (Colgan *et al.,* 1996; Bond *et al.,* 2000; Mizrahi and Moore, 2000; He and Moore, 2005), ubiquitinated (Mizrahi and Moore, 2000), and methylated (Valentini *et al.,* 1999; Boisvert *et al.,* 2003). While these modifications have not been associated with alternative poly(A) site choice, their impacts on the activities of the respective subunits have been defined, and it is possible that they might contribute to alternative poly(A) site choice. Such modifications are often linked with signaling events, thus providing a possible link to environmental and developmental cues. However, the corresponding plant proteins have not been studied in this regard, and it is not possible to determine the relevance of these mechanisms to gene expression in plants.

A different means for linking the activity of the polyadenylation complex to regulatory signals is revealed in studies of AtCPSF30. As mentioned above, AtCPSF30 interacts with calmodulin, and the RNA binding activity of the protein is inhibited by calmodulin in a calcium-dependent fashion (Delaney *et al*., 2006). This property should render responsive to

calcium/calmodulin polyadenylation signals whose function is mediated by AtCPSF30. Thus, processes that trigger calmodulin signaling events should affect the activity of the polyadenylation complex by inhibiting (if even transiently) the activity of AtCPSF30. It is interesting to recall that AtCPSF30 is not essential. This implies that there are polyadenylation complexes in plants that can function without AtCPSF30, and thus that can function when CPSF30 is inhibited by calmodulin. If these different complexes also have different polyadenylation signal preferences, then the possibility arises that the effects of calmodulin on AtCPSF30 might be manifest as a shift in poly(A) site choice in genes whose signals are recognized by AtCPSF30-containing complexes. These latter points are speculation, of course, since there are no studies that address these (or other) possibilities. Nonetheless, the inhibition of AtCPSF30 by calmodulin provides a possible paradigm for regulated polyadenylation in plants; polyadenylation factor subunits (such as AtCPSF30) may be directly impacted by signaling molecules in ways that alter the functioning of the polyadenylation apparatus.

The connections between stimuli and 3′-end processing may be less direct, mediated not by direct modulation of the core polyadenylation apparatus, but rather by auxiliary factors that act in processing in gene-specific manners. An example of this sort of mechanism is represented by the differential processing of FCA-encoded transcripts. This differential processing reflects mechanisms that control FCA protein levels and activities, and the cues for such alternative processing are those that serve to alter the activities of the FCA protein. One such cue is the hormone abscisic acid (ABA). In addition to its interaction with FY, FCA also binds ABA in a saturable manner with kinetics consistent with action as a hormone receptor (Razem *et al*., 2006). In the presence of ABA, the binding of FCA to FY is inhibited, as is the alternative processing that down-regulates the production of full-length FCA-encoding mRNAs (Razem *et al*., 2006). Thus, FCA provides a molecular link between ABA-mediated signaling, alternative polyadenylation, and the regulation of flowering time. This example demonstrates that the regulation of poly(A) site choice may be accomplished by regulation of the levels or activities of auxiliary, non-polyadenylation factor subunits that act, as does FCA, to recruit the polyadenylation machinery to noncanonical poly(A) sites in a transcript. The scope of this regulatory mechanism in plants is not known, but will be revealed as a greater understanding of the plant polyadenylation apparatus is attained.

Acknowledgements. The author thanks Drs. Quinn Li, Deane Falcone, Kalyan Vemaraju, and Blake Meyers for making available preprints and unpublished observations that helped in crafting this review, and Carol Von Lanken, Drs. Suryadevara Rao, and Balasubrahmanyam Addepalli for reviewing the manuscript. Research performed in the author's laboratory and cited herein was supported by USDA NRI grant #99-35301-7904 and NSF grant MCB-0313472.

References

Addepalli, B., Meeks, L.R., Forbes, K.P. and Hunt, A.G., 2004, Novel alternative splicing of mRNAs encoding poly(A) polymerases in *Arabidopsis, Biochim. Biophys. Acta* **1679**:117-128.

Bai, C. and Tolias, P. P.,1998, *Drosophila* clipper/CPSF 30K is a post-transcriptionally regulated nuclear protein that binds RNA containing GC clusters, *Nucl. Acids Res*. **26**:1597-1604.

Baillat, D., Hakimi, M.A., Naar, A.M., Shilatifard, A., Cooch, N. and Shiekhattar, R., 2005, Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II, *Cell* **123**:265-276.

Barabino, S.M., Hubner, W., Jenny, A., Minvielle-Sebastia, L. and Keller, W., 1997, The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins, *Genes Dev*. **11**: 1703-1716.

Barabino, S. M., Ohnacker, M. and Keller, W., 2000, Distinct roles of two Yth1p domains in 3′-end cleavage and polyadenylation of yeast pre-mRNAs, *EMBO J*. **19**:3778-3787.

Bassett, C.L., Artlip, T.S., and Callahan, A.M., 2002, Characterization of the peach homologue of the ethylene receptor, PpETR1, reveals some unusual features regarding transcript processing, *Planta* **251**:679-688.

Belostotsky, D.A. and Rose, A.B., 2005, Plant gene expression in the age of systems biology: integrating transcriptional and post-transcriptional events, *Trends Plant Sci*. **10**:347-353.

Bentley, D.L., 2005, Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors, *Curr. Opin. Cell Biol*. **17**:251-256.

Boisvert, F.M., Cote, J., Boulanger, M.C. and Richard, S., 2003, A proteomic analysis of arginine-methylated protein complexes, *Mol. Cell. Proteomics* **2**:1319-1330.

Bond, G.L., Prives, C. and Manley, J.L., 2000, Poly(A) polymerase phosphorylation is dependent on novel interactions with cyclins, *Mol. Cell Biol*. **20**:5310-5320.

Boss, P.K., Bastow, R.M., Mylne, J.S. and Dean, C., 2004, Multiple pathways in the decision to flower: enabling, promoting, and resetting, *Plant Cell* **16 Suppl**:S18-31.

Brodsky, A.S. and Silver, P.A., 2000, Pre-mRNA processing factors are required for nuclear export, *RNA* **6**:1737-1749.

Brown, K.M. and Gilmartin, G.M., 2003, A mechanism for the regulation of pre-mRNA 3′ processing by human cleavage factor Im, *Mol. Cell* **12**: 1467-1476.

Buratowski, S., 2005, Connections between mRNA 3′ end processing and transcription termination, *Curr. Opin. Cell Biol*. **17**:257-261.

Caballero, J.J., Giron, M.D., Vargas, A.M., Sevillano, N., Suarez, M.D. and Salto, R., 2004, AU-rich elements in the mRNA 3′-untranslated region of the rat receptor for advanced glycation end products and their relevance to mRNA stability, *Biochem. Biophys. Res. Commun*. **319**:247-255.

Calvo, O. and Manley, J.L., 2003, Strange bedfellows: polyadenylation factors at the promoter, *Genes Dev*. **17**:1321-1327.

Colgan, D.F., Murthy, K.G., Prives, C. and Manley, J.L., 1996, Cell-cycle related regulation of poly(A) polymerase by phosphorylation, *Nature* **384**:282-285.

Delaney, K.J., Xu, R., Zhang, J., Li, Q.Q., Yun, K.-Y., Falcone, D.F., and Hunt, A.G., 2006, Calmodulin interacts with and regulates the RNA binding activity of an *Arabidopsis* polyadenylation factor subunit, *Plant Physiol*. in press.

Dichtl, B., Blank, D., Sadowski, M., Hubner, W., Weiser, S. and Keller, W., 2002, Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination, *EMBO J*. **21**:4125-4135.

Dichtl, B. and Keller, W., 2001, Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor, *EMBO J*. **20**:3197-3209.

Dominski, Z., Yang, X.C. and Marzluff, W.F., 2005, The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing, *Cell* **123**:37-48.

Elliott, B.J., Dattaroy, T., Meeks-Midkiff, L.R., Forbes, K.P. and Hunt, A.G., 2003, An interaction between an *Arabidopsis* poly(A) polymerase and a homologue of the 100 kDa subunit of CPSF, *Plant Mol. Biol*. **51**:373-384.

Flaherty, S.M., Fortes, P., Izaurralde, E., Mattaj, I.W. and Gilmartin, G.M., 1997, Participation of the nuclear cap binding complex in pre-mRNA 3′ processing, *Proc. Natl. Acad. Sci*. USA **94**:11893-11898.

Forbes, K.P., Addepalli, B. and Hunt, A.G., 2006, An *Arabidopsis* Fip1 homolog interacts with RNA and provides conceptual links with a number of other polyadenylation factor subunits, *J. Biol. Chem*. **281**:176-186.

Giranton, J.L., Ariza, M.J., Dumas, C., Cock, J.M. and Gaude, T., 1995, The S locus receptor kinase gene encodes a soluble glycoprotein corresponding to the SKR extracellular domain in *Brassica oleracea*, *Plant J*. **8**: 827-834.

Graber, J.H., Cantor, C.R., Mohr, S.C. and Smith, T.F., 1999, *In silico* detection of control signals: mRNA 3′-end-processing sequences in diverse species, *Proc. Natl. Acad. Sci*. USA **96**:14055-14060.

Gross, S. and Moore, C.L., 2001, Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3′-end formation, *Mol. Cell Biol*. **21**:8045-8055.

Gunderson, S.I., Beyer, K., Martin, G., Keller, W., Boelens, W.C. and Mattaj, L.W., 1994, The human U1A snRNP protein regulates polyadenylation *via* a direct interaction with poly(A) polymerase, *Cell* **76**:531-541.

Gunderson, S.I., Vagner, S., Polycarpou-Schwarz, M. and Mattaj, I.W., 1997, Involvement of the carboxyl terminus of vertebrate poly(A) polymerase in U1A autoregulation and in the coupling of splicing and polyadenylation, *Genes Dev*. **11**: 761-773.

Gunther, C.V. and Riddle, D.L., 2004, Alternative polyadenylation results in a truncated daf-4 BMP receptor that antagonizes DAF-7-mediated development in *Caenorhabditis elegans*, *J. Biol. Chem*. **279**:39555-39564.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., Salzberg, S.L. and White, O., 2003, Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies, *Nucl. Acids Res*. **31**:5654-5666.

Hammell, C.M., Gross, S., Zenklusen, D., Heath, C.V., Stutz, F., Moore, C. and Cole, C.N., 2002, Coupling of termination, 3′ processing, and mRNA export, *Mol. Cell Biol*. **22**:6441-6457.

Hansen, W.R., Barsic-Tress, N., Taylor, L. and Curthoys, N.P., 1996, The 3′-nontranslated region of rat renal glutaminase mRNA contains a pH-responsive stability element, *Am. J. Physiol*. **271**:F126-131.

Hatton, L.S., Eloranta, J.J., Figueiredo, L.M., Takagaki, Y., Manley, J.L. and O'Hare, K., 2000, The *Drosophila* homologue of the 64 kDa subunit of cleavage stimulation factor interacts with the 77 kDa subunit encoded by the suppressor of forked gene," *Nucl. Acids Res*. **28**:520-526.

He, X. and Moore, C., 2005, Regulation of yeast mRNA 3′ end processing by phosphorylation, *Mol. Cell* **19**:619-629.

Hua, J., Sakai, H., Nourizadeh, S., Chen, Q.G., Bleeker, A.B., Ecker, J.R., and Meyerowitz, E.M., 1998, EIN4 and ERS2 are members of the putative ethylene receptor gene family in *Arabidopsis*, *Plant Cell* **10**:1321–1332.

Hunt, A.G., 1994, Messenger RNA 3′ End Formation in Plants, *Ann. Rev. Plant Physiol. Plant Mol. Biol*. **45**:47-60.

Iida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., Konagaya, A. and Shinozaki, K., 2004, Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences, *Nucl. Acids Res*. **32**: 5096-5103.

Imai, Y., Matsuo, N., Ogawa, S., Tohyama, M. and Takagi, T., 1998, Cloning of a gene, YT521, for a novel RNA splicing-related protein induced by hypoxia/reoxygenation, *Brain Res. Mol. Brain Res*. **53**:33-40.

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.

Kaufmann, I., Martin, G., Friedlein, A., Langen, H. and Keller, W., 2004, Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase, *EMBO J*. **23**:616-626.

Kikuchi, S., *et al*.., 2003, Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice, *Science* **301**:376-379.

Klahre, U., Hemmings-Mieszczak, M. and Filipowicz, W., 1995, Extreme heterogeneity of polyadenylation sites in mRNAs encoding chloroplast RNA-binding proteins in *Nicotiana plumbaginifolia*, *Plant Mol. Biol*. **28**:569-574.

Kleiman, F.E. and Manley, J.L., 2001, The BARD1-CstF-50 interaction links mRNA 3′ end formation to DNA damage and tumor suppression, *Cell* **104**:743-753.

Kolev, N. G. and Steitz, J.A., 2005, Symplekin and multiple other polyadenylation factors participate in 3′-end maturation of histone mRNAs, *Genes Dev*. **19**:2583-2592.

Kyburz, A., Sadowski, M., Dichtl, B. and Keller, W., 2003, The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3′-end formation, *Nucl. Acids Res*. **31**:3936-3945.

Li, Q. and Hunt, A.G., 1995, A near-upstream element in a plant polyadenylation signal consists of more than six nucleotides, *Plant Mol. Biol*. **28**:927-934.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J. B. and Bentley, D.L., 2002, Functional interaction of yeast pre-mRNA 3′ end processing factors with RNA polymerase II, *Mol. Cell* **9**:1101-1111.

Loke, J.C., Stahlberg, E.A., Strenski, DG., Haas, B.J., Wood, P.C. and Li, Q.Q., 2005, Compilation of mRNA polyadenylation signals in *Arabidopsis* revealed a new signal element and potential secondary structures, *Plant Physiol*. **138**:1457-1468.

MacDonald, C.C., Wilusz, J. and Shenk, T., 1994, The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location, *Mol. Cell Biol*. **14**:6647-6654.

MacDonald, M.H., Mogen, B.D. and Hunt, A.G., 1991, Characterization of the polyadenylation signal from the T-DNA-encoded octopine synthase gene, *Nucl. Acids Res*. **19**:5575-5581.

Meeks, L.R., 2005, Isolation and Characterization of the Four *Arabidopsis thaliana* Poly(A) Polymerase Genes, *Plant Physiology*, Lexington, KY, University of Kentucky, **Ph.D.**

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.

Meyers, B.C., Vu, T.H., Tej, S.S., Ghazal, H., Matvienko, M., Agrawal, V., Ning, J. and Haudenschild, C. D., 2004, Analysis of the transcriptional complexity of *Arabidopsis thaliana* by massively parallel signature sequencing, *Nat. Biotechnol*. **22**:1006-1011.

Millevoi, S., Geraghty, F., Idowu, B., Tam, J.L., Antoniou, M. and Vagner, S., 2002, A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3′-end processing, *EMBO Rep*. **3**:869-874.

Miyamoto, S., Chiorini, J.A., Urcelay, E. and Safer, B., 1996, Regulation of gene expression for translation initiation factor eIF-2 alpha: importance of the 3′ untranslated region, *Biochem. J*. **315 (Pt 3)**:791-798.

Mizrahi, N. and Moore, C., 2000, Posttranslational phosphorylation and ubiquitination of the *Saccharomyces cerevisiae* Poly(A) polymerase at the S/G(2) stage of the cell cycle, *Mol. Cell Biol*. **20**:2794-2802.

Mogen, B.D., MacDonald, M.H., Leggewie, G. and Hunt, A.G., 1992, Several distinct types of sequence elements are required for efficient mRNA 3′ end formation in a pea rbcS gene, *Mol. Cell Biol*. **12**:5406-5414.

Morlando, M., Greco, P., Dichtl, B., Fatica, A., Keller, W. and Bozzoni, I., 2002, Functional analysis of yeast snoRNA and snRNA 3′-end formation mediated by uncoupling of cleavage and polyadenylation, *Mol. Cell Biol*. **22**:1379-1389.

Murthy, K.G. and Manley, J.L., 1995, The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3′-end formation, *Genes Dev*. **9**:2672-2683.

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.

Nedea, E., He, X., Kim, M., Pootoolal, J., Zhong, G., Canadien, V., Hughes, T., Buratowski, S., Moore, C.L. and Greenblatt, J., 2003, Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3′-ends, *J. Biol. Chem*. **278**: 33000-33010.

Niwa, M., Rose, S.D. and Berget, S.M., 1990, In vitro polyadenylation is stimulated by the presence of an upstream intron, *Genes Dev*. **4**:1552-1559.

Proudfoot, N., 2004, New perspectives on connecting messenger RNA 3′ end formation to transcription, *Curr. Opin. Cell Biol*. **16**:272-278.

Qu, X., Qi, Y. and Qi, B., 2002, Generation of multiple mRNA transcripts from the novel human apoptosis-inducing gene hap by alternative polyadenylation utilization and the translational activation function of 3′ untranslated region, *Arch. Biochem. Biophys*. **400**:233-244.

Quesada, V., Macknight, R., Dean, C. and Simpson, G.G., 2003, Autoregulation of FCA pre-mRNA processing controls *Arabidopsis* flowering time, *EMBO J*. **22**: 3142-3152.

Razem, F.A., El-Kereamy, A., Abrams, S.R., and Hill, R.D., 2006, The RNA-binding protein FCA is an abscisic acid receptor, *Nature* **439**:290-294.

Rothnie, H.M., 1996, Plant mRNA 3′-end formation, *Plant Mol. Biol*. **32**:43-61.

Shell, S. A., Hesse, C., Morris, S.M., Jr. and Milcarek, C., 2005, Elevated levels of the 64-kDa cleavage stimulatory factor (CstF-64) in lipopolysaccharide-stimulated macrophages influence gene expression and induce alternative poly(A) site selection, *J. Biol. Chem*. **280**:39950-39961.

Simon, P., Schott, K., Williams, R.W. and Schaeffel, F., 2004, Posttranscriptional regulation of the immediate-early gene EGR1 by light in the mouse retina, *Eur. J. Neurosci*. **20**:3371-3377.

Simpson, G.G., Dijkwel, P.P., Quesada, V., Henderson, I. and Dean, C., 2003, FY is an RNA 3′ end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition, *Cell* **113**:777-787.

Skadsen, R.W. and Knauer, N.S., 1995, Alternative polyadenylation generates three low-pI alpha-amylase mRNAs with differential expression in barley, *FEBS Lett*. **361**:220-224.

Stoilov, P., Rafalska, I. and Stamm, S., 2002, YTH: a new domain in nuclear proteins, *Trends Biochem. Sci*. **27**:495-497.

Takagaki, Y. and Manley, J.L., 1998, Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation, *Mol. Cell* **2**:761-771.

Takagaki, Y. and Manley, J.L., 2000, Complex protein interactions within the human polyadenylation machinery identify a novel component, *Mol. Cell Biol*. **20**: 1515-1525.

Takagaki, Y., Seipelt, R.L., Peterson, M.L. and Manley, J.L., 1996, The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation, *Cell* **87**:941-952.

Tang, G., Zhu, X., Gakiere, B., Levanony, H., Kahana, A. and Galili, G., 2002, The bifunctional LKR/SDH locus of plants also encodes a highly active monofunctional lysine-ketoglutarate reductase using a polyadenylation signal located within an intron, 2002, *Plant Physiol*. **130**:147-154.

Tantikanjana, T., Nasrallah, M.E., Stein, J.C., Chen, C.H. and Nasrallah, J.B., 1993, An alternative transcript of the S locus glycoprotein gene in a class II pollen-recessive self-incompatibility haplotype of *Brassica oleracea* encodes a membrane-anchored protein, *Plant Cell* **5**:657-666.

Tian, B., Hu, J., Zhang, H. and Lutz, C.S., 2005, A large-scale analysis of mRNA polyadenylation of human and mouse genes, *Nucl. Acids Res*. **33**:201-212.

Touriol, C., Morillon, A., Gensac, M.C., Prats, H. and Prats, A.C., 1999, Expression of human fibroblast growth factor 2 mRNA is post-transcriptionally controlled by a unique destabilizing element present in the 3′-untranslated region between alternative polyadenylation sites, *J. Biol. Chem*. **274**:21402-21408.

Touriol, C., Roussigne, M., Gensac, M.C., Prats, H. and Prats, A.C., 2000, Alternative translation initiation of human fibroblast growth factor 2 mRNA controlled by its 3′-untranslated region involves a Poly(A) switch and a translational enhancer, *J. Biol. Chem*. **275**:19361-19367.

Valentini, S.R., Weiss, V.H. and Silver, P.A., 1999, Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3′-end formation, *RNA* **5**:272-280.

Venkataraman, K., Brown, K.M. and Gilmartin, G.M., 2005, Analysis of a non-canonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition, *Genes Dev*. **19**:1315-1327.

Wahle, E., 1991, A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation, *Cell* **66**:759-768.

Wahle, E. and Ruegsegger, U., 1999, 3′-End processing of pre-mRNA in eukaryotes, *FEMS Microbiol. Rev*. **23**:277-295.

Wallace, A.M., Dass, B., Ravnik, S.E., Tonk, V., Jenkins, N.A., Gilbert, D. J., Copeland, N.G. and MacDonald, C.C., 1999, Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells, *Proc. Natl. Acad. Sci*. USA **96**:6763-6768.

Wallace, A.M., Denison, T.L., Attaya, E.N. and MacDonald, C.C., 2004, Developmental distribution of the polyadenylation protein CstF-64 and the variant tauCstF-64 in mouse and rat testis, *Biol. Reprod*. **70**:1080-1087.

Wang, S.W., Asakawa, K., Win, T.Z., Toda, T. and Norbury, C.J., 2005, Inactivation of the pre-mRNA cleavage and polyadenylation factor Pfs2 in fission yeast causes lethal cell cycle defects, *Mol. Cell Biol*. **25**:2288-2296.

Xiao, Y.L., Smith, S.R., Ishmael, N., Redman, J.C., Kumar, N., Monaghan, E.L., Ayele, M., Haas, B.J., Wu, H.C. and Town, C.D., 2005, Analysis of the cDNAs of hypothetical genes on *Arabidopsis* chromosome 2 reveals numerous transcript variants, *Plant Physiol*. **139**:1323-1337.

Xu, R., Ye, X. and Quinn Li, Q., 2004, AtCPSF73-II gene encoding an *Arabidopsis* homolog of CPSF 73 kDa subunit is critical for early embryo development, *Gene* **324**:35-45.

Yan, J. and Marr, T.G., 2005, Computational analysis of 3′-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat, *Genome Res*. **15**:369-375.

Yao, Y., Song, L., Katz, Y. and Galili, G., 2002, Cloning and characterization of *Arabidopsis* homologues of the animal CstF complex that regulates 3′ mRNA cleavage and polyadenylation, *J. Exp. Bot*. **53**:2277-2278.

Zhao, J., Hyman, L. and Moore, C., 1999, Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis, *Microbiol. Mol. Biol. Rev*. **63**:405-44.

5
An Overview of Small RNAs

Jean-Michel Hily and Zongrang Liu

5.1. Small RNAs: Targets and Mechanisms

For a long time, it was thought that RNA molecules were essential molecules only for carrying the genetic information from the nucleus to the cytoplasm and helping in the protein synthesis process. Current studies have shown that most RNAs do not encode proteins. The role and diversity of these numerous non-coding RNAs is yet to be understood. Recent discoveries have indicated that RNAs are more than inactive molecules. For example, a class of small (≈ 21 nucleotides [nt]) RNA molecules is involved in many of the cell's controls, from defense mechanisms (against pathogens and other mobile genetic elements such as transposons) to regulation of the expression of their own genes during development. Among these 21 nt RNAs, microRNAs (miRNAs) and short interfering RNAs (siRNAs), are the two major types. Both diverged from their origin (miRNAs are endogenous small RNAs while siRNAs derive from exogenous dsRNA) but not from their function (Simard and Hutvagner, 2005). The finding of such small RNAs adds a "twist" to the general models of gene regulation and has opened an active field of research.

This mechanism of down-regulation that promotes degradation of the mRNA target, was first described in plants and named "gene silencing". The negative regulation of the target gene arises from either one of two mechanisms: processing by cleavage or translational inhibition of the cognate mRNA. Such inhibition of protein synthesis, which was later seen in worms, flies and other organisms, came to be known as RNA interference (RNAi). To date, all multicellular organisms examined, plants and animals, exhibit such miRNAs. The production and function of small RNAs require a common set of proteins. The finding of these important small RNAs resulted in the release in April 2005 of the miRNA registry, an online database that coordinates and records miRNAs (http://microrna.sanger.ac.uk/sequences/). So far, 117 microRNAs have been identified in *Arabidopsis thaliana* and grouped in 40 families (miRNAs differing by 3 or less nucleotides). Only one miRNA (miR402) gene has been identified within an intron (Sunkar and Zhu, 2004). The remaining endogenous small RNAs cloned from *Arabidopsis*

have not yet been classified. It has been shown in plants that small RNAs are involved in proper cell development and patterning (Rhoades *et al*., 2002; Parizotto *et al*., 2004), as well as in accurate cell division in yeast (Lippman and Martienssen, 2004), mouse (Bernstein *et al*., 2003) and fruit fly (Pal-Bhadra *et al*., 2004). Further insights are needed to develop an understanding of small RNA-mediated regulation and its importance in plant biology.

5.1.1. Distinguishing Between the Small RNAs: siRNA, miRNA, and Other Small RNAs

Different classes of small RNAs play an active role in RNAi (Table 5.1). Although they are similar in size (21–27 nt) and perform interchangeable biochemical functions, small RNAs have distinct characteristics and precursors and can be separated into two big families, siRNA and miRNA. Individual siRNAs are typically less conserved than miRNAs, most likely reflecting their different origins and operational modes.

5.1.1.1. The Origin of Small RNAs

siRNA sequences are rarely conserved, while many miRNAs are conserved across species, underlining a potential evolutionary preservation of their role in gene regulation (Carrington and Ambros, 2003). However, there is no report of any miRNA gene that is conserved between kingdoms, suggesting that miRNAs are evolutionarily ancient (at least 400 million years old [Floyd and Bowman, 2004]). Despite the evolutionary origins of miRNAs not being well understood (Voinnet, 2004), information about their biogenesis, activity and targets is accruing at a rapid pace.

Most of the miRNAs are predominantly located in intergenic regions, and exceed by far the numbers that match genes, pseudogenes, transposons and retrotransposons (Lu *et al*., 2005). By using computational approaches, complemented with experimental analysis, these authors have shown that miRNAs are highly concentrated in the pericentromeric regions of each chromosome, similar to the long repeats observed previously and inversely proportional to mRNA abundance distribution. Many regions of the genome that were before considered as inactive and featureless, like the pericentromeric region which generally corresponds to transposon and retrotransposon loci and is often called a genomic "cemetery", were found to have

TABLE 5.1. Classification of small RNAs.

Classes	Sub-class	Length	Action mechanism
miRNA	miRNA	≈22 (19-25)	Translation repression, mRNA cleavage
siRNA	siRNA	21	mRNA cleavage
	tasiRNA	21-22	mRNA cleavage
	rasiRNA	24-26	Histone and DNA modification

substantial small RNA activity. This cluster organization is usually located downstream from a canonical TATA box motif and seems to be critical for the coordinate regulation of miRNA biogenesis (Xie *et al*., 2005). Comparison of two small RNA libraries (from *Arabidopsis thaliana* inflorescence and seedlings) indicated preferential accumulation in specific tissue types, suggesting the existence of a mechanism that regulates miRNA expression. This may reflect a greater variety or increased use of small RNAs in specialized cell types.

5.1.1.2. The Precursor Molecule

siRNAs generally derive from exogenous sources (introduced dsRNA, viruses, transgenes) and recognize targets with perfect complementarity. miRNA typically derives from the distinct loci that they regulate and show less than perfect complementarity. In contrast with siRNAs that come from long dsRNA molecules (100–1000 bp), miRNAs originate from capped and polyadenylated primary RNA transcripts named pri-miRNA (Jones-Rhoades and Bartel, 2004). Those stem-loops or hairpins (≈70 nt) will successively be processed first into pre-miRNA and then into miRNA:miRNA* duplexes. Usually only a single-stranded stable and active miRNA is liberated from each stem-loop precursor. Only a single mature miRNA is generated from miRNA hairpin precursors while a multitude of siRNAs accumulate from the dsRNA precursor.

5.1.1.3. The Target

Another difference between miRNA and siRNA is in the target sequence. siRNA typically specifies "auto-silencing", with silencing of the same locus or highly conserved ones, while miRNA can silence very different families of genes. A mutation in siRNA sequence would change the targeted sequence and its regulation, while a mutation in miRNA sequence is rarely followed with a different selection pressure of its target. The putative targets of miRNAs in plants are predominantly regulatory genes, such as transcription factors, F-box proteins and ubiquitin-conjugating enzymes (Table 5.2). But targets that fall into other gene categories imply that miRNA-regulated genes may be involved in many pathways in plant biology and not just development. In mammalian cells, the number of genes coding for miRNA represents nearly 1% of the predicted genes in those genomes, a fraction similar to that of other large gene families with regulatory roles.

5.1.1.4. Other Small RNAs

Recently, another type of small RNA has been found in plants and worms, named endogenous *trans*-acting siRNA (tasiRNA). These tasiRNAs cleave mRNA *in trans*, indicating that the target gene is different from the gene of origin, however, their role and cellular function remain unclear (Vazquez *et al*, 2004a). Like tasiRNA, rasiRNA (repeat-associated siRNA) presumably

TABLE 5.2. *Arabidopsis* miRNAs.

miRNA	Target gene class	References
156	SYB transcription factor	1, 2
159	MYB transcription factor	1, 3, 4
JAW-319	TCP transcription factor	4
160	ARF transcription factor	1, 2
161	PPR transcription factor	1, 5
162	DCL1 RNAi enzyme	6
163	SAMT	5
164	NAC transcription factor	1, 2, 7, 8
165	HD-ZIP III	9, 10
167	ARF transcription factor	1, 2
168	AGO1 RNAi enzyme	1, 11
169	CCAAT transcription factor	12
170-171	SCR transcription factor	1, 13
172	AP2 transcription factor	2, 14, 15
173	TAS1, TAS2	16
390	TAS3	16
393	TIR1/F-box - bHLH transc. factor	12
394	F-box	12
395	ATP sulfurylase, AST	12, 16
396	GRF	12
397	Laccase	12
398	Superoxide dismutase	12
399	transporter E2-UBC	16
403	AGO2 RNAi enzyme	16
406	Spliceosomal protein	17
447	2PGK	16

1: Vazquez *et al*., 2004b; 2: Kasschau *et al*. 2003; 3: Achard *et al*., 2004; 4: Palatnik *et al*., 2003; 5: Allen *et al*., 2004; 6: Xie *et al*., 2003; 7: Mallory *et al*., 2004; 8: Laufs *et al*., 2004; 9: Tang *et al*., 2003; 10: Emery *et al*., 2003; 11: Vaucheret *et al*., 2004; 12: Jones-Rhoades and Bartel, 2004; 13: Llave *et al*., 2002b; 14: Aukerman and Sakai, 2003; 15: Chen, 2004; 16: Allen *et al*., 2005; 17: Sunar *et al*., 2004.

derives from long dsRNA and therefore belongs to the siRNA family (Table 5.1). rasiRNA leads to transcriptional silencing by remodeling the heterochromatin through histone 3 (Lys9) methylation (Volpe *et al*., 2003).

5.1.2. Two Distinct Stages of RNAi: Initiator and Effector Phases

Biogenesis and action of small RNAs is dependant upon the activity of different proteins that have been identified by homology between animals and plants. The initial synthesis of small RNAs is mediated by DICER, but they can only be activated, and then recognize and interact with the target when they are associated with the RISC (RNA inducing silencing complex) protein complex.

5.1.2.1. The Dicer Protein and the Initiator Phase

The first report of the presence of small RNAs was made in the plant kingdom, where plants that showed posttranscriptional gene silencing (PTGS) activated either by a transgenic or a viral sequence, accumulated specific small RNA known as siRNA (Hamilton and Baulcombe, 1999). Fire *et al.* (1998) showed evidence, in worms, that small double stranded RNAs were more active by dramatically inhibiting expression of genes displaying enough similarities with those small RNAs. Similar observations were made in plants using transformation (Waterhouse *et al.*, 1998). This mechanism was then thought to be part of the organism's cell machinery when a protein named DICER was identified. DICER (Denli and Hannon, 2003), an RNase III enzyme, is a protein that converts long exogenous dsRNA, by endonucleolytic cleavage, to small RNAs. The sequence and structure of the small RNAs generated determine both their function and the identity of their target genes.

As discussed earlier, siRNAs derive from hundred or thousand bp long dsRNA, whereas miRNAs derive from long, capped and polyadenylated RNAs transcripts named pri-miRNA that will be processed into pre-miRNA and then into mi-RNA:miRNA*. In plants, both steps are performed in the nucleus by a protein named DCL1 (DICER-LIKE 1) (Papp *et al.*, 2003) and also, based on homology with animal Dicer protein, by HEN1 and HYL1 (Park *et al.*, 2002; Vazquez *et al*, 2004b). The biogenesis of plant miRNA is shown in Fig. 5.1. All current evidence suggests that small RNAs (both siRNA and miRNA) are generated from dsRNA but accumulate in functional complexes as single-stranded RNA. Each miRNA and siRNA duplex can potentially give rise to two single-stranded active molecules, but only the mature miRNA and the siRNA "guide strand" accumulate *in vivo* and enter the functional complex. miRNA* and the siRNA "passenger strand", will be largely destroyed. This idea is well supported by experiments using P19 protein, a suppressor of the RNA silencing defenses that plants display against viral infection (see Section 5.1.5. of this chapter). P19 binds and stabilizes double stranded small RNA, which leads to the accumulation of miRNA:miRNA* duplexes in plants (Chapman *et al.*, 2004; Dunoyer *et al.*, 2004). Interestingly, a large fraction of plant miRNA targets are either known to be, or homologous to, transcription factors involved in cell fate determination. This strongly supports the idea that miRNAs are important in the proper regulation of plant development. The role of miRNA in plant development was demonstrated by the discovery of the *dcl1* mutant which displays conversion of the floral meristem to an undetermined meristem and also shows multiple other defects in plant development, such as delayed flower timing (Schauer *et al.*, 2002). The *dcl1* mutant has reduced levels of miRNA and shares phenotypic similarities with other mutants like *hen1, ago1* or *hyl1* (Lu and Fedoroff, 2000; Park *et al.*, 2002; Kidner and Martienssen, 2003). This data suggests that miRNAs function as posttranscriptional negative

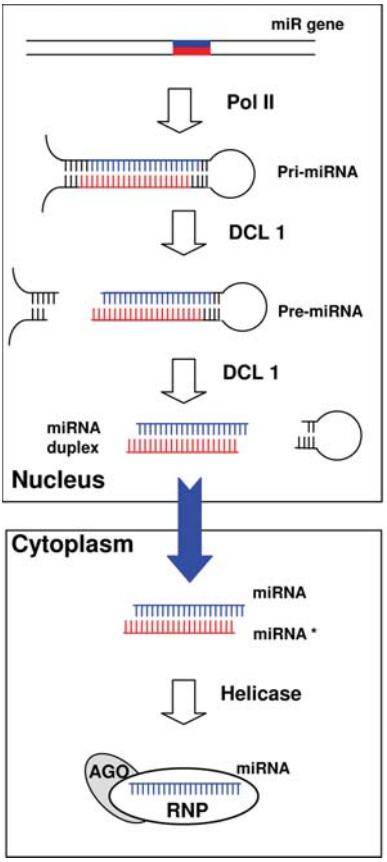


FIGURE 5.1. Plant miRNA biogenesis pathway. miRNA genes (*miR*) are transcribed by RNA polymerase II (Pol II) to generate the primary transcripts (pri-miRNAs). DCL1 catalyzes the processing of pri-miRNAs to form the miRNA duplexes. The duplex is separated, and the mature miRNA is selected and loaded onto the effector complex (Ribonucleic Protein, RNP), whereas the other strand (miRNA*) is degraded. (*See Color Plates*)

regulators to control meristem identity and cell proliferation, organ polarity and separation (*e.g.*, establishment of adaxial-abaxial polarity controlled by miRNA 165/166-mediated regulation of PHV/PHB genes, Tang *et al.*, 2003), as well as other developmental process (Bartel 2004, He and Hannon, 2004). These miRNAs show temporal and tissue-specific expression patterns (Llave *et al.*, 2002a; Reinhart *et al.*, 2002) and seem to interfere with expression of mRNA-encoding factors that control developmental timing, stem cell maintenance, and other developmental or physical processes. In addition, plants expressing a viral protein that alters miRNA accumulation have similar developmental defects (Kasshau *et al.*, 2003).

5.1.2.2. The Argonaute Protein and the Effector Phase

Despite the fact that Argonaute (Ago) and Dicer proteins have been shown to co-immunoprecipitate by binding through direct interaction between the RNase-III domain of DICER and the PIWI domain from Ago proteins,

dsRNA processing and target-mRNA degradation were found to be biochemically separable activities (Bern-stein *et al*., 2001; Hammond *et al*., 2001). These activities occur in two distinct phases termed "initiator" and "effector". RISC (RNA inducing silencing complex) is part of the "effector" process. RISC assembly has mainly been studied *in vitro* for *Drosophila* S2 cells or embryos and also in HeLa cells. We do not know yet if this data can be generalized for other organisms. We know that silencing pathways share, so far, a common set of proteins to produce and amplify such small RNAs. One of them is a member of the Argonaute family of proteins that seems to be a key protein in the RNA silencing effector complex (Tabara *et al*., 1999, 2002; Hammond *et al*., 2001; Shi *et al*., 2004). Ago proteins are defined by the presence of PAZ and PIWI domains. The PAZ domain is thought to bind single stranded RNA (Lingel *et al*., 2003; 2004), while the PIWI domain is structurally similar to the endoribonuclease-H domain that cleaves RNA. RISC contains single-stranded RNA, and therefore siRNA unwinding must occur at some point during the RISC assembly pathway. This process seems to be an energy-consuming procedure and may be catalyzed by an RNA unwindase enzyme. This event is important in the activation of RISC by the selection and loading of the functional siRNA (guide strand), while the passenger strand will be discarded and degraded. Two hypotheses on the selection of the active siRNA have been proposed. A few models envision that the selection could be largely dictated by the relative thermal stability of base pairing at the two ends. This "thermodynamic asymmetry" makes sense in the RISC where the PAZ domain seems to be the best candidate for binding the 3′ end, while the "seed" sequence (2–8 nt) matches the target sequence with the 5′ phosphate resting near the PIWI domain. The lack of a 5′ phosphate on the seed strand prevents stable association of the siRNA with the complex (RISC), and its presence is required for an siRNA to function. The other hypothesis comes from the dual roles of DICER that not only generates siRNAs but also escorts them into RISC. It seems that the selection could be influenced by the direction that DICER takes to produce siRNA and only the strand with its 3′ terminus at the processed end enters the RISC (Elbashir *et al*., 2001a).

5.1.3. Operational Modes and Functions

To prevent expression of a protein product, two steps are targeted: translation (through alteration in mRNA stability and through direct translation repression of protein synthesis) or transcription (through chromatin modification). Those mechanisms are explained in Fig. 5.2.

5.1.3.1. mRNA Cleavage

When siRNA or miRNA pairs extensively with its RNA target, the protein complex cleaves specifically across from nucleotides 10 and 11 of the small

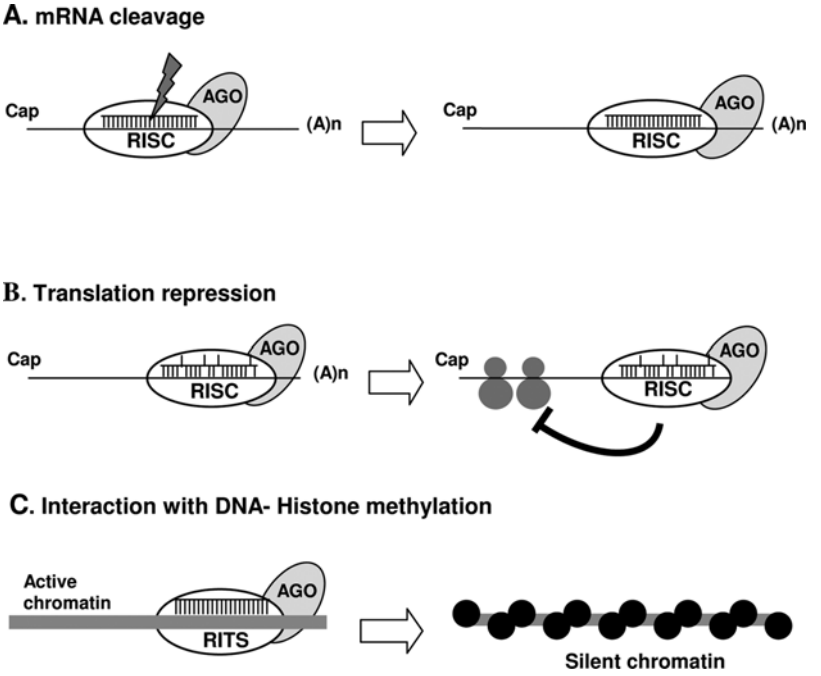


FIGURE 5.2. Model for RNA silencing pathways. A: Extensive complementarity of small RNAs in coding region or UTR induces target mRNA cleavage. B: Partial match of small RNA for the target, acts as guide molecule in translational control. C: small RNAs incorporated in RITS complex associate with chromatin and induce histone/DNA methylation.

RNA guide (Elbashir *et al*., 2001b). This is actually the dominant mechanism by which plant miRNAs regulate their target genes (Llave *et al*., 2004b; Tang *et al*., 2003). This cleaving process is carefully managed and is the consequence of the way small RNAs are bound to a specific member of the Argonaute protein family. The multidomain Argonaute proteins contain two RNA-binding domains: the PIWI and PAZ domains, that seem to affect their function in controlling gene expression. The 5′ end of an miRNA or siRNA (named "seed" sequence) contributes disproportionately to target RNA binding. To perform its task, a high-to-perfect complementarity is required between the target mRNA and the 5′-half of the miRNA (Parizotto *et al*., 2004; Doench and Sharp, 2004). Complementarities between the small RNA and its target help to determine if the small RNA directs target cleavage or represses mRNA translation, because an A-form helix must be formed with the target RNA at the center of an siRNA guide strand for cleavage to occur. In the absence of such a helix, the central region of the small RNA cannot pair with the target, and translation repression may ensue. miRNA cleavage in plants seems to tolerate up to 5 mismatches

between the miRNA and its target (Palatnik *et al*., 2003), but additional mismatches into the miRNA-binding site of the target reduce cleavage (Llave *et al*., 2002b). The putative targets are predominantly regulatory genes, such as transcription factors and F-Box proteins, but recent data implicates miRNA in many facets of plant biology from ubiquitin machinery to more structural and physiological processes (Lewis *et al*., 2003; John *et al*., 2004; Sunkar and Zhu, 2004). After cleavage of the mRNA, the miRNA remains intact and can guide the recognition and destruction of additional messages.

5.1.3.2. Translation Repression

When siRNAs or miRNAs pair only partially with their targets, including mismatches in the "seed" or central sequence, they cannot direct mRNA cleavage. Instead, they block translation of the mRNA into protein (Olsen and Ambros, 1999; Doench *et al*., 2003). They do this by either directing degradation of the polypeptide as it emerges from the ribosome or by "freezing" the ribosome after initiation so that ribosomal scanning is inhibited (Olsen and Ambros, 1999). Another possibility that could explain how small RNAs affect protein synthesis is by altering the stability of the target mRNA, making it more susceptible to degradation. New studies have proposed a speculative model for translation repression, whereby small RNAs, when bound to the Ago 2 protein associated with an enzyme that removes the 5′ 7-methylguanosine cap (characteristic of newly synthesized mRNA), can stimulate the movement of the mRNA from the cytosol to sites of mRNA destruction called "P-bodies" (Liu *et al*., 2005; Sen and Blau 2005). By sequestrating mRNA in P-bodies, small RNAs would block mRNA translation. This process would reduce the translational rate of the mRNA without affecting its steady-state abundance until the mRNAs were actually degraded in the P-bodies.

The most detailed and well understood example of an miRNA working as a translational repressor is miRNA172 which regulates APETALA2 (AP2, a class A gene) and controls floral organ identity and floral stem cell proliferation. Xuemei Chen (2004) first showed that the AP2 protein level in *hen1* and *dcl1* mutants defective in miRNA, was approximately 2–3 times lower than that in wild type, while AP2 mRNA abundance was similar in all genotypes. This observation suggested that AP2 was regulated at the translational or posttranslational level by a miRNA(s) absent in *hen1* and *dcl1*. The MIR172 gene under the double enhancer 35S (cauliflower mosaic virus) promoter was then introduced into wild type plants to determine the effect of elevated miRNA172 on AP2 expression. It appeared that transgenic plants showed a floral homeotic phenotype similar to those of AP2 loss-of-function mutants, but in addition displayed rosette leaves that curled upward, suggesting that AGAMOUS (AG, a class C gene) expression was simultaneously elevated. This data confirmed that miRNA172 regulated AP2 but also, influenced other floral patterning genes, like AG in leaves. In untransformed

plants, miRNA172 appears to be strongly present in the floral primordia until stage 7 when it accumulates specifically in the inner two floral whorls. As a result, this accumulation may cause the AP2 protein to be concentrated in the outer two whorls, where AP2 acts to specify perianth identities. miRNA172 acts like an AP2 regulator and appears to operate at the translation level, rather than at the level of RNA stability.

Complementarity may not be the only factor that influences the way in which miRNA regulates its target. Perhaps the context of the miRNA binding site plays a role in determining its mode of action. Another possibility is that different mechanisms of regulation may operate in different cell types. Lu *et al*. (2005) showed that the number of distinct siRNAs from the inflorescence was at least four times that of those from seedlings, which may reflect a greater variety of specialized cell types or increased use of small RNAs in all cell types within the inflorescence. After observing a lack of accumulation of virus-derived siRNAs in roots *versus* leaves, Andika *et al*. (2005) proposed that RNA silencing pathways may be fundamentally different in the aerial parts of the plant compared to the roots. This observation confirmed that the small RNA world and its regulatory role is broader and more complex than previously thought.

5.1.3.3. Chromatin Modification

The effect of RNA silencing pathways is not limited to cytoplasmic processes such as mRNA turnover and protein synthesis. In the last few years, some groups have shown that small RNA and RNAi participate in a genetic phenomenon known as epigenetic control (Xie *et al*., 2004), as well as contribute to the initiation of heterochromatin formation. Small RNAs seem to regulate and guide nuclear events that adjust the shape of the chromatin, including histone modification (H3 is modified by methylation at lysine 9) and DNA methylation. These changes in chromatin shape can result in transcriptional silencing (Wassenegger *et al*., 1994; Mette *et al*., 2000) and lead to gene inactivation (Reinhart and Bartel, 2002; Volpe *et al*., 2002; Zilberman *et al*., 2003). In plants, siRNAs that are needed to methylate cytosine and H3K9 are "diced" by DCL3 (DICER-LIKE protein 3) from dsRNAs initiated by the RNA-dependant RNA polymerase (RdRp), RDR2 (Xie *et al*., 2004). Then, the siRNA is unwound and the resulting small single-stranded RNA is loaded onto the RITS complex (RNA-induced transcriptional silencing) that contains a member of the Argonaute protein family, *e.g*., AGO4, among others in plants (Volpe *et al*., 2002; Zilberman *et al*., 2004). The RITS complex apparently recruits histone-modification factors. DNA and chromatin-modifying enzymes that bind in a large complex to siRNA, to guide the assembly of heterochromatin. It is not known whether the RITS-complex recognizes DNA directly or whether it hybridizes to newly-formed transcripts. It seems that siRNAs interact directly with DNA (Jones *et al*., 2001) but only in a short interval during transcription of a sequence when the DNA

is unpaired; when DNA is paired again behind the polymerase, the site becomes inaccessible to the siRNA. In plants, "transcriptionally silenced" DNA seems to be transcribed by a specialized DNA-dependant RNA-polymerase, RNA polymerase IV (Herr *et al*., 2005; Kanno *et al*., 2005). This RNA pol IV appears to provide a constant source of template for the RdRp that is necessary to manufacture siRNAs.

Bao *et al*., 2004, reported that genes encoding two *Arabidopsis* miRNA targets, PHABULOSA and PHAVOLUTA, are heavily methylated, and mutations that disrupt miRNA complementarity correlate with decreased methylation and presumably increased transcription. Even if the mechanism involved is not yet clear, these results suggest that miRNA may have direct or indirect access to transcriptional silencing pathways.

5.1.4. Amplification of the Silencing Triggers

In plants, RNA dependent RNA polymerase enzymes are required for silencing initiated by single-stranded RNA triggers (Dalmay *et al*., 2000), but perhaps not for dsRNA triggers (Waterhouse *et al*., 1998). One of the first hypotheses on how the RNA silencing pathway is initiated, was developed because of the detection of "aberrant" RNAs. A likely candidate for an aberrant RNA sensor was a class of RNA silencing proteins that copy single-stranded RNA into dsRNAs. These RdRps are found in nearly every eukaryote with a functioning RNA silencing pathway and are thought not only to trigger RNA silencing, but also amplify and sustain the mechanism. The model that seems to be the best to date, is that dsRNAs are "diced" into "primary siRNAs" that will be used as primers by the RdRps to amplify more dsRNA and create abundant "secondary siRNAs" (Sijen *et al*., 2001). This amplification can lead to the spreading of siRNA in both directions (5′ and 3′) along the target gene. siRNA initiation that is locally happening in one cell, can then be linked to amplification in other cells *via* plasmodesmata and systemic silencing. This has been observed in worms and plants, in which locally initiated silencing can be inherited and spread systemically to distant parts of the organism (Boerjan *et al*., 1994). The nature of the mobile silencing systemic signal is not yet well defined but due to its specificity, is thought to have a nucleic acid component that is probably RNA. By observing protein-bound siRNA flowing through the vascular system (which initiates systemic silencing), Yoo *et al*. (2004) have shown that siRNAs function as both intra and extracellular signaling molecules. This process is thought to provide a systemic antiviral response that immunizes tissues that are not yet infected.

5.1.5. A Natural Defense Mechanism

It was thought at first that the only purpose of siRNA was to defend plants against RNA viruses and transposons. It has now been shown that plants use RNA silencing as a natural antiviral defense (Lecellier and Voinnet, 2004).

Because plants lack the protein-based adaptive immunity seen in animals, RNA silencing can be seen as a form of immune system that operates at the nucleic acid level to fight off invading elements. However, this system is not triggered by the host but by features and sequences within the pathogen's genome. Upon infection of viruses, viral DNA or RNA is used as template to generate dsRNA that will then be recognized by the silencing machinery. Mutation in some RNAi key genes, such as DCL2, RDR6/SGS2, increases susceptibility and sensitivity toward viruses (Mourrain *et al.*, 2000). There is no clear evidence that animal viruses also induce siRNA, and it seems unlikely that RNA silencing is a general anti-viral defense mechanism, at least in mammals. The lack of DCL protein diversification in mammals, compared to plants, supports the hypothesis that progressively more advanced animals lost this feature in favor of a more elaborate active defense. This hypothesis, however, does not exclude participation of RNA-silencing in antiviral responses, and miRNA might just be the molecule involved in such a response. This has been demonstrated with animal viruses such as Epstein-Barr virus, which encodes miRNA in its genome (Pfeffer *et al.*, 2004), and regulates viral and host gene expression to optimize infectivity in host cells. Plant viruses also have evolved various strategies to counteract the host's antiviral RNA silencing (Baulcombe, 2004). The first identified strategy involves suppressor factors of silencing which are encoded by both DNA and RNA viruses (Voinnet, 2005). For instance, tombusviral p19 protein binds tightly to siRNA duplexes and prevents them from being incorporated into an active RISC (Lakatos *et al.*, 2004; Dunoyer *et al.*, 2004, Zamore, 2004). Other viruses use indirect interference with silencing-effector molecules, illustrated by the potyviral helper-component proteinase (Hc-Pro). A third option employed by viruses to counteract silencing is by making themselves inaccessible to RNAi machinery through protective secondary structure in their RNA or through compartmentalization that hides them from the RNA silencing mechanism.

5.2. Using RNAi Technology as a Molecular Tool

RNAi technology has emerged as an effective method in reverse genetics for silencing gene expression in eukaryotic cells. This technology allows scientists to observe the effect of the loss or reduction of mRNA level of an individual gene and therefore to investigate gene function. Compared to other methods that knock-down the expression of specific genes, *e.g.*, antisense oligodeoxynucleotides, ribozymes, insertional mutagenesis, and chemical mutagenesis), RNAi seems to be one of the most powerful tools. Many applications have been found in basic science to help in the characterization of gene function, signaling pathway analysis and target validation. Nevertheless, the application of RNAi technology in medical research to develop therapeutic methods for controlling human diseases may be its most valuable contribution. In this part we will discuss different methods used to activate the silencing pathway and some potential applications.

5.2.1. Methods of Induction of Gene Silencing

One of the biggest challenges facing research in this area is in the delivery of the active molecule that will trigger the RNAi pathway. While feeding worms with *Escherichia coli* expressing the active molecule may be possible (Timmons and Fire, 1998), the delivery of dsRNA that activates the RNAi pathway in plants is much more complex. Synthetic siRNA duplexes (21 nt) or dsRNAs can be delivered to plant cells in several ways: microprojectile bombardment, micro-injection, infiltration of plant tissue with *Agrobacterium*, or virus induced gene silencing. These various techniques each have advantages and disadvantages, but the weakest point is the fact that their effect is transient. The most reliable method to date for increasing the silencing effect is the stable transformation of constructs with intron spliced hairpin RNA (ihpRNA)-expressing transgenes; however, this technique requires efficient transformation techniques.

5.2.1.1. Biolistic, Virus and Agro-infection Delivery

Bombarding cells with particles coated with dsRNA, siRNA, or DNA that encode hairpin constructs, as well as sense or antisense RNA, activates the gene silencing pathway. Injection of *Agrobacterium* carrying similar DNA constructs into the intercellular spaces of leaves (known as *Agro*-inoculation) also triggers RNA silencing. This latter technique is mainly used in herbaceous species which led to the discovery and a better understanding of the silencing signal (Voinnet and Baulcombe, 1997). But the effect of both methods does not last more than a few days. An alternative to those procedures takes advantage of the virus genome's ability to carry and induce RNAi against foreign sequences and has been introduced as virus-induced gene silencing (VIGS) technology. Usually, 300–800 nt fragment target gene sequences are used and cloned into T-DNA (for transfer DNA from *A. tumefaciens*) plasmids modified to carry the viral sequences; these plasmids can be stably integrated into the plant genome (Ratcliff *et al*., 2001). The major limitation of this technology is the scarcity of virus-host systems that can be used. So far, most of the VIGS systems have been tested on *Nicotiana* species. Those experiments gave highly reproducible results using different vectors producing mild virus symptoms that do not interfere with or mask the visible effects resulting from silencing the target gene. In the last couple of years, significant efforts have been made to find new vectors to generate RNAi in important crop plants; however, this technology is still in its early development stage.

5.2.1.2. Stable Transformation

Initially, RNA silencing (or "posttranscriptional gene silencing" as it was named at that time) was generated by stable transformation using sense or antisense over-expressed constructs. The ratio of plants displaying characteristics of gene silencing was nevertheless really low. Waterhouse *et al*. (2000) designed an innovative vector expressing a self-complementary

hairpin (hpRNA) construct that efficiently produced siRNA and therefore activated the RNAi pathway in almost 100% of the transformed plants (Smith *et al.*, 2000). The hpRNAs were composed of a promoter, the targeted gene with two inverted sequences separated by a spacer, and a terminator. Thus, the RNA transcribed from such a transgene will hybridize with itself to form a hairpin structure, acting as a substrate for the generation of siRNA. Later work has shown that the use of an intron sequence separating the two inversely-repeated copies of the target (ihpRNA), efficiently enhances the capacity of silencing (Wesley *et al.*, 2001). Most of the ihpRNA constructs have been designed under the control of the constitutive 35S promoter, but tissue-specific or inducible promoters have also been successfully adopted to efficiently direct silencing (Byzova *et al.*, 2004; Wielopolska *et al.*, 2005). Moreover, this ihpRNAi technology could allow multiple genes to be silenced with a single construct, as suggested by experiments where effective silencing of two genes has been obtained from one hpRNAi construct (Allen *et al.*, 2004).

5.2.2. Functional Genomic Tools to Understand Essential Regulation of Key Developmental Processes

RNAi provides an important tool for functional genomics research that makes it possible to target and shut down any gene. The best example of the application of this technology is the use of the pHELLSGATE high-throughput gene silencing vector (Helliwell *et al.*, 2002) currently being used by the *Arabidopsis* genome RNAi knock-out line analysis (AGRIKOLA) consortium to create a collection of silencing vectors for all the available *Arabidopsis* genes [GSTs: gene sequence tags] (Hilson *et al.*, 2003). The aim of this type of study is to identify and assign a function to each gene, because many gene knockouts in plants do not produce a phenotype under normal growth conditions.

Nevertheless, many challenges in plant functional genomics are currently problematic. Most functional genomics research is conducted in model plants such as *Arabidopsis*, rice and maize. However, practical application of this research usually needs to be extended to non-model plants of economic importance. There are multiple limitations in applying RNAi technology across different plant species. First, knowledge of the target gene sequence is required, but only the *Arabidopsis thaliana*, *Oryza sativa* and *Populus trichocarpa* genomes have been completely sequenced, although there are programs currently sequencing the maize, soybean and tomato genomes which should be available in the near future. The second weakness is in the delivery methods of dsRNA, which are not always compatible with the plant species of choice. So far, *Arabidopsis* remains the model plant, but recent advances in VIGS, ihpRNA and siRNA delivery have allowed the development of other plant systems, including *solanaceae* (Ryu *et al.*, 2004), barley (Hein *et al.*, 2005), maize (Cigan *et al.*, 2005), poplar (Naylor *et al.*, 2005) and poten-

tially any transformable species. It will be possible to target other more recalcitrant species as improved transformation technology becomes available.

5.2.3. Improvement of Plant Characteristics

RNAi technology may be used to generate improved crop varieties in terms of disease and insect resistance, enhancing nutritional qualities and other properties of agronomic or horticultural interest. Table 5.3 displays examples of genes silenced through ihpRNA and their phenotypes in non-model plants. Some of these examples are discussed in more detail below.

5.2.3.1. Nutritional Improvement

Kusaba *et al*., (2003) have recently made an important contribution to agriculture by applying RNAi technology to improve rice plants. They were able to reduce the level of glutenin in rice grains and create a new rice variety named LGC-1 (Low Glutenin Content-1). Glutenin is a major seed storage protein accounting for 60% of total endosperm protein in rice, which is barely digested by the human body. LGC-1 is beginning to be used as a low-protein rice cultivar in diet therapy for patients with kidney disease who must restrict protein intake (Mochizuki and Hara, 2000).

Others have attempted to increase tomato fruit nutritional value by suppressing an endogenous photomorphogenesis regulatory gene, *DET1* (DETIOLATED 1). The DET1 transcription factor is a negative regulator of photomorphogenic responses, including downregulation of pigment formation. By reducing DET1 using fruit-specific promoters combined with RNA interference (RNAi) technology, Davuluri *et al*. (2005) have shown that DET1 transcripts were indeed specifically degraded in transgenic fruits. DET1-downregulated tomato fruit showed significant increases in all three

TABLE 5.3. Example of validated performances using ihpRNA technology in non model plant.

Plant	Phenotype	References
Potato	PLRV resistance	1
	Reduced enzymatic browning	2
Barley, wheat	BYDV resistance	3
Cotton	High stearic and oleic oil	4
Plum	PPV resistance	5
Ryegrass	Hypo-allergenic	6
Opium poppy	Morphine replacement with nonnarcotic alkaloids	7
Tomato	Slowed ripening process	8
Apple	Hypo-allergenic	9

1: Waterhouse *et al*., 1998; 2: Wesley *et al*., 2001; 3: Wang *et al*., 2000; 4: Liu *et al*., 2002; 5: Hily *et al*., in press; 6: Petrovska *et al*., 2004; 7: Allen *et al*., 2004; 8: Xiong *et al*., 2005; 9: Gilissen *et al*., 2005.

classes of photonutrient products (carotenoids, chlorogenic acid and flavonoids) which are precursors of vitamin A and other antioxidants. Such results could be beneficial for human health.

An Australian research team genetically modified the fatty acid composition of cottonseed oil using the hairpin RNA-mediated gene silencing method to down-regulate expression of two key fatty acid desaturase genes, *ghSAD-1* and *ghFAD2-1,* in seed (Liu *et al.*, 2002). Silencing *ghSAD-1* and/or *ghFAD2-1* to various degrees favors the development of cottonseed oils having novel combinations of palmitic, stearic, oleic, and linoleic contents that can be used in margarines and deep frying oils without hydrogenation. This should make possible the development of cottonseed fats and oils that satisfy different application requirements, in particular lines with fatty acid profiles matching those of valuable specialty confectionery fats such as cocoa (*Theobroma cocoa*) butter.

5.2.3.2. Agronomical Improvement

One possible application of RNAi involves the down-regulation of key enzymes in any biosynthetic pathway. This new technology has actually been tested in various economically important species, and has been associated with limiting the quantity of lignin in *Corchorus* to aid in the production of high quality paper in a more cost-effective manner. Another application of this technology is to reduce the production of CaMXMT1, an enzyme involved in the major caffeine synthetic pathway in coffee plants. Preliminary results indicate that the method can be practically applied to produce decaffeinated coffee varieties (Ogita *et al.*, 2004).

5.2.3.3. Disease Resistance

The most successful application of gene silencing for crop improvement so far has been the engineering of virus resistance, particularly in crops where resistance genes have not been identified in wild germplasm. This technology of improving pest resistance is based on a pathogen-derived resistance (PDR) approach (Sanford and Johnston, 1985). The concept of obtaining resistance to parasite/virus by transformation with genes derived from the genome of the pathogen is highly effective and relies on the production of dsRNA through the transcription of a self-complementary hairpin structure that interferes with the parasite/virus at the RNA level. Generally, fragments of 300 to 800 nucleotides derived from the target genome efficiently trigger the RNAi pathway. Nonetheless, fragments as short as 23–60 nucleotides have been successfully tested (Thomas *et al.*, 2001). Any part of the genome is able to activate the RNAi defense system. To fight viruses, most of the sequences used are derived from the viral coat protein (CP). For this technology to be very effective, Tenllado *et al.* (2004) estimated that a 500-fold excess of dsRNA over PMMoV viral RNA would be the threshold required for interference with virus replication under experimental conditions. This contrasts

with the few molecules of dsRNA sufficient to trigger RNAi reported in worms (Fire *et al.*, 1998). This discrepancy could be easily explained by the counter-defense naturally developed by some viruses (see section 4.1.5.). As discussed previously, the best way to deliver dsRNA into plant cells is through stable transformation of an ihpRNA construct. This strategy has been successfully tested in various organisms against numerous viruses in herbaceous plants. The technology has also been used to fight viruses in woody perennial trees. Researchers at the USDA Appalachian Fruit Research Station and INRA in France were able to transform plum trees (*Prunus domestica L.*) with an ihpRNA containing the *plum pox virus* (PPV) coat protein sequence. Two lines were obtained and inoculated with PPV, and no virus could be detected in either line (Hily *et al.*, unpublished result).

5.2.3.4. Biosafety

Questions concerning the potential ecological impact of transgenic plants have been raised, and the potential risks include complementation (ability of the transgene to restore a loss-of-function or to confer a gain-of-function to the wild-type virus when transgene and virus genome are brought together), heterologous encapsidation (when using coat protein transgenes), synergy (ability of the transgene to aid virus survival *in trans*), and unwanted genetic flow between organisms. RNAi technology is considered to be a "biosafe" technology, eliminating certain risks thought to be associated with transgenic plants carrying first generation constructs (binary vectors and sense or antisense genes). For example, earlier strategies for obtaining viral resistant plants required expression of the protein product from the transgene of interest and risked heteroencapsidation through protein-protein interactions between target and non-target viral gene products, as was observed when a non-aphid transmissible strain of zucchini yellow mosaic potyvirus became aphid transmissible from a transgene expressing a plum pox potyvirus capsid protein (Lecoq *et al.*, 1993). Since RNAi triggers the formation of dsRNA molecules that will target and facilitate the degradation of the gene of interest, as well as the transgene itself, it avoids problems arising from synthesis of protein products. Furthermore, this technology also allows the use of short gene sequences, as well as non-coding regions of genes, thus limiting undesirable recombination events.

5.3. Conclusion

RNA interference has added a new dimension to the regulation of gene expression by different types of RNAs. Recently labeled "Technology of the Year in 2002" (Couzin, 2002) and by Fortune Magazine's "Billion Dollar Breakthrough" (Stipp, 2003), this new field of research has emerged as a powerful research tool to understand unknown gene function and has also

been proven to be useful for producing improved crops. In plants, RNAi is thought to be a way of regulating endogenous genes, but also because plants lack the protein-based adaptive immunity, it is believed that the sequence-specific RNA silencing mechanism is an ancient form of defense evolved to counteract invading pathogens. The common denominator among these different roles is the trigger molecule, a dsRNA that will activate the RNAi pathway and lead to the degradation of the target mRNA or suppression of protein translation from the target mRNA. Studies of the pathway have revealed some interesting mechanistic details regarding physical and functional connections between initiator and effector phases; nevertheless, Ago family proteins and some tiny RNAs seem to be the core elements of this pathway. The term "small" RNA, that defines their physical properties, is somewhat misleading considering the importance of these "riboregulators" in key developmental and defense pathways. Future studies will likely find other significant roles for this class of RNA and suggest novel strategies for regulating gene expression to improve crop traits and to fight human diseases for which there are currently only limited therapies.

Acknowledgements. The authors would like to extend their appreciation to Dennis Bennett for consultation and assistance in this work. Post-doctoral program, class of 2005, of USDA-ARS Headquarter funding for J.M.H. is gratefully acknowledged.

References

Achard, P., Herr, A., Baulcombe, D.C., and Harberd, N.P., 2004, Modulation of floral development by gibberellin-regulated microRNA, *Development* **131**:3357-3365.

Allen, E., Xie, Z., Gustafson, A.M., Sung, G.H., Spatafora, J.W., and Carrington, J.C., 2004, Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana, Nat. Genet*. **36**:1282-1290.

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C., 2005, MicroRNA-directed phasing during trans-acing siRNA biogenesis in plants, *Cell* **121**:207-221.

Allen, R.S., Millgate, A.G., Chitty, J.A., Thistleton, J., Miller, J.A.C., Fist, A.J., Gerlach, W.L., and Larkin, P.J., 2004, RNAi-mediated replacement of morphine with the nonnarcotic alkaloid reticuline in opium poppy, *Nat. Biotech*. **22**: 1559-1566.

Andika, I.B., Kondo, H., and Tamada, T., 2005, Evidence that RNA silencing-mediated resistance to *Beet necrotic yellow vein virus* is less effective in roots than in leaves, *Molec. Plant-Microbe Interact*. **18**:194-204.

Aukerman, M.J., and Sakai, H., 2003, Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2-like* target genes, *Plant Cell* **15**: 2730-2741.

Bao, N., Lye, K.W., and Barton, M.K., 2004, MicroRNA binding sites in *Arabidopsis* class III HD-ZIP mRNAs are required for methylation of the template chromosome, *Dev. Cell* **7**:653-662.

Bartel, B.P., 2004, MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell* **116**:281-297.

Baulcombe, D.C., 2004, RNA silencing in plants, *Nature* **431**:356-363.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J., 2001, Role for a bidentate ribonuclease in the initiation step of RNA interference, *Nature* **409**: 293-296.

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J., 2003, Dicer is essential for mouse development, *Nat. Genet*. **35**:215-217.

Boerjan, W., Bauw, G., Van Montagu, M., and Inze, D., 1994, Distinct phenotypes generated by overexpression and suppression of S-adenosyl-L-methionine synthetase reveal developmental patterns of gene silencing in tobacco, *Plant Cell* **6**:1401-1414.

Byzova, M., Verduyn, C., De Brouwer, D., and De Block, M., 2004, Transforming petals into sepaloid organs in *Arabidopsis* and oilseed rape: implementation of the hairpin RNA-mediated gene silencing technology in an organ-specific manner, *Planta* **218**:379-387.

Carrington, J.C., and Ambros, V., 2003, Role of microRNAs in plant and animal development, *Science* **301**:336-338.

Cigan, A.M., Unger-Wallace, E., and Haug-Collet, K., 2005, Transcriptional gene silencing as a tool for uncovering gene function in maize, *Plant J*. **43**:929-940.

Chapman, E.J., Prokhenovsky, A.I., Gopinath, K., Dolja, V., and Carrington, J.C., 2004, Viral RNA silencing suppressors inhibits the microRNA pathway at an intermediate step, *Genes Dev*. **18**:1179-1186.

Chen, X., 2004, A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development, *Science* **303**:2022-2025.

Couzin, J., 2002, Small RNAs make big Splash, *Science* **298**:2296-2297.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D., 2000, An RNA-dependant RNA polymerase gene in *Arabidopsis* is required for post-transcriptional gene silencing mediated by a transgene but not by a virus, *Cell* **101**: 543-553.

Davuluri, G.R., van Tuinen, A., Fraser, P.D., Manfredonia, A., Newman, R., Burgess, D., Brummell, D.A., King, S.R., Palys, J., Uhlig, J., Bramley, P.M., Pennings, H.M., and Bowler, C., 2005, Fruit-specific RNAi-mediated suppression of DET1 enhances carotenoid and flavonoid content in tomatoes, *Nat. Biotechnol*. **23**:890-895.

Denli, A., and Hannon, G., 2003, RNAi: an evergrowing puzzle, *Trends Biochem. Sci*. **28**:196-201.

Doench, J.G., Petersen, C.P., and Sharp, P.A., 2003, siRNAs can function as miRNAs, *Genes Dev*. **17**:438-442.

Doench, J.G., and Sharp, P.A., 2004, Specificity of microRNA target selection in translational repression, *Genes Dev*. **18**:504-511.

Dunoyer, P., Lecellier, C.H., Parizotto, E.A., Himber, C., and Voinnet, O., 2004, Probing the microRNA and small interfering RNA pathways virus-encoded suppressors of RNA silencing, *Plant Cell* **16**:1235-1250.

Elbashir, S.M., Lendeckel, W., and Tuschl, T., 2001a, RNA interference is mediated by 21- and 22-nucleotides RNAs, *Genes Dev*. **15**:188-200.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T., 2001b, Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate, *EMBO J*. **20**:6877-6888.

Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L., 2003, Radial patterning of *Arabidopsis* shoots by class III *HD-ZIP* and *KANADI* genes, *Curr. Biol*. **13**:1768-1774.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C., 1998, Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*, *Nature* **391**:806-811.

Floyd, S.K., and Bowman, J.L., 2004, Gene regulation: ancient microRNA target sequences in plants. *Nature* **428**:485-486.

Gilissen, L.J., Bolhaar, S.T., Matos, C.I., Rouwendal, G.J., Boone, M.J., Krens, F.A., Zuidmeer, L., Van Leeuwen, A., Akkerdaas, J., Hoffmann-Sommergruber, K., Knulst, A.C., Bosch, D., Van de Weg, W.E., and Van Ree, R., 2005, Silencing the major apple allergen Mal d 1 by using the RNA interference approach, *J. Allergy Clin. Immunol*. **115**:364-369.

Hamilton, A.J., and Baulcombe, D.C., 1999, A species of small antisense RNA in posttranscriptional gene silencing in plants, *Science* **286**:950-952.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J., 2001, Argonaute2, a link between genetic and biochemical analysis of RNAi, *Science* **293**:1146-1150.

He, L., and Hannon, G.J., 2004, MicroRNAs: Small RNAs with a big role in gene regulation, *Nat. Rev. Genetics* **5**:522-531.

Hein, I., Barciszewska-Pacak, M., Hrubikova, K., Williamson, S., Dinesen, M., Soenderby, I.E., Sundar, S., Jarmolowski, A., Shirasu, K., and Lacomme, C., 2005, Virus-induced gene silencing-based functional characterization of genes associated with powdery mildew resistance in barley, *Plant Physiol*. **138**:2155-2164.

Helliwell, C.A., Wesley, S.V., Wielopolska, A.J. and Waterhouse, P.M., 2002, High-throughput vectors for efficient gene silencing in plants, *Funct. Plant Biol*. **29**:1217-1255.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C., 2005, RNA polymerase IV directs silencing of endogenous DNA, *Science* **308**:118-120.

Hilson, P., Small, I., and Kuiper, M.T., 2003, European consortia building integrated resources for *Arabidopsis* functional genomics, *Curr. Opin. Plant Biol*. **6**:426-429.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S., 2004, Human MicroRNA targets, *PLoS. Biol*. **2**:e363.

Jones, L., Ratcliff, F., and Baulcombe, D.C., 2001, RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance, *Curr. Biol*. **11**:747-757.

Jones-Rhoades, M.W., and Bartel, D.P., 2004, Computational identification of plant microRNAs and their targets, including a stress-induced miRNA, *Mol. Cell* **14**:787-799.

Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J., 2005, Atypical RNA polymerase subunits required for RNA-directed DNA methylation, *Nat. Genet*. **37**:761-765.

Kasschau, K.D., Xie, Z., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A., and Carrington, J.C., 2003, P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function, *Dev. Cell* **4**:205-217.

Kidner, C.A., and Martienssen, R.A., 2003, Macro effects of microRNAs in plants, *Trends Genet*. **19**:13-16.

Kidner, C.A., and Martienssen, R.A., 2005, The role of ARGONAUTE1 (AGO1) in meristem formation and identity, *Dev. Biol*. **280**:504-517.

Kusaba, M., Miyahara, K., Iida, S., Fukuoka, H., Takano, T., Sassa, H., Nishimura, M., and Nishio, T., 2003, Low glutelin content1: a dominant mutation that suppresses the glutelin multigene family *via* RNA silencing in rice, *Plant Cell* **15**:1455-1467.

Lakatos, L., Szittya, G., Silhavy, D., and Burgyan, J., 2004, Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses, *EMBO J*. **23**:876-884.

Laufs, P., Peaucelle, A., Morin, H., and Traas, J., 2004, MicroRNA regulation of *CUC* gene is required for boundary size control in *Arabidopsis* meristems, *Development* **131**:4311-4322.

Lecellier, C.H., and Voinnet, O., 2004, RNA silencing: no mercy for viruses? *Immunol. Rev*. **98**:285-303.

Lecoq, H., Ravelonandro, M., Wips-Scheibel, C., Monsion, M., Raccah, B., and Dunez, J., 1993, Aphid transmission of an non-aphid transmissible strain of zucchini yellow mosaic potyvirus from transgenic plants expressing the capsid protein of plum pox potyvirus, *Mol. Plant-Microb. Int*., **6**:403-406.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B., 2003, Prediction of mammalian microRNA targets, *Cell* **115**:787-798.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M., 2003, Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain, *Nature* **426**:465-469.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M., 2004, Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain, *Nat. Struct. Mol. Biol*. **11**:576-577.

Lippman, Z., and Martienssen, R., 2004, The role of RNA interference in heterochromatic silencing, *Nature* **431**:364-370.

Liu, Q., Singh, S.P., and Green, A.G., 2002, High-stearic and High-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing, *Plant Physiol*. **129**:1732-1743.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., Parker, R., 2005, MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies, *Nat. Cell Biol*. **7**:719-723.

Llave, C., Kasshau, K.D., Rector, M.A., and Carrington, J.C., 2002a, Endogenous and silencing-associated small RNAs in plants, *Plant Cell* **14**:1605-1619.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C., 2002b, Cleavage of *Scarecrow-like* mRNA targets directed by a class of *Arabidopsis* miRNA, *Science* **297**:2053-2056.

Lu, C., and Fedoroff, N., 2000, A mutation in the *Arabidopsis* HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin, *Plant Cell* **12**:2351-2366.

Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D., Meyers, B.C., and Green, P.J., 2005, Elucidation of the small RNA component of the transcriptome, *Science* **309**: 1567-1569.

Mallory, A.C., Dugas, D.V., Bartel, D.P., and Bartel, B., 2004, MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs, *Curr. Biol*. **14**:1035-1046.

Mette, M.F., Aufatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J.M., 2000, Transcriptional gene silencing and promoter methylation triggered by double stranded RNA, *EMBO J*. **19**:5194-5201.

Mourrain, P., Beclin, C., Elmayan, T., Feuerbach, F., Godon, C., Morel, J., Jouette, D., Lacombe, A., Nikic, S., Picault, N., Remoue, K., Sanial, M., Vo, T., and Vaucheret, H., 2000, *Arabidopsis* SGS2 and SGS3 genes are required for post-transcriptional gene silencing and natural virus resistance, *Cell* **101**:533-542.

Mochizuki, T., and Hara, S., 2000, Usefulness of low protein rice on diet therapy in patients with chronic renal failure, *Jpn. J. Nephrol*., **42**:24-29.

Naylor, M., Reeves, J., Cooper, J.I., Edwards, M.L., and Wang, H., 2005, Construction and properties of a gene-silencing vector based on Poplar mosaic virus (genus *Carlavirus*), *J. Virol. Methods* **124**:27-36.

Ogita, S., Uefuji, H., Morimoto, M., and Sano, H., 2004, Application of RNAi to confirm theobromine as the major intermediate for caffeine biosynthesis in coffee plants with potential for construction of decaffeinated varieties, *Plant Mol. Biol*. **54**:931-941.

Olsen, P.H., and Ambros, V., 1999, The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation *Dev. Biol*. **216**:671-680.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D., 2003, Control of leaf morphogenesis by microRNAs, *Nature* **425**:257-263.

Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A., and Elgin, S.C., 2004, Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery, *Science* **303**:669-672.

Papp, I., Mette, M.F., Aufsatz, W., Daxinger, L., Schauer, S.E., Ray, A., Van Der Winden, J., Matzke, M., and Matzke, A.J., 2003, Evidence for nuclear processing of plant microRNA and short interfering RNA precursors, *Plant Physiol*. **132**: 1382-1390.

Parizotto, E.A., Dunoyer, P., Rahm, N., Himber, C., and Voinnet, O., 2004, In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. *Genes Dev*. **18**: 2237-2242.

Park, W., Li, J., Song, R., Messing, J., and Chen, X., 2002, CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*, *Curr. Biol*. **12**:1484-1495.

Petrovska, N., Wu, X., Donato, R., Wang, Z.Y., Ong, E.-K., Jones, E., Forster, J., Emmerling, M., Sidoli, A., O'Hehir, R., and Spangenberg, G., 2004, Transgenic ryegrasses (*Lolium* spp.) with down-regulation of main pollen allergens, *Mol. Breed*. **14**:489-501.

Pfeffer, S., Zavolan, M., Grasser, F.A., Chien, M., Russo, J.J., Ju, J., John, B., Enright, A.J., Marks, D., Sander, C., and Tuschl, T., 2004, Identification of virus-encoded microRNAs, *Science* **304**:734-736.

Ratcliff, F., Martin-Hernandez, A.M., and Baulcombe, D.C., 2001, Technical Advance. Tobacco rattle virus as a vector for analysis of gene function by silencing *Plant J*. **25**:237-245.

Reinhart, B.J., and Bartel, D.P., 2002, Small RNAs correspond to centromere heterochromatic repeats, *Science* **297**:1831.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P., 2002, MicroRNAs in plants, *Genes Dev*. **16**:1616-1626.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P., 2002, Prediction of plant microRNA targets, *Cell* **110**:513-520.

Ryu, C.M., Anand, A., Kang, L., and Mysore, K.S., 2004, Agrodrench: a novel and effective agroinoculation method for virus-induced gene silencing in roots and diverse Solanaceous species, *Plant J*. **40**:322-331.

Sanford, J.C., and Johnston, S.A., 1985, The concept of parasite-derived resistance-deriving resistance genes from the parasite's own genome. *J. Theor. Biol*. **113**: 395-405.

Schauer, S.E., Jacobsen, S.E., Meinke, D.W., and Ray, A., 2002, DICER-LIKE1: blind men and elephants in *Arabidopsis* development, *Trends Plant Sci*. **7**:487-491.

Sen, G.L., and Blau, H.M., 2005, Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies, *Nat. Cell Biol*. **7**:633-636.

Shi, H., Djikeng, A., Tschudi, C., and Ullu, E., 2004, Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance, *Mol.Cell Biol*. **24**:420-427.

Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H.A., and Fire, A., 2001, On the role of RNA amplification in dsRNA-triggered gene silencing, *Cell* **107**:465-476.

Simard, M.J., and Hutvagner, G., 2005, RNA silencing, *Science* **309**:1518.

Smith, N.A., Singh, S.P., Wang, M.B., Stoutjesdijk, P.A., Green, A.G., and Waterhouse, P.M., 2000, Total silencing by intron-spliced hairpin RNAs, *Nature* **407**:319-320.

Stipp, D., 2003, Biotech's Billion dollar Breakthrough: a technology called RNAi has opened the door to major new drugs, *Fortune* **147**:96.

Sunkar, R., and Zhu, J.K., 2004, Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*, *Plant Cell* **16**:2001-2019.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C., 1999, The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*, *Cell* **99**:123-132.

Tabara, H., Yigit, E., Siomi, H., and Mello, C.C., 2002, The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in *C. elegans*, *Cell* **109**:861-871.

Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D., 2003, A biochemical framework for RNA silencing in plants, *Genes Dev*. **17**:49-63.

Tenllado, F., Llave, C., and Diaz-Ruiz, J.R., 2004, RNA interference as a new biotechnological tool for the control of virus diseases in plants, *Virus Res*. **102**:85-96.

Thomas, C.L., Jones, L., Baulcombe, D.C., and Maule, A.J., 2001, Size constraints for targeting post-transcriptional gene silencing and for RNA-directed methylation in *Nicotiana benthamiana* using a potato virus X vector, *Plant J*. 25:417-425.

Timmons, L., and Fire, A., 1998, Specific interference by ingested dsRNA, *Nature* **395**:854.

Vaucheret, H., Vazquez, F., Crete, P., and Bartel, D.P., 2004, The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development, *Genes Dev*., **18**:1187-1197.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P., 2004a, Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs, *Molec. Cell* **16**:69-79.

Vazquez, F., Gasciolli, V., Crete, P., and Vaucheret, H., 2004b, The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing, *Curr. Biol*. **14**:346-351.

Voinnet, O., 2005, Induction and suppression of RNA silencing: insights from viral infections, *Nat. Rev. Genet*. **6**:206-220.

Voinnet, O., 2004, Shaping small RNAs in plants by gene duplication, *Nat. Genet*. **36**:1245-1246.

Voinnet, O., and Baulcombe, D.C., 1997, Systemic signaling in gene silencing, *Nature* **389**:553.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A., 2002, Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi, *Science* **297**:1833-1837.

Volpe, T., Schramke, V., Hamilton, G.L., White, S.A., Teng, G., Martienssen, R.A., and Allshire, R.C., 2003, RNA interference is required for normal centromere function in fission yeast, *Chromosome Res*. **11**:137-146.

Wang, M.-B., Abbott, D., and Waterhouse, P.M., 2000, A single copy of a virus derived transgene encoding hairpin RNA gives immunity to barley yellow dwarf virus, *Molec. Plant Path*. **1**:401-410.

Wassenegger, M., Heimes, S., Riedel, L., and Sanger, H.L., 1994, RNA-directed de novo methylation of genomic sequences in plants, *Cell* **76**:567-576.

Waterhouse, P.M., Graham, M.W., and Wang, M.B., 1998, Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA, *Proc. Natl. Acad. Sci*. USA **95**:13959-13964.

Wesley, S.V., Helliwell, C.A., Smith, N.A., Wang, M., Rouse, D.T., Liu, Q., Gooding, P.S., Singh, S.P., Abbott, D., Stoutjesdijk, P.A., Robinson, S.P., Gleave, A.P., Green, A.G., and Waterhouse, P.M., 2001, Construct design for efficient, effective and high-throughput gene silencing in plants, *Plant J*. **27**:581-590.

Wielopolska, A., Townley, H., Moore, I., Waterhouse, P. and Helliwell, C., 2005, A high-throughput inducible RNAi vector for plants, *Plant Biotech. J*. **3**:583-590.

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., and Carrington, J.C., 2005, Expression of *Arabidopsis* miRNA genes, *Plant Physiol*. **138**:2145-2154.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C., 2004, Genetic and functional diversification of small RNA pathways in plants, *PLoS Biol*. **2**:E104.

Xie, Z., Kasschau, K.D., and Carrington, J.C., 2003, Negative feedback regulation of *Dicer-Like1* in *Arabidopsis* by microRNA-guided mRNA degradation, *Curr. Biol*. **13**:784-789.

Xiong, A.S., Yao, Q.H., Peng, R.H., Li, X., Han, P.L., and Fan, H.Q., 2005, Different effects on ACC oxidase gene silencing triggered by RNA interference in transgenic tomato, *Plant Cell Rep*. **23**:639-646.

Yoo, B.C., Kragler, F., Varkonyi-Gasic, E., Haywood, V., Archer-Evans, S., Lee, Y.M., Lough, T.J., and Lucas, W.J., 2004, A systemic small RNA signaling system in plants, *Plant Cell* **16**:1979-2000.

Zamore, P.D., 2004, Plant RNAi: How a viral silencing suppressor inactivates siRNA, *Curr. Biol*. **14**:R198-200.

Zilberman, D., Cao, X., and Jacobsen, S., 2003, ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation *Science* **299**: 716-719.

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E., 2004, Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats, *Curr. Biol*. **14**:1214-1220.

6
Control of Gene Expression by mRNA Transport and Turnover

Carole L. Bassett

6.1. Introduction

Newly synthesized mRNA is neither naked nor static. Transcription is coupled to the assembly of mRNPs (messenger ribonucleoprotein particles) so that RNA-binding factors which are recruited to active transcription sites are available for immediate construction of the mRNP complex. These factors specify the processing, export, subcellular location, and stability of the mRNA. Assembly of the mRNP proceeds in an orderly fashion beginning with the association of the cap-binding protein complex. This step is followed by the splicing-dependent assembly of proteins at the exon junctions of intron-containing genes and the addition of adaptor proteins to facilitate RNA export. In this way it can be said that transcription is physically and functionally coupled to the pre-mRNA processing events that lead to the final, export-competent mRNP. The relationships among these processes are summarized in Figure 6.1.

6.2. mRNA Transport and Localization

mRNA transport, localization, and translation are dynamic processes, requiring a constant restructuring of the mRNA-protein complex (reviewed in Van De Bor and Davis, 2004). These processes, which have been documented in many eukaryotic cell types (Kindler *et al*., 2005), share components with each other, as well as with other aspects of RNA metabolism, including transcription, translation and turnover. Sorting out the significance of interactions between components of the various processes is a monumental task facing biologists, since many such interactions are transient and rely on differential posttranslational modification of select proteins. Multiple genes encoding proteins associated with transport and localization add another dimension of complexity, as some can act as 'backup' genes for more than one process.

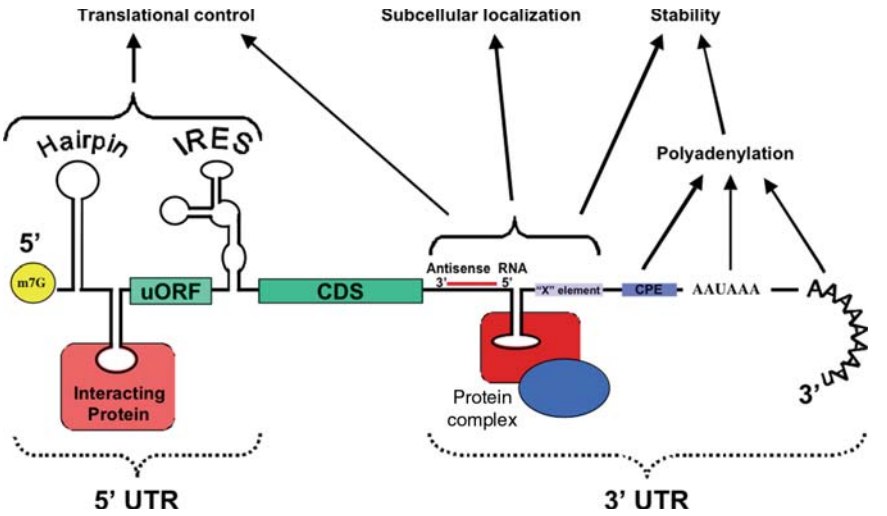


FIGURE 6.1. The generic structure of a eukaryotic mRNA, illustrating some posttranscriptional regulatory elements that affect gene expression. Abbreviations (from 5′ to 3′): UTR, untranslated region; m7G, 7-methyl-guanosine cap; hairpin, hairpin-like secondary structures; uORF, upstream open reading frame; IRES, internal ribosome entry site; CDS, coding sequence; "x" element, zipcode; CPE, cytoplasmic polyadenylation element; AAUAAA, polyadenylation signal. Reproduced with permission, from Mignone *et al*., 2002, *Genome Biology*, 3(3), 0004.1-0004.10. © *Genome Biology*. (*See Color Plates*)

6.2.1. mRNA Transport

Numerous studies have shown that the transport machinery is loaded onto the mRNA co-transcriptionally (Lei *et al*., 2001; Maniatis and Reed, 2002; Neugebauer, 2002). In addition, although splicing is not a requirement for mRNA transport *per se*, it can enhance transport of certain intron-containing transcripts (Le Hir *et al*., 2001; Wiegand *et al*., 2003). Furthermore, some proteins of the EJC (Exon Junction Complex) complex are required for mRNA transport, but they bind equally to intron-containing and intronless mRNAs (Linder and Stutz, 2001). The most crucial event for the acquisition of export competency appears to be 3′ end cleavage and polyadenlyation (Hilleren *et al*., 2001; Hammell *et al*., 2002). Several proteins have been linked to both export and 3′ processing. Hrp1p and Nab2p are shuttling proteins that have been associated with poly(A) tailing (Kessler *et al*., 1997; Hector *et al*., 2002). Although it would seem from these studies that interaction between the poly(A) binding protein and Hrp1 or Nab2 proteins is a given, the molecular mechanisms underlying transport and poly(A) tail coupling are poorly understood.

Studies of the Balbiani ring granules (mRNPs) of *Chironomus tentans* indicate that mRNPs exit the nucleus in a 5′ to 3′ manner (Mehlin *et al*., 1992).

Although most transcripts exit the nucleus without selectivity regarding which nucleopore is used, there are recent examples where such specificity has been shown to occur. As an example, consider the recent observation that when β2 tubulin transcripts aggregated at the posterior side of the nucleus in *Chlamydomonas*, nuclear pore complexes were also seen to preferentially accumulate posteriorly (Colon-Ramos *et al*., 2003).

As stated earlier, the transport particle is dynamic with many protein-RNA and protein-protein interactions taking place, although certain core proteins have been linked to general transport (reviewed in Farina and Singer, 2002; Palacios, 2002; Figure 6.2.). In addition to a "general" export pathway, there is evidence for specific pathways that facilitate export of subsets of mRNAs. For example, the eukaryotic initiation factor 5A (eIF5A) was originally thought to be associated with translation. However, subsequent studies failed to confirm a direct role in protein synthesis. It seems that eIF5A is not really an initiation factor at all, but rather has a major role in mRNA transport (Rosorius *et al*., 1999). The protein is activated posttranslationally by the conversion of a specific lysine to the unusual amino acid, hypusine. In *Arabidopsis* there are three isoforms, namely AteIF-5A1, AteIF-5A2, and AteIF-5A3. AteIF-5A1 and 2 are associated with various aspects of cell death, *e.g*. senescence and wounding, while AteIF-5A3 appears to be involved in cell division (Thompson *et al*., 2004). It is thought that eIF5A plays a role in the selective transport of a particular mRNA or subset of mRNAs (Hanauske-Abel *et al*., 1995). eIF-5A isoforms are found in tomato (Wang *et al*., 2001) and tobacco (Chamot and Kuhlemeier, 1992), and there is a recent report (Zhou *et al*., 2006) of the isolation of three eIF5A genes from Chinese Spring wheat.

The general mRNA export pathway consists of a number of core components (illustrated in Figure 6.2). Yra1p/REF/Aly (Yra1p in yeast; REF or Aly in mammals) is an export adaptor that partners with Sub2p/UAP56, an ATPase/DEAD-box RNA helicase and associates with mRNA during transcription. UAP56 and REF/Aly are also components of the EJC associated with splicing. This is consistent with the observation that, although splicing can enhance mRNA export, it is not required, since intronless mRNAs are as efficiently exported as spliced mRNAs. The Yra1p-Sub2p complex recruits the heterodimeric Mex67p-Mtr2p protein which releases Sub2p and facilitates export by mediating interaction of the mRNBP complex with nucleoporins lining the nucleopore. Interestingly, Yra1p is indispensable in yeast (Strasser and Hurt, 2000), whereas REF is nonessential in both *Drosophila* and *Caenorhabditis elegans* (Gatfield and Izaurralde, 2002). These observations and several other lines of evidence suggest that there may be other adaptors in *Drosophila* and *C. elegans* that complement the function of REF. The most recent evidence links Yra1p and Sub2p with a tetromeric complex called THO *via* interaction between Sub2p and one of the THO monomers, Hpr1p. The Sub2p-Yra1 dimer can be co-purified with THO in a complex called TREX. THO apparently facilitates the binding of Sub2p-Yra1p to the nascent mRNA. Another critical component is the heterodimeric Thp1p-Sac3 complex in which

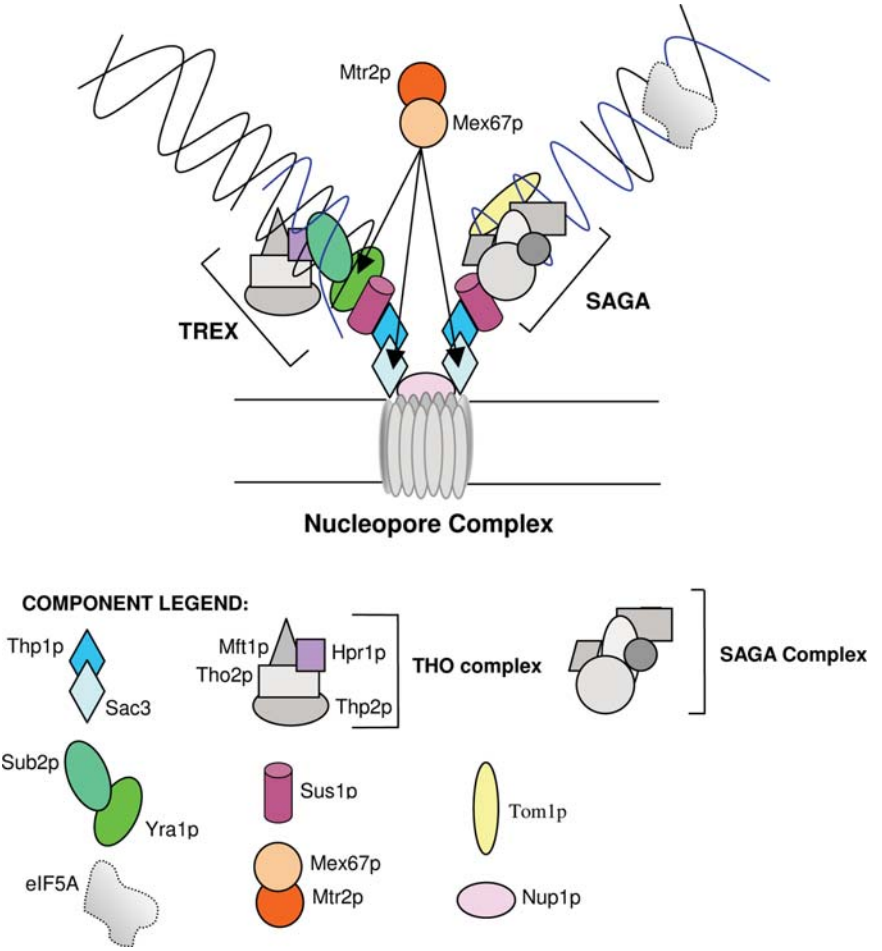


FIGURE 6.2. Core components of the mRNA export machinery. The TREX and SAGA complexes couple export to transcription initiation and the early steps of elongation. It is thought that SAGA directs export of a subset of mRNAs. The TREX complex consists of the THO complex coupled to the Sub2p and Yra1p dimer through interaction between Yra1p and Hpr1p. Tom1p associates with the SAGA complex, but does not co-purify with it. Protein components of the general export pathway are identified under the legend in the figure. DNA strands are shown in black; RNA strands are shown in dark blue. After Vinciguerra and Stutz, 2004. (*See Color Plates*)

Sac3 tethers Thp1p to the nuclear periphery. Association with the newly synthesized mRNBP complex is through a recently identified protein called Sus1p which interacts with Thp1p and Yra1p and the SAGA complex (Figure 6.2). The SAGA complex is a large intranuclear histone acetylase complex associated with transcription initiation of a subset of RNA Pol II genes (Lee *et al.*,

2000). Rsp5p and Tom1p are both E3 ubiquitin ligases required for mRNA export in yeast. Tom1p is physically associated with the SAGA complex, but it is not clear if Tom1p and Rsp5p are components of the same or different export pathways. Several other protein components required for or ancillary to mRNA export have also been identified and the list keeps growing. The reader is directed to two recent very good reviews (Vinciguerra and Stutz, 2004; Aguilera, 2005) for more in-depth information.

No homologs of the THO, or Sub2/UAP56 export proteins have been reported in plants to date. However, several laboratories have identified protein partners associating with the tomato bushy stunt virus (TBSV) P19 polypeptide that have high homology to Aly proteins (Storozhenko *et al*., 2001; Park *et al*., 2004; Uhrig *et al*., 2004). One study identified four Aly homologs in *Arabidopsis* and two in tobacco (Uhrig *et al*., 2004). Interestingly, like their animal counterparts, the plant Aly proteins localized to the nucleus; however, unlike other eukaryotic Aly proteins, five of the six plant Aly proteins were also found associated with both the nucleoplasm and nucleolus. Infection with TBSV altered the location of Aly2 from the nucleus to the cytoplasm in epidermal leaf cells, but did not affect P19-induced PTGS. At present it is not clear exactly what role(s) the plant Aly proteins play in addition to their possible association with selective mRNA transport.

6.2.2. mRNA Localization

Intracellular localization of mRNA is a common mechanism for targeting proteins to specific cellular regions, particularly in polarized cell types. Most often the localization process occurs *via* microtubules (tubulin polymers) or microfilaments (F-actin) using so-called molecular motors to drive movement along the cytoskeleton. Based on studies in *Drosophila*, yeast and cultured mammalian cells, three classes of molecular motors responsible for active mRNA transport have been reported (reviewed in Tekotte and Davis, 2002): myosin V associated with transport of yeast *Ash1* mRNA (Bertrand *et al*., 1998), Kinesin I associated with transport of *Drosophila oskar* mRNA (Brendza *et al*., 2000), and Dynein associated with transport of several *Drosophila* mRNBPs (Wilkie and Davis, 2001; Januschke *et al*., 2002). However, there are also examples of mRNA transport by diffusion (see Forrest and Gavis, 2003, for example). Furthermore, mRNA localization signals are complex; *cis*-acting elements that bind transport factors are probably not truly sequence-specific, but also utilize aspects of secondary and tertiary structure. This explains why it is difficult to find consensus sequences associated with such binding sites. Also, subsets of mRNAs may co-localize, for example Vg1 and VegT, without sharing obvious sequence or structural features (Bubunenko *et al*., 2002).

Specific localization of mRNAs in plants has not been as extensively characterized as it has in vertebrates and invertebrates. Nevertheless, the subject has recently experienced renewed interest as indicated by numerous reviews related to the plant cytoskeleton and molecular motors (Okita and Choi,

2002; Klyachko, 2005). In plants intracellular movement appears to depend more on actin filaments than on microtubules, in contrast to observations in animal models. As a result, plants may rely more on myosins as the principal motors for cytoskeletal movement than on kinesins and dyneins. Interestingly, the *Arabidopsis* genome contains some 60 kinesin genes, compared to 6 in yeast and 40 in humans (Reddy and Day, 2001; Lee and Liu, 2004). Furthermore, although lower plants apparently have genes encoding dynein, the *Arabidopsis* genome has none (Lawrence *et al*., 2001).

The establishment of polarity in the egg cell and embryo sac of lower and higher plants has been well documented (reviewed in Souter and Lindsey, 2000). In *Arabidopsis thaliana* (Mansfield and Briarty, 1991) and *Capsella bursa-pastoris* (Schulz and Jensen, 1968) apical-basal polarity is obvious within the egg cell itself before the first zygotic division takes place. For example, in *Arabidopsis* a large vacuole is asymmetrically located toward the micropyle end of the egg cell with the chalazal (opposite) end being more cytoplasmic. Additional studies using the brown alga, *Fucus,* as a model for fertilization and embryogenesis have provided some insight as to how plant polarity in early development is initiated and maintained. Correlated with formation of the zygotic axis is an asymmetric distribution of RNAs, including the mRNA for G-actin (F-actin monomer) which accumulates at the opposite pole from where the F-actin is located (Bouget *et al*., 1996). Recently, Han *et al*. (2006) have used transmission electron microscopy to follow polysome formation in pine egg cells before and after fertilization. Their studies indicated that upon fertilization, all polysomes were attached to microtubules until formation of the polarized embryo. At that time the polysomes became disaggregated and randomly distributed in the cytoplasm. These observations were associated exclusively with microtubules and no association of polysomes with microfilaments was observed.

A role for RNA trafficking associated with microtubules has also been suggested by global analysis of tubulin-binding proteins in *Arabidopsis* (Chuong *et al*., 2004). The results of this study identified several types of RNA-binding proteins and translation factors that are consistent with the existence of RNA localization and translation machineries operating in association with the cytoskeleton. Compartmentalization of protein synthesis and storage has long been known for cereal seed storage proteins. Recently, the localization of mRNAs encoding rice seed storage proteins has been studied in detail (Crofts *et al*., 2004; Crofts *et al*., 2005). These studies have led to the recognition of the key role mRNA localization plays in determining the site of seed storage protein accumulation.

6.2.3. RNA Granules

Some mRNAs are available for immediate translation after transport into the cytoplasm. After the mRNA is translated it undergoes deadenylation and the polysomes are disassembled. The deadenylated mRNA is then either targeted

for degradation or is stored. Other mRNAs are not translated immediately after synthesis, but may be stored pre- or post-transport, awaiting developmental or environmental signals to initiate translation. mRNAs that are translationally silenced are assembled into RNA granules comprising different classes based on the composition of protein components making up the granule (reviewed in Anderson and Kedersha, 2002). Proteins commonly associated with mRNA granules include helicases, certain ribosomal subunit proteins, mRNA binding proteins, translation factors, decay enzymes, and scaffolding proteins. Four classes of mRNA granules are recognized based on studies in yeast and mammalian cells: germ cell granules, stress granules, neuronal granules, and processing bodies. Although each class has a different repertoire of proteins, the classes do share some components and may use similar mechanisms to control mRNA translation and turnover (Table 6.1.).

The class of germ cell granules has been well documented during animal embryonic development, including *Xenopus laevis* (germinal granules), *Drosophila melanogaster* (called polar granules) and *Caenorhabditis elegans* (called P granules). These granules contain maternal mRNAs required for germ cell specification (Schisa *et al*., 2001; Leatherman and Jongens, 2003) by directing the timing of maternal mRNA translation required to promote germ cell development in the early embryo. Different aspects of mRNA metabolism are represented in the proteins making up germ cell granules (Table 6.1.) consistent with the role of these granules in regulating maternal mRNA expression.

TABLE 6.1. Protein composition of mRNA granules.

Role of associated proteins	mRNA granule class: Germ cell granules	Stress granules[1]	Neuronal granules	Processing bodies[1]	References
TRANSLATION					
Ribosomal proteins					
40S	ND[2]	+	+	–[3]	Kedersha *et al*. 2002; Krichevsky and Kosik 2001
60S	ND	–	+	–[3]	Krichevsky and Kosik 2001
Eukaryotic initiation factors					
eIF2	ND	+	+	–[3]	Kedersha *et al*., 2002; Krichevsky and Kosik, 2001; Anderson and Kedersha, 2006
eIF3	ND	+	ND	–	Kedersha *et al*., 2002; Kedersha *et al*., 2005
eIF4E[1]	+	+	+	+	Amiri *et al*., 2001; Kedersha *et al*., 2002; Smart *et al*., 2003; Kedersha *et al*., 2005;

(*Continued*)

TABLE 6.1. Protein composition of mRNA granules. — Cont'd

Role of associated proteins	mRNA granule class: Germ cell granules	Stress granules[1]	Neuronal granules	Processing bodies[1]	References
eIF5A	+	ND	ND	ND	Anderson and Kedersha, 2006
eIF4E-T	ND	ND	ND	+	Ferraiuolo *et al*., 2005
PABP[1]	ND	+	–	–	Kedersha *et al*., 1999; Krichevsky and Kosik, 2001; Kedersha *et al*., 2005
DECAY					
Decapping enzymes					
DCP1 and 2[1]	+	–	ND	+	Bashkirov *et al*., 1997; Bailey-Serres, 1999; Ingelfinger *et al*., 2002; Kedersha *et al*., 2005; Lall *et al*., 2005
CGH-1	+	+	ND	+	Cougot *et al*., 2004; Navarro and Blackwell, 2005; Wilczynska *et al*., 2005
CAR-1	+[4]	ND	ND	ND	Anderson and Kedersha, 2006
Exonucleases					
XRN1	ND	+	ND	+	Bashkirov *et al*., 1997; Kedersha *et al*., 2005
Exosome Components[1]	ND	ND	ND	+	Leib *et al*., 1998; Graham *et al*., 2006
STABILITY					
HuR	ND	+	+	+[3]	Gallouzi *et al*., 2000; Atlas *et al*., 2004
TTP	ND	+	ND	+	Stoecklin *et al*., 2004
OTHER					
RNA binding proteins					
Staufen	ND	+	+	+[3]	Thomas *et al*., 2005
SMN	ND	+	ND	ND	Hua and Zhou, 2004
G3BP	ND	+	+	–	Tourriere *et al*., 2003, Atlas *et al*., 2004; Kedersha *et al*., 2005
Si RNAs					
GW182	ND	–	ND	+	Eystathioy *et al*., 2003; Kedersha *et al*., 2005;
Argonaute	ND	+[3]	ND	+	Liu *et al*., 2005

Taken from Andersen and Kedersha, 2006.
[1] Also found in plants (Bailey-Serres, 1999)
[2] ND = not determined
[3] From unpublished data cited in Andersen and Kedersha, 2006
[4] CAR-1 is related to proteins (Lsm proteins) that regulate mRNA splicing, decapping and decay

The existence of germ cell granules has not been demonstrated in plants. Egg and pollen formation are quite distinct from their counterparts in the animal kingdom. However, since this aspect of mRNA metabolism has not been as extensively explored in plants, it is possible that such granules exist in plant germ cells or that homologs of genes whose products make up such granules exist in plants, but perform divergent functions. Indeed, *aubergine*, a PAZ domain protein associated with polar bodies, has a role both in polar body formation and in gene silencing. The *Arabidopsis aubergine*-like genes associated with PTGS, *argonaute* and *zwille*, have preserved the gene silencing function.

A second class of mRNA granules relates to environmental stress responses which include heat shock, exposure to UV irradiation, hypoxia, and oxidative stress. Under these conditions the translation of "housekeeping" mRNAs is circumvented in favor of the translation of stress responsive genes like molecular chaperones and damage repair enzymes. Consistent with this observation is the discovery that most stress granules contain stalled 48S preinitiation complexes. For example, cells that have been heat shocked contain most of their mRNAs in stress granules, leaving the heat shock mRNAs free for translation. Selective recruitment of mRNAs into these granules is thought to regulate their stability and delay their translation (Anderson and Kedersha, 2002). Figure 6.3. illustrates differences in translation between normal cells and cells experiencing stress. Therefore, stress granules appear to act as staging centers for sorting and/or remodeling specific mRNA transcripts for reinitiation, decay or packaging into untranslated mRNPs. Unlike other classes of granules, stress granules are absent from normally growing cells and are rapidly induced within minutes of exposure to the stress stimulus.

Stress granules have been described in plants. Their association with heat shock as heat shock granules or HSGs was first described in tomato leaves (Nover *et al*., 1983). In addition to the association of mRNAs with HSGs, they were also shown to contain small heat-shock proteins (HSPs) (Nover and Scharf, 1984). Furthermore, recent evidence suggests that a form of HsfA2, a transcription factor responsive to heat shock, is also associated with HSGs (Scharf *et al*., 1998). Analysis of the HSGs indicates that they contain a distinct subset of proteins that differs from other mRNA binding proteins, for example, like those seen in prosomes (Nover *et al*., 1989). This suggests that the RNA content of plant HSGs is selective, *i.e*., although they contain mRNAs encoding housekeeping genes, newly synthesized heat-shock mRNAs are specifically excluded. However, other studies indicate that isolated RNPs bind both stress-induced and repressed mRNAs, suggesting that selectivity is not dependent on the proteins (Stuger *et al*., 1999). The HSG-associated RNAs were translationally inactive, but could be translated *in vitro*, and could also be translated after the cells had recovered from the stress.

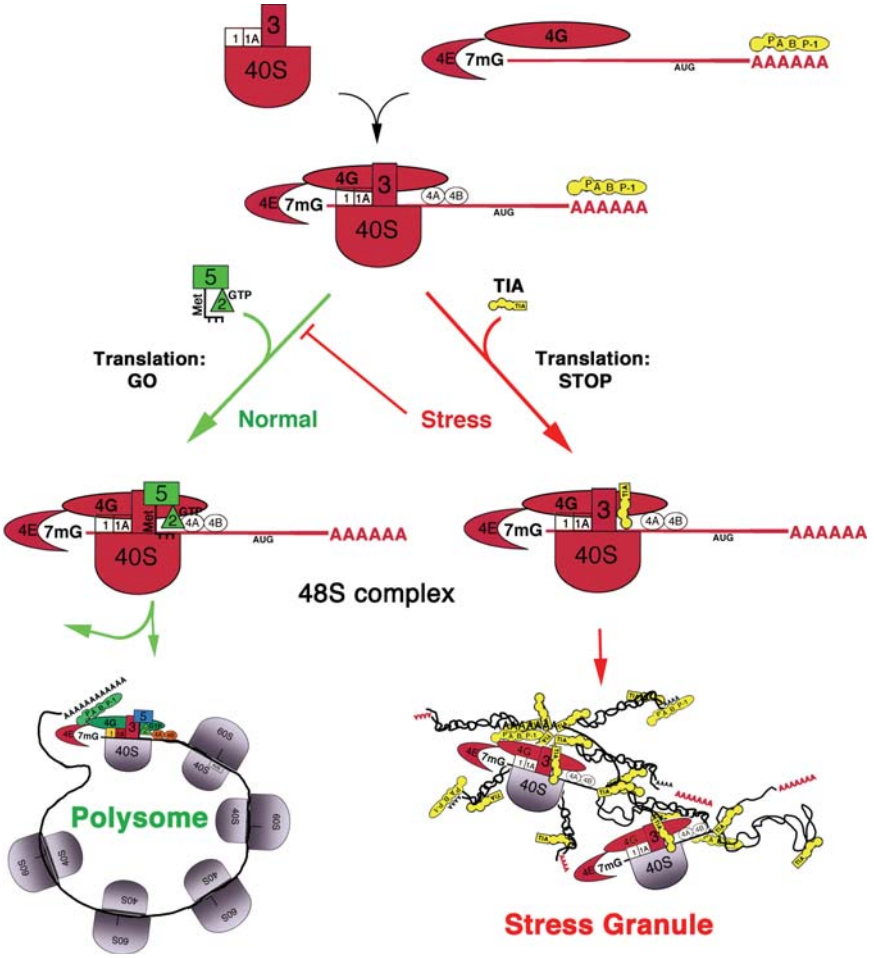


FIGURE 6.3. Translational initiation in the absence or presence of stress. In the absence of stress, the eIF2–GTP–tRNAMet ternary complex (green) is available to form a canonical 48S preinitiation complex at the 5′ end of capped transcripts (green arrow: normal) and scanning begins. Upon recognition of the initiation codon by the anticodon of tRNAMet, eIF5 promotes GTP hydrolysis and early initiation factors are displaced by the 60S ribosomal subunit. As additional ribosomes are added to the transcript, the mRNA is converted into a polysome (bottom left). In stressed cells, phosphorylation of eIF2a depletes the stores of eIF2–GTP–tRNAMet, and the cytoplasmic amount of TIA-1 (yellow) increases. Under these conditions, TIA-1 is included in a non-canonical, eIF2/eIF5-deficient preinitiation complex that is translationally silent. TIA-1 auto-aggregation then promotes the accumulation of these complexes at discrete cytoplasmic foci known as SGs (composed of all components of the 48S pre-initiation complex except eIF2 and eIF5). Translational components that are present in SGs are shown in red; non-translation proteins present in SGs are yellow, those absent from SGs are shown in green. Reproduced with permission, from Kedersha and Anderson, 2002, Biochemical Society Transactions 30:963-969. © The Biochemical Society. (*See Color Plates*)

Neuronal granules comprise the third class of mRNA granules. This class has evolved to facilitate the long distance transfer of signals from the cell body to the synaptic junctions to activate protein synthesis near the cell surface. Neuronal granules contain translationally silenced transcripts that are transported to the dendritic synapses of neurons where they are released and translated in response to specific stimuli (Krichevsky and Kosik, 2001; reviewed in Kosik and Krichevsky, 2002). Although intact ribosomes are associated with neuronal granules, the transcripts are translationally repressed until the granules reach the synapse where a stimulus activates translation (Huttelmaier *et al*., 2005). Since long distance transfer of signals in plants operates through a different mechanism(s), it is unlikely that neuronal-type granules would be found in plants, unless they are performing a function unrelated to their role in signal transfer between neurons.

The fourth class of granule is called the processing body (P-body) type. It is a distinct cytoplasmic granule containing components of the 5′→3′ mRNA decay machinery, the nonsense-mediated decay (NMD) pathway (Sheth and Parker, 2006) and the RNA-induced silencing complex (Jabri, 2005; Sen and Blau, 2005; see Section 6.4.2.5 and Chapter 5). There is a nuclear version that contains the cytoplasmic P-body core proteins, as well as additional components exclusively associated with the nucleus (Allmang *et al*., 1999). These complexes were hypothesized to represent a discrete type of mRNA decay focus, and this was confirmed by studies in yeast demonstrating that mRNAs blocked from 5′→3′ exonuclease attack accumulated at these foci (Sheth and Parker, 2006). P-bodies are sites of both translational repression and mRNA turnover (reviewed in Bruno and Wilkinson, 2006), two distinct processes that may share certain protein components. That these processes are separate is illustrated by the observation that translatable mRNAs can be released from P-bodies, indicating that they have not been substantially degraded (Bhattacharyya *et al*., 2006). On the other hand, aberrant mRNAs are targeted to P-bodies where they undergo rapid decay (Sheth and Parker, 2006). Liu *et al*. (2005) have recently shown that microRNAs direct targeted mRNAs to P-bodies, although it is not completely clear how the targeted messages are repressed.

Aside from studies on silencing in plants suggesting the presence of P-bodies, there is little information regarding this subset of granules in plants. Since P-bodies are found in a wide variety of eukaryotes and perform a basic cellular function, it is likely that they also occur in plants. A recent bioinformatic study of the Scd6p family of proteins revealed that certain regions of the protein were conserved across a wide variety of organisms, including *Arabidopsis* (Anantharaman and Aravind, 2004). The authors speculate that since at least one member of the Scd6p family (Rap55) has been shown to associate with cytoplasmic RNPs (Lieb *et al*., 1998), it is possible that other members of this family associated with mRNA degradation could also be localized to P-bodies. Another piece of circumstantial evidence suggesting that plants might have P-bodies comes from studies in *Drosophila* showing

that components of the exosome (see Section 6.4.1.) localize to cytoplasmic foci that may in fact be P-bodies (Graham *et al*., 2006; see also Table 1). Since *Arabidopsis* also has exosome components (Table 6.2, Chekanova *et al*., 2000; Chekanova *et al*., 2002), it is possible that they likewise associate with P-bodies. Although it seems logical that plants should utilize P-bodies as do Metazoans, there is as yet no direct evidence that they do.

6.2.4. Nuclear Compartments

Subnuclear compartmentation in the form of "speckles" (also known as splicing factor compartments or SFCs) has been shown to be a property of the splicing machinery. In mammalian cells SFCs can occupy as much as 20% of the total nuclear volume (Misteli, 2000). Splicing takes place concurrently with transcription and appears to be only tangentially associated with SFCs, yet both mRNA and splicing factors are found closely coupled with SFCs. It

TABLE 6.2. Core components of the exosome complex.

Yeast[1] proteins	Description	*Drosophila*[2]	Human[3]	*Arabidopsis*[4]
Rrp4p[5]	Novel distributive hydrolytic exonuclease	dRrp4	hRrp4p	AtRrp4 [At1g03360]
Rrp40p	Novel hydrolytic exonuclease	dRrp40	hRrp40p	[At4g32175; At2g25355]
Rrp41p/Ski6p	RNase-PH-like[6]; processive, phosphorolytic enzyme	dRrp41	hRrp41p	AtRrp41 [At3g46210]
Rrp42p	RNase-PH-like	dRrp42	hRrp42p	[At3g07750]
Rrp43p	RNase-PH-like		NF[7]	[At1g60080]
Rrp44p/Dis3	RNase II-like; processive hydrolytic enzyme	taz	hDis3p	NF
Rrp45p	RNase-PH-like	dRrp45	PM-Scl75	[At3g60500; At3g12990]
Rrp46p	RNase-PH-like	dRrp46	hRrp46p	NF
Mtr3p	RNase-PH-like	dMtr3	NF	NF
Cs14p	S1-type RNA-binding domain	dCs14	hCs14p	NF
Rrp6p[8]	*E. coli* RNaseD-like	dRrp6	PM-Sc1100[3]	[At1g54440; At2g32415]

[1] Allmang *et al*., 1999
[2] Cairrao *et al*., 2005; GenBank
[3] Allmang *et al*., 1999
[4] Chekanova *et al*., 2000; Chekanova *et al*., 2002; [GenBank – note there are numerous exonucleases annotated in the *Arabidopsis* genome, some of which may represent the missing yeast components]
[5] Rrp4p: ribosomal RNA processing 4 protein
[6] RNase-PH: an *E. coli* inorganic phosphate-dependent exoribonuclease that releases nucleoside diphosphates, not monophosphates
[7] NF: not found; sequences homologous to yeast mtr3 and Rrp43p were found, but they were more closely related to Rrp45p.
[8] Associates exclusively with nuclear version of exosome

is thought that the SFCs may act as a reservoir for splicing factors that can be recruited to genes actively being transcribed. However, the role of mRNAs found in the SFCs is at present unclear.

SFCs (speckles) have also been described in plants. A distinctive speckled organization of the splicing factor atRSp31 was seen in differentiated *Arabidopsis* root cells (Docquier *et al*., 2004). Although the speckles resembled aspects of mammalian SFCs, it is not clear whether they are structurally the same. Interestingly, the speckled organization of the atRSp31 protein was thought to reflect a demand for splicing factors in response to actively transcribed genes, a role also attributed to mammalian SFCs.

Cajal bodies are associated with pre-mRNA splicing, 3′ maturation of histone mRNAs, biogenesis, modification, and transport of snoRNPs, trafficking of telomerase RNA, and pre-rRNA processing (Shaw and Brown, 2004; Cioce and Lamond, 2005; Fu and Collins, 2006). They contain basal transcription factors, cell cycle factors, splicing small nuclear ribonucleoproteins (snRNPs), and nucleolar factors such as small nucleolar ribonucleoproteins (snoRNPs) (Gall *et al*., 1999; Matera, 1999). Although splicing snRNPs which are components of the spliceosome can be found in Cajal bodies, these particles differ from spliceosomes in that they do not contain non-snRNP splicing factors or nascent pre-mRNA (Dietz, 1998). The snRNPs in the Cajal body appear to be derived specifically from newly assembled particles prior to associating with SFCs (Sleeman and Lamond, 1999). This observation suggests that Cajal bodies may be involved in coordinating the assembly of nuclear snRNPs and might be part of an snRNP maturation pathway involving both types of particles (Sleeman and Lamond, 1999; Jady *et al*., 2003).

Cajal bodies have also been described in plants (Moreno Diaz de la Espina *et al*., 1980). Their function in plants has been essentially the same as that proposed from studies in animal cells, although an additional role in the transport of snoRNA precursors has been suggested (Shaw *et al*., 1998; Boudonck *et al*., 1999). A recent study in maize and *Arabidopsis* revealed Cajal bodies (see below) in quiescent cells and germinating embryos of both species (Docquier *et al*., 2004). In maize, resumption of mitotic activity in the meristem is related to a decrease in Cajal body formation. Cajal bodies in tobacco and *Arabidopsis* decrease directly after cell division (Boudonck *et al*., 1999).

The latest study implicates plant Cajal bodies in RNA-directed DNA methylation (RdDM), a major mechanism of gene silencing in plants (see below and Chapter 5). Li *et al*. (2006) co-localized the *Arabidopsis* ARGONAUTE4 (AGO4) with proteins known to be associated with Cajal bodies. They hypothesize that AGO associates with the Cajal body where it complexes with siRNAs formed by the action of other components of the silencing machinery. After associating with siRNA, the AGO complex leaves the Cajal body and can then activate the appropriate DNA methylase to complete gene silencing by RdDM.

6.3. mRNA Binding Factors

Throughout their "lives" mRNAs form mRNA particles by association with numerous proteins and microRNAs (miRNAs), some of which are stably bound and some of which participate in temporary dynamic combinations that affect the structure and function of the mRNA (reviewed in Moore, 2005; Wilusz and Wilusz, 2004). Most of these components have multiple roles and their influence on mRNA stability may be an accessory function. A few mRNA binding proteins (mRNBPs) are non-sequence specific, like most of the Y-box proteins, while the rest typically bind select RNAs using specific sequence targets generally found in the 5′ or 3′ untranslated regions (UTRs). This selectivity in binding provides a level of control of gene expression essentially equivalent to more commonly recognized controls on transcription itself. In so doing, specific groups of RNAs carrying the same recognition sequences can be regulated by the same signaling pathways, thus allowing a rapid and coordinated response to diverse stimuli. On the other hand, the importance of this aspect of regulation, *i.e*. posttranscriptional, is underscored by the frequent observation that RNA and protein levels do not always correlate, and indicate that changes in RNA levels measured by popular molecular techniques, such as microarrays, real-time PCR and RNA blotting, should never be presumed to reflect protein levels or activity.

6.3.1. mRNPs

Proteins are associated with mRNAs from the beginning to the end. Proteins may associate without any apparent sequence specificity, as do, for example, the cap forming proteins, or they may bind to discrete sequences within the RNA, particularly at exon/intron junctions and in the untranslated regions. Some proteins provide structural stability to the mRNA to delay degradation or to assist in transport from the nucleus, while others may aid in targeting a particular mRNA for degradation.

6.3.1.1. mRNA Capping Proteins

Addition of the 7-methylguanylated mRNA cap (m7GpppN) begins when the nascent mRNA contains around 25 nt, a length that allows it to protrude from the RNA exit channel of the RNA polymerase II (Pol II) (Moteki and Price, 2002). Two transcription elongation factors bind to the Pol II causing a pause in transcription and allowing the 5′ end of the pre-mRNA to be capped by the sequential action of an RNA 5′ triphosphatase, a guanylyltransferase, and a methyltransferase. The capping machinery and elongation factors are displaced, and a third elongation factor, P-TEFb, associates with the C-terminal domain of the Pol II. This association which leads to massive phosphorylation of Ser5 disrupts the interaction between Pol II and N-TEF (a negative transcription elongation factor complex) and stimulates the

resumption of transcription (Marshall *et al*., 1996; Ping and Rana, 2001). The pause appears to serve as a checkpoint to prevent extension of improperly capped mRNAs. Binding of the cap-binding protein complex (CBC) occurs when the mRNA cap is in place. CBC is composed of the cap-binding protein, CBP20, and CBP80 which associates with the carboxyterminus of the Pol II to ensure high-affinity binding (Lejeune *et al*., 2002). Once the first intron is transcribed, the core splicing machinery is recruited to the intron, and the exon-junction complex assembles onto the exon-exon junction created by the looping out of the intron.

Capping reactions are essential for cell growth and maintenance, since initiation of translation generally requires the mRNAs to have a 5′ cap. Although a cap-independent pathway exists, it is generally associated with non-capped viral RNAs and utilizes an internal ribosome entry site (IRES) to by-pass the cap requirement (Pelletier and Sonenberg 1988). In fungi, the three cap-synthesizing activities are found in separate enzymes, but in animals and plants the triphosphatase and guanyltransferase functions are encoded by a single bifunctional enzyme (Shuman, 2000). Null mutations in any one of the three enzymes in *Saccharomyces cerevisiae* are lethal; however, in *Candida albicans* the triphosphatase and methyltransferase activities can be eliminated without significantly affecting survival (Dunyak *et al*. 2002). Analysis of mRNAs from these mutants indicates the presence of aberrant cap structures, one in the triphosphatase mutants (GppppN) and one (GpppN) in the methyltransferase-minus mutants. Apparently the guanine portion of the cap structure is the critical element for cap recognition in subsequent cap-requiring processes in this species. It has been previously demonstrated that capping is associated primarily with mRNA stability (protection at the 5′ end) and translation initiation (interaction with eIF4).

In plants there are three different cap binding proteins, one of which, eIFiso4E, appears to be unique to plants (reviewed in Bailey-Serres, 1999). Homologs to the plant eIF4E and cap binding protein (CBP) are also found in mammals. Wheat eIF4E and eIFiso4E are differentially expressed during development in *Arabidopsis* (Rodriguez *et al*., 1998), raising the possibility that translation shows selectivity towards certain subsets of mRNAs during embryogenesis.

6.3.1.2. Structural and Regulatory mRNBPs

Many proteins bind to the noncoding portions of mRNAs and influence stability or cellular location of the bound mRNA. This binding may be nonspecific or it may be sequence and/or structure specific. Numerous elements that are implicated in mRNA regulation through the binding of these proteins have been identified in the 5′ and 3′ untranslated regions. A searchable database listing the various sequence elements identified to date can be found at http://www.ba.itb.cnr.it/UTR/. It has been predicted that hundreds of mRNBPs are encoded in a typical eukaryotic genome (Costanzo *et al*., 2001);

however, only a few of these proteins have been characterized and even fewer have been assigned specific functions. Some of the better characterized mRNBPs are discussed below.

At least five RNA binding motifs are recognized in mRNBPs. The RNP domain is primarily associated with proteins that participate in mRNA splicing (Nagai *et al*., 1995). The KH motif is found primarily, but not exclusively, in hnRNP proteins (Siomi *et al*., 1993). RRM (RNA recognition motif) and dsRBD (double-stranded RNA-binding domain) proteins interact with double-stranded RNA and will not bind to either single-stranded RNA or dsDNA (St. Johnston *et al*., 1992). The fifth domain is perhaps the oldest recognized RNA binding domain: the S1 domain named after the ribosomal S1 protein of *E. coli* (Subramanian, 1983).

6.3.1.3. *cis* Sequences and Proteins Associated with the 3′ UTR

Perhaps the most well known *cis-trans* factor combination in the 3′ UTR is that of the poly(A)-binding protein (PABP) and the poly(A) tail found in most eukaryotic mRNAs. It has long been recognized that the PABP contributes to the stability of the mRNA and that its absence leads to disassembly of the poly(A) tail, which is the first step in mRNA degradation by the two major pathways (discussed in Section 6.4.1). It is interesting to note that in yeast and insects there is only a single gene encoding PABP (Sachs *et al*., 1987; Sigrist *et al*., 2000; Thakurta *et al*., 2002), whereas two genes are found in *C. elegans* and three in humans (Mangus *et al*., 2003). Surprisingly, eight genes have been identified in *Arabidopsis* (Belostotsky, 2003) and they exhibit a wide range of expression patterns (Hilson *et al*., 1993; Belostotsky and Meagher, 1993; Belostotsky and Meagher, 1996; Palanivelu *et al*., 2000b; Belostotsky, 2003). Both *Arabidopsis* PAB2 and PAB5 genes complemented a yeast *pab1* mutant restoring viability, but only PAB2 was able to complement the poly(A) shortening and defect in translation initiation seen with *pab1* mutants (Palanivelu *et al*., 2000a). Complementation of *pab1* with PAB3 suggests a role in mRNA biogenesis and export (Chekanova *et al*., 2001; Chekanova and Belostotsky, 2003). The PAB C-terminal domain (PABC) has been shown to mediate protein-protein interactions and to recruit proteins to the mRNA-RNP complex. Six of the *Arabidopsis* PABs have the conserved PABC domain, and a recent study has identified some of the interacting partners *via* the yeast two-hybrid system (Bravo *et al*., 2005), suggesting that the relatively large number of PABPs and interacting proteins may have evolved in plants to provide greater flexibility in modulating expression of specific transcripts. Furthermore, they also demonstrated that PAB2 and PAB5 are essential for viability.

Another example of the importance of 3′ UTR sequences in mRNA function is the use of specific "zipcodes" designed to deliver transcripts to different cellular compartments (Singer, 1993). The zipcodes consist of short sequences recognized by specific proteins that direct the flow of information

in the cell. Disruption of this flow has been associated with human genetic diseases, such as myotonic dystrophy which results from expansion of a triplet repeat in the 3′ UTR that causes retention of the mRNA in the nucleus (Davis *et al*., 1997).

Some sequence-specific mRNBPs provide increased stability to the mRNA substrate by binding to certain sequence elements, usually in the 3′ UTR. For example, enhanced stability of the β-globin mRNA derives from a combination of factors that include binding of the α-complex, a member of the hnRNP-E family of mRNA binding proteins, to a specific stem-loop structure which enhances binding of the poly(A) binding protein and simultaneously blocks access to the 3′ UTR by an erythroid cell-specific endoribonuclease (Jiang *et al*., 2006). Likewise, the Elav-like family of proteins in mammalian cells binds AU-rich elements (AREs) in the 3′ UTRs of several mRNAs and increases their stability (Joseph *et al*., 1998).

AU-rich elements in the 3′ UTR have long been known to regulate addition of the poly(A) tail and thus stabilize mRNAs. This process was discussed in detail in Chapter 4. However, some AREs target specific mRNAs or classes of mRNAs for degradation. AREs are grouped experimentally into classes. Class I and II AREs have multiple copies of the element, AUUUA. Class I elements regulate cytoplasmic deadenylation in an organized fashion resulting in the formation of mRNAs with shorter, (30–60 nucleotides) poly(A) tails that are subsequently degraded completely. Class II AREs control an asynchronous process of deadenylation in the cytoplasm which results in the complete removal of the poly(A) tail. Class III AREs do not have the consensus AUUUA sequence, but do possess a U-rich region. They govern kinetics of deadenylation similar to that of the class I elements. Proteins that bind specifically to these elements include the ARE/poly(U)-binding/degradation factor (AUF1), tristetraprolin (TTP), T-cell internal antigen 1/TIA-1-related (TIA-1/TIAR); others have also been implicated. There is recent evidence that ARE-targeted mRNAs are degraded primarily through the 5′→3′ degradation pathway in mammals (Stoecklin *et al*., 2006).

TTP regulates expression of the tumor necrosis factor alpha (TNF-α) through its binding to a conserved ARE in the 3′ UTR of the TNF-α mRNA (Lai *et al*., 1999). The mechanism through which TTP acts to destabilize TNF-α mRNA is by the recruitment of factors involved in both deadenylation and exosome degradation. TTP has also been implicated in the posttranscriptional regulation of other mRNAs, including COX-2 (Phillips *et al*., 2004), interleukin-2 (Raghavan *et al*., 2001), and inducible nitric oxide synthase (Fechir *et al*., 2005).

The PUF (*Drosophila* Pumilio and *Caenorhabditis* FBF) family of proteins suppresses mRNA expression through the regulated localization, translation and/or degradation of select mRNAs in diverse organisms (Wickens *et al*., 2002). In every case examined to date, PUF requires the interaction of another protein partner to fulfill its regulatory role, and the identity of the partners varies with the specific mRNA and organism. In addition to yeasts,

insects, and mammals, PUF proteins have also been found in *A. thaliana*, rice, and poplar, indicating that this family of proteins has been well conserved in eukaryotes (Wickens *et al*., 2002). The plant sequences have only recently been deposited in GenBank, and at present there is no information regarding their substrates or mechanism(s) of action in these plants compared to their animal homologs.

6.3.1.4. *cis* Sequences and Proteins Associated with the 5′ Leader

The most common association between the 5′ leader of a given mRNA and specific proteins is the interaction between the mRNA 5′ leader and ribosome through its various initiation factors. In eukaryotes this interaction is for the most part cap-dependent; however, in some viruses and possibly in certain cellular mRNAs a cap-independent ribosome binding site provides an alternative mechanism for translation initiation. This site, known as the internal ribosome entry site (IRES), has been well documented with respect to viral mRNAs, but is still somewhat controversial regarding its use in cellular mRNAs (Kozak, 2005). Several excellent reviews are available covering conventional, alternative, and selective translation initiation (Jackson, 2005; Pisarev *et al*., 2005; Jang, 2006; Bailey-Serres, 1999), and therefore the topics will not be covered here. In addition, upstream AUGs also interact with the translation machinery and influence translation. mRNAs that contain these upstream open reading frames (uORFs) generally encode regulatory proteins such as kinases and transcription factors. Regulation of the translation of these mRNAs provides a back-up mechanism for keeping their expression at relatively low levels, since higher expression could interfere with critical cellular functions. Regulation of translation by these sequences was discussed in detail in Chapters 1 and 2.

CUG binding proteins (CUGBPs) bind to repeats of CUG found in the 5′ and 3′ UTRs of certain mammalian mRNAs. In one case CUGBP1 has been shown to regulate translation of the CCAAT enhancer binding protein beta (C/EBPβ) transcription factor by binding CUG repeats in the 5′ leader of the C/EBPβ mRNA and influencing the choice of initiation codons which alters the levels of the three C/EBPβ isoforms (Timchenko *et al*., 1999). HuR, a member of the Elav-like protein family that binds AU-rich elements in the 3′ UTRs of select mRNAs, can also bind to a U-rich element within the 5′ leader of a specific mRNA (p27) and regulate its translation (Millard *et al*., 2000).

A recent report of a CAUU-repeat sequence in the 5′ leader of the tobacco *Ferredoxin-1* (*Fed-1*) mRNA that regulates decline of the mRNA in the dark might qualify as a potential site for an mRNBP (Bhat *et al*., 2004). Evidence that the dark decline is independent of polysomal association strongly suggests that this event operates through another mechanism, and not translation *per se*. Of interest is the observation that a minimum of four repeats and a maximum of eleven repeats were required for the full dark

effect. This is similar to the expansion of mammalian CUG repeats which at some threshold level can cause disease.

6.3.1.5. Other mRNA Binding Proteins and Factors

Serine/arginine rich (SR) proteins are a class of mRNBPs with one or more highly conserved RRM domains, usually at the N-terminus, and an arginine/serine-rich (RS) domain at the C-terminus. SR proteins in animals function as essential splicing factors, although some are shuttled between the nucleus and cytoplasm and thought to be involved with export (Huang and Steitz, 2001), stability (Lemaire *et al*., 2002) and/or translation (Sanford, *et al*., 2004). In *Arabidopsis* SR proteins comprise a family of 19 proteins ranging from 21–45 kDa (Lorković and Barta, 2002; reviewed in Reddy, 2004). Four of these are similar to ASD/SF2 (linked to mRNA stability; Lemaire *et al*., 2002) except they have a C-terminal extension unique to plants. Although one of these SR proteins, SR1, failed to complement HeLa cells, it did promote splice site switching in HeLa nuclear extracts (Lazar *et al*., 1995). Five *Arabidopsis* SR proteins are similar to SC35, and three are like 9G8 in humans. The remaining seven are unique to plants. Some of the *Arabidopsis* SR proteins have homologs in other plants, and a few of these have been shown to complement splicing-deficient extracts or alter splicing *in vivo* (Lopato *et al*., 1996; Lopato *et al*., 1999, Kalyna *et al*., 2003).

One mechanism prokaryotes utilize to regulate gene expression is through binding of specific metabolites to certain RNA domains (aka riboswitches). It is thought that this is an ancient mechanism for regulating gene expression, possibly predating the emergence of proteins. An example of a well characterized riboswitch is the thiamine pyrophosphate (TPP) sensor found in the 5′ leader of *E. coli* genes, *thiM* and *thiC,* that regulate biosynthesis of thiamine (Winkler *et al*., 2002). Until recently it was not known whether this type of regulation was restricted only to prokaryotes. However, Sudarsan *et al*. (2003) have shown that this method of genetic control appears to have been evolutionarily conserved, being found in fungi and plants. They demonstrated that, like TPP regulation of the thiamine operon in *E. coli*, the TPP cofactor can bind the mRNA of a thiamine biosynthetic gene from *Arabidopsis* and alter its structure, thus potentially acting as a sensor of thiamine levels. In contrast to *E. coli* where the *cis*-binding sequences are located in the 5′ leader, the *Arabidopsis* TPP binding sequences are located in the 3′ UTR.

6.3.1.6. miRNA

MicroRNAs (miRNAs) are short (ca. 21 nts) noncoding, single-strand RNA species found in diverse organisms. They are synthesized from primary transcripts that are first processed in the nucleus to produce a hairpin RNA of *ca*. 70 nt which is exported to the cytoplasm. In the cytoplasm, DICER, a nuclease associated with silencing, cuts the hairpin, and one of the two resulting

strands is incorporated into the RNA inducing silencing complex (discussed in detail in Chapter 5).

miRNAs down-regulate gene expression by one of two mechanisms depending on the degree of complementarity between a given miRNA and its target mRNA. If complementarity is weak, miRNAs act by translational repression (Reinhart *et al*., 2000); if complementarity is strong, they act by directing cleavage of the target RNA Yekta *et al*., 2004). RNAi and miRNA share some common cellular machinery during their maturation, and the mechanisms by which they function are similar, *i.e*. endogenous miRNAs can cleave mRNAs with perfect complementarity and likewise, introduced siRNA can translationally repress mRNAs having imperfect complementary binding sites (Doench *et al*., 2003; Saxena *et al*., 2003).

As was discussed previously in Section 6.3.1.3, AU-rich elements in the 3′ UTRs of many mRNAs can regulate the degradation of these transcripts by binding certain proteins that aid in recruiting polypeptides associated with decay pathways. Jing *et al*. (2005) have shown a specific miRNA is also involved in contributing to the instability of ARE-containing mRNAs in *Drosophila via* DICER-activated degradation. On the other hand, repression of translation at the initiation step is the mechanism utilized by the Let-7 miRNA in human cells (Pillai *et al*., 2005).

In *Arabidopsis* there are at least 42 distinct families of miRNAs with a total of 118 members identified (Jones-Rhoades *et al*., 2006). Characterization of miR172 responsible for repression of APETALA2 (AP2) has revealed a dual mechanism for its effects on flowering. Chen (2004) first showed that a translational repression mechanism similar to Let-7 was involved with AP2 regulation. However, recent research has shown that RISC-mediated mRNA degradation also plays a key role in the regulation of AP2 (Schwab *et al*., 2005). The best explanation for these observations is a scenario where both mechanisms contribute significantly to the overall abundance of AP2. Another report implicates sequences in the 3′ UTR of the soybean glutamine synthase (GS1) gene in both mRNA turnover and translational repression (Ortega *et al*., 2006). Comparison of a construct containing the complete GS1 gene with one where the entire 3′ UTR was deleted revealed that differences in mRNA turnover were influenced by nitrogen status. In contrast, the effect of these same constructs on translation repression operated by a mechanism that was independent of nitrogen status and nodule formation. The possibility that a miRNA might be involved in the observed translation repression of GS1 has not yet been tested, but it would seem to be one logical avenue to explore.

6.4. mRNA Turnover

The fact that mRNAs encoding functionally related proteins appear to be coordinately regulated after transcription is the basis of an hypothesis that mRNA particles behave like posttranscriptional “operons” (Keene and

Tenenbaum, 2002). mRNA function, *i.e.*, translatability, is related to mRNA abundance in the cytoplasm, availability (location) for engaging the translation machinery, and efficiency in initiating translation. mRNA abundance in turn depends on transcription, transport and turnover rates. Cytoplasmic RNA turnover is governed by exonucleases that typically act processively and endonucleases associated with silencing. RNA turnover is influenced by the predetermined half life of each RNA and by the presence of premature stop codons (nonsense mediated decay or NMD) or the absence of any stop codon (nonstop mRNA decay).

6.4.1. General mRNA Decay

Most of the general mRNA decay machinery is found in P-bodies (discussed in Section 6.2.3.). These particles surround mRNAs that are not actively being translated and degrade them. The mechanism(s) of removal of an mRNA from the translation pool are not currently known, except for that associated with miRNAs and the RISC complex. Interestingly, mRNAs found in stress granules that are not actively being translated are not targets for degradation while associated with these granules (Kedersha *et al.*, 2005). Indeed, under stress conditions selection of individual RNAs for degradation may depend on staging through the SGs.

Based primarily on studies in yeast, two pathways of general mRNA degradation are known in eukaryotic cells. In both pathways mRNA degradation typically begins with deadenylation of the poly(A) tail by poly(A)-specific 3′→5′ exoribonucleases variously designated as PARN [poly(A) ribonuclease], DAN [deadenylation nuclease] and PAN [poly(A) nuclease] (Boeck *et al.*, 1996; Couttet *et al.*, 1997; Korner *et al.*, 1998). This step is the rate-determining step in general mRNA degradation.

Degradation by the first pathway begins at the 3′ end and may depend on the presence of A/T-rich elements (AREs) in the 3′ UTR (Mukherjee *et al.*, 2002), thus exhibiting a type of sequence-specificity. Catalysis continues in a 3′→5′ manner driven by a multi-subunit complex of 3′→5′ exonucleases termed the exosome. Although the exosome is usually associated with ribosomal RNA processing in the nucleus, a subset of exonucleases have been shown to be important for cytoplasmic mRNA degradation (Anderson and Parker, 1998). This pathway is also found in the nucleus where it operates in conjunction with a 5′→3′ exonuclease and competes with splicing (Bousquet-Antonelli *et al.*, 2000). A comparison of core exosome components is illustrated in Table 6.2.

The second pathway (the deadenylation-dependent decapping pathway) begins with removal of the 7-MeCap by the DCP1-DCP2 complex (Long and McNally, 2003; Jacobson, 2004), followed by 5′→3′ degradation by XRN1 [in yeast] (Stevens, 2001). A mutational block in DCP1 or XRN1 is not lethal; mRNA degradation proceeds *via* the 3′→5′ pathway, albeit at a slower rate overall. Likewise, mutations in the genes of proteins associated with the

3′→5′ pathway are not lethal, but result in the slow accumulation of mRNAs (Ridley *et al*., 1984). In contrast, mutation of components in both pathways simultaneously is lethal (Johnson and Kolodner, 1995), confirming that these two pathways are primary routes of mRNA degradation.

A third pathway (deadenylation-independent decapping pathway) exists in which RNAs are initially cleaved by endonucleases such as DICER. This pathway is associated with nonsense mediated decay and RNAi (see below) and, in association with subsequent degradation of the RNA fragments generated by the XRN1 and exosome pathways, is the major vehicle for the surveillance and silencing functions (Gatfield and Izaurralde, 2004; Orban and Izaurralde, 2005). Other endonucleases also contribute to general mRNA degradation, although in a minor way. These enzymes often exhibit sequence specificity towards the 3′ UTR of select messages. Some operate independently of deadenylation, for example, the transferrin receptor mRNA (Binder *et al*., 1994), while others require deadenylation of the mRNA substrate, *i.e*. degradation of the α-globin mRNA by ErEN (Rogers *et al*., 2002).

Regarding deadenylation of mRNAs in yeast and *C. elegans*, it seems that at least one of the PARN genes can be reduced or eliminated without obvious consequences on development or morphology. In *Arabidopsis* there is one putative PARN gene and one PARN-like gene that is missing sequence elements required for activity. When the latter gene was knocked out, no apparent effect on phenotype was observed (Reverdatto *et al*., 2006). In contrast, when the putative PARN homolog was knocked out, an embryo-lethal phenotype was obtained, suggesting that this gene is required for some aspect of development. Interestingly in the absence of the PARN homolog, some, but not all, of the embryo-specific transcripts became hyperadenylated. This suggests that deadenylation of a specific subset of mRNAs is the key to regulating embryonic development in wild type plants.

6.4.2. mRNA Surveillance

Early observations on bacterial gene expression using nonsense mutations to abolish specific proteins indicated that the lack of protein was due to the enhanced degradation of the cognate mRNA. Similar observations in eukaryotes suggested that all organisms possessed a "search and destroy" system for recognizing and eliminating faulty mRNAs. The obvious rationale for such a system is the prevention of the accumulation of truncated polypeptides, since it is well documented that truncated polypeptides can interfere in a dominant manner with the activity of the hetero- or homopolymeric proteins with which they naturally associate. Therefore, some mechanism(s) must be responsible for recognition and elimination of aberrant mRNAs.

Two types of aberrant mRNAs are typically found, the first being those RNAs containing one or more premature termination (nonsense) codons and the second being those RNAs lacking any stop codons. The nonsense-mediated decay (NMD) pathway removes the former mRNAs, while the

nonstop-mediated (NSD) pathway eliminates the latter. Another type of surveillance pathway recently discovered is associated with the removal of mRNAs which cause ribosome stalling. Yet a fourth recently described pathway relates to the removal of mRNAs whose protein products are unfolded or misfolded on the endoplasmic reticulum. The processes impacting these pathways are summarized in Table 6.3. All of these pathways are designed to prevent the accumulation of nonfunctional translation machinery or aberrant polypeptide products.

6.4.2.1. Nonsense-mediated Decay

Probably the first case of nonsense-mediated decay (NMD) reported in the literature was that of the effect of premature termination codons (PTCs) on URA-3 mRNAs in yeast (Losson and Lacroute, 1979). Since then, we now understand that NMD is a decay pathway for a variety of aberrant mRNAs that arise from either mutation or defects in pre-mRNA splicing which alter the normal spatial relationship between the termination codon and other RNA features (Oliveira and McCarthy, 1995; Muhlrad and Parker, 1999). NMD (reviewed in Maquat, 2004) promotes decay of targeted mRNAs through both the 5′→3′ and 3′→5′ pathways in mammals and yeast, and in most cases requires prior deadenylation of the mRNAs (Chen and Shyu, 2003; Lejeune *et al*., 2003; Mitchell and Tollervey, 2003; Muhlrad and Parker, 1999). In contrast, NMD in *Drosophila* is initiated by endonucleolytic cleavage followed by rapid degradation of the resulting 5′ (exosome degradation) and 3′ (XRN1 degradation) fragments (Gatfield and Izaurralde, 2004).

Apart from a simple "house cleaning" function, NMD has also been implicated in translation control (for example, Nott *et al*., 2004), DNA repair (Brumbaugh *et al*., 2004) and the regulation of a subset of transcripts encoding transcription factors (Mendell *et al*., 2004), enzymes associated with amino acid metabolism (Mendell *et al*., 2004), and pseudogenes, transposable elements and retroviruses (Mitrovich and Anderson, 2005). It has also recently been implicated in X chromosome inactivation *via* regulation of the

TABLE 6.3. Processes required for different types of surveillance.

Required biological feature	NMD	NSD	NG	IRE1
Translation	+[a]	+	+	+
Splicing	+[b]	ND	ND	ND
Deadenylation	+	+	–	ND
Ribosomal Stalling	–	+	+	+
cis elements	+[c]	–	+[d]	+[e]

[a] + indicates the feature is required; – indicates the feature is not required; ND indicates the requirement for the feature has not been determined.
[b] may be more dependent on splicing in mammalian cells than in fungi, insects or plants
[c] Includes premature termination codon as a *cis*-element
[d] Shown with CGS1 gene in *Arabidopsis* (Onouchi *et al*., 2005)
[e] located in the 3′ UTR

noncoding *Xist* RNA (Ciaudo et al., 2006). Although this pathway operates from yeast to humans, it is non-essential in yeast and *C. elegans* (Weischenfeldt, *et al*., 2005), and has features that are distinct between yeast and humans. For example, PTC recognition which is dependent on splicing in mammals, is independent of splicing in yeast and insects. Most mammalian systems studied to date suggest that NMD takes place in the nucleus, although a minority occurrence in the cytoplasm has been noted (Cheng and Maquat, 1993; Moriarty *et al*., 1998).

A *cis*-acting element (DSE) that destabilizes mRNAs when downstream of a nonsense codon is required to distinguish the normal stop codon from the PTC in yeast (Zhang *et al*., 1995), whereas in mammals the PTC is recognized by its relative position (>50 nucleotides upstream of the last exon-exon junction) (Zhang *et al*, 1998). This implies that NMD in mammals requires an intron to act on the target mRNA; in support of this is the observation that intronless mRNAs are immune to NMD (Maquat and Li, 2001). Further suggestion that splicing is required comes from studies that indicate the EJC deposited some 20–24 nucleotides upstream of the last exon-exon junction after splicing is complete may be a marker required for NMD recognition (Le Hir *et al*., 2000a and 2000b).

Little is known about NMD in plants. In contrast to observations in mammalian systems, several reports have suggested that PTC-containing mRNAs transcribed from intronless genes are degraded in plants, suggesting NMD could proceed by an EJC-independent mechanism or by more than one mechanism, one of which is EJC-independent (Voelker *et al*., 1990; van Hoof, and Green, 1996; Isshiki *et al*., 2001). Homologs for most of the genes encoding proteins associated with the EJC in other organisms have been found in the *Arabidopsis* genome (Pendle *et al*., 2005). Hori and Watanabe (2005) have shown that one of the key required proteins in mammalian NMD, UPF3, is also required for NMD in *Arabidopsis*. Recently, through the use of mutations in upf1 and upf3, a requirement of these NMD-associated proteins has also been demonstrated in *Arabidopsis* (Arciga-Reys *et al*., 2006). In the upf1 mutants, NMD of both intron-containing and intron-less gene mRNAs was inhibited. In addition, phenotypic traits indicative of developmental problems were seen in both the upf1 and upf3 mutants, suggesting that both are involved with NMD.

6.4.2.2. Nonstop-mediated Decay

mRNAs lacking stop codons are one of several mechanisms that can cause ribosomal stalling. Such mRNAs can arise as a consequence of mutation (probably rare), aberrant processing, as intermediates in mRNA degradation and possibly through premature polyadenylation (addition of poly(A) directly to coding sequence). Failure to encounter a termination codon prevents the ribosome from releasing from the mRNA. If this were allowed to continue, the cell would soon exhaust its supply of free ribosomes, and initi-

ation of new polypeptides would cease. In addition, if the ribosome were to release the truncated polypeptides, negative effects on protein activity might soon follow, as truncated polypeptides often exhibit dominant negative influence on multi-subunit protein complexes. A recent study using nonstop and derivative transcripts in yeast and HeLa cells indicates that this process occurs primarily in the cytoplasm by a mechanism that is distinct from NMD (Frischmeyer *et al*., 2002; Inada and Aiba., 2005). In yeast, nonstop mRNAs are degraded by the exosome in a 3′→5′ manner (van Hoof *et al*., 2002).

Release of stalled ribosomes in bacteria and eukaryotic DNA-containing organelles appears to occur by a *trans*-translation pathway (Keiler *et al*., 1996; Williams, 2002; Gueneau de Novoa and Williams, 2004; Jacob *et al*., 2004) involving a unique RNA known as tmRNA (transfer messenger RNA). tmRNA is part tRNA, capable of being charged with alanine, and part messenger RNA, possessing a pseudo-ORF which lacks AUG but possesses a stop codon. The stalled ribosome transfers its nascent polypeptide to the alanine-charged tmRNA, resulting in ejection of bound mRNA and resumption of the reading frame within the tmRNA positioned in the ribosome A-site. Translation then proceeds on the tmRNA to an in-frame stop codon, whereupon the protein now containing the tmRNA "tag" is released. Interestingly, the tag directs the newly synthesized peptide for proteolysis, thus the stalled ribosome is freed and the incomplete polypeptide eliminated. Furthermore, in addition to directing aberrant polypeptides for proteolysis, tmRNA also facilitates the degradation of the ribosome-bound mRNA by releasing the stalled ribosomes (Yamamoto *et al*., 2003; Sunohara *et al*., 2004).

6.4.2.3. No-go mediated Decay

This recently discovered pathway (Doma and Parker, 2006) in yeast is associated with stalled ribosomes. Stalling on polysomes can occur as a result of mRNA secondary structure, the presence of rare codons, and pseudoknots. The no-go decay (NGD) pathway is independent of deadenylation, but is dependent on translation. Products formed by NGD were indicative of endonucleolytic cleavage and were not present in cells defective in the Dom34 product similar to other factors involved in translation termination, suggesting that Dom34 recognizes the stalled ribosome and triggers mRNA cleavage. This pathway is thought to provide a mechanism for releasing stalled ribosomes from damaged mRNA.

Consistent with this hypothesis is the recent observation that sequences in the *Arabidopsis* CGS1 mRNA not only trigger endonucleolytic cleavage, but also mediate ribosome stalling (Onouchi *et al*., 2005). In this case a specific element, MTO1, when translated in the presence of S-adenosylmethionine causes the ribosome to pause downstream at the codon for ser-94. Subsequent cleavage of the protruding 5′ mRNA results in posttranscriptional down-regulation of this gene. This is reported to be the first example of nascent peptide-mediated translation elongation arrest leading to mRNA degradation in eukaryotes and may represent yet another mechanism for dealing with stalled ribosomes.

6.4.2.4. The IRE1 System

The endoplasmic reticulum (ER) can recover from the accumulation of misfolded proteins by increasing its folding capacity *via* inositol-requiring enzyme-1 (IRE1) which detects misfolded proteins and activates an X-box-binding (XBB) transcription factor. Activation of the XBB protein is by endonucleolytic cleavage of its cognate mRNA. Recently, another function of IRE1 has been described which operates independently of the XBB protein and mediates the rapid degradation of a subset of mRNAs based on their association with the ER membrane and on the sequence of the polypeptides they encode (Hollien and Weissman, 2006).

Enrichment of prolamine RNAs on prolamine protein bodies of the ER (PB-ER) has been described in developing seeds of rice plants. This enrichment of prolamine RNAs has been suggested to be an important step for the efficient folding and assembly of prolamine polypeptides (Okita and Rogers, 1996). Several aspects of this process appear to relate to the IRE1 surveillance system: 1) both require translation of the mRNAs associated with the ER, 2) neither requires the RNA region encoding the signal peptide of the cognate polypeptides, and 3) both require an RNA "zipcode" located in the 3′ UTR for localization to the ER (Choi *et al*., 2000).

6.4.2.5. Other Types of Surveillance

Another type of surveillance mechanism is reflected in the silencing of specific genes. This form of inspection has been described in plants, animals and fungi (Fagard *et al*., 2000) and therefore appears to be an ancient mechanism in eukaryotes for eliminating foreign nucleic acids. The process is termed posttranscriptional gene silencing (PTGS) in plants, RNAi (RNA interference) in animals, and quelling in fungi. The pathway can operate through three distinct mechanisms, two of which occur posttranscriptionally, namely targeted RNA degradation and translation inhibition (Fagard *et al*., 2000, Hannon, 2002). The third mechanism acts directly on transcription by histone and/or DNA methylation (Volpe *et al*., 2002; Chan *et al*., 2005). Although these mechanisms are distinct they share numerous components. Some of these components and mechanisms related to plant PTGS are discussed in greater detail in Chapter 5. As mentioned earlier, this process is associated with processing bodies and has recently been linked to Cajal bodies by the observation that a major component of silencing (ARGONAUTE4) can associate with them (see section 6.2.4). This latter location correlates with the methylation-dependent aspect of silencing, while association with PBs may be more closely linked to posttranslational mechanisms.

A type of surveillance system unique to plants is the S-locus self-incompatibility system that features an S-locus RNase and governs the success or failure of self-pollination in several plant families (recently reviewed by Kao and Tsukamoto, 2004). Although actual surveillance is presumably at the level of protein-protein recognition, the mechanism whereby control is exerted is through pollen tube RNA degradation, and this degradation

appears to be non-sequence-specific. A T2-like RNase expressed in the pistil is the female component of the incompatibility response and is responsible for the pollen-specific degradation of RNA.

6.5. Summary and Prospectus

At the beginning of this chapter it was stated that transport and localization of mRNAs is highly complex with considerable overlap between various processes related to RNA metabolism. There are direct interactions of components of RNA transcription with splicing components, as well as with export factors. In addition splicing factors are also associated with translational regulation and surveillance pathways. These interactions are further complicated by the recognition that some, if not most, proteins have multiple functions and interact with diverse partners representing different metabolic processes. The implication is that cellular metabolism is directly and indirectly linked to changes in gene expression through a complicated web of interactions. The challenge will be to understand all these interactions, and how they determine the state of the cell in response to a variety of internal and external signals.

One area ripe for exploration is that of predicting mRNA secondary and tertiary structures, and how they influence gene expression. Prediction programs are improving, but they have not yet reached a point where empirical determinations can be eliminated. The interaction of mRNAs or mRNBPs with small molecules like lipids, metals, *etc*. could be another source of influence *via* RNA-protein interactions or perhaps by effects on secondary/tertiary structure. Opportunities for RNA-RNA or DNA-RNA interactions through secondary structures have not yet been explored.

Identifying and sorting out the large plant gene families encoding small RNAs is an aspect of gene regulation that needs to be developed further. Assigning mRNA targets to these microRNAs will be critical to our understanding of what processes they control and to determining the mechanism(s) employed to exert their influence. There may also be more examples of metabolites that directly bind to mRNAs and provide feedback regarding their synthesis and/or degradation. One exciting area of needed research is that of phylogenetic transcriptome analysis emphasizing mRNA turnover and transport as potential mechanisms where differences among organisms have evolved. The amount of information derived from genomic studies has been overwhelming, and it is not expected to diminish, at least in the near future. Despite the difficulties in dealing with genetic "information overload", our understanding of the complexity of living organisms has increased exponentially in the past decade, and it will not be long before the agricultural benefits from this new knowledge can be applied to the food and nutrition challenges that are anticipated to occur as the human population continues to increase.

References

Aguilera, A., 2005, Cotranscriptional mRNP assembly: from the DNA to the nuclear pore, *Curr. Opin. Cell Biol*. **17**:242-250.

Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P., 1999, The yeast exosome and human PM-Scl are related complexes of 3′→5′ exonucleases, *Genes Dev*. **13**:2148-2158.

Amiri, A., Keiper, B.D., Kawasaki, I., Fan, Y., Kohara, Y., Rhoads, R.E., and Strome, S., 2001, An isoform of eIF4E is a component of germ granules and is required for spermatogenesis in *C. elegans*, *Development* **128**:3899-3912.

Anantharaman, V., and Aravind, L., 2004, Novel conserved domains in proteins with predicted roles in eukaryotic cell-cycle regulation, decapping and RNA stability, *BMC Genomics* **5**:45.

Anderson, P., and Kedersha, N., 2002, Stressful initiations, *J. Cell Sci*. **115**:3227-3234.

Anderson, P., and Kedersha, N., 2006, RNA granules, *J. Cell Biol*. **172**:803-808.

Anderson, J.S.J., and Parker, R., 1998, The 3′ to 5′ degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3′ to 5′ exonucleases of the exosome complex, *EMBO J*. **17**:1497-1506.

Arciga-Reys, L., Wootton, L., Kieffer, M., and Davies, B., 2006, UPF1 is required for nonsense-mediated mRNA decay (NMD) and RNAi in *Arabidopsis, Plant J*. **47**:480-489.

Atlas, R., Behar, L., Elliott, E., and Ginzburg, I., 2004, The insulin-like growth factor mRNA binding-protein IMP-1 and the Ras-regulatory protein G3BP associate with tau mRNA and HuD protein in differentiated P19 neuronal cells, *J. Neurochem*. **89**:613-626.

Bailey-Serres, J., 1999, Selective translation of cytoplasmic mRNAs in plants. *TRENDS Plant Sci*. **4**:142-148.

Bashkirov, V.I., Scherthan, H., Solinger, J.A., Buerstedde, J.M., and Heyer, W.D., 1997, A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates, *J. Cell Biol*. **136**:761-773.

Belostotsky, D.A., 2003, Unexpected complexity of poly(A)-binding protein gene families in flowering plants. Three conserved lineages that are at least 200 million years old and possible auto- and cross-regulation, *Genetics* **163**:311-319.

Belostotsky, D.A., and Meagher, R.B., 1993, Differential organ-specific expression of three poly(A)-binding-protein genes from *Arabidopsis thaliana, Proc. Natl. Acad. Sci. USA* **90**:6686-6690.

Belostotsky, D.A., and Meagher, R.B., 1996, A pollen-, ovule-, and early embryo-specific poly(A) binding protein from *Arabidopsis* complements essential function in yeast, *Plant Cell* **8**:1261-1275.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M., 1998, Localization of ASH1 mRNA particles in living yeast, *Mol. Cell* **2**:437-445.

Bhat, S., Tang, L., Krueger, A.D., Smith, C.L., Ford, S.R., Dickey, L.F., and Petracek, M.E., 2004, the Fed-1 (CAUU)4 element is a 5′ UTR dark-responsive mRNA instability element that functions independently of dark-induced polyribosome dissociation, *Plant Mol. Biol*. **56**:761-773.

Bhattacharyya, S.N., Habermacher, R., Martine, U., Closs, E., and Filipowicz, W., 2006, Relief of microRNA-mediated translational repression in human cells subjected to stress, *Cell* **125**:1111-1124.

Binder, R., Horowitz, J.A., Basilion, J.P., Koeller, D.M., Klausner, R.D., and Harford, J.B., 1994, Evidence that the pathway of transferrin receptor mRNA degradation involves an endonucleolytic cleavage within the 3 UTR and does not involve poly(A) tail shortening, *EMBO J*. **13**:1969-1980.

Boeck, R., Tarun, Jr., S., Rieger, M., Deardorff, J.A., Müller-Auer, S., and Sachs, A.B., 1996, The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity, *J. Biol. Chem*., **271**:432-438.

Boudonck, K., Dolan, L., and Shaw, P.J., 1999, The movement of coiled bodies visualized in living plant cells by the green fluorescent protein, *Mol. Biol. Cell* **10**: 2297-2307.

Bouget, F.-Y., Gerttula, S., and Quatrano, R.S., 1996, Localization of actin mRNA during the establishment of cell polarity and early cell divisions in *Fucus* embryos, *Plant Cell* **8**:189-201.

Bousquet-Antonelli, C., Presutti, C., and Tollervey, D., 2000, Identification of a regulated pathway for nuclear pre-mRNA turnover, *Cell* **102**:765-775.

Bravo, J., Aguilar-Henonin, L., Olmedo, G., and Guzmán, P., 2005, four distinct classes of proteins as interaction partners of the PABC domain of *Arabidopsis thaliana* poly(A)-binding proteins, *Mol. Gen. Genomics* **272**:651-665.

Brendza, R.P., Serbus, L.R., Duffy, J.B., and Saxton, W.M., 2000, A function for kinesin I in the posterior transport of *oskar* mRNA and Staufen protein, *Science* **289**:2120-2122.

Brumbaugh, K.M., Otterness, D.M., Geisen, C., Oliveira, V., Brognard, J., Li, X., Lejeune, F., Tibbetts, R.S., Maquat, L.E., and Abraham, R.T., 2004, The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells, *Mol. Cell* **14**:585-598.

Bruno, I., and Wilkinson, M.F., 2006, P-bodies react to stress and nonsense, *Cell* **125**:1036-1038.

Bubunenko, M., Kress, T.L., Vempati, U.D., Mowry, K.L., and King, M.L., 2002, A consensus RNA signal that directs germ layer determinants to the vegetal cortex of *Xenopus* oocytes, *Dev. Biol*. **248**:82-92.

Cairrao, F., Arraiano, C., and Newbury, S., 2005, *Drosophila* gene tazman, an orthologue of the yeast exosome component Rrp44p/Dis3, is differentially expressed during development, *Dev. Dyn*. **232**:733-737.

Chamot, D., and Kuhlemeier, C., 1992, Differential expression of genes encoding the hypusine-containing translation initiation factor, eIF-5A, in tobacco, *Nucl. Acids Res*. **20**:665-669.

Chan, S.W., Henderson, I.R., and Jacobsen, S.E., 2005, Gardening the genome: DNA methylation in *Arabidopsis thaliana*, *Nat. Rev. Genet*. **6**:351-360.

Chekanova, J.A., and Belostotsky, D.A., 2003, evidence that poly(A) binding protein has an evolutionarily conserved function in facilitating mRNA biogenesis and export, *RNA* **9**:1476-1490.

Chekanova, J.A., Dutko, J.A., Mian, I.S., and Belostotsky, D.A., 2002, *Arabidopsis thaliana* exosome subunit AtRrp4p is a hydrolytic 3′→5′ exonuclease containing S1 and KH RNA-binding domains, *Nucl. Acids Res*. **30**:695-700.

Chekanova, J.A., Shaw, R.J., and Belostotsky, D.A., 2001, Analysis of an essential requirement for the poly(A) binding protein function using cross-species complementation, *Curr. Biol*. **11**:1207-1214.

Chekanova, J.A., Shaw, R.J., wills, M.A., and Belostotsky, D.A., 2000, Poly(A) tail-dependent exonuclease AtRrp41p from *Arabidopsis thaliana* rescues 5.8 S

rRNA processing and mRNA decay defects of the yeast ski6 mutant and is found in an exosome-sized complex in plant and yeast cells, *J. Biol. Chem*. **275**: 33158-33166.

Chen, C.Y., and Shyu, A.B., 2003, Rapid deadenylation triggered by a nonsense codon precedes decay of the RNA body in a mammalian cytoplasmic nonsense-mediated decay pathway, *Mol. Cell Biol*. **23**:4805-4813.

Chen, X., 2004, A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development, *Science* **303**:2022-2025.

Cheng, J., and Maquat, L.E., 1993, Nonsense codons can reduce the abundance of nuclear mRNA without affecting the abundance of pre-mRNA or the half-life of cytoplasmic mRNA, *Mol. Cell. Biol*. **13**:1892-1902.

Choi, S.-B., Wang, C., Muench, D.G., Ozawa, K., Franceschi, V.R., Wu, Y., and Okita, T.W., 2000, Messenger RNA targeting of rice seed storage proteins to specific ER subdomains, *Nature* **407**:765-767.

Chuong, S.D.X., Good, A.G., Taylor, G.J., Freeman, M.C., Moorhead, G.B.G., and Muench, D.G., 2004, Large-scale identification of tubulin-binding proteins provides insight on subcellular trafficking, metabolic channeling, and signaling in plant cells, *Mol. Cell. Proteomics* **3**:970-983.

Ciaudo, C., Bourdet, A., Cohen-Tannoudji, M., Dietz, H.C., Rougeulle, C., and Avner, P., 2006, Nuclear mRNA degradation pathway(s) are implicated in *Xist* regulation and X chromosome inactivation, *PLoS Genet*. **2**:874-881.

Cioce, M., and Lamond, A.I., 2005, Cajal bodies: a long history of discovery, *Annu. Rev. Cell. Dev. Biol*. 21:105-131.

Colon-Ramos, D.A., Salisbury, J.L., Senders, M.A., Shenoy, S.M., Singer, R.H., and Garcia-Blanco, M.A., 2003, Asymmetric distribution of nuclear pore complexes and the cytoplasmic localization of beta2-tubulin mRNA in *Chlamydomonas reinhardtii, Dev. Cell* **4**:941-952.

Costanzo, M.C., Crawford, M.E., Hirschman, J.E., Kranz, J.E., Olsen, P., *et al*., 2001, YPD, PombePD and WormPD: Model organism volumes of the BioKnowledge library, an integrated resource for protein information, *Nucl. Acids Res*. **29**:75-79.

Cougot, N., Babajko, S., and Seraphin, B., 2004, Cytoplasmic foci are sites of mRNA decay in human cells, *J. Cell Biol*. **165**:31-40.

Couttet, P., Fromont-Racine, M., Steel, D., Pictet, R., and Grange, T., 1997, Messenger RNA deadenylation precedes decapping in mammalian cells, *Proc. Nat. Acad. Sci. USA* **94**:5628-5633.

Crofts, A.J., Washida, H., Okita, T.W., Ogawa, M., Kumamaru, T., and Satoh, H., 2004, Targeting of proteins to endoplasmic reticulum-derived compartments in plants: The importance of RNA localization, *Plant Physiol*. **136**:3414-3419.

Crofts, A.J., Washida, H., Okita, T.W., Satoh, M., Ogawa, M., Kumamaru, T., and Satoh, H., 2005, The role of mRNA and protein sorting in seed storage protein synthesis, transport, and deposition, *Biochem. Cell Biol*. **83**:728-737.

Davis, B.M., McCurrach, M.E., Taneja, K.L., Singer, R.H., and Housman, D.E., 1997, Expansion of a CUG trinucleotide repeat in the 3′ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts, *Proc. Natl. Acad. Sci USA* **94**:7388-7393.

Dietz, H., 1998. Polishing the cutting edge of gems. *Nat. Genet*. **20**:321-322.

Docquier, S., Tillemans, V., Deltour, R., and Motte, P., 2004, Nuclear bodies and compartmentalization of pre-mRNA splicing factors in higher plants, *Chromosoma* **112**:255-266.

Doench, J.G., Petersen, C.P., and Sharp, P.A., 2003, siRNAs can function as miRNAs, *Genes Dev*. **17**:438-442.

Doma, M.K., and Parker, R., 2006, Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation, *Nature* **440**:561-564.

Dunyak, D.S., Everdeen, D.S., Albanese, J.G., and Quinn, C.L., 2002, Deletion of individual mRNA capping genes is unexpectedly not lethal to *Candida albicans* and results in modified mRNA cap structures, *Eukaryotic Cell* **1**:1010-1020.

Eystathioy, T., Jakymiw, A., Chan, E.K., Seraphin, B., Cougot, N., and Fritzler, M.J., 2003, The GW182 protein colocalizes with mRNA degradation associated proteins hDcp1 and hLSm4 in cytoplasmic GW bodies, *RNA* **9**:1171-1173.

Fagard, M., Boutet, S., Morel, J.-B., Bellini, C., and Vaucheret, H., 2000, AGO1, QDE-2, and RDE-1 are related proteins required for post-transcriptional gene silencing in plants, quelling in fungi, and RNA interference in animals, *Proc. Nat. Acad. Sci. USA* **97**:11650-11654.

Farina, K.L., and Singer, R.H., 2002, The nuclear connection in RNA transport and localization, *Trends Cell Biol*. **12**:466-472.

Fechir, M., Linker, K., Pautz, A., Hubrich, T., Forstermann, U., Rodriguez-Pascual, F., and Kleinert, H., 2005, Tristetraprolin regulates the expression of the human inducible nitric-oxide synthase gene, *Mol. Pharmacol*. **67**:2148-2161.

Ferraiuolo, M.A., Basak, S., Dostie, J., Murray, E.L., Schoenberg, D.R., and Sonenberg, N., 2005, A role for the eIF4E-binding protein 4E-T in P-body formation and mRNA decay, *J. Cell Biol*. **170**:913-924.

Forrest, K.M., and Gavis, E.R., 2003, Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in *Drosophila*, *Curr. Biol*. **13**:1159-1168.

Frischmeyer, P.A., van Hoof, A., O'Donnell, K., Guerrerio, A.L., Parker, R., and Dietz, H.C., 2002, Mechanism that eliminates transcripts lacking termination codons, *Science* **295**:2258-2261.

Fu, D., and Collins, K., 2006, Human telomerase and Cajal body ribonucleoproteins share a unique specificity of Sm protein association, *Genes Dev*. **20**:531-536.

Gall, J.G., Bellini, M., Wu, Z., and Murphy, C., 1999, Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes, *Mol. Biol. Cell* **10**:4385-4402.

Gallouzi, I.E., Brennan, C.M., Stenberg, M.G., Swanson, M.S., Eversole, A., Maizels, N., and Steitz, J.A., 2000, HuR binding to cytoplasmic mRNA is perturbed by heat shock. *Proc. Natl. Acad. Sci. USA* **97**:3073-3078.

Gatfield, D., and Izaurralde, E., 2002, REF1/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export, *J. Cell. Biol.* **159**: 579-588.

Gatfield, D., and Izaurralde, E., 2004, Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in *Drosophila*, *Nature* **429**:575-578.

Graham, A.C., Kiss, D.L., and Andrulis, E.D., 2006, Differential distribution of exosome subunits at the nuclear lamina and in cytoplasmic foci, *Mol. Biol Cell* **17**:1399-1409.

Gueneau de Novoa, P., and Williams, K.P., 2004, The tmRNA website: reductive evolution of tmRNA in plastids and other endosymbionts, *Nucl. Acids Res*. **32**: D104-D108.

Hammell, C.M., Gross, S., Zenklusen, D., Heath, C.V., Stutz, F., Moore, C., and Cole, C.N., 2002, Coupling of termination, 3′ processing, and mRNA export, *Mol. Cell. Biol*. **22**:6441-6457.

Hanauske-Abel, H.M., Slowinska, B., Zagulska, S., Wilson, R.C., Staiano-Coico, L., Hanauske, A.-R., McCaffrey, T., and Szabo, P., 1995, Detection of a sub-set of polysomal mRNAs associated with modulation of hypusine formation at the G1-S boundary: Proposal of a role for eIF-5A in onset of DNA replication, *FEBS Lett*. **366**:92-98.

Han, Y., Yu, J., Guo, F., and Watkins, S.C., 2006, Polysomes are associated with microtubules in fertilized eggs of Chinese pine (*Pinus tabulaeformis*), *Protoplasma* **227**:223-227.

Hannon, G.J., 2002, RNA interference. *Nature* **418**:244-251.

Hector, R.E., Nykamp, K.R., Dheur, S., Anderson, J.T., Non, P.J., Urbinati, C.R., Wilson, S.M., Minvielle-Sebastiz, L., and Swanson, M.S., 2002, Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export, *EMBO J*. **21**:1800-1810.

Hilleren, P., McCarthy, T., Rosbash, M., Parker, R., and Jensen, T.H., 2001, Quality control of mRNA 3′ end processing is linked to the nuclear exosome, *Nature* **413**:538-542.

Hilson, P., Carroll, K.L., and Masson, P.H., 1993, Molecular characterization of PAB2, a member of the multigene family coding for poly(A)-binding proteins in *Arabidopsis thaliana*, *Plant Physiol*. **103**:525-533.

Hollien, J., and Weissman, J. S., 2006, Decay of endoplasmic reticulum-localized mRNAs during the unfolded protein response, *Science* **313**:104-107.

Hori, H., and Watanabe, Y., 2005, UPF3 suppresses aberrant spliced mRNA in *Arabidopsis, Plant J*. **43**:530-540.

Hua, Y., and Zhou, J., 2004, Survival motor neuron protein facilitates assembly of stress granules, *FEBS Lett*. **572**:69-74.

Huang, Y., and Steitz, J.A., 2001, Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA, *Mol. Cell* **7**:899-905.

Huttelmaier, S., Zenklusen, D., Lederer, M., Dictenberg, J., Lorenz, M., Meng, X., Bassell, G.J., Condeelis, J., and Singer, R.H., 2005, Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1, *Nature* **438**:512-515.

Inada, T., and Aiba, H., 2005, Translation of aberrant mRNAs lacking a termination codon or with a shortened 3′-UTR is repressed after initiation in yeast, *EMBO J*. **24**:1584-1595.

Ingelfinger, D., Amdt-Jovin, D.J., Luhrmann, R., and Achsel, T., 2002, The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrn1 in distinct cytoplasmic foci, *RNA* **8**:1489-1501.

Isshiki, M., Yamamoto, Y., Satoh, H., and Shimamoto, K., 2001, Nonsense-mediated decay of mutant *waxy* mRNA in rice, *Plant Physiol*. **125**:1388-1395.

Jabri, E., 2005, P-bodies take a RISC. *Nat. Struct. Mol. Biol*. **12**:564.

Jackson, R.J., 2005, Alternative mechanisms of initiating translation of mammalian mRNAs, *Biochem. Soc. Trans*. **33**:1231-1241.

Jacob, Y., Seif, E., Paquet, P-O., and Lang, B.F., 2004, Loss of the mRNA-like region in mitochondrial tmRNAs of jakobids, *RNA* **10**:605-614.

Jacobson, A., 2004, Regulation of mRNA decay: decapping goes solo. *Mol. Cell*. **12**:1-2.

Jady, B.E., Darzacq, X., Tucker, K.E., Matera, A.G., Bertrand, E., and Kiss, T., 2003, Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm, *EMBO J*. **22**:1878-1888.

Jang, S.Y., 2006, Internal initiation: IRES elements of picornaviruses and hepatitis c virus, *Virus Res*. **119**:2-15.

Januschke, J., Gervais, L., Dass, S., Kaltschmidt, J.A., Lopec-Schier, H., Johnston, D.S., Brand, A.H., Roth, S., and Guichet, A., 2002, Polar transport in the *Drosophila* oocyte requires dynein and kinesin I cooperation, *Curr. Biol*. **12**:1971-1981.

Jiang, Y., Xu, X-S., and Russell, J.E., 2006, A nucleolin-binding 3′ untranslated region element stabilizes β-globin mRNA *in vivo*, *Mol. Cell. Biol.* **26**:2419-2429.

Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., Di Padova, F., Lin, S-C., Gram, H., and Han, J., 2005, Involvement of microRNA in AU-rich element-mediated mRNA instability, *Cell* **120**:623-634.

Johnson, A.W., and Kolodner, R.D., 1995, Synthetic lethality of *sep1* (*xrn1*) *ski2* and *sep1* (*xrn1*) *ski3* mutants of *Saccharomyces cerevisiae* is independent of killer virus and suggests a general role for these genes in translation control, *Mol Cell Biol*, **15**:2719-2727.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B., 2006, MicroRNAs and their regulatory roles in plants, *Ann. Rev. Plant Biol*. **57**:19-53.

Joseph, B., Orlian, M., and Furneaux, H., 1998, p21 mRNA contains a conserved element in its 3′-untranslated region that is bound by Elav-like mRNA-stabilizing proteins, *J. Biol. Chem*. **273**:20511-20516.

Kalyna, M., Lopato, S., and Barta, A., 2003, Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development, *Mol. Biol. Cell* **14**:3565-3577.

Kao, T.H., and Tsukamoto, T., 2004, The molecular and genetic bases of S-Rnase-based self-incompatibility, *Plant Cell* **16**:Suppl:S72-S83.

Kedersha, N., and Anderson, P., 2002, Stress granules: sites of mRNA triage that regulate mRNA stability and translatability, *Biochem. Soc. Trans*. **30**:963-969.

Kedersha, N., Chen, S., Gilks, N., Li, W., Miller, I.J., Stahl, J., and Anderson, P., 2002, Evidence that ternary complex (eIF2-GTP-tRNA(i)(Met))-deficient preinitiation complexes are core constituents of mammalian stress granules, *Mol. Biol. Cell* **13**:195-210.

Kedersha, N.L., Gupta, M., Li, W., Miller, I., and Anderson, P., 1999, RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2a to the assembly of mammalian stress granules, *J. Cell Biol*. **147**:1431-1441.

Kedersha, N., Stoecklin, G., Ayodele, M., Yacono, P., Lykke-Andersen, J., Fritzler, M.J., Scheuner, D., Kaufman, R.J., Golan, D.E., and Anderson, P., 2005, Stress granules and processing bodies are dynamically linked sites of mRNP remodeling, *J. Cell Biol*. **169**:871-884.

Keene, J.D., and Tenenbaum, S.A., 2002, Eukaryotic mRNPs may represent post-transcriptional operons, *Mol. Cell* **9**:1161-1167.

Keiler, K.C., Waller, P.R., and Sauer, R.T., 1996, Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA, *Science* **271**:990-993.

Kessler, M.M., Henry, M.F., Shen, E., Zhao, J., Gross, S., Silver, P.A., and Moore, C.L., 1997, Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3′-end formation in yeast, *Genes Dev*. **11**:2545-2556.

Kindler, S., Wang, H., Richter, D., and Tiedge, H., 2005, RNA transport and local control of translation, *Annu. Rev. Cell Dev. Biol*. **21**:223-245.

Klyachko, N.L., 2005, The cytoskeleton and intracellular motility in plants, *Russian J. Plant Physiol*. **52**:786-795.

Korner, C.G., Wormington, M., Muckenthaler, M., Schneider, S., Dehlin, E. and Wahle, E., 1998, The deadenylating nuclease (DAN) is involved in poly(A) tail removal during the meiotic maturation of *Xenopus* oocytes, *EMBO J*. **17**:5427-5437.

Kosik, K.S., and Krichevsky, A.M., 2002, The message and the messenger: delivering RNA in neurons, *Sci. STKE*. **126**:pe16.

Kozak, M., 2005, A second look at cellular mRNA sequences said to function as internal ribosome entry sites, *Nucl. Acids Res*. **33**:6593-6602.

Krichevsky, A.M., and Kosik, K.S., 2001, Neuronal RNA granules: a link between RNA localization and stimulation-dependent translation, *Neuron*, **32**:683-696.

Lai, W.S., Carballo, E., Strum, J.R., Kennington, E.A., Phillips, R.S. and Blackshear, P.J., 1999, Evidence that tristetraprolin binds to AU-rich elements and promotes the deadenylation and destabilization of tumor necrosis factor alpha mRNA, *Mol. Cell. Biol*. **19**:4311-4323.

Lall, S., Piano, F., and Davis, R.E., 2005, *Caenorhabditis elegans* decapping proteins: localization and functional analysis of Dcp1, Dcp2, and DcpS during embryogenesis, *Mol. Biol. Cell* **16**:5880-5890.

Lawrence, C.J., Morris, N.R., Meagher, R.B., and Dawe, R.K., 2001, Dyneins have run their course in plant lineage, *Traffic*, **2**:362-363.

Lazar, G., Schaal, T., Maniatis, T., and Goodman, H.M., 1995, Identification of a plant serine-arginine-rich protein similar to the mammalian splicing factor SF2/ASF, *Proc. Natl. Acad. Sci. USA* **92**:7672-7676.

Leatherman, J.L., and Jongens, T.A., 2003, Transcriptional silencing and translational control: key features of early germline development, *Bioessays*, **25**:326-335.

Lee, T.I., Causton, H.C., Holstege, F.C., Shen, W.C., Hannett, N., Jennings, *E.G*., Winston, F., Green, M.R., Young, R.A., 2000, Redundant roles for the TFIID and SAGA complexes in global transcription, *Nature* **405**:701-704.

Lee, Y.R.J. and Liu, B., 2004, Cytoskeletal motors in *Arabidopsis*. Sixty-one kinesins and seventeen myosins, *Plant Physiol*. **136**:3877-3883.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J., 2001, The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay, *EMBO J*. **20**:4987-4997.

Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J., 2000a, The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions, *EMBO J*. **19**:6860-6869.

Le Hir, H., Moore, M.J., and Maquat, L.E., 2000b, Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions, *Genes Dev*. **14**:1098-1108.

Lei, E.P., Krebber, H., and Silver, P.A., 2001, Messenger RNAs are recruited for nuclear export during transcription, *Genes Dev*. **15**:1771-1782.

Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L.E., 2002, The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling, *EMBO J*, **21**:3536-3545.

Lejeune, F., Li, X., and Maquat, L.E., 2003, Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities, *Mol. Cell* **12**: 675-687.

Lemaire, R., Prasad, J., Kashima, T., Gustafson, J., Manley, J.L., and Lafyatis, R., 2002, Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins, *Genes Dev*. **16**:594-607.

Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bematavichute, Y.V., Chan, S. W-L., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E., 2006, An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*, *Cell* **126**:93-106.

Lieb, B., Carl, M., Hock, R., Gebauer, D., and Scheer, U., 1998, Identification of a novel mRNA-associated protein in oocytes of *Pleurodeles waltl* and *Xenopus laevis, Exp. Cell Res*. **245**:272-281.

Linder, P., and Stutz, F., 2001, mRNA export: traveling with DEAD box proteins, *Curr. Biol*. **11**:R961-R963.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R., 2005, MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies, *Nat. Cell Biol*. **7**:719-723.

Long, R.M., and McNally, M.T., 2003, mRNA decay:X (*XRN1*) marks the spot. *Mol. Cell*. **11**:1126-1128.

Lopato, S., Gattoni, R., Fabini, G., Stevenin, J., and Barta, A., 1999, A novel family of plant splicing factors with Zn knuckle motif: examination of RNA binding and splicing activities, *Plant Mol. Biol*. **39**:761-773.

Lopato, S., Mayeda, A., Krainer, A.R., and Barta, A., 1996, Pre-mRNA splicing in plants: characterization of Ser/Arg splicing factors, *Proc. Natl. Acad. Sci. USA* **93**:3074-3079.

Lorković, Z.J., and Barta, A., 2002, Genome analysis: RNA recognition motif (RRM) and K homology (KH) domain RNA-binding proteins from the flowering plant *Arabidopsis thaliana, Nucl. Acids Res*. **30**:623-635.

Losson, R., and Lacroute, F., 1979, Interference of nonsense mutations with eukaryotic messenger mRNA stability, *Proc. Natl. Acad. Sci. USA* **76**:5134-5137.

Mangus, D.A., Evans, M.C., and Jacobson, A., 2003, Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression, *Genome Biol*. 4:223.1-223.14.

Maniatis, R., and Reed, R., 2002, An extensive network of coupling among gene expression machines, *Nature* **416**:499-506.

Mansfield, S.G., and Briarty, L.G., 1991, Early embryogenesis in *Arabidopsis thaliana*. II. The developing embryo, *Can. J. of Bot*. **69**:461-476.

Maquat, L.E., and Li, X., 2001, Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay, *RNA* **7**:445-456.

Maquat, L.E., 2004, Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol*. **5**:89-99.

Marshall, N.F., Peng, J., Xie, Z., and Price, D.H., 1996, Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase, *J. Biol. Chem*. **271**:27176-27183.

Matera, A.G., 1999, Nuclear bodies: multifaceted subdomains of the interchromatin space, *Trends Cell Biol*. **9**:302-309.

Mehlin, H., Daneholt, B., and Skoglund, U., 1992, Translocation of a specific premessenger ribo-nucleoprotein particle through the nuclear pore studies with electron microscope tomography, *Cell*, 69:605-613.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C., 2004, Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise, *Nat. Genet*. **36**:1073-1078.

Mignone, F., Gissi, C., Liuni, S., and Pesole, G., 2002, Untranslated regions of mRNAs, *Genome Biology* **3**:0004.1-0004.10.

Millard, S.S., Vidal, A., Markus, M., and Koff, A., 2000, A U-rich element in the 5′ untranslated region is necessary for the translation of p27 mRNA, *Mol. Cell. Biol*. **20**:6947-6959.

Misteli, T., 2000, Cell biology of transcription and pre-mRNA splicing: nuclear architecture meets nuclear function, *J. Cell Sci.* **113**:1841-1849.

Mitchell, P., and Tollervey, D., 2003, An NMD pathway in yeast involving accelerated de-adenylation and exosome-mediated 3′–5′ degradation, *Mol. Cell* **11**:1405-1413.

Mitrovich, Q.M., and Anderson, P., 2005, mRNA surveillance of expressed pseudogenes in *C. elegans*, *Curr. Biol.* **15**:963-967.

Moore, M., 2005, From birth to death: The complex lives of eukaryotic mRNAs, *Science*, **309**:1514-1518.

Moreno Diaz de la Espina, S., Sanchez-Pina, M.A., Risueño, M.C., Medina, F.J., and Fernández-Gómez, M.E., 1980, The role of plant coiled bodies in the nuclear RNA metabolism, *Electron Microsc.* **2**:240-241.

Moriarty, P.M., Reddy, C.C., and Maquat, L.E., 1998, Selenium deficiency reduces the abundance of mRNA for Se-dependent glutathione peroxidase 1 by a UGA-dependent mechanism likely to be nonsense codon-mediated decay of cytoplasmic mRNA, *Mol. Cell. Biol.* **18**:2932-2939.

Moteki, S., and Price, D., 2002, Functional coupling of capping and transcription of mRNA, *Mol. Cell* **10**:599-609.

Muhlrad, D., and Parker, R., 1999, Aberrant mRNAs with extended 3′ UTRs are substrates for rapid degradation by mRNA surveillance, *RNA* **5**:1299-1307.

Mukherjee, D., Gao, M., O'Connor, J.P., Raijmakers, R., Pruijn, G., Lutz, C.S., and Wilusz, J., 2002, The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements, *EMBO J.* **21**:165-174.

Nagai, K., Oubridge, C., Ito, N., Avis, J., and Evans, P., 1995, The RNP domain – a sequence-specific RNA-binding domain involved in processing and transport of RNA, *Trends Biochem. Sci.* **20**:235-240.

Navarro, R., and Blackwell, T.K., 2005, Requirement for P granules and meiosis for accumulation of the germline RNA helicase CGH-1, *Genesis* **42**:172-180.

Neugebauer, K.M., 2002, On the importance of being co-transcriptional, *J. Cell Sci.* **115**;3865-3871.

Nott, A., Le Hir, H., and Moore, M.J., 2004, Splicing enhances translation in mammalian cells: an additional function of the exon junction complex, *Genes Dev.* **18**:210-222.

Nover, L., K.-D., Scharf, and Neumann, D., 1983. Formation of cytoplasmic heat shock granules in tomato cell cultures and leaves, *Mol. Cell. Biol.* **3**:1648-1655.

Nover, L., and Scharf, K.D., 1984, Synthesis, modification and structural binding of heat-shock proteins in tomato cell cultures, *Eur. J. Biochem.* **139**:303-313.

Nover, L., Scharf, K.-D., and Neumann, D., 1989, Cytoplasmic heat shock granules are formed from precursor particles and are associated with a specific set of mRNAs, *Mol. Cell. Biol.* **9**:1298-1308.

Okita, T.W., and Choi, S-B., 2002, mRNA localization in plants: targeting to the cell's cortical region and beyond, *Curr. Opin. Plant Biol.* **5**:553-559.

Okita, T.W., and Rogers, J.C., 1996, Compartmentation of proteins in the endomembrane system of plant cells, *Ann. Rev. Plant Physiol. & Plant Mol. Biol.* **47**:327-350.

Oliveira, C.C., and McCarthy, J.E., 1995, The relationship between eukaryotic translation and mRNA stability: A short upstream open reading frame strongly inhibits translational initiation and greatly accelerates mRNA degradation in the yeast *Saccharomyces cerevisiae*, *J. Biol. Chem.* **270**:8936-8943.

Onouchi, H., Nagami, Y., Haraguchi, Y., Nakamoto, M., Nishimura, Y., Sakurai, R., Nagao, N., Kawasake, D., Kadokura, Y., and Naito, S., 2005, Nascent peptide-

mediated translation elongation arrest coupled with mRNA degradation in the CGS1 gene of *Arabidopsis*, *Genes Dev*. **19**:1799-1810.

Orban, T.I., and Izaurralde, E., 2005, Decay of mRNAs targeted by RISC requires XRN1, the Ski complex and the exosome, *RNA* **11**:459-469.

Ortega, J.L., Moguel-Esponda, S., Potenza, C., Conklin, C.F., Quintana, A., and Sengupta-Gopalan, C., 2006, The 3′ untranslated region of a soybean cytosolic glutamine synthetase (GS1) affects transcript stability and protein accumulation in transgenic alfalfa, *Plant J*. **45**:832-846.

Palacios, I.M., 2002, RNA processing: splicing and the cytoplasmic localization of mRNA, *Curr. Biol*. **12**:R50-R52.

Palanivelu, R., Belostotsky, D.A., and Meagher, R.B., 2000a, *Arabidopsis thaliana* poly (A) binding protein 2 (PAB2) functions in yeast translational and mRNA decay processes, *Plant J*. **22**:187-198.

Palanivelu, R., Belostotsky, D.A., Meagher, R.B., 2000b, Conserved expression of *Arabidopsis thaliana* poly (A) binding protein 2 (PAB2) in distinct vegetative and reproductive tissues, *Plant J*. **22**:199-210.

Park, J-W., Faure-Rabasse, S., Robinson, M.A., Desvoyes, B., and Scholthof, H.B., 2004, The multifunctional plant viral suppressor of gene silencing P19 interacts with itself and an RNA binding host protein, *Virology* **323**:49-58.

Pelletier, J., and Sonenberg, N., 1988, Internal initiation translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA, *Nature*, **334**:320-325.

Pendle, A.F., Clark, G.P., Boon, R., Lewandowska, D., Lam, Y.W., Andersen, J., Mann, M., Lamond, A.I., Brown, J.W., and Shaw, P.J., 2005, Proteomic analysis of the *Arabidopsis* nuclelolus suggests novel nucleolar functions, *Mol. Biol. Cell* **16**:260-269.

Phillips, K., Kedersha, N., Shen, L., Blackshear, P.J., and Anderson, P., 2004. Arthritis suppressor genes TIA-1 and TTP dampen the expression of tumor necrosis factor α, cyclooxygenase 2, and inflammatory arthritis, *Proc. Natl. Acad. Sci. USA* **101**:2011-2016.

Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W., 2005, Inhibition of translational initiation by Let-7 microRNA in human cells, *Science* **309**:1573-1576.

Ping, Y-H., and Rana, T.M., 2001, DSIF and NELF interact with RNA polymerase II elongation complex and HIV-1 Tat stimulates P-TEFb-mediated phosphorylation of RNA polymerase II and DSIF during transcription elongation, *J. Biol. Chem*. **276**: 12951-12958.

Pisarev, A.V., Shirokikh, N.E., and Hellen, C.U., 2005, Translation initiation by factor-independent binding of eukaryotic ribosomes to internal ribosomal entry sites, *Comptes Rendus. Biol*. **328**:589-605.

Raghavan, A., Robison, R.L., McNabb, J., Miller, C.R., Williams, D.A., and Bohjanen, P.R., 2001, HuA and tristetraprolin are induced following T cell activation and display distinct but overlapping RNA binding specificities, *J. Biol. Chem*. **276**:47958-47965.

Reddy, A.S.N., 2004, Plant serine/arginine-rich proteins and their role in pre-mRNA splicing, *TRENDS in Plant Sci*. **9**:541-547.

Reddy, A.S.N., and Day, I.S., 2001, Kinesins in the *Arabidopsis* genome: a comparative analysis among eukaryotes, *BMC Genomics*, **2**:2.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvita, H.R., and Ruvkun, G., 2000, The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*, *Nature* **403**:901-906.

Reverdatto, S.V., Dutko, J.A., Chekanova, J.A., Hamilton, D.A., and Belostotsky, D.A., 2006, mRNA deadenylation by PARN is essential for embryogenesis in higher plants, *RNA* **10**:1200-1214.

Ridley, S.P., Sommer, S.S., and Wickner, R.B., 1984, Superkiller mutations in *Saccharomyces cerevisiae* suppress exclusion of M2 double-stranded RNA by L-A-HN and confer cold sensitivity in the presence of M and L-A-HN, *Mol Cell Biol*. **4**:761-770.

Rogers, N.D., Wang, Z., and Kiledjian, M., 2002, Characterization and purification of a mammalian endoribonuclease specific for the globin mRNA, *J. Biol. Chem*. **277**:2597-2604.

Rodriguez, C.M., Freire, M.A., Camilleri, C., and Robaglia, C., 1998, The *Arabidopsis thaliana* cDNAs coding for eIF4E and eIF(iso)4E are not functionally equivalent for yeast complementation and are differentially expressed during plant development, *Plant J*. **13**:465-474.

Rosorius, O., Reichart, B., Kratzer, F., Heger, P., Dabauvalle, M.C., and Hauber, J., 1999, Nuclear pore localization and nucleocytoplasmic transport of eIF5A: evidence for direct interaction with the export receptor CRM1, *J. Cell Sci*. **112**: 2369-2380.

Sachs, A.B., Davis, R.W., and Kornberg, R.D., 1987, A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability, *Mol. Cell Biol*. **7**:3268-3276.

Sanford, J.R., Gray, N.K., Beckmann, K., and Cáceres, J.F., 2004, A novel role for shuttling SR proteins in mRNA translation, *Genes Dev*. **18**:755-768.

Saxena, S., Jonsson, Z.O., and Dutta, A., 2003, Small RNAs with imperfect match to endogenous mRNA repress translationzzz: Implications for off-target activity of small inhibitory RNA in mammalian cells, *J. Biol. Chem*. **278**:44312-44319.

Scharf, K.D., Heider, H., Hohfeld, I., Lyck, R., Schmidt, E., and Nover, L., 1998, The tomato Hsf system: HsfA2 needs interaction with HsfA1 for efficient nuclear import and may be localized in cytoplasmic heat stress granules, *Mol. Cell. Biol*. **18**:2240-2251.

Schisa, J.A., Pitt, J.N., and Priess, J.R., 2001, Analysis of RNA associated with P granules in germ cells of *C. elegans* adults, *Development* **128**:1287-1298.

Schulz, R., and Jensen, W.A., 1968, *Capsella* embryogenesis: the egg, zygote and young embryo, *Amer. J. Bot*. **55**:807-819.

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M., and Weigel, D., 2005, Specific effects of microRNAs on the plant transcriptome, *Dev. Cell* **8**:517-527.

Sen, G.L., and Blau, H.M., 2005, Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies, *Nat. Cell Biol*. **7**:633-636.

Shaw, P.J., Beven, A.F., Leader, D.J., and Brown, J.W.S., 1998, Localization and processing from a polycistronic precursor of novel snoRNAs in maize. *J. Cell Sci*. **111**:2121-2128.

Shaw, P.J., and Brown, J.W., 2004, Plant nuclear bodies, *Curr. Opin. Plant Biol*. **7**:614-620.

Sheth, U., and Parker, R., 2006, Targeting of aberrant mRNAs to cytoplasmic processing bodies, *Cell* **125**:1095-1109.

Shuman, S., 2000, Structure, mechanism, and evolution of the mRNA capping apparatus, *Prog. Nucleic Acid Res. Mol. Biol*. **66**:1-40.

Sigrist, S.J., Thiel, P.R., Reiff, D.F., Lachance, P.E., Lasko, P., and Schuster, C.M., 2000, Postsynaptic translation affects the efficacy and morphology of neuromuscular junctions, *Nature* **405**:1062-1065.

Singer, R.H., 1993, RNA zipcodes for cytoplasmic addresses, *Curr. Biol*. **3:**719-721.

Siomi, H., Matunis, M.J., Michael, W.M., and Dreyfuss, G., 1993, The premessenger RNA-binding K-protein contains a novel evolutionarily conserved motif, *Nucl. Acids Res*. **21**:1193-1198.

Sleeman, J.E., and Lamond, A.I., 1999, Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway, *Curr. Biol*. **9**:1065-1074.

Smart, F.M., Edelman, G.M., and Vanderklish, P.W., 2003, BDNF induces translocation of initiation factor 4E to mRNA granules: evidence for a role of synaptic microfilaments and integrins, *Proc. Natl. Acad. Sci. USA* **100**:14403-14408.

Souter, M., and Lindsey, K., 2000, Polarity and signaling in plant embryogenesis. *J. Exp. Bot*. **51**:971-983.

Stevens, A., 2001, 5′-exoribonuclease 1:XRN1, *Methods Enzymol*. **342**:251-259.

St. Johnston, D., Brown, N.H., Gall, J.G., and Jantsch, M., 1992, A conserved double-stranded RNA-binding domain, *Proc. Natl. Acad. Sci. USA* **89**:10979-10983.

Stoecklin, G., Mayo, T., and Anderson, P., 2006, ARE-mRNA degradation requires the 5′→3′ decay pathway, *EMBO Rep*. **7**:72-77.

Stoecklin, G., Stubbs, T., Kedersha, N., Blackwell, T.K., and Anderson, P., 2004, MK2-induced tristetraprolin:14-3-3 complexes prevent stress granule association and ARE-mRNA decay, *EMBO J*. **23**:1313-1324.

Storozhenko, S., Inze, D., Van Montagu, M., and Kushnir, S., 2001, *Arabidopsis* coactivator ALY-like proteins, DIP1 and DIP2, interact physically with the DNA-binding domain of the Zn-finger poly (ADP-ribose) polymerase, *J. Exp. Bot*. **52**:1375-1380.

Strasser, K., and Hurt, E., 2000, Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export, *EMBO J*. **19**:410-420

Stuger, R., Ranostaj, S., Materna, T., and Forreiter, C., 1999, Messenger RNA-binding properties of nonpolysomal ribonucleoproteins from heat-stressed tomato cells, *Plant Physiol*. **120**:23-32.

Subramanian, A.R., 1983, Structure and functions of ribosomal protein S1, *Prog. Nucl. Acid Res. Mol. Biol*. **28**:101-142.

Sudarsan, N., Barrick, J.E., and Breaker, R.R., 2003, Metabolite-binding RNA domains are present in the genes of eukaryotes, *RNA* **9**:644-647.

Sunohara, R., Jojima, K., Tagami, H., Inada, T., and Aiba, H., 2004, Ribosome stalling during translation elongation induces cleavage of mRNA being translated in *Escherichia coli, J. Biol. Chem*. **279**:15368-15375.

Tekotte, H., and Davis, I., 2002, Intracellular mRNA localization: motors move messages, *Trends Genet*. **18**:636-642.

Thakurta, A.G., Ho Yoon, J., and Dhar, R., 2002, *Schizosaccharomyces pombe* spPABP, a homologue of *Saccharomyces cerevisiae* Pab1p, is a non-essential, shuttling protein that facilitates mRNA export, *Yeast* **19**:803-810.

Thomas, M.G., Tosar, L.J., Loschi, M., Pasquini, J.M., Correale, J., Kindler, S., and Boccaccio, G.L., 2005, Staufen recruitment into stress granules does not affect early mRNA transport in oligodendrocytes, *Mol. Biol. Cell* **16**:405-420.

Thompson, J.E., Hopkins, M.T., Taylor, C., and Wang, T-W., 2004, Regulation of senescence by eukaryotic translation initiation factor 5A: implications for plant growth and development, *TRENDS Plant Sci*. **9**:174-179.

Timchenko, N.A., Welm, A.L., Lu, X., and Timchenko, L.T., 1999, CUG repeat binding protein (CUGBP1) interacts with the 5 region of C/EBP mRNA and regulates translation of C/EBP isoforms, *Nucl. Acids Res*. **27**:4517-4525.

Tourriere, H., Chebli, K., Zekri, L., Courselaud, B., Blanchard, J.M., Bertrand, E., and Tazi, J., 2003, The RasGAP-associated endoribonuclease G3BP assembles stress granules, *J. Cell Biol*. **160**:823-831.

Uhrig, J.F., Canto, T., Marshall, D., and MacFarlane, S.A., 2004, Relocalization of nuclear ALY proteins to the cytoplasm by the tomato bushy stunt virus P19 pathogenicity protein, *Plant Physiol*. **135**:2411-2423.

Van De Bor, V., and Davis, I., 2004, mRNA localization gets more complex, *Curr. Opin. Cell Biol*. **16**:300-307.

van Hoof, A., Frischmeyer, P.A., Dietz, H.C., and Parker, R., 2002, Exosome-mediated recognition and degradation of mRNAs lacking a termination codon, *Science* **295**:2262-2264.

van Hoof, A., and Green, P.J., 1996, Premature nonsense codons decrease the stability of phytohemagglutinin mRNA in a position-dependent manner, *Plant J*. **10**:415-424.

Vinciguerra, P., and Stutz, F., 2004, mRNA export: an assembly line from genes to nuclear pores, *Curr. Opin. Cell Biol*. **16**:285-292.

Voelker, T.A., Moreno, J., and Chrispeels, M.J., 1990, Expression analysis of a pseudogene in transgenic tobacco: a frameshift mutation prevents mRNA accumulation, *Plant Cell* **2**:255-261.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A., 2002, Regulation of heterochromatic silencing and histone He lysine-9 methylation by RNAi, *Science* **297**:1833-1837.

Wang, T-W., Lu, L., Wang, D., and Thompson, J.E., 2001, Isolation and characterization of senescence-induced cDNAs encoding deoxyhypusine synthase and eucaryotic translation initiation factor 5A from tomato, *J. Biol. Chem*. **276**:17541-17549.

Weischenfeldt, J., Lykke-Andersen, J., and Porse, B., 2005, Messenger RNA surveillance: Neutralizing natural nonsense, *Curr. Biol*. **15**:R559-R562.

Wickens, M., Bernstein, D.S., Kimble, J., and Parker, R., 2002, A PUF family portrait: 3′ UTR regulation as a way of life, *Trends Genet*. **18**:150-157.

Wiegand, H.L., Lu, S., and Cullen, B.R., 2003, Exon junction complexes mediate the enhancing effect of splicing on mRNA expression, *Proc. Natl. Acad. Sci. USA* **100**:11327-11332.

Wilczynska, A., Aigueperse, C., Kress, M., Dautry, F., and Weil, D., 2005, The translational regulator CPEB1 provides a link between dcp1 bodies and stress granules, *J. Cell Sci*. **118**:981-992.

Wilkie, G.S., and Davis, I., 2001, *Drosophila wingless* and pair-rule transcripts localize apically by dynein-mediated transport of RNA particles, *Cell* **105**:209-219.

Williams, K.P., 2002, Descent of a split RNA, *Nucl. Acids Res*. **30**:2025-2030.

Wilusz, C.J., and Wilusz, J., 2004, Bringing the role of mRNA decay in the control of gene expression into focus, *TRENDS in Genet*. **20**:491-497.

Winkler, W., Nahvi, A., and Breaker, R.R., 2002, Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression, *Nature* **419**:952-956.

Yamamoto, Y., Sunohara, T., Jojima, K., Inada, T., and Aiba, H., 2003, SsrA-mediated trans-translation plays a role in mRNA quality control by facilitating degradation of truncated mRNAs, *RNA* **9**:408-418.

Yekta, S., Shih, I.H., and Bartel, D.P., 2004, MicroRNA-directed cleavage of HOXB8 mRNA, *Science* **304**:594-596.

Zhang, S., Ruiz-Echevarria, M.J., Quan, Y., and Peltz, S.W., 1995, Identification and characterization of a sequence motif involved in nonsense-mediated mRNA decay, *Mol. Cell. Biol*. **15**:2231-2244.

Zhang, J., Sun, X., Qian, Y., LaDuca, J.P., and Maquat, L.E., 1998, At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation, *Mol. Cell. Biol.* **18**:5272-5283.

Zhou, J.P., Yand, Z.J., Feng, J., Chi, S.H., Liu, C., and Ren, Z.L., 2006, [Cloning and analysis of a gene encoding wheat translation initiation factor, eIF5A], *Yi Chuan* **28**:571-577 [article in Chinese].

Subject Index

Color Plates

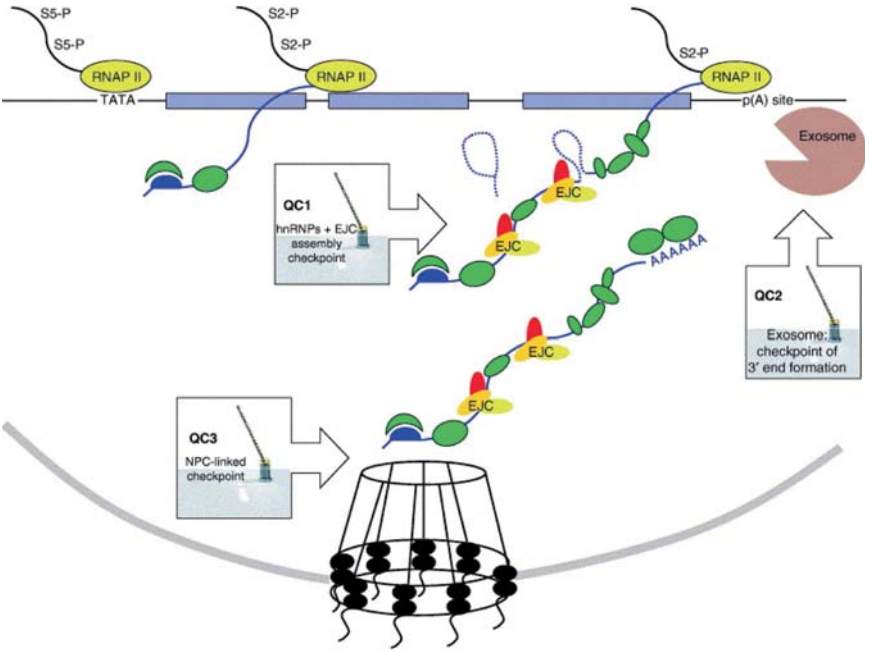


FIGURE 1.1. Key features and quality control checkpoints of the 'mRNA factory'. In the CTD (C-terminal domain) of RNAP II, the changing phosphorylation of repeat residues Serine 5 (S5) and Serine 2 (S2) during transcription regulates the dynamic interactions of the CTD with enzymes of mRNA maturation (capping, splicing and polyadenylation factors, not shown). The 5′ cap is added early in transcription, whereas splicing of introns usually occurs cotranscriptionally but can happen later as well. Splicing leads to a deposition of the exon junction complex (EJC) upstream of the splice junctions. Numerous hnRNPs and other RNA binding proteins (represented as green symbols of different shapes and sizes), including cap binding protein eIF4E and poly(A) binding protein, associate with maturing transcripts; the complement of these factors is likely to be distinct for different mRNAs. The appropriate assembly of hnRNP proteins and completion of 3′ end formation is monitored at quality control checkpoints QC1 and QC2, respectively; messages improperly matured at these stages are subject to retention at the transcription site and degradation by the exosome complex. The nuclear pore complex-linked checkpoint, QC3, ensures that only spliced transcripts are exported from the nucleus and causes mRNA to be kept at the site of transcription when the nuclear-pore complex-linked export step is blocked. Reproduced with permission, from Belostotsky and Rose, 2005, Trends in Plant Science, 10(7), 347–353, © Elsevier Ltd.

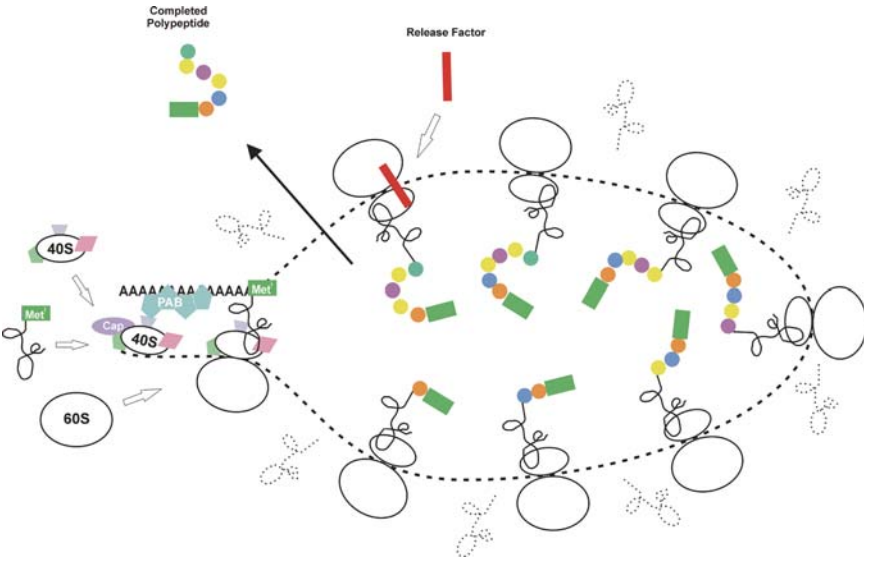


FIGURE 1.2. Summary of Eukaryotic Translation. The small ribosomal subunit (40S) associates with initiation factors (shown as colored shapes) and binds to the mRNA (dotted line) via the cap (red circle). This initiation complex then recruits the 60S ribosomal subunit to begin translation. A bridge formed between eIF4G and the poly (A) binding protein (purple pentagons) is thought to facilitate translation initiation. Various elongation factors modulate the rate of elongation of the nascent polypeptide. Different charged tRNAs (shown with the charged amino acid as a colored or gray rectangle) alternately bind the A (acceptor) and P (peptidyl) sites to position the correct amino acid specified by the triplet codon and to form the covalent peptide linkage. When the stop codon is encountered, release factors dissociate the completed polypeptide from the ribosomes, and the ribosomal subunits and accessory factors are recycled.

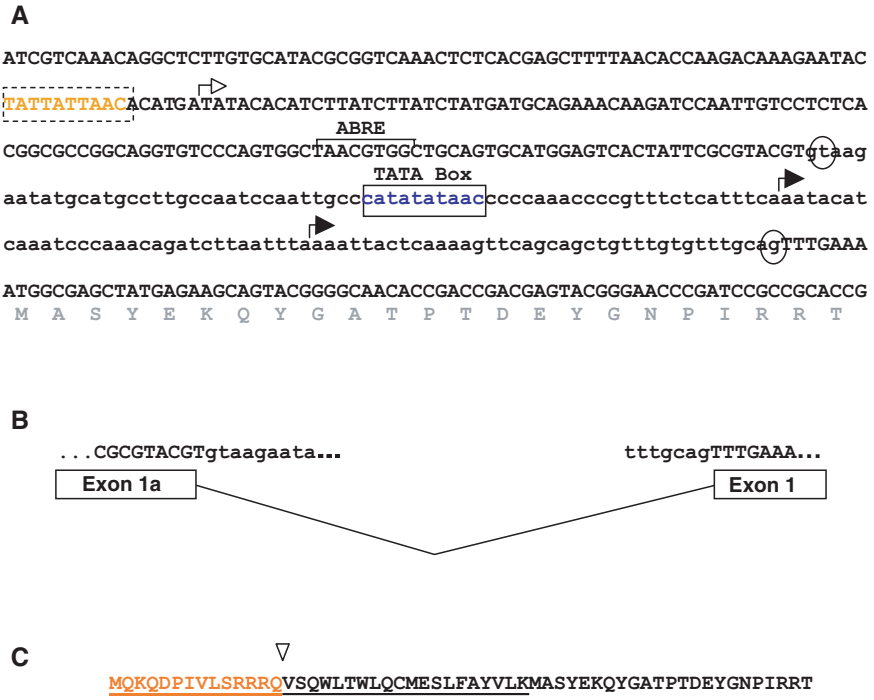


FIGURE 2.2. The promoter region of a peach dehydrin gene is illustrated. A. Sequence of the 490 bases upstream of the canonical translation start site is shown. The 5′ leader intron is shown in lower case lettering, and the TATA element (in blue) that regulates expression in bark is outlined by a solid 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.

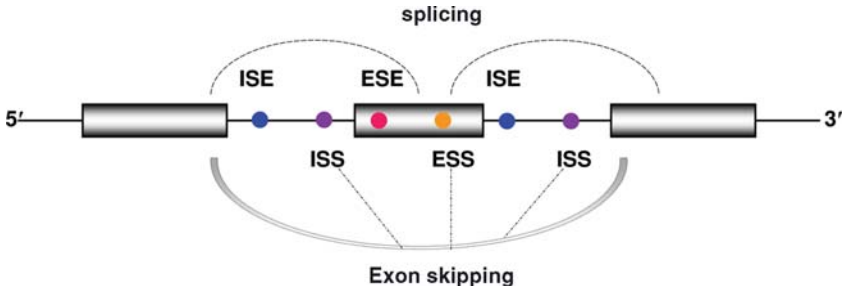


FIGURE 3.1. Enhancer and silencer elements at introns and exons, and splicing consequences. ISE: Intronic splicing enhancer; ESE: Exonic splicing enhancer; ISS: Intronic splicing silencer; ESS: Exonic splicing silencer.

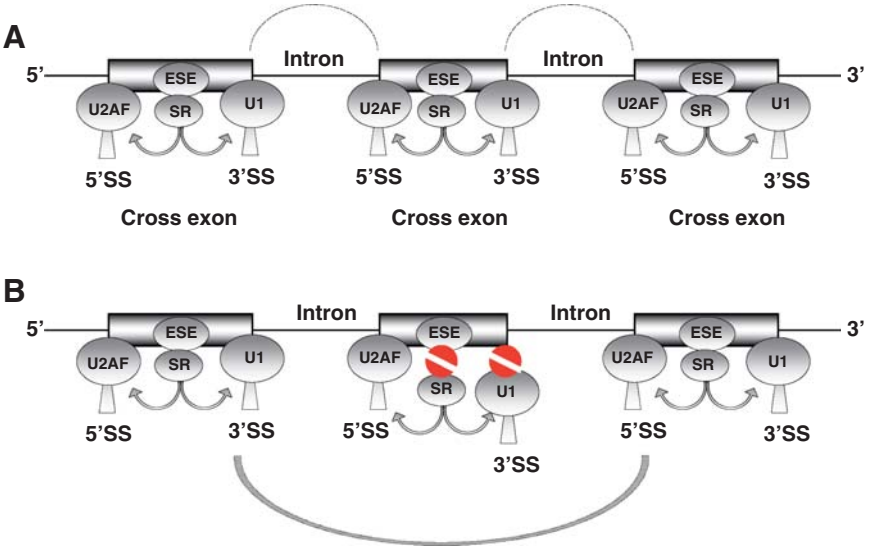


FIGURE 3.3. Mechanism of exon definition during splicing. (A) Exon definition requires the simultaneous recognition of 5′ and 3′ SS across exon. After exon definition, mechanisms at the intron assure that the 3′ end of one exon is joined to the 5′ end of the next exon. (B) Disturbance in the interaction between any components of the splicing machinery (shown in red) can induce exon skipping.

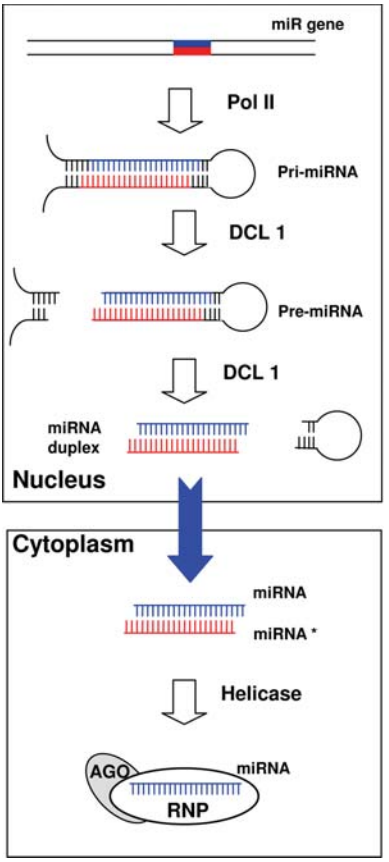


FIGURE 5.1. Plant miRNA biogenesis pathway. miRNA genes (*miR*) are transcribed by RNA polymerase II (Pol II) to generate the primary transcripts (pri-miRNAs). DCL1 catalyzes the processing of pri-miRNAs to form the miRNA duplexes. The duplex is separated, and the mature miRNA is selected and loaded onto the effector complex (Ribonucleic Protein, RNP), whereas the other strand (miRNA*) is degraded.

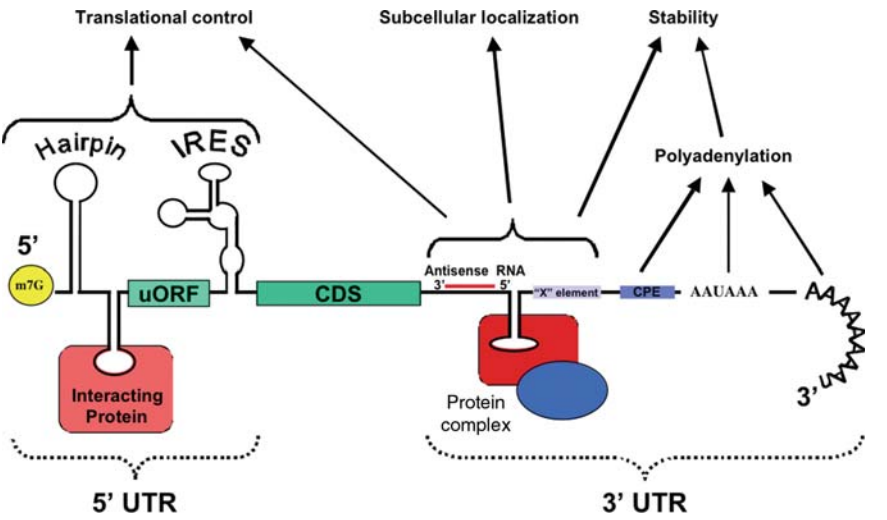


FIGURE 6.1. The generic structure of a eukaryotic mRNA, illustrating some posttranscriptional regulatory elements that affect gene expression. Abbreviations (from 5′ to 3′): UTR, untranslated region; m7G, 7-methyl-guanosine cap; hairpin, hairpin-like secondary structures; uORF, upstream open reading frame; IRES, internal ribosome entry site; CDS, coding sequence; "x" element, zipcode; CPE, cytoplasmic polyadenylation element; AAUAAA, polyadenylation signal. Reproduced with permission, from Mignone *et al*., 2002, *Genome Biology*, 3(3), 0004.1-0004.10. © *Genome Biology*.

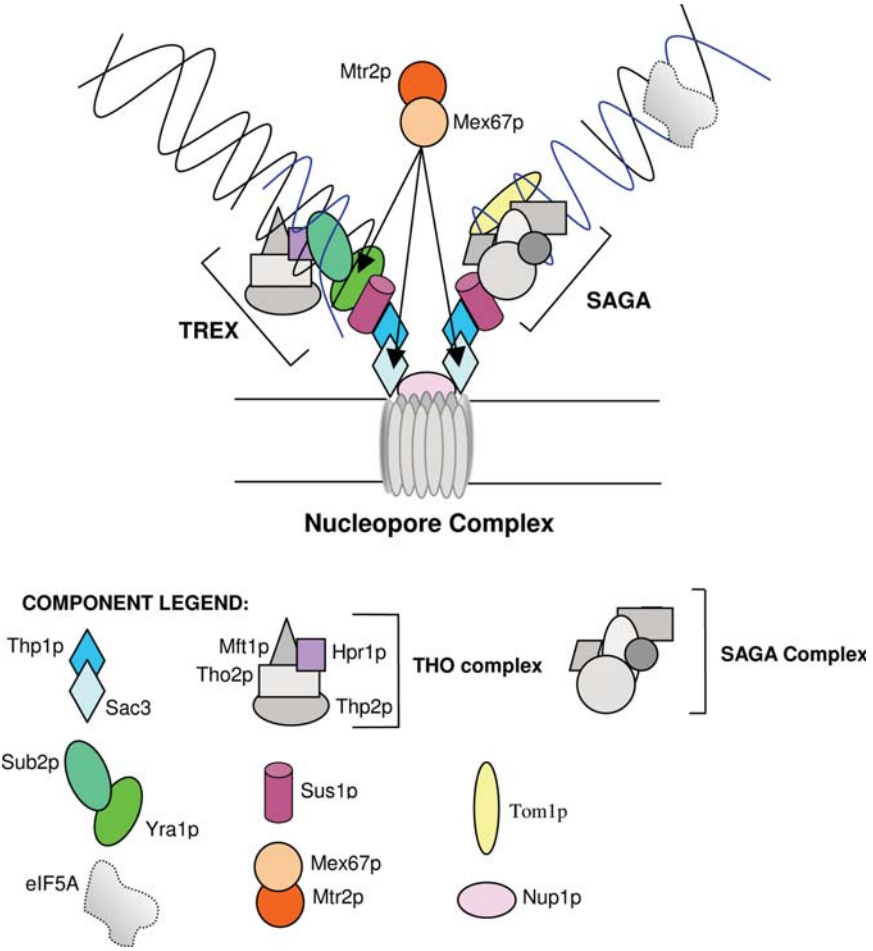


FIGURE 6.2. Core components of the mRNA export machinery. The TREX and SAGA complexes couple export to transcription initiation and the early steps of elongation. It is thought that SAGA directs export of a subset of mRNAs. The TREX complex consists of the THO complex coupled to the Sub2p and Yra1p dimer through interaction between Yra1p and Hpr1p. Tom1p associates with the SAGA complex, but does not co-purify with it. Protein components of the general export pathway are identified under the legend in the figure. DNA strands are shown in black; RNA strands are shown in dark blue. After Vinciguerra and Stutz, 2004.

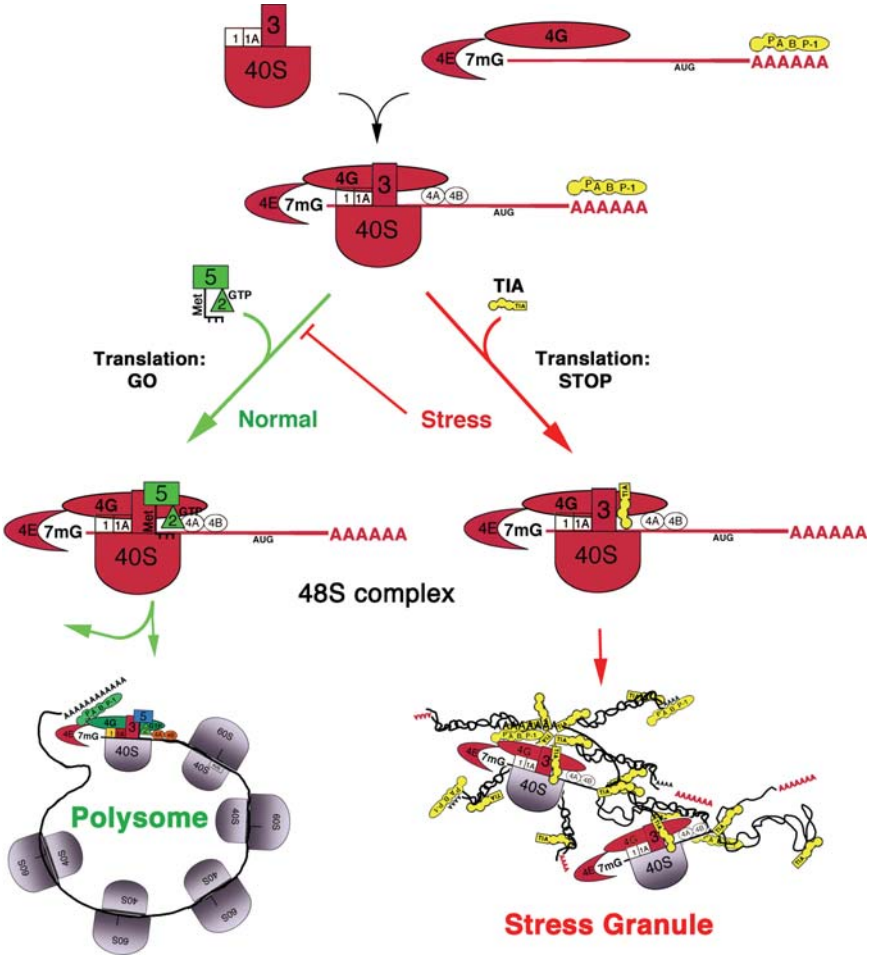


FIGURE 6.3. Translational initiation in the absence or presence of stress. In the absence of stress, the eIF2–GTP–tRNAMet ternary complex (green) is available to form a canonical 48S preinitiation complex at the 5′ end of capped transcripts (green arrow: normal) and scanning begins. Upon recognition of the initiation codon by the anticodon of tRNAMet, eIF5 promotes GTP hydrolysis and early initiation factors are displaced by the 60S ribosomal subunit. As additional ribosomes are added to the transcript, the mRNA is converted into a polysome (bottom left). In stressed cells, phosphorylation of eIF2a depletes the stores of eIF2–GTP–tRNAMet, and the cytoplasmic amount of TIA-1 (yellow) increases. Under these conditions, TIA-1 is included in a non-canonical, eIF2/eIF5-deficient preinitiation complex that is translationally silent. TIA-1 auto-aggregation then promotes the accumulation of these complexes at discrete cytoplasmic foci known as SGs (composed of all components of the 48S pre-initiation complex except eIF2 and eIF5). Translational components that are present in SGs are shown in red; non-translation proteins present in SGs are yellow, those absent from SGs are shown in green. Reproduced with permission, from Kedersha and Anderson, 2002, Biochemical Society Transactions 30:963-969. © The Biochemical Society.