Review TRENDS in Plant Science Vol.10 No.7 July 2005

Plant gene expression in the age of systems biology: integrating transcriptional and post-transcriptional events

Dmitry A. Belostotsky[1] and Alan B. Rose[2]

[1]Department of Biological Sciences, State University of New York at Albany, 1400 Washington Ave, Albany, NY 12222, USA
[2]Molecular and Cellular Biology, University of California Davis, 1 Shields Ave, Davis, CA 95616, USA

The extensive mechanistic and regulatory interconnections between the various events of mRNA biogenesis are now recognized as a fundamental principle of eukaryotic gene expression, yet the specific details of the coupling between the various steps of mRNA biogenesis do differ, and sometimes dramatically, between the different kingdoms. In this review, we emphasize examples where plants must differ in this respect from other eukaryotes, and highlight a recurring trend of recruiting the conserved, versatile functional modules, which have evolved to support the general mRNA biogenesis reactions, for plant-specific functions. We also argue that elucidating the inner workings of the plant 'mRNA factory' is essential for accomplishing the ambitious goal of building the 'virtual plant'.

Exploring the plant 'mRNA factory'

Progress in our understanding of eukaryotic mRNA biogenesis has been largely driven by mechanistic studies performed using *in vitro* systems in yeast and mammals. By contrast, the comparable *in vitro* tools in plants do not yet exist. Instead, most of the available knowledge is derived from three main sources: inferences based on bioinformatic analyses, examination of the behavior of DNA or RNA-based reporter constructs in cells or whole plants, and analysis of mutants. Curiously, many of the mutants that have allowed us a glimpse into the inner workings of the plant 'mRNA factory' have been identified serendipitously. A reluctance to seek such mutants directly might have been partly because of the relative paucity of the appropriate molecular assays and a view that the respective genes are essential for viability – an assumption that might seem reasonable, but nonetheless has not always proven valid.

The paradigm

The synthesis and maturation of mRNA, and the assembly of export-competent mRNP complexes (see Glossary), has been shown in recent years to be carried out by an integrated 'mRNA factory' that comprises RNA polymerase II (RNAP II) and numerous processing and export factors [1–8]. As in a real factory, a complex network of quality control checkpoints must be passed for the product (mRNA) to be successfully exported and enter cytoplasmic transactions (Figure 1).

The most important element that facilitates the interconnection of multiple steps of mRNA biogenesis is the C-terminal domain (CTD) of the largest subunit of RNAP II. Acquisition of the CTD was an important step in the evolution of complex patterns of regulated gene expression because it provided a scaffold for extensive interactions with factors involved in all steps of mRNA production [9]. The resulting close association of

Glossary

AtCPL1–AtCPL4: *Arabidopsis* proteins related to FCP (C-terminal domain phosphatase-like).
CBC: nuclear cap-binding complex, composed of CBP80 and CBP20 subunits.
CF1, CF2: cleavage factors 1 and 2, parts of the mRNA 3′ end processing machinery.
CPSF: cleavage and polyadenylation specificity factor.
CStF: mammalian cleavage stimulation factor.
CTD: C-terminal domain of RNAP II.
eIF4F, eIFiso4F: cytoplasmic cap binding complexes (heterodimers of eIF4E/eIF4G and eIFiso4E/eIFiso4G, respectively).
EJC: exon junction complex, deposited as a consequence of splicing.
FCP: TFIIF-associating component of CTD phosphatase. Dephosphorylates Serine 2 of the CTD.
GT: guanylyltransferase component of the capping enzyme.
GUS: β-glucuronidase.
hnRNP: heterogeneous nuclear ribonucleoprotein.
mRNP: messenger ribonucleoprotein.
MT: N^7G methyltransferase component of the capping enzyme.
NMD: nonsense-mediated decay, a specialized mRNA degradation pathway. Targets mRNAs containing premature stop codons.
PABP: major poly(A) binding protein, predominantly cytoplasmic at steady state but has an important role in nuclear 3 end processing as well.
PABPN: nuclear-specific poly(A) binding protein, regulates PAP.
PAP: poly(A) polymerase.
P-TEFb: transcription elongation factor composed of cyclin T and CDK9 kinase.
RNAP II: RNA polymerase II.
RT: RNA triphosphatase component of the capping enzyme.
SCP: small CTD phosphatase, dephosphorylates Serine 5 of CTD.
snRNA: small nuclear RNAs, key constituents of the spliceosome.
SR proteins: Ser/Arg-rich proteins participating in many aspects of splicing.
TAR: *trans*-activating response region.
TBRP: *trans*-activation response RNA-binding protein.
TFIIH: transcription initiation factor that contains CDK7 kinase and cyclin H.

Corresponding authors: Belostotsky, D.A. (dab@albany.edu); Rose, A.B. (abrose@ucdavis.edu).
Available online 13 June 2005

 1360-1385/$ - see front matter doi:10.1016/j.tplants.2005.05.004

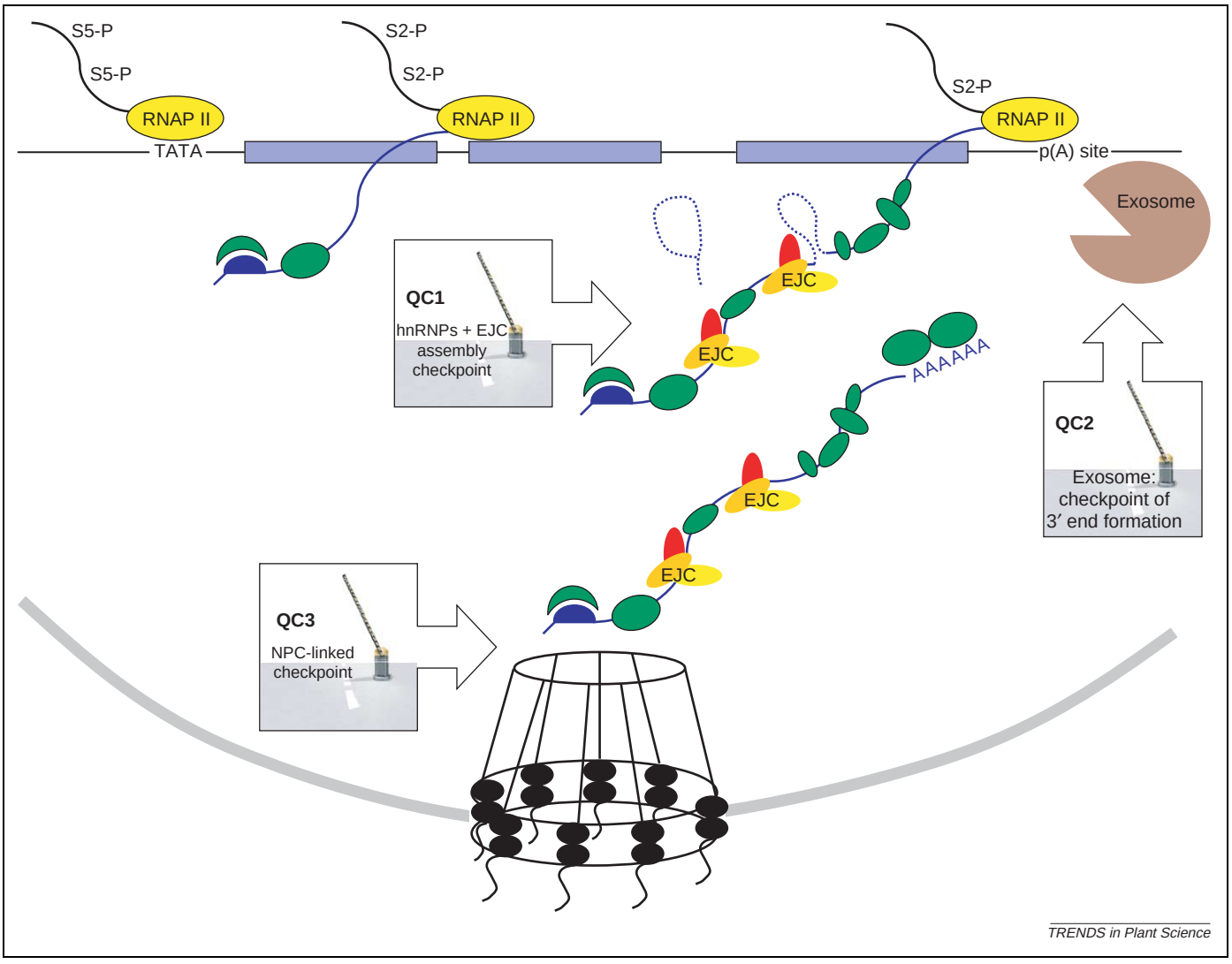

Figure 1. Key features and quality control checkpoints of the 'mRNA factory'. In the CTD of RNAP II, the changing phosphorylation of repeat residues Serine 5 (S5) and Serine 2 (S2) during transcription regulates the dynamic interactions of the CTD with enzymes of mRNA maturation (capping, splicing and 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 [69].

RNA-processing enzymes increases the efficiency of mRNA biogenesis by concentrating the key players near the site of action. The unique structure of the CTD, which comprises numerous tandem heptapeptide repeats with a consensus sequence YSPTSPS [10], provides ample opportunities for regulation via phosphorylation and conformational isomerization by peptidyl-prolyl *cis/trans* isomerases. The rich repertoire of conformational and phosphorylation states allows the CTD to not only modulate the activity of the polymerase, but also to recruit different multiprotein complexes involved in mRNA processing at specific points in the transcription cycle.

The CTD is hypophosphorylated during recruitment into preinitiation complexes. Phosphorylation of the CTD, particularly on Serine 5 by the CDK7 kinase of the initiation factor TFIIH (Figure 2), coincides with transcription initiation and also promotes the binding of mRNA capping factors. Capping occurs when the nascent transcript is 20–30 nt long; at this step, some of the capping enzyme subunits dissociate from the polymerase. The transition into elongation is marked by the removal of the phosphate groups from Serine 5 by the small CTD phosphatase (SCP) while Serine 2 phosphorylation increases as a result of the complex interplay between the transcription elongation factor P-TEFb (composed of cyclin T and CDK9 kinase) and phosphatase FCP (Figure 2). Concomitantly, a subset of splicing and polyadenylation factors adheres to the CTD, either by direct binding or via transfer from the preinitiation complex [11]. The polyadenylation factors that bind to the CTD near the start of a gene are carried along with the transcription machinery to the 3′ end where they and other factors that join the complex at the later stages of the transcription cycle participate in processing the 3′ end of the mRNA. Splicing can occur

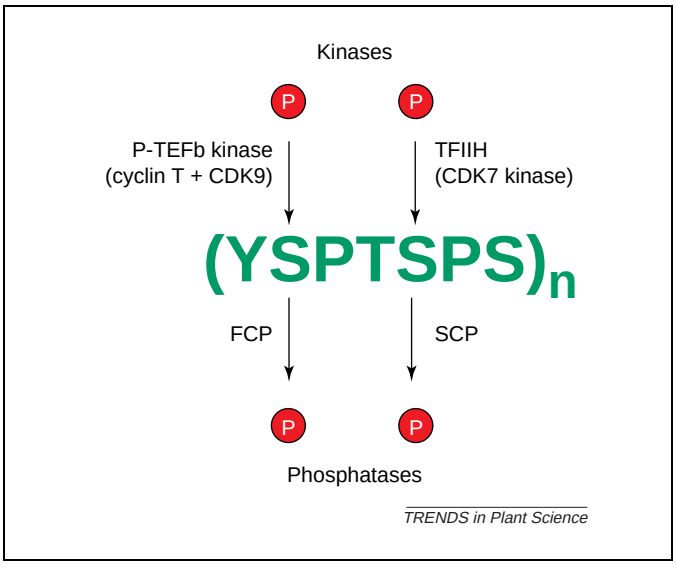

Figure 2. Factors regulating CTD phosphorylation. The consensus sequence of the heptapeptide repeats that constitute the CTD is shown and mammalian enzymes that phosphorylate and dephosphorylate the serine residues at positions 2 and 5 are highlighted. For simplicity, only those factors mentioned in the text are indicated. The complete details of all the enzymes that can modify the CTD and the residues on which they act remain unclear, and the substrate specificities of the enzymes shown are not absolute *in vitro*.

any time after an intron is transcribed until the mRNA is exported, but is usually completed while the nascent RNA is still tethered to the template DNA by the polymerase [12]. As a consequence of splicing, a collection of proteins known as the exon junction complex (EJC) is deposited 20–24 nt upstream of the former site of the intron [13]. The EJC has diverse roles in different species, such as facilitating mRNA export by direct interactions with proteins that associate with the nuclear pore complex, promoting association of the mRNA with polysomes, and helping to identify aberrant mRNAs containing premature stop codons for degradation via nonsense-mediated mRNA decay (NMD) [14].

To what extent does the scenario just described (based primarily on results from yeast and mammals) also occur in plants? The presence in plants of genes encoding presumptive homologs of most of the aforementioned factors that govern the processes of capping, splicing, polyadenylation, export and mRNA surveillance (listed on the accompanying website www.albany.edu/faculty/dab/IPGE_HUB.html) is consistent with the view that it should be fundamentally similar in plants. However, there are some remarkable and puzzling omissions, as discussed below. In addition, the overall similarity of the parts list contrasts with the many known differences of the gene expression reactions between plants and other organisms. Finally, the relatively mild consequences of mutations in several genes that are thought to be essential suggests that either many processes in plants can be carried out by functional homologs that lack sequence similarity or that plants are more different than we realize.

RNA polymerase II and its regulatory factors

The CTD in *Arabidopsis* has 40 repeats, 36 of which match the YSPTSPS consensus, and thus falls in the middle of the range defined by yeast and human CTDs (26 and 52 repeats, respectively). Consequently, the *Arabidopsis* CTD is capable of binding multiple factors, although the network of interactions might not be as extensive as it is in mammals. However, the machinery that governs CTD phosphorylation in plants might be more complex than it is in other kingdoms. For example, *Arabidopsis* differs from both fungi and vertebrates in having at least two CDK7-like kinases capable of phosphorylating Serine 5 [15]. Moreover, each of these can be found in two distinct, large complexes that further differ in their ability to phosphorylate the CTD. On top of this, a distinct kinase (At4g28980) acts upstream to activate both of these CDK7-like kinases. Therefore, a somewhat shorter CTD in *Arabidopsis* could be more than compensated for by the complex interplay of multiple activities that modify it.

In contrast to the TFIIH-associated CDK7-like kinases, the similarity of the presumptive *Arabidopsis* homologs of other kinases and phosphatases that govern the CTD phosphorylation (Figure 2) to their respective mammalian counterparts is low and/or limited in length. Nevertheless, At5g10270 and At5g64960 are likely to represent *bona fide* homologs of CDK9, a subunit of a kinase (P-TEFb) that can phosphorylate Serine 2 of the repeats, because they interact in two hybrid assays with the other P-TEFb subunit, cyclin T [16]. At least four genes (AtCPL1– AtCPL4) contain regions of similarity to the catalytic domain of FCP, and AtCPL1 and AtCPL2 specifically dephosphorylate Serine 5 but not Serine 2 *in vitro* [17]. Remarkably, mutations in two of these (AtCPL1/FRY2 and AtCPL3) result in increased expression of a subset of cold- and abscisic acid (ABA)-regulated genes [18,19].

How can a generic RNAP II regulatory factor such as FCP have such a specific function in plant stress response? In contrast to classical FCPs, the AtCPL proteins contain additional domains suggestive of function. AtCPL1 (as well as AtCPL2) are unique in that they also contain RNA binding domains resembling the human TRBP protein that binds the HIV TAR RNA. TAR RNA, in turn, is the interaction site for the HIV Tat protein that associates with the CTD kinase that hyperphosphorylates RNAP II. Extending the analogy, one could envision that AtCPL1 binds to a subset of messages in a sequence-specific fashion and regulates the mRNA biogenesis at specific loci via modulating the RNAP II phosphorylation state in *cis*. However, AtCPL3 is unique among all known CTD phosphatases in having a region homologous to CES1 (capping enzyme suppressor) in yeast. Thus, AtCPL3 could be directly regulating the 5′ cap status of specific messages.

Capping enzymes and cap binding factors

The mRNA 5′ cap is a focal point of many documented regulatory pathways [20]. Formation of the mature 5′ cap involves three distinct enzymatic activities: RNA triphosphatase (RT), guanylyltransferase (GT) and N7G methyltransferase (MT). The RT and GT activities are a property of a single, bifunctional enzyme in humans but reside in separate polypeptides in yeast. Moreover, the human and fungal enzymes are regulated differently by the RNAP II CTD, and the requirements for the interaction between the RT and GT subunits are also different, even within the

kingdom fungi [21]. Plants, like humans, have bifunctional RT/GT enzymes but, unlike humans, *Arabidopsis* has three distinct RT/GT genes with distinct expression patterns [22], which might indicate their functional specialization.

The cap binding complex (CBC), a heterodimer of CBP20 and CBP80, binds to the cap during transcription and is subsequently replaced by translation initiation factor eIF4E in the cytoplasm following a 'pioneer' round of translation that occurs while the mRNA is still associated with CBC [23]. Unexpectedly, the *Arabidopsis* mutant *abh1*, which is defective in CBP80, was isolated in a screen for ABA hypersensitivity: it shows ABA-hypersensitive regulation of seed germination, stomatal closure and cytosolic calcium increases in guard cells [24,25]. It is also remarkable that even though the *Arabidopsis* CBP80 is encoded by a single-copy, ubiquitously expressed gene, the *abh1* mutation has no major effect on plant growth and development beyond the ABA responses and alters the expression of only a small subset of genes. The relatively minor consequences of eliminating CBP80 are particularly surprising because CBC is thought to participate in multiple steps of mRNA metabolism, including the splicing of cap-proximal introns, 3′ end processing and the export of snRNAs.

Another unique aspect of the machinery that interfaces with the mRNA 5′ cap in plants is the presence of the two highly divergent cytoplasmic cap binding complexes, eIF4F and eIFiso4F (eIF4E/eIF4G and eIFiso4E/eIFiso4G heterodimers, respectively). The cap affinity and abundance of the two isoforms are different, and eIF4F is more efficient in supporting cap-independent translation and translation of structured RNA than is eIFiso4F *in vitro* [26], hinting at possible functional specialization *in vivo*. Indeed, the eIFiso4E subunit of eIFiso4F plays a nonredundant role in the replication cycle of plant potyviruses. Surprisingly, plants lacking eIFiso4E otherwise grow and develop completely normally [27,28].

Splicing

Homologs of most of the protein and RNA factors involved in splicing in other eukaryotic systems can be identified in the *Arabidopsis* genome, suggesting overall conservation of the basic splicing mechanisms [29–32]. The conserved sequences found at the 5′ and 3′ splice sites and branchpoints are also similar in plants, yeast and mammals; their importance is supported by parallel mutational analyses in all three systems. Given these similarities, it might seem puzzling that plant introns are poorly spliced in mammals and *vice versa*.

This incompatibility with the mammalian splicing machinery might be because splicing in plants is more complex. For example, *Arabidopsis* contains almost twice as many Ser/Arg (SR) proteins than the human genome does [33]. SR proteins have multiple roles in mediating constitutive and alternative splicing through numerous protein–protein interactions between themselves and involving the splicing apparatus [34]. Plant SR proteins help define exons and introns, recruit spliceosomal proteins onto pre-mRNA, influence splice site choice by binding to exonic and intronic splicing enhancer and silencer signals, and/or bridge the 5′ and 3′ splice sites. Thus, improper interpretation of the divergent instructions encoded in typical plant and mammalian introns might preclude splicing by the heterologous apparatus.

Additional considerations are that plant introns often lack recognizable polypyrimidine tracts (but are usually U-rich throughout, particularly in dicots), and the way in which introns are distinguished from exons might also be different. In vertebrates, the relatively short exons are usually separated by large introns and they are differentiated primarily by factors that bind to exon sequences (exon definition [35]). By contrast, plant introns are generally short and are identified either mostly by intron-binding factors (intron definition) or through a combination of exon and intron definition [29–32]. Thus, plant introns might have additional features that are not properly recognized in mammalian systems.

Introns are known to significantly stimulate gene expression in plants, mammals, yeast, nematodes and insects, suggesting widely conserved interactions between splicing and other steps of gene expression. A connection between splicing and transcription, particularly involving CTD phosphorylation, has been demonstrated in mammals and yeast [36]. Transcript initiation is stimulated by interactions between U1 snRNA, which anneals to the 5′ splice site in the first step of spliceosome assembly, and a kinase that phosphorylates Serine 5 of the CTD [37]. Furthermore, introns also stimulate RNAP II elongation via interaction of snRNPs with the elongation factor TAT-SF1; TAT-SF1 in turn binds to P-TEFb, which phosphorylates Serine 2 of the CTD [38].

Evidence for a similar connection between splicing and transcription in *Arabidopsis* comes from *in vivo* studies in which the position of a single intron was varied throughout a stably integrated transgene. The intron-mediated stimulation of mRNA accumulation declined with distance from the promoter until it was lost entirely about a kilobase from the start of the gene [39]. Although these findings are most consistent with an effect of introns on transcript elongation, direct evidence for intron-mediated stimulation of transcription in plants has not been reported [40,41].

The *Arabidopsis* protein HUA2 is unique to plants but contains homology to factors that interact with components of the general transcription machinery and has similarity to the domain in a rat SR protein that mediates its interaction with the CTD [42]. Furthermore, HUA2 interacts with a homolog of the yeast Prp40p splicing factor, which is known to bind to the phosphorylated CTD [43]. Thus, HUA2 might also link splicing and transcription. The *hua2* mutation affects the expression of only a few genes, including *AG* [42], *FLC* and *FLM* [44]. This illustrates a recurring trend of adaptation of the functional modules, which originally evolved to serve the general mRNA biogenesis reactions, for various plant-specific processes by combining conserved domains with sequence-specific RNA binding domains and/or protein-interacting domains.

Polyadenylation

Formation of the 3′ end involves cleavage of the nascent transcript and addition of a poly(A) tail whose length is

species-specific and to some extent message-specific [45]. The cleavage event is guided by *cis*-acting sequence determinants that act as binding sites for the two key multisubunit factors (called CPSF and CStF in mammals) that associate with cleavage factors. Poly(A) polymerase (PAP) then adds the tail whose length is limited by PABPN, a nuclear protein that controls the processivity of PAP, and/or by exonucleolytic trimming of the newly synthesized tails in a manner dependent on PABP – a conserved poly(A) binding protein. Importantly, an *Arabidopsis* PABP, PAB3, partially complements this function in yeast [46].

Although the *cis* requirements for polyadenylation in plants are reasonably well understood [47], the reaction mechanism and the composition of complexes that carry it out remain to be elucidated. The *Arabidopsis* genome contains many homologs of the factors found in yeast and mammals (www.albany.edu/faculty/dab/IPGE5.html). Among the few that have been studied, both copies of CPSF73-like genes are essential [48] (Q. Li, personal communication). However, so far, none of these factors have been experimentally proven to act in polyadenylation *in planta*.

In yeast and mammals, the functional coupling between polyadenylation and transcription is extensive and includes (but is not limited to) direct interactions of several polyadenylation factors with phosphorylated CTD [49]. Such interactions, if any, are yet to be demonstrated in plants. However, genetic dissection of flowering time control in *Arabidopsis* has led to an unexpected insight into the plant-specific factors interfacing with the polyadenylation apparatus. The *Arabidopsis* gene *FCA* encodes an RNA-binding protein that promotes flowering by counteracting the floral repressor FLC (FLOWERING LOCUS C) [50]. *FCA* expression is negatively autoregulated through alternative processing of its own pre-mRNA, brought about by the FCA protein promoting cleavage and polyadenylation within *FCA* intron 3 [51]. The resulting truncated transcript encodes an inactive polypeptide. Efficient selection of this promoter-proximal polyadenylation site depends on the interaction of FCA with FY, a homolog of the yeast Pfs2p subunit of the cleavage and polyadenylation factor (CPF). Moreover, the FCA–FY interaction is also required for the downregulation of the floral repressor *FLC* [52]. Thus, it is likely that FCA controls 3′ end formation of a select subset of transcripts in an FY-dependent manner.

The Pfs2p in yeast and the CstF subunit CstF50 in mammals are believed to be functionally equivalent in spite of having dissimilar sequences [53]. *Arabidopsis* has genes encoding homologs of both proteins, *FY* (At5g13480) and At5g60940, which raises the possibility that plants possess two different polyadenylation complexes, perhaps with distinct functions.

mRNA export

At all steps of mRNA biogenesis and export, messages are engaged in dynamic association with many proteins so that export occurs in the form of large messenger ribonucleoprotein (mRNP) complexes. Export is mediated by nuclear export receptors, the most important and conserved of which is TAP (Mex67 in yeast) [54]. Such receptors are recruited to the mRNA via adaptor proteins and help dock the mRNP substrate to specific sites at the nuclear pore complexes. The mRNP is then translocated through the central aqueous channel of the pore, followed by disengagement and remodeling of the mRNP on the cytoplasmic side. Homologs of many of the key players involved in these processes can be recognized in the *Arabidopsis* genome, with the striking exception of Mex67p/TAP. Thus, the factor responsible for the recruitment of the mRNA to the nuclear pore complex in plants remains mysterious.

mRNA export in fungi and metazoa is intimately connected to splicing. On the one hand, splicing acts to promote export through interactions between the EJC proteins and export factors such as TAP/Mex67p [14]. Conversely, unspliced transcripts are actively retained in the nucleus [55]. Whether either mechanism operates in plants remains unclear. On the other hand, there is tentative evidence that mRNA export in plants is linked to transcript elongation because the CTD kinase subunit of the *Arabidopsis* P-TEFb interacts with a RNA binding protein similar to the yeast and mammalian hnRNPs that engage in nucleocytoplasmic shuttling [16].

The yeast Dbp5p protein is a RNA helicase believed to be involved in remodeling or disassembly of the mRNP cargo on the cytoplasmic side of the nuclear pore. Two mutant alleles of the *Arabidopsis* homolog of yeast Dbp5p affect plant responses to cold, but in opposite ways [56]: *los4-2* is chilling resistant and heat sensitive, whereas *los4-1* is chilling sensitive. The *los4-2* allele also shows an ABA hypersensitive response, which led the authors to hypothesize that this specific allele influences the export of transcripts involved in ABA signal transduction [56]. Nevertheless, mRNA export is also globally perturbed in *los4-1* at all temperatures and in *los4-2* at elevated temperatures only.

mRNA export in animals and yeast is also coupled to proper mRNA 3′ end processing. In tobacco, *GUS* mRNA whose 3′ ends are generated by ribozyme-directed cleavage, rather than by normal processing, is preferentially detected in the nucleus [57]. This finding seems to indicate that proper mRNA 3′ end processing is also required for export in plants, although an alternative possibility (i.e. that mRNA is exported from the nucleus but rapidly degraded in the cytoplasm) cannot be ruled out.

Translation

Several of the EJC proteins remain bound to the mRNA in the cytoplasm where they enhance translation by promoting its association with ribosomes. They also facilitate the recognition of premature termination codons because they mark the exon–exon junctions, which typically occur no further than 50 nucleotides downstream of authentic stop codons in mammals. In this context, the peculiarities of the few known examples of NMD in plants are intriguing. First, NMD of the *waxy* mRNA in rice depends on splicing of an intron that is upstream of the premature stop codon [58], whereas in mammals downstream introns trigger NMD. Second, most of the genes shown to be subject to NMD in plants have no introns [59–62], demonstrating that plant NMD might be totally independent from intron

recognition. Perhaps *cis*-acting sequence determinants play an important role in NMD in plants, as they do in yeast [63].

Nevertheless, *Arabidopsis* has homologs of most of the components of the EJC in mammals, including several that are missing in yeast. In plants, as in mammals, the inclusion of an intron in a gene apparently provides a modest (two to fourfold) increase in translation [39,64–66]. The magnitude of the enhancement in *Arabidopsis* was similar even for introns that differed widely in their effect on mRNA accumulation and was relatively insensitive to changes in intron position [39]. These findings are consistent with a role for the EJC in either enhanced export or more efficient ribosome association in plants. Interestingly, the plant homolog of EJC protein Magoh has been identified as a pollen guidance mutant [67], providing yet another example of a seemingly generic mRNA biogenesis factor acting in a specific pathway.

Prospects

Understanding the mechanistic details of the fundamental reactions of gene expression as well as their interconnections will be essential for making the concept of the 'virtual plant' [68] a reality. The most striking feature of the evidence gained to date is the contrast between the plausible expectation, based on sequence homology, that gene expression operates by conserved mechanisms in all eukaryotes, and the frequent failure of plant mutants to display the phenotypes predicted by analogy with mammals and yeast. The mild consequences of mutating a factor thought to be essential, or lethality caused by disrupting a gene that is dispensable in other organisms, are instructive in demonstrating that bioinformatic analysis must be coupled with direct 'invasive' experimentation to generate an accurate understanding of gene expression in plants.

Even though conventional *in vitro* assays for many mRNA processing reactions might never be an option, perhaps because essential but delicate interconnections between steps are much more difficult to preserve during preparation of extracts from plant cells than from mammalian or yeast cells, several alternatives are likely to continue providing useful information. Foremost among these is the analysis of complete and partial loss-of-function mutants in genes whose sequence suggests a role in gene expression. As indicated by the precedents cited above, such analyses can yield a few surprises. Cross-species complementation, both *in vivo* and *in vitro*, will be another useful tool to determine the extent that processes are conserved. Finally, chromatin immunoprecipitation has the potential to detect changes in the transcription machinery (and associated factors) as it traverses a gene. Establishing the biochemical details of the various steps of gene expression in plants, and demonstrating the functional interactions between them, is a major challenge that must be met to accomplish the goal of the 2010 Project – to determine the function of all the genes in *Arabidopsis*.

Acknowledgements

Work in our laboratories is supported by grants from USDA (NRICGP grants 2003–35301–13218 to A.B.R. and 2003–35304–13210 to D.A.B.) and NSF (grant MCB-0424651 to D.A.B.).

References

1 Calvo, O. and Manley, J.L. (2003) Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev.* 17, 1321–1327
2 Bentley, D. (2002) The mRNA assembly line: transcription and processing machines in the same factory. *Curr. Opin. Cell Biol.* 14, 336–342
3 Hirose, Y. and Manley, J.L. (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev.* 14, 1415–1429
4 Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature* 416, 499–506
5 Proudfoot, N.J. *et al*. (2002) Integrating mRNA processing with transcription. *Cell* 108, 501–512
6 Reed, R. and Hurt, E. (2002) A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* 108, 523–531
7 Orphanides, G. and Reinberg, D. (2002) A unified theory of gene expression. *Cell* 108, 439–451
8 Jensen, T.H. *et al*. (2003) Early formation of mRNP. License for export or quality control? *Mol. Cell* 11, 1129–1138
9 Stiller, J.W. and Hall, B.D. (2002) Evolution of the RNA polymerase II C-terminal domain. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6091–6096
10 Shilatifard, A. (2004) Transcriptional elongation control by RNA polymerase II: a new frontier. *Biochim. Biophys. Acta* 1677, 79–86
11 Dantonel, J.C. *et al*. (1997) Transcription factor TFIID recruits factor CPSF for formation of 3′ end of mRNA. *Nature* 389, 399–402
12 Neugebauer, K.M. (2002) On the importance of being co-transcriptional. *J. Cell Sci.* 115, 3865–3871
13 Le Hir, H. *et al*. (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
14 Tange, T.O. *et al*. (2004) The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.* 16, 279–284
15 Shimotohno, A. *et al*. (2004) The plant-specific kinase CDKF;1 is involved in activating phosphorylation of cyclin-dependent kinase-activating kinases in *Arabidopsis*. *Plant Cell* 16, 2954–2966
16 Barroco, R.M. *et al*. (2003) Novel complexes of cyclin-dependent kinases and a cyclin-like protein from *Arabidopsis thaliana* with a function unrelated to cell division. *Cell. Mol. Life Sci.* 60, 401–412
17 Koiwa, H. *et al*. (2004) *Arabidopsis* C-terminal domain phosphatase-like 1 and 2 are essential Ser-5-specific C-terminal domain phosphatases. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14539–14544
18 Xiong, L. *et al*. (2002) Repression of stress-responsive genes by FIERY2, a novel transcriptional regulator in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10899–10904
19 Koiwa, H. *et al*. (2002) C-terminal domain phosphatase-like family members (AtCPLs) differentially regulate *Arabidopsis thaliana* abiotic stress signaling, growth, and development. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10893–10898
20 Richter, J.D. and Sonenberg, N. (2005) Regulation of cap-dependent translation by eIF4E inhibitory proteins. *Nature* 433, 477–480
21 Takagi, T. *et al*. (2002) Divergent subunit interactions among fungal mRNA 5′-capping machineries. *Eukaryot. Cell* 1, 448–457
22 Yamada, K. *et al*. (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* 302, 842–846
23 Ishigaki, Y. *et al*. (2001) Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* 106, 607–617
24 Hugouvieux, V. *et al*. (2002) Localization, ion channel regulation, and genetic interactions during abscisic acid signaling of the nuclear mRNA cap-binding protein, ABH1. *Plant Physiol.* 130, 1276–1287
25 Hugouvieux, V. *et al*. (2001) An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in *Arabidopsis*. *Cell* 106, 477–487
26 Gallie, D.R. and Browning, K.S. (2001) eIF4G functionally differs from eIFiso4G in promoting internal initiation, cap-independent translation, and translation of structured mRNAs. *J. Biol. Chem.* 276, 36951–36960
27 Lellis, A.D. *et al*. (2002) Loss-of-susceptibility mutants of *Arabidopsis thaliana* reveal an essential role for eIF(iso)4E during potyvirus infection. *Curr. Biol.* 12, 1046–1051
28 Duprat, A. *et al*. (2002) The *Arabidopsis* eukaryotic initiation factor (iso)4E is dispensable for plant growth but required for susceptibility to potyviruses. *Plant J.* 32, 927–934

29 McCullough, A.J. and Schuler, M.A. (1997) Intronic and exonic sequences modulate 5′ splice site selection in plant nuclei. *Nucleic Acids Res.* 25, 1071–1077
30 Lorkovic, Z.J. *et al*. (2000) Pre-mRNA splicing in higher plants. *Trends Plant Sci.* 5, 160–167
31 Reddy, A.S.N. (2001) Pre-mRNA splicing in plants. *CRC Cri Rev Plant Sci* 20, 523–571
32 Brown, J.W. and Simpson, C.G. (1998) Splice site selection in plant pre-mRNA splicing. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 77–95
33 Reddy, A.S. (2004) Plant serine/arginine-rich proteins and their role in pre-mRNA splicing. *Trends Plant Sci.* 9, 541–547
34 Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA* 6, 1197–1211
35 Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270, 2411–2414
36 Furger, A. *et al*. (2002) Promoter proximal splice sites enhance transcription. *Genes Dev.* 16, 2792–2799
37 Kwek, K.Y. *et al*. (2002) U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nat. Struct. Biol.* 9, 800–805
38 Fong, Y.W. and Zhou, Q. (2001) Stimulatory effect of splicing factors on transcriptional elongation. *Nature* 414, 929–933
39 Rose, A.B. (2004) The effect of intron location on intron-mediated enhancement of gene expression in *Arabidopsis*. *Plant J.* 40, 744–751
40 Rose, A.B. and Beliakoff, J.A. (2000) Intron-mediated enhancement of gene expression independent of unique intron sequences and splicing. *Plant Physiol.* 122, 535–542
41 Rose, A.B. (2002) Requirements for intron-mediated enhancement of gene expression in *Arabidopsis*. *RNA* 8, 1444–1453
42 Chen, X. and Meyerowitz, E.M. (1999) HUA1 and HUA2 are two members of the floral homeotic AGAMOUS pathway. *Mol. Cell* 3, 349–360
43 Morris, D.P. and Greenleaf, A.L. (2000) The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* 275, 39935–39943
44 Doyle, M.R. *et al*. (2005) HUA2 is required for the expression of floral repressors in *Arabidopsis thaliana*. *Plant J.* 41, 376–385
45 Proudfoot, N. and O'Sullivan, J. (2002) Polyadenylation: a tail of two complexes. *Curr. Biol.* 12, R855–R857
46 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
47 Li, Q.Q. and Hunt, A.G. (1997) The polyadenylation of RNA in plants. *Plant Physiol.* 115, 321–326
48 Xu, R. *et al*. (2004) AtCPSF73-II gene encoding an *Arabidopsis* homolog of CPSF 73 kDa subunit is critical for early embryo development. *Gene* 324, 35–45
49 Proudfoot, N. (2004) New perspectives on connecting messenger RNA 3′ end formation to transcription. *Curr. Opin. Cell Biol.* 16, 272–278
50 Macknight, R. *et al*. (1997) FCA, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. *Cell* 89, 737–745
51 Quesada, V. *et al*. (2003) Autoregulation of FCA pre-mRNA processing controls *Arabidopsis* flowering time. *EMBO J.* 22, 3142–3152
52 Simpson, G.G. *et al*. (2003) FY is an RNA 3′ end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. *Cell* 113, 777–787
53 Ohnacker, M. *et al*. (2000) The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3′-end-processing complex. *EMBO J.* 19, 37–47
54 Stutz, F. and Izaurralde, E. (2003) The interplay of nuclear mRNP assembly, mRNA surveillance and export. *Trends Cell Biol.* 13, 319–327
55 Legrain, P. and Rosbash, M. (1989) Some *cis*- and *trans*-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell* 57, 573–583
56 Gong, Z. *et al*. (2005) A DEAD box RNA helicase is essential for mRNA export and important for development and stress responses in *Arabidopsis*. *Plant Cell* 17, 256–267
57 Buhr, T. *et al*. (2002) Ribozyme termination of RNA transcripts down-regulate seed fatty acid genes in transgenic soybean. *Plant J.* 30, 155–163
58 Isshiki, M. *et al*. (2001) Nonsense-mediated decay of mutant waxy mRNA in rice. *Plant Physiol.* 125, 1388–1395
59 Petracek, M.E. *et al*. (2000) Premature termination codons destabilize ferredoxin-1 mRNA when ferredoxin-1 is translated. *Plant J.* 21, 563–569
60 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
61 Voelker, T.A. *et al*. (1990) Expression analysis of a pseudogene in transgenic tobacco: a frameshift mutation prevents mRNA accumulation. *Plant Cell* 2, 255–261
62 Jofuku, K.D. *et al*. (1989) A frameshift mutation prevents Kunitz trypsin inhibitor mRNA accumulation in soybean embryos. *Plant Cell* 1, 427–435
63 Maquat, L.E. (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.* 5, 89–99
64 Mascarenhas, D. *et al*. (1990) Intron-mediated enhancement of heterologous gene expression in maize. *Plant Mol. Biol.* 15, 913–920
65 Nott, A. *et al*. (2003) A quantitative analysis of intron effects on mammalian gene expression. *RNA* 9, 607–617
66 Nott, A. *et al*. (2004) Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes Dev.* 18, 210–222
67 Johnson, M.A. *et al*. (2004) *Arabidopsis hapless* mutations define essential gametophytic functions. *Genetics* 168, 971–982
68 Chory, J. *et al*. (2000) National Science Foundation-Sponsored Workshop Report: "The 2010 Project" functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. *Plant Physiol.* 123, 423–426
69 Jensen, T.H. *et al*. (2001) A block to mRNA nuclear export in *S. cerevisiae* leads to hyperadenylation of transcripts that accumulate at the site of transcription. *Mol. Cell* 7, 887–898