Pre-mRNA splicing: life at the centre of the central dogma

Jeremy R. Sanford and Javier F. Caceres

MRC Human Genetics Unit, Edinburgh, EH4 2XU, UK (e-mail: jeremy.sanford@hgu.mrc.ac.uk; javier.caceres@hgu.mrc.ac.uk)

Journal of Cell Science 117, 6261-6263 Published by The Company of Biologists 2004 doi:10.1242/jcs.01513

Numerous studies have demonstrated extensive coupling among different steps in eukaryotic gene expression, the best example being the intimate connection between transcription and RNA processing. More recently, new connections between nuclear and cytoplasmic steps in post-transcriptional gene expression have been established. The central illustration depicts the central dogma of eukaryotic gene expression and highlights the basic steps composing post-transcriptional regulation.

In the nucleus, the C-terminal domain (CTD) of the RNA polymerase II (Pol II) large subunit coordinates many RNA processing events by providing a platform for factors involved in different steps of RNA processing (reviewed by MMB and Reed, 2002). Splicing of pre-mRNAs can occur co-transcriptionally (for a review, see Neugebauer, 2002; Beyer and Osheim, 1998), and splicing factors that are enriched in interchromatin granule clusters (IGCs) are recruited to the sites of active transcription (Misteli et al., 1997; Lamond and Spector, 2003).

Components of the capping and polyadenylation machinery also associate with the CTD and these interactions facilitate 5’ end cap formation and polyadenylation in vivo (Maniatis and Reed, 2002). Here, we focus on the central role of pre-mRNA splicing in coordinating many different steps of the gene expression cascade.

The spliceosome

Eukaryotic genes are interrupted by intervening sequences known as introns, which are removed by a macromolecular machine, termed the spliceosome, to form the mature mRNA. The major spliceosome consists of five small nuclear ribonucleoproteins particles (snRNPs) U1, U2, U4, U5 and U6, and a large number of protein splicing factors (reviewed by Kramer, 1996).
Spliceosome assembly follows a carefully orchestrated stepwise pathway (reviewed by Will and Luhrmann, 2001). This is initiated by recognition of the 5′ and 3′ splice sites by the U1 snRNP and the heterodimeric U2 snRNP auxiliary factor (U2AF), respectively, generating the E (early) complex. The pre-spliceosomal A complex is generated by the recruitment of the U2 snRNP to the branch point adenosine (BP), in an ATP-dependent manner. Subsequently the U4-U6-U5 tri-snRNP joins the pre-spliceosome to form the B complex, which is resolved to the catalytic C complex following a series of RNA-RNA and RNA-protein rearrangements at the heart of the spliceosome, resulting in the release of U1 and U4. Although the catalytic centre of the spliceosome has not been fully defined, current evidence strongly suggests that pre-mRNA splicing is accomplished by protein-assisted RNA catalysis (reviewed by Nilsen, 2000).

Many components of this large and complex macromolecular machine have now been identified by proteomic approaches. Indeed, recent studies have led to the identification of over 300 putative spliceosomal protein components (reviewed by Jurica and Moore, 2003); however, many unexpected proteins present, with no apparent direct connection to splicing, still await functional characterization. Some of the additional proteins identified in these complexes have known associations with other aspects of RNA processing, including transcription and mRNA export. This is consistent with the intimate coupling of different steps in gene expression.

**Alternative splicing**

Alternative splicing enables a single gene to increase its coding capacity, allowing the synthesis of structurally and functionally distinct protein isoforms. Usually, alternative exons have suboptimal splicing signals, and their inclusion is modulated by trans-acting factors that recognize an arrangement of positive (splicing enhancers) and/or negative (splicing silencers) cis-acting sequence elements, which can be either exonic or intronic (reviewed by Smith and Valcarcel, 2000; Caceres and Kornблиhtt, 2002).

The highly conserved serine- and arginine-rich (SR) protein family plays an important role in this process (Graveley, 2000). The SR proteins have a modular structure consisting of one or two copies of an RNA-recognition motif (RRM) that determines their RNA-binding specificity, followed by a C-terminal domain rich in alternating serine and arginine residues (the RS domain). SR proteins bound to exonic splicing enhancers (ESEs) promote recruitment of U2AF to the polypyrimidine tract (poly Y) and activate an adjacent 3′ splice site. Alternatively, they might facilitate splicing by recruitment of coactivators or antagonize the negative activity of hnRNP proteins recognizing exonic splicing silencer (ESS) elements (reviewed by Blencowe, 2000; Hastings and Krainer, 2001). Differences in the activities or amounts of general splicing factors and/or gene-specific splicing regulators during development or in different tissues are thought to cause differential patterns of splicing. Moreover, signal transduction pathways can modulate alternative splice site selection in vivo by regulating the concentration, activity and/or subcellular localization of splicing regulatory proteins (reviewed by Stamm, 2002).

Another mechanism of regulation of alternative splicing is imparted by the transcriptional machinery and involves the processivity of Pol II (Roberts et al., 1998; de la Mata et al., 2003). For instance, a slow Pol II, and/or the presence of internal transcriptional pause sites, results in inclusion of the alternative exon harbouring a weak 3′ splice site. By contrast, when the same pre-mRNA is transcribed by a highly processive Pol II, the weak alternative 3′ splice site is unable to compete with the stronger downstream 3′ splice site, which results in skipping of the alternative exon.

**Pre-mRNA splicing and human disease**

Aberrant pre-mRNA splicing can cause human disease. Approximately 15% of mutations that cause genetic disease affect pre-mRNA splicing (Krawczak et al., 1992), often targeting enhancer and silencer sequences, as well as conserved splicing signals, including the 5′ and 3′ splice sites and the branch site. Recent studies have emphasized the importance of what were previously characterized as silent single nucleotide polymorphisms located in coding regions (cSNPs). In fact, cSNPs can directly affect different elements that influence exon recognition, for instance, by disrupting or creating exonic splicing enhancers and/or silencers (reviewed by Pagani and Baralle, 2004). A number of disease mutations and cSNPs that disrupt at least one of the binding sites for the SR proteins SF2/ASF, SRp40, SRp55 and SC35 have been found in genes such as the breast cancer-associated gene BRCA1 (Liu et al., 2001) and in SMN, the gene implicated in spinal muscular atrophy (SMA) (Cartegni and Krainer, 2002).

**Life after splicing**

Many of the proteins that are deposited onto the pre-mRNA remain bound after pre-mRNA splicing is completed and can influence downstream events. This link is provided by a multiprotein complex, the exon junction complex (EJC), that is deposited in a temporal fashion on the mRNA as a consequence of the splicing reaction in a sequence-independent, position-dependent manner. The EJC serves as a binding platform for factors involved in mRNA export and nonsense-mediated decay (NMD) (Le Hir et al., 2001; Lykke-Andersen et al., 2001). The components of the EJC include the splicing-associated factors SRm160 and RNP51, Y14 and its binding partner magoh, elF4AIIa (a DEAD-box RNA helicase that is a close relative of the translation factors elF4A1 and 2), and Upf3 (reviewed by Dreyfuss et al., 2002; Tange et al., 2004). The coupling between splicing and mRNA export is mediated by REF/Aly, which is recruited to the EJC by the conserved DEAD-box helicase UAP56 and promotes recruitment of the bona fide cellular mRNA export factor, the heterodimer TAP-p15, to cellular mRNP (Zhou et al., 2000; Le Hir et al., 2001). The observation that the REF proteins are dispensable for bulk mRNA export in *Drosophila* and in *C. elegans* suggested that additional adaptor proteins can recruit TAP-p15 to cellular mRNP (Gatfield and Izaurralde, 2002; Longman et al., 2003; MacMorris et al., 2003). Interestingly, shuttling SR proteins have
been shown to act as mRNA export adaptors in mammalian cells (Huang et al., 2003). Nonsense mutations that introduce premature termination codons (PTCs) target the mRNA for degradation by NMD (for a review, see Maquat, 2004). The EJC functions as a molecular signal to trigger NMD, since the presence of a PTC causes the ribosome to terminate prematurely, leaving downstream EJC marks that are not removed from the mRNA, which in turn recruit the NMD machinery and trigger mRNA degradation.

Cytoplasmic activities

Substantial evidence shows a clear effect of introns on gene expression. The presence of an intron not only influences 3′ end processing but also significantly enhances the translation of cytoplasmic mRNAs (Lu and Cullen, 2003; Nott et al., 2003). This stimulatory effect of introns on translation may, in part, be attributable to the EJC (Wiegand et al., 2003; Nott et al., 2004). Interestingly, a subset of SR proteins that shuttle continuously from the nucleus to cytoplasm are associated with translating ribosomes and stimulate translation both in vivo and in vitro. Thus, shuttling SR proteins play multiple roles in the post-transcriptional expression of eukaryotic genes by coupling splicing patterns in the nucleus with translational control in the cytoplasm (Sanford et al., 2004). Finally, physiological evidence for coupling of nuclear and cytoplasmic steps in post-transcriptional gene expression has recently been observed during Drosophila development. Here, splicing of the Oskar pre-mRNA is required for both the translation and proper localization of this mRNA at the posterior pole of the Drosophila oocyte (Hachet and Ephrussi, 2004). These new studies thus provide a compelling link between nuclear and cytoplasmic steps in post-transcriptional gene expression and suggest a central role for pre-mRNA splicing in coordinating the recruitment of mRNA-binding proteins that regulate downstream events.

We thank Heidi Sutherland and Ian Smyth (MRC HGU) for comments on the manuscript. We acknowledge support from the Caledonian Research Foundation (J.R.S.) and the MRC (J.F.C.). We apologise to colleagues whose work was not cited directly owing to space limitations.

References

Beyer, A. L. and Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2, 754-765.

Blencowe, B. J. (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* 25, 106-110.

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

Cartegni, L. and Krainer, A. R. (2002). Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMD2 causes spinal muscular atrophy in the absence of SMD1. *Nat. Genet.* 30, 377-384.

de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D. and Kornblitt, A. R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell.* 12, 525-532.

Dreyfuss, G., Kim, V. N. and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell Biol.* 3, 105-209.

Gatfield, D. and Izaurralde, E. (2002). REFI/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export. *J. Cell Biol.* 159, 579-588.

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

Hachet, O. and Ephrussi, A. (2004). Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. *Nature* 428, 959-963.

Hastings, M. L. and Krainer, A. R. (2001). PremRNA splicing in the new millennium. *Curr. Opin. Cell Biol.* 13, 302-309.

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

Jurica, M. S. and Moore, M. J. (2003). Pre-mRNA splicing. *Asash in a sea of proteins. Mol. Cell* 12, 5-14.

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

Krawczak, M., Reiss, J. and Cooper, D. N. (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.* 90, 41-54.

Lamond, A. I. and Spector, D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* 4, 605-612.

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.

Liu, H. X., Cartegni, L., Zhang, M. Q. and Krainer, A. R. (2001). A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat. Genet.* 27, 55-58.

Longman, D., Johnston, I. L. and Caceres, J. F. (2003). The REFI/Aly proteins are dispensable for mRNA export and development in Caenorhabditis elegans. *RNA* 9, 881-891.

Lu, S. and Cullen, B. R. (2003). Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *RNA* 9, 618-630.

Lykke-Andersen, J., Shu, M. D. and Steitz, J. A. (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNP51. *Science* 293, 1836-1839.

MacMorris, M., Brocker, C. and Blumenthal, T. (2003). UAP56 levels affect viability and mRNA export in Caenorhabditis elegans. *RNA* 9, 847-857.

Maniatis, T. and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature* 416, 499-506.

Maquat, L. E. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.* 5, 89-99.

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.

Neugebauer, K. M. (2002). On the importance of being co-transcriptional. *J. Cell Sci.* 115, 3865-3871.

Nilsen, T. W. (2000). The case for an RNA enzyme. *Nature* 408, 782-783.

Nott, A., Meislin, S. H. and Moore, M. J. (2003). A quantitative analysis of intron effects on mammalian gene expression. *RNA* 9, 607-617.

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.

Pagan, F. and Baralle, F. E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. *Nat. Rev. Genet.* 5, 389-396.

Roberts, G. C., Gooding, C., Mak, H. Y., Proudfoot, N. J. and Smith, C. W. (1998). Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res.* 26, 5568-5572.

Sanford, J. R., Gray, N. K., Beckmann, K. and Caceres, J. F. (2004). A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.* 18, 755-768.

Smith, C. W. and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* 25, 381-388.

Stamm, S. (2002). Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. *Hum. Mol. Genet.* 11, 2409-2416.

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

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.

Will, C. L. and Luhrmann, R. (2003). Spliceosomal UsRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* 13, 290-301.

Zhou, Z., Luo, M. J., Staesser, K., Katahira, J., Hurt, E. and Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* 407, 401-405.