Nuclear post-transcriptional control of gene expression

S A Akker, P J Smith and S L Chew

Department of Endocrinology, St Bartholomew’s Hospital, London EC1A 7BE, UK

ABSTRACT

The mammalian nucleus has considerable control over nascent transcripts. The basic mechanisms of post-transcriptional processing are well understood and recently some of the principles underlying the regulation of nuclear processing events have been elucidated. Here we review the recent progress in identification of signalling pathways that modulate the action of key RNA-binding proteins which regulate splicing, and the mechanisms of action of the C-terminal domain of RNA polymerase II that co-ordinate transcription with nuclear mRNA processing events.

INTRODUCTION

The human genome contains about 40 000 genes (International Human Genome Sequencing Consortium 2001, Venter et al. 2001). This is a relatively small number of genes considering the complexity of humans; for example, Saccharomyces cerevisiae and Caenorhabditis elegans genomes have 6000 (Goffeau et al. 1996) and 19 000 (C. elegans Sequencing Consortium 1998) genes respectively. The substantially increased human biological repertoire is provided by interactions between genes (epistasis), regulation of transcriptional programmes and control of post-transcriptional processing. This review will consider post-transcriptional events in the nucleus. Once a pre-messenger RNA (pre-mRNA) has been produced by transcription, substantial changes must occur before the mRNA is ready for export from the nucleus to the translation machinery. In addition, the kinetics of post-transcriptional processing must be co-ordinated with transcription in order to ensure efficient gene expression. Furthermore, variation in post-transcriptional processing provides substantial mRNA and protein diversity, with multiple isoforms generated from single genes. For example, the single human insulin-like growth factor-I (IGF-I) gene has three alternatively spliced isoforms (Jansen et al. 1983, Rotwein 1986, Chew et al. 1995), and since there are two promoters each expressing alternative signal peptides (Tobin et al. 1990), six peptide variants are made. Of the post-transcriptional events, splicing and polyadenylation are the major processes generating diversity, with alternative splicing being quantitatively more important than alternative polyadenylation (Claverie 2001). Regulation of mRNA stability and export contributes to the expression levels of a gene (Staton et al. 2000), but these processes have less influence on mRNA and protein heterogeneity. RNA editing is another nuclear process that generates protein diversity and this has been comprehensively reviewed elsewhere (Smith et al. 1997a, Holland et al. 1999). This review focuses on the regulation of pre-mRNA splicing and polyadenylation.

SPlicing

As organisms become more complex, the proportion of genes containing introns rises. Although fewer than 5% of S. cerevisiae genes contain introns, about 26% of expressed transcripts are derived from these genes (Ares et al. 1999). Thus, the presence of splicing confers an expression advantage. In humans, most genes contain introns. Furthermore, up to 60% of genes are alternatively spliced (International Human Genome Sequencing Consortium 2001, Kan et al. 2001). This generates substantial protein diversity (Black 2000).
The basic mechanisms of intron removal are well understood (Staley & Guthrie 1998, Reed & Oalandjian 2000). The intron is cleaved from the exons at the 5′ and 3′ ends, called the splice sites. The 5′ splice site consists of a short intronic sequence that loosely fits a consensus of GURAGU (G, guanine; U, uracil; A, adenine; and R, purine). The 3′ splice site consists of three elements: a branchsite (consensus YNYURAY, where Y is a pyrimidine); a stretch of pyrimidines (the polypyrimidine tract); and, finally the sequence CAG or UAG. These sequence elements must be recognised by the spliceosome, a multi-unit complex of proteins and RNA. The RNA components are small nuclear RNAs, U1, U2, U4, U5 and U6, assembled into ribonucleoprotein particles (snRNPs). Initially, the 5′ splice site of major introns is bound by U1 snRNP via Watson–Crick base-pairing (Fig. 1). A stepwise assembly of the spliceosome then occurs around the splice sites. The polypyrimidine tract of the 3′ splice site is bound by U2 snRNP and the U4/U6.U5 snRNPs are recruited. Correct selection of the splice sites is vital for gene expression, and regulating the use of different possible splice sites is fundamental to alternative splicing. Although necessary for splicing, the interaction between the U1 snRNP and the 5′ splice site is insufficient to account for the fidelity and flexibility of splicing because many sequences that match the 5′ splice site consensus are present in introns and exons and are bound by U1 snRNP, but are never used (Eperon et al. 1993, Sun & Chasin 2000). The main method of selecting the correct splice sites involves the co-ordinate recognition of nearby 3′ and 5′ splice sites, usually across an exon (called exon definition). The problem arises in explaining why other splice site-like sequences are not used for splicing even if they bind U1 and lie close to the sites that are used. This problem is usually explained by the surrounding sequence context (i.e. regulatory elements), which modulates the recruitment of the spliceosome.

Major advances in understanding regulated alternative splicing have come with the identification of some sequence elements involved in promoting exon selection (enhancers) or repressing splicing (silencers). The factors functioning through several regulatory sequences have been isolated. Many important non-snRNP proteins are involved in spliceosome assembly and function. A number of these also regulate splice site selection, in particular, a family of splicing factors characterised by RNA-recognition motifs and domains containing serine and arginine (SR) repeats, the SR proteins (Graveley 2000). SR proteins are required for general or constitutive splicing and are crucial mediators of regulated alternative splicing. Thus, members of the SR protein family bind and function at vertebrate exonic enhancers (Lavigne & al. 1993, Sun et al. 1993, Ramchatesingh et al. 1995, Gontarek & Derse 1996, Du et al. 1997, Selvakumar & Helfman 1999). Consensus exonic sequence motifs for several SR proteins have been derived experimentally (Liu et al. 1998, 2000, Schaal & Maniatis 1999) and are useful in predicting function (Liu et al. 2001).

Hormonal activation of signalling pathways can lead to modulation of the action of splicing factors and subsequent alteration in splice site choice. Several endocrine genes are subject to hormonally regulated alternative splicing (Chew 1997, Webster & Huang 1999) (Table 1), although these have only been partially characterised. A well-studied model is the splicing of three alternative exons of the fibronectin pre-mRNA (Magnuson et al. 1991, Inoue et al. 1999), where regulation reflects a
balance of splicing factors binding to several enhancers and silencers (Lavigueur et al. 1993, Caputi et al. 1994, Huh & Hynes 1994, Staalfa et al. 1997). A hormonal stimulus, insulin, changes splicing in rat fibronectin exon EIIB and is associated with increased levels of the rat SR protein, SRp40 (Du et al. 1997). This is not an isolated example, and there are several alternative splicing systems where SR proteins regulate exon selection. Alternative splicing in the mouse SRp20 pre-mRNA changes in response to serum stimulation or withdrawal. This change in splicing involves the SR proteins SF2/ASF (splicing factor-2/alternative splicing factor) and SRp20 (Jumaa & Nielsen 1997, Jumaa et al. 1997). In a different system, alternative splicing of CD4 and CD45 pre-mRNAs occurs in response to cytokine-induced T-cell differentiation (Screaton et al. 1995), via protein kinase C (PKC) and Ras pathways (Konig et al. 1998, Lynch & Weiss 2000). SF2/ASF and other SR proteins alter CD44 and CD45 splicing (Lemaire et al. 1999, ten Dam et al. 2000, Wang et al. 2001). Our recent data indicate the presence of an exonic splicing enhancer in a regulated exon of the human IGF-I gene and efficient splicing to this exon requires both the enhancer and the SR protein SF2/ASF (P J Smith & S L Chew, unpublished observations). These data are consistent with the well-documented splicing of growth hormone pre-mRNA. The removal of the last intron in this transcript is dependent on an exonic enhancer element and the SR protein, SF2/ASF (Sun et al. 1993). More recently, several SF2/ASF functional sites have been mapped in the downstream exon (Dirksen et al. 2000) and the actions of SF2/ASF at the enhancer is synergistic with another SR protein family member, 9G8, when the concentrations of SF2/ASF are low (Li et al. 2000). Thus, local concentrations and ratios of SR proteins are important in determining enhancement of splicing in alternative splicing systems.

Although more enhancers have been studied and characterised experimentally, recent evidence shows

---

**Table 1. Catalogue of hormonally regulated alternative splicing events**

<table>
<thead>
<tr>
<th>Alternatively spliced mRNA</th>
<th>Stimulus</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insulin receptor</td>
<td>Dexamethasone</td>
<td>Kosaki &amp; Webster (1993), Norgren et al. (1993, 1994a)</td>
</tr>
<tr>
<td>Cal/CGRP</td>
<td>Dexamethasone</td>
<td>Norgren et al. (1994a), Huang et al. (1996)</td>
</tr>
<tr>
<td>PKC beta</td>
<td>Insulin</td>
<td>Huang et al. (1994, 1996), Norgren et al. (1994b), Sell et al. (1994), Wiersma et al. (1997)</td>
</tr>
<tr>
<td>IGF-I</td>
<td>Growth hormone</td>
<td>Cote &amp; Gagel (1986)</td>
</tr>
<tr>
<td>FGF-R</td>
<td>Cytokines</td>
<td>Chalfant et al. (1995)</td>
</tr>
<tr>
<td>TNFα</td>
<td>2-Aminopurine</td>
<td>Jaworski et al. (1996)</td>
</tr>
<tr>
<td>PTP1B</td>
<td>PDGF, EGF, bFGF</td>
<td>Shifrin &amp; Neel (1993)</td>
</tr>
<tr>
<td>TGFβ, β-globin</td>
<td>src</td>
<td>Neel et al. (1995)</td>
</tr>
<tr>
<td>Hac1</td>
<td>UPR</td>
<td>Cox &amp; Walter (1996)</td>
</tr>
<tr>
<td>hPMCA2</td>
<td>Calcium</td>
<td>Zacharias &amp; Strehler (1996)</td>
</tr>
<tr>
<td>CD44</td>
<td>Phytobaemagglutinin</td>
<td>Screaton et al. (1995)</td>
</tr>
<tr>
<td>CD45</td>
<td>TPA, PDGF, IGF-I</td>
<td>Matter et al. (2000)</td>
</tr>
<tr>
<td>Fibronectin EIIB (rat)</td>
<td>Via hnRNP A1</td>
<td>Screaton et al. (1995)</td>
</tr>
<tr>
<td>Fibronectin ED (human)</td>
<td>Concanavalin A</td>
<td>Konig et al. (1998)</td>
</tr>
<tr>
<td>Kv3.1 channel</td>
<td>PKC and ras</td>
<td>Lynch &amp; Weiss (2000)</td>
</tr>
<tr>
<td>Agrin</td>
<td>Insulin, via HRS</td>
<td>Du et al. (1997)</td>
</tr>
<tr>
<td>MHC-B</td>
<td>TGFβ1, vitD, RA</td>
<td>Magnuson et al. (1991), Inoue et al. (1999)</td>
</tr>
<tr>
<td>BK channel</td>
<td>NGF</td>
<td>Smith et al. (1997b)</td>
</tr>
<tr>
<td>NR1</td>
<td>Serum/cell cycle</td>
<td>Itoh &amp; Adelstein (1995)</td>
</tr>
<tr>
<td>TRbeta</td>
<td>Hypophysectomy/ACTH</td>
<td>Jumaa et al. (1997)</td>
</tr>
<tr>
<td></td>
<td>Alcohol</td>
<td>Xie &amp; McCobb (1998)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hardy et al. (1999), Winkler et al. (1999)</td>
</tr>
</tbody>
</table>

bFGF, basic fibroblast growth factor; Cal/CGRP, calcitonin/calcitonin gene-related peptide; EGF, epidermal growth factor; FGF-R, fibroblast growth factor receptor; hPMCA2, human plasma membrane Ca-ATPase; Kv3.1, potassium voltage-gated channel; MHC-B, myosin heavy chain II-B; NGF, nerve growth factor; NRI, N-methyl-D-aspartate receptor subunit 1; PDGF, platelet-derived growth factor; PTP1B, phosphotyrosine-1B; TGF, tumour growth factor; TNF, tumour necrosis factor; TPA, 12-tetradecanoate 13-acetate; TRbeta, thyroid hormone receptor-β; UPR, unfolded protein response.
that the predominant mechanism of splicing may involve silencers (Fairbrother & Chasin 2001). This may be to ensure that multiple illegitimate splice sites and false exons contained within large introns are repressed and do not disrupt the reading frame. However, predicting silencer function is not possible at present, and the few silencers characterised have a diverse set of sequences (see references in Chew et al. 2000). Several protein factors have been identified that function at silencers, and these include members of the hnRNP (heterogeneous nuclear RNA particle) family. Recently, hnRNP A1 has been shown to regulate exon silencing in CD44 pre-mRNA and to be a target of oncogenic signalling pathways (Matter et al. 2000). Alternative splicing of the STREX exon in the BK channel pre-mRNA in the adrenal medulla or neuronal tissue is regulated by hypophysectomy, adrenocorticotrophin therapy or neuronal depolarisation (Xie & McCobb 1998). The calmodulin kinase IV pathway has recently been shown to repress splicing of the STREX exon through an element within a 54-nucleotide intronic region of the STREX 3′ splice site (Xie & Black 2001). A splicing factor called polypyrimidine tract binding protein (PTB) regulates alternative splicing of several genes by blocking the binding of factors such as U2AF65 to the polypyrimidine tract. However, calmodulin kinase IV activation does not alter the phosphorylation of PTB, so the splicing factors through which this pathway influences the silencer element are presently unknown.

The ratio of the SR protein SF2/ASF to hnRNP A1 determines splice site selection in several genes (Graveley 2000) (Fig. 2) and this is a mode by which external signals may alter splicing patterns. For example, a stress-induced p38 MAP-kinase signalling pathway induces hnRNP A1 phosphorylation, changes its localisation, and this switches alternative splicing (van der Houven van Oordt et al. 2000). The activity of SR proteins may be regulated by several mechanisms (Manley & Tacke 1996, Graveley 2000), including phosphorylation by kinases (Misteli 1999, Murray 1999), cellular localisation (Misteli et al. 1998) and varying tissue concentrations (Hanamura et al. 1998).

There are therefore two potential mechanisms for regulating alternative splicing by signalling pathways in differentiated tissues: (i) via a change in the ratio of ubiquitous splicing factors such as SF2/ASF, hnRNP A1; or (ii) through the use of splicing factors that are specific to a particular tissue or signalling pathway. These mechanisms are not mutually exclusive. Tissue-specific splicing factors are described, particularly in the context of neural and testes alternative splicing, and are closely related to general splicing factors (Venables et al. 1999, Markovtsov et al. 2000). To date, there is no evidence of a splicing factor specific to a signalling pathway.

### POLYADENYLATION/CLEAVAGE

The components of the polyadenylation machinery have been isolated and characterised (Barabino & Keller 1999). Two multi-protein complexes are involved: cleavage and polyadenylation specificity factor (CPSF) binds the AAUAAA motif, while cleavage stimulation factor (CstF) binds the downstream GU-rich region. CstF consists of three subunits of 77, 64 and 50 kDa (CstF-77, CstF-64 and CstF-50). CPSF, CstF, two cleavage factors (CF I_m and CF II_m) and poly(A) polymerase cleave the pre-mRNA, and then CPSF and poly(A) polymerase add the poly(A) tail of between 20–200 (A) nucleotides. The efficiency of polyadenylation and length of the tail may be regulated by the
action of poly(A) binding factor II. Several patterns of regulation are possible. Firstly, alternative polyadenylation/cleavage signals may be used. This changes the length of the 3′ untranslated region included in the mRNA. Second, the length of the poly(A) tail can vary. Third, the same polyadenylation signal is used, but the site of cleavage changes, as in the thyroglobulin pre-mRNA (Pauws et al. 2001).

Hormonal stimuli regulate polyadenylation and cleavage site selection and the length of the polyadenylation tail (Santra & Carter 1999). A good example of regulation of polyadenylation site usage is the action of follicle-stimulating hormone (FSH) during spermatogenesis. FSH stimulation promotes usage of an upstream polyadenylation/cleavage site in the cAMP-responsive element modulator-tau (CREM-τ) pre-mRNA, resulting in the exclusion of an instability element and an increase in CREM-τ levels (Foulkes et al. 1993). A comprehensive review of alternative poly(A) site selection has been published (Edwalds-Gilbert et al. 1997). An example of the regulation of poly(A) tail length is the effect of bromocriptine on the rat prolactin pre-mRNA (Carter et al. 1993).

The mechanisms by which external signals regulate polyadenylation or cleavage are not clear. More is understood about how cellular differentiation, growth control and DNA repair processes interact with the polyadenylation/cleavage machinery. Progress has been made in the context of B-lymphocyte differentiation, where the level of CstF-64 regulates polyadenylation site selection. The binding of CstF-64 and alternative polyadenylation/cleavage site selection can be blocked by hnRNP F (Veraldi et al. 2001). A link between cellular growth control and polyadenylation and cleavage is indicated by the modulation of the phosphorylation status and function of poly(A) polymerase by cyclin-dependent kinases (Colgan et al. 1998). Cyclin B(1) binds poly(A) polymerase directly (Bond et al. 2000). The polyadenylation/cleavage machinery is also regulated by DNA repair and tumour suppression mechanisms. Thus, the breast cancer gene product BRCA1 interacts with a BRCA1-associated RING domain protein (BARD1). DNA damage inhibits polyadenylation via the formation of a complex between BARD1/BRCA1 and CstF-50 (Kleiman & Manley 2001).

CO-ORDINATION

There is now substantial detail about the coupling of transcription to splicing and polyadenylation. The C-terminal domain (CTD) of RNA polymerase II (pol II) directs splicing and polyadenylation factors to the pre-mRNA (McCracken et al. 1997, Hirose & Manley 1998, 2000, Misteli & Spector 1999). This co-ordination between transcription and splicing also influences alternative splicing. Thus, the nature of the transcriptional promoter and complex influences splice site selection in the fibronectin pre-mRNA, via recruitment of SR proteins SF2/ASF and 9G8 (Cramer et al. 1999). A change in the conformation of the CTD of pol II may be the mode by which nuclear hormones and their receptors influence alternative splicing (Fig. 3). SR proteins and exonic splicing enhancers also function in both steps of the splicing reaction (Chew et al. 1999), perhaps to ensure co-ordination of the different steps of splicing during up-regulation of gene expression. The co-ordination between splicing and polyadenylation is well illustrated in the calcitonin/CGRP

![Figure 3. Schematic showing possible pathways involved in the regulation of splicing. The primary signals include insulin, serum stimulation, stress responses or nuclear hormones (vit D, vitamin D; RA, retinoic acid; T3, thyroid hormones; dex, dexamethasone). The areas of uncertainty are shown by question marks. For example, it is not clear if PKC and ras pathways function via SR protein kinases (SRPK) or through an unknown set of intermediates. Other abbreviations: PP2Cg, protein phosphatase 2 gamma; clk/sty, clk/sty kinase; CamKIV, calmodulin kinase IV.](https://www.endocrinology.org)
pre-mRNA, where polyadenylation and splicing factors interact to regulate the use of alternative terminal exons (Lou & Gagel 1999).

DISRUPTION

Disruption of existing splice sites or introduction of new splice sites via DNA sequence mutations may result in incorrect pre-mRNA splicing leading to genetic disease. One of the earliest understandings of a mechanism of genetic disease was in the splicing defect of a thalassaemic globin gene (Treisman et al. 1982). It is now clear that DNA mutations resulting in abnormal splicing cause a substantial proportion of genetic disease (Krawczak et al. 1992). The commonest functional consequence of genetic mutations in many familial endocrine diseases is aberrant splicing; e.g. in CYP21B (Speiser et al. 1992, Kapelari et al. 1999), NFI (Ars et al. 2000) and MENIN genes (Mutch et al. 1999). DNA mutations affecting splicing may be classified into those that disrupt the splice sites themselves (Krawczak et al. 1992), or those that change nonsplice site sequences (Valentine 1998). In the former class, mutations at 5’ splice sites may cause activation of nearby cryptic 5’ splice sites, or skipping of the entire adjacent upstream exon (Robberson et al. 1990). In the latter class, mutations of non-splice site sequences may disrupt regulatory elements for nearby splice sites (Liu et al. 2001). Another mode of disrupting normal splicing is through mutations that activate the splicing of a false exon which is normally never expressed, and several examples occur in cystic fibrosis (Friedman et al. 1999). In an endocrine example, we showed a point mutation in a false exon was necessary and sufficient for splicing of the false exon, thus disrupting the growth hormone receptor mRNA and causing Laron syndrome (Metherell et al. 2001).

SUMMARY

Recent insights have added greatly to our understanding of the mechanisms governing the regulation of post-transcriptional mRNA nuclear processing. SR proteins and hnRNPs are important regulators of pre-mRNA splicing and bind pre-mRNA at key regulatory elements. Signalling pathways alter splicing and the properties of such RNA-binding proteins. The details of the intermediates between the signalling pathways and splicing protein factors are still unclear (Fig. 3). There may be a specific set of proteins that directly couple signalling cascades and RNA-binding proteins. Alternatively, the action of signalling on RNA-binding protein function may be indirect, perhaps via effects on the cell cycle, as several kinases and phosphatases associated with modulating phosphorylation status of splicing proteins are also cell cycle-regulated factors (Burns & Gould 1999). The action of some of the steroid hormones on alternative splicing may be indirect, through changes in the structure of the transcriptional complex and the configuration of the CTD of pol II. Thus, there is still much work required to fully elucidate the molecular mechanisms and importance of pre-mRNA splicing in regulating gene expression. Insight into these mechanisms will have an impact on our understanding of certain genetic endocrine diseases and perhaps in development of novel therapies for the future.

REFERENCES


Chew SL, Liu H-X, Mayeda A & Kramer AR 1999 Evidence for the function of an exonic splicing enhancer after the
first catalytic step of pre-mRNA splicing. PNAS 96 10655–10660.


Claverie JM 2001 Gene number. What if there are only 30 000 human genes? Science 291 1255–1257.


Cote GJ & Gagal RF 1986 Dexamethasone differentially affects the levels of calcitonin and calcitonin gene-related peptide mRNAs expressed in a human medullary thyroid carcinoma cell line. Journal of Biological Chemistry 261 15524–15528.


Eperon IC, Ireland DC, Smith RA, Mayeda A & Krainer AR 1993 Pathways for selection of the 5′ splice sites by U1 snRNPs and SF2/ASF. EMBO Journal 12 3607–3617.


Hardy PA, Chen W & Wite PA 1999 Chronic ethanol exposure and withdrawal influence NMDA receptor subunit and splice variant mRNA expression in the rat cerebral cortex. Brain Research 819 33–39.


Jumaa H & Nielsen PJ 1997 The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. EMBO Journal 16 5077–5085.


Kapelari K, Ghanati Z, Wollmann H, Ventz M, Ranke MB, Koller R & Peters H 1999 A rapid screening for steroid 21-hydroxylase mutations in patients with congenital adrenal...
Konig H, Ponta H & Herrlich P 1998 Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. EMBO Journal 17 2904–2913.


Selvakuma M & Helfman DM 1999 Exonic splicing enhancers contribute to the use of both 3′ and 5′ splice site usage of rat beta-tropomyosin pre-mRNA. *RNA* **5** 378–394.


**RECEIVED 20 April 2001**

**ACCEPTED 15 June 2001**