Alternative pre-mRNA splicing: the logic of combinatorial control

Christopher W.J. Smith and Juan Valcárcel

Alternative splicing of mRNA precursors is a versatile mechanism of gene expression regulation that accounts for a considerable proportion of proteomic complexity in higher eukaryotes. Its modulation is achieved through the combinatorial interplay of positive and negative regulatory signals present in the RNA, which are recognized by complexes composed of members of the hnRNP and SR protein families.

EUKARYOTIC mRNAs ARE transcribed as precursors containing intervening sequences (introns). These sequences are subsequently removed such that the flanking regions (exons) are spliced together to form mature mRNA. Alternative splicing pathways generate different mRNAs encoding distinct protein products, thus increasing the coding capacity of genes. Alternative splicing can also act as an on–off gene expression switch by the introduction of premature stop codons.

Figure 1 illustrates examples of different modes of alternative splicing, which have dramatic biological consequences. Combinations of these basic events in complex transcription units can generate a multitude of protein isoforms. A remarkable example is the cell-specific expression of a subset of 576 possible alternatively spliced forms of a K^+ channel mRNA. These are expressed in a gradient along the 10 000 sensory-receptor cells present in the inner ear of birds, which enables perception of different sound frequencies[1].

Alternative splicing is often tightly regulated in a cell-type- or developmental-stage-specific manner. Coordinated changes in alternative splicing patterns of multiple pre-mRNAs are an integral component of gene expression programmes like those involved in nervous system differentiation[2] and apoptotic cell death[3]. Splice-site choice must therefore be tightly regulated in time and space. The purpose of this article is to review recent evidence from a variety of mammalian and *Drosophila* genes which supports the notion that even simple decisions can result from a complex interplay between signals in the RNA and *trans*-acting factors that assemble on them. (Excellent reviews on more general aspects of splicing regulation can be found in Refs 4,5.)

C.W.J. Smith is at the Dept of Biochemistry, University of Cambridge, 80 Tennis Court Road, Old Addenbrookes Site, Cambridge, UK CB2 1GA; and **J. Valcárcel** is at the Gene Expression Programme, European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany.
Emails: cwjs1@mole.bio.cam.ac.uk; juan.valcarcel@embl-heidelberg.de

Needles in a haystack

The question of splice-site choice is intimately connected to the problem of normal recognition of constitutive splice sites[6]. A feature shared by both regulatory sequences and splice-site signals is that they are usually short and often degenerate. The information content of their primary sequence is therefore rather limited (Fig. 2a). Even the best computer programs are only 50% accurate in predicting actual splice sites over multiple, equally good candidate sequences that are not used. Splice-site recognition represents a daunting problem, underscored by the fact that 15% of human genetic diseases are caused by mutations that destroy functional splice sites or generate new ones[7].

Initial recognition of splice sites involves crosstalk between multiple, relatively weak interactions that contribute to establishing complexes that commit the pre-mRNA to splicing[8]. The splicing reaction occurs in the spliceosome, a complex composed of five small nuclear ribonucleoprotein particles (snRNPs) and 50–100 polypeptides, many of which are not associated with snRNPs (Refs 9,10). The early steps of spliceosome assembly, which provide the main targets for regulation, involve recognition of the consensus elements at both ends of the intron (Fig. 2a). U1 snRNP binds to the 5′ splice site, splicing factor 1 (SF1, also known as branch-point-binding protein or BBP) to the branch point[11], whereas the 65 and 35 kDa subunits of U2 snRNP auxiliary factor (U2AF) recognize the polypyrimidine tract and 3′ AG, respectively[12] (Fig. 2b). Bridging interactions between U1 snRNP bound to the 5′ splice site and SF1/U2AF bound to the 3′ splice-site region necessarily exist, but their molecular nature has not yet been fully established. Candidate bridging factors in higher eukaryotes are members of the serine–arginine (SR) family of splicing factors[13,14]. These proteins contain N-terminal RNA recognition motifs, which mediate binding to pre-mRNA. Their C-terminal arginine–serine-rich (RS) domains mediate protein–protein interactions with similar RS domains in U2AF[35] and U1 snRNP 70K protein, thereby promoting U2AF and U1 snRNP binding to splice sites (Fig. 2b)[4,5,14]. One class of sequences bound by SR proteins are exon splicing enhancers (ESEs), which are often purine rich and play stimulatory roles in both constitutive and regulated splicing[15]. Different ESEs are recognized by specific subsets of SR proteins[14].

Vertebrate exons are usually short and separated by long introns that can be many kilobases in length. For the majority of internal exons, which are between 50 and 300 nucleotides in length, initial recognition of splice sites is enhanced by interactions between the 3′ and 5′ splice sites across the exon in a process termed exon definition[16] (Fig. 2b). How the pairing between splice sites across exons is subsequently swapped to pairing of sites across the (usually) much longer introns is an important question; this step is another potential target for regulating alternative splicing. A second question is how are the terminal 5′ and 3′ exons defined? The 7-methylguanosine triphosphate (m^7Gppp) cap structure at the 5′ end of the transcript promotes recognition of the first 5′ splice site, and polyadenylation signals at the 3′ end promote the use of the last 3′ splice site[17]. These processing events all occur co-transcriptionally, and the functional coupling between them reflects the integration of all steps of mRNA synthesis in mRNA 'factories'. The hyperphosphorylated C-terminal domain (CTD) of the

0968–0004/00/$ – See front matter PII: S0968-0004(00)01604-2

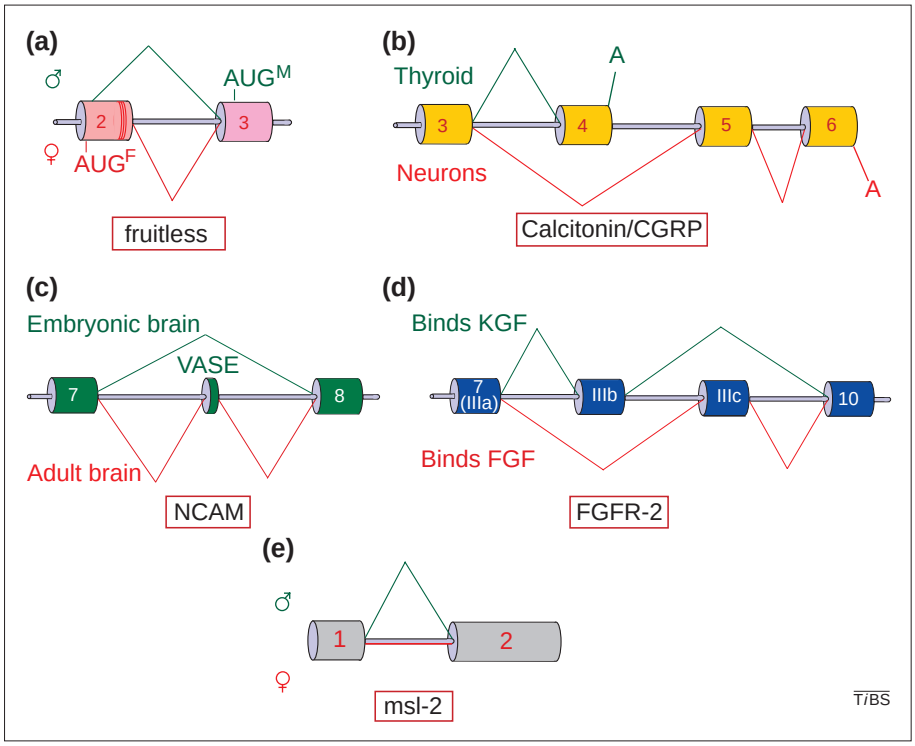

Figure 1

Different modes of alternative splicing and examples of its biological consequences. **(a)** Alternative 5′ splice-site use in the *Drosophila* gene *fruitless* governs sexual orientation and behaviour. Male (green) and female (red) patterns of splicing, as well as translation initiation codons giving rise to long open reading frames, are indicated. Red lines in exon 2 represent binding sites for Tra (for 'Transformer') and Tra-2. **(b)** Alternative 3′ splice-site usage, associated with differential use of polyadenylation sites (represented by A) in the vertebrate gene for calcitonin and calcitonin-gene-related peptide (CGRP) generates a calcium homeostatic hormone in the thyroid gland or a vasodilator neuropeptide in the nervous system. Processing patterns in green are found in thyroid, those in red are found in neurons. **(c)** Differential inclusion or skipping of the variable alternatively spliced exon (VASE) in the gene for neural cell adhesion molecule (NCAM) in embryonic (green) versus adult (red) rat brain, represses or promotes axon outgrowth during development. **(d)** Mutually exclusive use of exons IIIb and IIIc in mammalian fibroblast growth factor receptor 2 (FGFR-2) changes its binding specificity for growth factors during prostate cancer progression. The pattern of splicing represented in green generates an mRNA encoding a receptor with high affinity for keratinocyte growth factor (KGF), whereas that in red generates a receptor with high affinity for FGF. **(e)** Female-specific retention of an intron at the 5′ untranslated region (UTR) of the gene *male-specific-lethal 2* (*msl-2*) allows export of the unspliced RNA to the cytoplasm. The protein Sex-lethal facilitates both intron retention in the nucleus and translational repression in the cytoplasm, thereby switching off *msl-2* expression, which controls X-chromosome dosage compensation.

large subunit of RNA polymerase II (Fig. 2b) binds capping, polyadenylation and SR splicing factors. In addition to delivering some processing factors to their sites of action, there is evidence that the CTD can play a more direct role in the processing reactions themselves[18].

Packing and remodelling RNA

Primary transcripts form densely packed ribonucleoprotein complexes, known as heterogeneous nuclear (hn)RNPs, by associating with a family of polypeptides known as hnRNP proteins[19,20]. These are a diverse group of nuclear RNA-binding proteins that are involved in multiple functions. They contain various types of RNA-binding motifs, as well as domains rich in glycine and other amino acids (but not RS domains), which might serve in both RNA binding and protein–protein interactions. The set and arrangement of hnRNP proteins bound to an RNA influences its fate within the nucleus. hnRNP proteins of the A, B and C families can assemble with nascent pre-mRNA into regular 40S structures, which have been compared with the nucleosomal packaging of DNA (Ref. 19). Although assembly of 40S structures has been argued to be nonspecific[19], the individual hnRNP proteins do have preferred binding sequences[20]. This could serve to 'phase' the packaging of the pre-mRNA, and this, in turn, could antagonize or assist splice-site selection. Such packaging could compete with assembly of early-splicing complexes across exons and might explain why exons larger than 300 nucleotides are recognized inefficiently[8]. hnRNP packaging can also bring together distant regions of the pre-mRNA and therefore assist splice-site pairing[21]. Several hnRNP and SR proteins accompany the mRNA from the nucleus and might influence mRNA nucleocytoplasmic transport, and even cytoplasmic translation, RNA localization and decay[20,22]. Many others are removed from the RNA before export. These changes in the hnRNP complement of the RNA are achieved in part by the processing complexes themselves. But, by analogy with chromatin, it is possible that specific remodelling machines might also play a role in altering the packaging of transcripts throughout their lifetime.

Antagonism in splice-site selection

Early biochemical studies indicated that hnRNP proteins could regulate splice-site choice[13]. SV40 virus large T and small t proteins are generated from the same pre-mRNA by the use of alternative 5′ splice sites. An excess of the SR protein alternative splicing factor (ASF, also known as splicing factor 2 or SF2) promoted the use of the proximal site. This observation was subsequently extended to other RNAs and SR proteins. A model that explains these observations is that under limiting concentrations of ASF/SF2, U1 snRNP binds only to functionally stronger splice sites. Thus, a weak 5′ splice site might not be selected, despite its proximity to the 3′ splice site. Higher levels of ASF/SF2 promote full occupancy of all 5′ splice sites by U1 snRNP, and under these conditions the 5′ splice site closest to the 3′ splice site is selected[4,13,23].

By contrast, hnRNP A1 antagonizes this activity of SR proteins, causing a shift to selection of distal 5′ splice sites. The molecular mechanisms underlying the effects of hnRNP A1, and whether these are related to RNA packaging (see above), are less well understood. Access to the proximal site by ASF/SF2 could be blocked by hnRNP A1. Indeed, binding of hnRNP A1 to certain exon splicing silencers (ESS) inhibits the use of adjacent 3′ splice sites[24,25]. Inhibition can even be maintained after replacement of the ESS by a heterologous binding site, if the cognate RNA-binding protein has been fused to the C-terminal

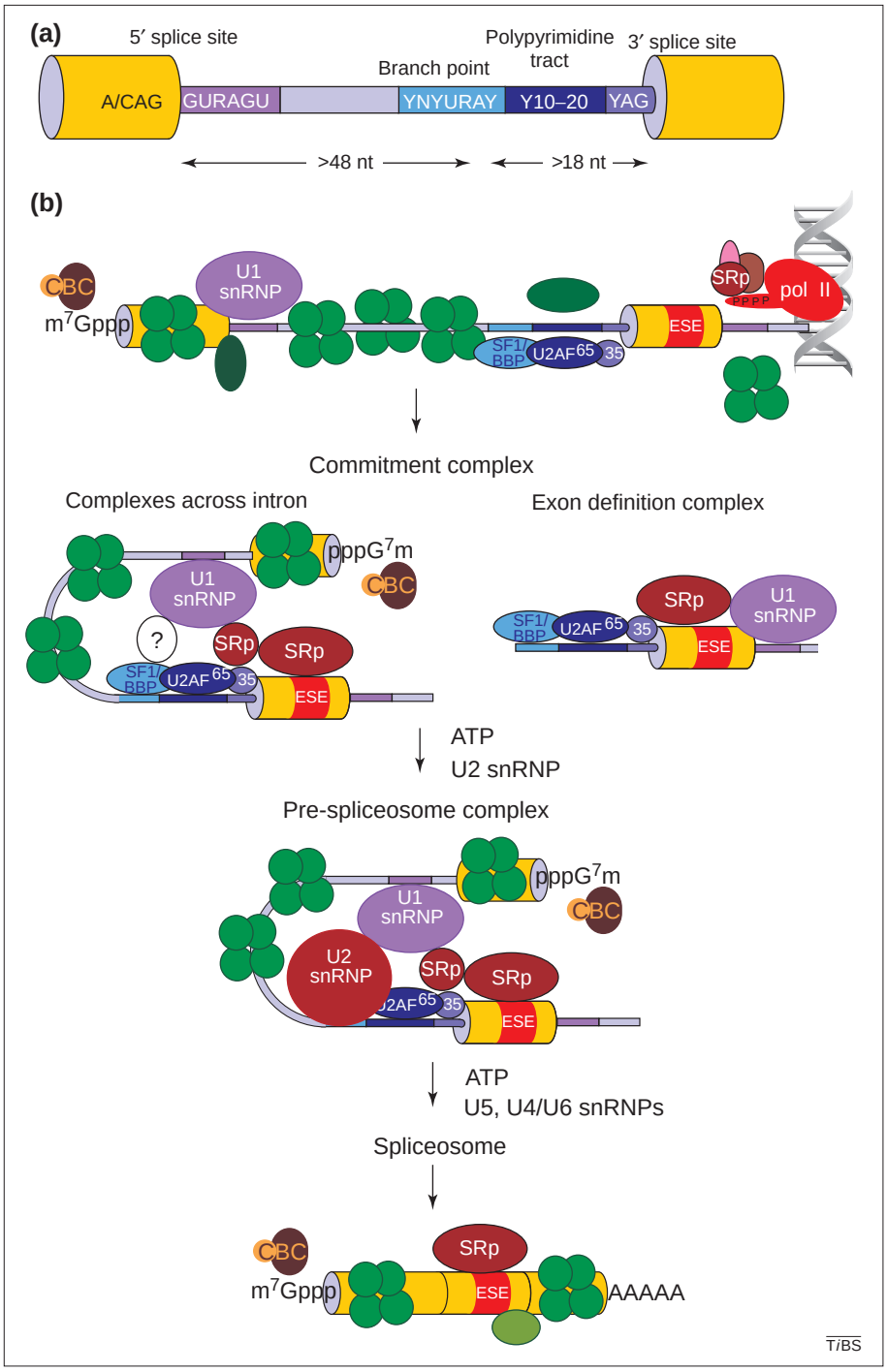

Figure 2

Splice-site elements and splicing complex assembly. **(a)** Consensus splice-site elements for a typical metazoan intron (Y, pyrimidine; R, purine; N, any nucleotide). The GU and AG dinucleotides at the intron termini and the A at the branch point are nearly invariant; the other positions show variable conservation. The minimal distances between the 5′ splice site and branch point, and between the 3′ splice site and branch point are indicated. Other sequences – enhancers and silencers – are able to influence the efficiency of selection of particular splice sites. **(b)** Nascent RNA transcribed by RNA polymerase II is immediately bound by various heterogeneous nuclear ribonucleoproteins (hnRNPs) (represented by the various green shapes), cap-binding complex (CBC, orange and brown) and splicing factors. The green tetramers represent general packaging by hnRNP A, B and C proteins. Binding of some hnRNP proteins might compete directly with splicing factors. The C-terminal domain of the large subunit of RNA polymerase II (pol II, red) is hyperphosphorylated (P) during transcription elongation and associates with RNA-processing factors, including capping enzymes (brown), polyadenylation factors (pink) and SR proteins (SRp), which are delivered to the nascent transcript[18]. U1 small nuclear (sn)RNP first binds the 5′ splice site[9], splicing factor 1/branch-point-binding protein (SF1/BBP) the branch point[11], 65-kDa U2 snRNP auxiliary factor ($U2AF^{65}$) the polypyrimidine tract[10], and $U2AF^{35}$ the 3′ splice site AG (Ref. 12). These interactions allow assembly of an early or commitment complex. Next, U2 snRNP binding to the branch point and ATP hydrolysis are required to form the pre-spliceosomal complex. Both complexes involve interactions that bridge the 5′ and 3′ splice sites. As indicated, these interactions might occur across the intron destined to be spliced (left). Alternatively, they might initially occur across exons (right). SR proteins promote assembly of commitment complexes by assisting binding of both U1 snRNP and U2AF, and they might form a bridge between the 5′ and 3′ splice sites. In the exon definition complexes, the SR proteins are depicted as binding to an exon splicing enhancer (ESE). An alternative intron-bridging interaction might occur between SF1/BBP and U1 snRNP components. Subsequent binding of U5 and U4/6 snRNPs allows the assembly of a mature spliceosome across the intron. The spliced mRNA is then released from the spliceosome, and changes in hnRNP composition can occur, which might facilitate nucleocytoplasmic export and even affect cytoplasmic events.

glycine-rich domain of hnRNP A1 (Ref. 24). The hnRNP A1 protein also regulates splicing of its own pre-mRNA by promoting skipping of an exon flanked by optimal hnRNP-A1-binding sites. Interaction between hnRNP A1 proteins bound to the flanking sites could promote splicing of the distal exons by bringing their splice sites closer together[21]. Tables 1 and 2 summarize known activities of hnRNP and SR proteins in splicing regulation.

A corollary of these findings is that variations in the relative concentrations of general splicing factors and hnRNPs can simultaneously affect splice-site choice in multiple RNAs (Ref. 13). These ratios could define a cellular code for establishing cell-specific patterns of splicing in multiple genes. They could also be subverted in disease. For example, changes in levels of SR proteins occur during tumour progression and correlate with changes in the inclusion of cassette exons in the *CD44* gene, which encodes a cell-adhesion molecule involved in cancer metastasis[26]. Another example is the expansion of CUG repeats in the 3′ untranslated region of the *DM* gene mRNA, which causes myotonic dystrophy. Titration of CUG-binding hnRNP proteins by the expanded repeats might alter splicing of some genes, thus explaining some of the associated pathologies[27].

Another mechanism of splicing regulation involves tissue- or developmental-stage-specific factors. Sex-lethal (SXL) is a *Drosophila* hnRNP-like protein

Table 1. Roles of hnRNP and other cellular proteins without RS domains in alternative splicing

Protein	Organism	Known target genes	Inhibition (−) or activation (+) of splicing	Mechanism	Refs[a]
Sxl	*Drosophila*	*Sxl*, *tra*, *msl-2*	−	Causes female-specific splicing of *Sxl*, *tra* and *msl-2* pre-mRNAs. Binds cooperatively to sites flanking *Sxl* exon 3, causing exon skipping (autoregulation), to pyrimidine tract of regulated *tra* 3′ splice site in competition with U2AF, and to sites adjacent to 5′ and 3′ splice sites of *msl-2* intron 1	a–d
PSI	*Drosophila*	Gene for P-element transposase	−	Part of germ-cell-specific complex along with hrp48 that stabilizes U1 snRNP binding to a nonfunctional pseudo-5′ splice site. Inhibits use of authentic 5′ splice site of intron 3. Expressed specifically in germ line	e,f 34
hrp48	*Drosophila*	Gene for P-element transposase	−	(see PSI)	e.g. 34
hnRNP A1	Mammalian	HIV tat, FGF receptor hnRNPA1 (autoregulation)	−	Binding to exon silencers of tat and FGFR-2 Promotes exon skipping or represses adjacent 3′ splice site Binds to intron elements flanking exon 7B of hnRNPA1. Causes exon skipping by 'looping out' sequences between the binding sites?	25,26 h,i,j 21
hnRNP-F	Mammalian	*Src*	+	Binds downstream intron splicing enhancer	k
hnRNP-H	Mammalian	*Src*	+	As above	l
hnRNP-H	Mammalian	β-Tropomyosin	−	Binds exon silencer. Represses adjacent 3′ splice site, promotes exon skipping	m
PTB (hnRNP-I)	Mammalian	*Src* and genes for α- and β-tropomyosins, α-actinin, $GABA_A$ γ2 receptor	−	Binds sequences flanking and within regulated exons, often within polypyrimidine tract. Causes exon skipping. Represses 3′ and 5′ splice sites	n–q 28–33
PTB (hnRNP-I)	Mammalian	Gene for calcitonin/CGRP	+	Binds to 'pseudo exon' enhancer that promotes use of calcitonin polyA site	47
CUG-BP (hNAB50)	Mammalian	Genes for cardiac troponin-T, CLCB (?), NMDA (?)	+	Binds CUG repeats downstream of cTnT exon 5, promoting inclusion. Binds pyrimidine tract of CLCB and NMDA	27,31
KSRP	Mammalian	*Src*	+	Binds downstream intron enhancer	35

[a]Numbers refer to literature cited in the text; letters a–q refer to additional citations available at http://www.bio.cam.ac.uk/dept/biochem/cwjs_tibs/
Abbreviations: CGRP, calcitonin-gene-related peptide; CLCB, clathrin light chain B; CUG-BP, CUG-binding hnRNP protein; FGFR-2, fibroblast growth factor receptor 2; $GABA_A$, gamma aminobutyric acid receptor; hrp48, *Drosophila* homologue of hnRNP A1; KSRP, KH-domain-containing factor; msl-2, male-specific-lethal-2; NMDA, *N*-methyl-*D*-aspartate; snRNP, small nuclear ribonucleoprotein; PSI, P-element somatic-cell-specific inhibitor; PTB, polypyrimidine-tract-binding protein; RS, arginine–serine-rich; Sxl, sex-lethal.

produced exclusively in female flies. It induces female-specific patterns of alternative splicing and represses translation of various target genes. This triggers regulatory cascades that control all aspects of sex determination[5]. The binding of SXL to a specific 3′ splice site polypyrimidine tract in *transformer* stops U2AF from binding to this site, thus diverting U2AF to an alternative site of lower affinity[5]. Like hnRNP A1, SXL also promotes an exon-skipping event in its own pre-mRNA by binding to both of the flanking introns, and in this case its glycine-rich N-terminal domain is important for cooperative binding to RNA and exon skipping[5].

Polypyrimidine-tract-binding protein (PTB or hnRNP-I) also binds to specific polypyrimidine tracts and appears to play a similar role to SXL (Refs 28–32). In some cases it has been suggested simply to compete with U2AF binding. Nevertheless, PTB also inhibits 5′ splice sites and in most cases where PTB is involved in regulation; exon skipping requires multiple PTB-binding sites in the vicinity of the regulated exon. Regulated exons might be packaged by PTB into complexes inaccessible to the splicing machinery. A number of neuron-specific exons are repressed in non-neuronal cells by PTB (Refs 30–33). In some cells where these exons are included, PTB levels are significantly reduced, whereas expression of a novel – potentially less inhibitory – neuronal PTB isoform is elevated[2,32,33]. It will be important to determine whether other tissue-specific PTB homologues with distinct regulatory activities exist, and also whether the three known alternatively spliced isoforms of PTB have distinct activities.

Regulatory complexes

Simple antagonism between individual hnRNP proteins and splicing factors might represent the basic operations of a more complex interplay between constitutive factors and dedicated regulators to achieve cell-type-specific splicing. The examples that follow illustrate this point.

Exon splicing silencer. P-element transposase is produced only in germline cells of *Drosophila*. Inhibition of intron 3 splicing in somatic cells leads to production of an inhibitor of transposase (Fig. 3a, Table 1). This is achieved by a complex bound to a regulatory sequence in exon 3, which promotes U1 snRNP assembly at a nonfunctional pseudo-5′ splice site and prevents its productive interaction with the authentic 5′ splice site. The complex contains at least four polypeptides, including a P-element somatic-cell-specific inhibitor (PSI), which contains RNA-binding domains with homology to those found in hnRNP K (K-homology or KH domains), and a *Drosophila* homologue of hnRNP A1 (hrp48), which targets the complex by binding to a second adjacent pseudo-5′ splice site[34].

Intron splicing enhancer. A complex that assembles on an intronic splicing enhancer (ISE), which promotes neuron-specific inclusion of a small exon in the mammalian gene *Src*, has

Table 2. Roles of proteins with RS domains in alternative splicing[a]

Protein	Organism	Known target genes	Inhibition (−) or activation (+) of splicing	Mechanism	References[b]
Tra	*Drosophila*	*dsx*	+	Induces cell-specific assembly of stable enhancer complex along with Tra-2 and other SR proteins.	36
		fru		Activates downstream female-specific 5′ splice site in collaboration with Tra-2	a
Tra-2	*Drosophila*	*dsx*, *fru*	+	Functions with tra (see above) in *dsx* and *fru* regulation	36
Tra-2	*Drosophila*	*Tra-2*	−	Causes intron retention (autoregulation)	b,c
RBP1 (d9G8?)	*Drosophila*	*dsx*	+	Functions with tra (see above) in *dsx* regulation	36,39
dSRp30 (dSF2?)	*Drosophila*	*dsx*	+	Functions with tra (see above) in *dsx* regulation	36
RSF1	*Drosophila*	Not known	−	Antagonizes SR proteins	d
SWAP	*Drosophila*	Gene for SWAP	−	Inhibits splicing of first two introns	e,f
SRp20	Mammalian	Gene for SRp20	+	Autoregulation. Promotes exon 4 inclusion. Antagonized by SF2/ASF	g
		Gene for calcitonin		Binds intron enhancer, activates polyadenylation (see PTB, Table 1)	h
SRp30a/SF2/ASF	Mammalian	Genes for bovine growth hormone, fibronectin	+	Binds exonic enhancer, activates upstream of 3′ splice site	i, j
		Gene for β-tropomyosin		Binds intron enhancer, activates upstream of 5′ splice site Antagonized by SC35	k
		Gene for CD45			l
SRp30a/SF2/ASF	Mammalian	Gene for SRp20	−	Antagonizes SRp20-activated exon inclusion	m
		Gene for RSV		Inhibits splicing by binding to negative regulatory signal	n
		Gene for adenovirus L1 transcription unit		Binds to enhancer element upstream of branch point. Sterically obstructs U2 snRNP binding	41
SC35	Mammalian	Gene for β-tropomyosin	−	Antagonizes SF2 at intron enhancer	k
9G8	Mammalian	Gene for fibronectin	+	Binds to exonic enhancer	j
SRp40	Mammalian	Gene for cTnT	+	Binds to exonic enhancer	o,p
		Gene for fibronectin			
SRp55	Mammalian	Gene for cTnT	+	Binds to exonic enhancer	o,q
		Gene for CD45			l

[a]The SR protein family is a subset of splicing factors that contain arginine–serine-rich (RS) domains. Most SR family members are able to influence alternative splicing decisions when overexpressed. The examples given here involve specific interactions of SR proteins with defined enhancer or repressor elements.
[b]Numbers refer to literature cited in the text; letters a–q refer to additional citations available at http://www.bio.cam.ac.uk/dept/biochem/cwjs_tibs/
Abbreviations: cTnT, cardiac troponin-T; *dsx, double sex; fru, fruitless*; PTB, polypyrimidine-tract-binding protein; RBP1, RNA-binding protein 1; RSF1, repressor splicing factor 1; SF2/ASF, splicing factor 2/alternative splicing factor ; SWAP, suppressor of white apricot; Tra, Transformer.

similarities with the complex involved in P-element regulation, even though both the sequence recognized and the regulatory outcome are completely different (Fig. 3b). Like the P-element complex – but unlike the majority of characterized exonic splicing enhancers, which bind SR proteins – the *Src* ISE also contains hnRNP proteins (H and F) and a KH-domain-containing factor (KSRP) related in sequence to PSI. Although none of the known components of the complex is neuron-specific, crosslinking of KSRP to the enhancer sequence is increased in extracts from neuronal cells, suggesting that the conformation or stability of the complex varies between cell types[35].

Exon splicing enhancer. An example of a regulated exonic splicing enhancer (ESE) for which the molecular architecture of the complexes formed is well characterized is found in the *Drosophila* gene *doublesex* (*dsx*) (Fig. 3c). Female-specific splicing of exon 4 requires a region located just over 300 nucleotides downstream of the 3′ splice site, which contains six 13-nucleotide repeats, as well as a purine-rich element (PRE). A complex containing Tra and Tra-2 proteins assembles on each of these sequences along with a specific member of the SR family – RBP1 on the repeats and dSF2/ASF on the PRE (Fig. 3d). Tra and Tra-2 possess RS domains, but they do not have the domain organization typical of SR proteins. Both RNA–protein and protein–protein contacts facilitate highly cooperative assembly of the complex on each *dsx* repeat. Tra acts as the key female cell-specific switch. Only in the presence of all three proteins can a stable enhancer complex form[36]. The distance of the enhancer from the 3′ splice site is crucial for cell-specific regulation by Tra. If this separation is reduced, Tra-2 and SR proteins can mediate activation from these elements without requiring the additional stabilization afforded by Tra. Perhaps Tra function is required to overcome the same constraints that impose the distance limit for exon definition. Assembly of the complex facilitates recruitment of splicing factors on the weak female-specific 3′ splice site. Absence of functional synergy between the enhancer repeats, demonstrated by varying their copy number, implies a single activation target[15]. Because RS domains tethered through heterologous RNA-binding motifs can mediate enhancer effects[37], the target could be stabilization of U2AF binding through RS-domain-mediated protein–protein interactions involving enhancer-bound components and U2AF35. This simple and attractive model is at present controversial as a general explanation for

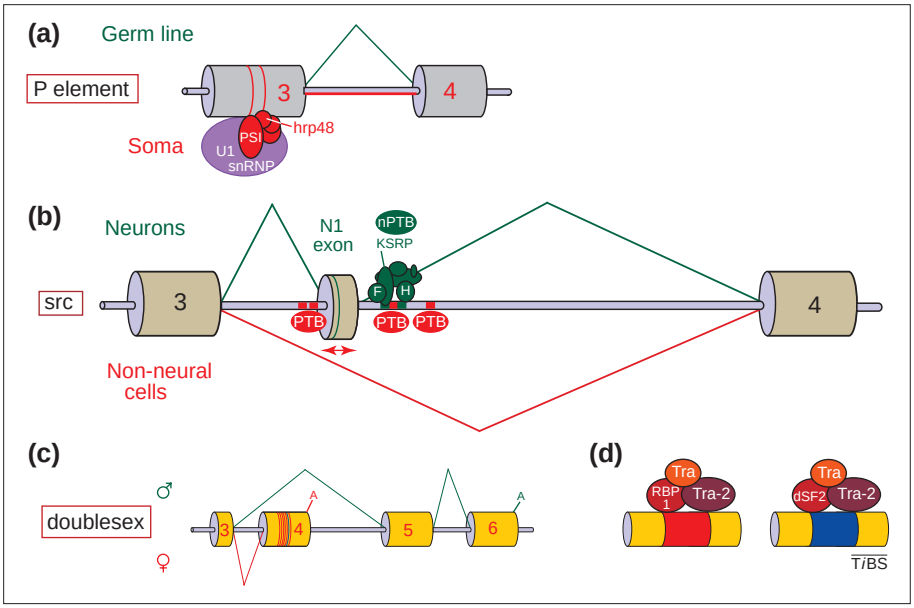

Figure 3

Regulatory complexes and combinatorial control. **(a)** Intron 3 retention in *Drosophila* P-element generates RNAs encoding a repressor of transposition, whereas splicing of intron 3 in the germline allows expression of transposase. The red lines in exon 3 represent pseudo-5′ splice sites that act as exonic silencers. They are bound by a multiprotein complex containing the hnRNP protein hrp48 and P-element somatic inhibitor (PSI, red ovals) in somatic cells that facilitates assembly of U1 small nuclear ribonucleoproteins (snRNP; magenta oval) on the upstream site[34]. **(b)** The 18-nucleotide N1 exon from the *Src* gene is inserted between exons 3 and 4 in neurons, generating a variant SH3 domain in the tyrosine kinase. *Cis*-acting sequences and *trans*-acting factors promoting exon inclusion or skipping are indicated with the same colour as the alternative splicing pathway. A short exon length is important to avoid exon inclusion in non-neural cells. Exon skipping is also promoted by polypyrimidine-tract-binding protein (PTB) binding upstream and downstream from the N1 exon. A multiprotein complex (green ovals) containing heterogeneous nuclear (hn)RNPs (F,H), KH-domain-containing factor (KSRP) and other polypeptides assembles on an enhancer downstream of exon N1, which has synergistic effects with a sequence at the 5′ end of exon N1. A neuron-specific form of PTB (nPTB) might also facilitate exon inclusion by antagonizing the inhibitory function of PTB. **(c)** Alternative 3′ splice site usage in the *Drosophila* gene *doublesex* allows production of sex-specific transcription factors with different C-terminal domains. A region of exon 4 containing six repeats of 13 nucleotides (represented by red lines), and a purine-rich element (PRE, represented by the blue line) acts as an exonic enhancer that favours exon 4 inclusion in females. **(d)** Each of the *dsx* repeats (red) and the PRE (blue) is bound by an SR protein, Tra-2 and the female-specific regulator Tra. The SR protein binding to the repeats is RNA-binding protein 1 (RBP1), whereas *Drosophila* alternative splicing factor/splicing factor 2 (dASF/SF2) binds the PRE. Tra-dependent cooperative assembly ensures that Tra-2 and the SR proteins only bind stably in female cells.

ESE function[38]. Another possible mechanism involves the *Drosophila* SR protein RBP1, which binds both the polypyrimidine tract and the repeats, and could conceivably replace U2AF at this particular 3′ splice site[39]. Human homologues of Tra-2 have been identified and shown to mediate activation by specific enhancers[40].

Sequences identical to the *dsx* enhancer stimulate the use of a downstream 5′ splice site in the gene *fruitless* (Fig. 1a)[5]. This type of enhancer complex can therefore promote coordinated changes in alternative splicing of different genes, all triggered by the same developmental regulator Tra, but involving different sets of molecular interactions depending on its position within the pre-mRNA. Position can have even more dramatic effects: an adenovirus SR-protein-binding enhancer just upstream of a branch point causes splicing inhibition by blocking access of U2 snRNP. The same element acts as an ESE when located downstream from the 3′ splice site[41]. Interestingly, this inhibitory function of SR proteins depends upon their phosphorylation state. A viral protein relieves inhibition of the 3′ splice site late during infection by inducing dephosphorylation of the bound SR proteins[42]. Post-translational modification can therefore regulate splice-site choice, and indeed a number of kinases and activation of particular signal transduction pathways have been shown to influence alternative splicing[43–45].

Multiple combinatorial control

Even simple tissue-specific decisions can reveal additional layers of complexity, in which complexes assembling on different RNA elements synergize or antagonize each other. As mentioned earlier, the *Src* gene contains a neuron-specific 18-nucleotide exon. At least six different RNA sequences have separable effects in exon skipping and inclusion (Fig. 3b). The short exon length and binding sites for PTB flanking the exon are important for exon skipping in non-neural cells[32,46]. The *Src* ISE described in the previous section synergizes with exonic sequences for neural-specific inclusion. Several of these sequences contain redundant signals, as well as overlapping elements with opposite effects in different cell lines. Although the tissue specificity of each set of signals is not complete, the balance of their mutual influences creates a 20-fold preference for exon inclusion in neurons[46].

An even more complex array of interactions arises when different RNA processing events influence each other. The gene for calcitonin and calcitonin-gene-related peptide (CGRP) is alternatively spliced and polyadenylated (Fig. 1b). Most cells include exon 4 and introduce a polyA tail at its 3′ end, whereas neurons skip exon 4 and use a polyadenylation site downstream. An ISE present downstream of exon 4 resembles a 'zero-length' exon: a 3′ splice site immediately followed by a 5′ splice site. Binding of splicing factors to the ISE inhibits exon 4 polyadenylation, without the actual occurrence of a splicing event, and this, in turn, allows exon 4 skipping. Recognition of the element by PTB prevents the assembly of splicing factors, favouring exon 4 polyadenylation and inclusion[47]. Thus, the repressive action of PTB on the ISE leads to activation of exon inclusion through polyadenylation. Another example of interconnection between regulated splicing and other RNA processing events occurs in the gene encoding the editing enzyme ADAR2 (adenosine deaminase that acts on double-stranded RNA). An alternative, out-of-frame 3′ splice site is created within ADAR2 pre-mRNA by the editing activity of ADAR2 itself, thus switching off expression of the protein in what appears to be an autoregulatory loop[48].

Finally, modulated competition between splicing and nucleocytoplasmic transport pathways facilitates the production of differentially processed RNAs during retroviral lifecycles. For instance, whereas early HIV-1 cytoplasmic

transcripts are highly spliced, viral Rev protein facilitates the direct export of unspliced RNAs at later stages of infection, thus allowing expression of structural genes and new viral genomic RNA (Ref. 49).

The preceding examples show how assemblies of tissue-specific factors and proteins of the SR and hnRNP families can have positive or negative roles depending on their precise location, composition and state of modification of their components. The rationale for building up regulatory complexes rather than using dedicated, gene-specific single factors might be twofold. First, as with splicing signals themselves, regulatory sequences are often short and degenerate, and therefore specific recognition could arise from multiple relatively weak interactions with different components of the complexes. Second, cooperative combinatorial assembly allows sensitive responses to variations in the concentration of a single factor, at the same time restricting its effects to only those cells that can provide all the factors required to form the complex[15].

The elaborate networks between regulatory sequences, complexes and different processing events afford the possibility of combinatorial control. This allows exquisite modulation of splice-site choice based upon the expression of cell-specific factors and variations in relative levels or degree of modification of constitutive factors.

Future challenges

The realization that cell-specific patterns of processing are achieved by an elaborate interplay of signals and complexes has two important implications for future work. First, understanding regulated splicing will still require a detailed, gene-specific analysis of the contributions of multiple sequences and factors and their mutual influences in different cell types. Second, an overall picture of cell-specific splicing will require quantification of the relative levels of expression of entire families of factors, an analysis that will be facilitated by the use of DNA microarrays.

One would expect signal transduction cascades to have similarly extensive effects on splicing regulation as they have on transcription regulation, but our knowledge of this is still rudimentary. The targets of these cascades can be processing factors themselves or other factors influencing processing, including the CTD of RNA polymerase II (Ref. 18).

Changes in the availability of competing splice sites and regulatory complexes caused by different transcription rates provides a means by which alternative splicing can be coupled to transcription[50,51]. Coupling has gained a new level of sophistication in recent reports that show the influence of promoter structure on the pattern of alternative splicing of nascent transcripts[52]. The molecular basis for this observation could be the deposition of specific processing factors on the CTD at the promoter – a possibility supported by the finding that ASF/SF2 interacts with a transcriptional co-activator[53]. It will be interesting to determine how far downstream these promoter-determined effects persist in genes with multiple introns; is the CTD imprinted with a specific complement of SR proteins at the promoter? Is this influenced by the precise state of phosphorylation of the CTD repeats? A combination of detailed work on particular systems, genome-wide thinking and integration with other steps of gene expression and cellular function will ultimately be required to understand fully the regulation of alternative splicing, along with its important developmental and medical implications.

Acknowledgements

We thank Doug Black, Javier Cáceres, Tom Cooper, Benoit Chabot, Fátima Gebauer, Iain Mattaj, Jim Patton and members of our laboratories for comments on the manuscript. Because of space limitations, we have cited reviews and recent articles that include extensive references to each particular topic rather than original references. We apologize to those colleagues whose work has not been cited more directly. Work in the laboratory of C.W.J.S. is supported by grants from the Wellcome Trust, Medical Research Council and British Heart Foundation.

References

1 Black, D.L. (1998) Splicing in the inner ear: a familiar tune, but what are the instruments? *Neuron* 20, 165–168
2 Grabowski, P.J. (1998) Splicing regulation in neurons: tinkering with cell-specific control. *Cell* 92, 709–712
3 Jiang, Z-H. and Wu, J.Y. (1999) Alternative splicing and programmed cell death. *Proc. Soc. Exp. Biol. Med.* 220, 64–72
4 Chabot, B. (1996) Directing alternative splicing: cast and scenarios. *Trends Genet.* 12, 472–478
5 López, A.J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* 32, 279–305
6 Black, D.L. (1995) Finding splice sites within a wilderness of RNA. *RNA* 1, 763–771
7 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
8 Reed, R. (1996) Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr. Opin. Genet. Dev.* 6, 215–220
9 Burge, C.B. *et al.* (1999) Splicing of precursors to mRNAs by the spliceosomes. In *The RNA World* (2nd edn), pp. 525–560, Cold Spring Harbor Laboratory Press
10 Krämer, A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* 65, 367–409
11 Berglund, J.A. *et al.* (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89, 781–787
12 Moore, M.J. (2000) Intron recognition comes of age. *Nat. Struct. Biol.* 7, 14–16
13 Cáceres, J.F. and Krainer, A.R. (1997) Mammalian pre-mRNA splicing factors. In *Eukaryotic mRNA Processing* (Krainer, A.R., ed.), pp. 174–212, Oxford University Press
14 Tacke, R. and Manley, J.L. (1999) Determinants of SR protein specificity. *Curr. Opin. Cell Biol.* 11, 358–362
15 Hertel, K.J. *et al.* (1997) Common themes in the function of transcription and splicing enhancers. *Curr. Opin. Cell Biol.* 9, 350–357
16 Berget, S. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.* 270, 2411–2414
17 Lewis J.D. *et al.* (1995) The influence of 5′ and 3′ end structures on pre-mRNA metabolism. *J. Cell Sci.* (Suppl.) 19, 13–19
18 Bentley, D. (1999) Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr. Opin. Cell Biol.* 11, 347–351
19 McAfee, J.G. *et al.* (1997) The packaging of pre-mRNA. In *Eukaryotic mRNA Processing* (Krainer, A.R., ed.), pp. 68–102, Oxford University Press
20 Krecic, A.M. and Swanson, M.S. (1999) hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* 11, 363–371
21 Blanchette, M. and Chabot, B. (1999) Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* 18, 1939–1952
22 Cáceres, J.F. *et al.* (1998) A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev.* 12, 55–66
23 Eperon, I.C. *et al.* (1993) Pathways for selection of 5′ splice sites by U1 snRNPs and SF2/ASF. *EMBO J.* 9, 3607–3617
24 Del Gatto-Konczak, F. *et al.* (1999) hnRNP A1 recruited to an exon *in vivo* can function as an exon splicing silencer. *Mol. Cell. Biol.* 19, 251–260
25 Caputi, M. *et al.* (1999) hnRNP A/B proteins are required for inhibition of HIV-1 pre-mRNA splicing. *EMBO J.* 18, 4060–4067
26 Stickeler, E. *et al.* (1999) Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis. *Oncogene* 18, 3574–3582
27 Philips, A.V. *et al.* (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* 280, 737–741
28 Valcárcel, J. and Gebauer, F. (1997) Post-transcriptional regulation: the dawn of PTB. *Curr. Biol.* 7, R705–R708
29 Southby, J. *et al.* (1999) Polypyrimidine tract binding protein functions as a repressor to regulate alternative splicing of alpha-actinin mutally exclusive exons. *Mol. Cell. Biol.* 19, 26699–26711
30 Perez, I. *et al.* (1997) Mutation of PTB binding sites causes misregulation of alternative 3′ splice site selection *in vivo*. *RNA* 3, 764–778
31 Zhang, L. *et al.* (1999) Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB. *RNA* 5, 117–130
32 Chan, R.C. and Black, D.L. (1997) The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. *Mol. Cell Biol.* 17, 4667–4676
33 Ashiya, M. and Grabowski, P.J. (1997) A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. *RNA* 3, 996–1015
34 Adams, M.D. *et al.* (1996) Biochemistry and regulation of pre-mRNA splicing. *Curr. Opin. Cell Biol.* 8, 331–339
35 Min, H. *et al.* (1997) A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes Dev.* 11, 1023–1036
36 Lynch, K.W. and Maniatis, T. (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila* doublesex splicing enhancer. *Genes Dev.* 10, 2089–2101
37 Graveley, B.R. and Maniatis, T. (1998) Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing. *Mol. Cell* 1, 765–771

38 Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human diseases. *Trends Biochem. Sci.* 25, 106–110
39 Heinrichs, V. and Baker, B.C. (1995) The *Drosophila* SR protein RBP1 contributes to the regulation of doublesex alternative splicing by recognizing RBP1 RNA target sequences. *EMBO J.* 14, 3987–4000
40 Tacke, R. *et al.* (1998) Human Tra2 proteins are sequence-specific activators of pre-mRNA splicing. *Cell* 93, 139–148
41 Kanopka, A. *et al.* (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* 381, 535–538
42 Kanopka, A. *et al.* (1998) Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins. *Nature* 393, 185–187
43 Wang, J. and Manley, J.L. (1997) Regulation of pre-mRNA splicing in metazoa. *Curr. Opin. Genet. Dev.* 7, 205–211
44 Du, C. *et al.* (1998) Protein phosphorylation plays an essential role in the regulation of alternative splicing and sex determination in *Drosophila*. *Mol. Cell* 2, 741–750
45 König, H. *et al.* (1998) Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. *EMBO J.* 17, 2904–2913
46 Modafferi, E.F. and Black, D.L. (1999) Combinatorial control of a neuron-specific exon. *RNA* 5, 687–706
47 Lou, H. *et al.* (1999) Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3′-terminal exon. *Mol. Cell. Biol.* 19, 78–85
48 Rueter, S.M. *et al.* (1999) Regulation of alternative splicing by RNA editing. *Nature* 399, 75–80
49 Izaurralde, E. *et al.* (1999) Viruses, microorganisms and scientists meet the nuclear pore. *EMBO J.* 18, 289–296
50 Roche, S.E. *et al.* (1995) P-element repressor autoregulation involves germ-line transcriptional repression and reduction of third intron splicing. *Genes Dev.* 15, 1278–1288
51 Roberts, G.C. *et al.* (1998) Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res.* 26, 5568–5572
52 Cramer, P. *et al.* (1999) Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol. Cell* 4, 251–258
53 Ge, H. *et al.* (1998) A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. *Mol. Cell* 2, 751–759

Signaling networks linking integrins and Rho family GTPases

Martin A. Schwartz and Sanford J. Shattil

Integrins and Rho family GTPases function coordinately to mediate adhesion-dependent events in cells. Recently, it has also become apparent that integrins regulate Rho GTPases and vice versa. Integrins and GTPases might therefore be organized into complex signaling cascades that regulate cell behavior.

THE RHO FAMILY GTPases, particularly Rho, Rac and Cdc42, have emerged as central coordinators of a remarkable variety of cellular functions. These functions include cytoskeletal organization, gene expression, cell cycle progression, membrane trafficking, cell adhesion, migration and polarity. The GTPases exert their effects through a host of effectors that interact with the GTP-bound, but not GDP-bound, Rho proteins, and are thereby activated or localized to specific subcellular compartments. The GTPases themselves are regulated by guanine nucleotide exchange factors (GEFs) and guanine nucleotide dissociation inhibitors that modulate exchange of GTP for GDP, as well as by GTPase-activating proteins (GAPs) that promote GTP hydrolysis. Loading of Rho family proteins with GTP (i.e. activation) can be induced by a variety of mitogens, cytokines and biological stresses (reviewed in Refs 1,2).

M.A. Schwartz and **S.J. Shattil** are in the Dept of Vascular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA; S.J. Shattil is also in the Dept Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA.
Emails: schwartz@scripps.edu; shattil@scripps.edu

Integrins are αβ heterodimers that mediate adhesion of cells to extracellular matrix (ECM) proteins such as fibronectin, laminin or collagen, and in some cases to counter receptors on other cells such as intracellular cell-adhesion molecule-1 (ICAM-1) and vascular cell-adhesion molecule-1 (VCAM-1). Both the α and β subunits contain large extracellular ligand-binding domains, single-pass transmembrane domains and short cytoplasmic domains that bind cytoskeletal and signaling proteins. In mammalian cells, more than 20 different αβ heterodimers have been identified, which vary in ligand-binding specificity and signaling properties. Binding of integrins to ligands within the ECM or on other cells triggers an increase in lateral clustering as well as occupancy of integrin ligand-binding sites. Integrin occupancy and clustering initiate not only adhesion and cytoskeletal organization via direct physical associations of integrins with other proteins, but also activate many intracellular signaling pathways that regulate cell migration, polarity, survival, growth, differentiation and gene expression (reviewed in Refs 3,4). Pathways stimulated by integrin-mediated adhesion include protein tyrosine and serine/threonine kinases, lipid kinases and membrane transporters, as well as GTP-binding proteins.

The central thesis of this article is that integrins and Rho family GTPases are intimately connected at multiple levels. As noted, integrins and Rho family GTPases control many of the same cellular events. Furthermore, it appears that integrins regulate Rho family GTPases and Rho family GTPases regulate integrins. The known interactions between integrins and Rho, Rac and Cdc42 will be discussed, and the significance of these interactions for anchorage-dependent cell functions will be considered.

Integrin regulation of Rho family GTPases

Early studies reported that cells plated on ECM proteins rapidly developed extended filopodia and lamellipodia, indicative of Cdc42 and Rac activity, respectively (reviewed in Ref. 5; Fig. 1). These protrusive structures contain focal complexes, which are localized clusters of integrins, and cytoskeletal and signaling proteins that are smaller than, but similar in composition to, focal adhesions and are characteristically found in contact with the ECM. As cells spread, larger integrin-based focal adhesions and actin stress fibers form, indicative of Rho activity. However, early studies left unanswered whether integrins could trigger biochemical activation of these GTPases or merely serve as attachment sites for GTPase-dependent structures. In fibroblasts, the Rac and Cdc42 effector p21-activated kinase (PAK) is rapidly activated by cell adhesion to fibronectin or plating on immobilized anti-integrin antibodies, which function as agonists in that context[6]. As PAK activity is stimulated by GTP-bound

PII: S0968-0004(00)01605-4