

# UCLA

## UCLA Previously Published Works

### Title

Mechanisms of Alternative Pre-messenger RNA Splicing

### Permalink

<https://escholarship.org/uc/item/2hg605wm>

### Journal

Annual Reviews of Biochemistry, 72

### Author

Black, Douglas L

### Publication Date

2003

Peer reviewed

Posted, with permission, from the *Annual Reviews of Biochemistry*, Volume 72  
(c)2003, by Annual Reviews, [www.annualreviews.org](http://www.annualreviews.org).

## **Mechanisms of Alternative Pre-Messenger RNA Splicing**

**Douglas L. Black**

Dept. of Microbiology, Immunology, and Molecular Genetics,  
Howard Hughes Medical Institute,  
UCLA  
6-762 MacDonald Building  
Box 951662,  
675 Charles Young Dr. South,  
Los Angeles, CA 90095-1662

Telephone: (310) 794-7644

FAX: (310) 206-8623

Email: [dougb@microbio.lifesci.ucla.edu](mailto:dougb@microbio.lifesci.ucla.edu)

<http://www.mimg.ucla.edu/faculty/black/index.html>

## ABSTRACT

Alternative pre-mRNA splicing is a central mode of genetic regulation in higher eukaryotes. Variability in splicing patterns is a major source of protein diversity from the genome. In this review, I describe what is currently known of the molecular mechanisms that control changes in splice site choice. I start with the best-characterized systems from the *Drosophila* Sex Determination pathway, and then describe the regulators of other systems about whose mechanisms there is some data. How these regulators are combined into complex systems of tissue specific splicing is discussed. In conclusion, very recent studies are presented that point to new directions for understanding alternative splicing and its mechanisms.

Key Words: Alternative Splicing, Regulatory Mechanisms, Protein Diversity, RNA Binding Proteins, Spliceosome.

## INTRODUCTION

The splicing reaction that assembles eukaryotic mRNAs from their much longer precursors provides a uniquely versatile means of genetic regulation. Alterations in splice site choice can have many different effects on the mRNA and protein products of a gene. Commonly, alternative splicing patterns determine the inclusion of a portion of coding sequence in the mRNA, giving rise to protein isoforms that differ in their peptide sequence and hence chemical and biological activity (1). Alternative splicing is a major contributor to protein diversity in metazoan organisms. Estimates of the minimum number of human gene products that undergo alternative splicing are as high as 60% (2). Moreover, many gene transcripts have multiple splicing patterns and some have thousands (3) (4). To understand this complexity of gene expression, we must study how changes in splice site choice come about.

In a typical multi-exon mRNA, the splicing pattern can be altered in many ways (Figure 1). Most exons are constitutive; they are always spliced or included in the final mRNA. A regulated exon that is sometimes included and sometimes excluded from the

mRNA is called a cassette exon. A special case of this is where multiple cassette exons are mutually exclusive - producing mRNAs that always include one of several possible exon choices but no more. In these systems, there must be special mechanisms that enforce the exclusive choice (5) (6). Exons can also be lengthened or shortened by altering the position of one of their splice sites. One sees both alternative 5' and alternative 3' splice sites. The 5' terminal exons of an mRNA can be switched through the use of alternative promoters and alternative splicing. Similarly, the 3' terminal exons can be switched by combining alternative splicing with alternative polyadenylation sites. Alternative promoters are primarily an issue of transcriptional control. Control of polyadenylation appears mechanistically similar to control of splicing, although it will not be discussed here (7). Finally, there are important regulatory events controlled by the failure to remove an intron, a splicing pattern called intron retention. Particular pre-mRNAs often have multiple positions of alternative splicing, giving rise to a family of related proteins from a single gene (Figure 1H).

Changes in splice site choice can have all manner of effects on the encoded protein. Small changes in peptide sequence can alter ligand binding, enzymatic activity, allosteric regulation, or protein localization. In other genes, the synthesis of a whole polypeptide, or a large domain within it, can depend on a particular splicing pattern. Genetic switches based on alternative splicing are important in many cellular and developmental processes, including sex determination, apoptosis, axon guidance, cell excitation and contraction, and many others. Errors in splicing regulation have been implicated in a number of different disease states. The roles played by alternative splicing in particular cellular processes and in disease have been reviewed extensively(1, 8-14) (15-18) (19, 20). Here we will focus on the common mechanistic features of a number of well-studied model systems.

The excision of the introns from a pre-mRNA and the joining of the exons is directed by special sequences at the intron/exon junctions called splice sites (21). The 5' splice site marks the Exon/Intron junction at the 5' end of the intron (Figure 2A). This includes a GU dinucleotide at the intron end encompassed within a larger, less conserved consensus sequence (21). At the other end of the intron, the 3' splice site region has three conserved sequence elements: the branchpoint, followed by a polypyrimidine tract,

followed by a terminal AG at the extreme 3' end of the intron. Splicing is carried out by the spliceosome, a large macromolecular complex that assembles onto these sequences, and catalyzes the two transesterification steps of the splicing reaction (Figure 2AB). In the first step, the 2' hydroxyl group of a special A residue at the Branchpoint attacks the phosphate at the 5' splice site. This leads to cleavage of the 5' exon from the intron and the concerted ligation of the intron 5' end to the Branchpoint 2' hydroxyl. This step produces two reaction intermediates, a detached 5' exon and an intron/3' exon fragment in a lariat configuration containing a branched A nucleotide at the branchpoint. The second transesterification step is the attack of the phosphate at the 3' end of the intron by the 3' hydroxyl of the detached exon. This ligates the two exons and releases the intron, still in the form of a lariat (Figure 2).

The spliceosome assembles onto each intron from a set of five small nuclear ribonucleoproteins (snRNPs) and numerous accessory proteins (Figure 2B) (21-23). During assembly, the U1 snRNP binds to the 5' splice site via base pairing between the splice site and the U1 snRNA. The 3' splice site elements are bound by a special set of proteins. SF1 is a branchpoint binding protein (also called BBP in Yeast). The 65 kilodalton subunit of the dimeric U2 Auxiliary Factor (U2AF) binds to the polypyrimidine tract. In at least some cases, the 35 kD subunit of U2AF binds to the AG at the intron/exon junction. The earliest defined complex in spliceosome assembly, called the E (early) or commitment complex, contains U1 and U2AF bound at the two intron ends (24). The E complex is joined by the U2 snRNP, whose snRNA base pairs at the branch point, to form the A complex. The A complex is joined by the U4/U5/U6 tri-snRNP to form the B complex. The B complex undergoes a complex rearrangement to form the C complex where the U1 snRNP interaction at the 5' splice site is replaced with the U6 snRNP and the U1 and U4 snRNPs are lost from the complex. It is the C complex that catalyzes the 2 chemical steps of splicing (Figure 2B). There is also a minor class of spliceosome that excises a small family of introns that use different consensus sequences (25). Only the major class spliceosome will be discussed here.

Changes in splice site choice arise from changes in the assembly of the spliceosome. In most systems, splice site choice is thought to be regulated by altering the binding of the initial factors to the pre-mRNA and the formation of early spliceosome

complexes. By the time the E complex is formed, it appears that the splice sites are paired in a functional sense and the defined intron is committed to being spliced. However, this need not always be the case and there is evidence in one system for the regulation of 3' splice site choice after the first catalytic step of splicing (i.e. after branch formation (see below) (26)).

The splice site consensus sequences are generally not sufficient information to determine whether a site will assemble a spliceosome and function in splicing. Other information and interactions are necessary to activate their use (27, 28). Introns can range in size from less than 100 nucleotides to hundreds of thousands of nucleotides. In contrast, exons are generally short and have a fairly narrow size distribution of 50 - 300 nucleotides. Commonly, spliceosomal components binding on opposite sides of an exon can interact to stimulate excision of the flanking introns (29). This process is called exon definition and apparently occurs in most internal exons (30). On top of this process, there are many non-splice site regulatory sequences that strongly affect spliceosome assembly. RNA elements that act positively to stimulate spliceosome assembly are called splicing enhancers. Exonic splicing enhancers are commonly found even in constitutive exons. Intronic enhancers also occur and appear to differ from exonic enhancers. Conversely, other RNA sequences act as splicing silencers or repressors to block spliceosome assembly and certain splicing choices. Again, these silencers have both exonic and intronic varieties. Some regulatory sequences create an RNA secondary structure that affects splice site recognition (31, 32) (33), but most seem to be protein binding sites. This review focuses on the non-spliceosomal pre-mRNA binding proteins that act through these splicing regulatory sequences.

## **THE SEX DETERMINATION GENES OF *DROSOPHILA* AS MODELS FOR SPLICING REGULATION**

### **Sex Lethal Protein is a Splicing Repressor.**

By far the best understood systems of splicing regulation come from the pathway of somatic sex determination in *Drosophila melanogaster*. A series of remarkable genetic studies identified the key regulators of the Sex Determination pathway as RNA binding

proteins that alter the splicing of particular transcripts (8, 18, 34). These studies provided an essential starting point for the biochemical analysis of splicing regulation (Figure 3).

The master regulatory gene at the top of the sex determination pathway encodes the RNA binding protein Sex Lethal. Sxl Protein is expressed specifically in female flies where it represses splicing patterns that would lead to male development. In the presence of Sxl, female splicing patterns are expressed, leading to gene products needed for female development. Sex-lethal contains two RNA binding domains (RBD) of the RNP consensus type (RNP-cs, also called an RRM) (Figure 4) (35) (36). In the crystal structure of the Sxl protein bound to RNA, the RNA interacting surfaces of these domains face each other to form a cleft that specifically interacts with a nine nucleotide U-rich element found in the Sxl target RNAs (37, 38). An additional N-terminal Glycine rich domain in Sxl influences the cooperative assembly of the protein onto multiple binding sites (39).

The downstream targets of the Sxl protein include transcripts from the Transformer (Tra) and Male-Specific Lethal 2 (Msl 2) genes (Figure 3AB). The Tra gene also encodes a splicing regulator. In the absence of Sxl, a splicing pattern is used that produces a truncated and inactive Tra protein. In female flies, where Sxl is present, Sxl binds to its recognition element within the 3' splice site of Tra exon 2, blocking recognition by U2AF (Figure 3A)(40) (41) (42, 43). This causes a shift in the 3' splice site to a position downstream, thus deleting a stop codon from the Tra mRNA and allowing translation of active Tra protein. This appears to be the simplest mechanism for altering a splicing pattern, where the Sxl protein directly competes with an essential splicing factor (U2AF) for its RNA binding site. However, Sxl regulation of Tra requires two other Drosophila genes, virilizer and Female-specific lethal 2D, so it is likely that there is more to this mechanism than is currently understood (44, 45).

A second target of Sxl is the gene MSL2, which regulates X Chromosome Dosage compensation in male flies (46, 47) (48) (Figure 3B). In females, Sxl binds to two sites in the first intron of Msl2 pre-mRNA: one in the polypyrimidine tract and one adjacent to the 5' splice site. Sxl binding to the polypyrimidine tract again blocks U2AF binding (49). Sxl binding near the 5' splice site blocks a regulatory factor called TIA-1 (see below) and binding of the U1 snRNP to the 5' splice site (50, 51). Both Sxl binding sites

are needed for the inhibition of splicing and retention of the first Msl2 intron in the final mRNA. Interestingly, the retained Msl2 intron is in the 5' UTR of the transcript and does not affect its open reading frame. It has been shown that the Sxl bound within this region blocks translation of the transcript in the cytoplasm (52) (53) (54). Thus, Sxl is affecting both the splicing of the transcript in the nucleus and its translation in the cytoplasm. There appear to be other examples in animal cells of predominantly nuclear splicing regulators having additional cytoplasmic functions (55) (56-58).

In addition to regulating transcripts downstream in the sex determination pathway, Sxl also autoregulates its own splicing to maintain the female splicing phenotype (8, 18, 34) (59) (Figure 3C). In male flies, the Sxl gene is transcribed but the inclusion of Sxl exon 3 introduces a premature stop codon to produce a truncated and inactive protein. In female flies, the same Sxl promoter is active, producing the same RNA precursor. However, in the early female embryo an additional *sxl* promoter is briefly active. This early promoter produces a transcript missing exon 3 that is translated into active Sxl protein (60) (61). This embryonic Sxl protein initiates the whole splicing cascade by repressing the splicing of *sxl* exon 3 from the constitutive promoter. The mechanism of *sxl* autoregulation is quite different from its activity on Tra and Msl2 (39, 62-64). Multiple Sxl binding sites flank exon 3 and are required for its repression. Sxl protein assembles cooperatively onto these sites and its Glycine rich Amino terminus is required for this cooperative binding. In this case, rather than blocking the binding of a particular factor, Sxl is creating an RNP structure that encompasses the whole repressed exon. How this structure actually represses splicing is the subject of some interest. The repression requires the Sans fille protein, the Drosophila homolog of the U1A and U2 B'' proteins, which may indicate interactions of Sxl with the snRNP's assembling on both sides of the exon (65, 66).

A recent study has given new insight into the mechanism of Sxl autoregulation (26, 67). Sxl exon 3 has an unusual 3' splice site containing two AG dinucleotides that flank the polypyrimidine tract (Figure 3C). The branchpoint is upstream from both of these AG dinucleotides. In the absence of Sxl, the polypyrimidine tract is bound by U2AF65, and the downstream AG by U2AF35. Interestingly, this downstream AG is required for the splicing of the exon but is not the site of exon ligation, which occurs at

the upstream AG (63). This upstream AG is bound by a factor called SPF45 and in the absence of SPF45 splicing occurs at the downstream AG (26). Thus, for this exon the role of U2AF in promoting spliceosome assembly at the 3' splice site has been separated from a role in defining the AG for exon ligation. Both in vivo and in vitro, the repression of the exon by Sxl requires SPF45. Most interestingly, in a model pre-mRNA carrying the Sxl 3' splice site, the first step of splicing (5' exon cleavage and lariat formation) occurs in vitro in the presence of Sxl, but the second step (3' splice site cleavage and exon ligation) is blocked. Thus, at least in this in vitro system, Sxl inhibits the second step of splicing but not the first step. Although it needs to be confirmed in vivo on a standard Sxl RNA, this result seems to imply that a spliceosome can be drastically rearranged after the first step of splicing and redirected to a distant 3' splice site. It will be interesting to find out if 3' splice site choice can commonly occur so late in the spliceosome assembly pathway, and if this will always require SPF45 and a similar arrangement of AG dinucleotides.

### **Transformer is a Splicing Activator**

Sex-lethal regulates the splicing of the Transformer transcript to produce the female specific Tra protein that in turn regulates the splicing of the Doublesex (dsx) and Fruitless (fru) mRNAs in the sex determination pathway (8, 18, 34). Unlike Sxl, Tra is a positive regulator of splicing that activates female specific splicing patterns in its targets (Figure 3DE).

Tra contains an extended RS domain, rich in Arginine/Serine dipeptides (Figure 4). These domains are a common feature of splicing regulatory proteins (see below). The Tra protein does not have an RNA binding domain and must cooperatively assemble with other proteins onto its target RNA sequences (68-70). In male flies, Tra is absent and mRNAs of the dsx gene are spliced from exon 3 to exon 5 to encode a transcription factor that determines male differentiation (71) (72). Conversely, the presence of Tra in females activates the splicing of exon 3 to exon 4 and results in an mRNA that encodes a transcriptional regulator leading to female differentiation. Dsx exon 4 carries an unusually weak 3' splice site that is not normally recognized by the U2AF protein. Within this exon is a splicing enhancer sequence whose activity is controlled by the

presence of Tra protein (73-77). This enhancer has a series of six 13-nucleotide repeat elements and an additional Purine rich element. The repeat element is bound by a trio of proteins including Tra, an SR protein RBP1, and an SR related protein called Tra-2 (76). SR Proteins comprise an important family of splicing regulators discussed in detail below. Tra-2 is not sex specific but is required for Tra dependent splicing regulation. Unlike Tra, both RBP1 and Tra-2 contain RNA binding domains and directly interact with the repeat sequence. Tra is required for Tra-2 binding to the repeat and the cooperative assembly of the complex. Interestingly, the Purine-rich element also binds Tra and Tra-2, but in this case they cooperate with a different SR protein: dSRp30 (76). Via its interactions with Tra and two different SR proteins, Tra-2 is induced to bind to two different regulatory sequences. The assembly of these complexes onto the *dsx* splicing enhancer is thought to stabilize U2AF binding to the upstream 3' splice site, leading to spliceosome assembly and splicing at exon 4 ((78) See also below and references (79) (80) for a discussion of this issue). Tra and its assembly with SR proteins and Tra-2 is the major paradigm for how exonic splicing enhancers work. However, the nature of the interactions that lead to U2AF binding are not completely clear and there is much to be learned about how splicing enhancement actually takes place.

Tra also positively regulates the splicing of the Fruitless gene transcript (Figure 3E)(81, 82) (83). An internal exon of *Fru* has a pair of alternative 5' splice sites. In male flies, the upstream 5' splice site is used. This gene product, encoding a BTB-ZF transcription factor, goes on to regulate male courtship behavior and other aspects of male sexual development. In females, Tra and Tra-2 activate the use of a downstream 5' splice site leading to a different mRNA. Like *dsx*, this activation of the female 5' splice site also appears to involve RBP1 and requires several copies of the same repeated element found in the *dsx* enhancer. However, here the repeats are just upstream of the activated 5' splice site. Thus, Tra and its cofactors Tra-2 and RBP1 can act through exonic splicing enhancers to stimulate splicing at either a 3' splice site or a 5' splice site.

## **COMPLEX SYSTEMS OF TISSUE-SPECIFIC SPLICING**

The *Drosophila* Sex Determination pathway has provided the central examples of how a choice between two possible splicing patterns is regulated. In addition to alterations by sex, metazoan organisms regulate the splicing of thousands of other transcripts depending on cell type, developmental state or external stimulus. The analysis of these systems, mostly in mammalian cells, has identified many of the same kinds of proteins seen in the Fly Sex Determination pathway (34, 69, 70). However, these proteins are often combined in complex ways into multiple layers of regulation. Many of these proteins can act either positively or negatively depending on their binding context. We will discuss individual proteins and regulatory elements according to their location in introns or exons and then describe the challenges to understanding how they are combined to give a precise pattern of regulation.

### **Exonic Regulatory Elements and the Proteins that Bind to Them.**

#### **Positive Regulation from Exons**

Exons often contain enhancer or silencer elements that affect their ability to be spliced. There are many exonic splicing enhancers (ESE's), similar to the *dsx* enhancer. These RNA regulatory elements are diverse in sequence and often embedded within nucleotides that also code for protein. Such enhancers have been identified through exon mutations that block splicing, through computational comparisons of exon sequences, and through the selection of sequences that activate splicing or that bind to splicing regulatory proteins - most notably the SR proteins (12, 20, 68, 84) (85).

The SR Proteins constitute the best-studied family of splicing regulators (80, 86, 87). The SR proteins have a common domain structure of one or two RNP-cs RNA binding domains followed by what is called an RS domain containing repeated Arginine/Serine dipeptides (Figure 4). The Serines in an RS domain can be highly phosphorylated. The family includes the proteins SRp20, SRp30c, 9G8, SRp40, SRp55, SRp70, and the best studied members ASF/SF2, and SC35 (80, 88). The SR family is also defined by particular properties in splicing (87). There are additional proteins in the splicing reaction that have RS domains but which serve different roles. These include U2AF, U1 70K, SRm160/300, Tra-2 and numerous others. Some of these SR related proteins are

general splicing factors and some are apparently specific inhibitors of splicing (89-91) (92) .

The true SR proteins have a wide range of activities in the splicing reaction (80). Splicing in cellular extracts requires the presence of at least one member of the family and the different members are generally interchangeable in their ability to fulfill this requirement. However, the proteins are not always interchangeable in vivo (93-96). A separate activity of these proteins that we focus on here is the activation of splicing through exonic splicing enhancers. This activity was first seen in the *dsx* enhancer as described above. It has become clear that many if not all exons, constitutive or regulated, contain ESE elements that bind to specific members of the SR family (12). The presence of these elements seems to be a general mechanism for defining exons.

Most naturally occurring ESE's have been shown to bind specific SR proteins (12, 20). The most commonly studied are purine-rich sequences sometimes given the consensus (GAR)<sub>n</sub> that are bound by the proteins ASF/SF2 and Tra-2 (97, 98). In addition to naturally occurring enhancers, in vitro selection has identified many optimal binding sequences for different SR family members. The binding sites for a given family member can be fairly degenerate. For example, only some of the selected sequences for ASF/SF2 are purine rich and show similarity to the GAR element (97). When these optimal sites are introduced into model splicing substrates, they act as SR protein dependent splicing enhancers. The families of SR protein recognition sequences generated through in vitro selection experiments have proven to be an effective means of predicting the location of ESE's in natural exons (12, 97, 99).

It is not clear whether all whether the activity of all ESE's is SR protein dependent. A recent computational analysis identified a number of elements with ESE activity (85). Similarly, functional selection of sequences that activate splicing has identified additional elements (100-103) (84). Some of these are known to be SR protein dependent and it will be interesting to see if any are not. An AC rich enhancer element is apparently mediated by two non-SR proteins YB-1 and p72. P72 is a member of the DEAD Box RNA helicase family, raising interesting questions about the mechanisms of splicing activation (104, 105).

The two domains of a SR protein are modular in function. The RBD targets the protein to a particular exonic element, with the different RBD's targeting different sequences (106-109). This RNPcs domain can be replaced with the RNA binding MS2 coat protein from the MS2 bacteriophage. When the MS2 Coat Protein binding site was placed in an enhancer dependent exon, an MS2/RS domain fusion protein activated splicing in vitro (110). Thus, just tethering the RS domain to the exon can activate splicing. The RS domains themselves are largely interchangeable (106, 111). One can switch the RS domains of different proteins and they will maintain their activity, both in vitro and in vivo. Moreover, a natural RS domain can be replaced with a synthetic sequence of 10 RS dipeptides and the protein will maintain its activity in at least some assays (112). RS domains have several proposed functions (80). The unphosphorylated domain is highly positively charged and may act as a counter ion to enhance protein affinity for RNA or RNA/RNA hybridization. However, phosphorylation is required for activity in splicing (113-115) (116-118), and there is clear evidence for phosphorylated RS domains engaging in protein/protein interactions (119). In biochemical experiments, ASF/SF2 was shown to interact with the RS domain containing U1 snRNP protein, U1 70K (120). Yeast two hybrid experiments showed that the RS domain could act as a protein/protein interaction domain for binding to other SR Proteins (121). It was further shown that this interaction in yeast required the presence of the SR protein kinase and presumably phosphorylation (122). Since RS domains are largely interchangeable in these assays, it may be that any RS domain can bind to any other. However, it has not been shown that two RS domains directly contact each other in these interactions rather than contacting other protein domains. Indeed, other tests of SR protein function both in vivo and in vitro indicate that there is specificity to these protein/protein interactions that maybe determined by contacts outside of the RS domain (123) (124) (111).

There are many questions about the mechanisms of SR protein action. As described above, an exonic enhancer can stimulate U2AF or U2 binding to weak 3' splice sites, and U1 snRNP binding to 5' splice sites (78, 125) (126) (127, 128) (129) (120) (79) (Figure 5). However, in naturally occurring enhancers, SR proteins apparently bind as a component of a large, multiprotein complex, as seen with dsx. The components of these exon complexes are not all identified, but can include such known splicing factors as the

large RS domain proteins SRm160/300, the mammalian homolog of Tra-2, the U1 snRNP, and the hnRNP's A1 and H (see below). Thus, the particular interactions of the SR protein that activate splicing are not entirely clear. Splicing activation by the dsx enhancer in vitro requires the RS domain on U2AF 35 (78). However, this domain is not required for splicing regulation in vivo, indicating that additional interactions take place between U2AF and the enhancer complex (130) (79). Splicing enhancement by at least some ESE's requires SRm160 and 300, two large splicing factors that also contain RS domains (131, 132) (133). The enhancer complex thus contains multiple RS domains, and which factors directly interact is not fully known. Careful kinetic analyses of the rate of splicing activated by repeated ESE's indicate that only one SR protein complex at a time can interact with the spliceosome at the 3' splice site (134). More structural data regarding the protein/protein interactions in these complexes is needed to develop more precise models of ESE function. Finally, in some systems, enhancer bound SR proteins may stimulate splicing by counteracting repressor molecules rather than enhancing U2AF binding (135, 136).

One of the most interesting aspects of SR protein function is their phosphorylation. Both hyper and hypo phosphorylation of SR proteins seems to inhibit their activity in vitro (113-115). In several biological contexts, the phosphorylation of SR proteins correlates with their activity (137, 138). Several kinases have been identified that phosphorylate SR proteins. The SR protein kinases (SRPK) 1 and 2 are conserved from humans to yeast and the activity of these kinases alter SR protein localization and protein/protein interactions (139-141) (142, 143) (144, 145). Clk/Sty is another interesting group of kinases that phosphorylate SR proteins. They show a different pattern of preferred phosphorylation sites than the SRK's and affect in vitro splicing activity, protein localization and U1 70K binding by ASF/SF2 (113, 119, 146, 147). Genetic studies of a *Drosophila* homolog of Clk/Sty called Doa clearly implicate the kinase in splicing regulation (148). Doa phosphorylates the *Drosophila* SR protein RBP1 as well as Tra and Tra-2. Doa mutations block enhancer dependent splicing of dsx exon 4. Interestingly, the splicing of Fruitless is not affected by Doa mutations. Thus, splicing events that use common regulators have different phosphorylation dependence for those regulators. As with other responses to signaling pathways, a widely expressed protein like

RBP1 or Tra-2 might control very precisely regulated splicing through its own precise modifications.

### **Negative Regulation in Exons**

In opposition to the positive effects of exonic enhancers, exonic silencer or repressor elements have been identified. The best characterized of these are bound by particular hnRNP proteins. The hnRNP proteins are a large group of molecules identified by their association with unspliced mRNA precursors (hnRNA), and are not a single family of related proteins (149).

The most studied of these proteins, HnRNP A1, has been implicated in several processes including splicing and, oddly, the maintenance of telomere length (150) (151, 152). HnRNP A1 contains two RNP-cs RNA binding domains and a Glycine rich auxiliary domain (Figure 4). It belongs to a family of related proteins arising from both multiple genes and alternative splicing (153) (154, 155). A crystal structure of the two RBD's of A1 bound to the telomeric repeat DNA d(TTAGGG)<sub>2</sub> shows that an A1 protein dimer forms a four RNP domain surface, which binds 2 DNAs or 4 copies of the element (156). The topology of each DNA strand extends from the N terminal RBD of one dimer subunit to the C terminal RBD of the other subunit. If similar interactions occur with RNA elements in the pre-mRNA, this structure would allow cooperative binding to multiple splicing silencer elements and looping of the RNA between binding elements. This A1 structure is also remarkably different from the double-RNP domain structures of Sxl and Nucleolin (38, 157). It seems that the many multi-RNP domain proteins will vary considerably in their RNA binding, oligomerization, and the orientation of their RNP-cs domains.

A1 was originally implicated in splicing as a factor that counteracted SR proteins in a splice site shifting assay in vitro (158, 159). Several transcripts have splicing patterns that are sensitive to the relative ratio of A1 to ASF/SF2, which changes between different tissues (160, 161). A1 has since been shown to bind to exonic splicing silencers (ESS) in the HIV, FGFR2 and other transcripts (162) (163). The protein is required for the silencing effects of these sequences in vitro. The silencing can be recapitulated by tethering just the Glycine rich domain to the exon via an MS2 fusion (163). Several

mechanisms have been proposed for A1 mediated splicing repression and its mechanism may differ between different transcripts (Figure 6). A1 could interfere directly with the assembly of spliceosomal components, it could block the exon bridging interactions that occur during exon definition, or it could block splicing activation by SR proteins binding to adjacent ESE's. There is evidence for all of these activities. A1 dependent silencer elements have also been found in introns (see below).

Several groups have studied the role of A1 in repressing exons in the HIV precursor RNA. The mechanistic analysis is most advanced for Tat exon 3. This exon contains several enhancer elements that bind the SR proteins SF2/ASF and SC35 (164) (136) (165-167). There are also A1 binding sites both within the exon and also adjacent to a site of branch formation upstream. In studies of this exon, the Krainer lab focused on the exonic elements that bind A1, SC35 and SF2/ASF (136). They used an S100 extract that is depleted for all the SR proteins and further depleted it for A1. In the absence of A1, either SC35 or SF2/ASF can activate Tat exon 3 splicing. The addition of A1 specifically inhibited SC35 but not SF2/ASF activated splicing. It was found that in addition to binding the ESS, A1 crosslinked to exonic RNA distal to the ESS in the region of the SR protein binding sites. This crosslinking occurred only if the ESS was present. It was proposed that specific A1 binding at the ESS nucleates the cooperative non-specific binding of A1 upstream. This higher order complex of multiple A1's would create a zone of inhibition along the RNA. SF2/ASF is proposed to block the propagation of this complex, whereas SC35 does not bind tightly enough to do this (136) (Figure 6A). This model is appealing because it provides a rationale for two long known modes of RNA binding by A1: specific binding to particular short RNA elements and non-sequence-specific binding to longer RNAs. On the other hand, this crosslinking experiment could also be interpreted to support cooperative binding to separated specific sites, as has been proposed for A1 autoregulation (see below). There are additional effects of A1 in inhibiting HIV Tat splicing through blocking U2 assembly (168)(see below). Additional experiments are needed to create a model that incorporates exonic A1 effects on SR protein activity and effects from other binding sites.

In addition to A1, other exonic splicing silencers have been shown to bind hnRNP H and its close relative hnRNP F (169) (170). H and F are a different structural family

from A1 and contain three RNP-cs type RNA binding domains that recognize G rich elements (171) (172). Interestingly, hnRNP H acts as a splicing repressor when bound to an ESS in  $\beta$ -tropomyosin, but as an activator when bound to a similar element in HIV Tat Exon 2 (170) (173). It also binds to splicing regulatory elements in introns.

### **Intronic Regulatory Elements**

Many splicing regulatory sequences are present in introns rather than exons. Binding sites for regulators are often found within the polypyrimidine tract or immediately adjacent to the Branchpoint or 5' splice site. However, splicing regulatory elements can also act from a distance, being found hundreds of nucleotides away from the regulated exon. Like exonic regulation, there are both positive and negative acting sequences comprising Intronic Splicing Enhancers and Silencers (ISE or ISS). Also like the exonic elements, groups of elements are often found clustered to make composite regulatory sequences that mediate both positive and negative regulation. These regulatory regions can be highly conserved between species, and are often identified by sequence alignments.

### **Positive Regulation from Introns**

There are several elements known to act as ISE's, but the proteins that mediate their effects are less well characterized than for ESE's. Some ISE's do appear to be SR protein dependent (174) (175). In other cases, SR proteins do not appear to directly bind to the regulatory element, and at least some ISE's appear to be mechanistically different from ESE sequences.

Some 5' splice sites are activated for splicing by a Uridine rich sequence immediately downstream. The regulated intron of *Drosophila* Msl-2 requires this element for splicing in heterologous HeLa cell extracts (50). Similarly, the K-SAM exon in the FGFR 2 transcript is activated by a U rich element, called IAS1, immediately adjacent to the K-SAM 5' splice site (176). In both cases, this element was found to bind the protein TIA-1 (Figures 4 and 5). Extracts depleted for TIA-1 are inhibited for Msl2 splicing. Overexpression of TIA-1 induces K-SAM splicing in transfected cells. Significantly, TIA-1 stimulates U1 snRNP binding to 5' splice sites that are dependent on

the U rich element for function (50). So far, this is the only intronic enhancer protein shown to directly effect spliceosome assembly. TIA-1 and its relative TIAR contain 3 RNP-cs domains and a C-terminal Prion-like aggregation domain (Figure 4).

Interestingly, the yeast homolog of TIA-1, NAM8, is required for the function of certain regulated 5' splice sites (177, 178). However, Yeast NAM8 is a component protein of the U1 snRNP, whereas in animal cells TIA-1 appears to mainly exist as a separate factor. TIA-1 and TIAR were discovered as regulators of apoptosis, and as components of the cytoplasmic mRNA granules formed in response to cellular stress (179). It will be very interesting to find a relationship between these various functions.

The CUGBP and ETR-Like Factors (CELF) are a large family of proteins generated from both multiple genes and alternative splicing (180) (181). These proteins have been implicated in many different aspects of RNA metabolism and been given many different names, including the Bruno-like factors in *Drosophila* and the *Drosophila* Splicing Factor ELAV (182-184) (185) (Figure 4). At least some members of this family activate splicing through intronic enhancer elements. One target of these proteins is chicken Cardiac Troponin T (TnT) (180) (186). TnT exon 5 is spliced in embryonic muscle but excluded from the TnT mRNA in the adult. This exon has been extensively analyzed and is controlled by a complex set of regulatory elements in both the exon and the flanking introns. Some of the positive intronic elements, called Muscle Specific Enhancers (MSE's), contain UG elements that bind to CELF family members, including ETR3 (180) (Figure 5). The overexpression of ETR3 in transfected cells, or the addition of ETR3 to in vitro splicing extracts, increases TNT exon 5 splicing in an MSE dependent manner. Interestingly, the loss of exon 5 inclusion during muscle development coincides with a change in the expressed isoforms of ETR3, which may be a key factor in determining exon 5 splicing during development (180).

There are a large number of CELF family members and the effects of these proteins is not limited to muscle cells. The NAPOR1 protein is a splice variant of ETR3 that is enriched in portions of the nervous system (187, 188). Overexpression of NAPOR in cells has divergent effects on two NMDA Receptor 1 exons (188). NR1 exon 5 splicing is decreased by NAPOR, whereas Exon 21 is increased. Again this effect requires particular intronic elements. Interestingly, the relative inclusion of exons 5 and

21 strongly correlates in vivo with expression of NAPOR. In forebrain, where NAPOR is highly expressed, NR1 exon 5 is mostly excluded from the mRNA and exon 21 is included. The reverse occurs in cerebellum, where NAPOR is low, exon 5 is included and exon 21 is excluded. Thus, whether this protein is a positive acting factor or a negative one depends on the exon and cellular context. The CELF proteins as a family are widely expressed and apparently take part in a wide variety of cellular activities. Models for the function of these proteins in splicing and their effects on spliceosome assembly will await better biochemical assays.

Another common ISE element is the hexanucleotide UGCAUG. This element, particularly when duplicated, strongly enhances splicing (189) (190) (191). The element is known to enhance the splicing of exons in the c-src, Fibronectin, Non-muscle Myosin Heavy Chain (NMHC), and Calcitonin transcripts (192) (190) (189, 193, 194) (195). For several enhancers, mutagenesis analyses indicate this to be the key element in their activity. It is usually found repeated several times downstream of the activated exon and these repeats can be found at some distance from the exon (>500 nt.) (191). UGCAUG was also shown to be the most common hexanucleotide downstream of a set of neuronally regulated exons (196). Although associated with tissue specific exons, the enhancement effect of the element is not necessarily tissue specific (189) (190). The protein that mediates the effects of the UGCAUG hexanucleotide has not been identified definitively.

Intronic enhancers are often made up of intricate combinations of positive and negative elements that assemble into large RNP complexes. The enhancer downstream of the neural specific N1 exon in c-src is an example of such a structure (190, 192, 197)(Figure 7). The most conserved portion of the enhancer, called the downstream control sequence (DCS), contains a GGGGG element that is needed for full enhancer activity, a CUCUCU element that is required for splicing repression, and the UGCAUG that is essential for enhancer activity (190, 198). In its native context, the DCS needs additional surrounding elements for enhancer activity. Alternatively, two copies of the DCS by itself, or three copies of the UGCAUG hexanucleotide are sufficient for strong splicing enhancement. The DCS assembles a large RNP complex that has been analyzed in some detail (172) (199-201) (202). The GGGGG element binds to hnRNP's H and F.

The CUCUCU element binds to the Polypyrimidine Tract Binding Protein or its neuronal homolog (PTB/nPTB; see below). The UGCAUG binds to the KH-type Splicing Regulatory Protein (KSRP) and at least one unidentified factor. RNA competition and immunodepletion or inhibition experiments indicate positive roles for hnRNP's F/H and KSRP in N1 exon splicing. However, it is not clear that these proteins are responsible for the strong enhancing effect of the UGCAUG.

### **Negative Regulation from Introns**

There are also intronic silencer elements. Some of these elements are bound by proteins that we have already discussed. SR proteins have been shown to bind to an intronic sequence near a branchpoint in the Adenovirus L1 mRNA (203). The binding of SR proteins to this sequence early in infection blocks the use of this 3' splice site and shifts splicing to an adjacent site. Late in infection, the cellular SR proteins become dephosphorylated and inactive in splicing repression, thus allowing the site to be used (114). Similarly an intronic binding site for SRp30c is inhibitory for the splicing of an hnRNP A1 exon (204). It appears that SR proteins can be either splicing activators or repressors, depending on where in the pre-mRNA they bind.

HnRNP A1 also binds to intronic elements as well as exonic (Figure 6BC). However unlike SR proteins, all of the A1 binding sites are inhibitory for splicing. Kjems and colleagues examined an A1 binding site adjacent to the branchpoint for HIV Tat exon 3, the same HIV exon analyzed for its A1-binding Exonic Silencer (see above) (168). As in the studies of the exonic elements, it was found that A1 is required for splicing repression. However in this system, full repression required both the exonic and the branchpoint elements. Interestingly, U2AF still binds to the polypyrimidine tract in an A1 repressed transcript but splicing is blocked at the assembly of U2 snRNP at the branchpoint. This may imply, at least for transcripts containing the branchpoint A1 binding site, that A1 is also acting at a later step than SR protein mediated splicing enhancement.

A1 also auto-regulates the splicing of its own transcript, but this appears to be by a different mechanism than seen with HIV Tat (205) (Figure 6C). A cassette exon (7b) is skipped in the production of A1, but included in the mRNA for A1b. This exon is

regulated through A1 binding sites in the introns surrounding it. Using a truncated RNA to model the splicing shift in vitro, it was shown that competing away A1 protein activated 7b splicing, and that addition of recombinant A1 could re-inhibit 7b splicing. A1 binding sites on both sides of the repressed exon are needed for full inhibition. This led to the model that A1 was using its dimerization ability to interact with the two elements simultaneously and loop out the region containing the repressed exon (Figure 6C). It was shown that A1 immobilized on an RNA affinity column could indeed interact with two RNA binding sites simultaneously. This simultaneous binding of two sites required the Glycine rich C-terminal of the protein. In the crystal structure, a dimeric A1 protein is missing the Gly domain (206). Thus the Gly domain may do more than form the dimerization interface, and the stoichiometry of the A1 in the RNA looping complex is not clear. This looping model is interesting for several reasons. First other splicing regulators may also use a looping out mechanism to sequester exons or splice sites away from the splicing apparatus. Second, such a loop provides an appealing model for the splicing of long introns. Introns can be extremely long and it is difficult to understand how the ends will find each other during spliceosome assembly. The presence of intronic A1 sites would allow large portions of intron to be looped out and kept away from the decision of where to splice. It will be interesting to develop other assays for this "intron bridging" effect, and to examine its role in the splicing of other transcripts.

Besides A1, the other splicing repressor commonly found associated regulated exons is the polypyrimidine tract binding protein, PTB (also discovered as hnRNP I) (207) (208, 209) (210). PTB contains 4 RNP-cs domains, as well as conserved N-terminal and linker peptides (Figure 4). PTB has been extensively studied for its roles in splicing regulation and in other processes such as viral translation (211) (212). By SELEX, PTB optimally binds UUCU elements placed within a larger pyrimidine rich region, although other combinations of Cytosines and Uridines also bind well (213) (214). PTB forms a dimer in solution containing 8 RBD's (215) (216). It is not clear how many CU rich elements are engaged by a protein monomer. How these many domains combine to interact with the RNA is an interesting unanswered question (215) (216) (217) (218).

PTB has been implicated in the repression of a wide range of vertebrate tissue specific exons (69) (212) (211). The depletion of PTB from in vitro splicing extracts can activate the splicing of such exons (219) (202). Added back PTB re-represses splicing and this repression is dependent on the presence of PTB binding sites in the pre-mRNA. Depletion of PTB in vivo by RNAi also leads to increased exon inclusion for several exons (Wagner and Garcia-Blanco in press). PTB binding sites are often found in the polypyrimidine tract of regulated 3' splice sites, but essential PTB repressor sites are also present elsewhere. The location of these binding sites has given rise to models for PTB mediated splicing repression that are reminiscent of those for Sex-lethal and hnRNP A1 (212) (1). In some transcripts, the high affinity of PTB for the polypyrimidine tracts is thought to allow PTB to out-compete U2AF for binding and thus inhibit splicing. Other transcripts have additional required PTB binding sites that are often in the intron downstream of the regulated exon. These additional PTB elements are frequently clustered with other, sometimes positive-acting, splicing regulatory elements (212). For these transcripts, PTB is thought to assemble into a higher order complex that bridges the two sets of sites (212) (Figure 7). Similar to the A1 model above, this will loop out the repressed exon and presumably sequester it from the splicing machinery. However, the simultaneous binding of two sites by a PTB monomer or dimer, as seen with A1, has not been demonstrated with PTB. Thus, the actual contacts that form the bridge are not clear and there could be additional proteins involved. Moreover, the clustering of PTB sites with enhancer elements indicates that the PTB may also prevent enhancer activity.

Though not ubiquitous, PTB is a widely expressed protein. Interestingly, PTB is most often implicated in the repression of a highly tissue specific exon in all cells outside of one particular tissue (69) (212) (211). For example, the Rat  $\beta$ -Tropomyosin exon 7 is apparently repressed by PTB in all tissues except skeletal muscle (220). A smooth muscle specific exon in  $\alpha$ -Actinin is repressed by PTB outside of smooth muscle cells (219). The neuron specific N1 exon of c-src is apparently repressed by PTB everywhere except neurons (221). An exception to this is the third exon of Rat  $\alpha$ -Tropomyosin, which is repressed by PTB specifically in smooth muscle, and active elsewhere (222) (223). Thus, PTB most often appears to be a general repressor whose effects are specifically blocked for particular transcripts in particular tissues.

Mechanisms that allow release of PTB repression are likely to play an important role in the tissue specificity of splicing. Several mechanisms for the release of PTB repression are proposed. PTB concentration could simply be reduced in specific tissues; For example, PTB is low in neurons (224) (225). Alternatively, splicing activators such as CELF proteins that compete with PTB could be specifically expressed (186). This is similar to splicing regulation by the ratio of SF2/ASF to hnRNP A1 that affects a number of transcripts (161). Finally, proteins that block PTB action can be expressed. These could include inactive or dominant negative splice variants of PTB or tissue specific PTB homologues that bind similar RNA elements (224) (226) (227) (228) (186) (229). These mechanisms are not mutually exclusive and all may occur in different systems. In the case of the c-src N1 exon, the loss of PTB binding to repressor elements seems to be an active process that requires ATP. There are at least 4 PTB binding elements, two in the N1 3' splice site and two within a splicing enhancer downstream (Figure 7). In neural extracts, the neural specific homolog of PTB (called nPTB or brPTB) replaces the standard PTB in binding to the downstream repressor sites (227) (224) (228). Interestingly, in these extracts nPTB/PTB still crosslinks to the upstream sites within the active 3' splice site (202). However, when ATP is added to the extract to initiate splicing, this PTB is stripped off. It is thought that the nPTB changes the PTB interaction with the repressor sites allowing some general ATP requiring activity to mediate its removal.

A better mechanistic understanding of PTB mediated splicing repression will require better structural information about the nature of the repressed complex, better biochemical assays for splicing repression, and genetic data on the loss of PTB activity in specific tissues.

### **Multifactorial Systems of Splicing Control**

One feature of the *Drosophila* Sex Determination transcripts appears to be different from tissue specific splicing systems. For Sxl, Tra and dsx splicing, there is an apparent default splicing pattern (male) that is chosen in the absence of a regulatory protein. The splicing pattern can be shifted to the female pattern by the introduction of a single specific factor, Tra or Sxl. In contrast, many systems of tissue specific splicing appear to be without a true default choice, or single determinative factor, but instead are

under a combination of positive and negative control by factors that are fairly widely expressed (69). This combinatorial control is apparent for both mammalian and *Drosophila* transcripts, and may reflect a condition where the ratio of splicing patterns must be adjusted under many different conditions or in multiple cell types.

To exemplify the many factors that contribute to a single splicing choice, we will describe just two systems, but similar results have been obtained in several others (69) (230) (176) (231) (232) (233). Both chicken cardiac Toponin T and mouse c-src mRNAs contain exons whose inclusion is limited to a very specific tissue. Chicken cTnT exon 5 is spliced in embryonic muscle, but skipped in adult (234). C-src exon N1 is spliced in neurons, but skipped in non-neuronal cells (Figure 7)(235) (236). Both exons are short (30 and 18 nucleotides respectively), which can lead to exon skipping, perhaps due to a loss of exon definition interactions between the flanking splice sites (237) (238). Improving the splice sites to better match the consensus, or increasing the exon length, can increase the unregulated inclusion of a short exon (237) (239) (238, 240). Thus, both exons have features that limit their recognition by constitutive splicing factors. Both cTnT exon 5 and c-src N1 have ESE's that bind to SR proteins, including SF2/ASF, SRp40, SRp55 and SRp75 for TnT exon 5 and SF2/ASF for N1, and that stimulate the splicing of these exons (239, 241); (N. Rooke and DB. Submitted). The N1 exonic enhancer also binds hnRNPs A1 and H (N. Rooke and DB. Submitted). Both cTnT exon 5 and c-src N1 contain numerous positive and negative regulatory elements in their flanking introns (Figure 7)(242, 243) (190, 192, 197, 244). For both exons, a limited combination of these elements is sufficient to determine proper tissue specific splicing of a heterologous exon. For TnT exon 5, members of the CELF protein family bind to some of the intronic enhancer elements, and activate splicing of the exon (180, 186). For c-src N1, CELF family members do not seem to activate splicing and the enhancer activity is dependent on a UGCAUG element, whose mediating factor is unknown (190). In both exons, PTB binding sites in the flanking introns repress splicing, and PTB is proposed to be a general repressor that turns off splicing in most tissues (202, 221) (186).

In the face of all of these general positive and negative influences, how the precise tissue specificity of these exons is produced is an interesting question. For cTnT exon 5, an embryonic muscle variant of ETR3 is thought to counteract PTB and activate

splicing specifically in embryonic muscle (180, 186). There is an ETR3 isoform seen on westerns whose developmental expression correlates beautifully with splicing of the exon. However, the identity of this isoform is not known and it is not clear why other CELF proteins, which activate the exon in transfection assays, do not show activity on the endogenous transcript. The enhancer for the N1 exon is active in more than just neuronal cells and is thus not the only source of tissue specificity (197). Instead, part of the specificity of splicing is thought to derive from the specific loss of PTB mediated repression in neurons (Figure 7)(202). This may be due to the substitution of PTB with its neuronal homologue nPTB, which does not repress N1 splicing in vitro (227). However, nPTB has not yet been shown to affect N1 splicing in vivo. Thus, for both cTnT exon 5 and src N1, the decision to splice is due to a balance of multiple positive and negative inputs. Protein over-expression experiments with wild type and dominant negative mutant proteins indicate that exon inclusion can be increased by the loss of repression activity, the gain of enhancement activity, or both (186). Moreover, splicing activation may occur for different reasons in different cells.

### **Highly Tissue Specific Regulatory Proteins**

Although a single critical factor has not been shown to determine the tissue specificity of splicing in any system, there are splicing regulatory proteins whose expression is restricted to certain cells. The Neuronal PTB protein, and the variants of ETR3, such as NAPOR, are examples of such factors (227) (188) (180, 186) (228). The best studied splicing regulators that show precise tissue specific expression are probably the neuronal proteins Nova-1 and Nova-2 (245, 246). These two related proteins were first identified as autoantigens in Paraneoplastic disorders. They each contain three RNA binding domains of the KH-type, similar to hnRNP K (Figure 4). The Nova proteins are expressed almost exclusively in neurons of the central nervous system. In vitro selection experiments identified preferred RNA binding sites for the two proteins (246, 247). A Nova-1 binding site was found adjacent to the regulated 3A exon in the Glycine Receptor  $\alpha 2$  and cotransfection experiments showed that Nova-1 over-expression did increase the inclusion of this exon. Nova-1 was also found to interact with neuronal PTB in yeast two hybrid assays (called Brain (br) PTB in this case) (228). A brPTB binding site present

upstream of the Nova-1 site for GlyRa2 exon 3A also affected exon 3A splicing. Most significantly, Nova-1 knockout mice were generated (248). These mice die postnatally due to a loss of spinal cord and brainstem neurons. The Nova-1 null mice show a two-fold decrease in GlyRa2 exon 3A inclusion and a three-fold decrease in GABA<sub>A</sub> ?2 alternative exon splicing. From the partial effect of the knockout, Nova-1 is likely not the single controlling factor for these exons. However, the lethality makes clear that the protein is controlling some essential function in the neurons where it is expressed. It will be interesting to find out whether the lethal phenotype is due to multiple partial changes in splicing, as seen for Gly R a2 and GABA<sub>A</sub> ?2, or instead is the result of a drastic change in one or a few transcripts.

There are other cell-type specific regulators of splicing found in other species. The Elav protein is expressed in all *Drosophila* neurons and has been shown to regulate transcripts from the Neuroglian, Erectwing, and Armadillo genes (182) (183, 184). Also in *Drosophila*, the Halfpint protein controls ovary specific splicing of transcripts from the Ovarian Tumor Gene (249). In *C. elegans*, the protein Mec-8 controls Unc52 splicing in body wall muscle (250). The availability of genetic analysis in these systems make them ripe for more detailed mechanistic studies.

### **Genetic Dissection of Complex Systems of Regulation**

It seems clear that the ratio of one splicing pattern to another for a typical alternatively spliced transcript is determined by the combination of factors. Most experiments using protein over-expression or mutation of specific protein binding sites show only partial effects on the use of a particular splicing pattern. It is thus difficult to assess the relative importance of one factor over another.

The biological roles of putative splicing regulatory proteins and their many of targets is perhaps best approached through genetic studies. Genetic analysis and RNA interference methods in *Drosophila* and *C. elegans* have already provided important information on the function of U2AF, SR proteins, and other factors in particular splicing events (130) (94) (93) (251). In mammals, this kind of data will come from genetic knockouts (such as for Nova-1 protein), from RNA interference experiments, or from

somatic cell knockout methods. These have only begun to be generated for the many known splicing regulatory proteins.

The *in vivo* role of ASF/SF2 was examined in the chicken lymphocyte cell line DT40, which exhibits high levels of homologous recombination (96). Transfection of insertion mutant alleles and selection of homologous recombinants allowed the recovery of heterozygous but not homozygous ASF/SF2 knockout cells. That the homozygous knockout was lethal was demonstrated by complementation with a transgene under tetracycline regulation. This gene when active allowed the recovery of cells where both endogenous genes were mutant. The repression of the transgene by Tetracycline in the homozygous knockout cells was again lethal, demonstrating that ASF/SF2 was indeed essential in these cells. Through complementation experiments with various mutant transgenes, it was demonstrated that the RNP domains of the protein were essential, while the RS domain could be replaced with the RS domain of several other SR proteins without the loss of viability (111). This demonstrated that the ASF/SF2 RS domain is a generic interaction domain that is not targeted differently in different SR proteins. These DT40 studies are a potentially fruitful approach to analyzing the function of many other splicing regulatory proteins that are expressed in these cells.

The DT40 studies indicate that ASF/SF2 is an essential gene, at least in this cell line. This could pose a problem for generating traditional mouse knockouts of SR proteins. However, the potential embryonic lethal phenotype of the double knockout can be circumvented using the Cre/LoxP method for generating knockouts in specific cell lineages. This has been used very effectively for the SR protein SC35 (95). For SC35, the homozygous null mutation did indeed have an embryonic lethal phenotype in mice. However, by generating the "floxed" allele with flanking LoxP sites and breeding these mice with mice expressing Cre recombinase specifically in the thymus, Fu and colleagues were able to examine SC35 function specifically in the T cell lineage. The SC35 mutant shows a specific T cell defect and changes in the splicing of CD45 transcripts. This method shows great promise for determining the roles of SC35 and other factors in any particular splicing event *in vivo*.

## **MECHANISMS OF INTRON RETENTION**

Most alternative splicing patterns involve a choice of one set of splice sites competing against another. Because the change involves two competing splicing pathways, any mechanism that alters the relative rates of spliceosome assembly for the two splicing patterns is a potential means of regulation. One alternative splicing pattern where this may not be the case is intron retention. Here, the choice is between splicing with intron excision and no splicing with the retention of an intron in the final mRNA. This partially spliced RNA product must then be exported to the cytoplasm. Thus for intron retention, the competition may be between splicing and mRNA transport rather than between two splicing patterns. Studying these systems may yield important information about how the cell determines when an mRNA is complete and ready for movement to the cytoplasm.

In fact, many factors that affect intron retention are the same as those affecting other alternative splicing patterns. For the *Drosophila* Msl-2 transcript, an intron is retained in the final mRNA in female flies (54). This intron retention requires Sxl protein binding to both the 5' and 3' splice sites. The blockage of both sites may be necessary to prevent any recognition of the intron as a substrate for the splicing reaction, thus allowing its transport. In this regard, it may be significant that the Sxl protein accompanies the transcript to the cytoplasm.

### **Drosophila P Element Splicing**

Another well characterized system of intron retention is the P Element transcript in *Drosophila* (252) (Figure 8). The P Element is a Transposon whose movement in the genome is restricted to the *Drosophila* Germline. The Transposase enzyme is encoded on the fully spliced P Element mRNA. In germ cells, all of the P element introns are excised and Transposase is expressed. In somatic cells, the P element third intron is retained in the mRNA, preventing translation of the Transposase enzyme and hence blocking somatic transposition. A series of genetic and biochemical analyses by the Rio lab have identified components of this regulatory system. The exon upstream from the third intron contains a regulatory sequence that inhibits its splicing in somatic cells. Interestingly, this sequence contains elements called Pseudo 5' splice sites that are similar to 5' splice

sites and bind the U1 snRNP, but which are not used for splicing (253). The regulatory region also contains binding sites for proteins, including the P element Somatic Inhibitor (PSI) and hrp48 (254-256). Hrp48 is a *Drosophila* homolog of hnRNP A and B. PSI has an interesting domain structure with four central KH-type RNA binding domains, flanked by large N and C-terminal regions (Figure 4). The C-terminal domain contains a novel repeated motif. PSI is homologous to several vertebrate proteins, including the KSRP that binds to the intronic enhancer for c-src N1 exon (200). In addition to splicing, these vertebrate proteins have been implicated in a variety of processes, including transcription, RNA localization, and RNA degradation (257) (57) (258). How all of these functions are reconciled with each other is not yet clear. The conserved C-terminal repeats in PSI bind to the U1 70K protein (259). PSI expression is excluded from the female germ line and both biochemistry and genetics experiments indicate that this restriction to somatic cells is an important determinant of the splicing inhibition (255, 256, 260). PSI null mutations are lethal. However, hypomorphic alleles that lack the U1 70K interaction sequence show misregulation of both P element splicing and the splicing of Squid gene transcripts (260). It is thought that in somatic cells, PSI binds to the repressor element and through its interaction with the U1 70K protein induces the binding of the U1 snRNP to the Pseudo 5' splice sites (Figure 8). This assembly in some way prevents use of the normal 5' splice site in the splicing reaction. Perhaps the purpose of the pseudo 5' splice sites is to interact with the 3' splice site in a non-functional manner and thus prevent its use with either the natural 5' splice site or any possible cryptic sites nearby. If such a non-functional spliceosome is being assembled, it will be interesting to determine what allows its disassembly and hence the transport of the mRNA retaining the intron to the cytoplasm.

### **Retroviral Splicing and the balance of splicing and transport**

The most familiar systems of intron retention are the retroviruses. In the simplest retroviruses the viral Envelope protein is encoded on a spliced mRNA (261-265). When this intron is retained, the unspliced transcript serves both as mRNA for the Gag and Polymerase proteins and as genomic RNA for packaging into virions. More complex retroviruses such as HIV create additional transcripts using alternative splice sites and

cassette exons to produce mRNAs for additional viral proteins, most notably the regulatory proteins Tat and Rev.

Retroviruses make use of many of the same splicing regulatory elements and factors described above. In addition to splicing enhancers and repressors, there are also RNA sequence elements that allow transport of unspliced RNAs to the cytoplasm (261-265). Some of these transport elements are constitutive, interacting directly with cellular factors (266-271). Other elements interact with viral proteins, such as HIV Rev, to allow viral control of the nuclear export process (264). It is thought that partially spliced RNAs are retained in the nucleus by the binding of particular spliceosomal components. The viral transport elements are able to override this retention and allow transport of any RNA that bears them. Retroviruses ensure that some unspliced RNA is available for export by making the viral introns relatively inefficient. The rate of viral splicing is carefully balanced to produce enough of both spliced mRNA and transportable unspliced RNA (272-274). Many features make retroviral splicing inefficient, including poorly recognized splice sites, and ESS elements as discussed above (for example see (275)). More unusual features are also seen. Rous Sarcoma Virus (RSV) has an interesting Negative Regulator of Splicing (NRS) downstream from the viral 5' splice site (276, 277). This multipartite element has binding sites for numerous factors, including SR proteins, the U1 snRNP, and the minor class snRNP U11. The sequence assembles into a large complex in cellular extracts (278-282). Both the SR proteins and U1 are important for the splicing inhibition. It is thought that this complex interacts with the downstream viral 3' splice sites to inhibit splicing. This is similar to the inhibition of P Element splicing and again it will be interesting to examine how and when this complex is removed from the transported RNA.

## **LONGSTANDING QUESTIONS AND EMERGING ISSUES**

Many of the most interesting questions regarding the mechanisms of alternative splicing are still unanswered. Recent results have identified important links between splicing and the events that precede and follow it in the gene expression pathway. Moreover, the biological roles played by alternative splicing in cellular differentiation

and adaptation, as well as in genetic variation between species, are only beginning to be examined.

### **How does the structure of the Pre-mRNP actually determine spliceosome assembly?**

Of the identified splicing regulators, relatively few are known to interact directly with spliceosomal components such as the U1 snRNP (TIA-1, PSI, SR proteins) or the U2AF protein (SR Proteins, Tra-2). These proteins all assemble into large pre-mRNP complexes and splicing commitment assays make clear that some pre-mRNP complexes determine subsequent spliceosome assembly. However, in nuclear extracts, a heterogeneous H complex containing many of these proteins forms on almost any RNA, and a majority of this material does not proceed through the splicing reaction (283). Because of their size and complexity, these complexes are difficult to define and isolate as homogenous assemblies. Thus, the actual structure of what is called the hnRNP or pre-mRNP complex is very unclear. At a more local level, how a multiple RBD protein interacts with RNA and cooperates with other proteins in assembling into a large pre-mRNP is also not understood. If we are to understand how proteins such as ASF/SF2 or PTB actually promote or block spliceosome assembly, the structures of complete multidomain proteins complexed with RNA will be essential.

There are also surely many splicing regulatory molecules yet to be discovered. Most known splicing regulators directly contact their target RNA. As in transcription, there are likely to be other splicing regulators that interact with other proteins but do not contact nucleic acid. Moreover, the regulators need not be proteins. Some of the recently discovered non-coding micro-RNA's are appealing candidates for regulators of the splicing reaction (284-287), although there is little evidence for this yet.

### **Coupling splicing to upstream and downstream processes**

Although studied *in vitro* using pre-synthesized RNAs, spliceosomes are thought to assemble *in vivo* onto the pre-mRNA as it is being synthesized (288) (289). Despite assembly concurrent with transcription, the kinetics of intron removal is variable and the actual excision of some introns may be completed after the polymerase reaches the end of the transcription unit (290). The cotranscriptional assembly of the spliceosome has

profound implications for the regulation of splice site choice. Recent studies have made clear the intimate relationship between pre-mRNA synthesis and processing (291) (292) (293, 294) (295). Simply changing the promoter on a gene can have a large effect on the rate of inclusion of a regulated exon (296). This is thought to be partially due to changes in the transcription elongation rate through the gene (297, 298). Given equal rates of spliceosome assembly at each of two competing splice sites, then by slowing the rate of the synthesis of the downstream choice (through a transcriptional pause site for example), the upstream site can be made to predominate. Although not proven to occur in an endogenous gene, such transcriptional regulation seems likely to be an important component of many splicing choices (See for example: (299-301)).

Genes where splicing seems particularly likely to occur during transcription are those with very long introns. It is a longstanding puzzle how a 5' splice site at the beginning of a 100 kb intron can be accurately joined to the correct 3' splice site so far downstream, rather than to an intervening cryptic site. This correct site is not only far away in the sequence, but will not be synthesized until over an hour after the 5' splice site (302). This problem is neatly resolved by the idea of recursive splice sites (303). The initial 5' splice site may splice to intermediate 3' splice sites along the intron as they are synthesized. These recursive intermediate sites are special in that they regenerate a 5' splice site as they are joined to the original site. Thus, the 5' exon may hop along the long intron, being re-spliced several times at these ratcheting points, before being joined to the final correct site at the end of the long intron. This last site would presumably not regenerate a 5' splice site, and thus terminate the re-splicing process. There is evidence for the use of recursive splice sites in several long introns (303). Interestingly, a resplicing mechanism offers another point of control for altering splice site choice. In the *Ubx* gene of *Drosophila*, a cassette exon containing a recursive splice site can be removed from the RNA even after joining to the upstream exon (303).

Besides indirect effects of the transcription reaction on splicing, there is also evidence for a more active role. This has given rise to the mRNA factory idea: that transcription and processing all take place in the same very large coordinated complex. Such mRNA factories have many appealing features for controlling gene expression. For

example, they would ensure that the order that introns assembled into spliceosomes would be the order they were transcribed.

The C-terminal domain (CTD) of RNA polymerase II has been shown to directly interact with a number of RNA processing factors (304, 305) (306). The CTD consists of a long series of heptapeptide repeats and is thus capable of interacting with multiple factors simultaneously. Deletion of the CTD disrupts the constitutive splicing reaction (307). The CTD is phosphorylated and this phosphorylation is altered as it progresses through gene presumably altering its interactions with processing factors (308) (309). Phosphorylated CTD can stimulate splicing in vitro (310, 311). Conversely, splicing factors directly affect elongation in vitro (312). Only with further study will a real picture emerge of how spliceosomes might be loaded onto the polymerase, and how they are deposited onto the nascent RNA.

It seems likely that some factors will bind directly onto the pre-mRNA rather than get there via the CTD, and that some introns will be removed after transcription is complete (290). However, proteins with RS domains have also been shown to interact with transcription factors that bind at promoters (313, 314). The WT1 protein seems to have a dual function with one splice variant acting on transcription and a very similar variant interacting with spliceosomes (315-317). The development of coupled in vitro transcription/splicing systems may enable us to look at the activity of such dual function proteins.

As discussed in relation to Retroviruses, splicing must also be controlled relative to downstream processes, most notably nuclear export (295, 318). How does the cell determine when the mRNA is finished and ready for export? In addition to the viral signals that allow export of unspliced RNA, there are also proteins that get deposited on the RNA during the splicing process (319) (320-322). The Exon Junction Complex (EJC) is a non-sequence-specific assembly, that is deposited upstream from an Exon/Exon junction after splicing has occurred. The EJC contains a number of proteins that interact with the Nuclear export pathway, and the Nonsense Mediated RNA Decay (NMD) pathway (295, 318). Interestingly, one of its components, UAP56, has also been implicated in the splicing reaction, providing a link with the process that is apparently required for its deposition (323). The connection created by the EJC between splicing

and the downstream fate of the mRNA allows for interesting speculation about additional roles for alternative splicing patterns. It is possible that there are multiple kinds of EJC's that differentially affect the later export, translation, localization or stability of the mRNA (see for example (324)). If altering the splicing pattern could alter the components of the EJC, the significance of a particular change in splicing could go well beyond a change in its sequence.

### **Splicing Regulation and Cell Physiology**

For practical reasons, most systems of alternative splicing analyzed at the molecular level exhibit stable differences in splicing between two different cell types. However, splicing can also be regulated within a given cell by external stimuli and growth conditions (325) (326) (327) (328) (329) (330) (331) (332) For a review see (333). Very little is known about how signal transduction pathways impinge on the splicing reaction. To observe a change in the expressed ratio of two spliced forms, the RNA must turn over. Thus, it is often difficult to separate changes in splicing from changes in mRNA stability. There are a few systems where results reflecting clear differences in splicing have been obtained. The cell surface molecule CD44 exhibits a large number of splice variants due to the differential inclusion of 10 variable exons (334). Several of these variable exons are heavily studied because their inclusion has been associated with the progression of certain tumors to an invasive phenotype (335). The inclusion of variable exon 5 is induced by activation of the PKC/Ras pathway and involves the downstream effector ERK (336) (337). Transcripts for the Protein Tyrosine Phosphatase CD45, an important lymphocyte marker, contain several exons that are included in B cells but variably excluded in T cells ((338) and references therein). During T Cell activation by antigen, the splicing of CD45 is altered such that these alternative exons are repressed. This splicing switch also involves the PKC/Ras pathway. In both these systems, exonic elements have been identified that are required for the inducible change in splicing. For CD44 exon 5, this element binds hnRNP A1 (339).

In the nervous system, a number of gene transcripts are known to alter their splicing in response to cell activity (340) (341, 342) (333). Some of these RNA's contain exons that are repressed in response to cell depolarization and to activation of CaM

Kinase IV. In the STREX exon of BK Potassium Channels and in Exon 5 of the NMDA Receptor 1, the CaM Kinase dependent repression requires a special sequence in the 3' splice site called a Calcium Responsive RNA Element (CaRRE) (343). This element is transferable and can confer CaM Kinase repression on an otherwise constitutive exon. The CaRRE is thus thought to be the actual target of the repressive effect, although the factors that bind to it are not known. The CaRRE element and those in CD44 and CD45 provide an essential starting point for following the splicing response back to the signaling system. It is not clear in these examples whether the signaling pathway is directly altering splicing regulatory proteins or inducing the expression of a splicing regulator. The CD44 switch is not blocked by protein synthesis inhibitors (336). The time-course of the effect (or our ability to measure it) is certainly long enough to allow for the latter possibility.

Besides the direct modification or the induced expression of splicing factors, splicing may also be regulated by alterations in the nucleo/cytoplasmic distribution of certain proteins. Changes in SR protein phosphorylation and localization are known to occur during early nematode development (138). HnRNP A1 is partially shifted to the cytoplasm in response to MKK/p38 signaling (344). In primary neuronal culture, Tra-2  $\beta$ 1 shows cytoplasmic accumulation in response to increasing the intracellular Calcium concentration with the drug Thapsigargin (345). A similar shift in the localization of both Tra-2 $\beta$ 1 and SR proteins was induced in response to Ischemia in mice. For both A1 and Tra-2  $\beta$ 1, shifting of the proteins to the cytoplasm was accompanied by moderate changes in the spliced isoform ratio of particular transcripts (344) (345). It is not clear that these splicing changes are the direct result of splicing factor relocation. However, such relocation is an appealing mechanism for inducible regulation and is common in systems of transcriptional regulation (346, 347).

Splicing is also likely to be regulated in response to cell proliferation. During mitosis, the nuclear envelope breaks down and the nuclear contents mix with the cytoplasm. Very little is known about what happens to the splicing reaction during this period, but most other features of gene expression are shut down. A recent study shed the first light on this process (Shin and Manley, *Cell* in press). Splicing extracts derived from mitotic cells were shown to have a specific inhibitor of splicing. This inhibitor was

identified as the dephosphorylated form of SRp38, an SR related protein. Thus, an attractive model is that a mitosis specific phosphatase dephosphorylates p38 to inhibit splicing during this phase of the cell cycle. Several important cell cycle regulators are alternatively spliced (348, 349). It will be interesting to test the dependence of these splicing patterns on SRp38.

### **Alternative Splicing and the Genome**

The prevalence of alternative splicing and our limited understanding of its mechanisms present a challenge for identifying all of the proteins available to an organism (3) (4). It is currently difficult to use genomic sequence to predict splicing patterns. Nevertheless, new genomic approaches will have a broad impact on our understanding of both the mechanisms and the biological roles of alternative splicing (350) (Woodley and Valcarcel in press).

The most studied systems of alternative splicing were chosen in part because of their conservation between species. This conservation provides some assurance of the physiological importance of the alternative transcripts. However, not all splicing patterns are conserved between different mammals or between different Dipteran Insects . Even within a given species, there appears to be significant variation in the relative use of particular splice variants (351) (352) (353). One explanation for this is that alternative splicing provides an advantageous mechanism for testing new protein sequences during evolution. A single point mutation can extend an exon or create a new exon. Such a transcript, encoding a new protein, may comprise only a few percent of the product mRNA. Thus, mutations that alter splicing can allow production of new proteins without significant loss of the wildtype protein. Although this might be advantageous for protein evolution, the high degree of variability in splicing makes it difficult to prove the significance of a splice variant that is not conserved across species.

The large number of splice variants also makes it difficult to study splicing regulation across the whole genome. One would like to examine whether a particular exon is coregulated with others, and to test how a whole ensemble of splice variants is altered by a particular condition. There is consequently a great deal of interest in the use

of DNA Micro-arrays to study splicing regulation (350) (Woodley and Valcarcel in press).

Several microarray approaches to splicing have been described, although there are technical limitations to all of them. The most straightforward system examined splicing across the yeast genome, using oligonucleotide probes that hybridized in exons, introns, or across splice junctions (354). This system was able to clearly categorize intron groups based on their different splicing factor dependence. In another approach, high density oligonucleotide arrays were generated where probes were tiled across the entire human genome or more densely across Chromosome 22 (355). Such probes do not give exon junction information, but only indicate the presence of a particular exon region in the mRNA population. This approach will not yield sufficient information to describe complicated splicing patterns and identify small exons. Nevertheless, it will give clear first identification of many new exons and their tissue specificity.

Oligonucleotide arrays depend on the hybridization properties of the probes being well matched across the array for  $T_m$  and other features. Some exons may have oligonucleotide probes that will not work in parallel with others on the same array. In a novel fiberoptic-array system, the differences in hybridization properties were minimized by using the ligation of two oligos to measure each spliced junction (356) (357). It is not clear whether this system can be scaled up to measure thousands of splice variants simultaneously, as is needed for a whole genome analysis. However, it has potential advantages for the simultaneous measurement of particular exon groups.

Even if it is not possible to put all exons on a single array, these methods hold promise for analyzing the combinatorial mechanisms controlling splicing. The co-variation of exon groups can be related to studies of splicing factor expression throughout an organism. This can be combined with genetic knockouts of individual splicing regulators and with the immunoprecipitation of specific transcripts with antibodies to regulatory proteins (358) (260). Micro-arrays promise to take studies of splicing regulation to the level of the whole genome. By studying alternative splicing at both the genomic and mechanistic levels, our understanding of splicing regulation can be integrated into the biology of cells and organisms.

## **ACKNOWLEDGMENTS**

I thank Benoit Chabot, Thomas Cooper, Xiang-Dong Fu, Brenton Graveley, Kristen Lynch, Don Rio, and Juan Valcarcel for advice on the manuscript, and Kim Simmonds for help in preparing it. My own work has been supported by the NIH, the Packard Foundation and as an Investigator of the Howard Hughes Medical Institute.

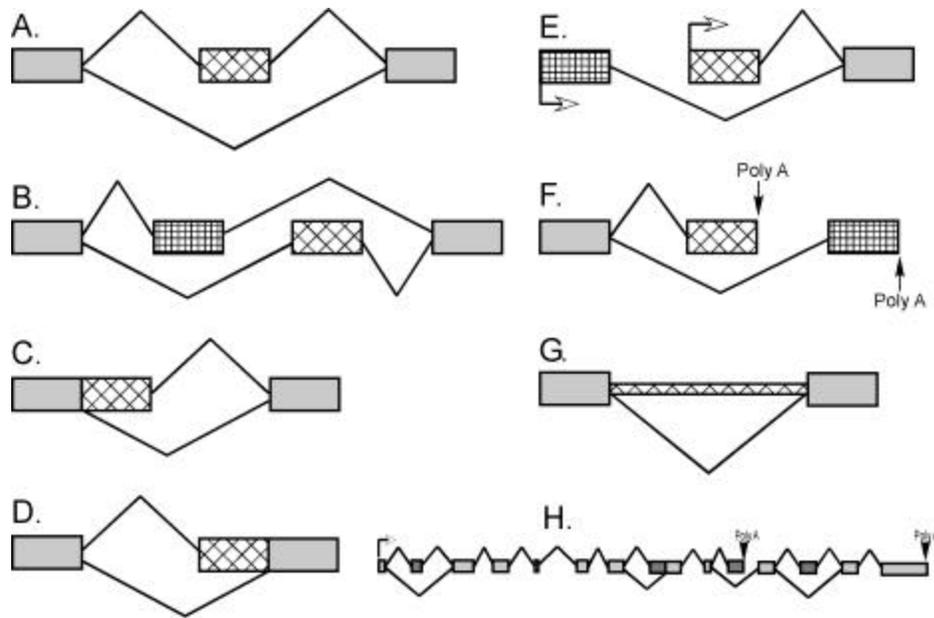


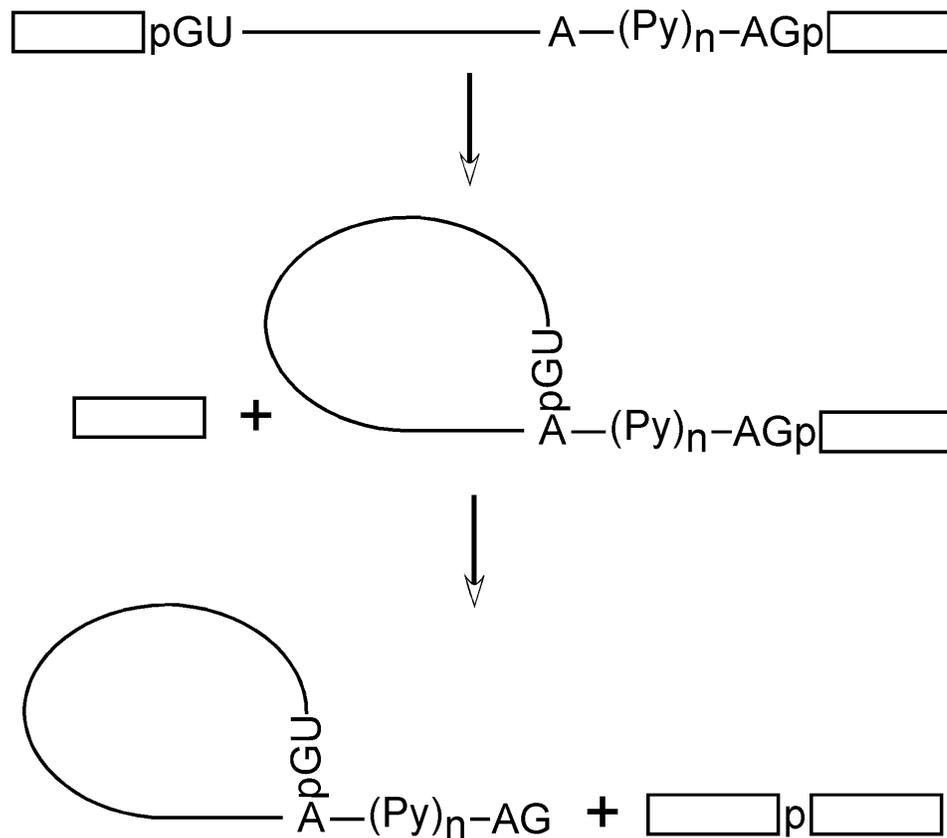
Figure 1.

**Figure 1. Patterns of Alternative Splicing**

Constitutive sequences that are present in all the final mRNAs are gray boxes. Alternative RNA Segments that may or may not be included in the mRNA are shown as hatched boxes.

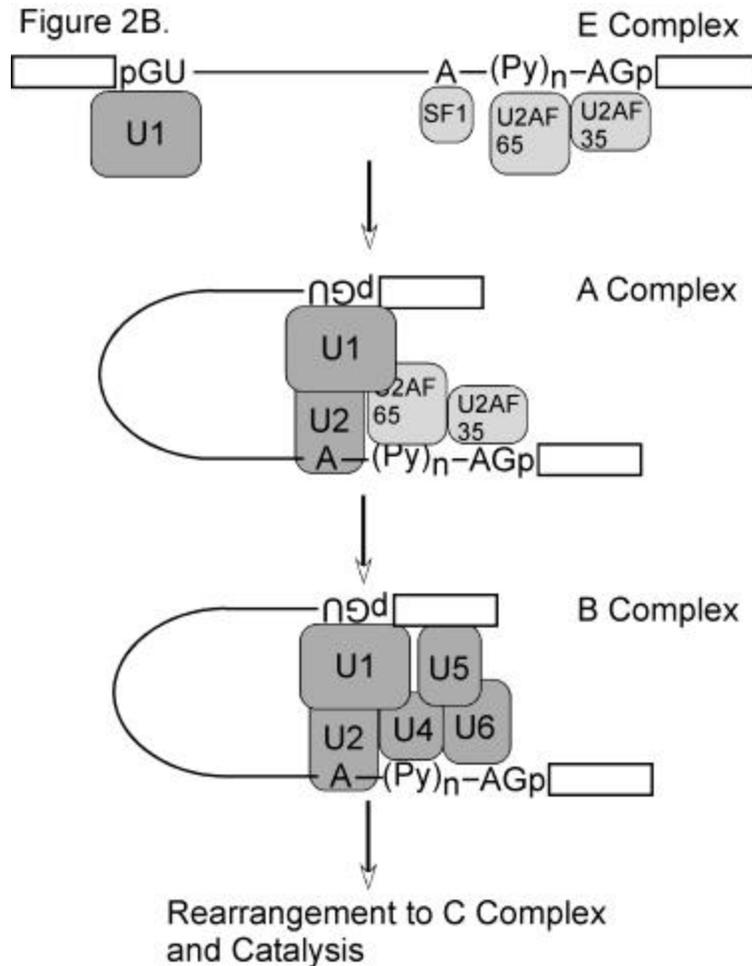
- A. A cassette exon can be either included in the mRNA or excluded.
- B. Mutually exclusive exons occur when two or more adjacent cassette exons are spliced such that only one exon in the group is included at a time.
- C. D. Alternative 5' and 3' splice sites allow the lengthening or shortening of a particular exon.
- E. F. Alternative Promoters and Alternative Poly A sites switch the 5' or 3' most exons of a transcript.
- G. A retained intron can be excised from the Pre-mRNA or can be retained in the translated mRNA.
- H. A single pre-mRNA can exhibit multiple sites of alternative splicing using different patterns of inclusion. These are often used in a combinatorial manner to produce many different final mRNAs.

Figure 2A.



**Figure 2A. Splicing takes place in two transesterification steps.**

The first chemical step of the reaction is the attack of the 2' Hydroxyl at the branch point A residue on the phosphate at the 5' end of the intron, within the 5' splice site. This results in two reaction intermediates: the detached 5' exon and an intron-3' exon fragment in a lariat structure. This contains an unusual phosphotriester linkage at the Branchpoint A residue. The second step of the reaction involves the attack of the 3' Hydroxyl of the detached exon on the phosphate at the 3' end of the intron, within the 3' splice site. This ligates the two exons and releases the intron lariat.

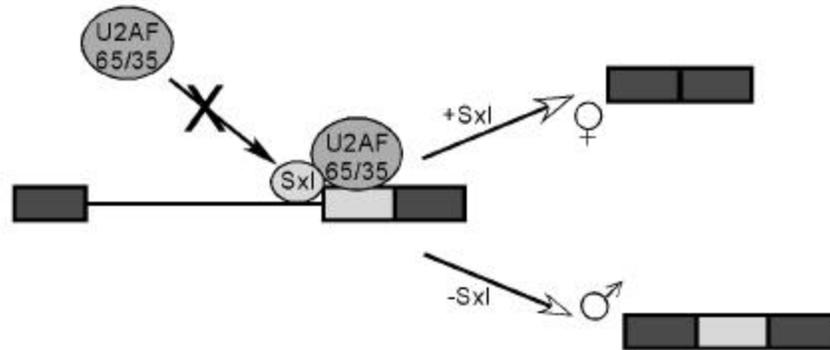


**2B. The spliceosome assembles onto each intron from a set of 5 small nuclear ribonucleoproteins.**

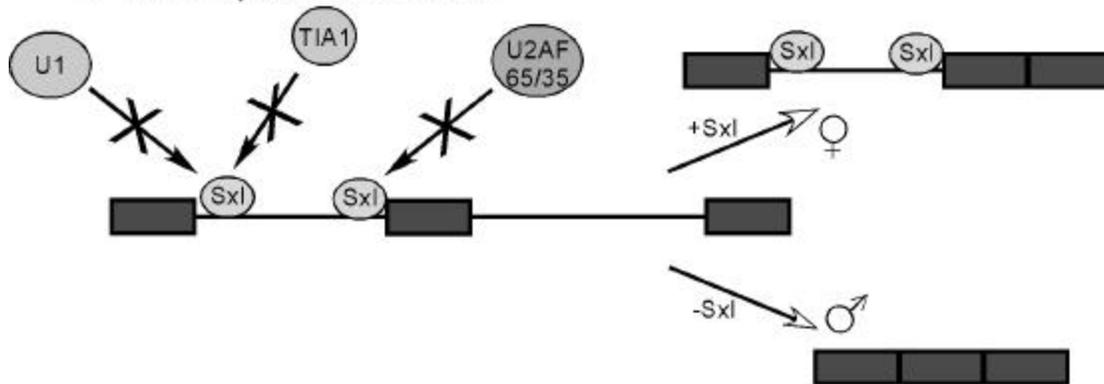
The conserved sequence elements that make up the splice sites are each recognized by different spliceosomal components. The Early (E) complex contains the U1 snRNP bound to the 5' splice site via RNA/RNA base pairing. Each element of the 3' splice site is bound by a specific protein, the branch point by SF1 (BBP), the Polypyrimidine Tract by U2AF 65, and the AG dinucleotide by U2AF 35. This complex also apparently contains the U2 snRNP not yet bound to the Branchpoint (24). The A complex forms when U2 engages the branchpoint via RNA/RNA basepairing. This complex is joined by the U4/5/6 Tri-snRNP to form the B complex. The B complex is then extensively rearranged to form the Catalytic C complex. During this rearrangement the interactions of the U1 and U4 snRNPs are lost and the U6 snRNP is brought into contact with the 5' splice site. The spliceosome contains many more component proteins than are shown here (359, 360) (361) (362) (363). Moreover, other pathways of assembly may be possible (364) (365).

Figure 3.

## A. Transformer



## B. Male Specific Lethal 2



## C. Sex Lethal Autoregulation

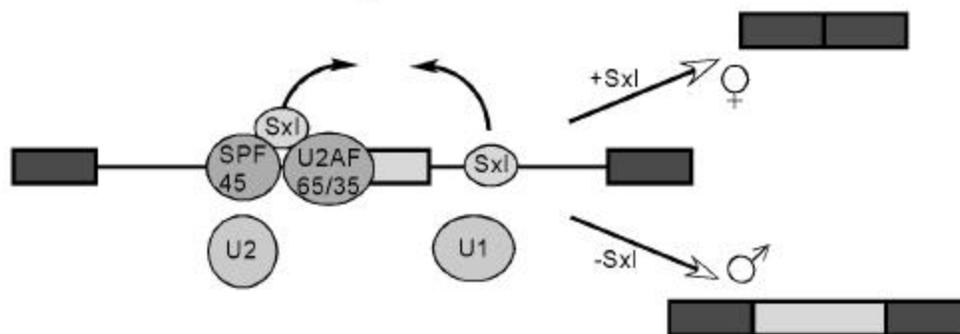
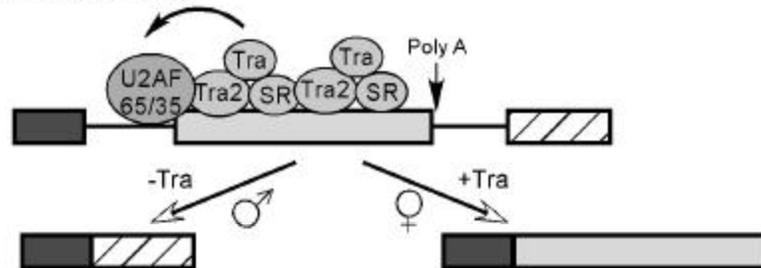
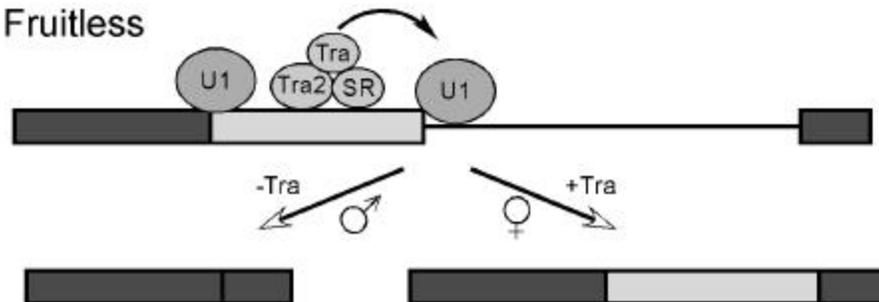


Figure 3

## D. Doublesex



## E. Fruitless

**Figure 3. Splicing Regulation in the *Drosophila* Sex Determination Cascade.**

- A. Sex Lethal represses a splicing pattern in the *Transformer* RNA. In male flies, the absence of Sxl protein allows U2AF binding to an upstream 3' splice site, producing the male spliced product. In female flies, Sxl protein binds to the upstream 3' splice site blocking U2AF binding. U2AF binds instead to an alternative downstream site causing splicing in the female specific pattern and producing an mRNA encoding active Transformer protein.
- B. Sxl Protein also binds to the *Msl2* transcript. Here, Sxl binds in the 3' splice blocking U2AF binding, and near the 5' splice site, blocking binding by the TIA-1 protein and the U1 snRNP. This causes retention of the intron in the final *Msl2* mRNA in female flies. In male flies, the absence of Sxl allows the mRNA to be fully spliced.
- C. In the *Sxl* transcript itself, the 3' splice site of the 3<sup>rd</sup> exon contains two AG dinucleotides. In male flies, the downstream AG is bound by U2AF and is required for activation of the exon for splicing. The upstream AG is bound by the SPF45 protein and is the site of exon ligation. Thus, in male flies, exon 3 is included in the mRNA that now encodes an inactive Sxl protein. In female

flies, Sxl protein derived from the activation of an early embryonic promoter is present. Sxl binds on both sides of the exon to repress its use. This produces some larger assembly of Sxl protein and requires the Snf protein. The result is exon skipping and the production of additional Sxl protein from the constitutive promoter. Under these conditions, it appears that lariat formation may still occur at the exon 3 3' splice site. See the references by Lallena (2002) and Graveley (2002) on this point.

- D. The Tra protein produced in female flies is a positive regulator of *Doublesex* splicing. *Doublesex* encodes two different transcription factors. Dsx Exon 4 contains a poor polypyrimidine tract causing exon 4 skipping in male flies. Exon 4 also contains a Tra dependent splicing enhancer. In females, Tra protein cooperatively assembles onto this enhancer with the Tra2 protein and two different SR proteins. This enhancer complex stimulates U2AF binding and splicing of exon 4, producing the female product of the dsx gene.
- E. The Tra protein also regulates the splicing of *Fruitless* mRNAs. *Fruitless* has a pair of alternative 5' splice sites. In male flies, the upstream most of these sites is used. In female flies, Tra, Tra2 and the SR protein RBP1 bind to an exonic enhancer to activate splicing of the downstream 5' splice site.

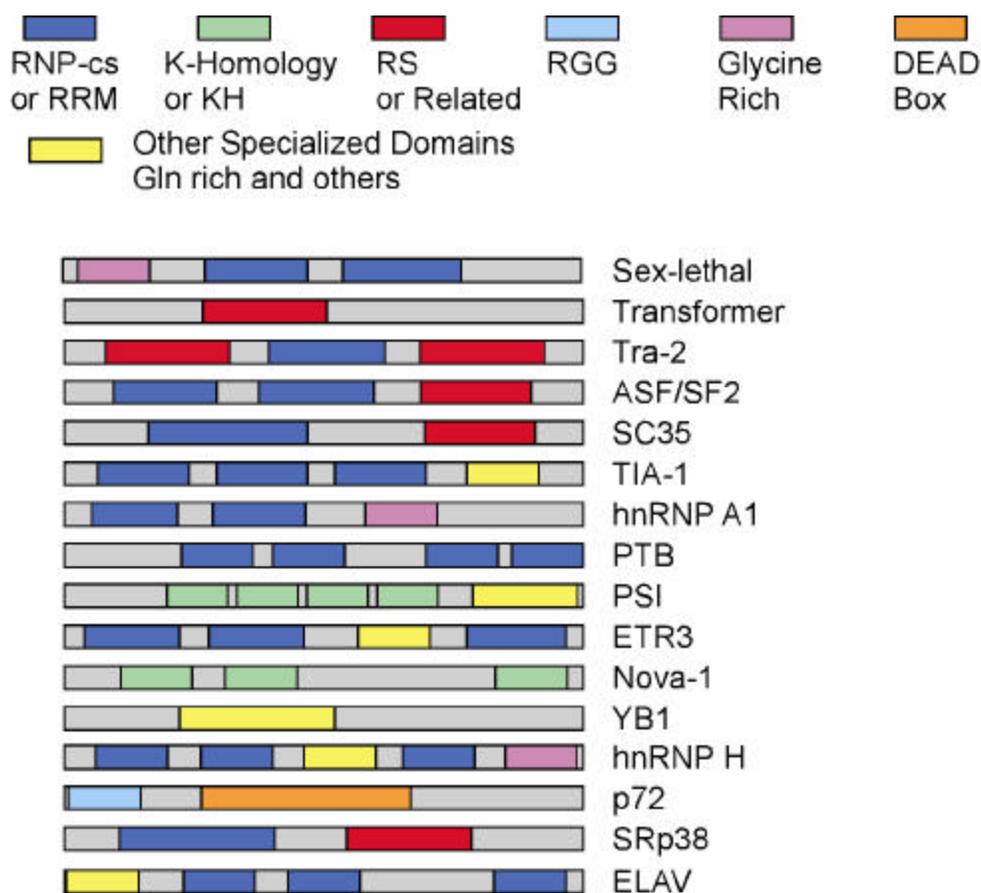


Figure 4.

**Figure 4. Domain Structures of Splicing Regulatory Proteins.**

Splicing regulatory proteins have common domains but combine them in many different ways. Nearly all have RNA binding domains of either the RNPcs (RRM) or KH type. However, these are repeated in different numbers and combined with different auxiliary domains such as RS or RS-like domains. Although structures of individual or pairs of domains have been solved, little is known about the overall structure of these multi-domain proteins. Shown here are some of the proteins discussed in the text, normalized to equal length. The positions and number of their known domains are indicated by color coded boxes.

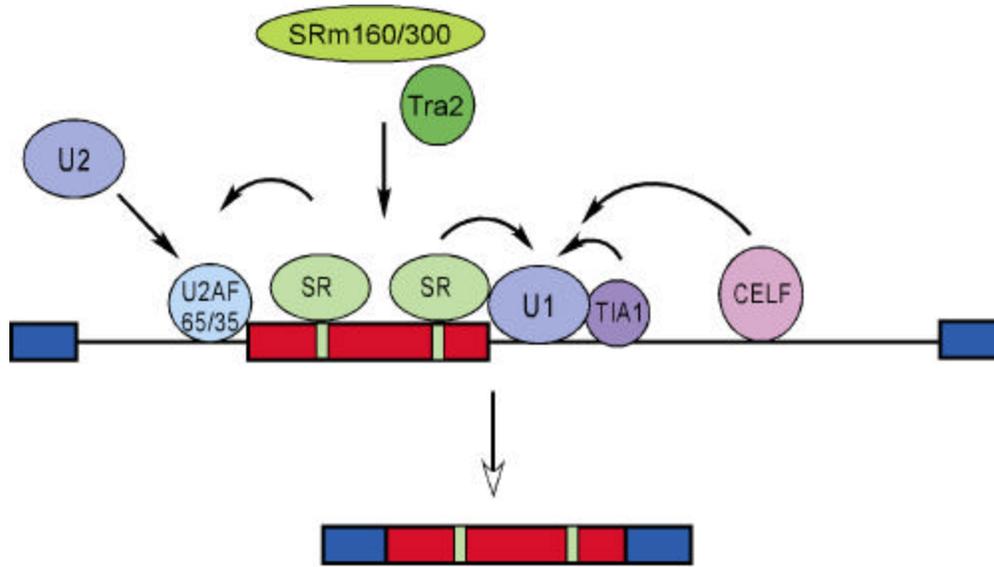


Figure 5.

### **Figure 5. Mechanisms of Splicing Activation**

Splicing activators are generally thought to interact with components of the spliceosome to stabilize their binding to adjacent splice sites. SR proteins bind to exonic splicing enhancer elements (green boxes) to stimulate U2AF binding to the upstream 3' splice site, or U1 snRNP binding to the downstream 5' splice site. SR proteins often need other factors to function such as the SRm160/300 proteins or Tra-2. Proteins that bind to intronic splicing enhancers include TIA-1 and the CELF proteins. TIA-1 binds immediately downstream from the 5' splice site to stimulate U1 binding. CELF protein binding sites can be more distant from the regulated exon and it is not known how they might interact with the spliceosome.

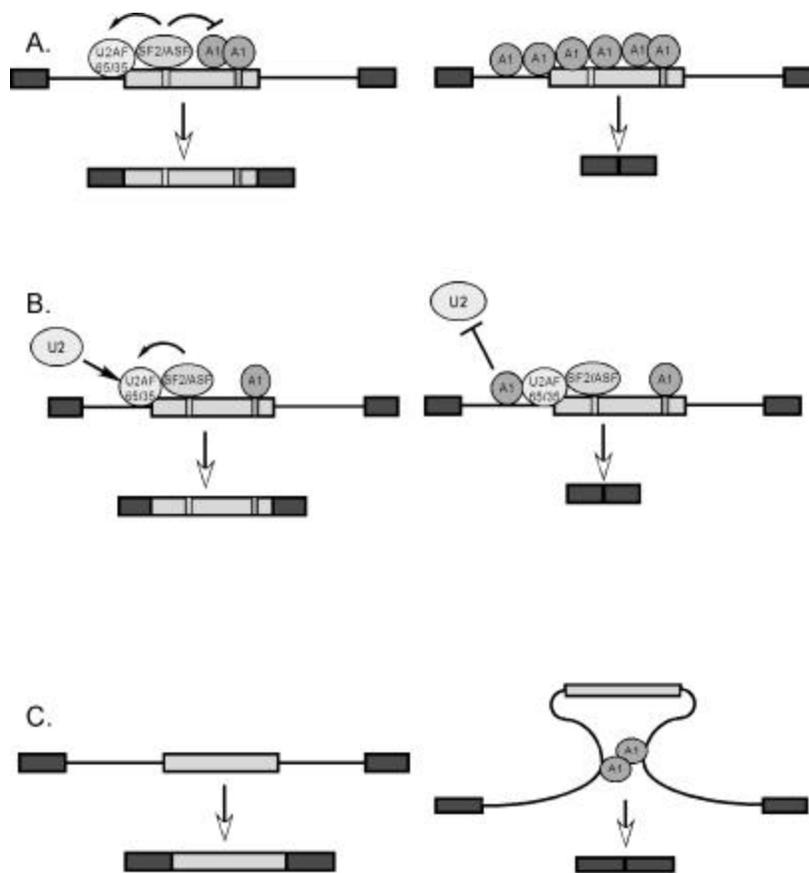
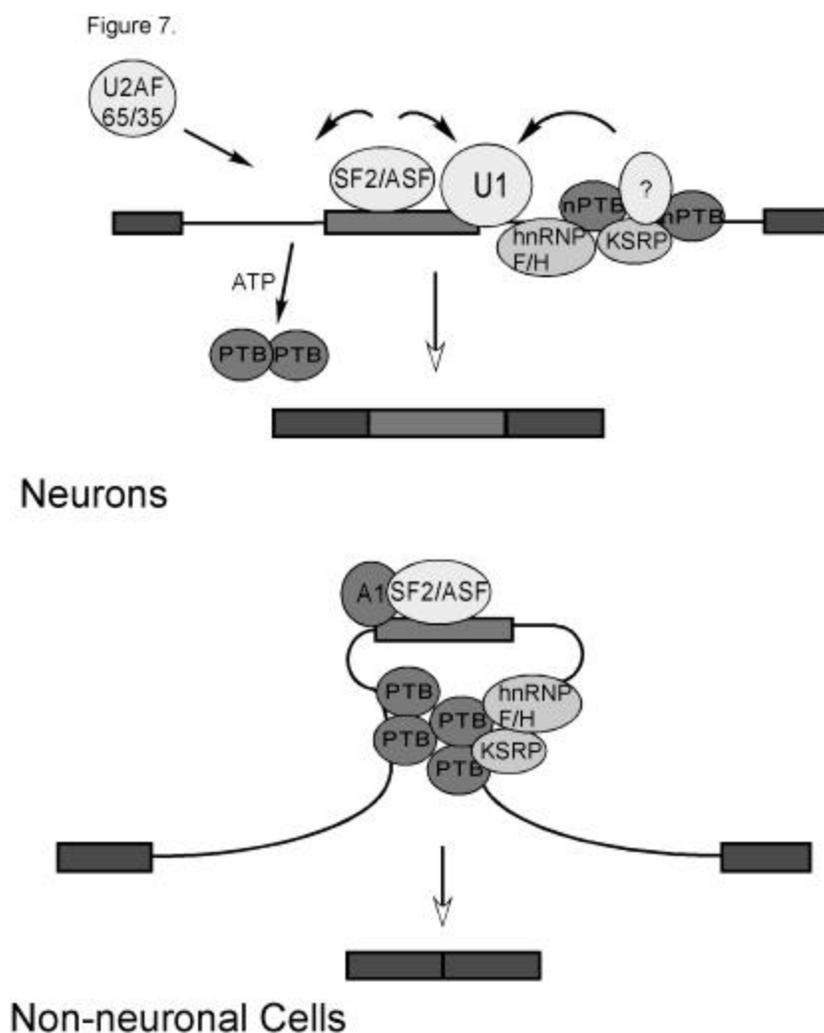


Figure 6.

**Figure 6. Models for Splicing Repression by hnRNP A1.**

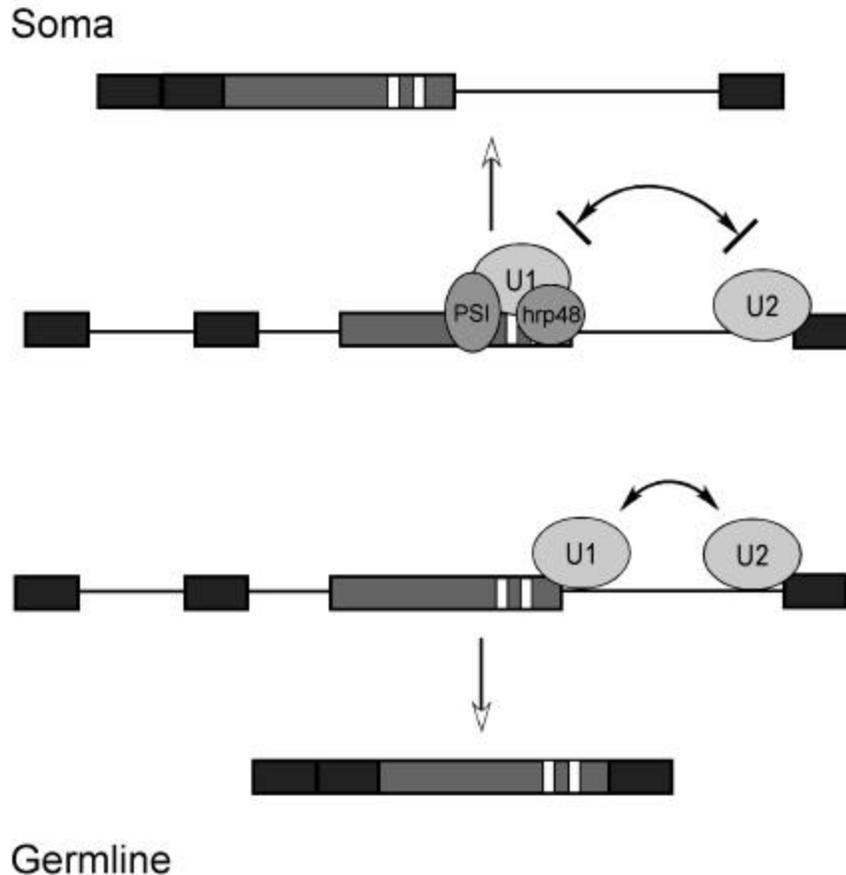
- A. In HIV Tat exon 3, A1 binds specifically to an exonic splicing silencer element. This is thought to nucleate the assembly of additional A1 molecules along the RNA, creating a zone of RNA where spliceosome assembly is repressed. The A1 repression can be blocked by the strong binding of SR proteins to ESE's within the exon. This presumably also stimulates spliceosome assembly at the upstream 3' splice site, allowing exon inclusion.
- B. In HIV Tat exon 3, there is an additional A1 binding site within the upstream intron adjacent to the branchpoint. This site blocks splicing in conjunction with the exonic A1 sites. A1 binding to this intronic element does not block U2AF binding to the 3' splice site, but does block U2 snRNP binding to the branch point and the formation of a spliceosomal A complex.
- C. HnRNP A1 also represses an exon in its own transcript using intronic binding sites. A1 bound to these sites can multimerize, thus looping out the exon and causing exon skipping.



**Figure 7. Combinatorial Control of Splicing in the c-src N1 exon.**

Tissue specific exons use a combination of positive and negative inputs to maintain their regulation. The N1 exon of the c-src gene is repressed by the PTB protein in non-neuronal cells. This PTB must bind to silencer elements on both sides of the exon and is thought to form a looping complex similar to hnRNP A1. PTB binds elements in the downstream intron within a large complex, containing the KSRP and hnRNP F/H proteins. In neurons, the PTB is replaced with the related nPTB protein and PTB binding to the upstream silencer elements is destabilized. There is an ATP dependent activity that removes the PTB from these upstream elements allowing use of this splice site. Under these conditions, the downstream regulatory region can act as a splicing enhancer. The stimulatory proteins for this enhancer are not yet known, but require the UGCAUG element. The N1 exon itself also contains regulatory elements that bind to the ASF/SF2 and hnRNP A1 proteins. Similar models have been developed for most other tissue specific exons that have been examined.

Figure 8.



**Figure 8. Regulation of Drosophila P Element Splicing.**

The third intron of the Drosophila P Element transcript is retained in the mRNA in somatic cells. In the Germline this intron is excised to make an mRNA encoding the P Element Transposase protein. The intron is repressed in somatic cells by the presence of the P element Somatic Inhibitor Protein (PSI). PSI can specifically interact with the U1 70K protein. PSI binds in the exon upstream from the retained intron in conjunction with U1 snRNP binding to two pseudo-5' splice sites. The protein hrp48 is also in this complex. The binding of U1 to the pseudo-5' splice sites prevents its binding to the correct site downstream and somehow blocks splicing. In the Germline, the absence of PSI allows normal splicing of the intron.

**Table of Known Splicing Regulators.**

| Protein         | Other Names        | Homologs             | Species    | Target Transcripts                    | Spliceosomal Target | Other Interactors  | Activator or Repressor | Domains     | References                   |
|-----------------|--------------------|----------------------|------------|---------------------------------------|---------------------|--------------------|------------------------|-------------|------------------------------|
| <b>Sxl</b>      |                    |                      | Dm         | Sxl, Tra, Msl2                        | U1/U2 via Snf       | Itself             | Repressor              | RNPcs       | (34)                         |
| <b>Tra</b>      |                    |                      | Dm         | dsx, Msl2                             | U2AF, U1            | Tra-2, SR Proteins | Activator              | RS          | (34)                         |
| <b>Tra-2</b>    |                    |                      | multiple   | multiple                              | U2AF                | SR Proteins        | Activator              | RNPcs<br>RS | (34)                         |
| <b>ASF/SF2</b>  |                    | SR Proteins          | multiple   | many                                  | U2AF, U1 snRNP      | Tra-2              | Activator              | RNPcs<br>RS | (80)                         |
| <b>TIA-1</b>    |                    | TIAR-1<br>NAM8       | multiple   | Msl2, FGFR2                           | U1                  |                    | Activator              | RNPcs       | (50)                         |
| <b>hnRNP A1</b> |                    | hnRNP A2, B<br>hrp48 | multiple   | many                                  | SR Proteins<br>U2AF | Itself             | Repressor              | RNPcs       | (136, 168, 205)              |
| <b>PTB</b>      | hnRNP I            | nPTB                 | multiple   | multiple                              |                     | Napor1<br>Nova1    | Repressor              | RNPcs       | (212)                        |
| <b>PSI</b>      |                    | FBP<br>KSRP          | multiple   | P element<br>Squid                    | U1                  | hrp48              | Repressor              | KH          | (260)                        |
| <b>ETR3</b>     | CELF2<br>Napor 1   | Multiple             | vertebrate | cTNT, NMDAR1,<br>ClChannel 1, Insu.R. |                     | PTB                | Activator              | RNPcs       | (180) (366, 367)             |
| <b>Nova 1</b>   |                    | Nova 2               | vertebrate | GlyR $\alpha$ 2 GABA <sub>A</sub> R?2 |                     | nPTB               | Activator              | KH          | (248)                        |
| <b>YB 1</b>     |                    |                      | vertebrate | CD44                                  |                     | p72                | Activator              | cold shock  | (105)                        |
| <b>hnRNP H</b>  |                    | hnRNP F<br>hnRNP H'  | vertebrate | HIV, src, $\beta$ Tm                  |                     | hnRNP F            | Both                   | RNPcs       | (169, 170, 173,<br>201, 368) |
| <b>p72</b>      |                    | p68                  | multiple   | CD44                                  |                     | YB1                | Activator              | DEAD Box    | (104)                        |
| <b>SRp38</b>    | SRrp35, 40<br>TASR |                      | vertebrate | many?                                 |                     |                    | Repressor              | RNPcs<br>RS | (90) (92)                    |
| <b>SRrp86</b>   |                    |                      | vertebrate |                                       | SR Proteins         |                    | Both                   | RNPcs<br>RS | (89)                         |
| <b>RSF1</b>     |                    |                      | Dm         |                                       | SR Proteins         |                    | Repressor              | RNPcs<br>RS | (91)                         |
| <b>ELAV</b>     |                    | Hu Proteins          | multiple   | Neuroglian Erectwing<br>Armadillo     |                     |                    | Repressor              | RNPcs       | (182)                        |
| <b>Mec8</b>     |                    |                      | C. Elegans | unc52                                 |                     |                    |                        | RNPcs       | (250)                        |
| <b>Fox1</b>     |                    |                      | C. Elegans | Xol-1                                 |                     |                    |                        | RNPcs       | (369)                        |
| <b>Quaking</b>  |                    | SAM68, SLM2          | vertebrate | MAG, PLP, MBP                         |                     |                    | Repressor              | KH          | (370) (371)                  |
| <b>SWAP</b>     |                    |                      | Dm         | Itself                                |                     |                    |                        | RS          | (372)                        |
| <b>Halfpint</b> |                    | FIR, PUF60           | multiple   | Otu, eIF4E, APP                       | U2AF                | Enc                | Activator              | RNPcs       | (249)                        |
| <b>p32</b>      |                    |                      | vertebrate | Adenovirus                            | SR Proteins         |                    | Repressor              |             | (373)                        |

**REFERENCES**

1. Grabowski PJ, Black DL. 2001. *Prog Neurobiol* 65: 289-308.
2. Modrek B, Lee C. 2002. *Nat Genet* 30: 13-9.
3. Black DL. 2000. *Cell* 103: 367-70
4. Graveley BR. 2001. *Trends Genet* 17: 100-7
5. Smith CW, Nadal-Ginard B. 1989. *Cell* 56: 749-58
6. Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, et al. 2000. *Cell* 101: 671-84
7. Colgan DF, Manley JL. 1997. *Genes Dev* 11: 2755-66.
8. Baker BS. 1989. *Nature* 340: 521-4
9. Black DL. 1998. *Neuron* 20: 165-8
10. Burgess RW, Nguyen, Q.T., Son, Y.J., Lichtman, J.W., Sanes, J.R. 1999. *Neuron* 23: 33-44.
11. Burke JF, Bright KE, Kellett E, Benjamin PR, Saunders SE. 1992. *Progress in Brain Research* 92: 115-25
12. Cartegni L, Chew SL, Krainer AR. 2002. *Nat Rev Genet* 3: 285-98.
13. Cooper TA, Mattox W. 1997. *Am J Hum Genet* 61: 259-66
14. Eckardt NA. 2002. *Plant Cell* 14: 743-7.
15. Jiang ZH, Wu JY. 1999. *Proceedings of the Society for Experimental Biology and Medicine* 220: 64-72
16. Matera AG. 1999. *Current Biology* 9: R140-2
17. Philips AV, Cooper TA. 2000. *Cellular and Molecular Life Sciences* 57: 235-49
18. Schutt C, Nothiger R. 2000. *Development* 127: 667-77
19. Caceres JF, Kornblihtt AR. 2002. *Trends Genet.* 18: 186-93
20. Blencowe BJ. 2000. *Trends Biochem. Sci.* 25: 106-10
21. Burge CB, Tuschl, T., and Sharp, P.A. 1999.: pp 525-60.
22. Nilsen TW. 2002. *Mol Cell* 9: 8-9.
23. Staley JP, Guthrie C. 1998. *Cell* 92: 315-26
24. Das R, Zhou Z, Reed R. 2000. *Mol Cell* 5: 779-87.
25. Nilsen TW. 1996. *Science* 273: 1813.

26. Lallena MJ, Chalmers KJ, Llamazares S, Lamond AI, Valcarcel J. 2002. *Cell* 109: 285-96.
27. Berget SM. 1995. *J Biol Chem* 270: 2411-4
28. Black DL. 1995. *Rna* 1: 763-71
29. Hoffman BE, Grabowski PJ. 1992. *Genes & Development* 6: 2554-68
30. Robberson BL, Cote GJ, Berget SM. 1990. *Mol Cell Biol* 10: 84-94.
31. Libri D, Piseri A, Fiszman MY. 1991. *Science* 252: 1842-5.
32. Libri D, Stutz F, McCarthy T, Rosbash M. 1995. *Rna* 1: 425-36.
33. Jacquenet S, Ropers D, Bilodeau PS, Damier L, Mouglin A, et al. 2001. *Nucleic Acids Res* 29: 464-78.
34. Lopez AJ. 1998. *Annu Rev Genet* 32: 279-305
35. Varani G, Nagai K. 1998. *Annual Review of Biophysics and Biomolecular Structure* 27: 407-45
36. Antson AA. 2000. *Curr Opin Struct Biol* 10: 87-94.
37. Crowder SM, Kanaar R, Rio DC, Alber T. 1999. *Proc. Natl. Acad. Sci. U.S.A.* 96: 4892-7
38. Handa N, Nureki O, Kurimoto K, Kim I, Sakamoto H, et al. 1999. *Nature* 398: 579-85
39. Wang J, Bell LR. 1994. *Genes Dev* 8: 2072-85
40. Sosnowski BA, Belote JM, McKeown M. 1989. *Cell* 58: 449-59
41. Inoue K, Hoshijima K, Sakamoto H, Shimura Y. 1990. *Nature* 344: 461-3
42. Valcarcel J, Singh R, Zamore PD, Green MR. 1993. *Nature* 362: 171-5
43. Granadino B, Penalva LO, Green MR, Valcarcel J, Sanchez L. 1997. *Proc Natl Acad Sci U S A* 94: 7343-8.
44. Granadino B, Penalva LO, Sanchez L. 1996. *Mol Gen Genet* 253: 26-31
45. Hilfiker A, Amrein H, Dubendorfer A, Schneiter R, Nothiger R. 1995. *Development* 121: 4017-26
46. Kelley RL, Solovyeva I, Lyman LM, Richman R, Solovyev V, Kuroda MI. 1995. *Cell* 81: 867-77.
47. Zhou S, Yang Y, Scott MJ, Pannuti A, Fehr KC, et al. 1995. *Embo J* 14: 2884-95.
48. Bashaw GJ, Baker BS. 1995. *Development* 121: 3245-58.
49. Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J. 1999. *Nature* 402: 838-41.

50. Forch P, Puig O, Kedersha N, Martinez C, Granneman S, et al. 2000. *Mol Cell* 6: 1089-98.
51. Forch P, Merendino L, Martinez C, Valcarcel J. 2001. *Rna* 7: 1185-91.
52. Bashaw GJ, Baker BS. 1997. *Cell* 89: 789-98.
53. Kelley RL, Wang J, Bell L, Kuroda MI. 1997. *Nature* 387: 195-9
54. Gebauer F, Merendino L, Hentze MW, Valcarcel J. 1998. *Rna* 4: 142-50
55. Gebauer F, Merendino L, Hentze MW, Valcarcel J. 1997. *Semin Cell Dev Biol* 8: 561-6.
56. Cote CA, Gautreau D, Denegre JM, Kress TL, Terry NA, Mowry KL. 1999. *Mol Cell* 4: 431-7
57. Gu W, Pan F, Zhang H, Bassell GJ, Singer RH. 2002. *J Cell Biol* 156: 41-51.
58. Kaminski A, Jackson RJ. 1998. *Rna* 4: 626-38
59. Hughson FM, Schedl P. 1999. *Nat Struct Biol* 6: 499-502.
60. Horabin JI, Schedl P. 1996. *Rna* 2: 1-10
61. Zhu C, Urano J, Bell LR. 1997. *Mol Cell Biol* 17: 1674-81
62. Horabin JI, Schedl P. 1993. *Mol Cell Biol* 13: 7734-46
63. Penalva LO, Lallena MJ, Valcarcel J. 2001. *Mol Cell Biol* 21: 1986-96.
64. Sakamoto H, Inoue K, Higuchi I, Ono Y, Shimura Y. 1992. *Nucleic Acids Res* 20: 5533-40
65. Salz HK, Flickinger TW. 1996. *Genetics* 144: 95-108.
66. Salz HK. 1992. *Genetics* 130: 547-54.
67. Graveley BR. 2002. *Cell* 109: 409-12.
68. Hertel KJ, Lynch KW, Maniatis T. 1997. *Curr Opin Cell Biol* 9: 350-7
69. Smith CW, Valcarcel J. 2000. *Trends Biochem Sci* 25: 381-8
70. Wang J, Manley JL. 1997. *Curr Opin Genet Dev* 7: 205-11
71. Coschigano KT, Wensink PC. 1993. *Genes Dev* 7: 42-54.
72. Burtis KC. 2002. *Science* 297: 1135-6.
73. Tian M, Maniatis T. 1994. *Genes Dev* 8: 1703-12
74. Tian M, Maniatis T. 1993. *Cell* 74: 105-14
75. Tian M, Maniatis T. 1992. *Science* 256: 237-40
76. Lynch KW, Maniatis T. 1996. *Genes Dev* 10: 2089-101
77. Hertel KJ, Lynch KW, Hsiao EC, Liu EH, Maniatis T. 1996. *Rna* 2: 969-81
78. Zuo P, Maniatis T. 1996. *Genes Dev* 10: 1356-68
79. Graveley BR, Hertel KJ, Maniatis T. 2001. *Rna* 7: 806-18.

80. Graveley BR. 2000. *Rna* 6: 1197-211
81. Ryner LC, Goodwin SF, Castrillon DH, Anand A, Vilella A, et al. 1996. *Cell* 87: 1079-89.
82. Heinrichs V, Ryner LC, Baker BS. 1998. *Mol Cell Biol* 18: 450-8
83. Anand A, Vilella A, Ryner LC, Carlo T, Goodwin SF, et al. 2001. *Genetics* 158: 1569-95.
84. Schaal TD, Maniatis T. 1999. *Mol Cell Biol* 19: 261-73
85. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. 2002. *Science* 297: 1007-13.
86. Caceres JF, Krainer AR. 1997. *Eukaryotic mRNA processing*: Oxford University Press, Oxford, UK, pp. 174-217
87. Manley JL, Tacke R. 1996. *Genes Dev* 10: 1569-79
88. Zahler AM, Lane WS, Stolk JA, Roth MB. 1992. *Genes Dev* 6: 837-47.
89. Barnard DC, Li J, Peng R, Patton JG. 2002. *Rna* 8: 526-33.
90. Cowper AE, Caceres JF, Mayeda A, Sreaton GR. 2001. *J Biol Chem* 276: 48908-14.
91. Labourier E, Bourbon HM, Gallouzi IE, Fostier M, Allemand E, Tazi J. 1999. *Genes Dev* 13: 740-53.
92. Yang L, Embree LJ, Tsai S, Hickstein DD. 1998. *J Biol Chem* 273: 27761-4.
93. Longman D, Johnstone IL, Caceres JF. 2000. *Embo J* 19: 1625-37.
94. Ring HZ, Lis JT. 1994. *Mol Cell Biol* 14: 7499-506.
95. Wang HY, Xu X, Ding JH, Bermingham JR, Jr., Fu XD. 2001. *Molecular Cell* 7: 331-42
96. Wang J, Takagaki Y, Manley JL. 1996. *Genes Dev* 10: 2588-99
97. Liu HX, Zhang M, Krainer AR. 1998. *Genes Dev* 12: 1998-2012
98. Tacke R, Manley JL. 1999. *Curr. Opin. Cell. Biol.* 11: 358-62
99. Cartegni L, Krainer AR. 2002. *Nat Genet* 30: 377-84.
100. Tian H, Kole R. 1995. *Mol Cell Biol* 15: 6291-8.
101. Tian H, Kole R. 2001. *J Biol Chem* 276: 33833-9.
102. Woerfel G, Bindereif A. 2001. *Nucleic Acids Res* 29: 3204-11.
103. Coulter LR, Landree MA, Cooper TA. 1997. *Mol Cell Biol* 17: 2143-50
104. Honig A, Auboeuf D, Parker MM, O'Malley BW, Berget SM. 2002. *Mol Cell Biol* 22: 5698-707.
105. Stickeler E, Fraser SD, Honig A, Chen AL, Berget SM, Cooper TA. 2001. *Embo J* 20: 3821-30.

106. Chandler SD, Mayeda A, Yeakley JM, Krainer AR, Fu XD. 1997. *Proc Natl Acad Sci U S A* 94: 3596-601.
107. Mayeda A, Sreaton GR, Chandler SD, Fu XD, Krainer AR. 1999. *Mol Cell Biol* 19: 1853-63.
108. Tacke R, Manley JL. 1995. *Embo J* 14: 3540-51
109. Zuo P, Manley JL. 1993. *Embo J* 12: 4727-37
110. Graveley BR, Maniatis T. 1998. *Mol Cell* 1: 765-71
111. Wang J, Xiao SH, Manley JL. 1998. *Genes Dev* 12: 2222-33
112. Zhu J, Krainer AR. 2000. *Genes Dev* 14: 3166-78.
113. Prasad J, Colwill K, Pawson T, Manley JL. 1999. *Mol Cell Biol* 19: 6991-7000.
114. Kanopka A, Muhlemann O, Petersen-Mahrt S, Estmer C, Ohrmalm C, Akusjarvi G. 1998. *Nature* 393: 185-7
115. Sanford JR, Bruzik JP. 1999. *Genes Dev* 13: 1513-8.
116. Mermoud JE, Cohen PT, Lamond AI. 1994. *Embo J* 13: 5679-88.
117. Mermoud JE, Cohen P, Lamond AI. 1992. *Nucleic Acids Res* 20: 5263-9.
118. Roscigno RF, Garcia-Blanco MA. 1995. *Rna* 1: 692-706.
119. Xiao SH, Manley JL. 1997. *Genes Dev* 11: 334-44
120. Kohtz JD, Jamison SF, Will CL, Zuo P, Luhrmann R, et al. 1994. *Nature* 368: 119-24
121. Wu JY, Maniatis T. 1993. *Cell* 75: 1061-70
122. Yeakley JM, Tronchere H, Olesen J, Dyck JA, Wang HY, Fu XD. 1999. *J Cell Biol* 145: 447-55.
123. Dauwalder B, Mattox W. 1998. *Embo J* 17: 6049-60
124. Xiao SH, Manley JL. 1998. *Embo J* 17: 6359-67
125. Wang Z, Hoffmann HM, Grabowski PJ. 1995. *Rna* 1: 21-35
126. Lavigne A, La Branche H, Kornblihtt AR, Chabot B. 1993. *Genes Dev* 7: 2405-17.
127. Bourgeois CF, Popielarz M, Hildwein G, Stevenin J. 1999. *Mol Cell Biol* 19: 7347-56.
128. Cote J, Simard MJ, Chabot B. 1999. *Nucleic Acids Research* 27: 2529-37
129. Selvakumar M, Helfman DM. 1999. *Rna* 5: 378-94.
130. Rudner DZ, Breger KS, Rio DC. 1998. *Genes Dev* 12: 1010-21.
131. Blencowe BJ, Issner R, Nickerson JA, Sharp PA. 1998. *Genes Dev* 12: 996-1009.
132. Li Y, Blencowe BJ. 1999. *J Biol Chem* 274: 35074-9.

133. Eldridge AG, Li Y, Sharp PA, Blencowe BJ. 1999. *Proc Natl Acad Sci U S A* 96: 6125-30.
134. Hertel KJ, Maniatis T. 1998. *Mol Cell* 1: 449-55
135. Kan JL, Green MR. 1999. *Genes Dev* 13: 462-71
136. Zhu J, Mayeda A, Krainer AR. 2001. *Mol Cell* 8: 1351-61
137. Sacco-Bubulya P, Spector DL. 2002. *J Cell Biol* 156: 425-36.
138. Sanford JR, Bruzik JP. 2001. *Proc Natl Acad Sci U S A* 98: 10184-9.
139. Yeakley JM, Tronchere H, Olesen J, Dyck JA, Wang HY, Fu XD. 1999. *Journal of Cell Biology* 145: 447-55
140. Siebel CW, Feng L, Guthrie C, Fu XD. 1999. *Proc Natl Acad Sci U S A* 96: 5440-5.
141. Yun CY, Fu XD. 2000. *Journal of Cell Biology* 150: 707-18
142. Wang HY, Arden KC, Bermingham JR, Jr., Viars CS, Lin W, et al. 1999. *Genomics* 57: 310-5.
143. Wang HY, Lin W, Dyck JA, Yeakley JM, Songyang Z, et al. 1998. *J Cell Biol* 140: 737-50.
144. Gui JF, Tronchere H, Chandler SD, Fu XD. 1994. *Proc Natl Acad Sci U S A* 91: 10824-8.
145. Gui JF, Lane WS, Fu XD. 1994. *Nature* 369: 678-82.
146. Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, et al. 1996. *Embo J* 15: 265-75
147. Tacke R, Chen Y, Manley JL. 1997. *Proc Natl Acad Sci U S A* 94: 1148-53
148. Du C, McGuffin ME, Dauwalder B, Rabinow L, Mattox W. 1998. *Molecular Cell* 2: 741-50
149. Krecic AM, Swanson MS. 1999. *Curr Opin Cell Biol* 11: 363-71
150. Ford LP, Wright WE, Shay JW. 2002. *Oncogene* 21: 580-3.
151. LaBranche H, Dupuis S, Ben-David Y, Bani MR, Wellinger RJ, Chabot B. 1998. *Nat Genet* 19: 199-202.
152. Dallaire F, Dupuis S, Fiset S, Chabot B. 2000. *J Biol Chem* 275: 14509-16.
153. Buvoli M, Cobianchi F, Bestagno MG, Mangiarotti A, Bassi MT, et al. 1990. *EMBO J*. 9: 1229-35
154. Burd CG, Dreyfuss G. 1994. *EMBO J*. 13: 1197-204
155. Mayeda A, Munroe SH, Caceres JF, Krainer AR. 1994. *Embo J* 13: 5483-95

156. Ding J, Hayashi MK, Zhang Y, Manche L, Krainer AR, Xu RM. 1999. *Genes Dev* 13: 1102-15.
157. Allain FH, Bouvet P, Dieckmann T, Feigon J. 2000. *Embo J* 19: 6870-81.
158. Mayeda A, Helfman DM, Krainer AR. 1993. *Mol Cell Biol* 13: 2993-3001
159. Mayeda A, Krainer AR. 1992. *Cell* 68: 365-75
160. Caceres JF, Stamm S, Helfman DM, Krainer AR. 1994. *Science* 265: 1706-9
161. Hanamura A, Caceres JF, Mayeda A, Franza BR, Jr., Krainer AR. 1998. *Rna* 4: 430-44.
162. Caputi M, Mayeda A, Krainer AR, Zahler AM. 1999. *EMBO J.* 18: 4060-7
163. Del Gatto-Konczak F, Olive M, Gesnel MC, Breathnach R. 1999. *Mol Cell Biol* 19: 251-60
164. Tange TO, Kjems J. 2001. *J. Mol. Biol.* 312: 649-62
165. Staffa A, Acheson NH, Cochrane A. 1997. *J Biol Chem* 272: 33394-401.
166. Amendt BA, Si ZH, Stoltzfus CM. 1995. *Mol Cell Biol* 15: 6480.
167. Amendt BA, Hesslein D, Chang LJ, Stoltzfus CM. 1994. *Mol Cell Biol* 14: 3960-70.
168. Tange TO, Damgaard CK, Guth S, Valcarcel J, Kjems J. 2001. *Embo J* 20: 5748-58.
169. Jacquenet S, Mereau A, Bilodeau PS, Damier L, Stoltzfus CM, Branlant C. 2001. *J Biol Chem* 276: 40464-75.
170. Chen CD, Kobayashi R, Helfman DM. 1999. *Genes and Development* 13: 593-606
171. Caputi M, Zahler AM. 2001. *J. Biol. Chem.* 276: 43850-9
172. Markovtsov V, Nikolic JM, Goldman JA, Turck CW, Chou MY, Black DL. 2000. *Mol. Cell. Biol.* 20: 7463-79
173. Caputi M, Zahler AM. 2002. *Embo J.* 21: 845-55
174. Hastings ML, Wilson CM, Munroe SH. 2001. *RNA* 7: 859-74
175. Gallego ME, Gattoni R, Stevenin J, Marie J, Expert-Bezancon A. 1997. *Embo J* 16: 1772-84
176. Del Gatto-Konczak F, Bourgeois CF, Le Guiner C, Kister L, Gesnel MC, et al. 2000. *Molecular and Cellular Biology* 20: 6287-99
177. Puig O, Gottschalk A, Fabrizio P, Seraphin B. 1999. *Genes Dev* 13: 569-80.
178. Spingola M, Ares M, Jr. 2000. *Mol Cell* 6: 329-38.

179. Kedersha NL, Gupta M, Li W, Miller I, Anderson P. 1999. *J Cell Biol* 147: 1431-42.
180. Ladd AN, Charlet N, Cooper TA. 2001. *Mol Cell Biol* 21: 1285-96.
181. Timchenko LT, Miller JW, Timchenko NA, DeVore DR, Datar KV, et al. 1996. *Nucleic Acids Res* 24: 4407-14
182. Lisbin MJ, Qiu J, White K. 2001. *Genes Dev* 15: 2546-61.
183. Koushika SP, Soller M, White K. 2000. *Molecular and Cellular Biology* 20: 1836-45
184. Koushika SP, Lisbin MJ, White K. 1996. *Curr Biol* 6: 1634-41
185. Good PJ, Chen Q, Warner SJ, Herring DC. 2000. *J Biol Chem* 275: 28583-92.
186. Charlet-B. N, Logan P, Singh G, Cooper TA. 2002. *Mol. Cell.* 9: 649-58
187. Choi DK, Ito T, Tsukahara F, Hirai M, Sakaki Y. 1999. *Gene* 237: 135-42.
188. Zhang W, Liu H, Han K, Grabowski PJ. 2002. *Rna* 8: 671-85.
189. Huh GS, Hynes RO. 1994. *Genes Dev* 8: 1561-74
190. Modafferi EF, Black DL. 1997. *Mol Cell Biol* 17: 6537-45
191. Lim LP, Sharp PA. 1998. *Mol Cell Biol* 18: 3900-6
192. Black DL. 1992. *Cell* 69: 795-807
193. Hedjran F, Yeakley JM, Huh GS, Hynes RO, Rosenfeld MG. 1997. *Proc Natl Acad Sci U S A* 94: 12343-7
194. Huh GS, Hynes RO. 1993. *Mol Cell Biol* 13: 5301-14
195. Guo N, Kawamoto S. 2000. *J Biol Chem* 275: 33641-9
196. Brudno M, Gelfand MS, Spengler S, Zorn M, Dubchak I, Conboy JG. 2001. *Nucleic Acids Res* 29: 2338-48.
197. Modafferi EF, Black DL. 1999. *Rna* 5: 687-706
198. Chan RC, Black DL. 1997. *Molecular and Cellular Biology* 17: 2970
199. Min H, Chan RC, Black DL. 1995. *Genes Dev* 9: 2659-71
200. Min H, Turck CW, Nikolic JM, Black DL. 1997. *Genes Dev* 11: 1023-36
201. Chou MY, Rooke N, Turck CW, Black DL. 1999. *Molecular and Cellular Biology* 19: 69-77
202. Chou MY, Underwood JG, Nikolic J, Luu MH, Black DL. 2000. *Mol Cell* 5: 949-57
203. Kanopka A, Muhlemann O, Akusjarvi G. 1996. *Nature* 381: 535-8
204. Simard MJ, Chabot B. 2002. *Mol Cell Biol* 22: 4001-10.
205. Blanchette M, Chabot B. 1999. *EMBO J.* 18: 1939-52

206. Xu RM, Jokhan L, Cheng X, Mayeda A, Krainer AR. 1997. *Structure* 5: 559-70
207. Patton JG, Mayer SA, Tempst P, Nadal-Ginard B. 1991. *Genes Dev* 5: 1237-51
208. Garcia-Blanco MA, Jamison SF, Sharp PA. 1989. *Genes Dev* 3: 1874-86
209. Gil A, Sharp PA, Jamison SF, Garcia-Blanco MA. 1991. *Genes Dev* 5: 1224-36
210. Ghetti A, Pinol-Roma S, Michael WM, Morandi C, Dreyfuss G. 1992. *Nucleic Acids Res* 20: 3671-8
211. Valcarcel J, Gebauer F. 1997. *Curr Biol* 7: R705-R8
212. Wagner EJ, Garcia-Blanco MA. 2001. *Mol Cell Biol* 21: 3281-8.
213. Singh R, Valcarcel J, Green MR. 1995. *Science* 268: 1173-6
214. Perez I, Lin CH, McAfee JG, Patton JG. 1997. *Rna* 3: 764-78
215. Pérez I, McAfee JG, Patton JG. 1997. *Biochemistry* 36: 11881-90
216. Oh YL, Hahm B, Kim YK, Lee HK, Lee JW, et al. 1998. *Biochem J* 331: 169-75
217. Conte MR, Grune T, Ghuman J, Kelly G, Ladas A, et al. 2000. *Embo J* 19: 3132-41
218. Liu H, Zhang W, Reed RB, Liu W, Grabowski PJ. 2002. *Rna* 8: 137-49.
219. Southby J, Gooding C, Smith CWJ. 1999. *Molecular and Cellular Biology* 19: 2699-711.
220. Mulligan GJ, Guo W, Wormsley S, Helfman DM. 1992. *J Biol Chem* 267: 25480-7
221. Chan RC, Black DL. 1997. *Mol Cell Biol* 17: 4667-76
222. Gooding C, Roberts GC, Moreau G, Nadal-Ginard B, Smith CW. 1994. *Embo J* 13: 3861-72
223. Gooding C, Roberts GC, Smith CW. 1998. *Rna* 4: 85-100
224. Zhang L, Liu W, Grabowski PJ. 1999. *Rna* 5: 117-30
225. Lillevali K, Kulla A, Ord T. 2001. *Mech Dev* 101: 217-20.
226. Wagner EJ, Carstens RP, Garcia-Blanco MA. 1999. *Electrophoresis* 20: 1082-6
227. Markovtsov V, Nikolic JM, Goldman JA, Turck CW, Chou MY, Black DL. 2000. *Mol Cell Biol* 20: 7463-79
228. Polydorides AD, Okano HJ, Yang YY, Stefani G, Darnell RB. 2000. *Proceedings of the National Academy of Sciences of the United States of America* 97: 6350-5
229. Wollerton MC, Gooding C, Robinson F, Brown EC, Jackson RJ, Smith CW. 2001. *Rna* 7: 819-32.
230. Ashiya M, Grabowski PJ. 1997. *Rna* 3: 996-1015

231. Southby J, Gooding C, Smith CW. 1999. *Mol Cell Biol* 19: 2699-711
232. Carstens RP, Wagner EJ, Garcia-Blanco MA. 2000. *Mol Cell Biol* 20: 7388-400
233. Expert-Bezancon A, Le Caer JP, Marie J. 2002. *J Biol Chem* 277: 16614-23.
234. Cooper TA, Ordahl CP. 1985. *J Biol Chem* 260: 11140-8.
235. Levy JB, Dorai T, Wang LH, Brugge JS. 1987. *Mol. Cell. Biol.* 7: 4142-5
236. Martinez R, Mathey-Prevot B, Bernardis A, Baltimore D. 1987. *Science* 237: 411-5
237. Black DL. 1991. *Genes Dev* 5: 389-402
238. Dominski Z, Kole R. 1991. *Molecular and Cellular Biology* 11: 6075-83
239. Xu R, Teng J, Cooper TA. 1993. *Mol Cell Biol* 13: 3660-74
240. Dominski Z, Kole R. 1992. *Molecular and Cellular Biology* 12: 2108-14
241. Ramchatesingh J, Zahler AM, Neugebauer KM, Roth MB, Cooper TA. 1995. *Mol Cell Biol* 15: 4898-907
242. Ryan KJ, Cooper TA. 1996. *Mol Cell Biol* 16: 4014-23
243. Cooper TA. 1998. *Mol Cell Biol* 18: 4519-25
244. Chan RC, Black DL. 1995. *Molecular and Cellular Biology* 15: 6377-85
245. Buckanovich RJ, Posner JB, Darnell RB. 1993. *Neuron* 11: 657-72
246. Yang YY, Yin GL, Darnell RB. 1998. *Proc Natl Acad Sci U S A* 95: 13254-9
247. Buckanovich RJ, Darnell RB. 1997. *Molecular and Cellular Biology* 17: 3194-201
248. Jensen KB, Dredge BK, Stefani G, Zhong R, Buckanovich RJ, et al. 2000. *Neuron* 25: 359-71
249. Van Buskirk C, Schupbach T. 2002. *Dev Cell* 2: 343-53.
250. Lundquist EA, Herman RK, Rogalski TM, Mullen GP, Moerman DG, Shaw JE. 1996. *Development* 122: 1601-10
251. Zorio DA, Blumenthal T. 1999. *Rna* 5: 487-94.
252. Rio DC. 1991. *Trends Genet* 7: 282-7.
253. Siebel CW, Fresco LD, Rio DC. 1992. *Genes Dev* 6: 1386-401.
254. Siebel CW, Kanaar R, Rio DC. 1994. *Genes Dev* 8: 1713-25
255. Siebel CW, Admon A, Rio DC. 1995. *Genes Dev* 9: 269-83
256. Adams MD, Tarnng RS, Rio DC. 1997. *Genes Dev* 11: 129-38
257. Duncan R, Bazar L, Michelotti G, Tomonaga T, Krutzsch H, et al. 1994. *Genes Dev* 8: 465-80

258. Chen CY, Gherzi R, Ong SE, Chan EL, Raijmakers R, et al. 2001. *Cell* 107: 451-64.
259. Labourier E, Adams MD, Rio DC. 2001. *Mol Cell* 8: 363-73.
260. Labourier E, Blanchette M, Feiger JW, Adams MD, Rio DC. 2002. *Genes Dev* 16: 72-84.
261. Kiss-Laszlo Z, Hohn T. 1996. *Trends Microbiol* 4: 480-5.
262. Boris-Lawrie K, Roberts TM, Hull S. 2001. *Life Sci* 69: 2697-709.
263. Fischer U, Pollard VW, Luhrmann R, Teufel M, Michael MW, et al. 1999. *Nucleic Acids Res* 27: 4128-34.
264. Pollard VW, Malim MH. 1998. *Annu Rev Microbiol* 52: 491-532
265. Cullen BR. 1998. *Virology* 249: 203-10.
266. Yang J, Bogerd HP, Wang PJ, Page DC, Cullen BR. 2001. *Mol Cell* 8: 397-406.
267. Yang J, Cullen BR. 1999. *Rna* 5: 1645-55.
268. Pasquinelli AE, Ernst RK, Lund E, Grimm C, Zapp ML, et al. 1997. *Embo J* 16: 7500-10.
269. Ernst RK, Bray M, Rekosh D, Hammarskjold ML. 1997. *Mol Cell Biol* 17: 135-44.
270. Gruter P, Taberero C, von Kobbe C, Schmitt C, Saavedra C, et al. 1998. *Mol Cell* 1: 649-59.
271. Saavedra C, Felber B, Izaurralde E. 1997. *Curr Biol* 7: 619-28.
272. Katz RA, Skalka AM. 1990. *Mol Cell Biol* 10: 696-704.
273. Fu XD, Katz RA, Skalka AM, Maniatis T. 1991. *Genes Dev* 5: 211-20.
274. Bouck J, Fu XD, Skalka AM, Katz RA. 1998. *J Biol Chem* 273: 15169-76.
275. Pongoski J, Asai K, Cochrane A. 2002. *J Virol* 76: 5108-20.
276. McNally MT, Beemon K. 1992. *J Virol* 66: 6-11.
277. Ogert RA, Lee LH, Beemon KL. 1996. *J Virol* 70: 3834-43.
278. McNally LM, McNally MT. 1998. *Mol Cell Biol* 18: 3103-11.
279. McNally LM, McNally MT. 1999. *J Virol* 73: 2385-93.
280. Fogel BL, McNally MT. 2000. *J Biol Chem* 275: 32371-8.
281. Gontarek RR, McNally MT, Beemon K. 1993. *Genes Dev* 7: 1926-36.
282. Hibbert CS, Gontarek RR, Beemon KL. 1999. *Rna* 5: 333-43.
283. Bennett M, Piñol-Roma S, Staknis D, Dreyfuss G, Reed R. 1992. *Mol Cell Biol* 12: 3165-75

284. Mourelatos Z, Dostie J, Paushkin S, Sharma A, Charroux B, et al. 2002. *Genes Dev* 16: 720-8.
285. Ambros V. 2001. *Cell* 107: 823-6.
286. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. *Science* 294: 853-8.
287. Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. *Science* 294: 858-62.
288. Osheim YN, Miller OL, Jr., Beyer AL. 1985. *Cell* 43: 143-51.
289. Misteli T, Spector DL. 1999. *Mol Cell* 3: 697-705.
290. Wetterberg I, Bauren G, Wieslander L. 1996. *Rna* 2: 641-51.
291. Hirose Y, Manley JL. 2000. *Genes Dev* 14: 1415-29.
292. Proudfoot NJ, Furger A, Dye MJ. 2002. *Cell* 108: 501-12.
293. Bentley D. 1999. *Curr Opin Cell Biol* 11: 347-51.
294. Bentley D. 2002. *Curr Opin Cell Biol* 14: 336-42.
295. Maniatis T, Reed R. 2002. *Nature* 416: 499-506.
296. Cramer P, Caceres JF, Cazalla D, Kadener S, Muro AF, et al. 1999. *Mol Cell* 4: 251-8.
297. Kadener S, Fededa JP, Rosbash M, Kornblihtt AR. 2002. *Proc Natl Acad Sci U S A* 99: 8185-90.
298. Kadener S, Cramer P, Nogues G, Cazalla D, de la Mata M, et al. 2001. *Embo J* 20: 5759-68.
299. Eperon LP, Graham IR, Griffiths AD, Eperon IC. 1988. *Cell* 54: 393-401.
300. Roberts GC, Gooding C, Mak HY, Proudfoot NJ, Smith CW. 1998. *Nucleic Acids Res* 26: 5568-72.
301. Ghosh S, Garcia-Blanco MA. 2000. *Rna* 6: 1325-34.
302. Thummel CS. 1992. *Science* 255: 39-40.
303. Hatton AR, Subramaniam V, Lopez AJ. 1998. *Mol Cell* 2: 787-96.
304. Corden JL, Patturajan M. 1997. *Trends Biochem Sci* 22: 413-6.
305. Morris DP, Greenleaf AL. 2000. *J Biol Chem* 275: 39935-43.
306. Robert F, Blanchette M, Maes O, Chabot B, Coulombe B. 2002. *J Biol Chem* 277: 9302-6.
307. McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, et al. 1997. *Nature* 385: 357-61.
308. Fong N, Bentley DL. 2001. *Genes Dev* 15: 1783-95.
309. Komarnitsky P, Cho EJ, Buratowski S. 2000. *Genes Dev* 14: 2452-60.

310. Hirose Y, Tacke R, Manley JL. 1999. *Genes Dev* 13: 1234-9.
311. Zeng C, Berget SM. 2000. *Mol Cell Biol* 20: 8290-301.
312. Fong YW, Zhou Q. 2001. *Nature* 414: 929-33.
313. Monsalve M, Wu Z, Adelmant G, Puigserver P, Fan M, Spiegelman BM. 2000. *Mol Cell* 6: 307-16.
314. Lai MC, Teh BH, Tarn WY. 1999. *J Biol Chem* 274: 11832-41.
315. Bickmore WA, Oghene K, Little MH, Seawright A, van Heyningen V, Hastie ND. 1992. *Science* 257: 235-7
316. Davies RC, Calvio C, Bratt E, Larsson SH, Lamond AI, Hastie ND. 1998. *Genes Dev* 12: 3217-25
317. Larsson SH, Charlier JP, Miyagawa K, Engelkamp D, Rassoulzadegan M, et al. 1995. *Cell* 81: 391-401
318. Reed R, Hurt E. 2002. *Cell* 108: 523-31.
319. Luo MJ, Reed R. 1999. *Proc Natl Acad Sci U S A* 96: 14937-42
320. Le Hir H, Gatfield D, Izaurralde E, Moore MJ. 2001. *Embo J* 20: 4987-97.
321. Le Hir H, Moore MJ, Maquat LE. 2000. *Genes Dev* 14: 1098-108.
322. Le Hir H, Izaurralde E, Maquat LE, Moore MJ. 2000. *Embo J* 19: 6860-9.
323. Luo ML, Zhou Z, Magni K, Christoforides C, Rappsilber J, et al. 2001. *Nature* 413: 644-7.
324. Palacios IM. 2002. *Curr Biol* 12: R50-2.
325. Berke JD, Sgambato V, Zhu P, Lavoie B, Vincent M, et al. 2001. *Neuron* 32: 277-87
326. Chalfant CE, Mischak H, Watson JE, Winkler BC, Goodnight J, et al. 1995. *Journal of Biological Chemistry* 270: 13326-32
327. Collett JW, Steele RE. 1993. *Developmental Biology* 158: 487-95
328. Rodger J, Davis S, Laroche S, Mallet J, Hicks A. 1998. *Journal of Neurochemistry* 71: 666-75
329. Shifrin VI, Neel BG. 1993. *Journal of Biological Chemistry* 268: 25376-84
330. Smith MA, Fanger GR, O'Connor LT, Bridle P, Maue RA. 1997. *Journal of Biological Chemistry* 272: 15675-81
331. Wang A, Cohen DS, Palmer E, Sheppard D. 1991. *Journal of Biological Chemistry* 266: 15598-601
332. Zacharias DA, Strehler EE. 1996. *Current Biology* 6: 1642-52
333. Stamm S. 2002. *Hum Mol Genet* 11: 2409-16.

334. Herrlich P, Morrison H, Sleeman J, Orian-Rousseau V, Konig H, et al. 2000. *Ann N Y Acad Sci* 910: 106-18; discussion 18-20.
335. Stickeler E, Kittrell F, Medina D, Berget SM. 1999. *Oncogene* 18: 3574-82.
336. Konig H, Ponta H, Herrlich P. 1998. *Embo Journal* 17: 2904-13
337. Weg-Remers S, Ponta H, Herrlich P, Konig H. 2001. *EMBO* 20: 4194-203
338. Lynch KW, Weiss A. 2000. *Molecular and Cellular Biology* 20: 70-80
339. Matter N, Marx M, Weg-Remers S, Ponta H, Herrlich P, Konig H. 2000. *J Biol Chem* 275: 35353-60.
340. Daoud R, Da Penha Berzaghi M, Siedler F, Hubener M, Stamm S. 1999. *European Journal of Neuroscience* 11: 788-802
341. Vallano ML, Lambolez B, Audinat E, Rossier J. 1996. *Journal of Neuroscience* 16: 631-9
342. Vallano ML, Beaman-Hall CM, Benmansour S. 1999. *Neuroreport* 10: 3659-64.
343. Xie J, Black DL. 2001. *Nature* 410: 936-9.
344. van der Houven van Oordt W, Diaz-Meco MT, Lozano J, Krainer AR, Moscat J, Caceres JF. 2000. *Journal of Cell Biology* 149: 307-16
345. Daoud R, Mies G, Smialowska A, Olah L, Hossmann KA, Stamm S. 2002. *J Neurosci* 22: 5889-99.
346. Cyert MS. 2001. *J Biol Chem* 276: 20805-8.
347. Vandromme M, Gauthier-Rouviere C, Lamb N, Fernandez A. 1996. *Trends Biochem Sci* 21: 59-64.
348. Robertson KD, Jones PA. 1999. *Oncogene* 18: 3810-20.
349. Sawa H, Ohshima TA, Ukita H, Murakami H, Chiba Y, et al. 1998. *Oncogene* 16: 1701-12.
350. Roberts GC, Smith CW. 2002. *Curr Opin Chem Biol* 6: 375-83.
351. Marden JH, Fitzhugh GH, Girgenrath M, Wolf MR, Girgenrath S. 2001. *Journal of Experimental Biology* 204: 3457-70
352. Gavrillov DK, Shi X, Das K, Gilliam TC, Wang CH. 1998. *Nat Genet* 20: 230-1.
353. Nissim-Rafinia M, Kerem B. 2002. *Trends Genet* 18: 123-7.
354. Clark TA, Sugnet CW, Ares M, Jr. 2002. *Science* 296: 907-10.
355. Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engel P, McDonagh PD. 2001. *Nature* 409: 922-7
356. Yeakley JM, Fan JB, Doucet D, Luo L, Wickham E, et al. 2002. *Nat Biotechnol* 20: 353-8.

357. Grabowski P. 2002. *Nat Biotechnol* 20: 346-7.
358. Brown V, Jin P, Ceman S, Darnell JC, O'Donnell WT, et al. 2001. *Cell* 107: 477-87.
359. Rappsilber J, Ryder U, Lamond AI, Mann M. 2002. *Genome Res* 12: 1231-45.
360. Neubauer G, King A, Rappsilber J, Calvio C, Watson M, et al. 1998. *Nat Genet* 20: 46-50.
361. Will CL, Luhrmann R. 1997. *Curr Opin Cell Biol* 9: 320-8.
362. Gottschalk A, Neubauer G, Banroques J, Mann M, Luhrmann R, Fabrizio P. 1999. *Embo J* 18: 4535-48.
363. Zhou Z, Licklider LJ, Gygi SP, Reed R. 2002. *Nature* 419: 182-5.
364. Stevens SW, Ryan DE, Ge HY, Moore RE, Young MK, et al. 2002. *Mol Cell* 9: 31-44.
365. Maroney PA, Romfo CM, Nilsen TW. 2000. *Mol Cell* 6: 317-28.
366. Charlet N, Savkur RS, Singh G, Philips AV, Grice EA, Cooper TA. 2002. *Mol Cell* 10: 45-53
367. Savkur RS, Philips AV, Cooper TA. 2001. *Nat Genet.* 29: 40-7
368. Fogel BL, McNally MT. 2000. *J. of Biol. Chem.* 275: 32371-8
369. Nicoll M, Akerib CC, Meyer BJ. 1997. *Nature* 388: 200-4.
370. Wu JI, Reed RB, Grabowski PJ, Artzt K. 2002. *Proc Natl Acad Sci U S A* 99: 4233-8.
371. Stoss O, Olbrich M, Hartmann AM, Konig H, Memmott J, et al. 2001. *J Biol Chem* 276: 8665-73.
372. Zachar Z, Chou TB, Kramer J, Mims IP, Bingham PM. 1994. *Genetics* 137: 139-50.
373. Petersen-Mahrt SK, Estmer C, Ohrmalm C, Matthews DA, Russell WC, Akusjarvi G. 1999. *Embo J* 18: 1014-24.