Alternatively Spliced Genes

Jane Y. Wu[1,2], *Liya Yuan*[1] *and Necat Havlioglu*[1]
[1] *Washington University School of Medicine, St. Louis, MO, USA*
[2] *John F. Kennedy Center for Research on Human Development, Vanderbilt University Medical Center, Nashville, TN, USA*

Keywords

Pre-mRNA
Nascent transcripts that are precursors of mature messenger RNAs.

Exon
Gene regions that are present in mature messenger RNA transcripts.

Intron
Intervening sequences that are removed from pre-mRNA transcripts and not represented in mature mRNA species.

Pre-mRNA Splicing
The process of removing introns from pre-mRNAs and ligating exons to form mature mRNA transcripts.

UsnRNPs
Uridine-rich small nuclear ribonucleoprotein particles.

Spliceosome
The macromolecular machine in which pre-mRNA splicing reactions take place.

5′ and 3′ Splice Sites
Sequences that spliceosomes recognize at 5′ and 3′ ends of an intron. They are also named splice donor and splice acceptor sites respectively.

Alternative Splicing
The process in which cells selectively use different combinations of splice sites to generate two or more mRNA transcripts from a single pre-mRNA species.

mcb200300189

Splicing Regulatory Elements
Specific sequences in introns or exons that are recognized by trans-acting factors to allow either enhancement or repression of splicing.

Splicing Regulators
Trans-acting factors that interact with splicing regulatory elements to either stimulate or inhibit splicing.

An important step of eukaryotic gene expression is pre-mRNA splicing, the process of removing intervening sequences (introns) from the nascent transcript (messenger RNA precursor, or pre-mRNA) (as depicted in •Fig. 1). The discovery of split genes in the viral genome and subsequent research in the field of pre-mRNA splicing have greatly advanced our understanding of mammalian gene regulation. Studies on pre-mRNA splicing have also facilitated sequence analyses of the human genome. With the completion of human genome sequencing, it is now further appreciated that pre-mRNA splicing and alternative splicing play critical roles in regulating gene expression and in enhancing genetic diversity.

Evolutionarily, the basic machinery for pre-mRNA splicing appears to be highly conserved among different species of metazoans. In *Saccharomyces cerevisiae*, although only a small percentage of genes undergo splicing, more than 100 genes have been identified that are either dedicated to or involved in pre-mRNA splicing. In mammals, pre-mRNA splicing is a crucial step for gene expression because the vast majority of mammalian transcription units contain one or more introns that must be accurately removed to form mature and functional messenger RNA (mRNA) species. In this chapter, we review the current knowledge about mammalian pre-mRNA splicing, with special emphasis on the aspects related to the pathogenesis or treatment of human diseases.

Q1

1 Pre-mRNA Splicing and Splicing Machinery

1.1 Splicing Machinery: Spliceosome

The biochemical reactions of pre-mRNA splicing occur in a macromolecular machine named the *spliceosome*. This large RNA–protein complex contains, in addition to the pre-mRNA substrate, several uridine-rich small nuclear ribonucleoprotein (UsnRNP) particles as well as a myriad of non-snRNP protein factors. The splicing machinery is similar in complexity and size to that of the protein synthesis machinery, the ribosome. In all cases studied so far, splicing reactions take place in the mature spliceosome inside the nucleus.

For processing the majority of introns (the major class, also called the U2-type intron), the spliceosome contains U1, U2, U4/6, and U5snRNPs. The splicing of the minor class of introns (also called

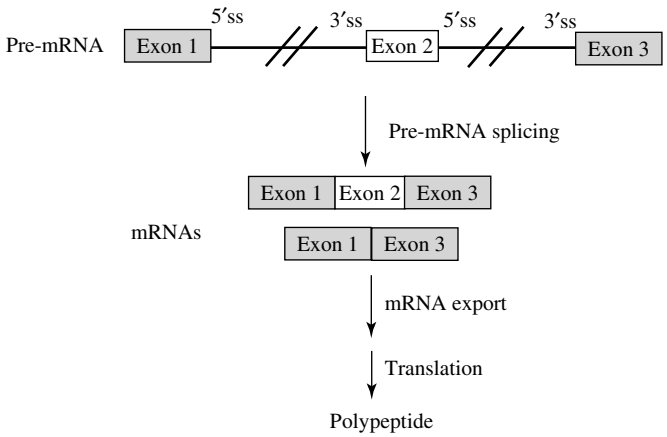

Fig. 1 The schematics for major posttranscriptional processes in eukaryotic gene expression. Different splice isoforms of mRNAs can be produced as a result of alternative selection of splice sites.

the U12-type) occurs in the spliceosome containing U11 and U12, in addition to U4atac, U6atac, and U5snRNPs (Figs. 2 and 3). This chapter will focus on the splicing of introns in the major class. This is because this class accounts for more than 99% of known introns. Most of our knowledge about pre-mRNA splicing has also come from studies of the major class of introns.

Comparison of genes coding for components of splicing machinery, including both snRNAs and protein factors, reveals a high degree of conservation through evolution. Amazing similarity is found among spliceosomal components from yeast, fruit fly, and human. A recent study using nanoscale microcapillary liquid chromatography tandem mass spectrometry to analyze partially purified human spliceosomes assembled on model splicing substrates revealed 145 distinct spliceosomal proteins. This suggests that the spliceosome is one of the most complex macromolecular machines for mammalian gene expression. Table 1 shows the currently known human proteins identified in functional spliceosomes assembled on model splicing substrates. The role of individual proteins in splicesome assembly and in splicing regulation will be discussed.

1.2 Splicing Signals

The sites for cleavage and ligation in splicing reactions are defined by conserved splicing signals in the pre-mRNA, including the 5′ splice site (5′ss), the branch site, the polypyrimidine tract, and the 3′ splice site (3′ss) consensus sequence. These cis-elements are important sites for RNA–RNA and protein–RNA interactions. Systematic analyses of both yeast and higher eukaryotic splicing signal sequences led to the classification of U2 and U12 types of introns based on the key UsnRNPs involved in the recognition of the branch sites in the corresponding introns (Fig. 2).

In general, the spliceosome for U2-type introns contains U1, U2, and U4/U6,

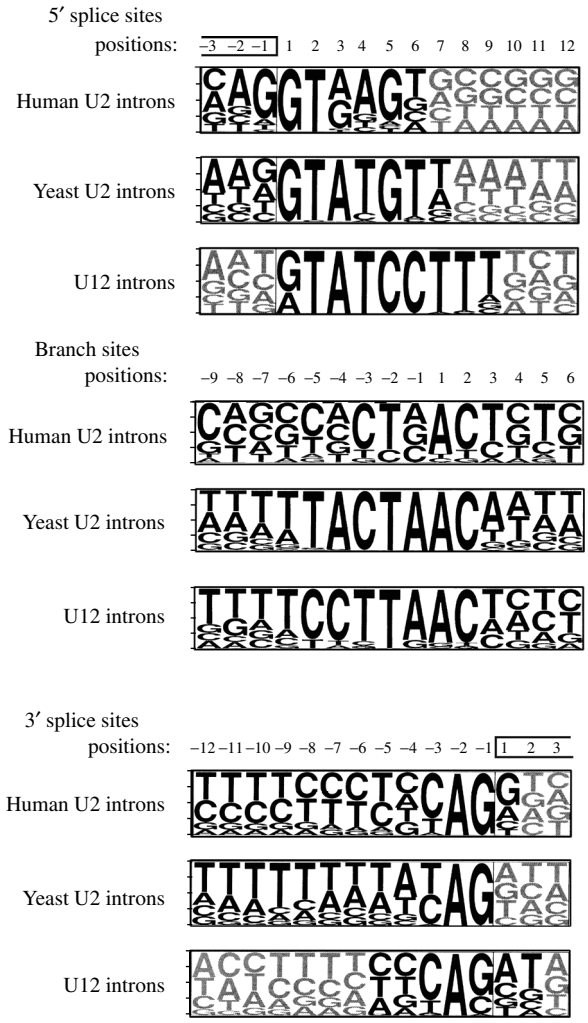

Fig. 2 The sequence features at the 5′ splice site, branch site, and 3′ splice site as derived from representative sets of introns of corresponding types. (Modified with permission from Burge, C.B., Tuschl, T.H., Sharp, P.A. (1999). in: The RNA World, Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds) *The RNA World*, Cold Spring Harbor Laboratory Press, New York, pp. 525–560). The frequency of each nucleotide (A, G, T, or C) at each sequence position is depicted by the heights of the corresponding letters in the diagram.

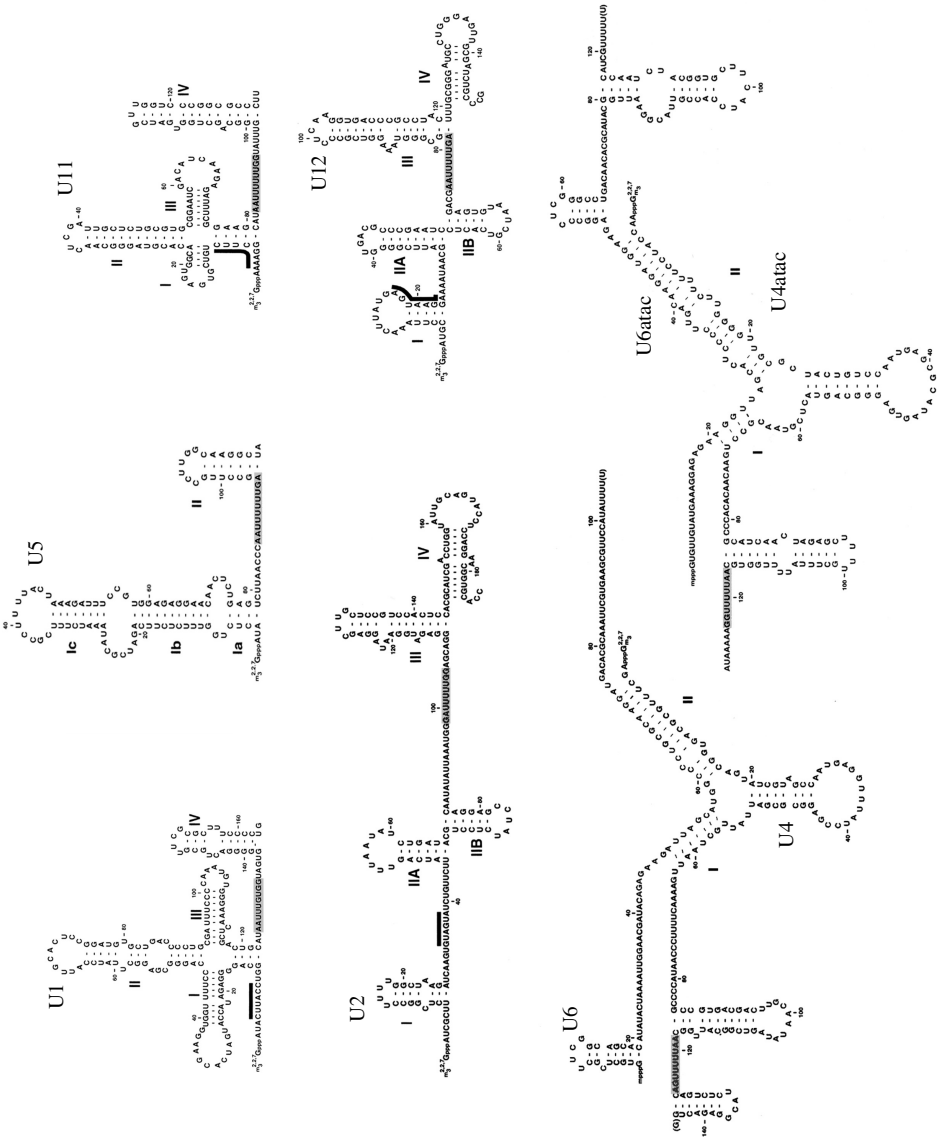
U1
U5
U11
U12
U2
U6
U4
U6atac
U4atac

Fig. 3 Sequences of human spliceosomal UsnRNAs and their interactions with splice sites. Panel (a) Sequences and secondary structures of spliceosomal snRNAs: U1, U2, U4/6 (for the U2 type), and U11, U12, U4atac/U6atac snRNA (for the U12 type) as well as U5snRNAs (for both classes). The 5′ to 3′ orientation of U4 and U4atac is from right to left, and that of U1, U11, U2, U12, and U5snRNAs is from left to right. The RNA helices are indicated by Roman numerals, and nucleotide positions by Arabic numerals. The shaded nucleotides are binding sites for Sm proteins, and regions marked with black lines form base-pairing interactions with the pre-mRNA. (Modified with permission from Burge, C.B., Tuschl, T.H., Sharp, P.A. (1999). in: The RNA World, Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds) *The RNA World*, Cold Spring Harbor Laboratory Press, New York, pp. 525–560). •Panel (b) Base-pairing interactions in the U2 and U12 types of spliceosome. The exon-intron boundaries at 5′ splice sites are marked by arrowheads. During early stages of spliceosome assembly, U1 or U11snRNP interacts with the 5′ splice site, and U2 or U12snRNP with the branch sites. These interactions are displaced in mature spliceosome by U6 or U6atac associated with U2 and U12 respectively. In these drawings, the distances between interacting nucleotides do not reflect corresponding physical distances in the spliceosome. (Adapted •from Tarn and Steitz, 1996b.)

Tab. 1 Proteins• detected in human spliceosomes.

Acce.#	*Protein name*	*Yeast homolog (SGD ORF)*	*Motifs*	*Cal. MWt*
A. snRNP proteins				
Sm/LSm core proteins				
P14678	Sm B/B'	SMB1 (YER029C)	Sm 1 and 2	24 610
P13641	Sm D1	SMD1 (YGR074W)	Sm 1 and 2	13 282
P43330	Sm D2	SMD2 (YLR275W)	Sm 1 and 2	13 527
P43331	Sm D3	SMD3 (YLR147C)	Sm 1 and 2	13 916
P08578	Sm E	SME1 (YOR159C)	Sm 1 and 2	10 804
Q15356	Sm F	SMX3 (YPR182W)	Sm 1 and 2	9725
Q15357	Sm G	SMX2 (YFL017W-A)	Sm 1 and 2	8496
Q9Y333	hLSm2	LSM2 (YBL026W)	Sm 1 and 2	10 835
Q9Y4Z1	hLSm3	LSM3 (YLR438C-A)	Sm 1 and 2	11 714
Q9Y4Z0	hLSm4	LSM4 (YER112W)	Sm 1 and 2	15 350
Q9Y4Y8	hLSm6	LSM6 (YDR378C)	Sm 1 and 2	9128
Q9UK45	hLSm7	LSM7 (YNL147W)	Sm 1 and 2	11 602
U1snRNP-specific proteins				
P08621	U1-70kD	SNP1 (YIL061C)	RS, 1 RRM	70 081
P09012	U1 A	MUD1 (YBR119W)	2 RRMs	31 280
P09234	U1 C	YHC1 (YLR298C)	ZF	17 394
AB034205	LUC7A	LUC7 (YDL087C)	RS-rich	51 466
U2snRNP-specific proteins				
AF054284	SAP155 (SF3b155)	HSH155 (YMR288W)	PP2A	1 45 815
Q13435	SAP145 (SF3b145)	CUS1 (YMR240C)	Pro-rich	97 657
AJ001443	SAP130 (SF3b130)	RSE1 (YML049C)		1 35 592
Q15459	SAP114 (SF3a120)	PRP21 (YJL203W)	SWAP, UBQ	88 886
Q15428	SAP62 (SF3a66)	PRP11 (YDL043C)	ZF	49 196
A55749	SAP61(SF3a60)	PRP9 (YDL030W)	ZF	58 849
Q15427	SAP49 (SF3b49)	HSH49 (YOR319W)	2 RRMs	44 386

(*continued overleaf*)

Tab. 1 (*continued*)

Acce.#	*Protein name*	*Yeast homolog (SGD ORF)*	*Motifs*	*Cal. MWt*
P09661	U2 A′	LEA1 (YPL213W)	Leu-rich	28 444
P08579	U2 B″	MSL1 (YIR009W)	2 RRMs	25 486
Q9Y3B4	p14	SNU17 (YIR005W)	1 RRM	14 584
U5snRNP-specific proteins				
AB007510	U5-220kD (PRPC8)	PRP8 (YHR165C)		2 73 785
O75643	U5-200kD	BRR2 (YER172C)	DExD, Helicase	1 94 479
BC002360	U5-116kD	SNU114 (YKL173W)	G domain	1 09 478
BC001666	U5-102kD	PRP6 (YBR055C)	HAT	1 06 925
BC002366	U5-100kD	PRP28 (YDR243C)	DExD, Helicase, RS	95 583
BC000495	U5-52kD	SNU40 (YHR156C)	GYF	37 646
AF090988	U5-40kD		WD40s	39 299
O14834	U5-15kD	DIB1 (YPR082C)		16 786
U4/U6snRNP-specific proteins				
T50839	U4/U6-90kD (HPRP3)	PRP3 (YDR473C)	PWI	77 529
BC007424	U4/U6-60kD (HPRP4)	PRP4 (YPR178W)	WD40s	58 321
AL050369	U4/U6-61kD (PRPF31)	PRP31 (YGR091W)	NOP	55 424
AF036331	U4/U6-20kD	CPH1 (YDR155C)	Cyclophilin	19 207
P55769	U4/U6-15.5kD	SNU13 (YEL026W)	Putative RRM	14 174
U4/U6.U5 tri-snRNP specific proteins				
T00034	Tri-snRNP 110kD	SNU66 (YOR308C)	RS	90 255
AF353989	Tri-snRNP 65kD	SAD1 (YFR005C)	RS, UBQ, ZF	65 415
X76302	Tri-snRNP 27kD		RS	18 859
B. Non-snRNP spliceosomal proteins				
SR proteins				
Q08170	SRp75		2 RRMs, RS	56 792
Q05519	p54/SFRS11		1 RRM, RS	53 542
Q13247	SRp55		2 RRMs, RS	39 568
Q13243	SRp40		2 RRMs, RS	31 264
Q07955	ASF/SF2		2 RRMs, RS	27 613
Q16629	9G8		1 RRM, RS	27 367
Q01130	SC35		1 RRM, RS	25 575
Q13242	SRp30c		2 RRMs, RS	25 542
AF057159	hTra2		1 RRM, RS	21 935
P23152	SRp20		1 RRM, RS	19 330
Other spliceosomal proteins with known motifs				
NP_057417	SRm300		RS-rich	2 51 965
P26368	U2 AF65	MUD2 (YKL074C)	RS, 3 RRMs	53 501
Q01081	U2AF35		RS	27 872
Y08765	SF1	BBP (Ylr116w)	ZF, KH	68 632
AB018344	hPrp5	PRP5 (YBR237W)	DExD, Helicase, RS	1 17 461
O60231	hPrp2	PRP2 (YNR011C)	DExD, HELICc	1 19 172
P38919	IF4N	FAL1 (YDR021W)	DExD, HELICc	46 833

Tab. 1 (*continued*)

Acce.#	*Protein name*	*Yeast homolog (SGD ORF)*	*Motifs*	*Cal. MWt*
P17844	p68	DBP2 (YNL112W)	DExD, Helicase	69 148
Q92841	p72	DBP2 (YNL112W)	DExD, HELICc	72 371
T00333	fSAP164		DExD	1 63 986
Q08211	RHA	(YLR419W)	DExD, HELICc, DSRM	1 40 877
O00571	DDX3	DBP1 (YPL119C)	DExD, HELICc, SR	73 243
Q9UJV9	Abstrakt		DExD, HELICc	69 738
P42285	fSAP118	MTR4 (YJL050W)	DExD	1 17 790
BC002548	fSAP113		DExD	1 13 671
NP_055792	fSAP152 (Acinus)		1 RRM	1 51 887
Q14011	CIRP (cold inducible RNA-binding)		1 RRM, Gly-rich	18 648
BC007871	SPF45		1 RRM, Gly-rich	44 962
NP_008938	CF I-68kD (cleavage factor)		1 RRM	59 209
AC004858	fSAP94		1 RRM, PWI	94 122
P52298	hCBP20	CBP20 (YPL178W)	1 RRM	18 001
BC003402	fSAP47	ECM2 (YBR065C)	1 RRM, ZF	46 896
BC006474	fSAPa		1 RRM, SWAP	Partial
AAB18823	TAT-SF1	CUS2 (Ynl286w)	2 RRMs	85 759
I55595	fSAP59 (transcription cofactor)		2 RRMs	58 657
U76705	IMP3		2 RRMs, 4 KH	63 720
NP_073605	OTT		3 RRMs	1 02 135
BC008875	PUF60		3 RRMs	58 171
Q15097	PAB2	PAB1 (YER165W)	4 RRMs, PABP	58 518
AF356524	SHARP (transcription cofactor)		4 RRMs	4 02 248
P41223	fSAP17	CWC14 (YCR063W)	ZF	16 844
O43670	ZNF207 (transcription factor)		ZF	50 751
AF044333	PLRG1	PRP46 (YPL151C)	WD40s	57 194
BC008719	hPrp19	PRP19 (YLL036C)	WD40s	55 181
BC006849	hTEX1	TEX1 (YNL253W)	WD40s	38 772
BC002876	fSAP57	PFS2 (YNL317W)	WD40s	57 544
BC003118	fSAP35		WD40s	34 849
AF083383	SPF38		WD40s	34 290
Other spliceosomal proteins with known motifs				
XM_087118	hFBP3		1 WW	24 297
AF049523	hFBP11	PRP40 (YKL012W)	2 WW, 1 FF	Partial
T08599	CA150		3 WW, 6 FF	1 23 960
XP_042023	CyP64	CPR3 (YML078W)	Cyclophilin, WD40	64 222
AF271652	PPIL3	CPR6 (YLR216C)*	Cyclophilin	18 155
S64705	PPIL2	CYP5 (YDR304C)*	Cyclophilin, Ubox	58 823

(*continued overleaf*)

Tab. 1 (*continued*)

Acce.#	*Protein name*	*Yeast homolog (SGD ORF)*	*Motifs*	*Cal. MWt*
BC003048	PPIL1	CYP2 (YHR057C)*	Cyclophilin	18 237
Q9UNP9	PPIE	CPH1 (YDR155C)*	Cyclophilin	33 431
AY029347	hPrp4 kinase		TyrKc	1 16 973
Q9NYV4	CrkRS	CTK1 (YKL139W)	TyrKc	1 64 155
T12531	TIP39	(YLR424W)*	Gly-rich	96 820
BC001403	CF I-25kD (cleavage factor)		NUDIX	26 227
Q09161	hCBP80	CBP80 (YMR125W)	MIF4G	91 839
AB046824	fSAPb	CWC22 (YGR278W)	MIF4G	Partial
T02672	fSAPc		ERM	Partial
AF255443	hCrn	CLF1 (YLR117C)	HAT	99 201
BC007208	XAB2 (transcription cofactor)	SYF1 (YDR416w)	HAT	1 00 010
Q13573	SKIP	PRP45 (YAL032C)	SKIP	61 494
A53545	hHPR1	HPR1 (YDR138W)	DEATH	75 627
AF083385	SPF30**		TUDOR	26 711
NP_055095	SPF31		DnaJ	30 986
Q13123	RED		RED	65 630
Spliceosomal proteins without known motifs				
NP_001244	CDC5L	CEF1 (YMR213W)		92 250
AAK21005	ASR2			1 00 665
AF441770	hTHO2	THO2 (YNL139C)		1 69 573
I39463	fSAP79			78 536
AK027098	fSAP24			23 671
BC001621	SNP70			69 998
BC006350	fSAP71			70 521
P55081	MFAP1			51 855
AJ276706	WTAP			44 243
O15355	PP2Cγ	PTC3 (YBL056W)		59 272
AJ279080	fSAP105 (putative transcription factor)			1 04 804
NP_056311	fSAP121			1 21 193
BC004442	fSAP33	ISY1 (YJR050W)		32 992
AF161497	CCAP2	CWC15 (YDR163W)		26 610
Q9Y5B6	GCFC (putative transcription factor)			29 010
NP_073210	DGSI			52 567
NP_110517	NAP (nuclear protein inducing cell death)			65 173
NP_056299	fSAP29			28 722
NP_037374	fSAP23 (putative transcription factor)			22 774
BC004122	fSAP11			10 870

Tab. 1 (*continued*)

Acce.#	*Protein name*	*Yeast homolog (SGD ORF)*	*Motifs*	*Cal. MWt*
BC000216	fSAP18			18 419
AF081788	SPF27			26 131
C. Late-acting spliceosomal proteins				
Catalytic step II and late-acting proteins				
Q92620	hPrp16	PRP16 (YKR086W)	DExD, Helicase, RS	1 40 473
Q14562	hPrp22	PRP22 (YER013W)	DExD, Helicase, RS	1 39 315
O43143	hPrp43	PRP43 (YGL120C)	DExD, Helicase	92 829
BC010634	hSlu7	SLU7 (YDR088C)	ZF	68 343
AF038392	hPrp17	PRP17 (YDR364C)	WD40s	65 521
BC000794	hPrp18	PRP18 (YGR006W)		39 860
Spliced mRNP/EJC proteins				
AF048977	SRm160		PWI, RS	93 519
Q13838	UAP56	SUB2 (YDL084W)	DExD, Helicase	48 991
PJC4525	RNPS1		1 RRM, Ser-rich	34 208
NM_005782	Aly	YRA1 (YDR381W)	1 RRM	26 861
Q9Y5S9	Y14		1 RRM	19 889
P50606	Magoh		Mago_nashi	17 164

Source: (Adapted •with permission from Zhou et al., 2002).
Note: A. snRNP proteins. B. Non-snRNP spliceosomal proteins. C. Late-acting spliceosomal proteins. In each list, spliceosomal proteins are grouped by their structural motifs. Corresponding •GenBank accession numbers are shown (Acce#), together with their yeast homologs (as SGD open reading frame, SGD ORF), sequence motifs, and calculated molecular weight (Cal. Mwt).
COLD: cold shock RNA-binding domain; Cyclophilin: cyclophilin type peptidyl–prolyl cis–trans isomerase; DEATH: domain found in proteins involved in cell death; DExD: DExD/H-like helicases superfamily; DSRM: double-stranded RNA-binding motif; ERM: ezrin/radixin/moesin family motif; fSAP: functional spliceosome-associated protein; FF: two conserved F residues; G domain: GTP-binding domain that contains a P-loop motif; GYF: contains conserved G-T-F residues; HAT: Half-A-TPR (tetratrico-peptide repeat); Gly-rich: peptide sequences rich in G, R, S residues; HELICc: helicase superfamily c-terminal domain; KH: hnRNP K homology RNA-binding domain; MIF4G: middle domain of eukaryotic initiation factor 4G; NOP: putative snoRNA binding domain; NUDIX: mutT-like domain; PABP: poly-adenylate binding protein, unique domain; PP2A: protein phosphatase 2A repeat; PWI: domain in splicing factors; RED: protein with extensive stretch of alternating R and E or D; RRM: RNA recognition motif; RS: arginine–serine-rich domains; SKIP: conserved domain found in chromatinic proteins; Sm: snRNP Sm proteins; SWAP: suppressor-of-white-apricot splicing regulator; TUDOR: a domain present in several RNA-binding proteins; TyrKc: tyrosine kinase, catalytic domain; UBQ: ubiquitin homologs; Ubox: modified RING finger domain; WD40s: WD40 repeats, structural repeats of the beta propeller domain; WW: domain with 2 conserved W residues, interacting with proline-rich polypeptides; ZF: Zinc finger domain.

whereas the U12-type of spliceosome involves U11, U12, and U4atac/U6atac, with both sharing U5snRNP. Although nucleotide sequences of U11, U12, and U4atac/U6atac snRNAs are different from their counterparts in the spliceosomes for U2-type introns, the predicted secondary structures of the corresponding snRNAs have striking similarity (Fig. 3a). In addition, the specific interactions between the spliceosomal snRNAs and pre-mRNA substrates also appear to be

highly similar, as shown in Fig. 3(b). Both U2- and U12-types of introns can be found in the same genes. There is evidence suggesting that the two types of spliceosomes may interact and share some protein components in addition to the common U5snRNP. The evolutionary origins of these two types of spliceosomes remain largely speculative.

U12-class introns contain more conserved sequences at the 5′ss. The sequences around branch sites of the U12-class introns are also highly conserved with the TCCTTAAC (the underlined A as the branch site) consensus signal located approximately 10 to 20 nucleotides upstream of the 3′ss. A typical U12-class intron lacks a polypyrimidine tract between the branch site and the 3′ss.

In the yeast, *S. cerevisiae*, only a small fraction of genes contain introns, and these are usually short introns (approximately 240 introns, averaging 270 nucleotides in length). Splicing signals in yeast introns are highly conserved and may contain sufficient information for defining splice junctions, especially considering the number of genes devoted to or associated with pre-mRNA splicing in the yeast genome.

In mammals, pre-mRNA transcripts are usually much longer and contain multiple introns of variable sizes. In humans, the average size of exons is 150 nucleotides, and that of introns is approximately 3500 nucleotides. Mammalian introns can be as large as 500 kbp. The basic splicing signals in mammalian pre-mRNAs are degenerate, especially in the case of U2-type of introns. The branch sites for U2-type introns are highly divergent. At both the 5′ss and 3′ss, only two nucleotides (/GT at the 5′ss and AG/ at 3′ss) are highly conserved. As a result, the nucleotide sequences surrounding the splice junctions, the 5′ss and 3′ss, usually contain only a limited amount of information. This is not sufficient for conferring the specificity required to achieve accurate splice site selection. In mammals, the recognition of not only exon–intron junction sequences but also the regulatory elements in intronic and exonic regions is important for defining splice junctions and maintaining splicing fidelity. In addition, multiple networks of interactions among the machineries for transcription, cap formation, splicing, and polyadenylation may also influence splice site selection. This high degree of degeneracy in the splicing signals in mammalian pre-mRNA transcripts provides the flexibility for alternative selection and pairing of different splice sites, a fundamental mechanism for regulating alternative splicing.

1.3 Spliceosomal UsnRNP Biogenesis

As essential subunits of the splicing machinery, spliceosomal UsnRNPs contain not only uridine-rich snRNAs but also a number of polypeptides. U1, U2, U4, and U5snRNAs are transcribed by RNA polymerase II as precursors containing additional 3′ nucleotides. After acquiring a monomethylated guanosine (m7G) cap structure, these pre-UsnRNAs are exported to the cytoplasm in a pathway dependent on the m7G cap, the cap-binding complex (CBC), RanGTP, and phosphorylated adaptor for RNA export. In the cytoplasm, pre-UsnRNAs interact with Sm proteins including B/B′, D3, D2, D1, E, F, and G (Table 1A) to form the snRNP core structure.

The Sm protein-binding sites in U1, U2, U4, and U5snRNAs are highly conserved,

containing two stem-loop structures flanking PuAU4-6GPu sequence (Fig. 3a). A number of proteins interacting with Sm proteins have been identified, including the protein product of SMN (survival of motor neuron) gene, SMN-interacting protein/Gemin2, Gemin3, and Gemin4. Genetic defects in the SMN gene cause spinal muscular atrophy (SMA), possibly by interfering with UsnRNP Sm core assembly and therefore deficiency in UsnRNP biogenesis.

Following UsnRNP Sm core assembly, the m7G cap of their snRNA is converted to the 2,2,7-tri-methylated guanosine (m3G), and the 3′ extra nucleotides of pre-UsnRNAs are removed. These core UsnRNPs are imported into the nucleus to be assembled into the spliceosome. This UsnRNP import process requires not only general nuclear import factors such as Importin-β but also specific factors such as Snurportin-1 that recognizes the m3G cap of the UsnRNP and interacts with general factors for snRNP import into the nucleus.

Before the association of UsnRNP particle-specific proteins, the UsnRNAs undergo internal modifications including pseudouridylation and 2′-O-methylation. Such modifications appear to be necessary for the assembly of a functional UsnRNP, as shown for the human U2snRNP. These posttranscriptional modifications are mediated by a small nucleolar (sno)RNA-guided mechanism through the action of snoRNPs. For example, U85 snoRNP directs both 2′-O-methylation and pseudouridylation of U5snRNA. Following these posttranscriptional internal modifications of snRNAs, the assembly of UsnRNPs is completed with the association with individual UsnRNP-specific proteins (Table 1A). Among these UsnRNP-specific proteins, U5-220 kDa (also named Prp8 or PRPC8) is a crucial spliceosomal protein that is most highly conserved through evolution. Prp8 interacts with sequences of all major splicing signals including 5′ss, branch site, and 3′ss. It has been proposed that Prp8 plays a critical role in catalysis by aligning 5′ss, and 3′ss at the catalytic center.

U6snRNP plays an important role at the catalytic center of the spliceosome, and its biogenesis has several unique features. The U6snRNA is transcribed by RNA polymerase III, and its cap structure is a γ-monomethyl group. The U6snRNP does not contain an Sm protein. Instead, U6snRNP contains seven Sm-like proteins (Lsm2, 3, 4, 5, 6, 7, and 8) interacting with the U-rich region at the 3′ end of the U6snRNA. The assembly of U6snRNP is believed to occur in the nucleus. The base-pairing interactions between U4snRNA and U6snRNA lead to the formation of the U4/U6snRNP, which associates with U5snRNP as the U4/U6.U5 tri-snRNP complex to join the spliceosome. The formation of U4/U6.U5 tri-snRNP complex and association of tri-snRNP with the spliceosome require PRPF31 (human homolog of yeast Prp31p) in addition to other proteins.

1.4 Spliceosome Assembly

A large number of studies show that mammalian spliceosomes are assembled on the pre-mRNA splicing substrate in an orderly fashion. A recent study, on the other hand, reported a penta-snRNP particle purified from the yeast extract. The finding of this ''preassembled'' penta-snRNP particle suggests the potential importance of concurrent multisite interactions during spliceosome assembly among spliceosomal components and different intronic as well as exonic regions.

Many components of spliceosomes are conserved between yeast and human. In fact, much of our knowledge of spliceosome assembly is based on genetic studies in yeast and on biochemical experiments using both yeast and mammalian systems. The mammalian homologs of key spliceosomal components have been identified (Table 1). Systematic proteomic studies of spliceosomes assembled on model splicing substrates confirm the high degree of conservation between mammalian and yeast spliceosomal proteins (Table 1).

Spliceosome assembly is a highly dynamic process, with multiple RNA–RNA, RNA–protein, and protein–protein interactions. During spliceosomal assembly and activation, there are a series of conformational rearrangements. This process involves not only changes in the RNA conformation but also remodeling of snRNPs and exchanges of protein factors. Studies over the past two decades have only provided a sketch of this multicomponent interactive process. At the catalytic center of the functionally mature spliceosome, the 5′ss and 3′ss must be juxtaposed precisely to ensure accurate cleavage and ligation.

The classical view of spliceosome assembly features stepwise interactions between the pre-mRNA substrate and different spliceosomal UsnRNPs. Immediately after transcription, nascent pre-mRNA transcripts interact with hnRNP (heterogeneous nuclear ribonucleoprotein) proteins (Table 2) to form *H complex*. Spliceosome assembly is initiated with the recognition of the 5′ splice site by U1snRNP in an ATP-independent manner. Efficient interaction between the U1snRNP and 5′ splice site requires U1snRNP-specific proteins (such as U-70 K) and possibly other proteins. Recognition of the branch site and 3′ splice site requires cooperative binding of SF1 (also named the *branch point binding protein*, BBP) to the branch site and binding of U2 auxiliary factor (U2AF65 and U2AF35 as a heterodimer) to the intronic sequence at the 3′ splice site. Splicing reactions can occur in the absence of detectable U1snRNP. However, efficient splicing of most introns appears to require U1snRNP. The formation of such an early complex (*E complex*) is considered to be the commitment step that directs the nascent pre-mRNA transcript into the splicing pathway.

A distinct feature of the formation of mammalian splicing commitment complex is the involvement of a family of proteins named SR proteins that are not found in the genome of *S. cerevisiae*. SR proteins share common structural features, with one or two RNA recognition motifs (RRMs) of the RNP-consensus (RNP-cs) type at the amino-terminus and a carboxyl domain rich in arginine and serine residues (RS domain, Table 1). These proteins play important roles in recognizing exons and mediating interactions between splice sites. SR proteins interact with exonic sequences and recruit splicing factors for the formation of ''cross-exon'' complex, (i.e. exon definition). SR proteins also interact with 5′ss and facilitate the ''cross-intron'' recognition by mediating protein–protein interactions between U1snRNP and factors associated with the branch site 3′ss including U2snRNP, U2AF, and possibly other proteins.

The binding of U1snRNP and SR proteins promotes the interaction of the 17S U2snRNP with the branch site sequence. The stable association of U2snRNP with the branch site to form

Tab. 2 Human H complex proteins.

Acce. #	*Protein name*	*Motifs*	*Cal. MWt*
Q13151	hnRNP A0	2 RRMs	30 841
P09651	hnRNP A1**	2 RRM	38 715
P22626	hnRNP A2/B1	2 RRMs	37 430
P51991	hnRNP A3	2 RRMs	39 686
P07910	hnRNP C1/C2	1 RRM	33 299
Q14103	hnRNP D0	2 RRMs	38 434
P26599	hnRNP I/PTB**	4 RRMs	57 221
Q07244	hnRNP K	KH	50 976
P14866	hnRNP L	3 RRMs	60 187
O43390	hnRNP R	3 RRMs	70 943
AL031668	hnRNP RALY	1 RRM	32 214
P35637	FUS/hnRNP P2	1 RRM	53 426
B54857	NF-AT 90 k	2 DSRMs, ZF	73 339
AJ271745	NFAR-1	DSRM, ZF	76 033
A54857	NF-AT 45 k	DSRM, ZF	44 697
AF037448	GRY-RBP	3 RRMs	69 633
P43243	Matrin3		94 623
O43684	hBUB3	WD40s	37 155
Q15717	HuR	3 RRMs	36 062
Q92804	TAFII68	1 RRM, ZF	61 830
P16991	YB1**	COLD	35 924
P16989	DBPA	COLD	40 060
P08107	HSP70	HSP70	70 052
P11142	HSP71	HSP70	70 898
P11021	GRP78	HSP70	72 116

The H complex proteins that are involved in alternative splicing regulation are marked with "**."

the *A complex* is ATP-dependent and requires the U2snRNP-specific proteins SF3a (SF3a60/SAP61 (Prp9), SF3a66/SAP62 (Prp11), and SF3a120/SAP114 (Prp21)). This process also requires non-snRNP splicing factors including hPrp5 (a putative ATP-dependent RNA helicase). Upon the integration of the U2snRNP complex, a U2snRNP-protein, p14, contacts the branch adenosine residue and interacts with other U2snRNP proteins. The next ATP-dependent step is the association of U4/U6.U5 tri-snRNP complex to form the *B1 complex*, in which the interaction of the 5′ss with U1snRNA is destabilized by U5-100 kDa (hPrp28), a putative RNA helicase. The 5′ss sequence is then engaged in interactions with U6snRNA around the intronic region and with U5snRNA at the exonic region. After escorting the U6snRNP into the B1 complex, the U4snRNP is released to form the *B2 complex*, a transition that requires ATP and possibly U5-200 kDa (human homolog of Brr2 in yeast, another putative RNA helicase). The next ATP-dependent transition is the formation of the *C1 complex*, a process involving the activity of yet another possible RNA helicase hPrp2 (human homolog of Prp2/Ynr011c). This

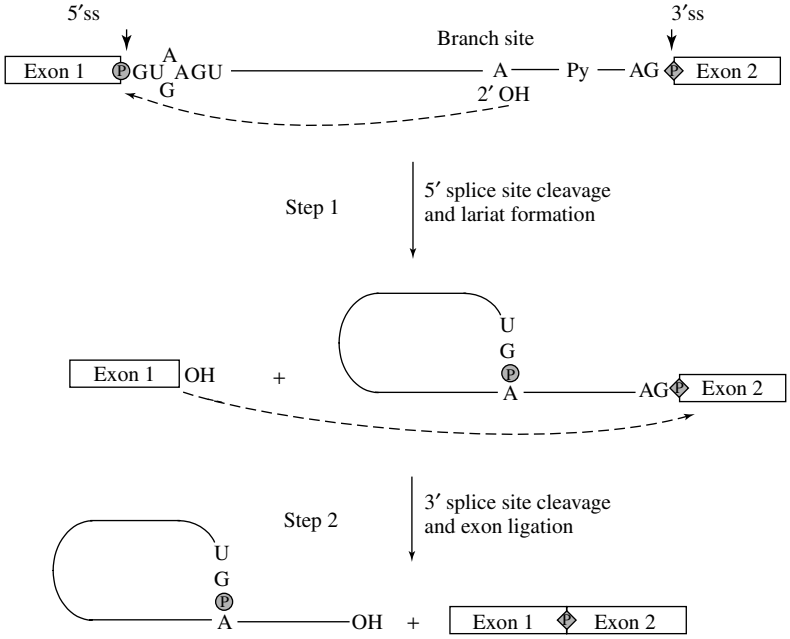

Fig. 4 The biochemical mechanisms of pre-mRNA splicing: two steps of transesterification. The phosphodiester linkages are indicated by the letter "p" inside a circle or a diamond.

may directly lead to the formation of the catalytic site for the *first step of splicing*, the cleavage at the 5′ss with the formation of a lariat intermediate (see Fig. 4). U2, U6, and U5snRNPs are associated with the catalytically active form of splicing complex in which the first step of splicing occurs. The *second step of splicing* (cleavage at the 3′ss and ligation of exons) requires an additional set of protein factors including hPrp16, hPrp17, hPrp18, and hSlu7 (Table 1). The human homolog of the yeast Prp16, hPrp16, contains an RS domain in addition to DEXD-box helicase and ATPase domains. The transition to form the *C2 complex* is again ATP-dependent. The catalytic center for the second step is also formed by U6snRNA, U2snRNA, and/or associated proteins. The splicing products, ligated exons, and the lariat intron, are then released from the spliceosome by another DEXD-box containing RNA helicase together with other factors. The release of the lariat intron and dissociation of U2, U5, and U6snRNPs from the *I complex* requires another putative RNA helicase, hPrp43. The released U6snRNP is reannealed with U4snRNP and associated with U5snRNP to form the U4/U6.U5 tri-snRNP complex to enter another cycle of spliceosome assembly. The lariat intron is debranched by the enzyme Dbr1. Many spliceosomal components are presumably recycled to form new spliceosomes.

It is clear that spliceosome assembly is a highly dynamic process involving multiple networks of RNA–RNA, protein–RNA,

and protein–protein interactions at each step. Our understanding of this complex process, as well as its regulation, remains limited.

1.5 Biochemical Mechanisms of pre-mRNA Splicing

The complex and dynamic process of spliceosome assembly culminates in the formation of the catalytic core in which 5′ss and 3′ss are precisely juxtaposed. The biochemical reactions of pre-mRNA splicing involve two transesterification steps (Fig. 4). The first step is the cleavage at the 5′ss and the formation of the lariat intermediate with a 2′-OH displacing a 3′-OH group. The second step is the cleavage at the 3′ splice site with concomitant ligation of the 5′ and 3′ exons in which a 3′-OH displaces another 3′-OH group. Although it remains controversial, these two steps of splicing reactions may take place in two catalytic sites. In addition to the chemical differences between the two reaction steps, the stereoselectivity of the transesterification reactions and the difference in the metal-ion dependence support the existence of two reaction sites.

Accumulating evidence suggests that pre-mRNA splicing may be RNA-catalyzed with U6snRNA at the center for catalysis and protein factors including PRP8 as components of the catalytic core. U6snRNA plays a critical role in catalysis and in interacting with pre-mRNA and other spliceosomal snRNAs. In yeast, U6snRNA contributes to spliceosomal catalysis by coordinating catalytic metals. Precise catalytic mechanisms of mammalian pre-mRNA splicing remain to be elucidated.

2 Alternative pre-mRNA Splicing

2.1 Alternative Splicing and its Role in Regulating Gene Activities and Generating Genetic Diversity

The vast majority of vertebrate genes contains at least one intron. It is estimated that more than 50% of human genes undergo *alternative splicing*, the process that generates distinct splicing products from the same pre-mRNA transcript by using different splice sites (Fig. 5). Alternative splicing is an important mechanism for regulating gene activities in eukaryotic species. Alternative splicing events can occur in the protein-coding regions or in the regulatory regions of genes, including both 5′ and 3′ untranslated regions. Alternative splicing can affect peptide coding capacity or influence mRNA stability or translational control of the transcription products. Alternative splicing regulation can also be coupled with transcription, polyadenylation, RNA editing, or mRNA exporting processes. A recent example for the coupling of splicing with other processes of gene regulation is that steroid hormone receptors can simultaneously regulate transcription and alternative splicing by recruiting coregulators involved in both processes. In addition, alternative splicing can affect posttranslational modifications of protein products.

2.1.1 Different Patterns of Alternative Splicing

A number of distinct alternative splicing patterns have been reported (Fig. 5). Most common alternative splicing events include exon inclusion/skipping, intron removal/retention, and alternative selection of competing 5′ or 3′ splice sites.

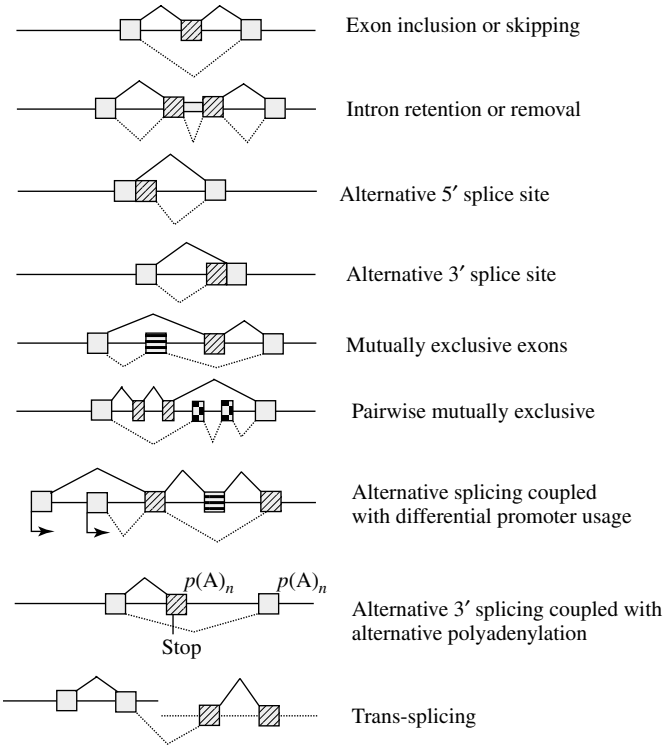

Fig. 5 Diverse patterns of alternative splicing. Alternatively spliced regions or exons are illustrated as shaded boxes.

More complex patterns of alternative splicing include mutually exclusive or cassette types of exon inclusion. Alternative selection of terminal exons can be coupled with differential promoter usage or polyadenylation. Furthermore, recent studies have documented possible alternative trans-splicing of mammalian genes, although it may occur only at a low frequency in mammals.

2.1.2 **Alternative Splicing and Genetic Diversity**

Alternative splicing regulates gene activities involved in every aspect of cell survival and function. It is a major mechanism for generating the complexity of mammalian proteomes. Alternative splicing contributes to proteome expansion by a number of mechanisms, such as the usage of distinct translation start sites, in-frame nucleotide deletion or insertion, changes in peptide sequence, and alternative usage of different translation stop codons. Such changes in peptide sequence or length may lead to the formation of proteins with distinct properties, including biochemical/biophysical characteristics, subcellular localization (secreted versus membrane associated, membrane-tethered versus cytoplasmic, cytoplasmic versus nuclear), posttranslational modifications (glycosylation, phosphorylation, or

lipid modification), or interactions with other cellular components.

Alternative splicing can be an excellent mechanism for generating functionally antagonistic products from the same genetic locus and for the fine-tuning of gene activities at the posttranscriptional level. For example, alternative splicing of a number of genes critical for cell death leads to the formation of both cell death–promoting and cell death–preventing splicing isoforms. These include genes encoding for death ligands, death receptors, Bcl-2 superfamily of death regulators, caspases, and other cell-death regulatory genes. Several human caspase genes utilize alternative splicing to produce protein products that either contain or lack their enzyme active sites, resulting in antagonistic activities in cell death.

The nervous system is a good example where alternative splicing is utilized to generate extreme functional diversity. A vast number of genes involved in neural development and function undergo complex alternative splicing. Some genes encoding neural receptors and axon guidance molecules can generate hundreds to even thousands of different splicing isoforms. Many ion channels and neurotransmitter receptors have different alternative splicing isoforms with distinct electrophysiological properties. Alternative splicing regulation has a significant impact on the proper function of the nervous system, including learning, memory, and behavior development.

2.2 Mechanisms Underlying Alternative Splicing Regulation

The fundamental mechanisms underlying alternative splicing are the intricate interactions among trans-acting splicing factors and cis-elements in the pre-mRNA substrates, leading to selection of different splice sites. These cis-elements include splice sites and splicing regulatory sequences (splicing enhancers or silencers). The sizes of the exons or introns, as well as secondary structures of the pre-mRNA also influence splice site selection. The highly degenerate nature of mammalian splicing signals and the existence of large numbers of trans-acting factors allow versatile RNA–protein and RNA–RNA interactions during different stages of spliceosome assembly. Such interactions can be modulated by both exonic and intronic regulatory elements. The recognition of splice sites and cis-acting splicing regulatory sequences is mediated by a network of interactions between pre-mRNA and trans-acting factors including snRNPs and non-snRNP splicing factors. A number of splicing regulators, both positive and negative, play important roles in alternative splicing regulation (Table 3). The final outcome of alternative splicing of a given gene, including the production and the delicate balance of distinct isoforms, is determined by combinatorial effects of multisite interactions among pre-mRNA, essential spliceosomal components, and regulatory factors.

Genetic and biochemical studies in *Drosophila* genes have tremendously advanced our understanding of alternative splicing regulatory mechanisms, although some mechanisms may not be conserved in corresponding mammalian systems. For example, *Drosophila* sex determination genes are under extensive regulation by alternative splicing, and this regulatory mechanism is not conserved in mammalian sex determination. In general, *Drosophila* genes contain more small introns than mammalian genes. Here,

we focus on mammalian alternative splicing regulation.

2.2.1 Splicing Signals and Splicing Regulatory Elements

Unlike the yeast spliceosome, mammalian spliceosomes have the daunting task of searching for small exons in the vast sea of introns. This process of ''exon recognition'' or ''exon definition'' is particularly remarkable because sequences at mammalian splice sites are so highly degenerate (see Fig. 2). A large number of sequence elements with similarity to

Tab. 3 Pre-mRNA splicing regulators.

Protein name	*Motifs*	*RNA-binding sites*
A. Splicing activators		
SRp75	2 RRMs, RS	
SRp55	2 RRMs, RS	ESEs
SRp40	2 RRMs, RS	
ASF/SF2	2 RRMs, RS	
9G8	1 RRM, RS	
SC35	1 RRM, RS	
SRp30c	2 RRMs, RS	
hTra2	1 RRM, RS	
SRp20	1 RRM, RS	
SRm300	RS-rich	
SRm160	PWI, RS	
SF1 (BBP)	ZF, KH, CCHC,	ISE
KSRP	KH	
NOVA1	3 KH	(UCAUY)3
rSLM-2	KH, STAR	
CUG-BP	RRMs	ISE
ETR3	3 RRMs,	ISE
Halfpint (D.m.)	RRM (homolog of PUF60)	
TIA-1	3 RRMs	U-rich
p72	DExD, Helicase, RGG	
YB1	Cold box	A/C-rich ESE
B. Splicing repressors		
hnRNP A1	2 RRMs, Gly-rich	ESSs or ISSs
hnRNP I/PTB	4 RRMs	U/C-rich
nPTB	4 RRMs	U/C-rich
SmPTB	4 RRMs	U/C-rich
Elav (D.m.)	3 RRMs	
Sxl (D.m.)	3 RRMs	
Mec8 (C.e.)	2 RRMs	
Fox1 (C.e.)	RRM	
PSI (D.m.)	4 KH	ISS
QKI-5	KH	ISS
SRp30c	2 RRMs, RS	
SRrp35	1 RRM, RS	
SRp38	1 RRM, RS	

Tab. 3 (*continued*)

Protein name	*Motifs*	*RNA-binding sites*
SRrp40	1 RRM, RS	
SPF45	1 RRM, Gly-rich	
RSF1 (D.m.)	1 RRM, GRS	
SWAP (D.m.)	RS	
C. Bifunctional splicing regulators		
p54/SFRS11	1RRM, RS	
SRrP86	1RRM, RS	
p32		
HnRNP H	3RRMs, Gly	ESE or ISE ESS
Napor1	3RRMs	

Note: A. Splicing activators. B. Splicing repressors. C. Bifunctional splicing regulators.
STAR: signal transduction and activation of RNA.
For other abbreviations, see footnotes for Table 1.

authentic splice sites (pseudosplice sites) can be found in both intronic and exonic regions). Further complicating the issue, some of these elements act as *cryptic splice sites* that are only used by the splicing machinery when the authentic splice sites are altered by mutations. Therefore, specific recognition of authentic splice sites and correct pairing of corresponding 5′ and 3′ splice sites is a central issue for both constitutive splicing and alternative splicing regulation. The intrinsic sequence degeneracy of mammalian splice sites determines that alternative splicing is a rule rather than an exception during mammalian gene expression.

In addition to splice sites, sequence elements in both intronic and exonic regions modulate alternative splicing. Such regulatory elements can either enhance or suppress splicing, and are hence named exonic splicing enhancers (ESEs) and intronic splicing enhancers (ISEs), or exonic splicing silencers (ESSs) and intronic splicing silencers (ISSs).

A number of ESE motifs have been identified using biochemical systematic evolution of ligands by exponential enrichment (SELEX) or bioinformatical approaches. A/G-rich (also called purine-rich) and A/C-rich elements are among ESE motifs characterized by biochemical studies. Proteins containing SR domains play a major role in recognizing A/G-rich ESEs and recruiting other spliceosomal components (including snRNPs and other protein factors), thereby promoting the usage of neighboring splice sites. A cold-box protein, YB-1, has been shown to enhance splicing by interacting with an A/C-rich ESE. Another ESE in HIV-1 tev-specific exon interacts with hnNRP H and SR protein SC35 to enhance splicing. The SELEX method has been used to identify preferred binding sequences for individual SR proteins, and optimal binding sites for individual SR proteins are degenerate. ESE •prediction programs based on SELEX and computational analysis of human genes have been developed (http://exon.cshl.org/ESE). These Q7

programs are useful in predicting alternative splicing patterns of natural pre-mRNA substrates in cells.

ISE elements have been studied in a number of genes, including c-src, β-tropomyosin, calcitonin/calcitonin gene-related peptide gene (CGRP), fibronectin, nonmuscle myosin heavy chain, cardiac troponin T (cTNT), FGFR-2, α2 subunit of glycine receptor, and other genes. Such ISEs may contain sequences similar to 5′ splice site, U-rich element adjacent to the regulated 5′ splice site, UGCAUG element, (UCAUY)3-containing sequences, or CUG-containing motif.

A number of ESSs have been characterized in different genes such as β-tropomyosin, CD44, and viral genes including HIV Tat, bovine papillomavirus type-1, and Rous sarcoma virus. They do not share any obvious sequence motifs. Their activities are often associated with interactions with proteins of the hnRNP family, including hnRNPA1, hnRNP H, and hnRNP F.

A variety of ISSs have been analyzed in alternatively spliced genes including hnRNP A1, fibroblast growth factor receptor 2 (FGFR2), caspase 2, $GABA_AR\gamma2$ (γ-aminobutyric acid receptor typeA γ2subunit), NMDA R1 (*N*-methyl-D-aspartate receptor R1 subunit;), clathrin light chain B, and HIV tat. Many of these ISSs contain extended polypyrimidine• tracts. Some splicing repressor elements contain sequences similar to authentic splice sites, or decoy splice sites. It has been proposed that such decoy splice sites mediate nonproductive interactions to suppress the usage of upstream 5′ splice site. Again, several hnRNP proteins, including polypyrimidine tract binding protein (PTB, also known as hnRNP I), hnRNP A1, and hnRNP H play important roles in splicing repression by ISSs.

One of the common features of these cis-regulatory elements is that they are not simple sequence elements that act independently of each other. Sequence elements have been identified that can act to stimulate the splicing of one exon but repress another exon (e.g. in FGFR2 gene). Some regulatory elements contain both enhancer and silencer domains. The splicing of IgM exons M1 and M2 is regulated by a sequence containing juxtaposed splicing enhancer and silencer elements. Alternative splicing of protein 4.1R pre-mRNA generates multiple isoforms. This alternative splicing event is critical for red blood cell membrane biogenesis during erythroid differentiation and is regulated by multiple cis-elements and trans-acting regulators.

Splicing regulatory elements usually contain multiple binding sites for splicing regulators and function by recruiting other spliceosomal components to form RNP-like complexes. An evolutionarily conserved 100 bp intronic suppressor element in the caspase 2 (casp-2) gene, In100, specifically inhibits its exon 9 splicing. Alternative splicing of this 61 bp exon 9 leads to the formation of two functionally antagonistic products, casp-2L and casp-2S. Casp-2L product promotes cell death, whereas casp-2S prevents cell death. The In100 element contains a decoy 3′ splice site juxtaposed to a PTB-binding domain, both of which contribute to the full activity of In100 in inhibiting exon 9 inclusion. The upstream portion of In100 contains a sequence with features of an authentic 3′ splice site (including a branch site, a polypyrimidine tract, and AG dinucleotide). However, this site is not used under normal conditions. This sequence is only recognized as a 3′ss when the site is isolated with the downstream PTB-binding domain deleted.

Biochemical and cell culture experiments show that this decoy 3′ acceptor site interacts nonproductively with the 5′ splice site of the alternative exon 9, thus repressing the efficient use of the 5′ splice site of exon 9 despite a high level of U1snRNP binding to this 5′ splice site. Downstream of the decoy 3′ splice site resides the second functional domain that interacts with PTB. The binding of PTB to CU-rich motifs within this downstream domain juxtaposed to the decoy 3′ acceptor site correlates well with the repressor activity of this domain. PTB can modulate recognition of the adjacent decoy 3′ acceptor site. In addition to factors interacting with an authentic 3′ splice site (such as U2AF or U2snRNP), PTB as well as other proteins interact with In100 and contribute to recognition of the In100 decoy 3′ splice site by the spliceosome as an intronic repressor element, rather than as an authentic 3′ splice site. The regulatory role of PTB in casp-2 alternative splicing and its mechanism of action appear to be distinct from other systems (●Sect. 2.2.2). A recent survey of known human genes involved in cell death regulation suggests that In100-like intronic elements (i.e. 3′ splice site juxtaposed to PTB-binding domains) may represent a general intronic splicing repressor motif. Such intronic elements may play a role in regulating alternative splicing of other cell-death genes.

Q9

Another example of the complexity of splicing regulatory elements is in the alternative splicing of exon 10 in the human *tau* gene. Alternative splicing of this exon is associated with the pathogenesis of dementias (●Sect. 3.1). In this case, both exonic and intronic regulatory elements play important roles in controlling exon 10 inclusion. In addition, in the exonic region, both positive and negative elements are involved and form a multidomain composite regulatory element.

Q10

2.2.2 Trans-acting Splicing Regulators

The recognition and selection of splice sites are determined during spliceosome assembly, especially at early steps of spliceosomal formation. Splicing activators facilitate interaction of U1snRNP with 5′ splice site and of U2snRNP with 3′ splice site, whereas splicing repressors suppress the recognition of splice sites. A number of proteins involved in spliceosome assembly also play important roles in regulating alternative splicing (see Table 1).

Several families of splicing activators have been reported (Table 3). SR proteins are among the best characterized splicing activators. The interactions between SR proteins and ESEs play a critical role in exon recognition by the spliceosome. ESEs are present in both constitutively and alternatively spliced exons. By mediating protein–protein and RNA–protein interactions during early steps of spliceosome assembly, SR proteins coordinate the communication between 5′ and 3′ splice sites and promote interactions between exonic enhancers and splice sites. Enhancer complexes are usually multicomponent complexes, containing not only SR proteins but also other RS domain–containing proteins such as U2AF. Some RS domain–containing proteins, such as SRm300 and SRm160, act as coactivators for ESE function. The RS domain in SR proteins can be differentially phosphorylated, providing another level of regulation. The phosphorylation status of SR proteins regulates protein–protein interactions, intracellular distribution, and activities of SR proteins. Differential phosphorylation of SR proteins has been shown to play a role in regulating gene expression during development.

SF1, KSRP, NOVA-1, and rSLM-2 are RNA-binding proteins containing heterogeneous nuclear ribonucleoprotein K-type homology (KH) domain. They can enhance splicing by interacting with ISEs. SF1, a protein important for branch site recognition during spliceosome assembly, binds to GGGGCUG repeats in an ISE to activate the recognition of a 6-bp micro exon in cTNT gene. KSRP interacts with UGCAUG sequence and stimulates the neuronal-specific exon inclusion in c-src. NOVA-1 enhances the splicing of exon E3A in the $\alpha 2$ subunit of the glycine receptor gene (GlyR$\alpha 2$). The rat Sam68-like mammalian protein (rSLM-2) is a member of the STAR (signal transduction and activation of RNA) protein family. It can influence the splicing pattern of the CD44v5, human transformer-2beta, and tau minigenes in transfected cells.

Several members of CUG-BP and ETR-like factors (CELF) family activate the splicing of genes including cTNT, muscle-specific chloride channel, insulin receptor, and NMDA R1. Some of these CELF proteins activate splicing by interacting with ISEs.

TIA-1, a mammalian homolog of yeast NAM8 protein, interacts with the U-rich intronic sequence adjacent to the 5′ splice site of the K-SAM alternative exon in the FGFR2 gene. The activation of this splice site is U1snRNP-dependent, and TIA-1 may function by facilitating U1-snRNP binding to the 5′ splice site.

Human Y-box binding protein (YB-1) was initially identified as a transcription factor interacting with single-stranded DNA in response to cold shock. YB-1 contains a five-stranded ß-barrel known as the ''COLD'' domain and accessory domains rich in basic amino acids and aromatic groups. This domain arrangement in YB-1 protein with specific interaction domain containing ß-sheets and a basic domain providing generic RNA binding is reminiscent of the modular structure of the SR proteins. YB-1 stimulates the splicing of the human CD44 alternative exon v4 by interacting with the A/C-rich element in the exonic splicing enhancer. Another protein, DEAD-box RNA helicase p72, has been reported to affect the splicing of alternative exons containing AC-rich exon enhancer elements. The mechanism of YB-1 or other proteins interacting with the A/C-rich type of ESEs in splicing activation remains to be elucidated.

HnRNP proteins hnRNPA1 A1 and PTB are among the best characterized splicing repressors. HnRNP A1 can interact with either ESSs or ISSs to prevent exon recognition by SR proteins or U2 assembly. A model has been proposed for interactions between hnRNP A1 and ESS based on studies on HIV Tat exon 3 splicing. In this model, the high-affinity binding of hnRNPA1 to ESS promotes nucleation of multiple A1 along the exon. The formation of this inhibition zone can be blocked by SF2/ASF, but not by another SR protein SC35, providing an explanation for differential antagonism between hnRNP A1 and different SR proteins. Two different models have been proposed for the function of hnRNPA1 in splicing silencing mediated by ISSs. In the case of HIV Tat intron 2 splicing, the interaction of hnRNP A1 with an alternative branch point sequence blocks the recognition of the branch site by U2snRNP. In the second model, the cooperative binding of hnRNP A1 to two intronic elements flanking the alternative exon inhibits exon inclusion by a looping-out mechanism during autoregulation of hnRNP A1 gene splicing.

PTB interacts with U/C-rich elements in ISSs in a number of genes including n-src, FGFR2, $GABA_AR\gamma 2$, tropomyosin,

NMDA R1 exon 5, clathrin light chain B, caspase-2, and calcitonin/CGRP genes. Depletion of PTB using an RNA interference approach demonstrates that PTB is a negative regulator of exon definition in cultured cells. PTB-binding sites are frequently located in the intronic regions upstream of the regulated 3′ splice sites, although functionally active PTB-binding sites are also found in intronic elements downstream of the alternatively spliced exons (such as in c-src and caspase-2 genes). Mechanisms by which PTB represses splicing are not clear yet. One model for repression by PTB is via competition with U2AF binding to the polypyrimidine tract to block early spliceosome formation. This model is based on studies of alternative splicing of $GABA_AR\gamma 2$, NMDA R1 exon 5, and clathrin light chain B genes. In these genes, the high-affinity PTB binding to the long polypyrimidine tracts immediately upstream of the neural-specific exons represses the inclusion of these exons in nonneural tissues. A similar repression mechanism may be used in suppressing the inclusion of muscle-specific exons in rat α- or β-tropomyosin. PTB exists in a range of different tissues as an abundant splicing repressor. The repressor activity of PTB may be modulated by other regulatory proteins. Two tissue-specific PTB-related proteins have been found, neuron-enriched (nPTB,) and smooth muscle–enriched (SmPTB). For example, nPTB can compete with PTB to promote inclusion of neuronal-specific N1 exon inclusion in c-src. The expression of SmPTB correlates with the smooth muscle–specific suppression of α-tropomyosin exon 3, which is included in nonsmooth muscle cells. The release of PTB suppression can be also achieved by cell-type specific CELF proteins such as ETR3 and Napor-1.

One exception to the general inhibitory activity of PTB has been reported where PTB stimulates the inclusion of an alternative 3′-terminal exon. In this case, the splicing regulation is coupled with alternative polyadenylation. In addition, PTB has also been implicated in translational control of viral transcripts. More comprehensive understanding of the biological roles of PTB in gene regulation requires further investigation.

Several proteins containing an RS domain and RRM-cs domains have also been reported to act as splicing repressors, including SRrp30c, SRrp35, SRp38, and SRrp40 (see Table 3). SR proteins are generally hyperphosphorylated by the kinase SRPK1. Dephosphorylated SRp38 is required for the splicing repression in mitotic cells. Another RS domain-containing protein, the *Drosophila* suppressor-of-white-apricot protein (SWAP) suppresses its own pre-mRNA splicing.

An RRM-containing spliceosomal protein, SPF45, represents a late-acting splicing regulator. In *Drosophila*, SPF45 blocks splicing at the second step by interacting with Sex-lethal (Sxl) protein, indicating that 3′ splice site recognition and splicing regulation can occur at the second catalytic step.

Proteins containing KH domains can also act as splicing repressors. For example, a nuclear isoform of quaking (qk) protein, QKI-5, regulates alternative splicing of myelin-associated glycoprotein gene by interacting with an intronic splicing repressor element. A *Drosophila* splicing repressor, P-element somatic inhibitor (PSI), is required for the soma-specific inhibition of splicing of P-element pre-mRNA both *in vitro* and *in vivo*.

Several splicing regulators can act as either positive or negative splicing regulators, depending on the sequence context

of the pre-mRNA substrates. We classify these splicing regulators as bifunctional splicing regulators. Some SR proteins or SR-domain containing proteins can repress splicing of certain pre-mRNAs but activate splicing of other substrates. SRrp86 is an 86-kDa related to SR proteins. It can function as either an activator or a repressor by regulating the activity of other SR proteins. Other proteins can repress splicing by interacting with SR proteins. For example, an ASF/SF2-interacting protein, p32, was shown to act as a splicing regulator by inactivating the function of ASF/SF2.

HnRNP H and the related protein hnRNP F contain three RRMs of RNP-cs type. HnRNP H interacts with G-rich elements in either splicing enhancers or silencers. When binding to a splicing enhancer in the HIV env gene, hnRNP H acts as an activator by promoting the assembly of a complex containing SC35 and U1snRNP. When interacting with ESSs in the β-tropomyosin pre-mRNA or HIV Tat exon 2, or with a negative regulator of splicing (NRS) in the Rous Sarcoma Virus (RSV) genome, hnRNP H serves as a splicing repressor.

Napor1, a splice variant of ETR3, is expressed at a high level in the forebrain but at a low level in the cerebellum. Overexpression of Napor1 exerts opposite effects on two different exons in NMDAR1, inhibiting exon 5 inclusion but stimulating exon 21 inclusion.

With more alternative splicing events characterized in detail, more cases of bifunctional splicing regulators may emerge. It is conceivable that these splicing regulators interact with different sequence elements and different spliceosomal components. When these interactions promote the productive recognition of splicing signals by the splicing machinery, such proteins act as splicing activators. On the other hand, when the same proteins compete with other splice activators or facilitate recognition of splicing silencers leading to nonproductive interactions between the spliceosome and splicing substrates, these proteins behave as splicing repressors.

2.3 Tissue-specific and Developmentally Regulated Alternative Splicing

Tissue-specific alternative splicing has been studied in mammalian systems for more than a decade. However, only a limited number of tissue-specific splicing regulators have been identified so far. These include Nova-1, Nova-2, nPTB, and SmPTB. Cell-type specific splicing regulators have also been found in *Drosophila* and *Caenorhabditis elegans*, such as neuronal-specific Elav (embryonic lethal and abnormal vision), ovary-specific Halfpint, and muscle-specific Mec8. From a large number of biochemical, molecular, and genetic studies, the emerging general theme is that no single factor dictates the tissue specificity of alternative splicing of any genes. Instead, multicomponent interactions among tissue-specific splicing regulators, pre-mRNA, and general splicing factors determine the specific alternative splicing pattern of a given gene in a tissue- or cell-type specific manner. On the other hand, genetic deletion of a single splicing regulator, even those highly tissue-specific ones, leads to defects in splicing of multiple target genes. For example, Nova-1 deletion in mice affects alternative splicing of $GABA_AR\gamma 2$, $Gly R\alpha 2$ and perhaps other unknown target genes. Similarly, the *Drosophila* neural-specific splicing regulator Elav has at least three known target genes, neuroglian (nrg), erectwing, and armadillo.

In addition to tissue-specific expression of splicing regulators, a number of other mechanisms modulate tissue-specific alternative splicing. Tissue-specific combination of different splicing regulators, relative concentrations of distinct splicing factors, and differential modification of splicing regulators all contribute to the tissue-specific alternative splicing of individual genes in a given tissue or cell type. For example, the level of the splicing repressor PTB in neural tissues is lower than in other tissues, providing an explanation for a more permissive environment for exon inclusion in a number of genes in neural tissues. Although most SR proteins are expressed in a wide range of tissues, their expression patterns vary in different tissues, both in different isoforms and at different levels. Many SR proteins have different isoforms either because of alternative splicing or posttranslational modifications. Kinases or phosphatases that regulate phosphorylation of different splicing factors can also be tissue specific. One example is the SR protein–specific kinase 2 (SRPK2), which is highly expressed in the brain and presumably capable of regulating activities of target SR proteins.

It is not clear yet how the extremely complex alternative splicing pattern of different genes in a given cell or a specific tissue is coordinated. In some cases, the alternative splicing events of different genes appear to be coregulated by the same protein. For example, the splicing regulator PTB can differentially recognize neural and nonneural substrates. The expression pattern of PTB in different regions of the brain at different developmental stages supports a role for PTB to act as an alternative splicing coordinator for different splicing target genes.

2.4 Regulation of Alternative Splicing in Response to Extracellular Stimuli

Extracellular signals can induce changes in alternative splicing of different genes. For example, growth factors or hormones stimulate alternative splicing of intracellular responsive genes. Treatment with growth factors induces changes in alternative splicing of phosphotyrosine phosphatase PTP-1B gene. The alternative splicing pattern of protein kinase C beta is changed by insulin. Activation of calmodulin-dependent kinases (CaM kinases) stimulates changes in alternative splicing of BK potassium channels. In most cases, however, signal transduction pathways involved in splicing regulation have not been characterized.

A number of genes important for the functioning of the nervous system show activity-dependent changes in their splicing. Alternative splicing of syntaxin 3 changes in response to induction of long-term potentiation. NMDAR1 alternative splicing is modulated by both pH and Ca^{2+}. Stress hormones regulate alternative splicing of potassium channels. A cis-acting sequence named *calcium-responsive RNA element* (CaRRE) in the stress-inducible exon of BK potassium channels and in NMDAR1 exon 5 has been identified. This element mediates CaM kinase-dependent repression of the inclusion of the CaRRE-containing exons. The mechanisms by which the CaRRE causes exon skipping remain to be elucidated.

Drastic changes in the cell growth environment are expected to affect RNA metabolism, including alternative splicing. For example, ischemia in mice induces changes in alternative splicing of

several genes examined. Accompanying such changes in alternative splicing are changes in the intracellular distribution of several splicing regulators. It is possible that these splicing regulators undergo changes in posttranslational modifications or at other levels induced by ischemia.

Chemical compounds, when administrated to animals or applied to cells in culture, can also cause changes in alternative splicings. Sodium butyrate has been tested in transgenic mice and shown to increase exon 7 inclusion of the SMN gene. Aclarubicin treatment induces changes in SMN gene alternative splicing in cultured cells. Treatment of A549 lung adenocarcinoma cells with cell-permeable ceramide, D-e-C(6) ceramide, downregulates the levels of Bcl-xL and caspase 9b splicing isoforms. The functional responses of nuclear splicing machinery to drug treatment opens the possibility of correcting aberrant or defective splicing using therapeutic agents (●Sect. 4.2). Q11

3 Pre-mRNA Splicing and Human Diseases

Aberrant pre-mRNA splicing has been implicated in the pathogenesis of a number of human diseases. Alterations, or dysregulation of either constitutive pre-mRNA splicing or alternative splicing can cause disease phenotypes. Molecular genetic studies of human diseases have revealed a wide range of mutations that cause diseases because of their effects on pre-mRNA splicing. Investigating the molecular nature of these genetic defects and pathogenetic mechanisms will facilitate both diagnosis of such diseases and development of new therapeutic approaches.

3.1 Splicing Defects in Human Diseases

Genetic defects that cause aberrant pre-mRNA splicing can be found in every category of disease, from malignancies affecting a certain tissue or organ to diseases or syndromes involving multiple systems. Splicing mutations have been identified in genes involved in the pathogenesis of diseases in every system. Detailed information about these mutations can be found in databanks including HGMD (the Human Gene Mutation Database, http://www.uwcm.ac.uk/uwcm/mg/) as well as in a large number of publications. As shown in ●Table 4, we use a few examples to illustrate the diverse range of disease phenotypes and genes involved. The current understanding of these splicing defects will be summarized. Q12

Malignancies are a major cause of mortality. Splicing mutations in genes critical for either cell proliferation or cell death result in malignancies in different systems. These include oncogenes, tumor suppressor genes, and genes involved in cell death. Only a few examples are included here to demonstrate the complexity of the involvement of defective/aberrant splicing in tumorigenesis. Cancer of the same tissue or organ can be caused by splicing mutations in different genes. For example, splicing defects in p53, p51, CD44, and members of epidermal growth factor receptor (EGFR) family lead to lung cancer, the most common cancer in humans. Breast cancer has been associated with splicing mutations in a number of genes, including HER-2/neu, p53, mdm2, and BRCA1 or BRCA2. Glioblastomas are also associated with splicing mutations in different genes. On the other hand, splicing defects in a single gene can lead to tumorigenesis in different tissues. For example,

Tab. 4 Examples of human diseases associated with aberrant or defective pre-mRNA splicing.

	Diseases	***Genes***	***Aberrant splicing***
Malignancies			
	Lung cancer	p53, p51, CD44, EGFR	Aberrant or defective splicing products
	Cancer in different tissues	p53, CD44	Defective or aberrant splicing isoforms
	Wilm's tumor	CD44	Aberrant splicing isoforms
	Breast cancer	HER2/neu	Aberrant splicing and imbalance of splicing isofors
	Breast and ovarian cancer	BRCA1, BRCA2	Exon skipping
	Breast cancer, rhabdomyosarcoma	Mdm2	Oncogenic variants lacking p53-binding domain
	Glioblastomas	Mdm2	Oncogenic variants lacking p53-binding domain
	Neurofibromatosis type 1	NF-1	Exon skipping
	Neuroblastoma	HUD, HUC, NNP-1, α-internexin	Production of tumor antigens
	AML	MLL-SEPTIN6	MLL-SEPTIN6 fusion-splicing variants
Cardiovascular diseases			
	Hypercholesterolemia	LDLR	Cryptic ss usage, exon skipping
	Hypertriglyceridemia	Hepatic lipase	Cryptic ss usage
	Marfan syndrome	Fibrillin-1	Cryptic ss usage, exon skipping
	Cardiomyopathy	cTNT	Aberrant splicing isoforms
	Hypertension	G-protein β3	Exon skipping
Metabolic diseases			
	Glycogen storage disease, type II	Lysosomal α-glucosidase	Cryptic ss usage
	Hereditary tyrosinemia, Type I	Fumarylacetoacetate hydrolase	Exon skipping
	Acute intermittent porphyria	Porphobilinogen deaminase	Exon skipping
	Ceruloplasmin deficiency	Ceruloplasmin	Cryptic ss usage
	Fabry's disease	Lysosomal α galactosidase A	Cryptic ss usage
	Tay–Sach's Disease	β-hexosaminidase	Intron retention, exon skipping
	Sandhof disease	β hexosaminidase β-subunit	Cryptic ss usage

(*continued overleaf*)

Tab. 4 (*continued*)

	Diseases	*Genes*	*Aberrant splicing*
Neurodegenerative diseases			
	FTDP-17	Tau	Imbalance of splicing isoforms
	Alzheimer's disease	Presenilin-1	Exon skipping
	Alzheimer's disease	Presenilin-2	Exon skipping
	Ataxia telangiectasia	ATM	Cryptic ss usage
	Multiple sclerosis	CD45	Imbalance of splicing isoforms
	Spinal muscular atrophy	SMN1, SMN2	Exon skipping
	Retinitis pigmentosa	HPRP3, PRPF31, PRPC8	Unknown
Psychiatric disorders			
	Schizophrenia	$GABA_AR\gamma 2$, NCAM	Imbalance of splicing isoforms
	Schizophrenia	NMDAR1	Aberrant alternative splicing
	ADHD	Nicotinic acetylcholine receptor	Aberrant alternative splicing
Other syndromes or diseases			
	Cystic fibrosis	CFTR	Exon skipping
	IGHD II	GH-1	Imbalance of splicing isoforms
	Frasier syndrome	WT 1	Imbalance of splicing isoforms
	Epilepsy	AMPA receptor	Imbalance of splicing isoforms
	Menkes disease	MNK	Exon skipping
	Beta-thalassemia	β-globin	Cryptic ss usage
	Metachromatic leukodystrophy	Arylsulfatase A	Cryptic ss usage
	Myotonic dystrophy	DMPK	(CUG)n expansion:aberrant splicing

CD44 aberrant splicing has been associated with the development of a range of different tumors, such as breast cancer, prostate cancer, lung cancer, and Wilm's tumor. CD44 is a polymorphic family of cell surface glycoproteins important for cell adhesion and migration. Aberrant splicing of CD44 has been associated with the invasive behavior and metastasis of several types of tumors. Alternative CD44 splicing variants have been correlated with the poor prognosis of tumors. Another gene is Mdm2, whose alternative splicing defects have been implicated

in oncogenesis in multiple tissues. Oncogenic splicing variants of Mdm2 that lack a domain important for its function in interacting with the tumor suppressor p53 have been found in breast cancer, glioblastoma, and rhabdomyosarcoma. Alterations in the balance of different FGFR2 splicing isoforms have been correlated with progression of prostate cancer. A significant fraction of neurofibromatosis type 1 (NF1) cases are associated with an aberrantly spliced NF1 gene. Production of tumor antigens as a result of alternative splicing has been associated with neuroblastoma and other tumors. The role of these tumor antigens in the development and metastasis of these tumors remains to be investigated. In pediatric acute myeloid leukemia (AML), splicing isoforms of gene fusion transcripts produced as a result of chromosomal translocation have been reported. Such aberrant gene fusion-splicing products may contribute to pathogenesis of tumors caused by chromosomal translocation. Finally, the development of drug resistance has been associated with changes in alternative splicing of members of multidrug resistance genes or genes involved in drug metabolism.

Cardiovascular diseases are among the leading causes of human death. Aberrant splicing of structural genes such as cTNT is associated with cardiomyopathy. Splicing defects resulting in the formation of a defective receptor, such as the low-density lipoprotein receptor (LDLR), lead to familial hypercholesterolemia (FH). A single-nucleotide change altering the function of intracellular signal transduction molecules can cause diseases affecting multiple systems. For example, a single-nucleotide polymorphism leading to the formation of a G-protein beta3 subunit splice variant has been associated with hypertension. Aberrant or defective splicing can disrupt the function of genes encoding a wide range of proteins important for the function of the cardiovascular system, from cell surface receptors to intracellular signaling molecules. Both structural genes and regulatory genes can be affected by splicing mutations. These examples show that splicing mutations can affect development/formation of the cardiovascular system and regulation of cardiovascular system function.

A range of metabolic diseases is caused by splicing mutations, affecting the production of a single metabolic enzyme. For example, the glycogen storage diseases, Sandhof's disease, and Fabry's disease can be caused by the usage of a single cryptic splice site in the corresponding genes. Hereditary tyrosinemia (type I) and acute intermittent porphyria can develop because of improper exon skipping in the genes encoding for respective metabolic enzymes.

The critical role of splicing regulation for the normal function of the nervous system is demonstrated by the large number of neurodegenerative and psychiatric disorders associated with aberrant splicing. Splicing defects in *tau* and presenilin genes have been identified in different types of dementia, including frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17) and Alzheimer's disease. Splicing mutations that cause cryptic splice site usage in the ataxia telangiectasia (ATM) gene have been identified in ATM patients. Multiple sclerosis has been associated with imbalances of different splicing isoforms of CD45. A number of studies have shown that genetic mutations in ubiquitously expressed protein factors that affect spliceosome formation can lead to diseases with specific neurological manifestations,

such as spinal muscular atrophy and retinitis pigmentosa. For example, mutations in SMN1 or SMN2 cause spinal muscular atrophy. Defects in genes essential for spliceosomal assembly, including HPRP3, PRPF31, and PRPC8, have been found in autosomal dominant retinitis pigmentosa. The molecular mechanisms underlying such neuronal-specific diseases caused by defects in the general splicing factors remain to be investigated.

Psychiatric disorders, including attention deficit/hyperactivity disorder (ADHD) and schizophrenia, have been associated with aberrant splicing. Imbalance of different splicing isoforms or aberrant splicing products of several genes were detected in the brain tissues of patients with schizophrenia. These genes include $GABA_AR\gamma 2$, NMDAR1, neuronal nicotinic acetylcholine receptors, and neural cell adhesion molecule (NCAM). Several other examples of diseases caused by splicing mutations or disregulation of alternative splicing are listed in Table 1, including cystic fibrosis, familial isolated growth hormone deficiency type II (IGHD II), Frasier's syndrome, Menkes disease, β-thalassemia, and metachromatic leukodystrophy. Changes in splicing isoforms of astroglial AMPA receptors have also been reported in human temporal lobe epilepsy. Again, a variety of aberrant splicing events in the corresponding genes cause the formation of defective gene products and consequently the disease phenotypes.

3.2 Molecular Mechanisms Underlying Splicing Defects Associated with Disease

Genetic defects in pre-mRNA splicing that cause human diseases can be classified into two categories, cis-acting mutations that alter expression or function of single genes and trans-acting defects that affect the components of splicing machinery or regulators of alternative splicing. Both cis-acting and trans-acting splicing mutations can result in clinical manifestations, either primarily involving a single tissue/organ or affecting multiple systems. These mutations can cause diseases by their direct effects on single genes or by indirect mechanisms such as disrupting expression or regulation of multiple genes.

Cis-acting splicing mutations can cause diseases by either producing defective/aberrant transcripts or by simply altering the delicate balance of different naturally expressed splicing isoforms. Development of disease phenotypes can be the result of loss of function of the involved genes or gain-of-function toxicities associated with the aberrant gene products. A large number of studies have focused on relationships between cis-acting splicing mutations and specific defects in the formation or function of the affected genes. However, the relationship is much less understood between genetic defects in trans-acting splicing factors (such as spliceosomal components or splicing regulators) and specific disease phenotypes.

At least four types of molecular mechanisms have been described for cis-acting splicing mutations that cause human diseases (Fig. 6). These mutations often lead to the formation of defective protein products or the loss of function of the genes involved as a result of RNA or protein instability. Exon skipping is perhaps the most commonly reported mechanism for the production of defective gene products. Activation of cryptic splice sites has also been frequently associated with the production of aberrant or defective gene products. The failure to remove introns or intron retention is another mechanism for the

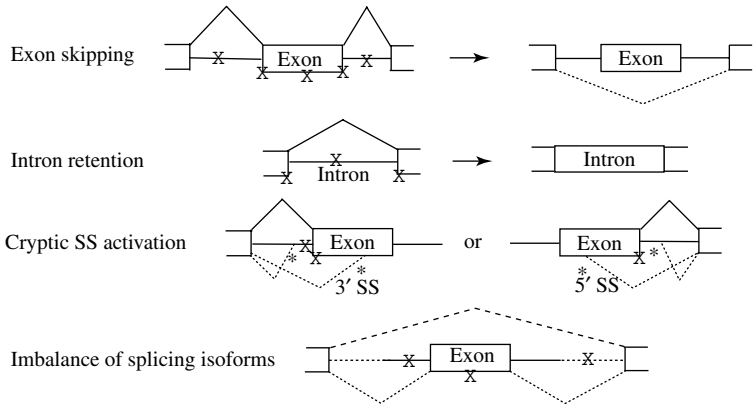

Fig. 6 Mechanisms for cis-acting splicing mutations. Positions of mutations are marked by "X," and they can be at the splice sites or within intronic or exonic sequences. The cryptic splice sites are marked by "*." Normal splicing events are illustrated by solid lines, and aberrant splicing events are depicted in dotted lines.

formation of truncated protein products or the complete loss of gene function, because the inclusion of intronic sequences often introduces premature stop codons in gene transcripts. Finally, disturbance in the balances among naturally occurring splicing isoforms can lead to human disease. These mechanisms are not exclusive of each other. Multiple mutations acting by different mechanisms have been found in the same genes, leading to similar disease phenotypes. For example, both exon skipping and cryptic splice site activation contributing to the formation of defective LDLR lead to hypercholesterolemia. Similarly, exon skipping or cryptic splice site usage in Fibrillin-1 gene have been associated with Marfan syndrome.

Exon skipping can be caused by mutations at the splice sites or in splicing enhancers (either ESEs or ISEs). Translationally silent mutations can be functionally significant in splicing. For example, third codon changes that do not affect peptide sequences have often been overlooked as disease-causing mutations. Recently, such mutations have been examined at the splicing level. Such "silent" mutations may cause significant disruption in splicing, because defective function of ESEs leads to either improper exon skipping or imbalance of natural splicing isoforms.

Activation of cryptic splice sites in a large number of genes has been associated with human pathogenesis (see examples in Table 4). Single-nucleotide changes at the authentic splice sites result in inactivation of authentic splice sites. The consequences can be failures in exon inclusion or intron removal in some genes, or selection/activation of cryptic splice sites in other genes.

Intron retention has been examined in a number of disease genes. The average size of introns in human genes is approximately 3 kbp. The retention of even a small single intron may have catastrophic effects on the function of the genes affected because of the formation of defective or truncated peptides. Alternatively, intron

retention can also cause a complete loss of expression of the mutated genes as a result of nonsense-mediated mRNA decay or instability of the gene products. Tay–Sach's disease can be caused by either intron retention or exon skipping in the splicing of the β-hexosaminidase gene.

Disturbance in the delicate balance of different natural splicing isoforms is now being recognized as a common mechanism for splicing diseases. It has been identified in neurodegenerative diseases, psychiatric disorders, and other diseases. These types of splicing defects can be caused by point mutations at splice junctions or in splicing regulatory elements (enhancers or silencers) located both in exon and intronic regions. FTDP-17, caused by mutations in the human *tau* gene, is a good example of such splicing defects. In addition to mutations that affect biophysical or biochemical function of Tau proteins, a large number of mutations that alter the ratio of natural splicing isoforms have been identified in FTDP-17 patients. The *tau* gene undergoes complex alternative splicing during the development of the nervous system. Six splicing isoforms are produced, three containing exon 10 and three lacking exon 10. These isoforms are named Tau4R and Tau3R respectively, because exon 10 encodes one of four microtubule-binding repeats. The balance among different Tau isoforms is important for normal brain function, and disruption of the balance of Tau isoforms leads to the development of FTDP-17. Maintaining the appropriate balance of different Tau isoforms is controlled by complex interactions between the splicing machinery and various splicing regulatory elements in the *tau* gene. At the 5′ splice site of exon 10, a stem-loop type of secondary structure modulates the recognition of this 5′ splice site by U1snRNP. A number of intronic or exonic mutations destabilizing this stem-loop structure leads to an increase in Tau4R production. Other regulatory elements residing in exon 10, either splicing enhancers or silencers, may be disrupted by point mutations or small deletions, leading to the imbalance of different Tau isoforms. These mutations may increase (such as N279K) or decrease (such as L284L, del280K) the activity of the exonic splicing enhancer.

Other examples of diseases caused by disruption of the balance of different alternative splicing isoforms include Frasier syndrome, familial isolated growth hormone deficiency type II (IGHD II) and atypical cystic fibrosis. Such splicing defects have also been implicated in the pathogenesis of other diseases such as multiple sclerosis and schizophrenia. In Frasier syndrome, intronic mutations are found to change the ratio of +KTS to – KTS isoforms of Wilms tumor suppressor (WT1) gene products. The formation of these two splicing isoforms is the result of selection of two competing 5′ splice sites in exon 9. These two splice sites are separated by 9 nucleotides encoding KTS (lysine-threonine-serine). The use of the upstream and the downstream 5′ splice sites produces – KTS and +KTS isoforms, and the balance of these isoforms is important for kidney and gonad development. Mutations that decrease the use of the downstream +KTS 5′ splice site with an increase in the – KTS isoform have been associated with the majority of Frasier syndrome cases. These examples clearly show the significant role of splicing defects in human pathogenesis.

Spinal muscular atrophy (SMA) is an example that demonstrates the complexity of mechanisms underlying diseases

caused by defects in trans-acting factors, such as proteins essential for spliceosomal snRNP biogenesis. SMA is a leading cause of infant mortality. Pathologically, it is characterized by degeneration of motor neurons in the anterior horn of the spinal cord leading to muscular atrophy. Genetic defects causing SMA include deletions or mutations in the survival of motor neuron genes (SMN). There are two copies of highly homologous SMN genes in humans, *SMN1* and *SMN2*. The *SMN1* gene is deleted or mutated in the majority of SMA patients. The SMN proteins are detected as distinctive speckles termed *gems* within the nucleus. SMN proteins contain an RNP1 motif and are critical for snRNP biogenesis and therefore, the formation of the spliceosome. Both SMN genes are expressed in a wide range of tissues. Although the predicted peptide sequences of *SMN1* and *SMN2* genes are identical, the gene products produced from the two genes in cells are different. A single translationally silent C to T nucleotide change at position +6 of exon 7 leads to the inefficient inclusion of exon 7 in the *SMN2* gene product, and thus the failure of *SMN2* to replace the function of *SMN1* and to provide protection against SMA. It is not clear why mutations in the *SMN* gene specifically cause motor neuron–specific disease. Although the *SMN2* gene is not usually mutated in SMA patients, studies on stimulating *SMN2* exon 7 alternative splicing suggest a potential therapeutic approach based on the activation of alternative splicing of genes homologous to mutated genes (●Sect. 4.2).

Q13

Myotonic dystrophy (DM), an autosomal dominant disease, is an example of trans-acting genetic mutations in which remarkable progress has been made. DM is a most common form of muscular dystrophy affecting both skeletal muscle and smooth muscles. Type I myotonic dystrophy (DM1) is caused by a CTG trinucleotide expansion in the 3′ untranslated region (3′UTR) of the DMPK (DM protein kinase) gene on chromosome 19q13.3. Type II myotonic dystrophy (DM2) is associated with a large CCTG repeat expansion in the intron 1 of the *ZNF9* gene. An ''RNA gain-of-function'' hypothesis has been proposed that these CTG or CCTG repeat expansions cause the formation of aberrant RNA transcripts containing large CUG or CCUG repeats. Such RNA transcripts containing long tracks of CUG/CCUG repeats may disrupt the normal function of certain RNA-binding proteins and induce secondary aberrant splicing of other genes. The RNA-binding proteins involved are likely to be the splicing regulators of CUG-BP family. The disruption of CUG-BP protein functions leads to aberrant splicing of genes including cardiac troponin T, insulin receptor, muscle-specific chloride channel, and tau. The aberrant splicing of these downstream genes can explain the cardiac phenotype, insulin resistance and myotonia found in DM patients. These findings demonstrate the complex roles of alternative splicing regulation in human pathogenesis.

Several trans-acting splicing defects have been reported in autosomal dominant retinitis pigmentosa (adRP). Recent genetic studies demonstrate that mutations in genes encoding general splicing factors such as HPRP3, PRPC8, or PRPF31 cause adRP. It remains unclear why these seemingly general defects in the splicing machinery cause such neuron-specific diseases. The underlying pathogenetic mechanisms await further investigation.

4
Perspectives on Diagnosis and Treatment of Diseases Caused by pre-mRNA Splicing Defects

4.1
Diagnosis of Human Diseases Caused by Splicing Defects

In the past three decades, significant advances have been made in the diagnosis of human genetic diseases, including diseases caused by splicing mutations. However, currently available molecular diagnostic tools remain limited for detecting splicing mutations.

Definitive diagnosis of splicing diseases should be based on the following criteria. First, the clinical manifestations should correlate with defects in a given gene. Second, corresponding splicing mutations are detected in genomic DNA samples of the patients. Third, aberrant or defective splicing products should be detected in the affected tissues or cells from patients. Detection of genomic mutations relies on DNA sequence analysis. In most cases, efforts in genomic DNA sequence analysis have been focusing on either exonic regions or near splice junctions. Detection of aberrant or defective splicing products from patient samples can be technically challenging. The most frequently used methods include RT-PCR and analyses of defective protein products. In general, these methods are still limited to research studies. Systematic approaches to detecting disease-causing splicing mutations are yet to be developed for diagnostic applications in clinics.

4.2
Potential Therapeutic Approaches

Molecular studies and mechanistic characterization of splicing mutations that cause human disease have led to the development of a number of potential therapeutic approaches. They can be classified into four types, oligonucleotide-, ribozyme-, trans-splicing- and chemical compound-based approaches. These approaches are still at the stage of research and development.

4.2.1 Oligonucleotide-based Approaches: Antisense, RNAi, and Chimeric Molecules

Several groups have tested oligonucleotide-based methods for designing therapies to treat diseases caused by aberrant pre-mRNA splicing. These approaches include using modified antisense oligonucleotides or RNA interference (RNAi). Antisense oligonucleotides could be used to restore the proper function of mRNAs that are disrupted by splicing mutations. Antisense oligonucleotides designed to block cryptic sites in the mutated globin gene in β-thalassemia were reported to increase wild-type mRNA and decrease aberrant mRNA. A lentiviral vector-based system using a modified U7snRNA containing a sequence blocking the aberrant splice sites, reduced aberrant pre-mRNAs and increased levels of the correctly spliced β-globin mRNA and protein. The production of stress-induced aberrant AChE-R mRNA was reduced by antisense oligonucleotides. Another study using the antisense oligonucleotide approach reported the reversal of the aberrant splicing pattern caused by mutations associated with FTDP-17 in the human *tau* gene. On the other hand, antisense oligonucleotides have been used to increase the inclusion of exon 7 of SMN2 in order to develop therapeutic agents for spinal muscular atrophy. 2'-O-Methylated antisense oligoribonucleotides were used to modify the splicing pattern of the dystrophin pre-mRNA in the mdx mouse model of Duchenne muscular dystrophy.

Recently, RNAi was used to selectively degrade alternatively spliced mRNA isoforms in *Drosophila* by treating cultured cells with dsRNA corresponding to an alternatively spliced exon.

One of most common splicing defects associated with human diseases is exon skipping. Because an RS domain can act as a splicing activator, small chimeric molecules containing a minimal synthetic RS domain covalently linked to an antisense moiety have been tested to target defective BRCA1 or SMN2 pre-mRNA transcripts and shown to restore splicing. These oligonucleotide-based approaches have shown feasibility in correcting splicing defects and treating the corresponding splicing diseases.

4.2.2 **Ribozymes**

Ribozymes are RNAs that catalyze biochemical reactions in cells, especially cleavage of other nucleic acids. Efforts have been made in developing derivatives from small naturally occurring RNAs, including the hammerhead, the hairpin, the tRNA processing ribonuclease P (RNase P), and group I and group II ribozymes as therapeutic agents. Ribozymes can be used to reduce aberrant or defective splicing transcripts. In addition, a trans-splicing group I ribozyme was shown to convert mutant transcripts to normal mRNAs in the beta-globin and p53 genes.

4.2.3 **SMaRT**

Spliceosome-mediated RNA trans-splicing (SMaRT) was developed utilizing the endogenous trans-splicing activity in mammalian cells to correct aberrant splicing. SMaRT-mediated repair was reported to partially correct splicing defects in cystic fibrosis transmembrane conductance (CFTR) gene in cultured cells and bronchial xenografts.

4.2.4 **Chemical Compounds**

A number of chemical compounds have been shown to interact with RNA and/or RNA-binding proteins. Using chemical compounds to correct pre-mRNA splicing defects is being actively explored as a new therapeutic approach. Although the underlying mechanisms remain unclear, aclarubicin and sodium butyrate increase the inclusion of exon 7 in SMN2 transcripts in fibroblasts derived from spinal muscular atrophy patients or in transgenic mice, suggesting the therapeutic potential of chemical compounds in treating diseases associated with defective splicing.

Other chemical compounds are being tested that are capable of modifying splicing regulators, for example, kinases or phosphatases, that influence SR protein functions.

5 Concluding Remarks

After more than two decades of studies, a general picture of mammalian pre-mRNA splicing and alternative splicing regulation has begun to emerge. The basic components of the mammalian splicing machinery have been identified. The highly dynamic process of spliceosome assembly involves multiple networks of RNA–RNA, RNA–protein and protein–protein interactions. Recognition of splice sites and splicing regulatory sequences has been investigated in a number of genes, leading to identification and characterization of both cis-acting regulatory elements and trans-acting factors. We have begun to appreciate the contribution of alternative pre-mRNA

splicing to creating genetic diversity, especially in mammals. A large number of splicing mutations that cause human diseases are being identified and characterized. We now have a glimpse of the complex picture of the involvement of aberrant or defective splicing in the pathogenesis of human diseases. Furthermore, efforts are being made to improve the diagnosis and treatment of diseases associated with pre-mRNA splicing defects.

Despite the significant progress, we are only at the beginning stage of understanding the molecular mechanisms controlling pre-mRNA splicing and alternative splicing regulation. A number of important questions remain to be addressed.

For the majority of genes, we have little knowledge about their complete expression profiles of different splicing isoforms in different cell types. We do not know how the splicing events of different genes are coordinated during development or in response to environmental changes. How a cell senses the environmental stimuli and responds by producing different splicing products remains largely unknown. Molecular pathways that transduce extracellular signals into the nuclear splicing machinery have not yet been delineated. Further development of new technologies such as using microarray approaches to examine splicing isoforms under different physiological or pathological conditions at the genome level may be important for understanding the role of alternative pre-mRNA splicing in the biology of mammalian cells.

The basic components of mammalian splicing machinery are now characterized at the molecular level. However, the dynamic interactions among different components of the splicing machinery during spliceosome assembly and recognition of splicing regulatory elements are far from being understood. Little is known about the structural basis of such multicomponent interactions. Furthermore, the catalytic mechanisms of pre-mRNA splicing remain to be elucidated.

On the basis of studies of the relatively few model genes, a number of cis-acting elements and trans-acting splicing regulators have been identified. The mechanisms by which the splicing machinery specifically recognizes authentic versus decoy, or pseudosplice sites remain unclear. Some of the splicing regulatory factors are also important players for spliceosome assembly. Spliceosomal proteins also play a role in other processes of gene regulation. Further work needs to be carried out to understand the relationship between pre-mRNA splicing and other processes of gene expression and regulation, including transcription, RNA editing, RNA transport, translational control, and posttranslational modification.

Studies of human diseases caused by splicing mutations or aberrant splicing have significantly advanced our understanding of human genetic diseases. Tremendous effort is still required to understand pathogenetic mechanisms underlying diseases caused by splicing defects, to develop efficient diagnostic tools and to design effective therapeutic approaches for these diseases.

Acknowledgment

We apologize to readers and our colleagues for having referred only to reviews instead of original research articles because of the space limit. The work in the authors' laboratory has been generously supported by Leukemia and Lymphoma Society, Muscular Dystrophy Association, Alzheimer's Association, the Society for Progressive Supranuclear Palsy, and NIH (AG17518,

GM53945 and EY014576). We thank Drs Chris Burge, Reinhard Lührmann, Robin Reed, Phil Sharp, and Joan Steitz for sharing information and for giving permissions to modify figures from their previous publications. We are grateful to Drs Edward Benz, Doug Black, James Bruzik, Mariano Garcia-Blanco, Warren Gish, David Ornitz, Arnold Strauss, and members of Wu laboratory for their critical reading of the manuscript. We thank German G. Leparc, Joe Gee Seo, and Ying Zhang for the data bank searches and Mrs. Angie Hantak for outstanding assistance.

Bibliography

Books and Reviews

Q14 Adams●, M.D., Rudner, D.Z., Rio, D.C. (1996) Biochemistry and regulation of pre-mRNA splicing, *Curr. Opin. Cell Biol.* **8**((3)), 331–339.

Bartel, F., Taubert, H., Harris, L.C. (2002) Alternative and aberrant splicing of MDM2 mRNA in human cancer, *Cancer Cell* **2**((1)), 9–15.

Baserga, S.J., Yang, X.D., Steitz, J.A. (1993) The Diverse World of Small Ribonucleoproteins, in: Gesteland, R.F., Atkins, J.F. (Eds) *The RNA World*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 359–381.

Berget, S.M. (1995) Exon recognition in vertebrate splicing, *J. Biol. Chem.* **2**, 2411–2414.

Bilbao, D., Valcarcel, J. (2003) Getting to the heart of a splicing enhancer, *Nat. Struct. Biol.* **10**(1), 6–7.

Black, D.L. (2000) Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology, *Cell* **103**(3), 367–370.

Black, D.L. (2003) Mechanisms of alternative pre-mRNA splicing, *Annu. Rev. Biochem.* **72**, 291–336.

Black, D.L., Grabowski, P.J. (2003) Alternative pre-mRNA splicing and neuronal function, *Prog Mol. Subcell Biol.* **31**, 187–216.

Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases, *Trends Biochem. Sci.* **25**(3), 106–110.

Brett, D., Pospisil, H., Valcarcel, J., Reich, J., Bork, P. (2003) Alternative splicing and genome complexity, *Nat. Genet.* **30**(1), 29–30.

Brow, D.A. (2002) Allosteric cascade of spliceosome activation, *Annu. Rev. Genet.* **36**, 333–360.

Bruzik, J.P., Van Doren, K., Hirsh, D., Steitz, J.A. (1988) Trans splicing involves a novel form of small nuclear ribonucleoprotein particles, *Nature* **335**(6190), 559–562.

Buchholz, T.A., Wazer, D.E. (2002) Molecular biology and genetics of breast cancer development: a clinical perspective, *Semin. Radiat. Oncol.* **12**(4), 285–295.

Burge, C.B., Tuschl, T.H., Sharp, P.A. (1999) in: Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds) *The RNA World*, Cold Spring Harbor Laboratory Press, New York, pp. 525–560.

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

Cartegni, L., Chew, S.L., Krainer, A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing, *Nat. Rev. Genet.* **3**(4), 285–298.

Chabot, B., LeBel, C., Hutchison, S., Nasim, F.H., Simard, M.J. (2003) Heterogeneous nuclear ribonucleoprotein particle A/B proteins and the control of alternative splicing of the mammalian heterogeneous nuclear ribonucleoprotein particle A1 pre-mRNA, *Prog. Mol. Subcell Biol.* **31**, 59–88.

Collins, C.A., Guthrie, C. (2000) The question remains: is the spliceosome a ribozyme? *Nat. Struct. Biol.* **7**(10), 850–854.

Cooper, T.A., Mattox, W. (1997) The regulation of splice-site selection, and its role in human disease, *Am. J. Hum. Genet.* **61**(2), 259–266.

Dredge, B.K., Polydorides, A., Darnell, R.B. (2001) The splice of life: alternative splicing and neurological disease, *Nat. Neurosci. Rev.* **2**, 43–50.

Faustino, N.A., Cooper, T.A. (2003) Pre-mRNA splicing and human disease, *Genes Dev.* **17**(4), 419–437.

Forch, P., Valcarcel, J. (2003) Splicing regulation in Drosophila sex determination, *Prog. Mol. Subcell Biol.* **31**, 127–151.

Fu, X.D. (1995) The superfamily of arginine/serine-rich splicing factors, *RNA* **1**(7), 663–680.

Goedert, M., Spillantini, M.G. (2001) Tau gene mutations and neurodegeneration, *Biochem. Soc. Symp.* **67**, 59–71.

Goldstrohm, A.C., Greenleaf, A.L., Garcia-Blanco, M.A. (2001) Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing, *Gene* **277**(1–2), 31–47.

Grabowski, P.J. (1998) Splicing regulation in neurons: tinkering with cell-specific control, *Cell* **92**(6), 709–712.

Grabowski, P.J., Black, D.L. (2001) Alternative RNA splicing in the nervous system, *Prog. Neurobiol.* **65**, 289–308.

Graveley, B.R. (2000) Sorting out the complexity of SR protein functions, *RNA* **6**(9), 1197–1211.

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

Holmila, R., Fouquet, C., Cadranel, J., Zalcman, G., Soussi, T. (2003) Splice mutations in the p53 gene: case report and review of the literature, *Hum. Mutat.* **21**(1), 101–102.

Keegan, L.P., Gallo, A., O'Connell, M.A. (2001) The many roles of an RNA editor, *Nat. Rev. Genet.* **2**(11), 869–878.

Konarska, M.M. (1998) Recognition of the 5′ splice site by the spliceosome, *Acta Biochim. Pol.* **45**(4), 869–881.

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

Ladd, A.N., Cooper, T.A. (2002) Finding signals that regulate alternative splicing in the post-genomic era, *Genome Biol.* **3**(11), •reviews0008.

Q15

Maniatis, T., Tasic, B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans, *Nature* **418**(6894), 236–243.

Maniatis, T., Reed, R. (2002) An extensive network of coupling among gene expression machines, *Nature* **416**(6880), 499–506.

Manley, J.L. (2002) Nuclear coupling: RNA processing reaches back to transcription, *Nat. Struct. Biol.* **9**(11), 790–791.

Manley, J.L., Tacke, R. (1996) SR proteins and splicing control, *Genes Dev.* **10**(13), 1569–1579.

Maroney, P.A., Romfo, C.M., Nilsen, T.W. (2002) Functional recognition of 5′ splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly, *Mol. Cell* **6**, 317–328.

Moore, M.J. (2000) Intron recognition comes of AGe, *Nat. Struct. Biol.* **7**(1), 14–16.

Mount, S.M. (2000) Genomic sequence, splicing, and gene annotation, *Am. J. Hum. Genet.* **67**(4), 788–792.

Mount, S.M., Salz, H.K. (2000) Pre-messenger RNA processing factors in the *Drosophila* genome, *J. Cell Biol.* **150**, F37–F43.

Nilsen, T.W. (2002) The spliceosome: no assembly required? *Mol. Cell* **9**(1), 8–9.

Nissim-Rafinia, M., Krem, B. (2002) Splicing regulation as a potential genetic modifier, *Trends Genet.* **18**, 123–127.

Paushkin, S., Gubitz, A.K., Massenet, S., Dreyfuss, G. (2002) The SMN complex, an assembly of some of ribonucleoproteins, *Curr. Opin. Cell Biol.* **14**(3), 305–312.

Proudfoot, N.J., Furger, A., Dye, M.J. (2002) Integrating mRNA processing with transcription, *Cell* **108**, 501–512.

Reed, R. (1996) Initial splice-site recognition and pairing during pre-mRNA splicing, *Curr. Opin. Genet. Dev.* **6**(2), 215–220.

Reed, R. (2000) Mechanisms of fidelity in pre-mRNA splicing, *Curr. Opin. Cell Biol.* **12**(3), 340–345.

Reed, R., Hurt, E. (2002) A conserved mRNA export machinery coupled to pre-mRNA splicing, *Cell* **108**, 523–531.

Roberts, G.C., Smith, C.W. (2002) Alternative splicing: combinatorial output from the genome, *Curr. Opin. Chem. Biol.* **6**(3), 375–383.

Rosonina, E., Blencowe, B.J. (2002) Gene expression: the close coupling of transcription and splicing, *Curr. Biol.* **12**(9), R319–R321.

Sharp, P.A. (1994) Split genes and RNA splicing, *Cell* **77**(6), 805–815.

Smith, C.W., Valcarcel, J. (2000) Alternative pre-mRNA splicing: the logic of combinatorial control, *Trends Biochem. Sci.* **25**(8), 381–388.

Soret, J., Tazi, J. (2003) Phosphorylation-dependent control of the pre-mRNA splicing machinery, *Prog. Mol. Subcell Biol.* **31**, 89–126.

Staley, J.P., Guthrie, C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things, *Cell* **92**(3), 315–326.

Steitz, J.A. (1984) Autoantibody probes for mammalian gene expression, *Harvey Lect.* **80**, 39–47.

Stoilov, P., Meshorer, E., Gencheva, M., Glick, D., Soreq, H., Stamm, S. (2002) Defects in pre-mRNA processing as causes of and predisposition to diseases, *DNA Cell Biol.* **21**(11), 803–818.

Tacke, R., Manley, J.L. (1999) Determinants of SR protein specificity, *Curr. Opin. Cell Biol.* **11**(3), 358–362.

Tarn, W.Y., Steitz, J.A. (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge, *Trends Biochem. Sci.* **22**(4), 132–137.

Tupler, R., Perini, G., Green, M.R. (2001) Expressing the human genome, *Nature* **409**(6822), 832–833.

Valadkhan, S., Manley, J.L. (2002) Intrinsic metal binding by a spliceosomal RNA, *Nat. Struct. Biol.* **9**(7), 498–499.

Valcarcel, J., Green, M.R. (1996) The SR protein family: pleiotropic functions in pre-mRNA splicing, *Trends Biochem. Sci.* **21**(8), 296–301.

Villa, T., Pleiss, J.A., Guthrie, C. (2002) Spliceosomal snRNAs: Mg(2+)-dependent chemistry at the catalytic core? *Cell* **109**(2), 149–152.

Wagner, E.J., Garcia-Blanco, M.A. (2001) Polypyrimidine tract binding protein antagonizes exon definition, *Mol. Cell. Biol.* **21**(10), 3281–3288.

Wang, J., Manley, J.L. (1997) Regulation of pre-mRNA splicing in metazoa, *Curr. Opin. Genet. Dev.* **7**(2), 205–211.

Will, C.L., Luhrmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function, *Curr. Opin. Cell Biol.* **13**(3), 290–301.

Wu, J.Y., Tang, H., Havlioglu, N. (2003) Alternative pre-mRNA splicing and regulation of programmed cell death, *Prog. Mol. Subcell Biol.* **31**, 153–185.

mcb200300189

QUERIES TO BE ANSWERED BY AUTHOR (SEE MARGINAL MARKS Q..)

IMPORTANT NOTE: Please mark your corrections and answers to these queries directly onto the proof at the relevant place. Do NOT mark your corrections on this query sheet.

Q1. Please clarify if Fig. 1 can be cited elsewhere in text so that the Introduction can be taken as the concise definition of the chapter.
Q2. Please provide Fig. 3b.
Q3. Please provide the complete details for this reference.
Q4. Some of accession numbers were underlined in the original manuscript. We have removed the same. Please clarify if it is fine.
Q5. Please provide the complete details for this reference.
Q6. We have changed ''genbank'' to ''GenBank''. Please clarify if this is fine.
Q7. We have changed ''predication'' to ''prediction''. Please clarify if this is fine.
Q8. We have deleted ''tract'' here. Please clarify if this is fine.
Q9. We have cited sect. 2.2.2 here. Please clarify if this is fine.
Q10. We have cited Sect. 3.1 here. Please clarify if this is fine.
Q11. We have cited Sect. 4.2 here. Please clarify if this is fine.
Q12. We have changed table 5 to table 4, as there are only 4 tables in this chapter. Please clarify if this is fine.
Q13. We have cited Sect. 4.2 here. Please clarify if this is fine.
Q14. As per the style of this encyclopedia, you are required to provide 50 to 100 primary research paper references. Please do so.
Q15. Please provide the page range for this reference.