

REVIEW

SR proteins and galectins: what's in a name?

Kevin C Haudek², Ronald J Patterson³,
and John L Wang^{1,2}

²Department of Biochemistry and Molecular Biology and ³Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI 48824, USA

Received on April 30, 2010; revised on June 15, 2010; accepted on June 15, 2010

Although members of the serine (S)- and arginine (R)-rich splicing factor family (SR proteins) were initially purified on the basis of their splicing activity in the nucleus, there is recent documentation that they exhibit carbohydrate-binding activity at the cell surface. In contrast, galectins were isolated on the basis of their saccharide-binding activity and cell surface localization. Surprisingly, however, two members (galectin-1 and galectin-3) can be found in association with nuclear ribonucleoprotein complexes including the spliceosome and, using a cell-free assay, have been shown to be required splicing factors. Thus, despite the difference in terms of their original points of interest, it now appears that members of the two protein families share four key properties: (a) nuclear and cytoplasmic distribution; (b) pre-mRNA splicing activity; (c) carbohydrate-binding activity; and (d) cell surface localization in specific cells. These findings provoke stimulating questions regarding the relationship between splicing factors in the nucleus and carbohydrate-binding proteins at the cell surface.

Keywords: carbohydrate-binding/lectins/RNA processing/spliceosome

Annotations: name and location

The availability of genome sequences of a variety of species, the development of high throughput screening methodologies such as DNA microarrays, and the refinement of computational algorithms for searching large databases have all stimulated more and more genome-wide investigations. In turn, the interpretation of the results of such studies yielding large data sets has depended on the annotation of genes, which suggests relationships in terms of subcellular localization, biochemical activity, or connectivity via a metabolic pathway. Most of the annotations in the databases, however, are predicated on two key tenets in the “classical” approach in naming a gene or pro-

tein: (a) the assay used in the initial purification of the protein; and (b) localization in a subcellular compartment consistent with expectations.

More recently, it has become evident that many proteins comprise multiple distinct domains and can interact with a bewildering number of other moieties ranging from other polypeptide sequences to carbohydrate and nucleic acid sequences. In addition, members of many different classes of proteins (secreted growth factors, transcription factors, cytoskeletal proteins, and glycolytic enzymes) have been identified in cellular compartments other than their “classical” site of action. In the present essay, we focus on two protein families that have recently been shown to share localization and functional similarities: carbohydrate-binding at the cell surface and RNA processing inside the nucleus.

Two protein families: SR proteins and galectins

Splicing factor 2/alternative splicing factor (SF2/ASF) was initially identified as an activity required for constitutive splicing (Krainer et al. 1990) and as a factor that could modulate 5'-splice site selection (Ge and Manley 1990). SF2/ASF (Ge et al. 1991; Krainer et al. 1991), now designated Sfrs1 (Table I), represents the prototype of a family of splicing factors containing the RS domain, rich in arginine and serine dipeptides. Additional RS-domain-containing proteins were identified by a monoclonal antibody (mAb 104) which recognizes a phosphoepitope shared by the SR proteins (Zahler et al. 1992). These include spliceosomal component 35 (SC35; Sfrs2), another factor required for spliceosome assembly (Fu and Maniatis 1990). The SR protein family now contains seven “canonical” members, using the criteria described by Long and Caceres (2009). In addition to this canonical SR family, other SR-related families have been described, such as non-canonical SR proteins, SR-related proteins, and RS-domain-containing proteins (for recent reviews, see Long and Caceres 2009; Shepard and Hertel 2009). In the present essay, we use the name SR protein family to refer to the seven “canonical” SR proteins. Although a common nomenclature and a numbering system has been designated for the seven members of the SR protein family, investigators in the field of splicing have continued, for the most part, using the original name corresponding to each specific SR protein (Table I). All eukaryotes express some SR family member or recognizable SR-like proteins.

The SR proteins exhibit a modular structure (Figure 1A) with one or two RNA recognition motifs (RRMs) that provide the basis for RNA-binding activity. In cases where it has been analyzed, individual SR proteins exhibit specific yet degener-

¹To whom correspondence should be addressed: Tel: +1-517-353-9542; Fax: +1-517-353-9334; E-mail: wangj@cns.msu.edu

Table I. Polypeptide molecular weight and subcellular localization of SR proteins^a

Name	Polypeptide molecular weight (kDa)	Nuclear	Cytoplasmic	Extracellular
Sfrs1 (ASF/SF2)	27	+	+	+
Sfrs2 (SC35)	24	+	^b	+
Sfrs3 (SRp20)	18	+	+	
Sfrs4 (SRp75)	54	+	+	
Sfrs5 (SRp40)	30	+	+	+
Sfrs6 (SRp55)	38	+	+	
Sfrs7 (9G8)	26	+	+	+

^aBlank boxes in the table represent “not known” due to lack of documented evidence.

^bUnlike the other members of the SR protein family, Sfrs2 does not undergo nucleocytoplasmic shuttling (see text).

ate RNA-binding specificities (Liu et al. 1998, 2000; Sanford et al. 2008). The carboxyl terminal RS domain is characterized as being “disordered” (Haynes and Iakoucheva 2006). Nevertheless, this domain participates in either protein–protein or protein–RNA interactions in assembly of the spliceosome, as well as in targeting the polypeptide to the nucleus (see below).

The galectins are a family of carbohydrate-binding proteins that share two key properties: (a) binding affinity for β-galactosides; and (b) conserved sequence elements in the saccharide-binding site (Barondes et al. 1994). To date, 15 mammalian galectins have been identified (Table II). Galectins have also been identified in many non-mammalian species, including birds, amphibians, fish, worms, sponges, and fungi. Screening the databases of genomic DNA sequences and expressed sequence tags has revealed additional candidates for membership in the mammalian galectin family, as well as putative galectins in plants and viruses (Cooper 2002).

Each member of the galectin family contains at least one domain of about 130 amino acids; this domain binds to saccharides and is designated the carbohydrate recognition domain (CRD) (Figure 1B). Based on the number and organi-

Table II. Polypeptide molecular weight and subcellular localization of galectins^a

Name	Polypeptide molecular weight (kDa)	Nuclear	Cytoplasmic	Extracellular
Galectin-1	14	+	+	+
Galectin-2	14	+	+	
Galectin-3	30	+	+	+
Galectin-4	36	+	+	+
Galectin-5	16		+	
Galectin-6	33		+	
Galectin-7	15	+	+	+
Galectin-8	35	+	+	+
Galectin-9	35		+	+
Galectin-10	17	+	+	+
Galectin-11	15	+	+	+
Galectin-12	36	+	+	
Galectin-13	16	+	+	
Galectin-14	17	+	+	+
Galectin-15	14		+	+

^aBlank boxes in the table represent “not known” due to lack of documented evidence.

zation of domains in the polypeptides, the galectins have been classified into subfamilies (Hirabayashi and Kasai 1993): (a) the prototype group contains one domain, the CRD; (b) the chimera group contains a proline (P)- and glycine (G)-rich domain (also about 130 amino acids) fused amino terminal to the CRD; and (c) the tandem repeat group contains two CRDs.

In the amino acid sequence of galectin-3 (Gal3), the sole representative of the chimera group, the P- and G-rich domain is characterized by multiple internal sequence homologies, each of which consists of a nine-residue repeat with a consensus sequence of Pro–Gly–Ala–Tyr–Pro–Gly, followed by three additional amino acids. Differential scanning calorimetry of the murine Gal3 polypeptide yielded distinct transition temperatures for the amino terminal, P- and G-rich domain (~40°C), and the carboxy-terminal CRD (~55°C), both in the full-length polypeptide and as isolated preparations of

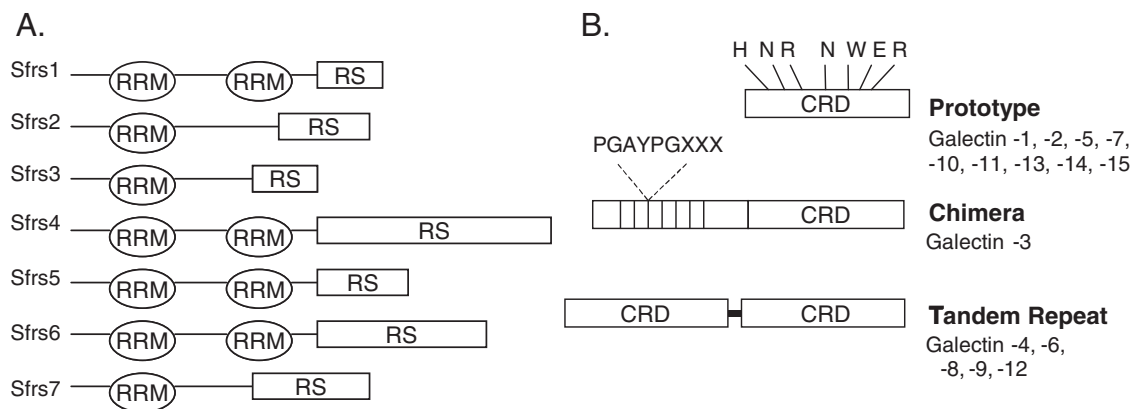


Fig. 1. Diagram illustrating the domain content and organization of the polypeptides of the SR splicing factor family and the galectin family. (A) The seven canonical members of the SR protein family are arranged in the order of the common nomenclature, Sfrs1 through Sfrs7. RRM, RNA recognition motif; RS, domain rich in arginine and serine dipeptides. (B) The 15 members of the galectin family are classified into three subfamilies, based on the number and arrangement of domains. CRD, carbohydrate recognition domain. Conserved amino acid residues in the CRD are highlighted. The amino terminal domain of galectin-3, the sole representative of the Chimera group, contains a repeating motif rich in proline and glycine residues. The single amino acid code is used; X denotes any amino acid.

individual domains (Agrwal et al. 1993). The lower transition temperature of the P- and G-rich domain of Gal3 suggests a “less-ordered” structure, like the RS domain.

Four properties shared by SR proteins and galectins

Binding of carbohydrates

Recently, Hatakeyama et al. (2009) reported that Sfrs1, Sfrs2, Sfrs5, and Sfrs7 are expressed at the cell surface of lung endothelial cells (see below). Even more striking was the observation that these SR proteins exhibited carbohydrate-binding activity. In previous studies, a short peptide (IELL-QAR), designated I-peptide, was shown to be a carbohydrate mimicry peptide and inhibited carbohydrate-dependent lung colonization of sialyl Lewis X-expressing B16-FTIII-M cells (B16 melanoma cells transfected with fucosyltransferase 3) in E-selectin and P-selectin doubly deficient mutant mice (Fukuda et al. 2000; Zhang et al. 2002). When a membrane protein fraction from rat lung was subjected to I-peptide affinity chromatography, the four SR splicing factors were isolated and identified by mass spectrometry. Recombinant Sfrs1 bound to B16-FTIII-M cells (expressing the sialyl LeX carbohydrate) but not to parental B16 cells. Thus, it appears that endothelial cells expressing cell surface SR splicing factors were responsible for the carbohydrate-dependent binding and lung colonization of the sialyl LeX-expressing B16-FTIII-M cells.

Using a glycan array, it was determined that Sfrs1 bound a series of fucosylated oligosaccharides, including H-type 1, H-type 2, LeX, LeY, and LeB structures (Hatakeyama et al. 2009). Sfrs1 also exhibited strong binding to α -rhamnose (6-deoxy-mannose); this was rationalized in terms of the structural similarity of this saccharide to fucose (6-deoxy-galactose). These results provoke questions and investigations at several levels. First, do other members of the SR protein family also bind carbohydrates? Second, what is the structural basis for the observed carbohydrate-binding activity of Sfrs1? Finally, since it was reported that Sfrs1 requires calcium for binding to saccharides, is there a calcium-dependent CRD as seen in the C-type lectins (Zelensky and Gready 2005)?

The carbohydrate-binding specificity of various purified galectins has been studied extensively and interpreted in terms of the three-dimensional structures which have been determined for CRDs derived from galectin-1 (Gal1), galectin-2, Gal3, galectin-7, and galectin-10 (see Leffler et al. 2004 for an overview). The CRDs all show a highly conserved tight fold, with two anti-parallel β -pleated sheets forming a sandwich-like structure (Rini and Lobsanov 1999). Amino acid side chains on one of these sheets form the core carbohydrate-binding site; the highly conserved residues responsible for saccharide-binding include H158, N160, R162, N174, W181, E184, and R186 (residue numbering from the human Gal3 sequence as reported in Seetharaman et al. 1998).

The composite picture derived from ligand binding and structural studies suggests that the binding site is composed of at least four subsites, each of which interacts with a single monosaccharide unit (Leffler and Barondes 1986; Knibbs et al. 1993; Hirabayashi et al. 2002). More recently, this has been refined to six subsites (designated as A, B, C, D, D', and E) by Stowell et al. (2008), who conducted a detailed dose-

response analysis of the binding of Gal1, galectin-2, and Gal3 to a microarray containing hundreds of structurally diverse glycans. The major and most conserved feature in the interaction of a galectin polypeptide with saccharides is the binding of a galactose residue in subsite C. This includes direct interaction of the monosaccharide with six of the seven most highly conserved amino acid residues listed above. The binding of a second monosaccharide unit in subsite D is also significant inasmuch as there is an approximately 100-fold higher affinity for lactose than for galactose. Subsite D allows for variations such as in lactose (Gal β 1-4Glc) versus lactosamine (Gal β 1-4GlcNAc). Differences in the specificities of the galectins are often reflected in the preference of subsite D for different monosaccharides. Other fine features of the specificity of galectins can be additionally accounted for by their ability to accommodate different saccharides (Gal, GalNAc, GlcNAc, Fuc, or NeuAc) or aglycone groups (e.g., sulfate) in subsite B, with further extensions in subsite A.

As an example, Gal3 binds type 1 or type 2 Gal β 1 \rightarrow 3(4) GlcNAc chains, and the affinity for straight-chain polylectosamine structures or complex-type branched glycans is increased over the simple disaccharide. Fucosylation or sialylation of terminal galactose residues or further substitution by (α 1 \rightarrow 3)-linked Gal or GalNAc does not affect binding or may even enhance it, whereas substitution (occupying subsite D') of the penultimate GlcNAc residue in subsite D drastically reduces binding. Thus, blood group A, H-type 1, or H-type 2 structures show high affinity, whereas LeA and LeX structures bind poorly. While Gal1 and galectin-2 share some common binding characteristics with Gal3, it is important to note that the study of Stowell et al. (2008) highlights the fact that, in mechanistic detail, each of the galectins differs in binding to glycans on microarrays and on cell surfaces. These differences provide the basis for unique and/or opposing biological activities.

Cell surface expression and extracellular localization

In addition to the isolation of Sfrs1, Sfrs2, Sfrs5, and Sfrs7 as carbohydrate-binding receptors, several other lines of evidence are consistent with the notion that SR proteins are expressed at the cell surface of endothelial cells (Hatakeyama et al. 2009). First, mice were injected with anti-Sfrs1 antibodies through the tail vein. Lung tissue sections prepared from such mice yielded fluorescent capillaries when visualized with fluorescence-conjugated secondary antibody. In contrast, analysis of similar sections from mice injected with irrelevant primary antibody yielded no fluorescence staining. Second, proteins exposed to the luminal surface of blood vessels of a mouse were labeled using a biotinylation reagent injected through the tail vein (Rajotte and Ruoslahti 1999). When the in vivo biotinylated lung endothelial surface proteins were immunoprecipitated with anti-Sfrs1, the precipitated Sfrs polypeptides can be detected by peroxidase-conjugated avidin. Finally, as discussed above, Sfrs1, Sfrs2, Sfrs5, and Sfrs7 bind to the carbohydrate mimic, I-peptide. Mice were injected intravenously, first with anti-Sfrs antibodies and then followed by I-peptide displaying phage particles (Fukuda et al. 2000). The binding of the I-phage to the lung was inhibited by anti-Sfrs but not by control normal goat immunoglobulin. All of these results indicate that the SR proteins are expressed on the luminal cell surface of lung capillaries.

With the discovery of Hatakeyama et al. (2009), the SR family of splicing factors joins a surprisingly long and rapidly growing list of diverse proteins that have been reported to exhibit dual localization (Figure 2A and B), being found in both the extracellular (cell surface and medium) and intracellular (cytoplasm and nucleus) compartments of cells (Arnoys and Wang 2007). The mechanism by which the SR proteins are externalized, however, must be unusual inasmuch as there does not appear to be a signal sequence to direct the polypeptide into the endomembrane pathway for cell surface expression and/or secretion. An unconventional secretion pathway has also been postulated for the galectin family (Hughes 1999; Nickel and Rabouille 2009), in which 10 of the members are found outside of cells (Table II and Figure 2C) despite the fact that the amino acid sequences of the polypeptides do not exhibit an obvious signal sequence.

The literature on the galectins has been dominated by studies focused on their activity on the extracellular side (for reviews, see Leffler et al. 2004 and Rabinovich et al. 2007). The present essay will illustrate three salient features using Gal3 as an example. First, Gal3 binds to oligosaccharide structures on cell surface or extracellular matrix proteins to modulate cell-to-cell or cell-substrate adhesion. For example, Gal3 induces neutrophil adhesion to laminin by bridging the cell surface carbohydrates to the adhesive glycoprotein (Kuwabara and Liu 1999). This kind of bridging function may be particularly important in organogenesis. The extracellular matrix protein hensin is responsible for the differentiation of a clonal, intercalated epithelial cell line to form columnar structures and develop specialized apical membrane structures (microvilli)

and functions (regulated endocytosis or exocytosis). Upon terminal differentiation, the epithelial cells externalize Gal3, which bundles hensin to maintain polarity in the differentiated state (Hikita et al. 2000).

Second, Gal3 binds to saccharide structures on certain cell surface receptors, leading to their cross-linking and activation which, in turn, induces signaling events and physiological responses from within the cell. The binding of Gal3 to CD98 on the T-lymphocyte cell line Jurkat results in the uptake of extracellular Ca^{2+} ions (Dong and Hughes 1996). It has been shown that Gal3 binds to CD32 (the low-affinity IgG receptor FcγRII) on peripheral blood mononuclear cells and induces down-regulation of the IL-5 gene (Cortegano et al. 2000). Gal3 also binds to the high-affinity IgE receptor FcεRI on mast cells and a rat basophilic leukemia cell line, modulating the IgE-dependent activation as assayed by degranulation and serotonin release (Zuberi et al. 1994).

Finally, Gal3 at the cell surface appears to play a supportive role in endocytosis by binding to cubilin, a ~400-kDa membrane protein implicated as the endocytic receptor for a variety of molecules (Crider-Pirkle et al. 2002). Interestingly, cubilin is found only in perforin-containing granules of uterine natural killer cells. By in situ hybridization, cubilin mRNA could be found in the yolk sac epithelium but not in the uterine natural killer cells. These results raise the possibility that yolk sac-derived cubilin is endocytosed by uterine natural killer cells via Gal3.

In the examples cited above to illustrate the three features of the extracellular galectin activity, Gal3 acts as a soluble protein binding to cell surface carbohydrates (Figure 2C). On the other

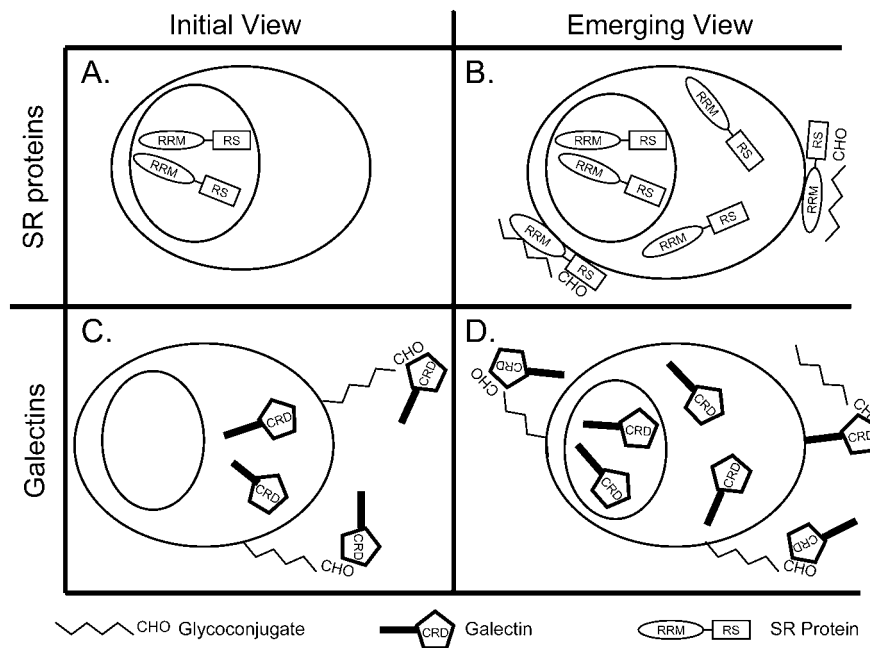


Fig. 2. Schematic comparison of the “initial” and the “emerging” views of the localization and activities of the SR splicing factor and the galectin protein families. (A) The SR proteins were initially identified as splicing factors in the nucleus. (B) It is now known that SR proteins are also found as components of mRNPs in the cytoplasm as well as cell-surface carbohydrate-binding proteins. (C) Galectins were initially studied as carbohydrate-binding proteins of the cell surface and extracellular matrix. (D) Galectins have now been localized in the nucleus, and at least two members of the family participate in pre-mRNA splicing. The possibility that cell surface galectin serves as a carbohydrate receptor is also raised in this depiction. RRM, RNA recognition motif; RS, domain rich in arginine and serine dipeptides; CRD, carbohydrate recognition domain; CHO, carbohydrate.

hand, the results of Hatakeyama et al. (2009) document SR proteins as carbohydrate-specific receptors, responsible for the sialyl LeX-dependent binding and lung colonization of B16 melanoma cells. Using magnetic glyco-nanoparticles, the binding of specific saccharides to the cell surface of various cell types and tissues has been documented both *in vitro* and *in vivo* via magnetic resonance imaging (van Kasteren et al. 2009; El-Boubbou et al. 2010). Although the identity of the carbohydrate-binding protein(s) responsible for the recognition of each of the specific glyco-nanoparticle has not been definitively established, the data are consistent with the possibility that cell surface SR splicing factors, along with the well characterized E-selectin, P-selectin, and mannose receptors, could collectively contribute to the observed binding. This, in turn, raises the question whether galectins might be cell-surface carbohydrate receptors as well (Figure 2D).

Partitioning between the nucleus and cytoplasm

Early immunofluorescence studies localized Sfrs2 to the nucleoplasm (Table I and Figure 2A) as well as to discrete regions in the nucleus, designated as speckles (Fu and Maniatis 1990). More recent studies have shown that all SR protein family members can be found in these speckles (Caceres et al. 1997; Sapra et al. 2009), which are thought to be sites to which factors involved in transcription and splicing coalesce for storage, recycling, or reactivation during periods of less active transcription (for a review on nuclear speckles, see Zhao et al. 2009). Additional studies have shown that SR proteins move to regions of active transcription and can be recruited to nascent mRNPs as RNA polymerase II is elongating the transcript (Misteli et al. 1997, 1998; Sapra et al. 2009). The RS domain (or even a short stretch of RS dipeptides) is sufficient for targeting to the nucleus; this is mediated by transportin-SR, which shows sequence similarity to the nuclear import receptors importin- α and importin- β (Caceres et al. 1997; Cazalla et al. 2002).

With the exception of Sfrs2, the SR splicing factors undergo nucleocytoplasmic shuttling (Caceres et al. 1998), binding to nascent, nuclear, and cytoplasmic mRNAs, suggesting a role for SR protein members in the biogenesis and activities of various mRNPs (Sapra et al. 2009; Zhong et al. 2009). The nuclear export of the SR proteins has been linked to mRNA export due to the interaction between TAP, an RNA export factor, and some members (Sfrs1, 3, 7) of the SR protein family (Huang et al. 2003). Hypophosphorylated Sfrs1 and 7 preferentially bind to TAP, and this form of Sfrs1 can be found associated with mature mRNPs (Huang et al. 2004; Lia and Tarn 2004). Interestingly, a recent report has implicated arginine methylation of Sfrs1 in the control of its subcellular localization; Sfrs1 mutants that are unable to be methylated and lack positive charge accumulate in the cytoplasm (Sinha et al. 2010). In the cytoplasm, the SR proteins have been shown to stimulate translation (Sanford et al. 2004, 2005), and Sfrs1 has been shown to be polysome-associated (Sanford et al. 2008).

Of the 15 galectins identified, 11 have been reported to be in the nucleus and cytoplasm of cells (Table II and Figure 2D). Several distinct activities have been reported for galectins in the cytoplasm, including membrane anchorage of the H-Ras oncogene product by Gal1 (Paz et al. 2001) and anti-apoptotic

activity of Gal3 (Yang et al. 1996). This diverse set of subjects has been reviewed (Liu et al. 2002; Liu 2004), and the present essay will focus only on those issues for which there is parallelism with the SR splicing factors.

The nuclear versus cytoplasmic distribution reported for each galectin most likely reflects the specific cell type and experimental conditions (source of cells under study, culture conditions, proliferation status of the cell/culture, or neoplastic transformation). For example, Gal3 has been reported to be exclusively cytoplasmic, predominantly nuclear, or distributed between the two subcellular compartments (for a recent review, see Haudek et al. 2010). In fact, Gal3 shuttles between the cytoplasm and nucleus (Davidson et al. 2002), and an importin- α/β -mediated nuclear localization signal and an exportin-1-mediated nuclear export signal have been identified on the polypeptide (Davidson et al. 2006; Li et al. 2006; Nakahara et al. 2006). In contrast to Sfrs1, in which a hypophosphorylated form is preferentially exported (Lia and Tarn 2004), it is the phosphorylated form of Gal3 that is transported during nuclear export (Tsay et al. 1999; Takenaka et al. 2004). On this basis, there must be exquisite mechanisms by which Gal3 localization, transport, and association with distinct subcellular components are regulated.

Like Sfrs2, Gal1 and Gal3 have been localized in the nucleoplasm and in speckles under fluorescence microscopy and to corresponding structures by electron microscopy (Hubert et al. 1995). In fact, double immunofluorescence studies have shown tight co-localization between Gal1, Gal3, and Sfrs2 (Vyakarnam et al. 1997, 1998). Particularly striking was the observation that Gal3 in the nucleus was dependent on the integrity of ribonucleoprotein complexes (Laing and Wang 1988). More recent studies have identified a Gal3-U1 small nuclear ribonucleoprotein (snRNP) complex, which provides a mechanism of incorporation of the galectin onto the pre-mRNA splicing substrate (Haudek et al. 2009; see below).

Splicing of pre-mRNA

Since its initial discovery, Sfrs1 and the other members of the SR protein family have been investigated intensely in terms of their participation in splicing (for recent reviews, see Long and Caceres 2009; Shepard and Hertel 2009; Zhong et al. 2009). In contrast, much less work and much less information have been accumulated for the galectins with regard to splicing. We will restrict our discussion of the galectins to Gal1 and Gal3, the two members that have been documented as splicing factors (Dagher et al. 1995; Vyakarnam et al. 1997). This disparity is opposite the situation with respect to carbohydrate-binding activity, in which much more has been learned regarding the galectins. Nevertheless, there appears to be significant overlap in terms of what we do know regarding the splicing activity of the two protein families (Table III).

Using cell-free splicing assays, the classic criteria of depletion–reconstitution have established Gal1 and Gal3, as well as all members of the SR family, as required factors in splicing. In general, the SR splicing factors function in defining both the 5'- and 3'-splice sites and bridge factors assembled at both sites to form stable complexes committed to the splicing pathway. Many members of the SR protein family display a redundant function in cell-free assays using standard pre-mRNA substrates (Fu et al. 1992; Zamore et al. 1992; Wang

Table III. Comparison of selected features in splicing displayed by galectins and SR proteins^a

ACTIVITY	GALECTIN FAMILY	SFRS FAMILY
Members active in pre-mRNA splicing	Gal1/Gal3	All
Depletion abolishes splicing activity	Yes	Yes
Recombinant protein reconstitutes splicing	Yes	Yes
Splicing activity redundant with other family members	Yes	Yes/No
Protein domain active in reconstitution	CRD	intact protein
Assembled onto spliceosome throughout pathway	Yes	Yes
Binds pre-mRNA directly	No	Yes
Salt concentration that releases protein from spliceosomes	>130 mM	>250 mM
Members on separate spliceosomes	Yes	No
Members of endogenous RNP complexes in the absence of pre-mRNA scaffold	Yes	Yes
Saccharides inhibit splicing activity	Yes	
Carbohydrate-binding site separate from splicing site	Yes	
Involved in alternative splicing		Yes

^aBlank boxes in the table represent “not known” due to lack of documented evidence.

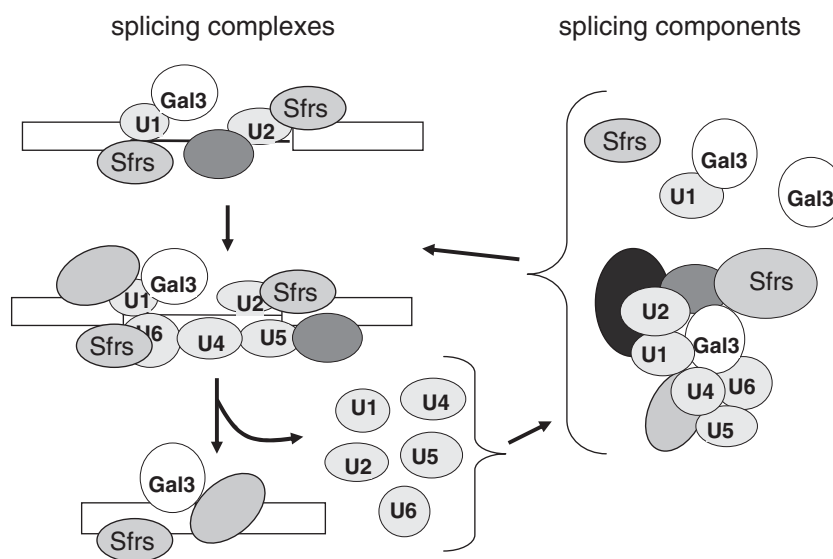


Fig. 3. Diagram illustrating the association of SR splicing factor(s) and galectin(s) with snRNPs and with the pre-mRNA substrate in the assembly of the spliceosome. The pre-mRNA is shown on the left side as two rectangular exons joined by a single-line intron. The SR proteins, represented by Sfrs, can interact with the pre-mRNA via its RNA recognition motif. Galectin-3 (Gal3) can enter the splicing pathway through its association with U1 snRNP, highlighted at the top on the right side. The unmarked ovals represent other components of the splicing machinery. The left side highlights the fact that Sfrs and Gal3 are associated with the pre-mRNA substrate from the early commitment complex through the active C complex, and the final mature mRNA product. Both Sfrs and Gal3 are found as free polypeptides as well as being associated with multiple snRNPs in larger complexes outside of the spliceosome, as schematically illustrated on the right side.

et al. 1996). However, individual SR proteins have distinct specificity and efficiency in splicing different pre-mRNA substrates, and SR proteins are known to affect splice site selection in a concentration-dependent manner (see review by Long and Caceres 2009).

Figure 3 highlights two important aspects of the association of Gal1/Gal3 and the SR proteins with the splicing machinery. First, members of both protein families are associated with the pre-mRNA substrate from the early commitment complex to the final active C complex (reviewed by Wang et al. 2006; Wahl et al. 2009), as depicted on the left side of the figure under splicing complexes. Second, the right side of the figure illustrates the fact that, in the absence of a pre-mRNA scaffold, SR proteins and Gal3 can be found either as free protein or in association with multiple ribonucleoprotein complexes, including complexes containing all five snRNPs (Peng et al. 2006; Haudek et al. 2009). This is significant because one

mechanism for the incorporation of Gal3 onto the forming spliceosome is through its association with the U1 snRNP, which recognizes the 5'-splice site of the pre-mRNA substrate (Haudek et al. 2009).

Table III also highlights three differences between galectins and the SR proteins. First, both domains of the SR protein interact directly with the pre-mRNA splicing substrate (Hertel and Graveley 2005), whereas it appears that neither Gal1 nor Gal3 binds pre-mRNA directly (Wang et al. 2006). Second, more than one SR family member can coexist on the same spliceosome (Wahl et al. 2009), while it appears that Gal1 and Gal3 assemble onto different spliceosomes (Wang et al. 2006). Finally, the stability of the association of each family with spliceosomes is very different in terms of ionic strength. Galectins can be released from spliceosomes at ~130 mM salt (Wang et al. 2006), and the SR proteins remain associated even when the salt concentration is 250 mM (Jurica et al. 2002;

Zhou et al. 2002). The significance of this disparate stability is intriguing.

The comparisons between SR proteins and galectins in Table III also pose questions that require future investigations. First, Dagher et al. (1995) had documented that, in cell-free assays, carbohydrate ligands of galectins inhibit splicing, whereas non-binding saccharides failed to yield the same effect. Do saccharides that bind the SR proteins modulate splicing in the cell-free assay (i.e. are the RNA-binding and carbohydrate-binding activities mutually exclusive for SR proteins)? Moreover, can saccharide ligands of both SR proteins and galectins serve as endogenous splicing regulators in intact cells? Second, site-directed mutants of Gal1, devoid of carbohydrate-binding activity, retained the ability to reconstitute splicing in a galectin-depleted nuclear extract, suggesting that saccharide-binding, per se, is not required for the splicing activity of galectins (Voss et al. 2008). Can the splicing and carbohydrate-binding activities of the SR proteins be similarly dissociated? Third, while SR proteins are known to be involved in splice site selection and alternative splicing, the question remains whether the splicing activity of the galectins is restricted to constitutive splicing or broader in scope to include alternative splicing.

Summary and perspectives

The studies cited in the present review document that two distinct families of proteins share the following four properties: (a) pre-mRNA splicing activity; (b) carbohydrate-binding activity; (c) nuclear localization; and (d) cell surface expression. The SR proteins SF2/ASF and SC35 were so named because they were initially discovered on the basis of their splicing activity and localized to the nucleus (Figure 2A). The surprise is the recent documentation of their carbohydrate-binding activity at the cell surface (Figure 2B). In contrast, the galectins were initially purified on the basis of their saccharide-binding activity and cell surface localization (Figure 2C). The unexpected findings are their association with the spliceosome in the nucleus (Figure 2D). Why have the carbohydrate-binding domains and cell surface localization properties of the two families been retained? Are the SR proteins lectins disguised as splicing factors, and are the galectins splicing factors masquerading as lectins? Is it just a coincidence that these two conserved families share such properties as carbohydrate-binding, cell surface localization, and splicing activity? Really, what is in a name?

Acknowledgments

The work carried out in the authors' laboratories has been supported by grants GM-38740 from the National Institutes of Health (JLW), MCB-0092919 from the National Science Foundation (RJP), and 06-IRGP-858 (JLW) and 09-CDFP-2001 (RJP) from the Michigan State University Intramural Research Program.

Abbreviations

ASF, alternative splicing factor; B16-FTIII-M cells, B16 melanoma cells transfected with fucosyltransferase 3; CRD, carbohydrate

recognition domain; Gal1, galectin-1; Gal3, galectin-3; RRM, RNA recognition motif; SC35, spliceosomal component 35; SF2, splicing factor 2; snRNP, small nuclear ribonucleoprotein complex.

References

- Agrwal N, Sun Q, Wang SY, Wang JL. 1993. Carbohydrate binding protein 35. I. Properties of the recombinant polypeptide and the individuality of the domains. *J Biol Chem.* 268:14932–14939.
- Arnoys EJ, Wang JL. 2007. Dual localization: proteins in extracellular and intracellular compartments. *Acta Histochem.* 109:89–110.
- Barondes SH, Castronovo V, Cooper DN, Cummings RD, Drickamer K, Feizi T, Gitt MA, Hirabayashi J, Hughes C, Kasai K, et al. 1994. Galectins: a family of animal β -galactoside-binding lectins. *Cell.* 76:597–598.
- Caceres JF, Misteli T, Sreaton GR, Spector DL, Krainer AR. 1997. Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol.* 138:225–238.
- Caceres JF, Sreaton GR, Krainer AR. 1998. A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev.* 12:55–66.
- Cazalla D, Zhu J, Manche L, Huber E, Krainer AR, Caceres JF. 2002. Nuclear export and retention signals in the RS domain of SR proteins. *Mol Cell Biol.* 22:6871–6882.
- Cooper DN. 2002. Galectinomics: finding themes in complexity. *Biochim Biophys Acta.* 1572:209–231.
- Cortegano I, del Pozo V, Cardaba B, Arrieta I, Gallardo S, Rojo M, Takai T, Palomino P, Liu FT, Di Mario U. 2000. Interaction between galectin-3 and the Fc γ RII induces down-regulation of IL-5 gene: implication of the promoter sequence IL-5REIII. *Glycobiology.* 10:237–242.
- Crider-Pirkle S, Billingsley R, Faust C, Hardy DM, Lee V, Weitlauf H. 2002. Cubilin, a binding partner for galectin-3 in the murine utero-placental complex. *J Biol Chem.* 277:15904–15912.
- Dagher SF, Wang JL, Patterson RJ. 1995. Identification of galectin-3 as a factor in pre-mRNA splicing. *Proc Natl Acad Sci USA.* 92:1213–1217.
- Davidson PJ, Davis MJ, Patterson RJ, Ripoche MA, Poirier F, Wang JL. 2002. Shuttling of galectin-3 between the nucleus and cytoplasm. *Glycobiology.* 12:329–337.
- Davidson PJ, Li SY, Lohse AG, Vandergaast R, Verde E, Pearson A, Patterson RJ, Wang JL, Arnoys EJ. 2006. Transport of galectin-3 between the nucleus and cytoplasm I. Conditions and signals for nuclear import. *Glycobiology.* 16:602–611.
- Dong S, Hughes RC. 1996. Galectin-3 stimulates uptake of extracellular Ca²⁺ in human Jurkat T-cells. *FEBS Lett.* 395:165–169.
- El-Boubbou K, Zhu DC, Vasileiou C, Borhan B, Prospero D, Li W, Huang X. 2010. Magnetic glyco-nanoparticles: a tool to detect, differentiate, and unlock the glyco-codes of cancer via magnetic resonance imaging. *J Am Chem Soc.* 132:4490–4499.
- Fu XD, Maniatis T. 1990. Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature.* 343:437–441.
- Fu XD, Mayeda A, Maniatis T, Krainer AR. 1992. General splicing factors SF2 and SC35 have equivalent activities *in vitro*, and both affect alternative 5' and 3' splice site selection. *Proc Natl Acad Sci USA.* 89:11224–11228.
- Fukuda MN, Ohyama C, Lowitz K, Matsuo O, Pasqualini R, Ruoslahti E, Fukuda M. 2000. A peptide mimic of E-selectin ligand inhibits sialyl Lewis X-dependent lung colonization of tumor cells. *Cancer Res.* 60:450–456.
- Ge H, Manley JL. 1990. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA *in vitro*. *Cell.* 62:25–34.
- Ge H, Zuo P, Manley JL. 1991. Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell.* 66:373–382.
- Hatakeyama S, Sugihara K, Nakayama J, Akama TO, Wong SM, Kawashima H, Zhang J, Smith DF, Ohyama C, Fukuda M, et al. 2009. Identification of mRNA splicing factors as the endothelial receptor for carbohydrate-dependent lung colonization of cancer cells. *Proc Natl Acad Sci USA.* 106:3095–3100.
- Haudek KC, Voss PG, Locascio LE, Wang JL, Patterson RJ. 2009. A mechanism for galectin-3 incorporation into the spliceosome through its association with U1 snRNP. *Biochemistry.* 48:7705–7712.
- Haudek KC, Spronk KJ, Voss PG, Patterson RJ, Wang JL, Arnoys EJ. 2010. Dynamics of galectin-3 in the nucleus and cytoplasm. *Biochim Biophys Acta.* 1800:181–189.

- Haynes C, Iakoucheva LM. 2006. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res.* 34: 305–312.
- Hertel KJ, Graveley BR. 2005. RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends Biochem Sci.* 30:115–118.
- Hikita C, Vijayakumar S, Takito J, Erdjument-Bromage H, Tempst P, Al-Awqati Q. 2000. Induction of terminal differentiation in epithelial cells requires polymerization of hensin by galectin-3. *J Cell Biol.* 151:1235–1246.
- Hirabayashi J, Kasai K. 1993. The family of metazoan metal-independent β -galactoside-binding lectins: structure, function, and molecular evolution. *Glycobiology.* 3:297–304.
- Hirabayashi J, Hashidate T, Arata Y, Nishi N, Nakamura T, Hirashima M, Urashima T, Oka T, Futai M, Muller WE, et al. 2002. Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. *Biochim Biophys Acta.* 1572:232–254.
- Huang Y, Gattoni R, Stevenin J, Steitz JA. 2003. SR splicing factors serve as adaptor proteins for TAP-dependent mRNA export. *Mol Cell.* 11: 837–843.
- Huang Y, Yario TA, Steitz JA. 2004. A molecular link between SR protein dephosphorylation and mRNA export. *Proc Natl Acad Sci USA.* 101: 9666–9670.
- Hubert M, Wang SY, Wang JL, Seve AP, Hubert J. 1995. Intranuclear distribution of galectin-3 in mouse 3 T3 fibroblasts: comparative analysis by immunofluorescence and immunoelectron microscopy. *Exp Cell Res.* 220:397–406.
- Hughes RC. 1999. Secretion of the galectin family of mammalian carbohydrate-binding proteins. *Biochim Biophys Acta.* 1473:172–185.
- Jurica MS, Licklider LJ, Gygi SR, Grigorieff N, Moore MJ. 2002. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA.* 8:426–439.
- Knibbs RN, Agrwal N, Wang JL, Goldstein IJ. 1993. Carbohydrate-binding protein 35. II. Analysis of the interaction of the recombinant polypeptide with saccharides. *J Biol Chem.* 268:14940–14947.
- Krainer A, Conway GC, Kozak D. 1990. Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev.* 4:1158–1171.
- Krainer A, Mayeda A, Kozak D, Binns G. 1991. Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70 K and *Drosophila* splicing regulators. *Cell.* 66:383–394.
- Kuwabara I, Liu FT. 1999. Galectin-3 promotes adhesion of human neutrophils to laminin. *J Immunol.* 156:3939–3944.
- Laing JG, Wang JL. 1988. Identification of carbohydrate binding protein 35 in heterogeneous nuclear ribonucleoprotein complex. *Biochemistry.* 27: 5329–5334.
- Leffler H, Barondes SH. 1986. Specificity of binding of three soluble rat lung lectins to substituted and unsubstituted mammalian β -galactosides. *J Biol Chem.* 261:10119–10126.
- Leffler H, Carlsson S, Hedlund M, Qian Y, Poirier F. 2004. Introduction to galectins. *Glycoconj J.* 19:433–440.
- Li SY, Davidson PJ, Lin NY, Patterson RJ, Wang JL, Arnoys EJ. 2006. Transport of galectin-3 between the nucleus and cytoplasm. II. Identification of the signal for nuclear export. *Glycobiology.* 16:612–622.
- Lia MC, Tarn WY. 2004. Hypophosphorylated ASF/SF2 binds TAP and is present in messenger ribonucleoproteins. *J Biol Chem.* 279:31745–31749.
- Liu FT. 2004. Double identity: galectins may not function as lectins inside the cell. *Trends Glycosci Glycotechnol.* 16:255–264.
- Liu FT, Patterson RJ, Wang JL. 2002. Intracellular functions of galectins. *Biochim Biophys Acta.* 1572:263–273.
- Liu HX, Zhang M, Krainer AR. 1998. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* 12: 1998–2012.
- Liu HX, Chew SL, Cartegni L, Zhang MQ, Krainer AR. 2000. Exonic splicing enhancer motif recognized by human SC35 under splicing conditions. *Mol Cell Biol.* 20:1063–1071.
- Long JC, Caceres JF. 2009. The SR protein family of splicing factors: master regulators of gene expression. *Biochem J.* 417:15–27.
- Misteli T, Caceres JF, Spector DL. 1992. The dynamics of a pre-mRNA splicing factor in living cells. *Nature.* 387:523–527.
- Misteli T, Caceres JF, Clement JQ, Krainer AR, Wilkinson MF, Spector DL. 1998. Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription *in vivo*. *J Cell Biol.* 143:297–307.
- Nakahara S, Hogan V, Inohara H, Raz A. 2006. Importin-mediated nuclear translocation of galectin-3. *J Biol Chem.* 281:39649–39659.
- Nickel W, Rabouille C. 2009. Mechanisms of regulated unconventional protein secretion. *Nature Rev Mol Cell Biol.* 10:148–155.
- Paz A, Haklai R, Elad-Sfadia G, Ballan E, Kloog Y. 2001. Galectin-1 binds to oncogenic H-Ras to mediate Ras membrane anchorage and cell transformation. *Oncogene.* 20:7486–7493.
- Peng R, Hawkins I, Link AJ, Patton JG. 2006. The splicing factor PSF is part of a large complex that assembles in the absence of pre-mRNA and contains all 5 snRNPs. *RNA Biol.* 3:69–76.
- Rabinovich GA, Toscano MA, Jackson SS, Vasta GR. 2007. Functions of cell surface galectin-glycoprotein lattices. *Curr Opin Struct Biol.* 17:513–520.
- Rajotte D, Ruoslahti E. 1999. Membrane dipeptidase is the receptor for a lung-targeting peptide identified by *in vivo* phage display. *J Biol Chem.* 274:11593–11598.
- Rini JM, Lobsanov YD. 1999. New animal lectin structures. *Curr Opin Struct Biol.* 9:578–584.
- Sanford JR, Gray NK, Beckmann K, Caceres JF. 2004. A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.* 18:755–768.
- Sanford JR, Ellis JD, Cazalla D, Caceres JF. 2005. Reversible phosphorylation differentially affects nuclear and cytoplasmic functions of splicing factor2/ alternative splicing factor. *Proc Natl Acad Sci USA.* 102:15042–15047.
- Sanford JR, Coutinho P, Hackett JA, Wang X, Ranahan W, Caceres JF. 2008. Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PlosOne.* 3:e3369.
- Sapra AK, Anko ML, Grishina I. 2009. SR protein family members display diverse activities in the formation of nascent and mature mRNPs *in vivo*. *Mol Cell.* 34:179–190.
- Seetharaman J, Kanigsberg A, Slaaby R, Leffler H, Barondes SH, Rini JM. 1998. X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-angstrom resolution. *J Biol Chem.* 273:13047–13052.
- Shepard PJ, Hertel KJ. 2009. The SR protein family. *Genome Biol.* 10: 242.1–242.9.
- Sinha R, Allemand E, Zhang Z, Karni R, Myers MP, Krainer AR. 2010. Arginine methylation controls the subcellular localization and functions of the oncoprotein splicing factor SF2/ASF. *Mol Cell Biol.* 30:2762–2774.
- Stowell SR, Arthur CM, Padmaia M, Mehta P, Slanina KA, Blixt O, Leffler H, Smith DF, Cummings R. 2008. Galectin-1, -2, and -3 exhibit differential recognition of sialylated glycans and blood group antigens. *J Biol Chem.* 283:10109–10123.
- Takenaka Y, Fukumori T, Yoshii T, Oka N, Inohara H, Kim HR, Bresalier RS, Raz A. 2004. Nuclear export of phosphorylated galectin-3 regulates its anti-apoptotic activity in response to chemotherapeutic drugs. *Mol Cell Biol.* 24:4395–4406.
- Tsay YG, Lin NY, Voss PG, Patterson RJ, Wang JL. 1999. Export of galectin-3 from nuclei of digitonin-permeabilized mouse 3 T3 fibroblasts. *Exp Cell Res.* 252:250–261.
- van Kasteren SI, Campbell SJ, Serres S, Anthony DC, Sibson NR, Davis BG. 2009. Glycanonparticles allow pre-symptomatic *in vivo* imaging of brain disease. *Proc Natl Acad Sci USA.* 106:18–23.
- Voss PG, Gray RM, Dickey SW, Wang W, Park JW, Kasai K, Hirabayashi J, Patterson RJ, Wang JL. 2008. Dissociation of the carbohydrate-binding and splicing activities of galectin-1. *Arch Biochem Biophys.* 478:18–25.
- Vyakarnam A, Dagher SF, Wang JL, Patterson RJ. 1997. Evidence for a role for galectin-1 in pre-mRNA splicing. *Mol Cell Biol.* 17:4730–4737.
- Vyakarnam A, Lenneman AJ, Lakkides KM, Patterson RJ, Wang JL. 1998. A comparative nuclear localization study of galectin-1 with other splicing components. *Exp Cell Res.* 242:419–428.
- Wahl MC, Will CL, Luhrmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell.* 136:701–718.
- Wang J, Takagaki Y, Manley JL. 1996. Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes Dev.* 10: 2588–2599.
- Wang W, Park JW, Wang JL, Patterson RJ. 2006. Immunoprecipitation of spliceosomal RNAs by antisera to galectin-1 and galectin-3. *Nucleic Acids Res.* 34:5166–5174.
- Yang RY, Hsu DK, Liu FT. 1996. Expression of galectin-3 modulates T-cell growth and apoptosis. *Proc Natl Acad Sci USA.* 93:6737–6742.
- Zahler AM, Lane WS, Stolk JA, Roth MB. 1992. SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* 6:837–847.
- Zamora PD, Patton JG, Green MR. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature.* 355:609–614.
- Zelensky AN, Gready JE. 2005. The C-type lectin-like domain superfamily. *FEBS J.* 272:6179–6217.
- Zhang J, Nakayama J, Ohyama C, Suzuki M, Suzuki A, Fukuda M, Fukuda MN. 2002. Sialyl Lewis X-dependent lung colonization of B16 melanoma cells through a selectin-like endothelial receptor distinct from E- or P-selectin. *Cancer Res.* 62:4194–4198.

- Zhao R, Bodnar MS, Spector DL. 2009. Nuclear neighborhoods and gene expression. *Curr Opin Genet Dev.* 19:172–179.
- Zhong XY, Wang P, Han J, Rosenfeld MG, Fu XD. 2009. SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol Cell.* 35:1–10.
- Zhou Z, Licklider LJ, Gygi SP, Reed R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature.* 419:182–185.
- Zuberi RI, Frigeri LG, Liu FT. 1994. Activation of rat basophilic leukemia cells by ϵ BP, an IgE-binding endogenous lectin. *Cell Immunol.* 156:1–12.