Advances in the Study of SR Protein Family

Xiaoyun Ma and Fuchu He*
Beijing Institute of Radiation Medicine, Beijing 100850, China

The name of SR proteins is derived from their typical RS domain that is rich in serine (Ser, S) and arginine (Arg, R). They are conserved in evolution. Up to now, 10 members of the SR protein family have been identified in humans. SR proteins contain one or two RNA binding motifs aside from the RS domain, and also possess special biochemical and immunological features. As to the functions of SR proteins, they facilitate the recruitment of the components of splicesome via protein-protein interaction to prompt the assembly of early splicesome; while in alternative splicing, tissue-specifically expressed SR protein along with the relative ratio of SR protein and heterogeneous nuclear ribonucleoprotein (hnRNP) is composed of two main regulative mechanisms for alternative splicing. Almost all of the biochemical functions are regulated by reversible phosphorylation.

Key words: SR proteins, splicesome, hnRNP, phosphorylation

Introduction

The splicing of pre-mRNA takes place in the splicesome. Except for the components of catalytic core, splicesome is mainly composed of small nuclear ribonucleoprotein (snRNP) and non-snRNP factors. Among the latter, SR proteins are studied more widely and thoroughly. SR proteins are named by their typical RS domain, which is composed of a certain number of tandemly repeated serine-arginine (Ser-Arg, SR) dipeptide. The meaning of SR and RS is identical, and for some historical reasons, the difference between the names of the proteins and their signature domain was formed. As to the function of SR proteins, they are involved in almost all steps of pre-mRNA splicing. They are necessary for constitutive splicing, and their function is also very important in the regulation of alternative splicing. Since the 1990's, the study on the function of SR proteins has become a very active field. This review focuses on the basic structural features, classification and identification, and functions in pre-mRNA and phosphorylation of the SR protein family.

Discovery and Nomenclature

The first two SR proteins were discovered in the biochemical studies on mammalian splicing. In 1985, Krainer and Maniatis identified a splicing factor required for splicing of a β-globin pre-mRNA *in vitro*, SF2 (splicing factor 2), and subsequently confirmed it as an essential splicing factor. In 1990, Ge and Manly discovered a splicing factor that could influence selection of $5'$ alternative splicing sites in an SV40 early pre-mRNA, and named it the alternative splicing factor (ASF). Then, ASF and SF2 were confirmed as an identical protein and called ASF/SF2 or SF2/ ASF (also named SRp30a). That is the prototype of SR protein. SC35 (SRp30b/PR264) is the second identified SR protein. They were shown to be members of a larger, evolutionarily conserved family of proteins. Up to now, 10 SR proteins have been identified in human. In addition to the two proteins discussed above, SRp20 (X16 in mouse, RBP1 in *Drosophila*), SRp40 (HRS in mouse), SRp55 (B52 in *Drosophila*), h9G8, SRp75, SRp30c, SRp46, and p54 were identified later. SR proteins or their homologues were also discovered in other species. Moreover, many SR-related or -like proteins, which are involved in multiple functions such as splicing, transcription, processing and transporting of mRNA, along with SR proteins make up the SR protein superfamily.

The nomenclature of SR proteins is based on their molecular weight and capped with "SR" according to the rule of Zahler. For example, ASF/SF2 was termed as SRp30a (Since there are more than one SR protein with molecular weight is 30 kDa, only the letter "a" is used in the name of ASF/SF2), likewise SC35 is called SRp30b.

* **Corresponding author.**
E-mail: hefc@nic.bmi.ac.cn

The Structural Feature, Subcellular Distribution and Identification of the SR Proteins

The structural feature and subcellular distribution

To date, the molecular weights of all identified SR proteins are between 20 kDa and 75 kDa. They contain one or two RNA-recognition motifs (RRM) (or RNP-type RNA-binding domain, RBD) in the N-terminus. The number of RBD in SR proteins is different. Some of them possess one, such as SC35 and SRp20, and others possess two, such as ASF/SF2, SRp40, SRp55 and SRp75 (Fig. 1). The domain has been discovered in many splicing factors as well as in some RNA-binding proteins that do not participate in splicing. RS-domain that is rich in serine and arginine is located at the C-terminus of SR proteins. Actually, RS-domain has been discovered in many proteins, a large number of which belong to splicing factors or splicing regulators. In addition, some SR proteins contain special structures such as zinc knuckle (a kind of conserved zinc finger) and RGG-box (rich in Gly) in human h9G8 and in *Arabidopsis thaliana* atRSZp22 (*1*, *2*).

SR proteins aggregate at a number of dots in nucleus of mammalian cells, making a typical nuclear structure and thus named nuclear speckles. In general, there are 20-50 speckles per cell (Fig. 2). Every speckle represents a storage and recycling site for SR proteins and other splicing factors. Such structures are dynamically distributed, depending on the transcriptional activity of cells. When the synthesis of mRNA is inhibited, the dots will become more obvious. Moreover, some SR proteins can shuttle between cytoplasm and nucleus from time to time.

Identification

Strictly speaking, SR proteins are not only conserved in their structure, but also endowed with some special biochemical and immunological features. Additionally they belong to essential splicing factors in function. First, in structure, they have RRM in N-terminus and RS domain in C-terminus. The primary structures of almost all SR proteins include one "GFAFVEFEDPRDAEDA" sequence (16-amino acid in the conserved RNP-type RNA binding domain) at the first RNA binding domain and another conserved sequence, RLIVENLSSRVSWQDLKD, in the second RNA binding domain, among which the last seven amino acids are absolutely conserved. Because of their special structure—serine and arginine distributed alternatively—SR proteins can be precipitated from nuclear extract of cells in 20 mM $MgCl_2$, and then be dissolved in a buffer containing 2 M NaCl and 20 mM EDTA. Moreover, all SR proteins can be recognized by a monoclonal antibody-MAb104 (recognizing phosphorylated serine within the RS domain). At last, it must be confirmed that they have activities of essential and alternative splicing factors by using both *in vitro* and *in vivo* assay systems. Only the protein that contains all of the features discussed above is named SR protein, also the theme discussed in this review.

The Functions of SR Proteins

RNA-binding property

Binding to specific RNA sequences is an important and universal function of SR proteins, in which RRM (zinc knuckle or RGG-box in some proteins) plays a significant role in specific binding to target sequences. The high-affinity binding sequences of SR proteins can be isolated by systematic evolution of ligands by exponential enrichment (SELEX). The isolated RNA sequences are often splicing enhancers located in the exon, and thus are named exon splicing enhancers (ESE), which are often the recognition and pre-mRNA binding targets of SR proteins, and are also the mutation targets in some human diseases. ESE of many SR proteins has been identified, such as in SF2/ASF and SC35.

The sequences recognized by different SR proteins are also distinct. The first evidence of interaction between SR proteins and specific RNA sequences is derived from a recombinant derivative of ASF/SF2 lacking the RS domain (ASFΔRS). In theory, deletion of this domain would become very basic and might enhance nonspecific bindings with RNA. However, ASFΔRS is still able to recognize two specific RNA fragments containing intact 5′ splicing sites, while mutations that disrupt the 5′ splicing site in each case can reduce or eliminate such binding. ASF/SF2 can bind a purine-rich sequence in the bovine growth hormone gene and also enhance splicing when added to nuclear extract. In contrast, SC35 can neither bind the element nor activate splicing, providing evidence that different SR proteins possess distinct RNA binding

specificities. Now, as the evidences are accumulates, it has been determined conclusively that SR proteins bind conserved RNA sequences specifically, and the binding is especially important in alternative splicing. Moreover, it has been demonstrated that ESEs could stimulate both constitutive and alternative splicing.

The functions of RS-domain

RS-domain was first discovered in splicing regulator transformer (Tra) and transformer-2 (Tra2) of *Drosophila*. Its function involves protein-protein interaction, sub-nuclear localization and regulation of RNA binding.

Protein-protein interaction within RS-domain is very important in early assembly of splicesome. Specifically, ASF/SF2 and SC35, ASF/SF2 (or other SR proteins) and other splicing factors (such as U1snRNP specific protein 70K and essential splicing factor U2AF35) are all found to interact with each other. In such a process, RS-domain or similar domain is necessary (*4, 5*). The direct interaction between SR proteins and components of splicing machinery suggest that there is a significant function of SR proteins in splicing. In fact, it is the interaction that enhanced recognition specificity of splicesome to splicing sites and facilitates recruitment of splicing machinery components, resulting eventually in promoting the assembly of early splicesome.

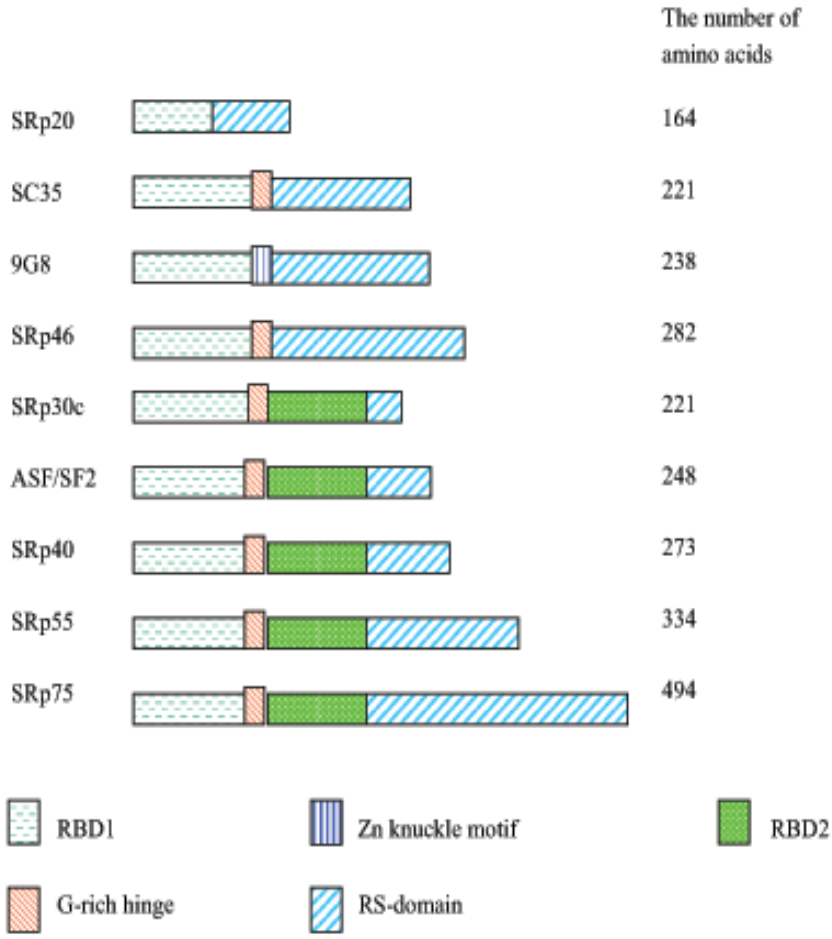

Fig. 1. Schematic representation of human SR proteins. Every member contains one or two RBD domain and one RS domain. (Note: This figure does not contain p54, which owned all the structural and biochemical features of SR protein, because its homology with other members is very low.)

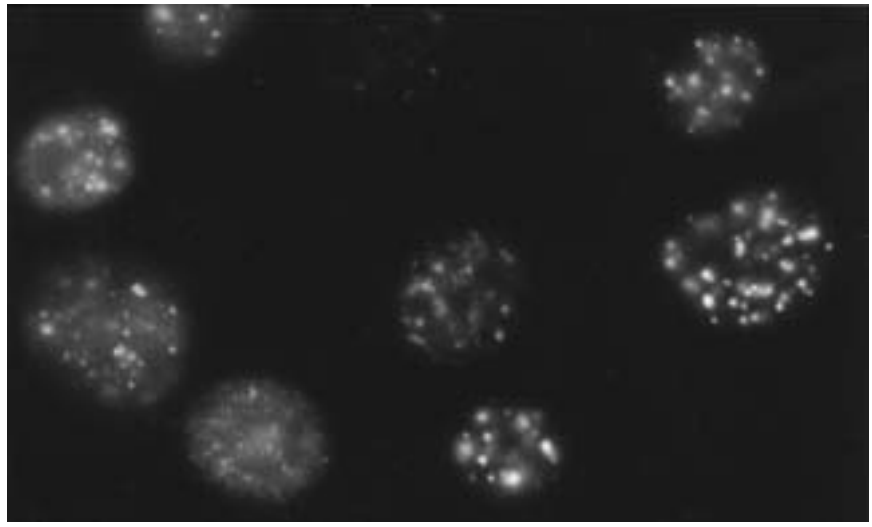

Fig. 2. The subcellular localization of SR proteins-nuclear speckles (the localization of SC-35 in CaSKi cells, (*3*)).

RS-domain also plays a role in subcellular targeting, serving as both a nuclear localization signal and a sub-nuclear localization signal, but the localization capacity of every RS-domain is quite different. Recently, it was demonstrated that two SR protein-specific nuclear import proteins—transportin-SR1 (TRN-SR1) and transportin-SR2 (TRN-SR2)—localized SR proteins to nuclear speckles via the interaction between their RS or RS-like domain and the RS domain of SR protein (*6, 7*). Therefore, the nature of the subnuclear localization function of the RS-domain still depends on protein-protein interaction in the domain.

Despite the fact that SR proteins are gathered in the nucleus, a portion of them shuttle between cytoplasm and nucleus. The shuttling SR proteins might escort the processed mRNA to its destination in the cytoplasm, in which the RS-domain plays an important role. Some RS-domains could also influence RNA binding or promote RNA-RNA annealing.

The functions of SR proteins

Simply put in one word, the functions of SR proteins are to promot early splicesome assembly in constitutive splicing and to regulate alternative splicing via their RRM binding to pre-mRNA and protein-protein interactions in their RS-domain.

Their function in splicesome assembly

Five snRNPs and multiple protein factors interact in turn with pre-mRNA to form a larger splicesome, in which SR proteins participate in nearly every step of the assembly. First, in the recognition of 5′-splicing site, SR proteins could promote formation of E complex containing U1snRNP and the binding of this complex to the 5′-splicing site, in which SR protein might function as a bridge between U1 snRNP and the 5′-splicing site to facilitate base pairing between U1 snRNA 5′ terminus and the splicing site. Next, SR proteins promote binding between U2 snRNPs with their related auxiliary factors and the branch point region of pre-mRNA to form complex A. Then via their binding to the 35-kDa subunit of U2AF (U2 snRNP auxiliary factor, binding at 3′-splicing site) and U1 70K (binding at 5′-splicing site), SR proteins such as SC35 form a bridge complex across two splicing sites. They facilitate recruitment of U4/U6 and U5 tri snRNP complex to form complex B. Ultimately the mature catalytically active complex C was formed (*8*). As discussed above, SR proteins act as a rope that links a series of splicing factors during assembly of early splicesome. The commonly accepted view is that SR proteins do not take part in specific selection of splicing sites, but just increase recognition specificity of other factors such as U1snRNP to splicing sites.

Their function in alternative splicing

Alternative splicing is an important mechanism in regulating the gene expression of cells and viruses. At certain concentrations of SR protein, U1snRNP can only bind to very strong splicing sites; therefore, weaker splicing sites might not be selected. However, under special conditions such as embryonal development, change of physiological state, virus infection,

weaker splicing sites might also be used (that is, alternative splicing), in which the function of SR protein is necessary.

At present, it is known that SR proteins regulate alternative splicing via two mechanisms. First, the expression of SR proteins is tissue-specific (*9*, *10*), and thus usage of an exon could be determined by whether or not one SR protein is expressed in the tissue. When SR protein binds to its specific sequence (commonly, ESEs members), it assists in recruiting components of splicing machinery, and then activates splicing of weaker intron via protein-protein interaction, which involves the interactions between SR superfamily members, SR protein and non-SR protein, and SR protein and putative nuclear matrix. It was recently demonstrated that the binding of SR protein to its target sequences and the comparative location of binding to splicing sites determines the possibility of using the splicing sites in pre-mRNA processing(*11*, *12*). Another main mechanism is that the relative level and activity of SR protein and hnRNP A/B that also regulate alternative splicing. Increased SR protein activates proximal 5′-splicing sites, while overexpressed RNP A1 activates distal 5′-splicing sites. In recent years, exonic splicing silencer (ESS) that inhibits usage of splicing sites has been discovered. The binding of certain hnRNP proteins to these sequences together with SR proteins binding to ESE sequences determines the selection of the splicing site. A natural stimulus that influences the ratio of these proteins in the nucleus is genotoxic stress, which induces hnRNP A1 phosphorylation by the p38-MAP kinase pathway and leads to accumulation of hnRNP A1 in the cytoplasm (*13*). Then the changed relative level of hnRNP A1 and SR protein in nucleus alters splice-site selection, thereby linking alternative splicing regulation to a specific signal transduction pathway. Moreover, under certain conditions, SR proteins antagonize each other in splice-site selection, for example, ASF/SF2 and SC35, ASF/SF2 and SRp2.

Other functions

Some SR proteins still bind with mRNA products, and a portion of them shuttle to the nucleus from the cytoplasm after splicing, implying that such proteins might have other functions, such as mRNA transport, or even translation of RNA in cytoplasm and localization of RNA, and so on. Indeed, ASF/SF2 could regulate the stability of mRNA and increase the half-life of certain mRNA.

Their redundant or specific function

The problem of whether the function of SR protein is redundant or non-redundant has always been debated in related fields, and moreover, the evidence supporting each view is probably increasing. Any single SR protein can activate splicing of several different pre-mRNAs, suggesting that there is little or no substrate specificity and that SR protein function is redundant. However, there is also some evidence demonstrating that SR protein function is non-redundant. Deletion of B52, the homologue of SRp55 gene in *Drosophila* leads to lethality during development, implicating that B52 is essential to normal development. The fact that SF2/ASF is necessary to viability of cells and that deletion of the gene could not be rescued by other SR proteins further explaine its unique function. Dasa Longman *et al.* (*14*) identified orthologues of all human SR proteins in *Caenorhabditis elegans*, and demonstrated that CeSF2/ASF is an essential protein during *C. elegans* early development by using dsRNAi technique. RNAi with other SR genes results in no obvious phenotype, indicating gene redundancy. Simultaneous interference of two or more SR proteins in certain combinations causes lethality or other developmental defects. As discussed above, SR protein function may be not only redundant but also unique in some conditions. In this way different SR protein may have distinct behavior.

Phosphorylation of SR Proteins

If in fact SR proteins play an essential role in every splicing step, phosphorylation modification will also be necessary for exerting SR protein function. Phosphorylation regulates the efficiency and specificity of interaction between SR proteins, subnuclear localization of SR proteins, and eventually influences splicing activity and specificity. All SR proteins are phosphorylated proteins, with most of such modifications taking place in the Ser residue of RS-domain. For example, the RS-domain of ASF/SF2 contains multiple phosphorylated sites. Reversible phosphorylation can regulate the interaction of protein—RNA and protein-protein, localization pattern and recruitment to transcriptional active sites of SR proteins. Such post-translation modification is very important and involved in the multiple physiological process such as early development of *Ascaris lumbricoides*, sex determination of *Drosophila*, infection of adenovirus, and so forth.

The first purified and sequenced SR protein kinase is SRPK1 (SR protein kinase 1), which is capable of phosphorylating Ser-residue within the RS-domain of several SR proteins. Moreover, the activity of such enzyme is highest in the M-phase of the whole cell cycle, suggesting the possibility that such phosphorylation is cell cycle-regulated. To date, many SR protein kinases have been identified, such as Clk/Sty1, 2, 3, 4, SRPK2, topoisomerase I, DSK1 of yeast, *etc.*

At the biochemical level, phosphorylation of RS-domain-containing splicing factors is important for initiating splicesome assembly and enhancing sequence-specific RNA binding. *In vitro*, phosphorylation of RS-domain is able to promote interactions between the proteins containing RS-domain and to inhibit the non-specific binding between proteins and RNA. On the other hand, de-phosphorylation is essential for dissociation of splicesome and release of splicing products (*15*). Therefore, the phosphorylation-dephosphorylation recycling of RS-domain likely accompanies with and regulates the splicing recycle. In addition, protein-protein interactions between RS-domain-containing splicing factors as well as their overall activity in pre-mRNA splicing can be regulated through reversible phosphorylation. In the living nucleus, SR proteins are released from speckles to nucleoplasm in a phosphorylation-dependent manner (*16*). Phosphorylation can both release SR protein from speckles and increase their affinity for pre-mRNA and for the splicing machinery. Similarly, dephosphorylation was reported to facilitate splicesome disassembly and to recycle SR protein back to speckles. During *A. lumbricodes* embryogenesis, the translocation of certain SR proteins to nucleus possibly requires partial dephosphorylation. In fact, SR protein kinase can induce redistribution of SR protein to the cytoplasm (*17, 18*). Recently two SR protein specific transporters, TRN-SR and TRN-SR2, have been identified. Both can interact with RS-domain of SF2/ASF, and such interaction requires phosphorylation of RS-domain. Moreover, the nucleus-cytoplasm shuttling activity of SR proteins also depends on their phosphorylated state. Thus, phosphorylation is the key step to nuclear import of SR proteins.

Summary and Prospects

SR proteins, conserved during evolution in metazoan, have classic RNA binding motifs, RS-domain and special biochemical and immunological features. The functions of the proteins are regulation of pre-mRNA splicing. Such regulation is really omnibearing and mutiple-orientational: from specific recognition of 5′-splicing sites, assembly of the early splicing complex, regulation of alternative splicing via tissue specific expression of SR protein and the relative ratio of SR protein and hnRNP, to regulation of stability of mRNA even escorting the processed mRNA to cytoplasm by nucleus-cytoplasm shuttling. At the same time, all the activities of SR proteins are controlled by phosphorylation-dephosphorylation of serine in RS-domain.

In recent years, the biological functions of SR proteins—regulation of splicing—have been studied. SR protein SC35 plays an important role in T cells' specific splicing regulation pathway. CD45, a kind of tyrosine phosphatase, is a transmembrane glucoprotein expressed on the surface of many leukemia cells. The gene has 8 splice variants, and antagonistic specific SR proteins regulate its splicing. After T-cells are activated, expressing pattern of SR proteins changes specifically and thus leads to the alteration of splicing pattern of CD45. It has also been demonstrated that deletion of SC35 in thymus could lead to developmental inhibition of T cells and the alteration of the splicing pattern of CD45 (*19*, *20*). It was discovered that hypophosphorylated SR protein and polyA-containing mRNA were largely accumulated in the nucleus after NB-506, an anti-tumor drug, was added to mouse leukemia cells P388. Further analysis showed that the drug inhibited phosphorylation of ASF/SF2 completely and, concomitantly assembly of splicesome (*21*). Therefore, it is easy to understand that the clarification biological function of SR proteins may influence the progress of biomedicine and possibly even clinical medicine.

The research on the functions and regulative mechanism of SR proteins is progressing rapidly, but some questions still remain to be elucidated. For example, despite the fact that regulative sequences of SR proteins have been identified in many genes, the biological function of such regulation is unknown. Moreover, all biochemical activities of SR protein are regulated by phosphorylation, but by which signal such phosphorylation is regulated? What is the relation between phosphorylation and the cell cycle? How do the integrate the redundant and unique functions of SR protein? *etc.* Except for SR protein family, most RS-domain-containing proteins are splicing factors or splicing regulation factors. The splicing takes place on the nuclear matrix. Indeed, two nuclear ma-

trix proteins containing RS-like domain, SRm160 and SRm300, have been identified as co-activators of pre-mRNA splicing. Thus, the study of SR protein superfamily functions may facilitate understanding of the functions of SR protein and mechanisms of pre-mRNA splicing as a whole.

Acknowledgements

We thank Professor Yuanmin Li and Dr. Yongxin Zhou for revising this manuscript.

References

1. Cavaloc, Y., *et al.* 1994. Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family. *EMBO J.* 13: 2639-2649.
2. Lopato, S., *et al.* 1999. A novel family of plant splicing factors with a Zn knuckle motif:examination of RNA binding and splicing activities. *Plant Mol. Biol.* 39: 761-773.
3. Zeng, C., *et al.* 1997. Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity. *EMBO J.* 16: 1401-1412.
4. Wu, J.Y. and Maniatis, T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* 75: 1061-1070.
5. Kohtz, J. *et al.* 1994. Protein-protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. *Nature* 368: 119-124.
6. Lai, M.C., *et al.* 2000. Human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *J. Biol. Chem.* 275: 7950-7957.
7. Kataoka, N., *et al.* 1999. Transportin-SR, a nuclear import receptor for SR proteins. *J. Cell Biol.* 145: 1145-1152.
8. Stojdl, D.F. and Bell, J.C. 1999. SR protein kinases: the splice of life. *Biochem. Cell Biol.* 77: 293-298. Review.
9. Screaton, G.R., *et al.* 1995. Identification and characterization of three members of the human SR family of pre-mRNA splicing factors. *EMBO J.* 14: 4336-4349.
10. Zahler, A.M., *et al.* 1993. Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science* 260: 219-222.
11. Graveley, B.R., *et al.* 1998. A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO J.* 17: 6747-6756.
12. Schaal, T.D. and Maniatis, T. 1999. Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol. Cell Biol.* 19: 1705-1719.
13. van der Houven van Oordt, W., *et al.* 2000. The MKK (3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. *J. Cell Biol.* 149: 307-316.
14. Longman, D., *et al.* 2000. Functional characterization of SR and SR-related genes in *Caenorhabditis elegans*. *EMBO J.* 19: 1625-1637.
15. Cao, W., *et al.* 1997. Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing *in vitro*. *RNA* 3: 1456-1467.
16. Sanford, J.R. and Bruzik, J.P. 1999. SR proteins are required for nematode trans-splicing *in vitro*. *RNA* 5: 918-928.
17. Sanford, J.R. and Bruzik, J.P. 2001. Regulation of SR protein localization during development. *Proc. Natl. Acad. Sci. USA* 98: 10184-10189.
18. Koizumi, J., *et al.* 1999. The subcellular localization of SF2/ASF is regulated by direct interaction with SR protein kinases (SRPKs). *J. Biol. Chem.* 274: 11125-11131.
19. ten Dam, G.B., *et al.* 2000. Regulation of alternative splicing of CD45 by antagonistic effects of SR protein splicing factors. *J. Immunol.* 164: 5287-5295.
20. Wang, H.Y., *et al.* 2001. SC35 plays a role in T cell development and alternative splicing of CD45. *Mol. Cell* 7: 331-342
21. Pilch, B., *et al.* 2001. Specific inhibition of serine- and arginine-rich splicing factors phosphorylation, spliceosome assembly, and splicing by the antitumor drug NB-506. *Cancer Res.* 61: 6876-6884.

This work was supported by the National Natural Science Foundation of China (30070285).

Received: 21 November, 2002
Accepted: 5 December, 2002