The EMBO Journal Vol.18 No.16 pp.4560–4570, 1999

Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing

Akila Mayeda[1], Joseph Badolato[2,3], Ryuji Kobayashi[4], Michael Q.Zhang[4], Edith M.Gardiner[2] and Adrian R.Krainer[4]

Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, Miami, FL 33136, [4]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA and [2]Bone and Mineral Research Program, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia

[3]Present address: Eli Lilly Australia, Pty. Ltd, West Ryde, NSW 2114, Australia

[1]Corresponding author
e-mail: amayeda@molbio.med.miami.edu

Biochemical purification of a pre-mRNA splicing activity from HeLa cells that stimulates distal alternative 3′ splice sites in a concentration-dependent manner resulted in the identification of RNPS1, a novel general activator of pre-mRNA splicing. RNPS1 cDNAs, encoding a putative nucleic-acid-binding protein of unknown function, were previously identified in mouse and human. RNPS1 is conserved in metazoans and has an RNA-recognition motif preceded by an extensive serine-rich domain. Recombinant human RNPS1 expressed in baculovirus functionally synergizes with SR proteins and strongly activates splicing of both constitutively and alternatively spliced pre-mRNAs. We conclude that RNPS1 is not only a potential regulator of alternative splicing but may also play a more fundamental role as a general activator of pre-mRNA splicing.

Keywords: alternative splicing/RNA-recognition motif/RNPS1/splicing activator/splicing factor

Introduction

Most mRNA molecules in higher eukaryotes must undergo splicing in the nucleus before they can carry the correct genetic information to the cytoplasm as templates for translation. Although a given constitutively spliced pre-mRNA contains many potential splice sites, only the correct splice site pairs are selected, generating identical mature mRNAs. The splicing process is therefore highly discriminating and precise, and it is sensitive to mutations in critical sequence elements of the pre-mRNA (reviewed in Reed and Palanjian, 1997; Burge *et al*., 1999). Splice site mutations often cause changes in the splicing pattern and are therefore responsible for a number of hereditary diseases (reviewed in Antoniou, 1995; Cooper and Mattox, 1997). On the other hand, there is flexibility in the splicing process for many pre-mRNAs, allowing alternative splice sites to be used, often in a regulated way in response to tissue-specific, physiologically or developmentally controlled states (reviewed in Wang *et al*., 1997; Lopez, 1998). This process, alternative splicing, is a common strategy for expanding the coding potential of individual genes and for the regulation of gene expression in higher eukaryotes.

Since the discovery of introns, significant progress has been made in understanding pre-mRNA splicing through the isolation and characterization of essential and regulatory splicing factors, which include small ribonucleoprotein particles (snRNPs), heterogeneous ribonucleoproteins (hnRNPs) and several non-snRNP protein factors (reviewed in Cáceres and Krainer, 1997; Reed and Palanjian, 1997; Will and Lührmann, 1997; Burge *et al*., 1999). Biochemical studies of the splicing complex, or spliceosome, together with extensive genetic studies in yeast, indicate that >50 proteins are essential for constitutive splicing (reviewed in Reed and Palanjian, 1997; Burge *et al*., 1999). In mammalian systems, many general and regulatory splicing proteins remain to be discovered and characterized.

Efforts to uncover the mechanisms of splice-site selection in either constitutively or alternatively spliced pre-mRNAs by identifying factors with antagonistic effects on splice-site selection have been reported (reviewed in Cáceres and Krainer, 1997). Several, but not all, SR proteins tend to stimulate the use of proximal splice sites when multiple alternative 5′ or 3′ splice sites are available (Ge and Manley, 1990; Krainer *et al*., 1990b; Mayeda *et al*., 1993; Cáceres *et al*., 1994; Bai *et al*., 1999). Proteins of the hnRNP A/B family can counteract the 5′ splice site modulatory activity of SR proteins, thereby stimulating distal alternative 5′ splice sites with varied efficiencies (Mayeda and Krainer, 1992; Cáceres *et al*., 1994; Mayeda *et al*., 1994; Yang *et al*., 1994).

We have found additional activities, distinct from either SR or hnRNP A/B proteins, which promote the selection of distal alternative 3′ splice sites *in vitro*. In the course of efforts to purify and identify the factors responsible for these activities, we discovered a novel pre-mRNA splicing factor that functions as a potent activator of both constitutively and alternatively spliced pre-mRNAs. We report the purification and identification of this protein, and demonstrate its unique functional properties in pre-mRNA splicing.

Results

Purification and identification of human RNPS1

hnRNP A/B proteins affect alternative 5′ splice-site selection, but they have no significant effect on alternative 3′ splice-site selection *in vitro* (Figure 1, lanes 5–12). However, using fractions derived from HeLa cell nuclear extract we found an activity, termed SF7, that switches the use of alternative 3′ splice sites from proximal to distal in a concentration-dependent manner (Figure 1,

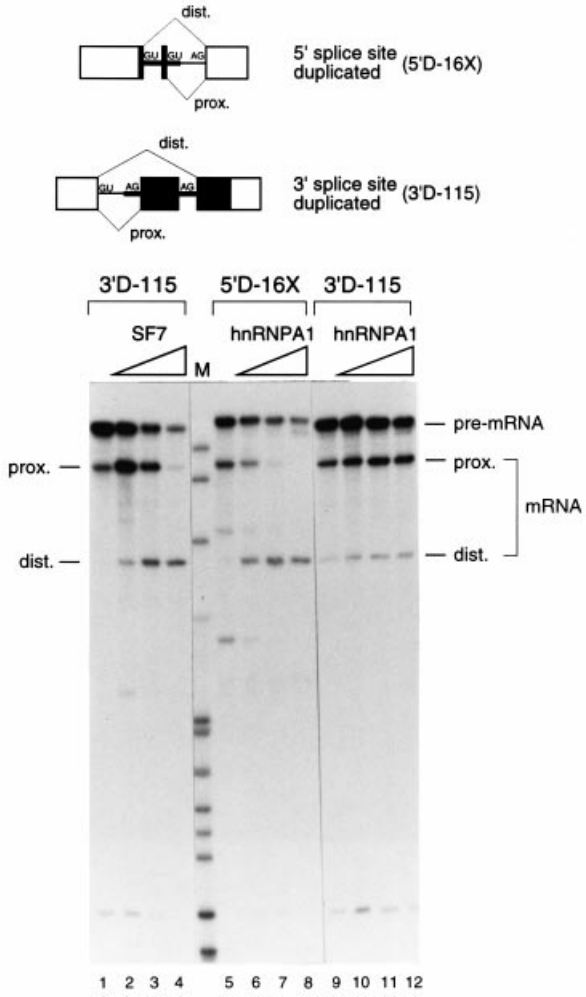

Fig. 1. Alternative 3′ splice-site switching activity (SF7). *In vitro* splicing with β-globin 3′- and 5′-splice site duplicated pre-mRNAs (3′D-115 and 5′D-16X). Splicing reactions included constant amounts of HeLa cell S100 extract (7 μl) plus purified SF2/ASF (5 pmol), and variable amounts of a partially purified SF7 fraction (0, 1, 2 and 4 μl of fraction III in lanes 1–4; see Figure 2A) or recombinant hnRNP A1 (0, 5, 10 and 20 pmol in lanes 5–8 and 9–12). The structure of the model pre-mRNAs with alternative splicing paths is shown schematically above the panel. The black shading indicates the duplicated intron/exon segments. The pre-mRNA and spliced mRNA products generated via use of proximal (prox.) and distal (dist.) alternative splice sites are indicated. The detailed structures of these substrates and the characterization of splicing products have been described (Reed and Maniatis, 1986; Krainer *et al*., 1990b). M: ^{32}P-labeled pBR322/*Hpa*II DNA markers.

lanes 1–4). To purify SF7, test fractions were assayed by adding them to *in vitro* splicing reactions consisting of HeLa cell S100 extract complemented with the SR protein SF2/ASF and measuring the stimulation of distal 3′ splice-site use with one of several model pre-mRNAs with duplicated 3′ splice sites. SF7 activity was purified by ammonium sulfate fractionation followed by density gradient centrifugation, which effectively separated it from SR proteins and snRNPs (Figure 2A, fraction III; data not shown). At this stage of purification, strong SF7 activity was recovered (Figure 1, lanes 1–4). Further purification on heparin–Sepharose and Mono Q columns resulted in the separation of SF7 activity into three independently active fractions, termed SF7A, SF7B and SF7C. Unexpectedly, the SF7A and SF7B fractions activated not only the distal 3′ splice site but also the proximal 3′ splice site, whereas SF7C activated only the distal 3′ splice site (data not shown).

SDS–PAGE analysis showed only five major polypeptides in the SF7A fraction, whereas the SF7B and SF7C fractions were still relatively crude (Figure 2B, fractions V-A and V-B; data not shown). We therefore focused on the SF7A fraction and obtained partial amino acid sequences for all the polypeptides detected in this fraction. The peptide sequences were then compared with known proteins or open reading frames (ORF) in the sequence databases, and identical matches were obtained in all cases (Figure 2A and B). The doublet that migrates at 38–40 kDa represents two forms of nucleophosmin (DDBJ/EMBL/GenBank accession No. M23613); the ~54 kDa polypeptide is calreticulin (DDBJ/EMBL/GenBank accession No. M84739); and the ~88 kDa polypeptide is the 80K-H protein kinase-C substrate (DDBJ/EMBL/GenBank accession No. J03075). None of these proteins have been linked to RNA metabolism. The partial sequence of the ~50 kDa polypeptide matched the human cDNA sequence of a ~34 kDa ORF termed E5.1 (DDBJ/EMBL/GenBank accession No. L37368) which has sequence features characteristic of an RNA-binding protein, i.e. a single RNA-recognition motif (RRM) with well conserved RNP-1 and RNP-2 submotifs (Figure 2D; Badolato *et al*., 1995). The RRM is preceded by an extensive serine-rich domain. A mouse homologue was originally identified as a cloned cDNA encoding a putative nucleic acid-binding protein of unknown function (Schmidt and Werner, 1993). Conveniently, the original mouse protein name, RNPS1 (for RNA-binding protein prevalent during S phase), is also an acronym for 'RNA-binding protein with serine-rich domain', and so here we adopt the same name for the human protein.

A notable feature of the RNPS1 ORF is the absence of cysteines, which led us to use covalent chromatography on thiopropyl-agarose to separate this protein from the more abundant contaminants (Figure 2C). As expected from their sequences, the contaminants bound quantitatively to the resin, whereas the ~50 kDa RNPS1 protein was recovered in the unbound fraction in an apparently homogeneous form (fraction VI). This protein proved to be sufficient for the activity observed in the SF7A fraction (see below).

Preparation of recombinant RNPS1

To confirm rigorously that the RNPS1 protein is responsible for SF7A activity and to address the discrepancy in the predicted and observed size of the protein, a full-length human E5.1 cDNA was expressed in baculovirus-infected insect cells. Recombinant RNPS1 (r-RNPS1) was recovered exclusively in the insoluble fraction of the Sf9 cell lysate (Figure 3A, lane 4). After denaturation with urea, the protein was purified to near homogeneity (Figure 3B, lane 2). Purified r-RNPS1 migrated at ~50 kDa on SDS–PAGE, i.e. it was indistinguishable from HeLa cell RNPS1 (Figure 3C, lanes 1 and 2). r-RNPS1 was also expressed in *Escherichia coli* and purified, but this form of the protein migrated faster on SDS–PAGE, with ~42 kDa apparent molecular weight (lane 3). The same expression plasmid, bearing a T7 promoter, was transcribed and translated in wheat germ extract, generating a labeled

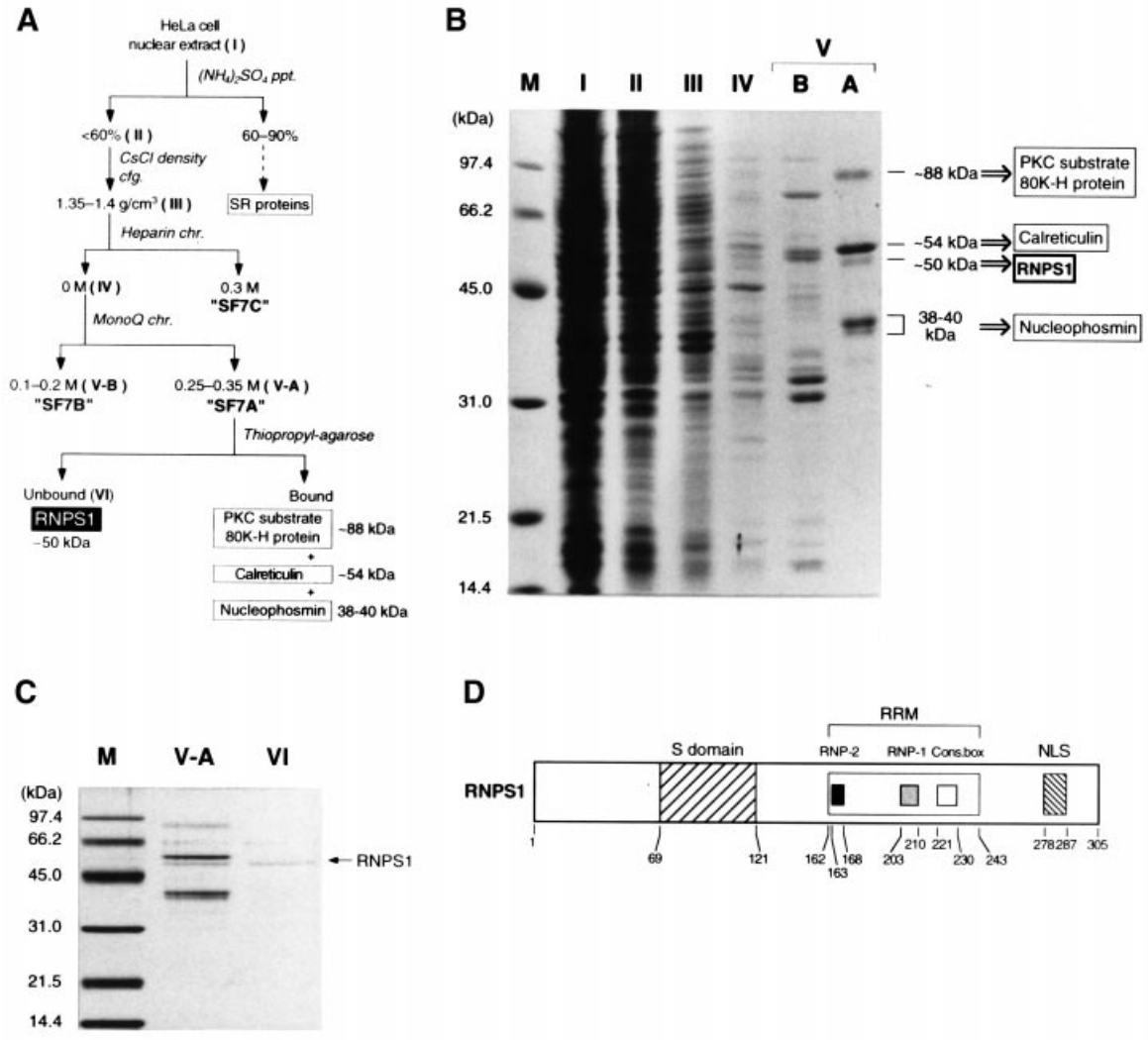

Fig. 2. Purification and identification of human RNPS1. (**A**) Purification profile of SF7 activity from HeLa cell nuclear extracts. Methods of purification are indicated in italics. Individual fractions are designated with Roman numerals. (**B**) Protein analysis of fractions (I, 3 μl; II, 2 μl; III, 2 μl; IV, 3 μl; V-B, 8 μl and V-A, 8 μl) by 12% SDS–PAGE and Coomassie Blue staining. M: molecular weight (kDa) markers. The five polypeptides identified by partial amino acid sequencing, corresponding to four different proteins, are indicated with their apparent molecular weights. (**C**) Final purification of fraction V-A by thiopropyl-agarose (fraction VI). The purified human RNPS1 polypeptide (~50 kDa) is indicated. (**D**) Primary structure of the RNPS1 polypeptide based on the E5.1 cDNA sequence (Badolato *et al*., 1995). The single canonical RNA-recognition motif (RRM) with conserved RNP-1 and RNP-2 submotifs, the extensive serine-rich domain (S domain), the C-terminal presumptive nuclear localization signal (NLS), and a phylogenetically conserved decapeptide box (Cons. box) are indicated. The numbers indicate amino acid positions from the initiation codon (Met1).

protein with the same electrophoretic mobility (~50 kDa) as HeLa cell RNPS1 and baculovirus-expressed r-RNPS1 (Figure 3D, lane 4). N-terminal amino acid sequence analysis showed that the *E.coli*-expressed protein initiated correctly and its N-terminus remained intact (data not shown). Therefore, bacterial expression results in a protein that lacks a post-translational modification or is perhaps truncated at the C-terminus through proteolysis or premature termination of translation attributable to rare codons. So far, we have been unable to distinguish between these two possibilities.

Effect of RNPS1 on alternative and constitutive splicing

In vitro splicing assays are commonly performed either with HeLa cell nuclear extract or with cytosolic S100 extract complemented with SR protein(s) (Mayeda and Krainer, 1999b). As in the case of SR proteins, we found that RNPS1 is much less abundant in S100 than in nuclear extract (Figure 3E, lanes 1 and 2). Therefore, to maximize the effects of exogenous RNPS1, we used the S100 extract in the presence of limiting amounts of SF2/ASF, which is essential for splicing (Krainer *et al*., 1990a).

Addition of purified baculovirus-expressed r-RNPS1 strongly activated both distal and proximal 3′ splice sites of a model β-globin pre-mRNA with duplicated 3′ splice sites, and the specific activity was equivalent to that of purified HeLa cell RNPS1 (Figure 4A, lanes 1–3). In contrast, *E.coli*-expressed r-RNPS1 had no stimulatory activity and was in fact inhibitory (lane 4). Therefore, full-length and/or post-translationally modified RNPS1 is necessary for splicing activity. Both baculovirus-expressed and HeLa cell RNPS1 also strongly stimulated splicing of constitutively spliced β-globin pre-mRNA (Figure 4A,

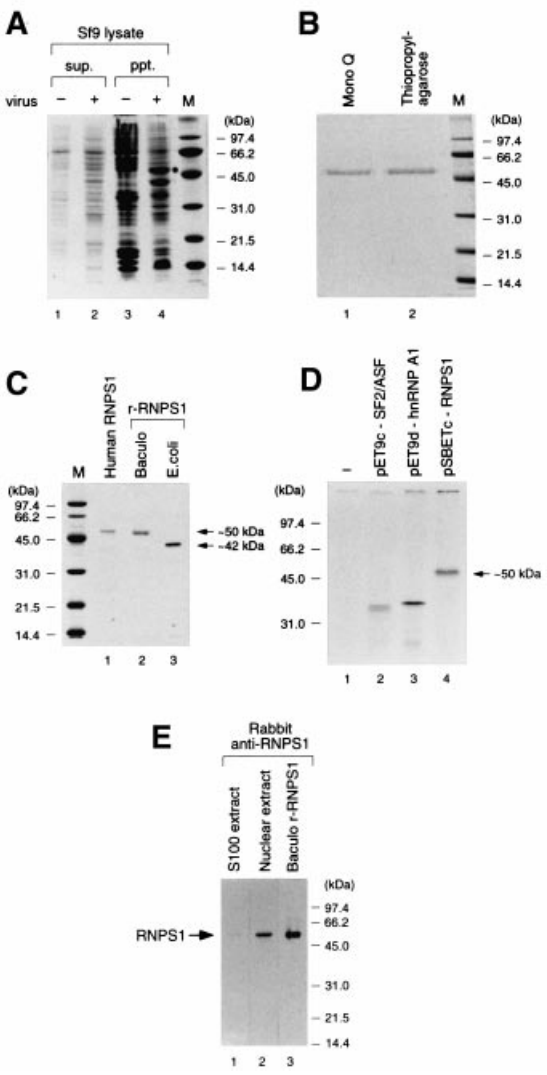

Fig. 3. Expression and purification of recombinant RNPS1 (r-RNPS1). (**A**) Expression of authentic r-RNPS1 in baculovirus. Sf9 cells infected with pVL-RNPS1 virus (+) or mock-infected (–) were lysed and centrifuged. Supernatant (sup.) and pellet (ppt.) fractions were analyzed by 12% SDS–PAGE and Coomassie Blue staining. Overproduced r-RNPS1 (~50 kDa) is indicated by a dot. (**B**) Purification of baculovirus r-RNPS1. The RNPS1 peak Mono Q fraction and thiopropyl-agarose unbound fraction were analyzed by 12% SDS–PAGE. (**C**) Comparison of native and recombinant RNPS1 proteins. Protein analysis of purified human (HeLa cell) RNPS1, baculovirus r-RNPS1 and *E.coli* r-RNPS1 by 12% SDS–PAGE. The apparent molecular weights of the proteins are indicated. (**D**) *In vitro* translation of *E.coli* expression plasmids in wheat germ extract. The indicated plasmids and a no-plasmid control (–) were used directly in a coupled transcription/translation system. The translated ^{35}S-labeled products were analyzed by 12% SDS–PAGE and autoradiography. (**E**) Immunoblotting analysis of RNPS1. Total proteins (~5 μg) from HeLa cell nuclear and cytosolic S100 extracts were probed with rabbit anti-RNPS1 antibody. Purified baculovirus r-RNPS1 (~5 pmol) was included as a control. The estimated abundance of RNPS1 in the nuclear and cytosolic extracts is ~0.3 and ~0.01 pmol/μg protein, respectively.

lanes 5–7). Similar stimulatory effects, with varied efficiencies, were also observed with four other well characterized constitutively spliced mini-gene transcripts derived from HIV-*tat*, immunoglobulin μ-chain (IgM), *Drosophila ftz* and chicken δ-crystallin (Figure 4B). We also observed a splicing stimulatory effect of RNPS1 when splicing was performed in dilute nuclear extract (data not shown). In contrast, RNPS1 had no effect on *in vitro* splicing of a pre-mRNA with an AT-AC intron (M.L.Hastings and A.R.Krainer, unpublished data). Therefore, the stimulatory function of RNPS1 appears to be specific for the major splicing pathway.

To test whether RNPS1 activity is influenced by the distance between alternative 3′ splice sites or by their sequence context, we tested several additional model β-globin pre-mRNAs in which these parameters vary systematically (Figure 4C). Both distal and proximal 3′ splice sites were markedly activated by addition of baculovirus r-RNPS1 in the case of pre-mRNAs with either duplicated 3′ splice sites in tandem (Figure 4C, lanes 1–8) or duplicated 3′ splice sites with substitutions in the middle exons (lanes 9–14), although the relative use of the proximal 3′ splice site increased as the size of the duplication increased. Remarkably, efficient use of the distal 3′ splice sites in the pre-mRNAs with the long duplications, 3′D-205, 3′D-115 and 3′S-188, was not observed previously under standard conditions, either in nuclear extract or in S100 extract plus SF2/ASF (Reed and Maniatis, 1986; data not shown). Baculovirus r-RNPS1 also significantly activated overall splicing of a model β-globin pre-mRNA with duplicated 5′ splice sites (5′D-16X, Figure 1) via both proximal and distal sites (data not shown).

We conclude that the RNPS1 protein is not sufficient for SF7 activity, i.e. full switching from the proximal to the distal 3′ splice site, as observed in partially purified fractions. It is likely that another factor(s), possibly present in the SF7C fraction, suppresses use of the proximal 3′ splice site, such that both activities are necessary for full splice-site switching activity.

Synergistic splicing activation by RNPS1 and SR proteins

The observation that RNPS1 stimulates overall splicing activity with a wide variety of pre-mRNA substrates suggests that it plays a general role in the splicing reaction. We took advantage of the fact that the S100 extract has only trace amounts of both RNPS1 and SR proteins (Figure 3E; Krainer *et al*., 1990a) to test for functional interactions between RNPS1 and SF2/ASF.

RNPS1 alone did not complement the S100 extract efficiently for splicing activity, as seen for example with the 3′D-55 pre-mRNA (Figure 5, lanes 1 and 2). However, when SF2/ASF was titrated into the reaction in limiting amounts to complement the S100 extract, RNPS1 behaved as a potent activator of the splicing reaction (lanes 3–8). This effect was not simply additive, but rather it showed a strong synergy between SF2/ASF and RNPS1. A similar synergistic effect was observed with the constitutively spliced β-globin pre-mRNA and with four other pre-mRNAs (Figure 5, lanes 9–11; Figure 4B; data not shown). The observed splicing stimulation is not simply attributable to overall RNA stabilization, since the accumulation of spliced mRNA increased over an order of magnitude, whereas the levels of pre-mRNA remained constant or even decreased somewhat (Figure 5, lanes 3–11). The synergistic activation appears not to be highly specific for a particular SR protein, since in addition to SF2/ASF, RNPS1 strongly activated β-globin pre-mRNA splicing in

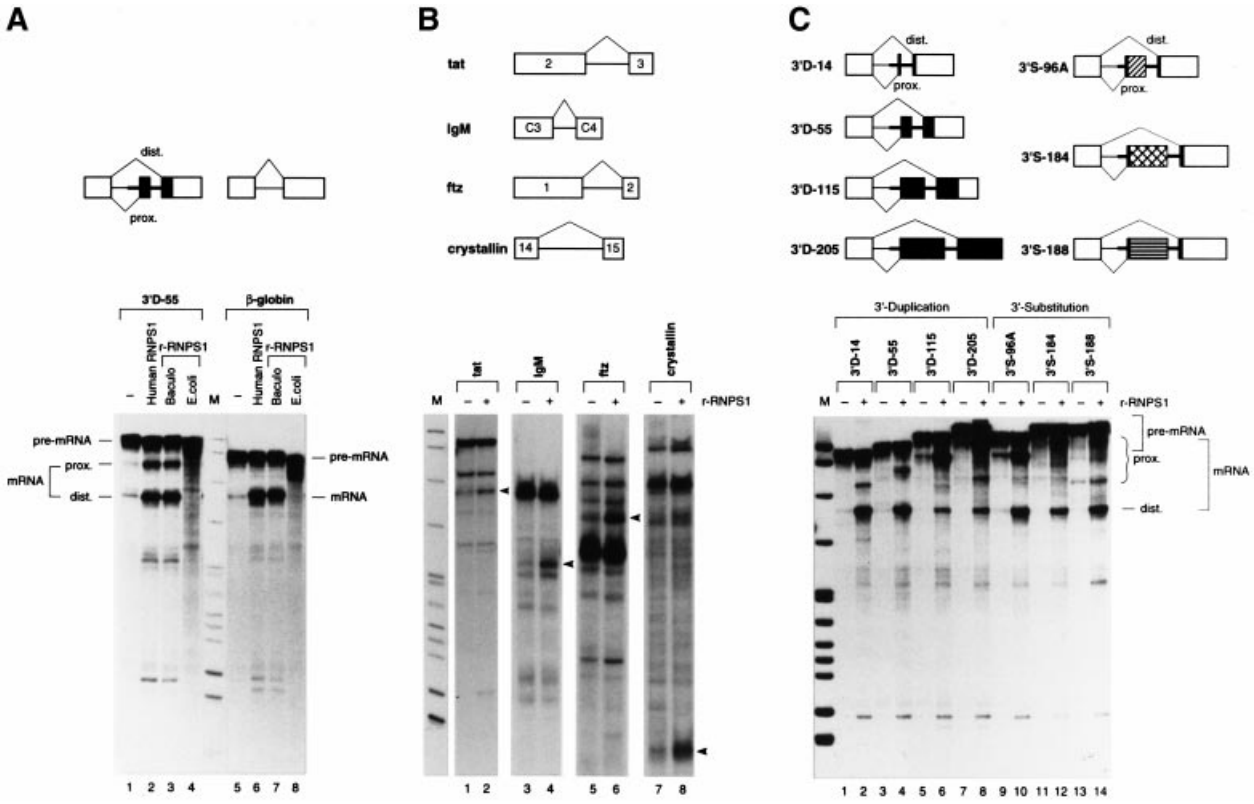

Fig. 4. Effect of RNPS1 on alternative and constitutive splicing *in vitro*. The structure of the different pre-mRNAs with possible splicing paths is shown schematically above each panel. The cross-hatched patterns in (C) denote the different exon segment substitutions (Reed and Maniatis, 1986) and other shading and labeling is as in Figure 1. (**A**) *In vitro* splicing with duplicated 3′ splice site pre-mRNA (3′D-55) and constitutive pre-mRNA (β-globin). Constant amounts of HeLa cell S100 extract (7 μl) and purified SF2/ASF (5 pmol) were incubated without (–) or with 10 pmol of purified human (HeLa cell) RNPS1, baculovirus r-RNPS1 or *E.coli* r-RNPS1 in the splicing reactions. (**B**) *In vitro* splicing with four other constitutive pre-mRNAs: HIV-*tat*, IgM (C3-C4), *Drosophila ftz* and δ-crystallin. Splicing conditions were as in (A), with 10 pmol of baculovirus r-RNPS1 (+). Spliced mRNA products are indicated by arrowheads. *In vitro* splicing of these substrates has been characterized previously (Rio, 1988; Sawa *et al.*, 1988; Mayeda *et al.*, 1999 and references therein). (**C**) *In vitro* splicing with 3′ splice site-duplicated/substituted β-globin pre-mRNAs. Splicing conditions were as in (B) but with 1.25 pmol of SF2/ASF.

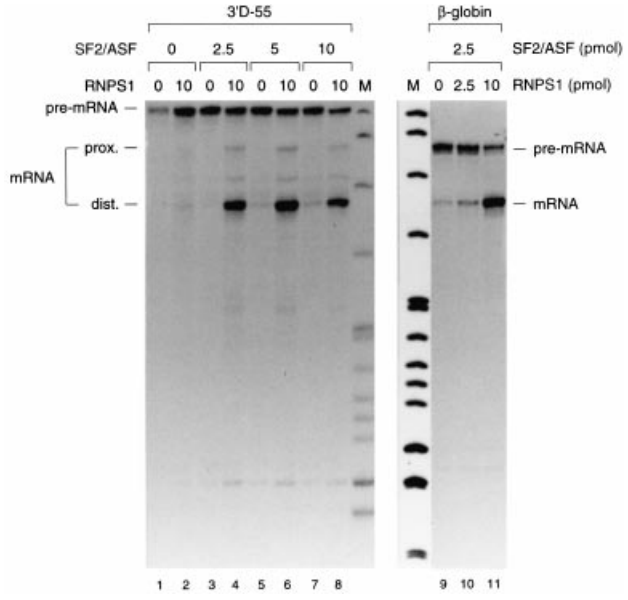

Fig. 5. Synergistic effect of RNPS1 and SF2/ASF on splicing *in vitro*. Duplicated 3′ splice site pre-mRNA (3′D-55) and wild-type pre-mRNA (β-globin) were spliced in S100 extract (7 μl) with the indicated amounts of SF2/ASF and baculovirus r-RNPS1. See Figure 4A for the structures of the substrates.

the presence of limiting SC35 (data not shown). SF2/ASF and SC35 represent two different kinds of SR proteins, with two RRMs or one RRM, respectively (Cáceres and Krainer, 1997). We conclude that RNPS1 functions as a potent splicing activator, allowing efficient splicing to proceed in the presence of limiting amounts of SR proteins.

Subcellular localization of RNPS1

RNPS1 is considerably more abundant in HeLa cell nuclear extract than in cytosolic S100 extract (Figure 3E), suggesting that it is predominantly or exclusively a nuclear protein. To investigate the subcellular localization of RNPS1, we transiently expressed epitope-tagged RNPS1 in HeLa cells and detected it by immunofluorescence microscopy with antibody against the epitope tag (Figure 6). This approach was used because our RNPS1 antibodies did not work well in immunofluorescence assays. The tagged RNPS1 localized exclusively to the nucleus, giving an intense speckled pattern and a diffuse nucleoplasmic staining, but it was excluded from the nucleoli (Figure 6a). This pattern is clearly distinct from that of hnRNP A1, which shows only a diffuse nuclear distribution (Figure 6d). The transiently expressed RNPS1 colocalized in the speckles with the endogenous SR protein SC35 (Figure 6b), as seen by double-label immunofluorescence (Figure 6c).

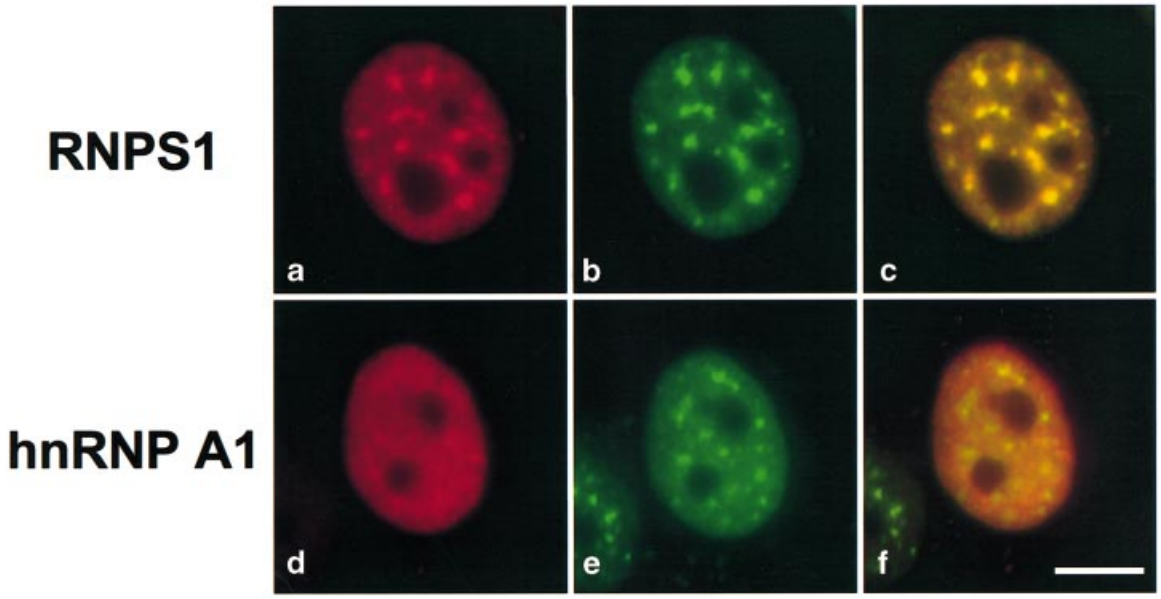

Fig. 6. Subcellular localization of transiently expressed RNPS1. HeLa cells were transiently transfected with plasmids expressing T7-tagged RNPS1 or hnRNP A1, and analyzed by indirect immunofluorescence microscopy. RNPS1 and hnRNP A1 were detected with an anti-T7 antibody, followed by Texas Red-conjugated secondary antibody (**a** and **d**). Endogenous SC35 was detected with a fluorescein-conjugated anti-SC35 antibody (**b** and **e**). Both images were superimposed and the yellow color indicates colocalization of the two proteins in nuclear speckles (**c**) but hnRNP A1 did not accumulate in speckles and showed only a diffuse nuclear pattern (**f**). Bar, 5 μm.

Association of RNPS1 with splicing complexes

To test whether RNPS1 is associated with active spliceosomes, an *in vitro* splicing reaction was incubated with immobilized anti-RNPS1 antibodies and the immunoprecipitated labeled RNAs were analyzed (Figure 7). RNPS1 preferentially bound (directly or indirectly) to the spliced mRNA product and to a lesser extent to the pre-mRNA (Figure 7, lane 3), as did SF2/ASF (lane 4; Hanamura *et al.*, 1998). In contrast, immunoprecipitation with anti-hnRNP A1 preferentially selected the intron-containing molecules (lane 5; Hanamura *et al.*, 1998). These data demonstrate that RNPS1 and SF2/ASF are stably associated with complexes containing mRNA, as are other SR proteins (Blencowe *et al.*, 1995).

Phylogenetic conservation of RNPS1

If RNPS1 plays an important role in pre-mRNA splicing, it should be highly conserved, as are other proteins involved in splicing. We searched the protein, ORF and expressed sequence tag (EST) databases with human RNPS1 peptide and DNA sequences. Presumptive homologues from mouse, rat, zebrafish (*Danio rerio*), fruit fly (*Drosophila melanogaster*) and nematode (*Caenorhabditis elegans*) were aligned with the human RNPS1 sequence (Figure 8).

Mouse RNPS1 was identified by differential screening as a cDNA corresponding to a transcript expressed at higher levels during the S phase of the cell cycle (Schmidt and Werner, 1993). As noted previously, the mouse and human proteins can be aligned and show ~99% amino acid identity if several apparent frameshift errors in the mouse sequence are corrected (Badolato *et al.*, 1995). Here, the correct full-length mouse cDNA sequence was retrieved from three overlapping EST clones (Figure 8), confirming the presence of sequencing errors in the original RNPS1 sequence. Partial cDNA sequences from rat and zebrafish are also available; the rat sequence is nearly

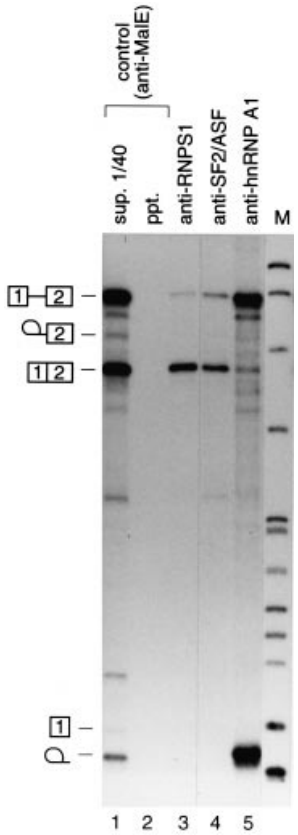

Fig. 7. Immunoprecipitation of splicing reactions. *In vitro* splicing reactions with labeled β-globin pre-mRNA were incubated and immunoprecipitated with the indicated antibodies immobilized on protein G–Sepharose. RNA was extracted from the pellets and analyzed by denaturing PAGE and autoradiography (lanes 2–5). One-fortieth of the total RNA recovered from the supernatant of an immunoprecipitation with the control maltose-binding protein (MalE) antibody (lane 1) reflects the initial relative abundances of the pre-mRNA, intermediates and products.

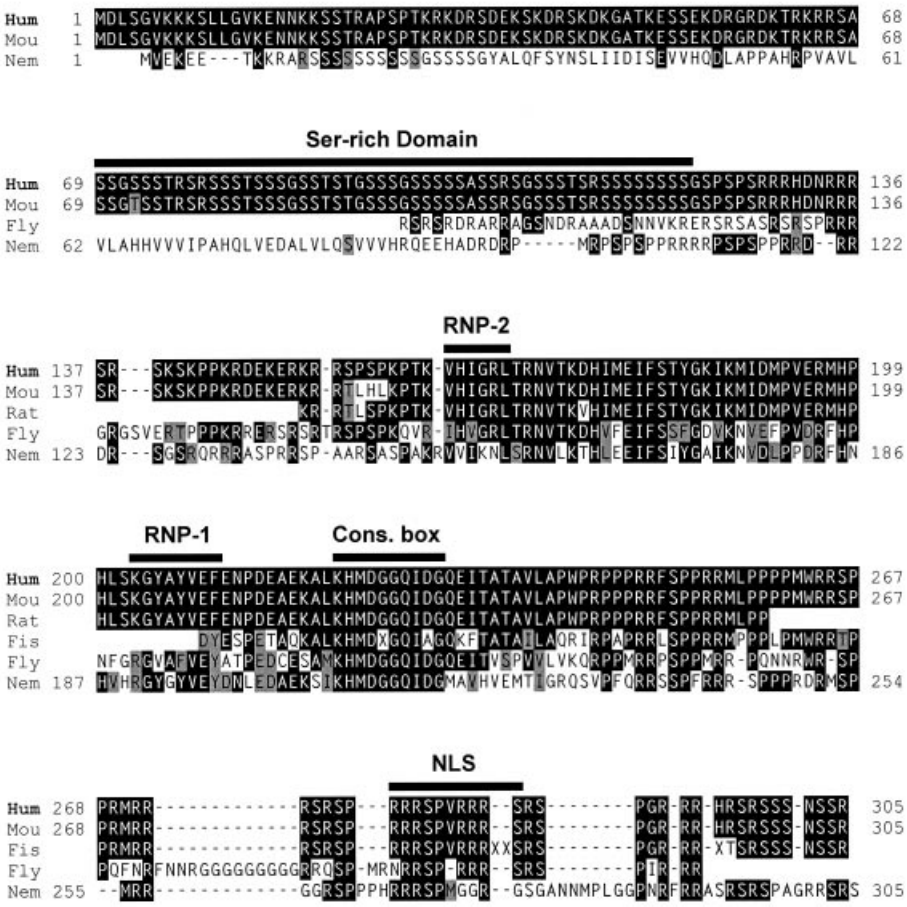

Fig. 8. Multiple alignment of presumptive RNPS1 homologues. Amino acid sequences of selected RNPS1 putative homologues from mouse (Mou), rat (Rat), zebrafish (Fis), fruit fly (Fly) and nematode (Nem) were aligned based on the human RNPS1 (Hum) sequence. Positions of identity or similarity to the human sequence are shaded black and gray, respectively. The similarity groupings applied were H~K~R, I~L~M~V, F~W~Y, S~T, D~E and N~Q. Gaps are indicated by dashes. Only partial sequences of the rat, zebrafish and fruit fly proteins were available. Three frameshifts in the zebrafish EST fragment are likely sequencing errors; the undefined amino acids resulting from restoring the correct frame are indicated by X. The serine (Ser)-rich domain, the RNP-1 and RNP-2 submotifs of the RRM, the putative nuclear localization signal (NLS) and a conserved decapeptide sequence (Cons. box) of unknown function are indicated by horizontal bars. Accession numbers for the source sequences are as follows: human (DDBJ/EMBL/GenBank, L37368), mouse (EMBL: X70067, corrected by DDBJ/EMBL/GenBank, AA839995, AA823056, AA930346), rat (DDBJ/EMBL/GenBank, H35524, AA687005), zebrafish (DDBJ/EMBL/GenBank, AA605765; three single-base deletions are likely sequencing errors), fruit fly (DDBJ/EMBL/GenBank, AA951371, AA942134) and nematode (DDBJ/EMBL/GenBank, U00052).

identical to the human and mouse sequences, whereas the zebrafish sequence is somewhat more divergent (~97 and ~78% identity, respectively).

The homologous sequences from lower metazoans, nematode and fruit fly (partial), are much more divergent, as expected, especially towards the N-terminus. The five presumptive homologues for which sufficient sequence is available have one canonical RRM with highly conserved RNP-1 and RNP-2 submotifs, as well as a potential nuclear localization signal (NLS) near the C-terminus (Dingwall and Laskey, 1991). The mammalian proteins have an extensive serine-rich region upstream of the RRM. The nematode sequence has a much shorter cluster of serines very near the N-terminus. The decapeptide sequence KHMDGGQIDG, downstream of the RNP-1 submotif, is perfectly conserved in five of the sequences. This decapeptide sequence is not present in any known PROSITE motifs, in other unrelated RRMs, or in any other proteins in the current databases.

Discussion

Function of RNPS1 in pre-mRNA splicing

The biochemical purification of SF7 activity from HeLa cells relied on a splice-site switching assay, in which fractions that promoted the use of distal 3′ splice sites were identified. The first protein we have identified as a result of this effort proved not to discriminate between distal and proximal sites, but rather to activate splicing in a more general manner. The splice-site switching activity observed in cruder fractions may reflect the cooperation of RNPS1 with additional unknown factors. Alternatively, RNPS1 does not contribute directly to SF7 activity but is instead a more general splicing factor whose activity became unmasked during the fractionation. Further purification of the SF7B and SF7C fractions will be required before the relationship of these activities to RNPS1 and their role in alternative 3′ splice site selection can be assessed.

Although we have shown that hnRNP A1 affects alternative 5′ but not 3′ splice site selection *in vitro* (Figure 1; Fu *et al*., 1992; Mayeda *et al*., 1993), this protein was recently shown to stimulate distal 3′ splice site use *in vivo* (Bai *et al*., 1999). This apparent discrepancy may reflect differences between *in vitro* and transient transfection assays or between the substrates used in each case.

The strong synergistic effect between RNPS1 and either SF2/ASF or SC35 is a distinctive property of RNPS1 that is not shared by SR proteins. The effect of different SR proteins in enhancer-dependent splicing is additive, rather than synergistic or cooperative (Hertel and Maniatis, 1998), and the same is the case for general splicing (data not shown). The splicing activator function of RNPS1 is reminiscent of the nuclear matrix protein complex SRm160/300, which also stimulates splicing in the presence of limiting SR proteins (Blencowe *et al*., 1998). However, two features of RNPS1 make it functionally unique and distinct from this nuclear matrix protein complex (Blencowe *et al*., 1998; B.Blencowe and P.Sharp, personal communication). First, both SRm160 and SRm300 are serine/arginine-rich proteins but they lack RRMs, whereas RNPS1 has a typical RRM, which is a prevalent RNA-binding motif present in SR proteins and many other splicing factors, as well as in other RNA-binding proteins. There is no significant sequence homology between SRm160/300 and RNPS1, although SRm160 has two short runs of serines and SRm300 has two very long stretches of consecutive serines. Consistent with the lack of an RRM, SRm160/300 interacts with the pre-mRNA indirectly through SR proteins and U1 and U2 snRNPs, and is therefore considered to be a splicing coactivator (Blencowe *et al*., 1998). A second important difference is that the splicing stimulatory effect of SRm160/300 is highly dependent upon the substrate. In contrast, RNPS1 stimulates splicing in a more general manner. SRm160/300 has little effect on splicing of substrates that are not highly dependent on U1 snRNP, such as β-globin pre-mRNA, whereas this substrate was the most responsive to RNPS1 among several substrates we tested. Therefore, the mechanisms of splicing activation and coactivation by RNPS1 and SRm160/300 are probably different.

RNPS1 bound stably to splicing complexes containing spliced mRNA, and it activated splicing of all conventional introns tested. However, whether RNPS1 is essential for basal splicing or its role is only stimulatory is unclear at the present time. Extensive immunodepletion of RNPS1 from S100 extract resulted in loss of basal splicing activity in the presence of limiting SF2/ASF, which could be restored by addition of baculovirus-expressed r-RNPS1, consistent with an essential role of RNPS1 (data not shown). However, splicing activity was also restored by adding excess SF2/ASF and thus we cannot distinguish between the cooperative effect of residual RNPS1 with SF2/ASF and a putative independent role of RNPS1 that might be required for splicing. We could not successfully immunodeplete RNPS1 from nuclear extract, where it is very abundant.

The $p34^{cdc2}$-related PITSLRE kinase isoform p110 interacts specifically with human RNPS1 *in vitro* and *in vivo* (Loyer *et al*., 1998). The C-terminal region of RNPS1, which appears to be important for the splicing activity (see below), is dispensable for the interaction with p110. This recent study also reported the anomalous mobility of RNPS1 on SDS–PAGE and the colocalization of RNPS1 with SC35. In contrast to our results, however, this study showed transiently expressed RNPS1 localized to only six or so mega-speckles, together with p110 and a fraction of SC35; moreover, there was only partial overlap in the localization of SC35 with RNPS1 in the mega-speckles, with SC35 concentrating at their periphery (Loyer *et al*., 1998). This previously reported localization pattern probably reflects overexpression of RNPS1, as noted by the authors. In our expression system, RNPS1 always gave the standard speckle plus diffuse nucleoplasmic pattern characteristic of SR proteins and many other splicing factors (reviewed in Misteli and Spector, 1998). The major structural component of the nuclear speckles is the interchromatin granule clusters (IGC), which are thought to represent storage or assembly sites for splicing components, from where they are supplied to nascent pre-mRNA transcripts. Mouse RNPS1 was recently shown to be enriched in purified IGC (D.L.Spector, personal communication). The observed localization of RNPS1 is consistent with a role in splicing, as suggested by the functional data *in vitro*.

RNPS1 was also identified in a yeast two-hybrid screen using the Clk/Sty kinase as bait (Colwill *et al*., 1996). The Clk/Sty kinase, which has its own RS domain, phosphorylates SF2/ASF and may regulate its splicing activity (reviewed in Misteli and Spector, 1998). In light of our present findings concerning the function of RNPS1 in splicing, phosphorylation of RNPS1 and SF2/ASF as well as physical interactions between RNPS1, SF2/ASF and Clk/Sty (perhaps modulated by phosphorylation) are plausible mechanisms to be explored.

Post-translational modification of RNPS1

We observed a pronounced difference in the electrophoretic mobilities of purified HeLa cell RNPS1, baculovirus r-RNPS1 or *in vitro*-translated r-RNPS1, compared with *E.coli*-expressed r-RNPS1. The faster-migrating protein expressed in bacteria not only lacked splicing stimulatory activity, but also had a dominant-negative effect, inhibiting splicing in nuclear extract, i.e. in the presence of wild-type RNPS1 (data not shown). Since RNPS1 migrates more slowly than expected from its deduced molecular weight and has multiple potential serine/threonine phosphorylation sites, we analyzed its state of phosphorylation. Treatment of baculovirus-expressed r-RNPS1 with bacterial alkaline phosphatase slightly increased its mobility (~50 versus ~48 kDa; data not shown), but the dephosphorylated protein still migrated more slowly than *E.coli*-expressed r-RNPS1 (~42 kDa). Conversely, post-translational modification of *E.coli* r-RNPS1 either in the splicing reaction with crude extracts or by incubation in wheat germ translation extract, resulted in a slight decrease in electrophoretic mobility (~42 versus ~43.5 kDa; data not shown), but this is much less than the mobility of HeLa cell, baculovirus-expressed or *in vitro*-translated RNPS1 (~50 kDa). There are potential glycosylation sites in the RNPS1 sequence, but treatment with *N*- and *O*-glycosidases did not affect the mobility of baculovirus r-RNPS1 (data not shown).

We conclude that there is a post-translational modification in native RNPS1—probably phosphorylation—while additional modifications of the protein as well as the integrity of *E.coli* r-RNPS1 at the C-terminus are presently uncertain. If there is indeed a truncation of the C-terminus of *E.coli* r-RNPS1, this portion of the molecule may be functionally important, since this form of the protein is not only inactive but also has a dominant negative effect on splicing *in vitro*.

Structure of RNPS1 and homologous proteins in other organisms

The multiple sequence alignment of presumptive RNPS1 homologues shows extremely high conservation of the protein sequence in mammals, whereas the fruit fly and nematode sequences are more degenerate, especially near the N-terminus (Figure 8). Therefore, any functional requirement for the serine-rich domain may be different in lower metazoans. In contrast, the C-terminal half, including the RRM, is highly conserved in all the sequences. This region includes a strikingly conserved decapeptide motif of unknown function, KHMDGGQIDG, near the C-terminus of the RRM, just downstream of the RNP-1 submotif (Figures 1D and 8). This portion of the RRM includes a part of helix α_2 and loop5 and is relatively degenerate among all RRMs except for the Gly-residues at positions five and ten of the decapeptide and the Ile-residue at position eight, which show partial conservation in other known RRMs (Birney *et al*., 1993). The decapeptide is therefore a distinctive signature for the RNPS1-like proteins and is likely to be functionally significant. The RRM of RNPS1 is only distantly related to the RRMs of SR proteins, and more closely resembles the RRM of the hnRNP C protein (Badolato *et al*., 1995).

We also searched the yeast sequence databanks for putative RNPS1 homologues, and the closest candidate, annotated as a possible splicing factor, was found in fission yeast (DDBJ/EMBL/GenBank accession Nos AL031158, AL022600). However, this sequence shows only partial conservation of the decapeptide motif and has much lower overall homology compared with the other listed homologues (data not shown). No functional data are available on this yeast protein. So far, no candidate RNPS1 homologues have been found in the plant sequence databases.

The human gene encoding RNPS1 has been cloned as 'gene G' (encoding an RNA-binding protein of unknown function) within the 700 kb region surrounding the polycystic kidney disease and tuberous sclerosis disease genes on chromosome 16p13.3 (Burn *et al*., 1996; DDBJ/EMBL/GenBank accession No. AC005212). Comparison of the genomic and cDNA sequences shows that the gene has seven introns, of which one interrupts the 5′ non-coding region. The expression of this gene was examined by Northern blotting analysis and a transcript of the expected size was observed, whose levels varied in different human tissues but not significantly during the cell cycle (Badolato *et al*., 1995). This pattern of expression contrasts with that of the mouse homologue, which was reported to be expressed at higher levels during S phase (Schmidt and Werner, 1993). No alternatively spliced transcripts of human RNPS1 have been detected in several cell lines examined (J.Badolato, unpublished data). The ubiquitous expression of RNPS1 is consistent with its role as a general splicing activator.

Materials and methods

In vitro splicing

m^7GpppG-capped ^{32}P-labeled pre-mRNA substrates were made by runoff transcription of linearized template DNA with SP6 RNA polymerase (Mayeda and Krainer, 1999b). β-globin and derivative plasmid templates have been described (Reed and Maniatis, 1986; Krainer *et al*., 1990b). HIV-*tat*, IgM (C3-C4), *Drosophila ftz* and δ-crystallin pre-mRNAs were transcribed from plasmids pSP64-HIV-1/*tat*23, pμC3-C4, pSPftz and pSP14-15, respectively (Sawa *et al*., 1988; Inoue *et al*., 1990; Mayeda *et al*., 1999).

HeLa cell S100 extract and purified SF2/ASF were prepared as described (Mayeda *et al*., 1993; Mayeda and Krainer, 1999a). *In vitro* splicing reactions in 25 μl with the indicated amounts of S100 extract, purified SF2/ASF, purified or recombinant RNPS1, and 20 fmol of ^{32}P-labeled pre-mRNA were incubated at 30°C for 3–4 h (Mayeda and Krainer, 1999b). The RNA products were analyzed by electrophoresis on a 5.5% polyacrylamide/7 M urea gel and autoradiography.

Purification and identification of human RNPS1

The purification outline and fraction designations are given in Figure 2A. Individual fractions were assayed for SF7 activity, i.e. the ability to stimulate use of the distal 3′ splice site of 3′D-55 or 3′D-115 pre-mRNA (Figure 4C) in HeLa cell S100 extract in the presence of 5–10 pmol SF2/ASF. Maximal splicing activity requires ~30 pmol of SF2/ASF.

SF7 activity was recovered in the 60% saturated $(NH_4)_2SO_4$ pellet of extract from sonicated nuclei (fraction I; Mayeda and Krainer, 1992) from a 60 l HeLa cell culture (~6×10^{10} cells). The pellet was resuspended in buffer A [20 mM HEPES–NaOH pH 8.0, 5% (v/v) glycerol, 0.2 mM EDTA, 1 mM dithiothreitol (DTT)] with 0.5 mM phenylmethylsulfonyl fluoride (PMSF) (fraction II; 42 ml). CsCl density gradient centrifugation was performed as described (Krainer *et al*., 1991) but on a larger scale, i.e. centrifugation in four polycarbonate tubes (25×89 mm, Nalgene) at 48 000 r.p.m. (Type 60Ti rotor, Beckman) for 24 h at 4°C. This procedure separates nucleic acids from the bulk of protein, facilitating ion exchange chromatography. Strong SF7 activity was recovered in fractions 3–5 from the top, out of fourteen 1.8 ml fractions. The pooled peak (fraction III) was dialyzed twice in 0.5 l of buffer B (buffer A + 6 M urea) with 0.5 mM PMSF. The dialysate was applied to a HiTrap Heparin column (5 ml; Pharmacia), which was washed with 40 ml of buffer B and eluted with 20 ml each of 0.3 M NaCl and 1 M NaCl in buffer B. SF7 activity was recovered in both the flow-through (fraction IV) and the 0.3 M salt eluate ('SF7C'). Fraction IV was loaded on a Mono Q column (1 ml; Pharmacia), washed with 30 ml buffer B, and eluted with a 40 ml linear gradient from 0 to 0.5 M NaCl in buffer B, followed by 5 ml of 1 M NaCl in buffer B. SF7 activity was recovered in each of two peaks, eluting at ~0.1–0.2 M salt (fraction V-B; 'SF7B') and 0.25–0.35 M salt (fraction V-A; 'SF7A').

To obtain partial amino acid sequences, fraction V-A was separated by preparative SDS–PAGE and five polypeptides were individually excised and digested with lysylendopeptidase (Wang *et al*., 1996). The eluted peptide fragments were separated by HPLC (Model 1090, Hewlett Packard) using a Vydac C_{18} column (Separations Group) and analyzed in an automated protein sequencer (Procise 494, Applied Biosystems). The peptide sequences obtained were used for database searches. Identification of the ~50 kDa polypeptide as E5.1 (human RNPS1) was based on three non-overlapping peptide sequences, LLGVK, VHIGRLTRNVTK and MIDMPVERMHPHL.

Covalent chromatography was used for final purification of RNPS1 to apparent homogeneity. One milliliter of fraction V-A was dialyzed twice in 0.5 l of 0.1 M KCl in buffer B lacking DTT. An aliquot of 0.4 ml of the dialysate was added to ~0.2 ml of thiopropyl-agarose (Sigma) and the suspension was rocked at 4°C overnight. The suspension was spun and the supernatant was filtered through a Micropure filter unit (Amicon) and dialyzed against 1 l of buffer A plus 0.1 M KCl and 0.5 mM PMSF (fraction VI). The essentially pure RNPS1 fraction was aliquoted and stored at –70°C.

The protein concentration of each fraction was determined by the dye-binding method (Bio-Rad), standardized with BSA.

Preparation of baculovirus-expressed recombinant RNPS1

The coding region of RNPS1 was amplified using the E5.1 cDNA as a template (Badolato *et al*., 1995) and the DNA primers RNPS1-N2 (GAGTCGGATCCATGGATTTATCAGGAGTGA) and RNPS1-C (CACACGGATCCTTATCGGGAGGAGTTG) by PCR with Vent DNA polymerase (New England Biolabs). The amplified 940 bp PCR fragment was subcloned into the *Bam*HI site of pVL1392 (PharMingen) generating the pVL-RNPS1 baculovirus expression construct, whose structure was verified by sequencing.

The pVL-RNPS1 plasmid was introduced into Sf9 insect cells using a BaculoGold transfection kit (PharMingen) and a high-titer viral stock was generated. Large scale amounts of the protein were generated by infecting ~2×10^7 Sf9 cells/T150 tissue culture flask with ~5 ml (~5×10^8 virus particles) of viral stock. The infected cells were incubated at 27°C for 4 days, centrifuged, resuspended in PBS and frozen at –70°C.

The frozen infected Sf9 cell suspension (~9 ml, ~3.9×10^7 cells) was centrifuged at 4500 r.p.m. for 20 min (H-6000A rotor, Sorvall). The cell pellet (~1 ml) was resuspended in 6 ml of buffer A plus 0.1 M KCl and 0.5 mM PMSF, and sonicated on ice. The lysate was centrifuged at 5000 r.p.m. for 20 min and full-length r-RNPS1 (~50 kDa) was recovered exclusively in the pellet. The insoluble proteins were dissolved and denatured by brief sonication in 6 ml of buffer B with 0.5 mM PMSF, followed by dialysis twice against 0.5 l of the same buffer. The dialysate (~6.3 ml) was fractionated on a Mono Q column under denaturing conditions as described above. Nearly pure r-RNPS1 eluted around 0.35 M NaCl. Thiopropyl-agarose was employed for further purification to remove trace amounts of contaminants (see above). The final purified fraction was dialyzed against buffer A containing 0.1 M KCl and 0.5 mM PMSF. The dialysate was centrifuged to remove insoluble material, and the supernatant was stored at –70°C in small aliquots. Approximately 0.4 mg of apparently homogeneous r-RNPS1 was obtained from ~3.9×10^7 infected Sf9 cells, compared with ~0.17 mg of purified native RNPS1 from ~6.0×10^{10} HeLa cells.

Preparation of E.coli-expressed recombinant RNPS1

The coding region of RNPS1 was amplified by PCR as above, except that the upstream primer RNPS1-N1 (GAGTCCATATGGATTTATCAG-GAGTGA) was used. The PCR-amplified 937 bp product was digested with *Nde*I and *Bam*HI, and subcloned in *Nde*I- and *Bam*HI-digested pSBETc (Schenk *et al*., 1995), to generate pSBETc-RNPS1, which was transformed into *E.coli* BL21(DE3)/plysS (Novagen). Transformants were grown and induced with IPTG as described (Mayeda and Krainer, 1992).

The washed frozen cells from a 1 l culture were resuspended in 15 ml of buffer A plus 0.1 M KCl and 0.5 mM PMSF, and lysed by sonication. Cell debris was removed by centrifugation and r-RNPS1 was recovered in the supernatant and denatured by dialyzing twice against 0.5 l of buffer B containing 0.5 mM PMSF. After removing precipitated material by centrifugation, the supernatant was fractionated by Mono Q chromatography under denaturing conditions as described above, except that the column was eluted with a 0–0.7 M NaCl linear gradient. The peak fractions containing r-RNPS1 (~42 kDa) eluted around 0.50–0.55 M NaCl. The pooled fractions were dialyzed against buffer A containing 0.1 M KCl, 0.5 mM PMSF and 20% (v/v) glycerol, and were stored at –70°C in small aliquots. Approximately 1.4 mg of nearly homogeneous r-RNPS1 was obtained from a 1 l bacterial culture.

In vitro coupled transcription/translation assays

Bacterial expression plasmids (0.5–2 μg each) pET9c-SF2/ASF (Krainer *et al*., 1991), pET9d-hnRNP A1 (Krainer and Mayeda, 1992) and pSBETc-RNPS1 (see above) were incubated with wheat germ extract, T7 RNA polymerase and 13 μCi of [^{35}S]methionine in 16 μl reactions as described by the manufacturer (Promega).

Preparation of antibodies

To obtain anti-RNPS1 polyclonal antibodies, female New Zealand white rabbits were immunized with purified r-RNPS1 expressed in *E.coli* according to standard procedures carried out by Covance Research Products, Inc. (Denver, PA, USA).

Immunoblotting assays

Aliquots of the HeLa cell nuclear and S100 extracts used for the splicing assays were separated by SDS–PAGE and electroblotted onto nitrocellulose. The membranes were blocked with 5% dry milk in TBST (10 mM Tris–HCl pH 8.0, 150 mM NaCl, 0.05% Tween 20), followed by incubation with rabbit anti-RNPS1 antiserum (1:2000 dilution). After washing four times with TBST, the membranes were incubated with anti-rabbit IgG conjugated to alkaline phosphatase (1:7500; Promega), followed by four washes with TBST. Immunoreactive proteins were visualized with BCIP/NBT reagents (Promega), as described by the manufacturer.

Cell culture and transient transfections

To generate an epitope-tagged mammalian expression plasmid, an RNPS1 (E5.1) cDNA fragment was PCR-amplified as described above, except that the upstream primer was RNPS1-N3 (GAGTCTCTAGAGATTTAT-CAGGAGTGA). The amplified fragment was digested with *Xba*I and *Bam*HI, and subcloned into the corresponding sites of the pCGNHCF$_{FL}$ T7-expression vector (Wilson *et al*., 1995) to generate pCGT7-RNPS1.

HeLa cells were grown on glass coverslips in 35-mm tissue culture dishes in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum at 37°C in 5% CO_2. Transient transfections were performed with 1 μg of plasmid pCGT7-RNPS1 or pCGT7-hnRNP A1 (Cáceres *et al*., 1997) in the presence of 10 μg Lipofectamine (Gibco-BRL), as described by the manufacturer.

Immunofluorescence assays

HeLa cells were fixed and permeabilized at 21 h after transfection and analyzed by indirect immunofluorescence essentially as described (Misteli and Spector, 1996). The primary antibody was anti-T7 tag monoclonal antibody (1:400 dilution; Novagen). The secondary and tertiary antibodies were Texas Red-conjugated goat anti-mouse IgG (1:200; Jackson Laboratories) and fluorescein (FITC)-conjugated anti-SC35 antibody (1:200), which was prepared using a FluorX Ab labeling kit (Amersham).

Immunoprecipitation assays

Immunoprecipitations of *in vitro* splicing reactions were done as described (Hanamura *et al*., 1998). Rabbit anti-RNPS1 antibody was described above. Anti-MalE (mAb105) and anti-SF2/ASF (mAb96) antibodies were described previously (Hanamura *et al*., 1998). Anti-hnRNP A1 (mAb9H10) antibody was a generous gift from G.Dreyfuss.

Sequence analysis of RNPS1 homologues

To assemble RNPS1-related protein sequences, full-length human RNPS1 protein and cDNA sequences (Badolato *et al*., 1995) were used as queries to search the non-redundant protein and DNA databases and various EST databases using the suite of BLAST 2.0 based tools (Altschul *et al*., 1997). EST fragments were extended by further BLAST 2.0 searches. Pairwise alignments of retrieved sequences were done using the FASTA-based global alignment program ALIGN (Pearson, 1990). Multisequence alignment was done using Clustal W (Thompson, 1994) with manual adjustment. Protein motifs were searched by Pfscan against PROSITE (Bairoch *et al*., 1997).

Acknowledgements

We thank Tom Misteli, Paul Mintz and David Spector for reagents and helpful advice on immunofluorescence assays; Michelle Hastings for sharing unpublished data; Shern Chew, Michael Murray and John Eisman for helpful discussions and critical reading of the manuscript; Robin Reed for the β-globin derivative plasmids; Kunio Inoue, Naoyuki Kataoka and Yoshiro Shimura for pSPftz and pSP14-15 plasmids; and Gideon Dreyfuss for anti-hnRNP A1 antibody. A.M., R.K. and A.R.K. were supported by grant CA13106 from the NCI. M.Q.Z. was supported by grant HG01696 from the NIH. J.B. and E.M.G. were supported by the National Health and Medical Research Council of Australia.

References

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Antoniou,M. (1995) Clinical defects in pre-mRNA processing. In Lamond,A.I. (ed.), *Pre-mRNA Processing*. R.G. Landes Company, Austin, TX, pp. 187–200.

Badolato,J., Gardiner,E., Morrison,N. and Eisman,J. (1995) Identification and characterisation of a novel human RNA-binding protein. *Gene*, **166**, 323–327.

Bai,Y., Lee,D., Yu,T. and Chasin,L.A. (1999) Control of 3′ splice site choice *in vivo* by ASF/SF2 and hnRNP A1. *Nucleic Acids Res.*, **27**, 1126–1134.

Bairoch,A., Bucher,P. and Hofmann,K. (1997) The PROSITE database, its status in 1997. *Nucleic Acids Res.*, **25**, 217–221.

Birney,E., Kumar,S. and Krainer,A.R. (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.*, **21**, 5803–5816.

Blencowe,B.J., Issner,R., Kim,J., Mccaw,P. and Sharp,P.A. (1995) New proteins related to the Ser-Arg family of splicing factors. *RNA*, **1**, 852–865.

Blencowe,B.J., Issner,R., Nickerson,J.A. and Sharp,P.A. (1998) A coactivator of pre-mRNA splicing. *Genes Dev.*, **12**, 996–1009.

Burge,C.B., Tuschl,T. and Sharp,P.A. (1999) Splicing of precursors to mRNAs by the spliceosomes. In Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds), *The RNA World*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 525–560.

Burn,T.C., Connors,T.D., Van Raay,T.J., Dackowski,W.R., Millholland, J.M., Klinger,K.W. and Landes,G.M. (1996) Generation of transcriptional map for a 700-kb region surrounding polycystic kidney disease type 1 (PKD1) and tuberous sclerosis type 2 (TSC2) disease genes on human chromosome 16p13.3. *Genome Res.*, **6**, 525–537.

Cáceres,J.F. and Krainer,A.R. (1997) Mammalian pre-mRNA splicing factors. In Krainer,A.R. (ed.), *Eukaryotic mRNA Processing*. IRL Press, Oxford, UK, pp. 174–212.

Cáceres,J.F., Stamm,S., Helfman,D.M. and Krainer,A.R. (1994) Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science*, **265**, 1706–1709.

Cáceres,J.F., Misteli,T., Screaton,G.R., Spector,D.L. and Krainer,A.R. (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J. Cell Biol.*, **138**, 225–238.

Colwill,K., Pawson,T., Andrews,B., Prasad,J., Manley,J.L., Bell,J.C. and Duncan,P.I. (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J.*, **15**, 265–275.

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

Dingwall,C. and Laskey,R.A. (1991) Nuclear targeting sequences—a consensus? *Trends Biochem. Sci.*, **16**, 478–481.

Fu,X.-D., Mayeda,A., Maniatis,T. and Krainer,A.R. (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.

Ge,H. and Manley,J.L. (1990) A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA *in vitro*. *Cell*, **62**, 25–34.

Hanamura,A., Cáceres,J.F., Mayeda,A., Franza,B.R.,Jr and Krainer,A.R. (1998) Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA*, **4**, 430–444.

Hertel,K.J. and Maniatis,T. (1998) The function of multisite splicing enhancers. *Mol. Cell*, **1**, 449–455.

Inoue,K., Hoshijima,K., Sakamoto,H. and Shimura,Y. (1990) Binding of the *Drosophila* sex-lethal gene product to the alternative splice site of transformer primary transcript. *Nature*, **344**, 461–463.

Krainer,A.R., Conway,G.C. and Kozak,D. (1990a) Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev.*, **4**, 1158–1171.

Krainer,A.R., Conway,G.C. and Kozak,D. (1990b) The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal sites. *Cell*, **62**, 35–42.

Krainer,A.R., Mayeda,A., Kozak,D. and Binns,G. (1991) Functional expression of cloned human splicing factor SF2: homology to RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. *Cell*, **66**, 383–394.

Lopez,A.J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.*, **32**, 279–305.

Loyer,P., Trembley,J.H., Lahti,J.M. and Kidd,V.J. (1998) The RNP protein, RNPS1, associates with specific isoforms of the $p34^{cdc2}$-related PITSLRE protein kinase *in vivo*. *J. Cell Sci.*, **111**, 1495–1506.

Mayeda,A. and Krainer,A.R. (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell*, **68**, 365–375.

Mayeda,A. and Krainer,A.R. (1999a) Preparation of HeLa cell nuclear and cytosolic S100 extracts for *in vitro* splicing. *Methods Mol. Biol.*, **118**, 309–314.

Mayeda,A. and Krainer,A.R. (1999b) Mammalian *in vitro* splicing assays. *Methods Mol. Biol.*, **118**, 315–322.

Mayeda,A., Helfman,D.M. and Krainer,A.R. (1993) Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF [published erratum appears in *Mol. Cell. Biol.*, 1993, **13**, 4458]. *Mol. Cell. Biol.*, **13**, 2993–3001.

Mayeda,A., Munroe,S.H., Cáceres,J.F. and Krainer,A.R. (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B group proteins. *EMBO J.*, **13**, 5483–5495.

Mayeda,A., Screaton,G.R., Chandler,S.D., Fu,X-D. and Krainer,A.R. (1999) Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. *Mol. Cell. Biol.*, **19**, 1853–1863.

Misteli,T. and Spector,D.L. (1996) Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors. *Mol. Biol. Cell*, **7**, 1559–1572.

Misteli,T. and Spector,D.L. (1998) The cellular organization of gene expression. *Curr. Opin. Cell Biol.*, **10**, 323–331.

Pearson,W.R. (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.*, **183**, 63–98.

Reed,R. and Maniatis,T. (1986) A role for exon sequences and splice-site proximity in splice-site selection. *Cell*, **46**, 681–690.

Reed,R. and Palanjian,L. (1997) Spliceosome assembly. In Krainer,A.R. (ed.), *Eukaryotic mRNA Processing*. IRL Press, Oxford, UK, pp. 103–129.

Rio,D.C. (1988) Accurate and efficient pre-mRNA splicing in *Drosophila* cell-free extracts. *Proc. Natl Acad. Sci. USA*, **85**, 2904–2908.

Sawa,H., Ohno,M., Sakamoto,H. and Shimura,Y. (1988) Requirement of ATP in the second step of the pre-mRNA splicing reaction. *Nucleic Acids Res.*, **16**, 3157–3164.

Schenk,P.M., Baumann,S., Mattes,R. and Steinbiss,H.-H. (1995) Improved high-level expression system for eukaryotic genes in *Escherichia coli* using T7 RNA polymerase and rare Arg-tRNAs. *Biotechniques*, **19**, 196–200.

Schmidt,G. and Werner,D. (1993) Sequence of a complete murine cDNA reflecting an S phase-prevalent transcript encoding a protein with two types of nucleic acid binding motifs. *Biochim. Biophys. Acta*, **1216**, 317–320.

Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL_W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.

Wang,R., Kobayashi,R. and Bishop,J.M. (1996) Cellular adherence elicits ligand-independent activation of the Met cell-surface receptor. *Proc. Natl Acad. Sci. USA*, **93**, 8425–8430.

Wang,Y.-C., Selvakumar,M. and Helfman,D. (1997) Alternative pre-mRNA splicing. In Krainer,A.R. (ed.), *Eukaryotic mRNA Processing*. IRL Press, Oxford, UK, pp. 242–279.

Will,C.L. and Lührmann,R. (1997) snRNP structure and function. In Krainer,A.R. (ed.), *Eukaryotic mRNA Processing*. IRL Press, Oxford, UK, pp. 130–173.

Wilson,A.C., Peterson,M.G. and Herr,W. (1995) The HCF repeat is an unusual proteolytic cleavage signal. *Genes Dev.*, **9**, 2445–2458.

Yang,X., Bani,M.R., Lu,S.J., Rowan,S., Ben-David,Y. and Chabot,B. (1994) The A1 and $A1^B$ proteins of heterogeneous nuclear ribonucleoparticles modulate 5′ splice site selection *in vivo*. *Proc. Natl Acad. Sci. USA*, **91**, 6924–6928.

Received April 12, 1999; revised and accepted July 2, 1999