SR proteins: a conserved family of pre-mRNA splicing factors

Alan M. Zahler, William S. Lane,[1] John A. Stolk, and Mark B. Roth[2]

Fred Hutchinson Cancer Research Center, Division of Basic Sciences, Seattle, Washington 98104 USA; [1]Harvard Microchemistry Facility, Cambridge, Massachusetts 02138 USA

We demonstrate that four different proteins from calf thymus are able to restore splicing in the same splicing-deficient extract using several different pre-mRNA substrates. These proteins are members of a conserved family of proteins recognized by a monoclonal antibody that binds to active sites of RNA polymerase II transcription. We purified this family of nuclear phosphoproteins to apparent homogeneity by two salt precipitations. The family, called SR proteins for their serine- and arginine-rich carboxy-terminal domains, consists of at least five different proteins with molecular masses of 20, 30, 40, 55, and 75 kD. Microsequencing revealed that they are related but not identical. In four of the family members a repeated protein sequence that encompasses an RNA recognition motif was observed. We discuss the potential role of this highly conserved, functionally related set of proteins in pre-mRNA splicing.

[*Key Words*: SR proteins; alternative splicing; RNA splicing; splicing factors]

Received January 21, 1992; revised version accepted March 2, 1992.

Studies of mRNAs from many different tissues and developmental stages show that regulation of RNA processing can lead to the expression of multiple proteins from single genes (Smith et al. 1989). Analysis of *Drosophila* development revealed the first four regulators of pre-mRNA splicing (Boggs et al. 1987; Chou et al. 1987; Amrein et al. 1988; Bell et al. 1988; Goralski et al. 1989). More recently, in vitro splicing has been used to biochemically identify pre-mRNA splicing factors from HeLa cells. In addition to several U small nuclear ribonucleoproteins (UsnRNPs) (for review, see Steitz et al. 1988), other factors have been isolated including SC35 (Spector et al. 1991), U2AF (Zamore and Green 1989, 1991), and SF2 (Krainer et al. 1991), which is identical to ASF (Ge et al. 1991). Complementation of a HeLa cell cytoplasmic S-100 extract that is splicing deficient was used to identify and purify SF2 (Krainer and Maniatis 1985; Krainer et al. 1990a). This factor is essential for cleavage of the 5′ splice site, the first step in splicing. In addition, it was found that SF2 influences splice site selection when pre-mRNAs containing alternative 5′ sites are spliced in vitro (Ge and Manley 1990; Krainer et al. 1990b). Structural analysis of SF2 and the genetically defined splicing regulators shows that these proteins contain either an RNA recognition motif (RRM), which includes two conserved sequences, RNP-1 and RNP-2 (Bandziulis et al. 1989), or sequences of consecutive alternating serine and arginine dipeptides, or both.

Previously, we described a monoclonal antibody, mAb104, that recognizes structures known to contain high levels of spliceosomal components, including sites of RNA polymerase II transcription on lampbrush chromosomes and B "snurposomes" in oocyte nuclei, and small granules in interphase nuclei of many vertebrate and invertebrate cell types (Roth et al. 1990). Therefore, we have been investigating the possibility that mAb104 binds proteins involved in pre-mRNA splicing. We found that mAb104 binds a conserved phosphoepitope on a family of electrophoretically separable proteins with molecular masses of 30, 40, 55, and 75 kD (Roth et al. 1991). In this paper we refer to this set of proteins as SR proteins because the characterized members contain sequences of consecutive serine (S) and arginine (R) dipeptides. Analysis of a cDNA encoding the 55-kD family member from *Drosophila*, dSRp55, shows that it is highly related to SF2 and, to a lesser extent, to the genetically defined *Drosophila* splicing regulators. dSRp55 and SF2 both contain amino-terminal RNA-binding domains, and carboxy-terminal serine- and arginine-rich domains. Subsequent functional tests have shown that dSRp55 is able to substitute for SF2 to complement the S-100 splicing-deficient human extract (Mayeda et al. 1992). This result shows that dSRp55 is either the *Drosophila* homolog of SF2 or that humans and *Drosophila* have two proteins that perform the same function. To distinguish between these two possibilities and to characterize further this group of proteins, we analyzed several SR proteins isolated from human HeLa tissue culture cells and calf thymus. In this paper we show that five SR proteins, plus one additional polypeptide of 20 kD, can be purified from all other cell proteins by a two-step precipitation procedure and that these proteins are

[2]Corresponding author.

all highly related in primary sequence. Preparative isolation of SR proteins from calf thymus allows us to demonstrate biochemically that four of the SR proteins can complement the S-100 splicing-deficient extract originally used to identify SF2.

Results

Purification of SR proteins from human cells

The use of magnesium chloride to selectively precipitate a set of proteins was first described by Roth et al. (1991). This property of SR proteins was found by observing that B snurposomes—1- to 3-μm-diam. ribonucleoprotein structures present in amphibian oocyte nuclei—are stabilized by magnesium (Roth et al. 1990). We have since improved this protocol (see Materials and methods) such that all mAb104 immunoreactive proteins (SR proteins) can be purified from a variety of animal cells and tissues, including mouse 10T1/2, mouse C2 myoblasts, human HeLa, *Drosophila* Kc and Schneider tissue culture cells, *Xenopus* liver, and calf thymus.

The striking aspect of this purification is that only two salt precipitations (ammonium sulfate and then magnesium chloride) are required to purify SR proteins from all other cell proteins. Steps in the purification of SR proteins from HeLa cells are outlined in Figure 1. Figure 1A shows a Coomassie-stained gel revealing the complexity of total proteins throughout the purification; Figure 1B is an immunoblot to show the enrichment of mAb104 antigens. After sonication and centrifugation to clear the extract of insoluble material, the complexity of proteins was high (A1). Presumably, SR proteins are rare because only the 30- and 75-kD family members were detectable with mAb104 (B1). These two SR proteins are enriched in the 65–90% ammonium sulfate precipitate, and the 55-kD protein is also apparent (A2/B2). Subsequent precipitation with magnesium chloride (20 mM) results in the selective and quantitative removal of mAb104 antigens from the supernatant (cf. A3 and A4, B3 and B4). Figure 1C shows the end result of the purification and is included to allow a more direct comparison between the Coomassie-stained proteins (C1) and the proteins detected by mAb104 (C2) in the magnesium pellet. There are five prominent polypeptides detectable by Coomassie stain or immunoblotting with mAb104 that have apparent molecular masses of ~20, 30, 40, 55, and 75 kD. In lanes C1 and C2 the 20-kD protein appears as

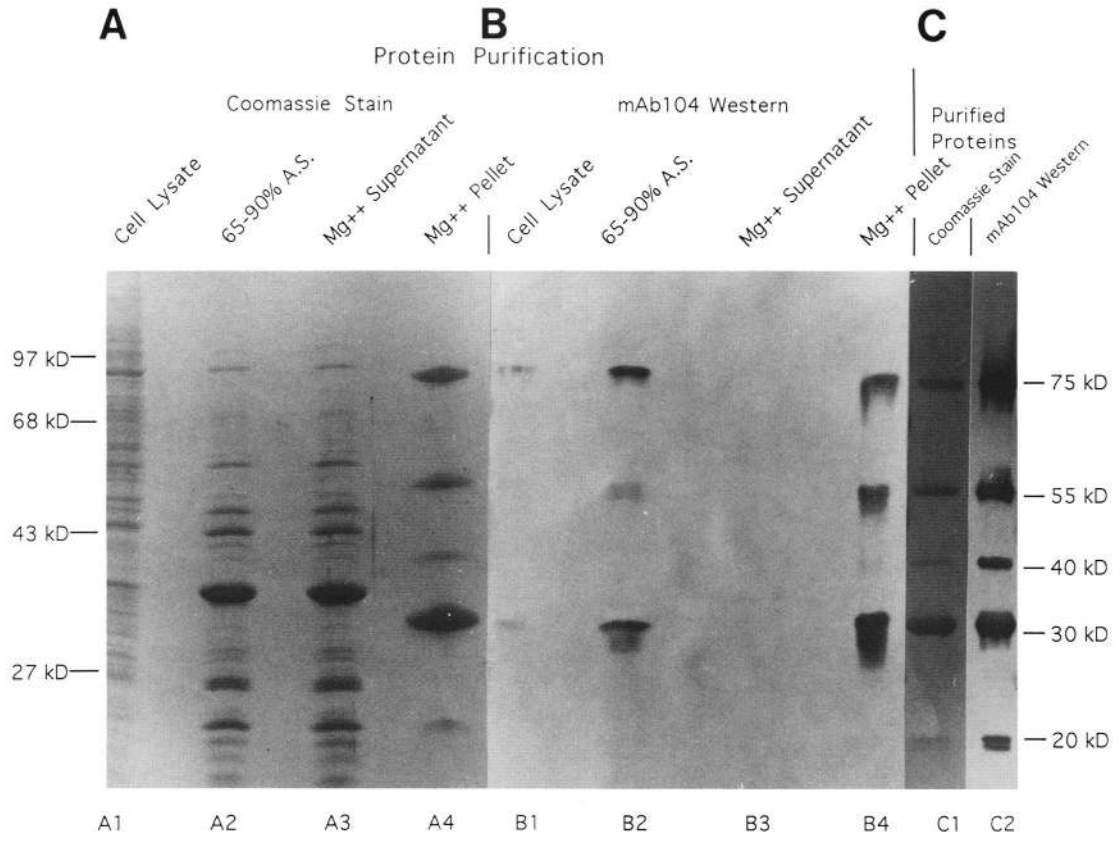

Figure 1. Purification of SR proteins from human HeLa cells. (*A*) Coomassie-stained SDS/PAGE-separated HeLa cell proteins at various steps in the purification of SR proteins (see Materials and methods). (*B*) An immunoblot of the same proteins stained with mAb104. For comparison, Coomassie-stained and mAb104-stained SR proteins are shown in *C*. (Lanes *A1/B1,A2/B2,A3/B3*) Ten micrograms of protein from the cell lysate, the 65–90% ammonium sulfate precipitate, and the soluble proteins after magnesium precipitation, respectively; (lanes *A4,C1*) 10 μg of SR proteins after SDS-PAGE separation and Coomassie staining; (lanes *B4,C2*) 0.5 μg of SR proteins on an immunoblot after staining with mAb104.

two closely spaced bands. Several minor bands or smearing can also be seen immediately above or below the 30-, 55-, and 75-kD proteins. This may be the result of different levels of phosphorylation, degradation of single proteins, or mixtures of different proteins in any one band. Comparison of lanes C1 and C2 also indicates that intensity of staining with mAb104 for any one protein is proportional to the total mass of protein detected by Coomassie staining. Because only two SR proteins are detectable in the whole-cell lysate, it is difficult to measure abundance and percentage recovery of the entire family of proteins in the purification. Given these limits, we estimate that at least 0.04% of cell proteins are SR proteins and that recovery is nearly quantitative.

Sizes of SR proteins are conserved

Previously, we showed that the relative migration of the 30-, 40-, 55-, and 75-kD SR proteins isolated from human HeLa and *Drosophila* Kc cells is identical (Roth et al. 1991). We applied the magnesium precipitation method for purifying SR proteins to many different animal cells and tissues and found that the sizes of SR proteins are conserved. Figure 2 shows a mAb104-stained immunoblot of purified proteins from mouse 10T1/2 cells (lane 1), human HeLa cells (lane 2), and calf thymus (lane 3). All three have immunoreactive proteins at 30, 40, and 55 kD. The 20-kD protein polypeptide was not detectable in calf thymus or HeLa cells because insufficient amounts of protein were electrophoresed in this experiment. The 75-kD polypeptide present in 10T1/2 and HeLa cells runs slightly faster in calf thymus (~70 kD). This may be the result of partial proteolysis; alternatively, this might indicate the presence of a different SR protein.

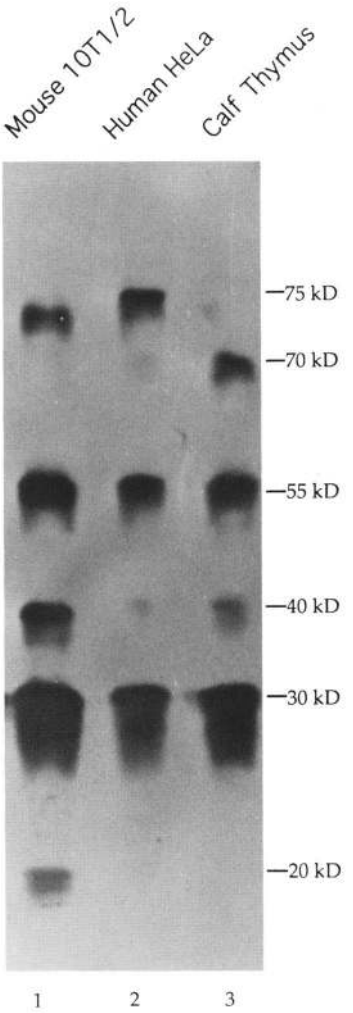

Figure 2. The sizes of SR proteins are highly conserved. (Lanes *1–3*) An mAb104-stained immunoblot of SR proteins isolated from mouse 10T1/2 cells, human HeLa cells, and calf thymus, respectively.

SR proteins are splicing factors

Initially, we found that magnesium-pelleted SR proteins from several sources were able to complement a HeLa cell cytoplasmic S-100 splicing-deficient extract (data not shown) (Dignam et al. 1983; Krainer and Maniatis 1985). We isolated four of the SR proteins from calf thymus (Materials and methods), an inexpensive source, and assayed each individual protein for the ability to complement this splicing-deficient extract. The results of separating preparative quantities of individual SR proteins are shown in Figure 3. A fraction (14 μl) of each purified protein was separated by SDS-PAGE and stained by Coomassie blue (Fig. 3A); less protein (2 μl) was used for an mAb104 immunoblot (Fig. 3B). The 30-kD protein was recovered in the highest abundance followed by the 55-, the 40-, and the 70-kD protein, respectively.

Each of the four purified SR proteins individually complements the cytoplasmic S-100 splicing-deficient extract using three different substrates. The products from in vitro splicing reactions using two of these substrates, *Drosophila fushi tarazu* (*ftz*) pre-mRNA and β-globin pre-mRNA, are shown in Figure 4, A and B, respectively. Both of these transcripts contain single introns with flanking exons and have been characterized previously using HeLa cell extracts (Ruskin et al. 1984; Siebel and Rio 1990). Figure 4, lanes 4–15, shows the products of splicing after incubation of the splicing-deficient extract with individual SR proteins and one of the pre-mRNA substrates. Increasing amounts of the 30-, 40-, 55-, and 70-kD SR proteins (1, 2, or 4 μl of the protein shown in Fig. 3, at a concentration of 0.3 pmole/μl of the 70-kD protein band, 1.4 pmole/μl of the 55-kD protein band, 0.7 pmole/μl of the 40-kD protein band, and 8 pmole/μl of the 30-kD protein band) were added to each 25-μl reaction, and the recovered products are shown in lanes 4–6, 7–9, 10–12, and 13–15, respectively. Note that splicing activity is proportionate to the amount of each SR protein added, indicating that SR proteins are limiting in the reaction. This would explain the slightly reduced levels of splicing by the less abundant 40-, 55-, and 70-kD SR proteins relative to the 30-kD SR protein (cf. Figs. 3 and 4). The amounts of SR proteins added may not be a very accurate quantitation of active protein because each pro-

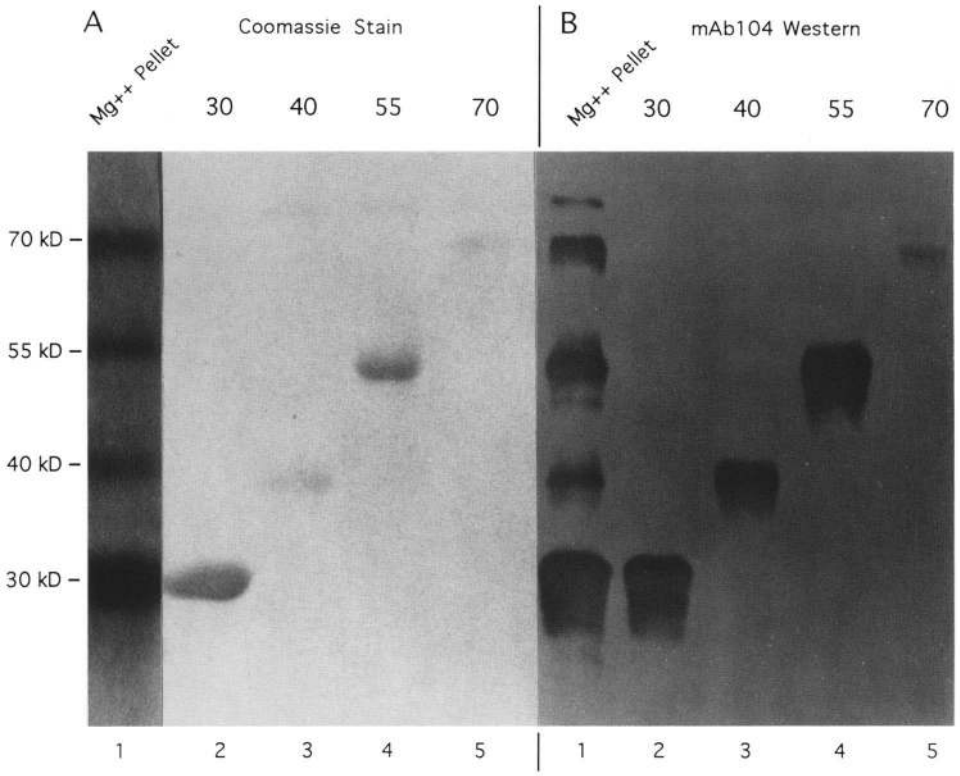

Figure 3. Isolation of individual SR proteins from calf thymus. (*A*,*B*) Coomassie-stained and mAb104-stained SR proteins, respectively. (*A*, lane *1*) The total amount of SR proteins isolated from 5 grams of calf thymus (see Materials and methods). One-tenth this amount was used for the immunoblot (*B*, lane *1*). From 150 grams of tissue, the recovered and renatured 40-, 55-, and 70-kD SR proteins were resuspended in 40 μl of buffer D; the 30-kD SR protein was resuspended in 100 μl. A fraction (14 μl) of the resuspended protein was separated by SDS-PAGE (*A*, lanes *2*–*5*); 2 μl of the same sample was used to detect the proteins by antibody staining (*B*, lanes *2*–*5*).

tein was denatured, precipitated, and renatured prior to testing for activity. The specific activities observed are comparable to the specific activity of *Escherichia coli*-produced SF2 (Krainer et al. 1991), which was treated in a similar way before testing for activity. We also found that the pre-mRNA substrate L/Ad (Bruzik and Steitz 1990) could be spliced in the S-100 extract using any one of the four SR proteins (data not shown). No obvious substrate specificity was observed for any of the SR proteins, although the 30-kD SR protein appeared to splice β-globin pre-mRNA slightly better than *ftz* when compared to the splicing observed for the 40-, 55-, and 75-kD proteins on these two substrates.

Sequence analysis of SR proteins

The discovery that four individual SR proteins are able to complement the same splicing-deficient extract suggests that these proteins are related at the level of primary sequence. To pursue this possibility, we microsequenced all five SR protein bands from human (HeLa) cells. The sequence information derived from this approach is listed in Figure 5A. In several cases, particular peptides identified more than one entry in the data base, all of which are mentioned in the text and in the legend to Figure 5. Thereafter, we use the following nomenclature in an attempt to simplify the description of individual SR proteins. A lowercase letter is used to describe the species of origin, followed by SR, followed by a lowercase p to indicate protein, followed by the SDS-PAGE molecular mass.

Two peptides of 9 and 21 amino acids were sequenced from the 20-kD SR protein. A search of the data base of all existing DNAs (translated) revealed an identical match with both peptides and a predicted translation product from a mouse cDNA of unknown function named X16. The gene encoding X16 mRNA was shown to be actively transcribed in several cell lines and differentially expressed in a variety of mouse tissues (Ayane et al. 1991). Unique polymerase chain reaction (PCR) primers corresponding to the mouse cDNA were made containing the leader and the complement of the sequence immediately after the termination codon. These primers were used to clone a cDNA encoding the human homolog of X16. The DNA sequence showed that this clone is different than mouse X16 at only 15 positions; because these DNA changes do not result in amino acid changes, the two predicted proteins are identical. The predicted protein sequence is shown in Figure 5A and is referred to as hSRp20.

Two peptides sequenced from the 30-kD SR protein band identified multiple entries in the data base, suggest-

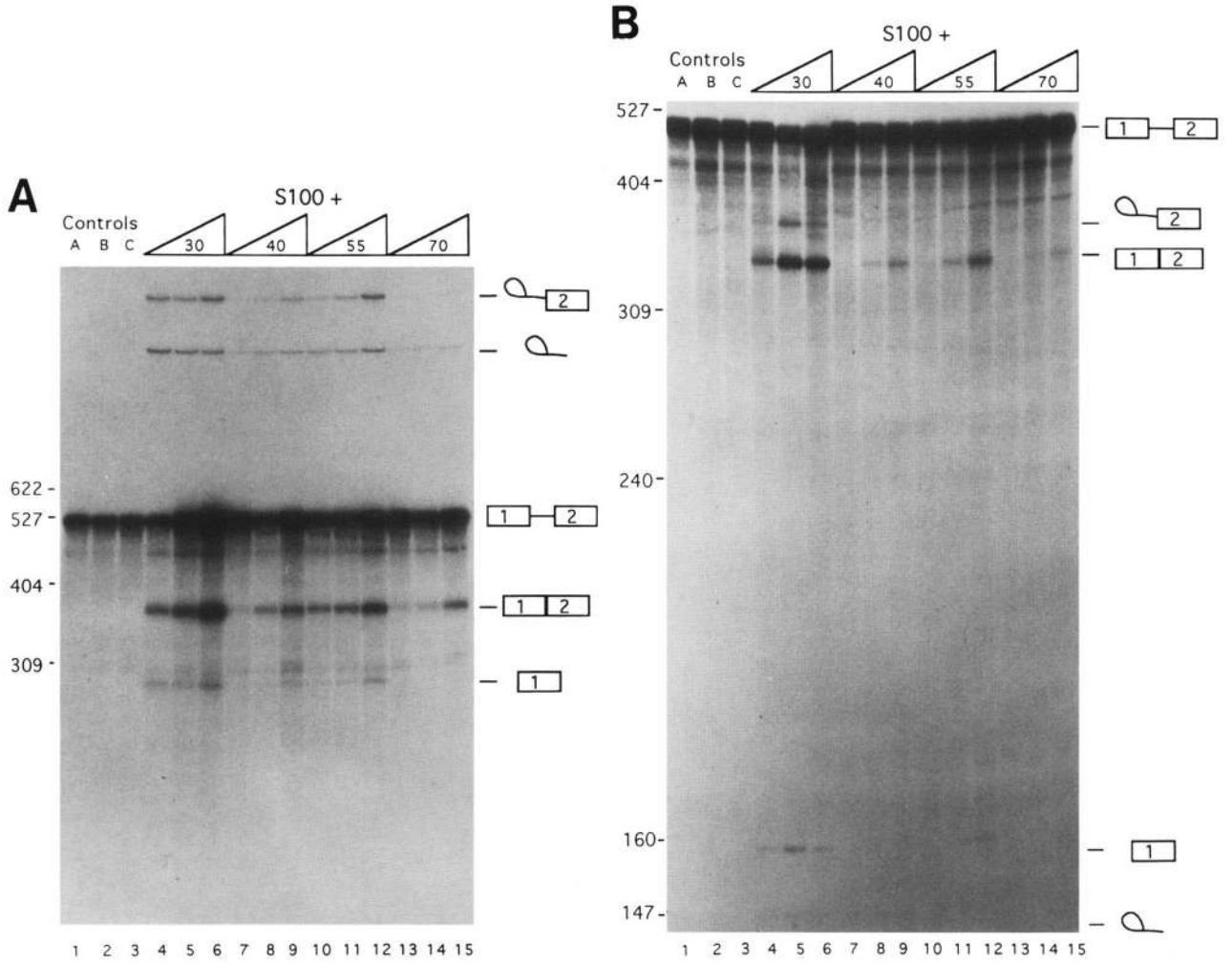

Figure 4. Functional assays of SR proteins. The products of in vitro splicing of *ftz* and β-globin pre-mRNA substrates are shown in *A* and *B*, respectively. All reactions contained S-100 splicing-deficient extract. (*A*,*B*) The first three lanes are control experiments; (lane *1*) 4 μl of buffer; (lane *2*) 4 μl of buffer derived from a mock protein purification and renaturation from acrylamide that did not contain any SR protein; (lane *3*) 100 ng of BSA that was used as a carrier in the SR protein elution and renaturation. (Lanes *4–6*,*7–9*,*10–12*, *13–15*) From reactions incubated with increasing amounts (1, 2, and 4 μl) of the 30-, 40-, 55-, and 70-kD SR proteins from calf thymus, respectively. Splicing products are diagramed at the positions they are predicted to migrate according to reports published previously (see Materials and methods). Note that there is a gel-drying artifact band that appears just below the substrate RNA in many of the lanes in *A* and just above the position of the lariat exon 2 band in lanes *10–15* of *B*.

ing that this band contained more than one protein. The first, a 19-amino-acid peptide in which 17 residues were identified, is found in human SF2 (Krainer et al. 1991), also known as ASF (Ge et al. 1991). This protein is referred to here as hSRp30a (Fig. 5A). A second peptide, 17 amino acids in length, is identical to the predicted translation product of PR264 (Vellard et al. 1992; GenBank accession no. X62447) and SC35 (Fu and Maniatis 1992). A portion of the PR264 and SC35 cDNA sequences corresponds to the antisense strand of the E_T exon of the c-*myb* proto-oncogene (Perbal and Vellard 1990; Vellard et al. 1991; Fu and Maniatis 1992). SC35 has been shown to be a spliceosome assembly factor that is required for pre-mRNA splicing (Fu and Maniatis 1990; Spector et al. 1991; Fu and Maniatis 1992). This protein is referred to here as hSRp30b (Fig. 5A). Comparison of the molar amount of each residue recovered from the first 10 Edman cycles of these two peptides indicates that the tryptic fragment sequenced from hSRp30b is ~5.8-fold more abundant than the tryptic fragment sequenced from hSRp30a. This may reflect a difference in the relative abundance of these two proteins in HeLa cells.

Peptides of varying lengths were sequenced from the 40-, 55-, and 75-kD proteins. All of these peptides are placed in Figure 5 to show their relationships to the predicted translation products from dSRp55, hSRp30a, hSRp30b, and hSRp20.

Two conclusions can be drawn from these data. First, the primary sequences of the family of SR proteins are highly related. Regions of similarity extend throughout

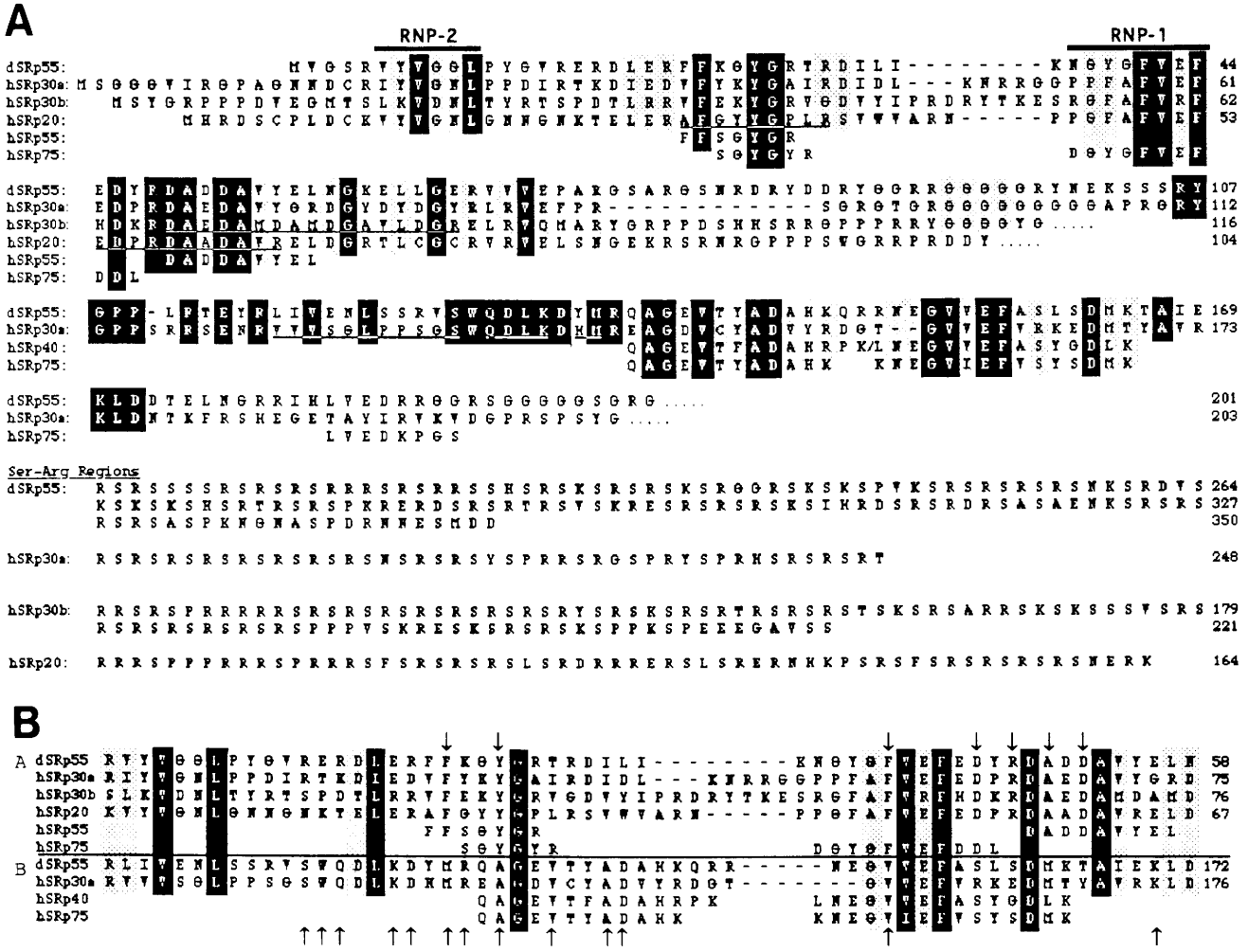

Figure 5. Amino acid sequence analysis of SR proteins. (*A*) Peptide sequences from each of seven different SR proteins. Several sequences identified the predicted translation products from more than one entry in the DNA data base, including dSRp55 (Roth et al. 1991), which is probably encoded by the same gene as B52 (Champlin et al. 1991), is referred to here as dSRp55; SF2 (Krainer et al. 1991), also known as ASF (Ge et al. 1991), is referred to here as hSRp30a; PR264 (Vellard et al. 1991), which is likely to be encoded by the same gene as SC35 (Fu and Maniatis 1992) and is referred to here as hSRp30b; and a mouse protein X16 (Ayane et al. 1991), which is identical to a human protein hSRp20 and is referred to here as hSRp20. In these cases, the full protein sequence is listed and the peptide used to identify it appears as an underlined sequence. Also included are several short peptide sequences from hSRp40, hSRp55, and hSRp75. The sequences used to identify the RNA recognition motif (RNP-1 and RNP-2) are indicated. All sequences are aligned to show identical amino acids (solid boxes) and amino acids where three of the four sequences are identical or highly related (stippled boxes). No attempt was made to align SR domains, which are listed separately below. (*B*) Alignment of two copies of a repeated sequence in dSRp55 and hSRp30a. The first copy of the repeated sequence (repeat A) is shown above the line and extends from amino acid 5 to 58 of dSRp55. The second copy of the repeat (repeat B) is shown below the line and extends from amino acid 116 to 172 of dSRp55. Partial amino acid sequences of hSRp40, hSRp55, and hSRp75 are also listed. Only repeat A is found in hSRp30b and hSRp20. The two repeats are aligned to show identical amino acids (solid boxes) and amino acids where 80% or more are identical or conservative changes (stippled boxes). Arrows indicate amino acids that are conserved in SR proteins but are specific to either repeat A or B.

the amino-terminal domain and are evident particularly in RRM, as well as sequences between the RRM and the S/R domains. Second, comparison of the *Drosophila* and human sequences shows that the primary sequences of SR proteins of the same size are highly conserved between vertebrates and invertebrates. Only 1 of the 16 sequenced amino acids from hSRp55 is different from dSRp55, whereas there are 7 differences when hSRp55 is compared with hSRp30a, 8 differences when hSRp55 is compared with hSRp30b, and 6 differences when hSRp55 is compared with hSRp20 over the same region.

Previously published analysis of dSRp55 (Champlin et al. 1991; Roth et al. 1991) and hSRp30a (Ge et al. 1991; Krainer et al. 1991) reported that these two SR proteins have a single RRM composed of one copy of RNP-2, followed by one copy of RNP-1 (Bandziulis et al. 1989). Our analysis of the family of SR protein primary sequences reveals that there is a direct repeat of RRM-related se-

quences in these two proteins (Fig. 5B). We refer to these two repeats as repeat A, which extends from amino acid 4 to 58 of dSRp55, and repeat B, which extends from amino acid 116 to 172 of dSRp55. Partial amino acid sequencing suggests that repeats A and B are also present in hSRp40 and hSRp75. Both repeats consist of 16 residues that are identical or well conserved among SR proteins. The spacing of the residues that define the repeats is conserved with one exception; the number of amino acids between the two glycines varies from 11 to 19 residues in repeat A and 12 to 14 residues in repeat B. Repeat B is not recognized as an RRM using a motif-searching program (Henikoff et al. 1990; Henikoff 1991; Wallace and Henikoff 1992). This suggests that if the second repeat is an RRM, it is quite different from other known RRMs. Repeats A and B almost certainly perform distinct functions because there are many amino acids that are found in all A repeats but are not found in B repeats and vice versa (see arrows in Fig. 5B). Direct biochemical tests will be required to determine whether either repeat is active for RNA binding.

Discussion

In this paper we characterize the structure and function of a family of splicing factors—SR proteins. We refer to these proteins as a family on the basis of the following criteria: (1) SR proteins are recognized by a monoclonal antibody (mAb104) specific for a shared phosphoepitope (the only possible exception is hSRp30b, which remains to be tested); (2) SR proteins copurify to apparent homogeneity in a two-step precipitation procedure; (3) SR proteins are closely related in primary protein sequence; (4) the apparent sizes of SR proteins on SDS-PAGE are conserved in the animal kingdom; and (5) four different SR proteins isolated from calf thymus can each complement a splicing-deficient cytoplasmic S-100 extract using three different pre-mRNA substrates. SR proteins appear to be distinct from snRNPs because mAb104, which precipitates SR proteins, cannot be used to coprecipitate snRNAs (Roth et al. 1991). It is reasonable to propose that SR proteins are a specific subset of proteins that interact with heterogeneous nuclear RNA (hnRNA). However, none of the characterized hnRNA-binding proteins share significant amino acid homology with SR proteins (Dreyfuss et al. 1988).

Analysis of the family of SR proteins began with the identifcation of a monoclonal antibody (mAb104) that binds both the lateral loops of amphibian lampbrush chromosomes and small granules in nuclei of all animal cells tested (Roth et al. 1990). Further analysis showed that the epitope involved a phosphate residue that we now know is a characteristic of SR proteins. For dSRp55, we found that serine was the only phosphorylated amino acid and that partial hydrolysis released substantial amounts of a small phosphorylated peptide. On the basis of its chromatographic properties, we suspect that this is a dipeptide containing phosphoserine and arginine. We also suspect that multiple epitopes recognized by mAb104 may be located in the carboxy-terminal SR domain. We predict that SRp40 and SRp75 will have SR domains because the tryptic reverse-phase high-performance liquid chromatography (HPLC) peptide maps of hSRp55, hSRp40, and hSRp75 show many peaks with short retention time similar to dSRp55. Furthermore, like dSRp55, these proteins show dramatic shifts in SDS-PAGE mobility after phosphatase treatment, which indicates multiple sites of phosphorylation (K. Neugebauer and M.B. Roth, unpubl.). The fact that mAb104 stains cells in mitosis more intensely than cells in interphase suggests that SR proteins may be hyperphosphorylated during mitosis, which could repress splicing by impeding an interaction of the basic SR domain with the acidic pre-mRNA. Alternatively, other regulated events during mitosis may change the conformation of SR proteins such that the mAb104 epitope is more accessible to the antibody. Further work will be required to distinguish between these possibilities and to determine their relationship to splicing activity.

Magnesium-dependent aggregation is the critical step that allows for extremely rapid purification of this set of splicing factors. A possible explanation for this interaction is that magnesium mediates ionic cross-linking between phosphoserines on adjacent SR proteins. Several lines of evidence suggest that this interaction may be important for the formation of supramolecular complexes of SR proteins in vivo. First, mAb104 stains large magnesium-stabilized complexes (B snurposomes) in the nuclei of amphibian oocytes (Roth et al. 1990). Second, the splicing-deficient extract used to demonstrate that SR proteins are splicing factors is prepared by centrifugation in the presence of magnesium using conditions that sediment SR protein aggregates. Finally, the fusion of serine- and arginine-rich domains from either of two *Drosophila* splicing regulators to β-galactosidase causes localization of the fusion protein to nuclear aggregates similar to those stainable by mAb104—referred to as the speckled network (Li and Bingham 1991). Because SR proteins are splicing factors, it raises the possibilty that the supramolecular complexes of SR proteins seen in cells may be active for splicing. B snurposomes (Wu et al. 1991) and the speckled network (Lerner et al. 1981; Reuter et al. 1984; Nyman et al. 1986; Spector and Smith 1986; Leser et al. 1989; Spector 1990) have been shown to contain many pre-mRNA splicing components. Although it is possible that these structures are active in splicing, direct observation of splicing has only been made for nascent transcripts that are still attached to the DNA template (Osheim et al. 1985; Beyer and Osheim 1988).

The structure of SR proteins appears to consist of two or three domains: one or two copies of an amino-terminal repeat spaced by a series of glycine residues that may form a flexible hinge, followed by a carboxy-terminal highly charged SR domain. dSRp55 and hSRp30a contain two repeats; hSRp20 and hSRp30b contain only one. Repeated specific amino acids shared by several SR proteins make it possible to distinguish between repeats A and B and show that hSRp20 and hSRp30b have only repeat A. The absence of repeat B in these two SR proteins sug-

gests that their function is quite different from the other four characterized SR proteins. Sequence analysis shows that repeat A contains an RRM as defined by homology with other RRM-containing proteins (Henikoff 1991). Using the same analysis, repeat B is not recognized as containing an RRM; however, it is conceivable that repeat B binds RNA but has diverged from the existing consensus. We propose that the two repeats arose by tandem duplication, diverged, and now perform distinct functions that are conserved in SRp30a, SRp40, SRp55, and SRp75. Variation between SR family members, other than the number of repeat domains, is evident throughout the proteins. Particularly divergent sequences are evident between repeat A and the series of glycine residues and between repeat B and the SR domain. Other differences include the length of the SR domain and the relatively small interruptions of this domain by nonbasic or serine residues. We suppose that the amino acid sequences that are shared between all family members are required for function and that the specificity of particular SR proteins is encoded by the variable domains. Strict conservation of the size of SR proteins in cells and tissues from many different animals suggests that when the primary protein sequences of SR proteins are compared between many organisms, a high degree of conservation will be revealed. At this time, it is only possible to compare a few SR protein sequences between species: the complete sequences of mouse (Ayane et al. 1991) and human SRp20, which show no amino acid changes; the complete sequences of human and chicken SRp30b, which show four amino acid changes; and 16 amino acids of human and *Drosophila* SRp55, in which one difference was observed. SR proteins are related in structure to genetically defined splicing regulators isolated from *Drosophila*, *suppressor-of-white-apricot*, *tra*, *tra-2*, and *sex-lethal*, which all contain RRMs, SR domains, or both.

All of the animal cells that we have examined appear to express at least six SR proteins. Consistent with the high level of amino acid sequence conservation among SR proteins from different species, the apparent sizes of SR proteins from vertebrates and invertebrates are virtually identical. Using the two-step purification described, we have detected SR proteins with molecular masses of ~20, 30, 40, 55, and 75 kD from mouse 10T1/2, mouse C2 myoblasts, human HeLa, and *Drosophila* Kc and Schneider tissue culture cells. The same proteins have been detected in *Caenorhabditis elegans*, *Xenopus* liver, and calf thymus. Our analysis of SR proteins suggests that the 20-, 40-, 55-, and 75-kD proteins from human HeLa cells are composed of distinct polypeptides and that the 30-kD band is composed of two different proteins. This is based on partial amino acid sequencing, which shows that the 30-kD protein band has sequences identical to two different proteins. Other results suggest that animals can express more than six SR proteins. We have seen mAb104 immunoreactive proteins of ~180 kD in *Drosophila* Kc tissue culture cells and several proteins that migrate between 100 and 150 kD in *Xenopus* oocytes.

Our studies show that at least four different proteins can complement the same S-100 splicing-deficient extract. On the basis of amino acid sequence analysis of human SR proteins and conservation of sizes of SR proteins between humans and cows, it is possible that the calf thymus 40-, 55-, and 70-kD bands contain single proteins and that the 30-kD protein band contains both SRp30a and SRp30b. These are two previously characterized splicing factors (Ge et al. 1991; Krainer et al. 1991; Spector et al. 1991; Fu and Maniatis 1992). SRp30a was the first protein shown to complement the S-100 splicing-deficient extract; therefore, SRp30b, if present, need not be active in this assay.

Why is there a family of structurally and functionally related splicing factors in animal cells? Sequence similarity between different family members in human cultured cells, as well as the degree of conservation of the family between vertebrates and invertebrates, suggests that specific SR proteins have distinct and essential roles in pre-mRNA splicing. We propose that SR proteins have different specificities for subclasses of pre-mRNAs and that regulation of the levels of SR proteins in different cell types contributes to the regulation of cell-specific splice choices. One prediction from this model is that the levels of specific SR proteins are regulated. The level of mRNA encoded by the mSRp20 gene was shown to be regulated in different tissues. The highest levels of mSRp20 mRNA were found in thymus, whereas lower levels were detected in testis, brain, spleen, and heart. In liver and kidney, mSRp20 mRNA was undetectable when normalized to a constitutive mRNA (Ayane et al. 1991). In addition, we have detected variation in the relative amounts of particular SR proteins in different calf tissues (K. Neugebauer and M. Roth, unpubl.). This model also predicts that changing the level of an SR protein changes splice site selection. In vitro splicing studies of SR proteins using several different substrates that have multiple 5′ splice sites support this idea. Increasing concentrations of specific SR proteins cause a shift from usage of distal to proximal 5′ splice sites for several different substrates (Ge et al. 1991; Krainer et al. 1991; Mayeda et al. 1992).

Materials and methods

Purification and microsequencing of SR proteins

Protein purifications generally started with 1×10^{10} HeLa cells or 100 grams of calf thymus. Cells and tissues were ground to a fine powder in liquid N_2 using a mortar and pestle. The powder was then tranferred to a beaker on ice, and 350 ml of isolation buffer was added (isolation buffer is 65 mM KCl, 15 mM NaCl, 10 mM HEPES at pH 7.6, 10 mM Na_2EDTA, 5 mM DTT, 5 mM potassium fluoride (KF), 5 mM β-glycerophosphate, 0.2 mM PMSF, and 2 μg/ml of aprotinin). Tissue culture cells were sonicated 10 times for 20 sec with a probe sonicator set at 50 W; thymus was sonicated continuously for 20 min. Homogenates were centrifuged for 20 min at 8000*g*, supernatants were transferred to a beaker, and ammonium sulfate was added to 65% of saturation. After 2 hr of stirring at 4°C, the extract was centrifuged at 8000*g* for 20 min. Supernatants were transferred to clean tubes and centrifuged again at 8000*g* for 20 min. Super-

natants were collected, and ammonium sulfate was added to 90% of saturation. Extracts were stirred at 4°C for 0.5–12 hr, followed by centrifugation at 25,000 rpm (85,000*g*) in an SW28 rotor at 4°C for 1 hr. Supernatants were removed, and pellets were rinsed with 90% ammonium sulfate in isolation buffer. Pellets were resuspended in 10 ml of dialysis buffer (65 mM KCl, 15 mM NaCl, 10 mM HEPES at pH 7.6, 1 mM Na_2EDTA, 2 mM DTT, 5 mM KF, 5 mM β-glycerophosphate, and 0.2 mM PMSF) and dialyzed against three changes of 1.4 liter of dialysis buffer over the course of 16 hr. The dialysate was recovered and stored in 1.0-ml aliquots at −80°C.

Dialysate aliquots were thawed and centrifuged for 15 min at 13,000*g*. Supernatants were transferred to clean tubes, and $MgCl_2$ was added to 20 mM. After a 1-hr incubation on ice, tubes were centrifuged at 13,000*g* for 30 min. After removal of the supernatants, the pellets were washed with 200 μl of 20 mM $MgCl_2$ dialysis buffer and resuspended in 20 μl of 5% glycerol buffer D (Dignam et al. 1983). Resuspended pellets containing the purified SR proteins were tranferred to clean tubes.

Immunoblotting with mAB104 was performed according to methods published previously (Roth et al. 1991).

Trypsin digestion, HPLC separation, and microsequencing of SR proteins

Purified SR proteins were separated by SDS-PAGE and electrotransferred to nitrocellulose membrane. Individual SR protein bands were excised and submitted to in situ digestion with trypsin (Aebersold et al. 1987), omitting the NaOH wash. The resulting peptide mixture was separated by narrow-bore HPLC chromatography using a Vydac C18 2.1 × 150-mm reverse-phase column on a Hewlett-Packard 1090 HPLC/1040 diode array detector. Optimum fractions from each peptide chromatogram were chosen on the basis of differential UV absorbance at 210, 277, and 292 nm, peak symmetry, resolution, and column retention time. Automated Edman degradation was performed on an Applied Biosystems 477A protein sequencer by standard methods, except that the reaction cartridge temperature was raised to 53°C during coupling with a commensurate decrease in R2 delivery and dry-down times. Details of strategies for the selection of peptide fractions and their microsequencing have been described (Lane et al. 1991).

Isolation of a cDNA encoding hSRp20

Oligonucleotides corresponding to the sequence immediately before the start codon of X16 and the reverse complement of the sequence immediately after the X16 stop codon were used to prime the synthesis of hSRp20 cDNAs from a mixture of cDNAs prepared from total HeLa cell RNA (Rupp and Wentraub 1991). The cDNAs were separated by agarose gel electrophoresis and identified by ethidium staining. The 530-bp cDNA was inserted into BS/KS+, and both strands were sequenced using the dideoxy method (Sanger et al. 1977).

Separation of purified of SR proteins

$MgCl_2$ pellets of 65–90% ammonium sulfate extracts from 150 grams of calf thymus were resuspended in 5% glycerol buffer D (Dignam et al. 1983) as described above. An equal volume of sample buffer (4% SDS, 125 mM Tris-Cl at pH 6.8, 20% glycerol, 10% β-mercaptoethanol) was added, and the proteins were separated by SDS–10% PAGE. Proteins were eluted from the gel by the methods of Hager and Burgess (1980), with several modifications. Protein bands were visualized by soaking the gel in 0.25 M KCl at 4°C for 5 min. Bands were cut from the gel with a razor, soaked in dH_2O for 5 min, and ground to a fine powder under liquid N_2 with a mortar and pestle. The powder was then transferred to an Eppendorf tube, along with 500 μl of protein gel elution buffer (0.1% SDS, 50 mM Tris-Cl at pH 8.0, 0.1 mM EDTA, 5 mM DTT, 0.15 M NaCl) and 5 μl of 10 mg/ml acetylated bovine serum albumin (BSA) (New England Biolabs). Proteins were eluted by rocking at room temperature overnight. The acrylamide was removed by centrifugation, and the supernatant was transferred to a clean tube. This was repeated three to five times to ensure that all of the acrylamide fragments were removed. Four volumes of acetone at −20°C were added to the supernatant, and the tubes were incubated on dry ice for 30 min. Samples were centrifuged at 13,000*g* for 20 min, and the supernatant was thoroughly removed without allowing the pellet to dry. Pellets were resuspended immediately in 5 μl of 6.0 M guanidine-HCl in buffer D and incubated at room temperature for 20 min. Buffer D (250 μl) was then added to each tube, and the proteins were renatured at room temperature for 1 hr. $MgCl_2$ was added to a final concentration of 20 mM, and the tubes were incubated on ice for 1 hr. SR protein aggregates were removed by centrifugation at 13,000*g* for 30 min as described above. Sedimented 40-, 55-, and 70-kD SR proteins were resuspended in 40 μl of 5% glycerol buffer D; the 30-kD SR protein was resuspended in 100 μl.

Preparation of ^{32}P-labeled RNA-splicing substrates and HeLa S-100 splicing extracts

Plasmid pGEM-2V61 (Rio 1988) containing the *Drosophila ftz* gene intron was a generous gift from Christian Siebel and Donald Rio. Plasmid SP64HβΔ6 containing a β-globin gene intron (Ruskin et al. 1984) was a generous gift from Adrian Krainer. Capped ^{32}P-labeled RNA transcripts were synthesized using the enzymes T7 RNA polymerase and SP6 RNA polymerase, respectively, in 100-μl reactions containing 500 μM ATP, CTP, GTP, UTP, and 7-methyl G(5′)pppG, and 100 μCi of 3000 Ci/mmole [α-^{32}P]UTP. Transcripts labeled with ^{32}P were purified on 5% acrylamide denaturing gels and eluted prior to addition to in vitro splicing reactions. HeLa cytoplasmic S-100 extracts were prepared according to the method of Dignam et al. (1983), with the exception that dialysis was done against 5% glycerol buffer D (Krainer et al. 1990a). The protein concentration of the final extract was ~10 mg/ml.

Cytoplasmic S-100 complementation in vitro splicing assays

Cytoplasmic S-100 complementation splicing assays were done as described previously (Krainer et al. 1991; Mayeda et al. 1992). Reactions contained 2.5% polyvinylalcohol, 20 mM HEPES at pH 7.6, 4 mM $MgCl_2$, 4 mM ATP, 5 mM phosphocreatine, 0.4 U/μl of RNase inhibitor, 10 μl of HeLa cytoplasmic S-100, and 5000 cpm of ^{32}P-labeled RNA substrate in a total volume of 21 μl. The final 4 μl of the 25-μl reactions contained controls or SR proteins in buffer D. Reactions were incubated at 30°C for 2.5 hr, digested with proteinase K for 30 min at room temperature, phenol- and chloroform-extracted, and ethanol-precipitated. Reactions with the SP64HβΔ6 substrate were run on 5% acrylamide denaturing gels for 3 hr at 45 W. Reactions with the pGEM-2V61 substrate were run on 12% acrylamide denaturing gels for 12 hr at 45 W. Gels were dried onto Whatmann 3MM paper prior to autoradiography.

Acknowledgments

We thank Jim Wallace and Steve Henikoff for computer sequence analysis. We gratefully acknowledge Renée Robinson

for her efforts and expertise in digestion and HPLC separation of peptides. We also thank Christian Siebel, Don Rio, Adrian Krainer, and Joan Steitz for advice and plasmid constructs. In addition, we thank Hal Weitraub for 10T1/2 cells and Andrew Koff for assistance in growing HeLa cells. We are indebted to Laurie Bohor Roth and Karla Neugebauer for their careful reading of the manuscript. A.M.Z. is a Burroughs-Wellcome Fund Fellow of the Life Sciences Research Foundation. This research was supported by grant 42786-02 to M.B.R. from the National Institute for General Medical Sciences.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Aebersold, R.H., J. Leavitt, R.A. Saavedra, L.E. Hood, and S.B.H. Kent. 1987. Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc. Natl. Acad. Sci.* **84:** 6970–6974.

Amrein, H., M. Gorman, and R. Nöthiger. 1988. The sex-determining gene *tra-2* of Drosophila encodes a putative RNA-binding protein. *Cell* **55:** 1025–1035.

Ayane, M., U. Preuss, G. Köhler, and P.J. Nielsen. 1991. A differentially expressed murine RNA encoding a protein with similarities to two types of nucleic acid binding motifs. *Nucleic Acids Res.* **19:** 1273–1278.

Bandziulis, R.J., M.S. Swanson, and G. Dreyfuss. 1989. RNA-binding proteins as developmental regulators. *Genes & Dev.* **3:** 431–437.

Bell, L.R., E.M. Maine, P. Schedl, and T.W. Cline. 1988. *Sex-lethal*, a Drosophila sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* **55:** 1037–1046.

Beyer, A.L. and Y.N. Osheim. 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes & Dev.* **2:** 754–765.

Boggs, R.T., P. Gregor, S. Idriss, J. Belote, and M. McKeown. 1987. Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the *transformer* gene. *Cell* **50:** 739–747.

Bruzik, J.P. and J.A. Steitz. 1990. Spliced leader RNA sequences can substitute for the essential 5′ end of U1 RNA during splicing in a mammalian in vitro system. *Cell* **62:** 889–899.

Champlin, D.T., M. Fräsch, H. Saumweber, and J.T. Lis. 1991. Characterization of a Drosophila protein associated with boundaries of transcriptionally active chromatin. *Genes & Dev.* **5:** 1611–1621.

Chou, T., Z. Zachar, and P.M. Bingham. 1987. Developmental expression of a regulatory gene is programmed at the level of splicing. *EMBO J.* **7:** 4095–4104.

Dignam, J.D., R.M. Lebovitz, and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11:** 1475–1489.

Dreyfuss, G., M.S. Swanson, and S. Piñol-Roma. 1988. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem. Sci.* **13:** 86–91.

Fu, X.-D. and T. Maniatis. 1990. Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343:** 437–441.

———. 1992. Isolation of a cDNA that encodes the mammalian splicing factor SC35. *Science* (in press).

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

Ge, H., P. Zuo, and J.L. Manley. 1991. Primary structure of the human splicing factor ASF reveals similarities with Drosophila regulators. *Cell* **66:** 373–382.

Goralski, T.L., J.E. Edström, and B.S. Baker. 1989. The sex determination locus *transformer-2* of Drosophila encodes a polypeptide with similarity to RNA binding proteins. *Cell* **56:** 1011–1018.

Hager, D.A. and R.R. Burgess. 1980. Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: Results with sigma subunit of Escherichia coli RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. *Anal. Biochem.* **109:** 76–86.

Henikoff, S. 1991. Playing with blocks: Some pitfalls of forcing multiple alignments. *New Biologist* **3:** 1–7.

Henikoff, S., J.C. Wallace, and J.P. Brown. 1990. Finding protein similarities with nucleotide sequence databases. *Methods Enzymol.* **183:** 111–132.

Kenan, D.J., C.C. Query, and J.D. Keene. 1991. RNA recognition: Towards identifying determinants of specificity. *Trends Biochem. Sci.* **16:** 214–220.

Krainer, A.R. and T. Maniatis. 1985. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell* **42:** 725–736.

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

———. 1990b. The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal splice sites. *Cell* **62:** 35–42.

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

Lane, W.S., A. Galat, M.W. Harding, and S.L. Schreiber. 1991. Complete amino acid sequence of the FK506 and rampamycin binding protein, FKBP, isolated from calf thymus. *J. Prot. Chem.* **10:** 151–160.

Lerner, E.A., M.R. Lerner, C.A. Janeway, and J.A. Steitz. 1981. Monoclonal antibodies to nucleic acid containing cellular constituents: Probes for molecular biology and autoimmune disease. *Proc. Natl. Acad. Sci.* **78:** 2737–2741.

Leser, G.P., S. Fakan, and T.E. Martin. 1989. Ultrastructural distribution of ribonucleoprotein complexes during mitosis. snRNP antigens are contained in mitotic granule clusters. *Eur. J. Cell Biol.* **50:** 376–389.

Li, H. and P.M. Bingham. 1991. Arginine/serine-rich domains of the *suv*^{a} and *tra* RNA processing regulators target proteins to a subnuclear compartment implicated in splicing. *Cell* **67:** 335–342.

Mayeda, A., A.M. Zahler, A.R. Krainer, and M.B. Roth. 1992. Two members of a conserved family of nuclear phosphoproteins are involved in general and alternative pre-mRNA splicing. *Proc. Natl. Acad. Sci.* **89:** 1301–1304.

Nyman, U., H. Hallman, G. Hadlaczky, I. Pettersson, G. Sharp, and N.R. Ringertz. 1986. Intranuclear localization of snRNP antigens. *J. Cell. Biol.* **102:** 137–144.

Osheim, Y.N., O.L. Miller Jr., and A.L. Beyer. 1985. RNP particles at splice junction sequences on Drosophila chorion transcripts. *Cell* **43:** 143–151.

Perbal, B. and M. Vellard. 1990. Intermolecular recombination of the c-myb proto-oncogene coding sequences in chicken: Clues for a mechanism. *C.R. Acad. Sci. III* **311:** 467–472.

Reuter, R., B. Appel, P. Bringmann, J. Rinke, and R. Lührmann. 1984. 5′ Terminal caps of snRNAs are reactive with antibodies specific for 2,2,7-trimethylguanosine in whole cells and nuclear matrices. *Exp. Cell Res.* **154:** 548–560.

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

Roth, M.B., C. Murphy, and J.G. Gall. 1990. A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *J. Cell. Biol.* **111:** 2217–2223.

Roth, M.B., A.M. Zahler, and J.A. Stolk. 1991. A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J. Cell. Biol.* **115:** 587–596.

Rupp, R.A. and H. Weintraub. 1991. Ubiquitous MyoD transcription at the midblastula transition precedes induction-dependent MyoD expression in presumptive mesoderm of X. laevis. *Cell* **65:** 927–937.

Ruskin, B., A.R. Krainer, T. Maniatis, and M.R. Green. 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* **38:** 317–331.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74:** 5463–5467.

Siebel, C.W. and D.C. Rio. 1990. Regulated splicing of the Drosophila P transposable element third intron in vitro: Somatic repression. *Science* **248:** 1200–1208.

Smith, C.W.J., J.G. Patton, and B. Nadal-Ginard. 1989. Alternative splicing in the control of gene expression. *Annu. Rev. Genet.* **23:** 527–577.

Spector, D.L. 1990. Higher order nuclear organization: Three-dimensional distribution of small nuclear ribonucleoprotein particles. *Proc. Natl. Acad. Sci.* **87:** 147–151.

Spector, D. and H.C. Smith. 1986. Redistribution of snRNPs during mitosis. *Exp. Cell Res.* **163:** 87–94.

Spector, D.L., X.-D. Fu, and T. Maniatis. 1991. Associations between distinct pre-mRNA splicing components and the cell nucleus. *EMBO J.* **10:** 3467–3481.

Steitz, J.A., D.L. Black, V. Gerke, K.A. Parker, A. Kramer, D. Frendeway, and W. Keller. 1988. UsnRNP assembly and transport. In *Structure and function of major and minor small nuclear ribonucleoproteins* (ed. M.L. Birnstiel), pp. 115–154. Springer, Berlin/New York.

Vellard, M., J. Soret, E. Viegas-Pequignot, F. Galibert, N.V. Cong, B. Dutrillaux, and B. Perbal. 1991. C-myb proto-oncogene: Evidence for intermolecular recombination of coding sequences. *Oncogene* **6:** 505–514.

Wallace, J.C. and S. Henikoff. 1992. PATMAT: A searching and extraction program for sequence, pattern and block queries and databases. *Comput. Appl. Biosci.* (in press).

Wu, Z., C. Murphy, H.G. Callan, and J.G. Gall. 1991. Small nuclear ribonucleoproteins and heterogeneous nuclear ribonucleoproteins in the amphibian germinal vesicle: Loops, spheres, and snurposomes. *J. Cell. Biol.* **113:** 465–483.

Zamore, P.D. and M.R. Green. 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl. Acad. Sci.* **86:** 9243–9247.

———. 1991. Biochemical characterization of U2 snRNP auxiliary factor: An essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J.* **10:** 207–214.

SR proteins: a conserved family of pre-mRNA splicing factors.

A M Zahler, W S Lane, J A Stolk, et al.

Genes Dev. 1992, **6:**
Access the most recent version at doi:10.1101/gad.6.5.837