MOLECULAR AND CELLULAR BIOLOGY, July 1993, p. 4023–4028
0270-7306/93/074023-06$02.00/0
Copyright © 1993, American Society for Microbiology
Vol. 13, No. 7

Human SR Proteins and Isolation of a cDNA Encoding SRp75

ALAN M. ZAHLER, KARLA M. NEUGEBAUER, JOHN A. STOLK, AND MARK B. ROTH*

Division of Basic Sciences, A3-015, Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington, 98104

Received 19 January 1993/Returned for modification 30 March 1993/Accepted 16 April 1993

SR proteins are a family of proteins that have a common epitope recognized by a monoclonal antibody (MAb104) that binds active sites of polymerase II transcription. Four of the SR family members have been shown to restore activity to an otherwise splicing-deficient extract (S100 extract). Here we show that two untested SR proteins, SRp20 and SRp75, can also complement the splicing-deficient extract. We isolated a cDNA encoding SRp75 and found that this protein, like other SR proteins, contains an N-terminal RNA recognition motif (RRM), a glycine-rich region, an internal region homologous to the RRM, and a long (315-amino-acid) C-terminal domain composed predominantly of alternating serine and arginine residues. The apparent molecular mass of dephosphorylated SRp75 is 57 kDa, the size predicted from the cDNA clone. We also detected mobility shifts after dephosphorylating SRp55, SRp40, SRp30a, and SRp30b; the sizes of the shifts are proportional to the length of the SR domain, suggesting that serines in this domain are phosphorylated.

Primary transcripts of protein-coding genes (precursor mRNAs) frequently contain introns that are removed during RNA maturation (6, 9). Many of the macromolecules involved in splicing introns from pre-mRNAs are found in spliceosomes composed of ribonucleoprotein particles (U-snRNPs) that contain the uridylic acid-rich small nuclear RNAs U1, U2, U4, U5, and U6 complexed with several proteins (for a review, see reference 22). Ordered binding of these U-snRNP complexes to the pre-mRNA facilitates the two sequential transesterification reactions of splicing (for a review, see reference 16). Many pre-mRNAs contain multiple introns that in some cases, are alternatively spliced (for a review, see reference 27). It is, therefore, of particular interest to learn how spliceosomes recognize these splice junctions. It follows that the initial step in splice site recognition is a likely determinant of splice site specificity.

Recently, a second group of proteins important for pre-mRNA splicing has been discovered (12, 13, 19, 21, 24, 30). These proteins, called SR proteins, are required for the first step of pre-mRNA splicing, yet they do not appear to be part of the U-snRNPs (24). Initially, SR proteins were identified as a group of at least five proteins (20-, 30-, 40-, 55-, and 75-kDa proteins) recognized by a monoclonal antibody (MAb104) that binds active sites of RNA polymerase II transcription (24). Understanding the structures and functions of SR proteins was accelerated by the development of a simple two-step salt precipitation procedure that allows SR proteins to be purified from all other cellular proteins (24, 30). This procedure made it possible to show that the sizes and numbers of SR proteins have been highly conserved throughout evolution, suggesting that SR proteins may have distinct roles in vivo (24).

Structural analysis of SR proteins for which cDNAs have been isolated suggests that the SR protein family can be divided into two subgroups (30), the first containing SR proteins with only one RNA recognition motif (RRM) (3) and the second containing proteins which have both an RRM and an additional highly homologous sequence, called the RRM-homologous sequence (RRMH). SR proteins have a sequence of alternating serine (S) and arginine (R) residues in the C terminus; it was for this shared characteristic that the family of SR proteins was named (30).

To date, cDNAs for four of the six SR proteins have been isolated (2, 8, 11, 13, 19, 24, 28, 30). In this report, we present the sequence of a cDNA that can encode human SRp75. SRp75 is the largest SR protein and contains an RRM, RRMH, and a very long (315-amino-acid) SR domain. Comparing the SRp75 protein sequence with the sequences of other SR proteins shows that the presence or absence of an RRMH domain and the length of the SR domain account for nearly all of the size differences between SR proteins. Thus, functional differences between SR proteins may be attributable to these two regions.

The function of SR proteins has been studied biochemically; four SR proteins (SRp30a, SRp30b, SRp40, and SRp55) can each complement the same splicing-deficient extract (S100) (12, 13, 19, 21, 30). Here we show that SR proteins are depleted in the S100 splicing-deficient extract compared with splicing-competent nuclear extract. We also show that two SR proteins not previously tested (SRp20 and SRp75) can also complement the splicing-deficient extract.

MATERIALS AND METHODS

Purification and separation of SR proteins from HeLa cells. SR proteins were purified from 10^{10} HeLa cells as previously described (30). The SR proteins were separated from one another and eluted from a preparative sodium dodecyl sulfate (SDS)-polyacrylamide gel. The eluted SR proteins were precipitated in 50% acetone at −20°C for 20 min and centrifuged in an Eppendorf microcentrifuge at −20°C for 15 min. Sedimented SR proteins were resuspended in 100 μl of 6.0 M guanidine HCl, dialyzed against 5% glycerol buffer D (10), reprecipitated with 20 mM $MgCl_2$, and resuspended in 5% glycerol buffer D. The purified SR protein preparations had the following concentrations determined by comparison of anti-SR protein MAb104 immunoblots (24) of these preparations to known SR protein standards: SRp20, 25 μg/ml; a mixture of SRp30a and SRp30b (SRp30a/b), 300 μg/ml; SRp40, 40 μg/ml; SRp55, 100 μg/ml; and SRp75, 150 μg/ml.

Preparation of splicing extracts and in vitro splicing reactions. HeLa nuclear extracts and cytoplasmic S100 extracts were prepared by the protocol of Dignam et al. (10), with the

* Corresponding author.

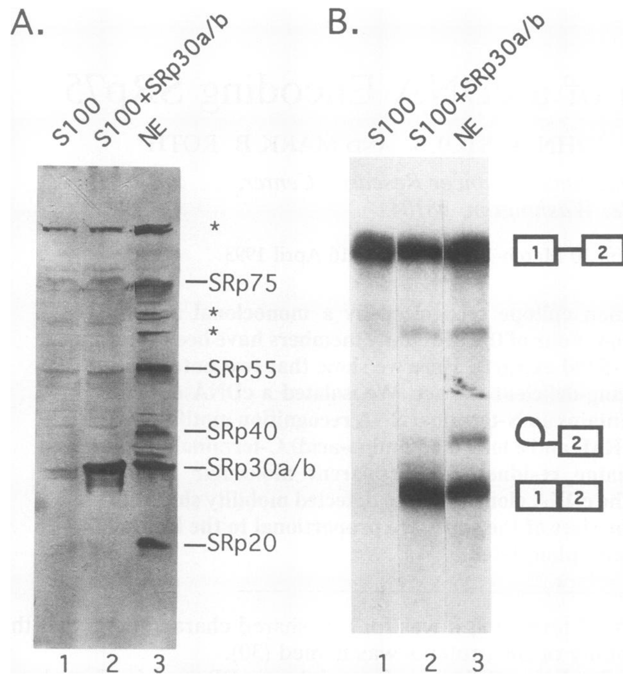

FIG. 1. (A) SR proteins in HeLa cytoplasmic S100 and nuclear extract. Anti-SR protein (MAb104) immunoblot of a SDS–10% polyacrylamide gel. Samples (100 μg) of HeLa cytoplasmic S100 extract alone (lane 1), HeLa cytoplasmic S100 extract to which 450 ng of calf thymus SRp30a/b had been added (lane 2), and HeLa nuclear extract (lane 3) were used. The positions of SR proteins are indicated. Asterisks indicate several of the MAb104-immunoreactive proteins which have not been characterized. (B) In vitro splicing activity of HeLa cytoplasmic S100 and nuclear extract. This panel shows the products of in vitro splicing reactions with the HβΔ6 RNA splicing substrate derived from human β-globin gene (26). The reaction mixtures contain 10 μl (100 μg) of S100 extract (lane 1), 10 μl of S100 extract (100 μg) with 450 ng of calf thymus SRp30a/b (lane 2), and 13 μl (100 μg) of HeLa nuclear extract (lane 4). Spliced products are indicated with exons, introns, and intron lariats shown as boxes, lines, and loops, respectively. NE, nuclear extract.

exception that final extracts were dialyzed against buffer D containing only 5% glycerol. The HeLa nuclear extract used in these experiments had a protein concentration of 7 mg/ml, and the HeLa cytoplasmic S100 extract had a protein concentration of 10 mg/ml.

In vitro splicing reaction conditions, preparation of ^{32}P-labeled HβΔ6 RNA splicing substrate derived from the human β-globin gene (26), and electrophoretic separation conditions have been previously described (30). Reaction mixtures contained 10 μl of cytoplasmic S100 extract or 13 μl of nuclear extract. SR proteins were added in a volume of 4 μl in buffer D.

Isolation of a cDNA encoding SRp75. A cDNA encoding SRp75 was isolated using information from sequencing eight tryptic fragments of purified SRp75 (30). Two degenerate 14-base oligonucleotides, corresponding to short stretches of amino acids on either end of the peptide QAGEVTY ADAHK not including seven central nucleotides, were used in a polymerase chain reaction (PCR) to amplify the DNA sequence encoding this portion of the protein (14). The reaction mixtures contained the following: 0.1 μg of each oligonucleotide (one of which was labeled at the 5′ end with ^{32}P); 1 μg of single-stranded DNA synthesized from HeLa cell RNA using reverse transcriptase and random hexamer primers (25); 50 mM KCl; 10 mM Tris (pH 8.3); 1.5 mM $MgCl_2$; 200 μM (each) dATP, dTTP, dCTP, and dGTP; and 2.5 U of AmpliTaq polymerase (United States Biochemical, Cleveland, Ohio). The total volume of each PCR was 25 μl. After initial denaturation at 94°C for 4 min, the reaction was carried out for 30 cycles, with 1 cycle consisting of at 94°C for 1 min, 50°C for 2 min, and 70°C for 1 min. The PCR products were separated on an 18% denaturing polyacrylamide gel. The expected 35-nucleotide product was excised and used in a second round of amplification. The second-round products yielded sufficient amounts of each strand to use in chemical DNA sequencing (5). A second round of PCR amplification using an oligonucleotide derived from the first amplification and an oligonucleotide derived from another SRp75 peptide sequence yielded a 285-bp DNA fragment that could encode part of SRp75. This 285-bp fragment was used to screen 1.5×10^5 plaques from a human (manca) cell poly(A)$^+$, oligo(dT)-primed cDNA library (gift of Brian McStay). Four positive phages were plaque purified, the inserts were subcloned into Bluescript, and the sizes were determined following *Eco*RI digestion. The largest cDNA (2,070 bp) was sequenced on both strands with Sequenase (United States Biochemical), and several deletions were prepared by exonuclease III and S1 nuclease. Subsequent analysis of the sequence using the DNA Strider program (20) revealed an open reading frame that could encode all eight peptides derived from the intact protein.

Dephosphorylation of HeLa cell SR proteins. Following precipitation in 20 mM $MgCl_2$, SR proteins purified from HeLa cells were resuspended in 0.1 M Tris (pH 8.8)–0.5 mM $MgCl_2$. Approximately 100 μg of SR protein was placed into each of three tubes, and calf intestinal alkaline phosphatase (1 U/μl; Boehringer Mannheim) was added to a final concentration of 0, 0.1, or 0.5 U/μl. Each tube was incubated for 2.5 h at 37°C. For analysis by SDS-polyacrylamide gel electrophoresis (PAGE), 85% of each reaction mixture was visualized by Coomassie blue staining and 15% was subjected to immunoblotting with MAb104 (23).

Nucleotide sequence accession number. The nucleotide sequence for the cDNA encoding SRp75 has been given GenBank accession number L14076.

RESULTS

SR proteins are less abundant in S100 splicing-deficient extract than in splicing-competent nuclear extract. The initial characterization of extracts capable of splicing introns from pre-mRNA substrates led to an effort to fractionate splicing-competent extracts. It was demonstrated that a splicing-deficient S100 cell extract could be complemented by the addition of a nuclear extract that was rendered splicing deficient by treatment with micrococcal nuclease. This result suggested that the nuclear extract contains an essential protein factor not present in the S100 splicing-deficient extract (18). More recently, it has been shown that S100 extract can be complemented for splicing by the addition of any one of several different SR proteins (21, 30). Consistent with this observation, SR proteins have been found to be insoluble in 4.5 mM magnesium chloride (22a), which is present in the extract during centrifugation at 100,000 × *g*.

We used an antibody that binds a conserved phosphorylation-dependent epitope on all SR proteins (MAb104) (23) to detect SR proteins in splicing-competent nuclear extract and splicing-deficient S100 extract. The results indicate that the relative concentration of each of the five most prominent SR

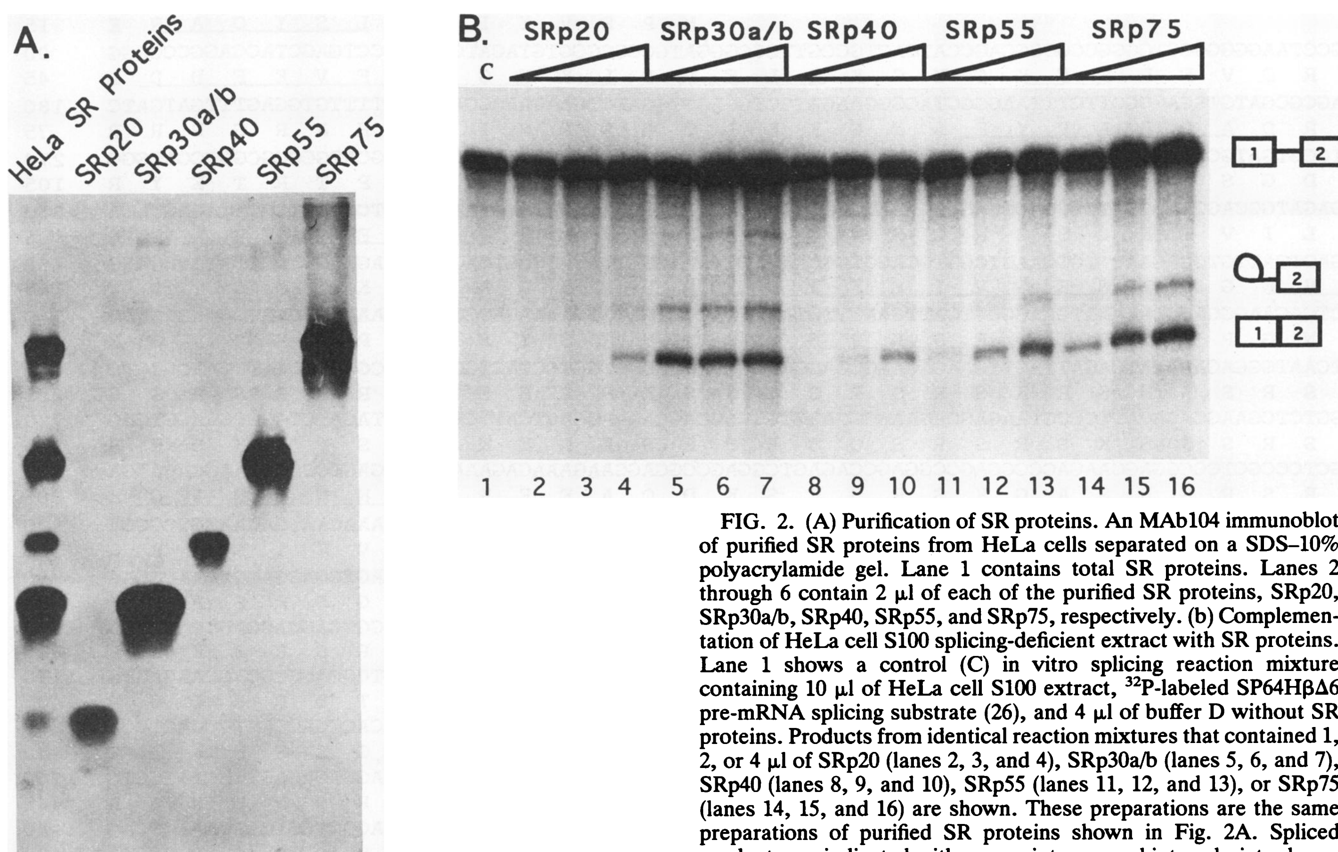

FIG. 2. (A) Purification of SR proteins. An MAb104 immunoblot of purified SR proteins from HeLa cells separated on a SDS–10% polyacrylamide gel. Lane 1 contains total SR proteins. Lanes 2 through 6 contain 2 μl of each of the purified SR proteins, SRp20, SRp30a/b, SRp40, SRp55, and SRp75, respectively. (b) Complementation of HeLa cell S100 splicing-deficient extract with SR proteins. Lane 1 shows a control (C) in vitro splicing reaction mixture containing 10 μl of HeLa cell S100 extract, ^{32}P-labeled SP64HβΔ6 pre-mRNA splicing substrate (26), and 4 μl of buffer D without SR proteins. Products from identical reaction mixtures that contained 1, 2, or 4 μl of SRp20 (lanes 2, 3, and 4), SRp30a/b (lanes 5, 6, and 7), SRp40 (lanes 8, 9, and 10), SRp55 (lanes 11, 12, and 13), or SRp75 (lanes 14, 15, and 16) are shown. These preparations are the same preparations of purified SR proteins shown in Fig. 2A. Spliced products are indicated with exons, introns, and intron lariats shown as boxes, lines, and loops, respectively.

proteins is lower in S100 splicing-deficient extract than in splicing-competent nuclear extract (Fig. 1A, lanes 1 and 3). It is possible that dephosphorylated SR proteins are present in these samples, because MAb104 binds an epitope that requires a phosphorylated antigen. We believe, however, that most of the SR proteins are phosphorylated, because we have not detected dephosphorylated SR proteins in cells when we use antibodies that bind SR proteins independent of phosphorylation (21a).

Previous results have shown that the addition of SR proteins to S100 splicing-deficient extract complements the extract for pre-mRNA splicing (12, 21, 30). The simplest explanation for this complementation is that SR proteins are limiting in the S100 extract. To determine whether this is the case, we added increasing amounts of SR protein (SRp30a/b) to the S100 splicing-deficient extract and compared the amount of SR protein needed to complement the S100 extract with the amount of SR protein in splicing-competent nuclear extract. Our results are consistent with this explanation. We found comparable amounts of spliced mRNA product in either S100 or nuclear extract (Fig. 1B, lanes 2 and 3) when comparable amounts of SR proteins are present (Fig. 1A, lanes 2 and 3).

SRp20 and SRp75 can complement the S100 splicing-deficient extract. It has been shown that any one of several different SR proteins, including SRp30a, SRp30b, SRp40, SRp55, and SRp70, can complement the S100 splicing-deficient extract (12, 13, 19, 21, 30). To determine whether the two other abundant SR proteins present in HeLa cells, SRp20 and SRp75, can also complement S100 splicing-deficient extract for splicing, we purified these two proteins, as well as other SR proteins, from HeLa cells, and separated them by SDS-PAGE (Fig. 2A). Addition of each SR protein to the S100 splicing-deficient extract allowed for splicing of a β-globin pre-mRNA substrate (Fig. 2B). The amount of spliced product appears to be related to the amount of SR protein added, except for SRp30a/b, which was added in amounts that saturate the splicing reaction. These results are consistent with earlier observations that each of the SR protein bands has similar specific activity in this assay (29, 30).

Isolation of a cDNA encoding SRp75. To further characterize this conserved family of pre-mRNA splicing factors, we used information from amino acid sequencing of several tryptic fragments to isolate a cDNA that could encode SRp75. This 2,070-nucleotide cDNA was sequenced (Fig. 3) and found to contain an open reading frame encoding a 57-kDa protein containing all eight tryptic peptide sequences we had obtained from purified SRp75. The difference in the SDS-PAGE-resolved size (75 kDa) and the predicted molecular size (57 kDa) can be explained by the phosphorylation of SRp75 (see below). More than half (52%) of the amino acids in SRp75 are serine or arginine.

Analysis of the predicted amino acid sequence revealed four domains in SRp75. Amino acids 1 through 71 contain an RRM found in other SR proteins as well as many proteins that interact with RNA (3, 17). The RRM contains two short sequences termed RNP-2 and RNP-1. The spacing between RNP-2 and RNP-1 ranges from 26 to 34 amino acids in 32 different RNA-binding proteins; the spacing between these two subdomains in SRp75 is 25 amino acids. Comparison of

```
                                                                    M   P   R   V   Y   I   G   R   L   S   Y   Q   A   R   E     15
GCCTAAGGGGGGCTGGGGCCGGGGCAGCCATCACTGCCGTTGCCGGGATGCCGCGGGTGTACATCGGCCGCCTGAGCTACCAGGCCCGGG     90
  R   D   V   E   R   F   F   K   G   Y   G   K   I   L   E   V   D   L   K   N   G   Y   G   F   V   E   F   D   D   L     45
AGCGCGATGTGGAGCGCTTCTTTAAGGGCTACGGGAAGATCCTGGAGGTGGATCTGAAGAACGGATATGGTTTTGTGGAGTTTGATGATC    180
  R   D   A   D   D   A   V   Y   E   L   N   G   K   D   L   C   G   E   R   V   I   V   E   H   A   R   G   P   R   R     75
TGCGTGATGCAGATGATGCTGTTTATGAACTGAATGGCAAAGACCTTTGTGGTGAGCGAGTAATTGTTGAGCATGCCCGCGGCCCACGGC    270
  D   G   S   Y   G   S   G   R   S   G   Y   G   Y   R   R   S   G   R   D   K   Y   G   P   P   T   R   T   E   Y   R    105
GAGATGGCAGTTACGGTTCTGGACGCAGTGGATATGGTTATAGAAGAAGTGGCCGAGATAAATATGGCCCTCCTACTCGCACAGAGTACA    360
  L   I   V   E   N   L   S   S   R   C   S   W   Q   D   L   K   D   Y   M   R   Q   A   G   E   V   T   Y   A   D   A    135
GACTTATTGTGGAGAATTTGTCAAGTCGGTGCAGCTGGCAAGACCTAAAGGATTATATGCGTCAGGCAGGAGAAGTGACTTATGCAGATG    450
  H   K   G   R   K   N   E   G   V   I   E   F   V   S   Y   S   D   M   K   R   A   L   E   K   L   D   G   T   E   V    165
CTCACAAGGGACGCAAAAATGAAGGGGTGATTGAATTTGTATCTTATTCTGATATGAAAAGAGCTTTGGAAAAGTTGGATGGAACTGAAG    540
  N   G   R   K   I   R   L   V   E   D   K   P   G   S   R   R   R   R   S   Y   S   R   S   R   S   H   S   R   S   R    195
TCAATGGGAGAAAAATCAGATTAGTTGAAGACAAGCCAGGGTCCAGACGACGCCGGTCCTACTCCAGAAGCCGGAGTCATTCAAGGTCTC    630
  S   R   S   R   H   S   R   K   S   R   S   R   S   G   S   S   K   S   S   H   S   K   S   R   S   R   S   R   S   G    225
GGTCTCGAAGCAGACATTCCCGTAAGAGCAGAAGCCGAAGTGGCAGCAGCAAAAGCAGTCATTCTAAGAGTAGATCTCGGTCCAGGTCGG    720
  S   R   S   R   S   K   S   R   S   R   S   Q   S   R   S   R   S   K   K   E   K   S   R   S   P   S   K   D   K   S    255
GCTCCCGCTCCCGGAGCAAGAGCCGGAGCCGGAGCCAGAGTCGGAGCCGGAGCAAGAAAGAGAAAAGCAGGAGCCCCAGCAAGGACAAGA    810
  R   S   R   S   H   S   A   G   K   S   R   S   K   S   K   D   Q   A   E   E   K   I   Q   N   N   D   N   V   G   K    285
GCCGCAGCCGCAGCCATAGCGCTGGCAAGAGCCGCAGCAAGAGCAAAGACCAAGCTGAAGAGAAGATCCAAAACAATGACAATGTCGGGA    900
  P   K   S   R   S   P   S   R   H   K   S   K   S   K   S   R   S   R   S   Q   E   R   R   V   E   E   E   K   R   G    315
AACCCAAGAGCCGGAGTCCTAGCAGGCATAAAAGTAAGAGCAAAAGTCGGAGCAGGAGTCAGGAGAGGAGAGTGGAGGAGGAGAAGCGAG    990
  S   V   E   Q   G   Q   E   Q   E   K   S   L   R   Q   S   R   S   R   S   R   S   K   A   G   S   R   S   R   S   R    345
GGAGTGTTGAGCAGGGGCAGGAGCAGGAGAAGAGCCTCCGCCAGAGTCGGAGCCGGAGCAGGAGCAAAGCGGGCAGCAGGAGCCGGAGCA   1080
  S   R   S   K   S   K   D   K   R   K   S   R   K   R   S   R   E   E   S   R   S   R   S   R   S   R   S   K   S   E    375
GGAGCCGCAGCAAGAGCAAGGACAAGAGGAAGAGCAGGAAGAGAAGCAGAGAGGAGAGCCGCAGTCGCAGTCGCAGCCGCAGCAAGAGTG   1170
  R   S   R   K   R   G   S   K   R   D   S   K   A   G   S   S   K   K   K   K   K   E   D   T   D   R   S   Q   S   R    405
AGAGGAGCAGAAAGCGAGGCAGCAAGCGAGACAGCAAGGCGGGCAGCAGCAAGAAGAAGAAGAAGGAAGACACTGACCGCTCCCAGTCCA   1260
  S   P   S   R   S   V   S   K   E   R   E   H   A   K   S   E   S   S   Q   R   E   G   R   G   E   S   E   N   A   G    435
GATCTCCATCCCGCTCCGTGTCAAAGGAGCGGGAACATGCCAAGTCTGAATCCAGCCAGAGGGAAGGTCGAGGAGAGAGTGAGAATGCTG   1350
  R   N   E   E   T   R   S   R   S   R   S   N   S   K   S   K   P   N   L   P   S   E   S   R   S   R   S   K   S   A    465
GCAGGAATGAGGAGACCCGGTCCAGGTCGAGATCCAATTCCAAATCGAAACCAAACCTTCCATCAGAATCACGCTCCAGATCAAAGTCAG   1440
  S   K   T   R   S   R   S   K   S   R   S   R   S   A   S   R   S   P   S   R   S   R   S   R   S   H   S   R   S   *    494
CTTCAAAAACCCGATCTCGGTCCAAGTCTAGATCCAGGTCTGCTTCCAGATCGCCCTCCCGATCTAGATCTAGGTCCCACTCAAGGTCCT   1530
AACTGGCTATGGCCACAGCTGGAACTACCCGAGAAGTCTTTTGTACATGTTTGGTAGCCGTAGCACAAGTGATTGGAGTAGAACATGTCA   1620
CTGCTGTACATTTTTAACTCCCCTAATGGTGTGTCTATAATTGTTAAATCTAAGTGCTTCCTCTCAGTAAAGCCTCCTGGCACCAGGCCT   1710
TCCTGCTCGACTGAAAAAAATTTTCTCTTTGAAAATCCCCTTTTACTCATGGCCCACAGTAGAATATCCAAAACGCCTTGGCTTTCAGGC   1800
CTGGCCTTTCCTACAGGGAGCTCAGTAACCTGGACGGCTCTAAGGCTGGAATGACCACATAGGTAGGTATGGTGAGTTCAACCATTTTTG   1890
CTCTTGAATTGATGCCCTTCGATGTATGCCATTTAGTGAAAGTGCTAAGTCTTAAGTTTCCTACCACTTTGGTTTCATATTTTTGGACTT   1980
AACAAAGTTGTGAATAGCACAGTCGAGGAAAATTGATACCTGCAGTAACCCATAGGAAATAAACTGTAGAGTTCCATATTCTGCGGCCGC   2070
```

FIG. 3. Nucleotide sequence of the human SR75 cDNA and the predicted amino acid sequence of SRp75. Underlined peptide sequences match those derived from the sequencing of tryptic fragments of purified SRp75.

the SRp75 RRM with the RRM found in other SR proteins shows that SRp75 is most closely related to SRp55 (24). Indeed, the RRM in human SRp75 is somewhat more related to the RRM in *Drosophila melanogaster* SRp55 than it is to the RRM found in human SRp20, SRp30a, and SRp30b.

A second domain within SRp75 extends from amino acids 72 to 94 and contains a large number of glycine residues (30%). Comparison of the glycine-rich region of SRp75 with the glycine-rich region in other SR proteins shows no other obvious homology.

Following the glycine-rich region there is a domain extending from amino acids 95 to 178 that is homologous to the RRM. This RRMH is composed of 20 identical or closely related amino acids that are similarly spaced in the RRM and the RRMH. While the RRMH is not recognized as an RNA-binding domain using motif recognition computer analysis, the RRMH is likely to be important for SR protein function, because it is conserved among four SR proteins, SRp30a, SRp40, SRp55, and SRp75 (30).

The remaining two-thirds of the SRp75 protein contains the SR domain. S and R refer to the single-letter designations of the two most abundant amino acids in SR proteins, serine (S) and arginine (R). The 315 amino acids in the SR domain of SRp75 include 108 serines and 116 basic amino acids (arginines and lysines). Serine and basic amino acids alternate throughout the SR domain; other amino acids with polar or charged R groups are interspersed with no obvious motifs detected. A 17-amino-acid stretch devoid of serine and arginine is present within the SR domain of SRp75 and is not found in any other SR protein.

SR proteins are posttranslationally modified by phosphorylation. To examine the difference between the apparent molecular weight of SRp75 resolved on SDS-PAGE and the size predicted from cDNA, we treated HeLa cell SR proteins with calf intestinal alkaline phosphatase. Figure 4 shows HeLa SR proteins incubated with two different concentrations of phosphatase. Note that the low concentration of phosphatase (Fig. 4A, lane 2) yields partially dephosphorylated SR proteins that migrate between their initial position (lane 1) and their final position (lane 3). A monoclonal antibody (MAb104) which is known to bind a phosphate-dependent epitope on SR proteins (23) recognizes partially dephosphorylated SR proteins (Fig. 4B, lane 2) but does not recognize completely dephosphorylated SR proteins (Fig. 4B, lane 3). We hypothesize that the majority of SR proteins are maximally phosphorylated, because we have not ob-

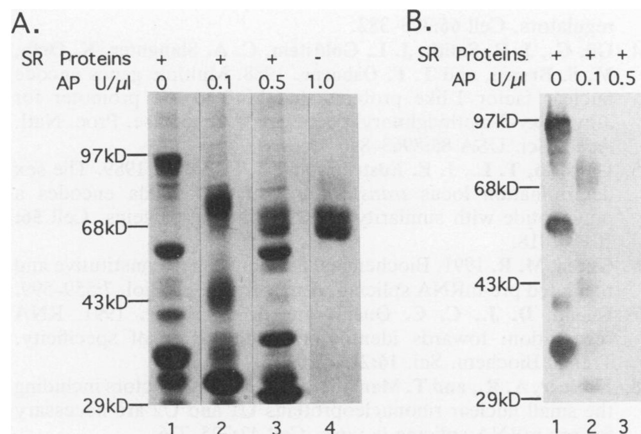

FIG. 4. Dephosphorylation of HeLa cell SR proteins. (A) SDS-PAGE-separated and Coomassie blue-stained SR proteins after treatment with 0 (lane 1), 0.1 (lane 2), and 0.5 (lane 3) U of phosphatase per µl and phosphatase alone (lane 4). (B) MAb104 immunoblot of one-fifth of the samples separated in lanes 1, 2, and 3 of panel A, respectively. Note that MAb104 recognizes partially dephosphorylated SR proteins in lane 2, but not the fully dephosphorylated SR proteins in lane 3.

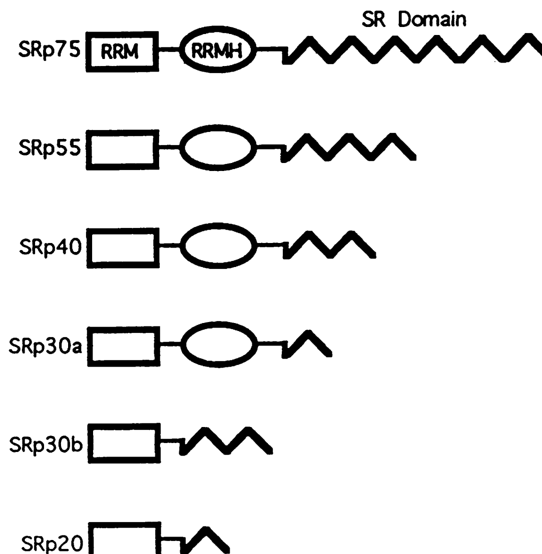

FIG. 5. Schematic diagram of SR proteins. RRM or rectangle indicates the RNA recognition motif; RRMH or oval indicates a sequence in SRp30a, SRp40, SRp55, and SRp75 that is 30% homologous to conserved amino acids in SR protein RRMs; the bold zigzag line indicates the region of the proteins that are composed predominantly of a series of alternating serine (S) and arginine (R) residues, namely, the SR domain; narrow horizontal lines indicate sequences that are variable between different SR proteins. The four motifs are drawn to scale (i.e., the length of the bold zigzag line indicates the relative length of the SR domain in each protein).

served partially or fully dephosphorylated SR proteins in whole-cell extracts, even when antibodies that bind SR proteins independently of phosphorylation are used (21a).

In the gel system used for these studies, phosphorylated SRp75 migrates with an apparent molecular mass of 93 kDa and completely dephosphorylated SRp75 migrates with an apparent molecular mass of 57 kDa which is consistent with the predicted molecular mass from the cDNA sequence. The predicted sizes of other SR proteins for which cDNAs have been isolated, including SRp20, SRp30a, SRp30b, and SRp55, are also similar to those of the dephosphorylated forms of these proteins. The large change in mobility and the lack of bands corresponding to partially dephosphorylated intermediates suggest that SR proteins are maximally phosphorylated at many sites. Immunostaining with MAb104, which binds phosphorylated SR proteins only shows intense staining of mitotic cells relative to interphase cells, indicating a possible cell cycle regulation of SR protein phosphorylation (23).

DISCUSSION

Characterization of a unique group of pre-mRNA splicing factors called SR proteins shows that this family is defined by certain shared characteristics. (i) SR proteins are recognized by a monoclonal antibody (MAb104) which binds a conserved phosphorylation-dependent epitope; (ii) SR proteins copurify to apparent homogeneity in a two-step salt precipitation procedure. (iii) SR proteins are very closely related in amino acid sequence. (iv) The apparent sizes of SR proteins by SDS-PAGE are conserved in the animal kingdom. (v) SR proteins have similar specific activities in complementing an S100 splicing-deficient extract when pre-mRNA substrates that contain single introns are used. Four SR proteins (SRp30a, SRp30b, SRp40, and SRp55) were recently shown to meet each of these criteria (12, 13, 19, 21, 30). The work presented here demonstrates that SRp20 and SRp75, the two most structurally divergent SR proteins, also share these properties with other members of the SR protein family.

In the last 3 years, cDNAs that encode five of the most abundant SR proteins have been isolated; for a sixth SR protein, SRp40, only partial amino acid sequence information is available (30). An additional SR protein identified in calf thymus, SRp70 (30), could be either a breakdown product of SRp75 or a different SR protein. This study and other published work allow for some general statements to be made about SR proteins. First, the SR protein family can be divided into two structurally distinct subgroups (Fig. 5): those that contain an RRM (SRp20 and SRp30b) and those that contain an RRM and an RRMH (SRp30a, SRp40, SRp55, and SRp75). For the proteins that contain an RRMH, a glycine-rich variable region of 18 to 34 amino acids separates the RRM and the RRMH. Second, within each subgroup, the differences in size can be accounted for almost entirely by differences in the length of the SR domain. Strict size conservation between vertebrates and invertebrates implies that functional differences between the SR proteins are likely to be determined by specific variations in these two domains.

The overlapping functions of SR proteins in the complementation of the S100 splicing-deficient extract with pre-mRNAs that contain single introns suggest that SR proteins could have common functions in vivo. On the other hand, the evolutionary conservation of the sizes and number of SR proteins supports the idea that they may have distinct functions in vivo. We have been attempting to study the functions of SR proteins using pre-mRNA substrates that can be alternatively spliced and have found distinct in vitro functions for some of the SR proteins. The results suggest that distinct SR proteins allow for the preferential use of different 5′ splice site when pre-mRNAs that contain multiple 5′ splice sites are used as substrates in the S100 comple-

mentation assay (29). Combined with the striking similarity between SR proteins and genetically defined splicing regulators isolated from *D. melanogaster* (1, 4, 7, 15) (both contain RRM and/or SR domains), these results support the model that SR proteins play a role in alternative splice site selection in vivo.

ACKNOWLEDGMENTS

We thank Brian McStay for providing us with a human cDNA library. We also thank Laurie Bohor Roth, Steve Hahn, and Jim Priess for editorial comments.

A.M.Z. is a Burroughs-Wellcome Fund fellow of the Life Sciences Research Foundation. K.M.N. is a fellow of the American Cancer Society. This work was supported by a grant to M.B.R. from the National Institutes of Health.

REFERENCES

1. **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.
2. **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.
3. **Bandziulis, R. J., M. S. Swanson, and G. Dreyfuss.** 1989. RNA-binding proteins as developmental regulators. Genes Dev. **3:**431–437.
4. **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.
5. **Bencini, D. A., G. A. O'Donovan, and J. R. Wild.** 1984. Rapid chemical degradation sequencing. BioTechniques **2:**4–5.
6. **Berget, S. M., C. Moore, and P. A. Sharp.** 1977. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. USA **74:**3171–3175.
7. **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.
8. **Champlin, D. T., M. Frasch, H. Saumweber, and J. T. Lis.** 1991. Characterization of a Drosophila protein associated with boundaries of transcriptionally active chromatin. Genes Dev. **5:**1611–1621.
9. **Chow, L. T., R. E. Gelinas, T. R. Broker, and R. J. Roberts.** 1977. An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell **12:**1–8.
10. **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.
11. **Fu, X.-D., and T. Maniatis.** 1992. Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. Science **256:**535–538.
12. **Fu, X.-D., A. Mayeda, T. Maniatis, and A. R. Krainer.** 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.
13. **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.
14. **Gil, G., J. R. Smith, J. L. Goldstein, C. A. Slaughter, K. Orth, M. S. Brown, and T. F. Osborne.** 1988. Multiple genes encode nuclear factor 1-like proteins that bind to the promoter for 3-hydroxy-3-methylglutaryl-coenzyme A reductase. Proc. Natl. Acad. Sci. USA **85:**8963–8967.
15. **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.
16. **Green, M. R.** 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu. Rev. Cell Biol. **7:**559–599.
17. **Kenan, D. J., C. C. Query, and J. D. Keene.** 1991. RNA recognition: towards identifying determinants of specificity. Trends Biochem. Sci. **16:**214–220.
18. **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.
19. **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.
20. **Marck, C.** 1988. 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucleic Acids Res. **16:**1829–1836.
21. **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. USA **89:**1301–1304.

21a. **Neugebauer, K. M.** Unpublished data.

22. **Reddy, R., and H. Busch.** 1988. Small nuclear RNAs: RNA sequences, structure, and modifications, p. 1–37. *In* M. L. Birnstiel (ed.), Structure and function of major and minor small nuclear ribonucleoproteins. Springer-Verlag, Berlin.

22a. **Roth, M. B.** Unpublished data.

23. **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.
24. **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.
25. **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.
26. **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.
27. **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.
28. **Vellard, M., A. Sureau, J. Soret, C. Martinerie, and B. Perbal.** 1992. A potential splicing factor is encoded by the opposite strand of the trans-spliced c-myb exon. Proc. Natl. Acad. Sci. USA **89:**2511–2515.
29. **Zahler, A. M., K. M. Neugebauer, J. A. Stolk, and M. B. Roth.** 1993. Alternative pre-messenger RNA splicing by SR proteins. Science **260:**219–222.
30. **Zahler, A. M., J. A. Stolk, W. S. Lane, and M. B. Roth.** 1992. SR proteins: a conserved family of pre-mRNA splicing factors. Genes Dev. **6:**837–847.