The EMBO Journal vol.13 no.11 pp.2639-2649, 1994

Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa factor of the serine/arginine protein family

Yvon Cavaloc[1], Michel Popielarz[1], Jean-Paul Fuchs, Renata Gattoni and James Stévenin[2]

Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Institut de Chimie Biologique, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cedex, France
[1]Equal first authors
[2]Corresponding author

Communicated by R.Lührmann

By adopting a monoclonal antibody approach, we have identified a novel splicing factor of 35 kDa which we have termed 9G8. The isolation and characterization of cDNA clones indicate that 9G8 is a novel member of the serine/arginine (SR) splicing factor family because it includes an N-terminal RNA binding domain (RBD) and a C-terminal SR domain. The RNA binding domain of 9G8 is highly homologous to those of the SRp20 and RBP1 factors (79-71% identity), but the homology is less pronounced in the cases of SF2/ASF and SC35/PR264 (45-37% identity). Compared with the other SR splicing factors, 9G8 presents some specific sequence features because it contains an RRSRSXSX consensus sequence repeated six times in the SR domain, and a CCHC motif in its median region, similar to the zinc knuckle found in the SLU7 splicing factor in yeast. Complete immunodepletion of 9G8 from a nuclear extract, which is accompanied by a substantial depletion of other SR factors, results in a loss of splicing activity. We show that a recombinant 9G8 protein, expressed using a baculovirus vector and excluding other SR factors, rescues the splicing activity of a 9G8-depleted nuclear extract and an S100 cytoplasmic fraction. This indicates that 9G8 plays a crucial role in splicing, similar to that of the other SR splicing factors. This similarity was confirmed by the fact that purified human SC35 also rescues the 9G8-depleted extract. The identification of the 9G8 factor enlarges the essential family of SR splicing factors, whose members have also been proposed to play key roles in alternative splicing.

Key words: factor cloning/*in vitro* complementation/monoclonal antibody/pre-mRNA splicing/SR splicing factor

Introduction

The splicing of nuclear pre-mRNA requires the recognition of each intron by multiple *trans*-acting factors and the formation of splicing complexes, called spliceosomes [reviewed in Green (1991) and Moore *et al.* (1992)]. Among the *trans*-acting factors, the U-small nuclear ribonucleoproteins (U snRNP) U1, U2, U5 and U4/U6 are the best characterized and their involvement in the formation of the spliceosome has been well defined in lower and higher eukaryotes (Guthrie, 1991; McKeown, 1993). As for the non snRNP *trans*-acting protein factors, only limited information is available in higher eukaryotes due to the great complexity of the splicing process. In contrast, in yeast a whole set of such factors, together with snRNP factors, have been identified by applying a genetic approach (Ruby and Abelson, 1991). Because this approach cannot be used in higher eukaryotes, the identification of protein factors has been particularly difficult. With the exception of the serine/arginine (SR) splicing factor family, only two other factors, the U2 auxiliary factor (U2AF) (Zamore *et al.*, 1992) and the pPTB-associated splicing factor (PSF) (Patton *et al.*, 1993), had been cloned and characterized until recently.

Relatively few approaches are available to identify splicing factors in higher eukaryotes. One consists of the biochemical fractionation of splicing activities that are tested in a minimal *in vitro* assay which analyzes an individual step of the splicing process. This method works well mainly for the early steps of the process in which a limited number of factors are involved. In this way U2AF was identified by its interaction with the pyrimidine-rich region of the 3′ splice site, which is a prerequisite for the interactions of U2 snRNP with the branch site (Ruskin *et al.*, 1988). By a similar strategy, several splicing activities required for the formation of the presplicing complex have been identified, but they still contain several polypeptides (Krämer and Utans, 1991; Brosi *et al.*, 1993b). Further analysis of proteins associated with the U2 snRNP (Brosi *et al.*, 1993a) and the presplicing complex (Bennett *et al.*, 1992) led to the identification of the mammalian homologs of the yeast PRP9 (Behrens *et al.*, 1993) and PRP11 factors (Bennett and Reed, 1993). Furthermore, by exploiting a cytoplasmic S100 extract which contains all the splicing machinery except a limited number of factors, a biochemical complementation assay allowed the purification and cloning of a splicing factor called SF2 (Krainer *et al.*, 1990a, 1991). This factor was also isolated and cloned by its ability to alter the selection of alternative 5′ splice sites and was called ASF, for alternative splicing factor (Ge and Manley, 1990; Ge *et al.*, 1991). SF2/ASF represents the first identified member of the SR splicing factor family.

Splicing factors have also been characterized by mAbs directed against components of nuclear substructures which contain the splicing machinery. A protein may be considered as a splicing factor when specific immunodepletion of a nuclear extract and subsequent complementation with the recombinant protein lead to inhibition and restoration of splicing, respectively. This approach has been successfully used to identify and clone another SR factor called SC35 (Fu and Maniatis, 1990, 1992b), which was cloned independently and termed PR264 (Vellard *et al.*, 1991, 1992). A mAb (mAb104) also allowed the characterization

© Oxford University Press

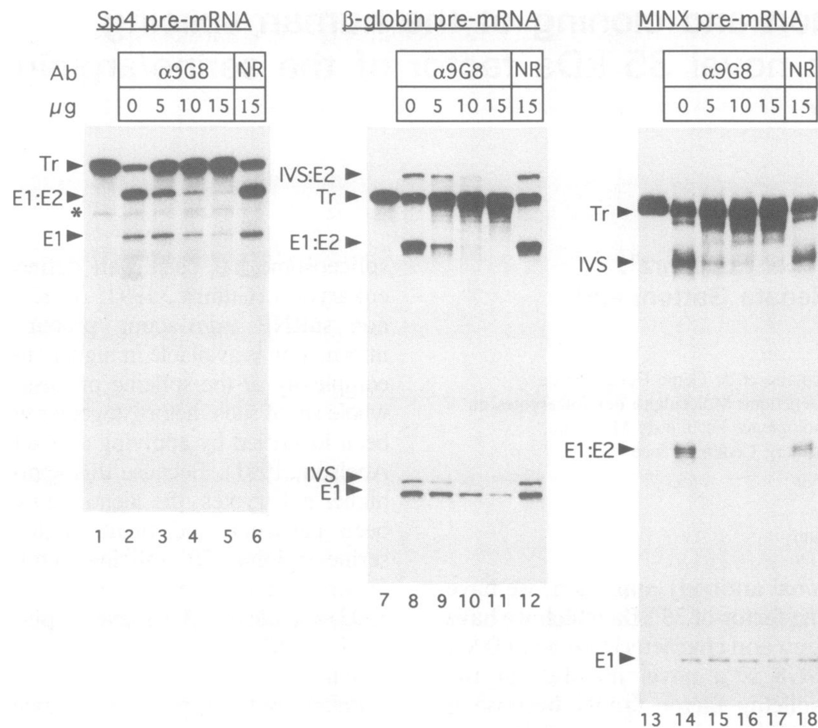

Fig. 1. Inhibitory effect of anti-9G8 on *in vitro* splicing. Splicing assays containing HeLa nuclear extract but no pre-mRNA were incubated for 30 min on ice in the presence or absence of purified mAbs (Ab) as indicated above each panel. Labeled pre-mRNAs (Sp4, β-globin *Bam*HI or MINX) were then added and the splicing reaction carried out for 2 h. Splicing products were analyzed by electrophoresis in a 5.2% (left panel) or 6% polyacrylamide, 8 M urea gel and the run length was adjusted for each substrate. Lanes 1, 7 and 13, initial pre-mRNA; lanes 2, 8 and 14, control splicing reaction without antibody; lanes 3–5, 9–11 and 15–17, splicing reaction in the presence of increasing amounts of anti-9G8; lanes 6, 12 and 18, splicing reaction in the presence of the non-relevant antibody (5G4). The transcript (Tr), pre-mRNA intermediates (E1 and IVS:E2) and final products (E1:E2 and IVS) are indicated on the left. In the left panel, the band indicated by an asterisk corresponds to a premature transcription termination product. In addition, the 13S IVS:E2 and IVS, which contain a limited number of C residues, are only poorly visible after overexposure of gels and are not detected in the exposure conditions that we used in Figures 1, 5 and 6.

of a protein associated with nascent transcripts, the SRp55 species, which was cloned in *Drosophila* (Roth *et al.*, 1991). Comparison of SRp55 with SF2/ASF (Mayeda *et al.*, 1992) and the development of a simple isolation method for SR splicing factors led to the identification of SR protein species of 20, 40, 55 and 75 kDa, in addition to the 30–35 kDa factors already cloned as SF2/ASF and SC35 (Zahler *et al.*, 1992).

Among the SR factor family, SF2/ASF and SC35 are the best characterized factors. They are required during the early steps of the splicing process, because the formation of the pre-splicing complex is abolished in their absence (Fu and Maniatis, 1990; Krainer *et al.*, 1990a). In addition, SC35 mediates the interactions of U1 and U2 snRNP with the 3′ region of the intron (Fu and Maniatis, 1992a), but it also intervenes much earlier to commit pre-mRNA substrates to the splicing pathway (Fu, 1993). Furthermore, using functional complementation assays, Zahler *et al.* (1992) have shown that each individual SR factor is sufficient to activate a splicing-deficient extract such as S100 in which SR factors are absent or only present at low levels. Thus, all SR factors appear to be functionally related, even if their ability to activate splicing does not appear to be completely equivalent (Zahler *et al.*, 1993a). Unexpectedly, this multiplicity of functionally related factors is in apparent contradiction with the fact that each SR factor is well conserved during evolution, thus strongly suggesting that each species has a specific and important role. Interestingly, SF2/ASF and SC35 are able to strongly influence *in vitro* 5′ splice site utilization when several 5′ splice sites are present in the pre-mRNA (Ge and Manley, 1990; Krainer *et al.*, 1990b; Fu *et al.*, 1992). Moreover, a systematic comparison of the ability of several SR factors to splice alternatively different pre-mRNA substrates shows that they are differentially active in modulating the alternative splicing (Zahler *et al.*, 1993a).

In a search for RNA binding factors involved in the modulation of alternative splicing of adenoviral E1A pre-mRNA during infection (Gattoni *et al.*, 1991), we have characterized a mAb, called 9G8, which is able to inhibit directly splicing *in vitro*. Here, we report the isolation of the 35 kDa factor recognized by the 9G8 antibody. cDNA clones encode a 238 amino acid polypeptide which represents a novel SR splicing factor that we have named 9G8. The recombinant 9G8 factor appears to complement SR factor-depleted extracts and to be functionally related to the other SR splicing factors.

Results

The 9G8 mAb inhibits splicing in vitro

A set of mAbs directed against heterogeneous nuclear ribonucleoproteins (hnRNP) were tested for their ability to inhibit splicing *in vitro*. One of them, the 9G8 mAb, an IgG1 which recognizes a 35 kDa antigen, clearly inhibits the splicing reaction. For a detailed analysis (Figure 1) we used several splicing substrates including the adenoviral E1A pre-mRNA (Sp4), a β-globin transcript covering the 5′ half of

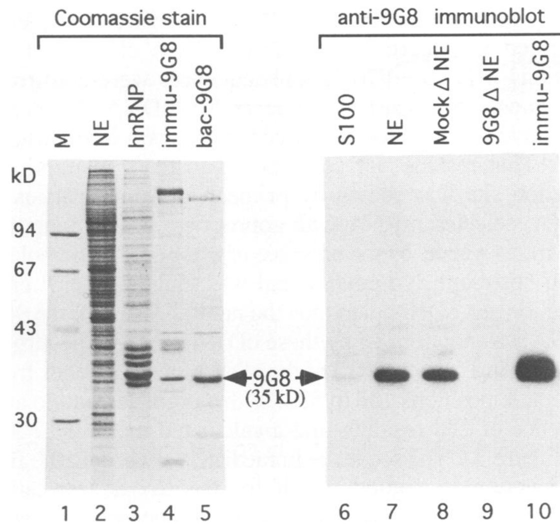

Fig. 2. SDS–polyacrylamide gel analysis of various subcellular preparations and of 9G8-containing fractions. The proteins in the indicated fractions are analyzed by SDS–polyacrylamide gel and stained by Coomassie blue (left panel) or by immunoblotting with anti-9G8 (right panel). On the left panel: lane 1, molecular mass markers (M); lane 2, crude HeLa nuclear extract; lane 3, purified HeLa hnRNP particles; lanes 4 and 5, 9G8 protein immunopurified from HeLa nuclear extract and baculovirus-infected Sf9 cytoplasmic fraction (see Materials and methods). Note that the weak bands of 42 and ~150 kDa common to immunopurified fractions in lanes 4 and 5 originate from anti-9G8 preparations. On the right panel: lane 6, HeLa S100 cytoplasmic extract (1.5×10^5 cells); lane 7, HeLa nuclear extract (3×10^5 cells); lanes 8 and 9, HeLa nuclear extracts (3×10^5 cells) depleted with the non-relevant antibody 5G4 (Mock Δ NE) or with anti-9G8 (9G8 Δ NE), respectively; lane 10, 9G8 protein immunopurified from 293 cells (1.5×10^6 cells).

the β-globin gene and the chimeric MINX transcript. Under the conditions chosen for the *in vitro* assays (see Materials and methods), all transcripts were spliced with an efficiency >50%, both in the absence of antibodies (lanes 2, 8 and 14) and in the presence of the highest amount of the non-relevant antibody used as a control (lanes 6, 12 and 18). As expected from the standard ionic conditions used, the 13S mRNA is the major product formed from the E1A transcript (lanes 2 and 6). The addition of increasing amounts of anti-9G8 results in considerable inhibition of splicing. For example, with the E1A transcript the level of 13S exon 1 and 13S mRNA decreases gradually (lanes 3–5) and splicing inhibition finally becomes almost complete. Similar results are obtained with the β-globin transcript (lanes 9–11) and the MINX transcript (lanes 15–17). However, with this latter transcript we note that the low level of exon 1 remains roughly constant with increasing amounts of anti-9G8, but the strong inhibition of the mRNA and IVS formation indicates that not only the second step but also the first step of splicing are severely impaired by the antibody. Thus the data in Figure 1 strongly suggest that the 9G8 antibody either directly or indirectly recognizes a factor involved in the splicing process.

Characterization of the 9G8 factor and cloning of its cDNA

Immunoblot analysis of a nuclear extract from HeLa cells using the 9G8 antibody shows that the recognized protein, which we refer to as the 9G8 factor, has an apparent molecular mass of ~35 kDa (Figure 2, lane 7), comparable with that of the SC35 factor (Fu and Maniatis, 1992a). The 9G8 factor is mainly present in the nuclear fraction, and only

```
1                                                      GTAGTGCCGCCGGGACTCTTGGCGGG
27   TGAAGGTGTGTGTCAGCTTTTGCGTCACTCGAGCCCTGGGCGCTGCTTGCTAAAGAGCCGAGCACGCGGGTCTGTCATC
106  ATG TCG CGT TAC GGG CGG TAC GGA GGA GAA ACC AAG GTG TAT GTT GGT AAC CTG GGA ACT
1    M   S   R   Y   G   R   Y   G   G   E   T   K   V   Y   V   G   N   L   G   T

166  GGC GCT GGC AAA GGA GAG TTA GAA AGG GCT TTC AGT TAT TAT GGT CCT TTA AGA ACT GTA
21   G   A   G   K   G   E   L   E   R   A   F   S   Y   Y   G   P   L   R   T   V

226  TGG ATT GCG AGA AAT CCT CCA GGA TTT GCC TTT GTG GAA TTC GAA GAT CCT AGA GAT GCA
41   W   I   A   R   N   P   P   G   F   A   F   V   E   F   E   D   P   R   D   A

286  GAA GAT GCA GTA CGA GGA CTG GAT GGA AAG GTG ATT TGT GGC TCC CGA GTG AGG GTT GAA
61   E   D   A   V   R   G   L   D   G   K   V   I   C   G   S   R   V   R   V   E

346  CTA TCG ACA GGC ATG CCT CGG AGA TCA CGT TTT GAT AGA CCA CCT GCC CGA CGT CCC TTT
81   L   S   T   G   M   P   R   R   S   R   F   D   R   P   P   A   R   R   P   F

406  GAT CCA AAT GAT AGA TGC TAT GAG TGT GGC GAA AAG GGA CAT TAT GCT TAT GAT TGT CAT
101  D   P   N   D   R   C   Y   E   C   G   E   K   G   H   Y   A   Y   D   C   H

466  CGT TAC AGC CGG CGA AGA AGA AGC AGG TCA CGG TCT AGA TCA CAT TCT CGA TCC AGA GGA
121  R   Y   S   R   R   R   R   S   R   S   R   S   R   S   H   S   R   S   R   G

526  AGG CGA TAC TCT CGC TCA CGC AGC AGG AGC AGG GGA CGA AGG TCA AGG TCA GCA TCT CCT
141  R   R   Y   S   R   S   R   S   R   S   R   G   R   R   S   R   S   A   S   P

586  CGA CGA TCA AGA TCT ATC TCT CTT CGT AGA TCA AGA TCA GCT TCA CTC AGA AGA TCT AGG
161  R   R   S   R   S   I   S   L   R   R   S   R   S   A   S   L   R   R   S   R

646  TCT GGT TCT ATA AAA GGA TCG AGG TAT TTC CAA TCC CCG TCG AGG TCA AGA TCA AGA TCC
181  S   G   S   I   K   G   S   R   Y   F   Q   S   P   S   R   S   R   S   R   S

706  AGG TCT ATT TCA CGA CCA AGA AGC AGC CGA TCA AAG TCC AGA TCT CCA TCT CCA AAA AGA
201  R   S   I   S   R   P   R   S   S   R   S   K   S   R   S   P   S   P   K   R

766  AGT CGT TCC CCA TCA GGA AGT CCT CGC AGA AGT GCA AGT CCT GAA AGA ATG GAC TGA AGC
221  S   R   S   P   S   G   S   P   R   R   S   A   S   P   E   R   M   D   *

826  TCTCAAGTTCACCCTTTAGGGAAAAGTTATTTTGTTTACATTATTATAAGGGATTTGTGATGTCTGTAAAGTGTAACCT
905  AGGAAAGATAATTCAACCATCTAATCAAAATGGATCTGGATTACTATGTAAATTCACAGCAGTAAG
```

Fig. 3. Nucleotide and deduced amino acid sequence of 9G8 cDNA. Nucleotides are numbered from the transcription initiation site (+1). The coding sequence extends from positions 106 to 822 (stop codon). The amino acid positions (+1 for the ATG codon) are numbered in bold and italic. The three tryptic peptides initially sequenced are underlined and those subsequently sequenced are YGGETK, VYVGNLGTGAGK, NPPGFAFVEFEDPR and VELSTGMPR. RNP-1 and RNP-2 of the RNA binding domain (RBD) are boxed in reverse type. The repeated RRSRSXSX consensus sequences in the SR domain are underlined with dashes. The invariable cysteine and histidine residues of the CCHC motif in the median region are circled. The EMBL/GenBank accession number for 9G8 is L22253.

low amounts are detected in the cytoplasmic S100 fraction (lane 6). The factor was immunopurified from nuclear extracts of human 293 cells. This was performed under stringent conditions because the extracts were diluted and brought to 1 M monovalent salt and 0.25% NP-40 (see Materials and methods). Gel electrophoresis of the eluted fractions (Figure 2, lane 4) revealed only a limited number of major bands, including an irregularly obtained high molecular weight band (>150 kDa) and a band which comigrates with the 9G8 immunoreactive band, as determined by immunoblot analysis (lane 10). The stained 9G8 band (estimated molecular mass of 35 kDa, compared with protein markers) migrates slightly above the A1 protein (34 kDa) of hnRNP (lane 3). Otherwise, analysis of the protein distribution in the nuclear extract and hnRNP (lanes 2 and 3) shows that 9G8 is not a major protein within these two preparations. A rough estimation indicated that in our HeLa nuclear extract preparations the 9G8 factor constitutes ~0.05% by weight of the nuclear proteins. On an SDS-polyacrylamide gel, the 9G8 protein present in the immunopurified fraction migrates as a well isolated band (lane 4), which allowed further purification by preparative gel electrophoresis. From this, we prepared tryptic peptides which were then separated by HPLC. The sequencing of a first set of major peaks gave an unambiguous amino acid sequence for three peptides (which are underlined in Figure 3). A comparison with protein sequence data banks indicated that the 9G8 peptides originate from a novel protein species, but nevertheless revealed a significant homology with the RNA binding domain (RBD) of the X16 protein (Ayane *et al.*, 1991), the murine equivalent of the hSRp20 species which belongs to the SR splicing factor family (Zahler *et al.*, 1992). Analysis of other HPLC peaks, performed subsequently, has revealed four additional peptide sequences (indicated in the legend to Figure 3), but no sequences equivalent to the SC35 and SF2/ASF factors were obtained from the HPLC peaks analyzed. We have also verified with polyclonal antibodies directed against C-terminal peptides of SC35 and SF2/ASF that these factors were undetectable in the immunopurified 9G8 fraction.

The N- and the C-terminal sequences of the first peptide set were selected to design degenerate primers which were then used for RT-PCR from poly(A)$^+$ mRNA of 293 cells. The sequence of a 92 bp sequence PCR product was consistent with the peptide sequences and confirmed the presence of a RBD. Because this relatively well conserved domain is found in many RNA binding proteins, some of which are present at high levels in the cell, we predicted that the screening of a cDNA library with a RBD sequence probe might lead to an excess of cDNA clones not relevant to the 9G8 sequences. Therefore, we first screened a human genomic DNA library with the 38 nucleotide (nt) QE203 probe (see Materials and methods for definition of the probes), which revealed two positive clones containing inserts of 15 and 17 kb. Sequencing of genomic subclones upstream and downstream from the 92 bp fragment extended the coding sequence and also included predicted intronic sequences. The extra coding sequence downstream from the RNP1 consensus sequence has still retained significant homology with the RBD of other SR factors, but was more divergent compared with other RNA binding proteins. We thus screened a cDNA library with a newly defined probe (QK7) in this region, which gave nine positive clones with insert size between 700 and 1900 bp. The DNA sequence presented in Figure 3 was derived from clone 3, from positions +17 to +970. These sequences were confirmed by sequencing at least two independent cDNA clones and a genomic subclone for the sequence upstream from position +17. That residue +1 corresponds to the transcriptional initiation site was shown by primer extension analysis of poly(A) selected mRNA with appropriate probes (data not shown), as well as by the presence of a TATA box homology 22 bp upstream. No poly(A) tail was found at the 3′ ends of the longest cDNA, and thus the natural 3′ end of the 9G8 mRNA was not defined by these cDNA clones. The largest open reading frame (ORF) of the cDNA clones extends from nucleotide positions 106 to 822, with a predicted amino acid sequence of 238 residues and a calculated mass of 27 350 Da (Figure 3). The sequence immediately preceding the first AUG codon is a good match for the Kozak translation initiation consensus and the deduced amino acid sequence was consistent with the sequence of all seven tryptic peptides determined previously.

Sequence features of the 9G8 factor

From the deduced amino acid sequence, the 9G8 factor appears to be a novel member of the SR factor family. It includes a consensus 80 amino acid RBD in the N-terminal portion and an SR domain in its C-terminal portion, separated by an atypical region of ~40 amino acids. As mentioned previously, the RBD of 9G8 exhibits a strong homology with that of hSRp20 (sequence identity 78%), as well as with that of RBP1, another small SR factor isolated from *Drosophila* (Kim *et al.*, 1992) (sequence identity 71%). However, the sequence identity with SF2/ASF and SC35/PR264 is only 45 and 37%, respectively (see also Figure 7 for an alignment of the RBD sequences). The SR domain of the 9G8 factor, which extends from residues 123 to 233, has a size which is similar to that of SC35, but longer than that of SF2/ASF or SRp20 (35 and 49 residues, respectively). In addition, we found a basic sequence **RRRRSRSR** at the N-terminus of the SR domain of 9G8 (six basic residues out of eight residues), which also exists in SR domains of SC35 (**RRRRRSRS**), hSRp20 (**RRRSPRRR**) and SRp55 (**RRRSRSRR**), but not of SF2/ASF or RBP1 (Figure 7). Such basic regions are also found in *Drosophila* factors, namely suppressor of white apricot and transformer, involved in the regulation of alternative splicing (Li and Bingham, 1991), and they may serve as nuclear localization signals (Dingwall and Laskey, 1991). In the SR domain of 9G8, we also found a set of arginine doublets, most of them being present in a sequence of eight residues with the consensus RRSRSXSX. These sequences, underlined with dashes in Figure 3, are repeated six times in the SR domain of 9G8 but are not present in other splicing factors. Finally, the atypical region separating the N- and C-terminal domains, which is located between amino acids 81 and 122, does not display good sequence homology with the median region of any other SR splicing factor. However, this region contains a Cys-X_2-Cys-X_4-His-X_4-Cys sequence which corresponds to the zinc knuckle motif of the CCHC family (Green and Berg, 1989).

Since the 9G8 has characteristics similar to those of SC35, it was surprising that Zahler *et al.* (1992) did not recover this factor when they isolated the SR factor family using a two-step procedure including an ammonium sulfate

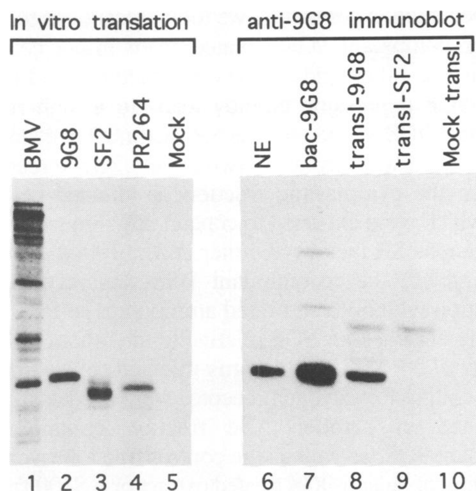

Fig. 4. *In vitro* expression of 9G8 and other SR cDNAs. *In vitro* translation in the presence (lanes 1–4, 8 and 9) or absence (mock, lanes 5 and 10) of capped mRNAs coding for 9G8 and other SR factors was performed in wheat germ extract with ^{35}S-labeled (lanes 1–5) or unlabeled (lanes 8–10) methionine. The proteins from these preparations and from other fractions were resolved on an SDS–polyacrylamide gel. The labeled translation products were visualized by fluorography (left panel) and the unlabeled products were analyzed by immunoblot with anti-9G8 (right panel). *In vitro* translation was in the presence of RNA of BMV (lane 1), 9G8 (lanes 2 and 8), SF2 (lanes 3 and 9), PR264/SC35 (lane 4) or in the absence of RNA (lanes 5 and 10). Proteins of nuclear extract (NE) and baculovirus-expressed 9G8 (bac-9G8) were analyzed in lanes 6 and 7, respectively.

fractionation followed by a $MgCl_2$ precipitation. However, we have observed that 9G8 is quantitatively recovered in the 60% saturated ammonium sulfate pellet of the nuclear extract (not shown), whereas the other described SR factors are recovered in the 65–90% saturated ammonium sulfate pellets (Zahler *et al.*, 1992). Thus, these data easily explain why 9G8 has not been identified in the previous characterization of SR factors.

Expression of the recombinant 9G8 factor

To analyze further the 9G8 factor we attempted to express the recombinant 9G8. Therefore, the 9G8 cDNA was cloned in an *in vitro* transcription vector and the resultant transcript, along with control transcripts from SF2/ASF and PR264/SC35, were translated either in a reticulocyte lysate (not shown) or in a wheat germ extract. After gel electrophoresis, the ^{35}S-labeled or non-radioactive proteins were detected by autoradiography and immunoblot analysis, respectively (Figure 4). Fluorography shows that the 9G8 transcript promotes the synthesis of a 35 kDa protein, whereas the SF2/ASF and SC35/PR264 products migrate significantly faster, with an apparent molecular mass of 33 and 34 kDa, respectively. Immunoblot analysis shows that the protein produced from the 9G8 transcript is recognized by the 9G8 antibody and comigrates with the 9G8 present in nuclear extract (lanes 6 and 8). This confirms that our cloned cDNA indeed encodes the 9G8 factor. In contrast, anti-9G8 did not reveal significant signals with the recombinant SF2/ASF (lane 9).

Our attempts to overexpress recombinant 9G8 in

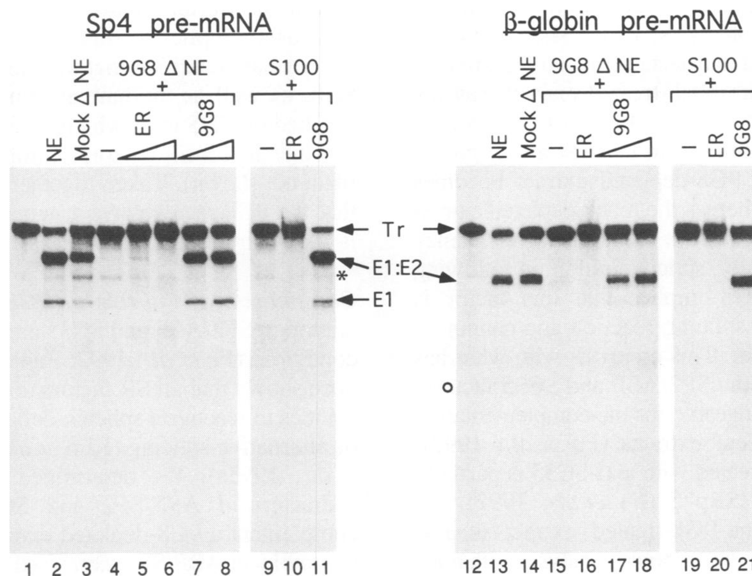

Fig. 5. *In vitro* complementation of 9G8-depleted extract or cytoplasmic S100 fraction. The Sp4 (left panel) or β-globin *Hinc*II (right panel) transcripts were spliced *in vitro* with various nuclear extracts or S100 preparations and either were or were not complemented with recombinant 9G8 or mock fractions. The splicing products were analyzed in a 5.2% polyacrylamide gel. Lanes 1 and 12, initial Sp4 and β-globin pre-mRNA; lanes 2 and 13, standard splicing in the presence of 7 μl nuclear extract (NE); lanes 3 and 14, nuclear extract depleted with non-relevant 5G4 IgG1 (Mock Δ NE). The anti-9G8-treated nuclear extract (9G8 Δ NE) in lanes 4–8 and 15–18 is complemented with the cytoplasmic fraction of Sf9 cells infected with baculovirus containing either the estrogen receptor cDNA (ER, lanes 5, 6 and 16), or the 9G8 cDNA (lanes 7, 8, 17 and 18). Lanes 9–11 and 19–21, splicing reaction in the presence of 10 μl S100 (lanes 9 and 19), complemented with the mock fraction (lanes 10 and 20) or recombinant 9G8 fraction (lanes 11 and 21). The estimated amounts of recombinant 9G8 used for complementation are 30 and 60 ng for the depleted nuclear extracts and 45 ng for the S100 fraction. The asterisk and circle alongside the panels indicate prematurely terminated transcript and an undetermined cleavage product, respectively. Note that the IVS/E2 intermediate of the β-globin transcript migrates here, exactly like the initial transcript, and that the E1 and IVS products, which migrate in the bottom of the gel, are not shown.

Escherichia coli by means of pET vectors (Novagen), or pQE vectors (Diagen/Qiagen) which are derived from the pDS56 vector (Gentz *et al.*, 1989), were unsuccessful, whereas the insertion of the 9G8 coding sequence into the GST fusion vector pGEX-3X (Pharmacia) allowed the production of a truncated fusion protein (data not shown). Therefore, we chose a eukaryotic expression system and inserted the 9G8 cDNA in a baculovirus vector to express the recombinant protein in insect cells. Immunoblot analysis of the proteins of infected cells showed that recombinant 9G8 gave a signal at least 50 times higher compared with the endogenous signal. Moreover, the amount of 9G8 expressed in insect cells is ~20–25 times higher than in human cells. After a 2 day infection, the 9G8 factor was roughly equally distributed between cytoplasmic and nuclear fractions. The recombinant 9G8 factor was then purified on anti-9G8–Sepharose, analyzed by SDS–polyacrylamide gel and revealed by staining or immunoblot analysis. Figure 2 shows that the immunopurified fraction is highly enriched for recombinant 9G8 factor (lane 5), which comigrates with the human 9G8 factor (compare lane 10 with lane 7).

Functional assays of the recombinant 9G8 protein

To check whether the recombinant 9G8 intervenes in splicing, we relied on nuclear extract depletion with the 9G8 antibody followed by complementation with the recombinant 9G8 protein. Nuclear extracts were incubated with anti-9G8–Sepharose in the presence of 0.42 M NaCl, which is expected to minimize non-specific interactions between nuclear components as well as non-specific adsorption of nuclear components on the antibody–Sepharose (Ast *et al.*, 1991). However, this treatment is less stringent than that used for 9G8 immunopurification. Depletion of 9G8 was monitored by immunoblot analysis, which shows that the level of 9G8 becomes almost undetectable in anti-9G8-treated extract, whereas that of the mock-treated extract was not affected (Figure 2, lanes 8 and 9). In Figure 5, parallel analysis of the splicing abilities with the Sp4 and β-globin transcripts shows that the 9G8-depleted extract becomes inactive (lanes 4 and 15), whereas the mock-depleted extract only revealed a slight decrease in the splicing efficiency (lanes 3 and 14). Theoretically, splicing inhibition following a specific depletion of 9G8 implies that this factor is absolutely required for the splicing reaction and cannot be replaced by other SR factors. This contrasts with what has been shown previously for the SF2/ASF and SC35 factors, which appear to be interchangeable for the complementation of SR factor-depleted nuclear extracts (Fu *et al.*, 1992). Because a nuclear extract treated with anti-SC35 is partially depleted in SF2/ASF and SRp75 (Fu *et al.*, 1992), we determined whether an anti-9G8-treated extract is also deficient in the SC35 factor. Immunoblot analysis with anti-SC35 showed that the treatment with anti-9G8 also results in an almost complete depletion of SC35 (data not shown). Thus, the loss of the splicing activity of anti-9G8-treated extract may be due to a significant depletion of certain, if not all, of the SR splicing factors, including the 9G8 factor. In agreement with the above and with the previous results (Fu and Maniatis, 1990), we have observed that an anti-9G8-treated extract loses its ability to promote the formation of the presplicing and splicing complexes (data not shown).

To test directly the role of 9G8 in splicing in the following complementation experiments, we took precautions to ensure that the recombinant 9G8 isolated from insect cells was devoid of the other endogenous SR factors. In fact, the efficient 9G8 expression already leads to a high ratio of recombinant 9G8 to endogenous SR factors. Their presence was minimized still further by two means: (i) we recovered 9G8 from the cytoplasmic fraction of infected cells (see earlier) which were expected to contain only limited amounts of endogenous SR factors (Krainer *et al.*, 1990a; Fu *et al.*, 1992); and (ii) the recombinant 9G8 was recovered by precipitation with 60% saturated ammonium sulfate, which selectively enriches for 9G8 relative to the other SR factors (Zahler *et al.*, 1992). Baculovirus-infected cells expressing a cDNA coding for estrogen receptor were used as a control and treated in parallel. The fraction containing the recombinant 9G8, as well as the control fraction, were used to complement an anti-9G8-treated extract or a S100 fraction, splicing experiments being performed with E1A and β-globin transcripts (Figure 5). The results show that the 9G8-depleted nuclear extract was well complemented with the 9G8-containing fraction (lanes 7 and 8), because ~30 ng of recombinant 9G8 was sufficient to restore the splicing activity. Interestingly, this amount is roughly equivalent to the quantity present in 10 μl of HeLa nuclear extract, indicating that recombinant 9G8 is very effective in complementing an SR factor-depleted extract. In contrast, the control ER-containing fraction could complement neither 9G8-depleted nuclear extract (lanes 5 and 6), nor the S100 fraction (lane 10), indicating that the insect cytoplasmic fractions, prepared as described above, did not contain significant amounts of endogenous SR factors. Exactly the same complementation results were obtained with the β-globin transcript (lanes 16–21), suggesting that our conclusion applies to transcripts other than E1A. Finally, we have also verified that the immunopurified recombinant 9G8, as well as the human immunopurified 9G8 (further purified on SDS gel), which are both expected to be devoid of other insect SR factors, complement a depleted extract (data not shown). Taken together, these results demonstrate that the 9G8 factor plays a general role in splicing, which is likely to be similar to that of the other SR factors.

Comparison of the role of 9G8 and other SR factors

Although SF2/ASF and SC35 are interchangeable in certain conditions (Fu *et al.*, 1992; Zahler *et al.*, 1992), it has also been shown that all SR factors do not have exactly the same abilities to reactivate splicing-deficient extracts in constitutive or alternative splicing (Kim *et al.*, 1992; Fu, 1993; Zahler *et al.*, 1993a). We determined to what extend the well characterized ASF/SF2 and SC35 factors are able to complement a 9G8-depleted extract. Therefore, we used a recombinant ASF/SF2 expressed in *E.coli* (Ge *et al.*, 1991) and a SC35 factor prepared from HeLa cells, according to Fu and Maniatis (1992b). The SC35 factor preparation is in fact expected to consist of a mixture of SC35 and ASF/SF2, where SC35 is predominant. However, it does not contain 9G8 as verified by using a polyclonal antibody directed against a C-terminal peptide of 9G8 (not shown). Both SC35 and ASF/SF2 preparations were functional, because they were able to complement an S100 fraction for its ability to splice the E1A transcript (Figure 6, lanes 11–14). We observed that the SC35 preparation efficiently rescues the 9G8-depleted extract (lanes 5 and 6), indicating

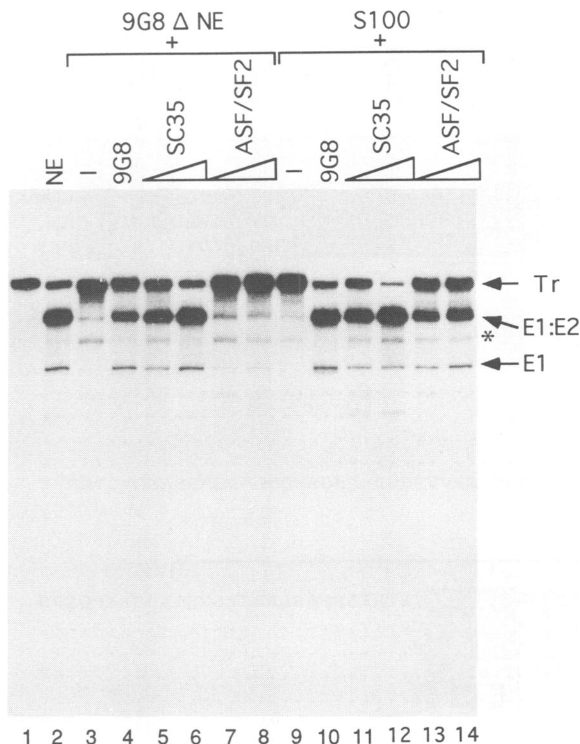

Fig. 6. Comparison of *in vitro* complementation of 9G8-depleted nuclear extract and cytoplasmic S100 fraction with 9G8, SC35 or ASF/SF2 factors. The Sp4 transcript is spliced *in vitro* and the splicing products are analyzed as described in Figure 5. Lane 1, the initial Sp4 transcript; lane 2, splicing in the presence of 7 μl crude nuclear extract; lanes 3–8, splicing in the presence of anti-9G8-treated extract (lane 3) complemented with 45 ng recombinant 9G8 (lane 4), with 30 and 60 ng purified SC35 (lanes 5 and 6) and with 75 and 150 ng *E.coli* recombinant ASF/SF2 (lanes 7 and 8). The same complementation assays are performed with the S100 fraction (lanes 9–14).

that 9G8 and SC35 have similar splicing activities. However, the recombinant ASF/SF2 was unable to rescue the same depleted extract (lanes 7 and 8), suggesting that the *E.coli*-expressed ASF/SF2 is not equivalent to either the native SC35 or the recombinant 9G8 preparation. Further biochemical complementation assays will be required to determine whether or not this reflects an intrinsic property of the ASF/SF2 factor.

Discussion

A mAb 9G8 which inhibits *in vitro* splicing was used to immunopurify the recognized factor. The partial sequencing of this factor allowed the isolation and characterization of the corresponding cDNA. Primary sequence analysis along with depletion–complementation experiments showed that 9G8 is a novel SR splicing factor. This 9G8 factor has an apparent molecular mass of 35 kDa, which is much higher than its theoretical molecular mass (27.3 kDa). This discrepancy, which is also observed for all other SR factors, is at least in part due to multiple phosphorylations occurring on serine residues (Roth *et al.*, 1991; our unpublished results). We have also shown that 9G8 is a low abundance protein and that its abundance is comparable with that of other SR factors which represent, on the whole, 0.04% of cellular proteins (Zahler *et al.*, 1992).

Comparison of 9G8 and other SR factors

Although 9G8 and SC35 share similar general characteristics, their resemblance is limited to the molecular mass and the size of different domains. Comparison of the amino acid sequence of the RBD of SR factors shows that the best homology occurs between 9G8 and the small SR factor SRp20 (Figure 7). Alignment of the RBD sequence according to Kenan *et al.* (1991) indicated that the determinants of the RNA binding specificity, at least for the U snRNP proteins analyzed up to now (Scherly *et al.*, 1990; Bentley and Keene, 1991; Mattaj, 1993), are located in loops L1 and L3, which are expected to be well-exposed sequences. Interestingly, we observe in Figure 7 that the amino acid differences between 9G8 and SRp20 are concentrated in loops L1 and L5. In contrast, the sequences of loop 3, which are the most variable within a set of >30 different RBD (Kenan *et al.*, 1991), are identical in 9G8, SRp20 and RBP1, but differ from those of SC35 and SF2/ASF. Thus, it is likely that members of the SR family do not have exactly the same RNA binding specificity. This agrees with the fact that single SR factors commit different pre-mRNAs to splicing with a pronounced substrate specificity (Fu, 1993).

Comparison of the sequences of SR domains shows that 9G8 contains, in addition to the highly basic sequence (RRRRSRSR), a set of arginine doublets, most of them being part of an eight amino acid sequence with the consensus RRSRSXSX, which is repeated six times in 9G8 (Figure 7). The X residue is very often a hydrophobic amino acid, because proline, leucine, isoleucine or alanine residues are found 10 times in the 12 possibilities. Interestingly, this consensus consists of a highly charged and hydrophilic part followed by a neutral and more hydrophobic part. Such repetitions, which are not found in the other SR factors, create a regular alternation of basic and non basic motifs, with a four-residue periodicity besides the general serine/arginine alternation, and might be the basis for specific properties of the SR domain of the 9G8 factor.

As with the other SR factors, 9G8 contains many putative phosphorylation sites and we have good evidence that 9G8 isolated from HeLa cells, as well as from baculovirus-infected cells, is phosphorylated (data not shown). The phosphorylation of the SR factors is likely to be general, as the antibody mAb104, which binds to an epitope phosphorylated on serine, recognizes most of the SR factors (Roth *et al.*, 1991; Zahler *et al.*, 1992). It is not known whether the phosphorylation of the different SR factors is absolutely required for activity. However, protein phosphorylation is known to intervene in a large variety of biological processes and there is some evidence that phosphorylation–dephosphorylation cycles occur on the U1 70 kDa protein and that these cycles are crucial for its *in vitro* splicing activity (Tazi *et al.*, 1993).

The median region located just downstream from the N-terminal RBD shows considerable divergence. Indeed, this region is expected to be highly flexible in ASF/SF2 or SC35 due to repeats of glycine or to the presence of glycine and proline residues, respectively. In contrast, we expect a more structured sequence in 9G8 because its median region contains a putative zinc knuckle of the CCHC family. This motif was initially encountered in the gag-encoded small nucleic acid binding proteins of retroviruses (Green and Berg, 1989), then in a human cellular nucleic acid binding

```
RNA Binding Domain

           Beta-1   Loop-1   Alpha-1   Loop-2  Beta-2   Loop-3   Beta-3 Loop-4 Alpha-2   Loop-5   Beta-4
9G8      TKVYVGNLGTGAGKGELERAFSYYGPLRTVWIARN.....PPGFAFVEFEDPRDAEDAVRGLDGKVICGSRVRVELST
RBP1     CKVYVGNLGSSASKHEIEGAFAKYGPLRNVWVARN.....PPGFAFVEFEDRRDAEDATRALDGTRCCGTRIRVEMSS
SRp20    CKVYVGNLGNNGNKTELERAFGYYGPLRSVWVARN.....PPGFAFVEFEDPRDAADAVRELDGRTLCGCRVRVELSN
SC35     TSLKVDNLTYRTSPDTLRRVFEKYGRVGDVYIPRDRYTKESRGFAFVRFHDKRDAEDAMDAMDGAVLDGRELRVQMA.
SF2/ASF  CRIYVGNLPPDIRTKDIEDVFYKYGAIRDIDL.KNR..RGGPPFAFVEFEDPRDAEDAVYGRDGYDYDGYRLRVEFPR
           RNP 2                                  RNP 1

Intermediate Zone

9G8      GMPRRSRFDRPPARRPFDPNDRCYECGEKGHYAYDCHRY........................................
RBP1     G..............................................................................
SRp20    G..............................................................................
SC35     ......RYGRPPDSHHSRRGPPPRRYGGGGYG...............................................
SF2/ASF  SGRGTGR.................GGGGGGGGAPRGRYGPPSRRSENRVVVSGLPPSGSWQDLKDHMREAGDVCYADVY

SR Domain

9G8      ......SRRRRSRSRSRSHSRSRGR.RYSRSRSRSRGRRSRSASPRRSRSISLRRSRSASLRRSRSGSIKGSRYFQSPS
RBP1     ..........RSRDRRRGEGGSSG..........RSGSGRYRSRSPRRSRS............................
SRp20    ..........EKRSRNRGPPPSWGR.R.PRDDYRRRSPPPRRRSPRR.RSF............................
SC35     RRSRSPRRRRRSRSRSRSRSRSRSRSRYSRSKSRSRTRSRSRST.SKSRS...........................AR
SF2/ASF  ..........................................P.RSPSY..............................

9G8      RSRSRSRSISRP.RSSRSKSRSPSPKR.....SRSPSGSPRRSASPERMD
RBP1     ...PRSRSFSRD.RRSRSDSRDRH
SRp20    .SRSRSRSLSRDRRRERSLSRERNHKP.....SRSFSRSRSRSRSNERK
SC35     RSKSKSSSVSRSRSRSRSRSRSRSPPPVSKRESKSRSRSKSPPKSPEEEGAVSS
SF2/ASF  .GRSRSRSRSRSRSRSRSNSRSRSYSPRRSRGSPRYSPRHSRSRSRT
```

Fig. 7. Alignment of the peptide sequences of five members of the SR splicing factor family. Peptide sequences of the three different domains of 9G8, RBP1, SRp20, SC35 and SF2/ASF were aligned according to Kenan *et al.* (1991) to emphasize amino acids identical in three out of five sequences (solid boxes). The secondary structure of RBD is indicated above the sequence alignment of this region. The positions of RNP-1 and RNP-2 are indicated under the alignment. The four conserved amino acids of the putative zinc knuckle of 9G8 are underlined. The repeated sequences RRSRSXSX in the SR region are highlighted with bold lines above.

protein (CNBP) (Rajavashisth *et al.*, 1989) and in a yeast splicing factor, SLU7, which mediates 3′ splice site choice (Franck and Guthrie, 1992; Mattaj, 1993). Significantly, the amino acid residues separating the invariable cysteine–histidine residues are well conserved because the **CGEKGH** sequence of 9G8 matches very well with that of HIV gag **CGKEGH** and SLU7 gene **CGEAGH**, and the terminal sequence **HYAYDC** of 9G8 matches well with that of CNBP (**HIARDC**). This is the first description of a characteristic CCHC motif in an SR splicing factor, and it is possible that this zinc knuckle is important for a specific function of 9G8. Putative zinc fingers of the CCHH class have been detected recently in PRP6, -9 and -11 in yeast, as well as in the mammalian splicing factor SAP62, homologous to the yeast PRP11 (Legrain and Choulika, 1990; Bennett and Reed, 1993). Some of these motifs are absolutely required for viability of yeast (Legrain and Choulika, 1990; Legrain and Chapon, 1993).

General role of SR factors in splicing

In our complementation experiments, we show that the recombinant 9G8 factor expressed in the baculovirus system, as well as the purified SC35 factor, are able to complement an anti-9G8-treated nuclear extract and an S100 extract. This indicates that 9G8 is indeed a splicing factor and that 9G8 appears to be functionally related to SC35. The finding of multiple SR splicing factors with the same general function in splicing appears to be puzzling, except if a more specific role attached to each of these factors is directly associated to this general function. However, the apparent interchangeability of the SR factors may be the result of the two particular situations in which the properties of these factors have been assessed. Complementation assays of splicing-deficient extracts (S100 extract or SR factor-depleted nuclear extract) may induce a bias, as limiting but detectable amounts of certain SR factors may persist in these extracts. For instance, ASF/SF2 and SC35 (Fu *et al.*, 1992), as well as 9G8 (this paper), are detected at low levels in S100 extracts. As suggested earlier (Fu *et al.*, 1992), if all SR proteins are part of common protein complexes, it is possible that the addition of any of the SR proteins drives assembly of the complexes so that they are no longer limiting for a complementing activity. A similar situation might also exist within the SR factor-depleted nuclear extracts, as an anti-SC35-treated nuclear extract still contains detectable levels of ASF/SF2 and SRp75, as pointed out previously (Fu *et al.*, 1992). Therefore, limiting levels of certain SR factors might help the complementation by individual SR factors, and thus mask some specific properties. This is well illustrated by our work because the *E.coli* recombinant ASF/SF2 is able to complement an S100 extract, but not an anti-9G8-treated extract, indicating that these two splicing-deficient preparations are not completely equivalent and that the recombinant ASF/SF2 and 9G8 factors used here do not have

the same complementation properties. It will be interesting to determine whether this reflects differences in the intrinsic properties of these factors (ASF/SF2 might require the presence of factors or activities which are not necessary for 9G8), or only differences in their post-translational modifications. Finally, although most of the SR factors exhibit similar abilities to complement an S100 fraction (Zahler *et al.*, 1992, 1993b), it has been shown that SC35, SRp20, ASF/SF2 and SRp55 have very different specificities in the commitment of different pre-mRNAs to the splicing pathway in conditions of competition (Fu, 1993). Thus, *in vivo*, it is possible that the SR factors are not as interchangeable in their function as thought previously.

An intriguing question which has been posed as a result of the analysis of the role of SC35 and 9G8 factors in splicing is why immunodepletion assays with mAbs (anti-SC35 or anti-9G8) result in a substantial depletion of most, if not all other, SR factors. Two possible explanations may be advanced for this feature. First, partial crossreactivities between anti-SC35 and SF2 (Fu *et al.*, 1992) as well as anti-9G8 and the SC35 factor, but not ASF/SF2 (unpublished data), have been noted. However, this explanation may not account for a significant depletion of all SR proteins, because anti-SC35 and anti-9G8 do not crossreact systematically with other SR factors in immunoblot analysis. A second possibility is that SR factors are present in nuclear extracts as common protein complexes (Fu *et al.*, 1992), a possibility also raised in recent studies by Wu and Maniatis (1993), so that the treatment of nuclear extracts with an antibody specific for a given SR factor may result in an important co-immunodepletion of all SR factors.

Specific role of SR splicing factors in alternative splicing

There is now much evidence that SR factors are involved in the modulation of alternative splicing *in vitro* (Ge and Manley, 1990; Krainer *et al.*, 1990b; Tian and Maniatis, 1993). Interestingly, this is consistent with the fact that SC35 and ASF/SF2 intervene very early in the splicing process at some steps (commitment complexes, interactions between U1 and U2 snRNP) for which a selection of 5′ splice sites may take place easily. Very recently, it has been proposed that preferential selection of downstream 5′ splice sites in the presence of SF2/ASF may be mediated by an increase of U1 snRNP binding to all 5′ splice sites, indiscriminately (Eperon *et al.*, 1993). Furthermore, it has been shown that the two RNA binding domains of SF2/ASF are sufficient to confer alternative splicing activity (Caceres and Krainer, 1993; Zuo and Manley, 1993). Moreover, a comparison of the SRp30b (SC35), SRp40, 55 and 75 factors showed that they are able to modulate, in a dose-dependent manner, the splicing of pre-mRNA containing several 5′ splice sites, but they have distinct effects on the alternative splicing of the tested E1A and SV40 early pre-mRNA (Zahler *et al.*, 1993a). By using derivatives of these pre-mRNA, we have shown recently that 9G8 is also involved in the modulation of alternative splicing *in vitro* (R.Gattoni and J.Stévenin, unpublished results).

Although direct proofs that SR factors play roles *in vivo* in the modulation of alternative splicing do not exist, several lines of evidence support this proposition. First, as noted earlier, the strict size and sequence conservation of the different SR proteins from *Caenorhabditis elegans* to human (Zahler *et al.*, 1992) strongly suggest that individual SR proteins have essential and distinct functions *in vivo*. Second, the observation that the individual SR factors are differentially expressed in various tissues (Zahler *et al.*, 1993a) is consistent with the idea that SR factors may participate to some cell- or tissue-specific regulation processes. In conclusion, we have identified a novel splicing factor which extends further the already large family of SR splicing factors. The 9G8 factor, by virtue of its specific characteristics, does not appear to have originated from a recent duplication of other SR factor genes. Detailed studies on the expression of the 9G8 factor compared with that of other SR splicing factors and its ability to modulate alternative splicing will define further the specific roles played by this factor.

Materials and methods

Antibodies

Hybridoma cell lines were obtained after immunizing BALB/c mice with purified hnRNP as described previously (Lutz *et al.*, 1988). Both hybridoma (9G8 and 5G4) which produce IgG1 antibodies were cultured in RPMI 1640 supplemented with 10% FCS, or expanded as ascites tumors in syngenic mice. The antibody 5G4 recognizes a 110 kDa nucleolar protein not involved in RNA splicing. For direct addition to *in vitro* splicing assays, the antibodies 9G8 and 5G4 were purified from ascites fluids using Bakerbond ABx (JT Baker). Antibodies were eluted with a linear gradient from 10 mM MES pH 5.6 to 100 mM KH_2PO_4 pH 7.4, as recommended by the manufacturer, and were concentrated by ammonium sulfate precipitation. For immunopurification of the 9G8 protein and immunodepletion, we used 9G8 and 5G4 antibodies from ascites fluids which were precipitated with 50% saturated ammonium sulfate. The pellets were redissolved in PBS and extensively dialysed. The immunoaffinity matrix was prepared by coupling 35–40 mg of antibodies per ml of swollen cyanogen bromide-activated Sepharose 4B CL according to the manufacturer (Pharmacia).

Cytoplasmic and nuclear extract preparation and immunodepletion

Nuclear and cytoplasmic S100 extracts were prepared from HeLa, HeLa S3 or 293 cells as described by Dignam *et al.* (1983). Nuclear extracts were either dialysed against buffer D for splicing assays or kept in buffer C to be used for subsequent immunodepletion or purification of 9G8 protein. For depletion, 300 μl of nuclear extracts were adjusted to 0.42 M NaCl and were incubated for 1 h at 0–4°C with 270 μl of anti-9G8 Sepharose. After centrifugation, the depleted nuclear extract was dialysed against buffer D for 3 h. Mock-depleted nuclear extract was prepared under the same conditions, with the 5G4-coupled Sepharose.

Splicing reactions and complementation assays

Splicing reactions were carried out in standard 25 μl reaction assays, with 10^5 c.p.m. labeled substrate corresponding to ~4 ng RNA, in the presence of 3.2 mM $MgCl_2$ and 60 mM KCl, as described previously (Schmitt *et al.*, 1987). After 2 h incubation at 30°C the reaction was stopped and the RNA was purified and analyzed on 5.2 or 6.0% polyacrylamide gels in 8 M urea, as described previously (Schmitt *et al.*, 1987). Antibody inhibition reactions were carried out in the same splicing reaction mixture as above, except that the labeled substrate was included only after a 30 min preincubation on ice with the indicated amounts of purified antibodies. Complementation assays of 9G8- or mock-depleted extracts or S100 fractions with recombinant 9G8 or mock fractions were carried out in standard splicing conditions. The amounts of nuclear extracts, S100 fractions, mAbs or recombinant proteins used are indicated in each figure legend.

Protein purification and peptide sequencing

Immunopurification of the 9G8 protein was carried out at 4°C starting from 5 ml of 293 cell nuclear extract that was then diluted 4-fold and adjusted to 20 mM HEPES pH 7.9, 0.9 M NaCl, 0.1 M KCl, 0.25% NP-40 and 5% glycerol. The diluted nuclear extract was passed three times through a 1 ml column of anti-9G8–Sepharose beads. After several washes, the retained material (~50% of the initial 9G8 protein) was eluted at pH 3 with a step gradient containing sodium citrate/urea at the following concentrations: 70 mM/2.1 M, 80 mM/2.4 M, 90 mM/2.7 M (two column

volumes each). Fractions containing the 9G8 protein were immediately neutralized with 1 M Tris-base, dialysed against buffer D and concentrated by precipitation with 60% saturated ammonium sulfate or 20 mM $MgCl_2$. An enriched fraction containing ~10 μg 9G8 protein was applied to an SDS−polyacrylamide gel and transferred to PVDF membrane (Immobilon-Millipore) by electroblotting. The portion of membrane bearing the 9G8 protein, as detected by Coomassie blue coloration, was digested *in situ* with trypsin. The proteolytic peptides were resolved by HPLC and submitted subsequently to microsequencing.

The human SC35 protein was purified as described (Fu and Maniatis, 1992b), except that Mono Q fractions containing SC35 were precipitated with 20 mM $MgCl_2$ and that the last step of centrifugation on a glycerol gradient was not performed. Recombinant ASF was expressed in *E.coli* with the pDS56-ASF plasmid (a kind gift from J.Manley) and purified as described (Ge *et al.*, 1991). Nuclear hnRNP were isolated according to the protocol described previously (Stévenin *et al.*, 1982).

Cloning and sequencing of 9G8 cDNA

Degenerate 17 base oligonucleotides deduced from the upstream (VYVGNLGTGAGK) and downstream (TVWIAR) peptides sequenced initially were used to prime cDNA synthesis from 1 μg HeLa cell poly(A)$^+$ RNA, and then were used for PCR. The resulting DNA fragment of 92 bp was chemically sequenced and used to design QE203, a 38 nt probe spanning positions 153−190 of the cDNA (Figure 3). It was labeled at the 5′ end with ^{32}P, and used to screen a human placenta cell genomic library in λ GEM12. Duplicate plaque lifts were prepared and probed on duplicate filters as described (Sambrook *et al.*, 1989). Two positive genomic clones containing inserts of 17 and 15 kb were isolated from 4 × 10^5 recombinant phages. After subcloning of the 17 kb insert into the vector pBluescript, sequencing downstream and upstream from the region corresponding to the 92 bp PCR-generated fragment was performed on plasmid DNA using the dideoxy method with custom-designed oligonucleotide primers and Sequenase DNA polymerase (Sambrook *et al.*, 1989). A new probe, QK7, from positions 276 to 314, was designed to screen a 293 cell random-primed cDNA library in λ ZAP II vector (Stratagene). Nine positive phages were obtained from 4 × 10^5 plaques and their cDNA inserts were rescued in pBluescript by *in vivo* excision, as described by the manufacturer. The inserts were characterized by restriction enzyme mapping and several were completely (the clone 3) or partially sequenced on both strands.

In vitro transcription and translation of SR factors

The synthesis of pre-mRNA substrates was performed from a plasmid Sp4, which contains the major part of the E1A unit of adenovirus 2 (Schmitt *et al.*, 1987), a plasmid which contains the rabbit β-globin gene from position −9 to 1201 (Gattoni *et al.*, 1991), and the MINX plasmid, which contains duplication of leader 2 of the major late transcription unit of adenovirus separated by a reconstituted intron (Zillmann *et al.*, 1988). For *in vitro* transcription, these plasmids were cleaved downstream from the inserted sequences, except for the β-globin plasmid which was cleaved at a *Bam*HI site at the end of exon 2 or at a *Hinc*II site 162 bp downstream from exon 2. Labeled capped RNAs were prepared by transcription using SP6 RNA polymerase incorporating [α^{32}P]CTP, and were purified as described previously (Schmitt *et al.*, 1987).

The 9G8, SF2 and PR264 coding sequences were inserted downstream from the T7 promoter in pTL2(glo), pET9c and pBluescript vectors, respectively, to generate pTL2(glo)-9G8, pET9c-SF2RS (a gift from A.Krainer) and pBS-PR264 (a gift from B.Perbal). pTL2(glo) is a derivative of the pTL2 vector (Perez *et al.*, 1993), which contains the β-globin 5′ untranslated sequence inserted upstream of several cloning sites. An *Xho*I−*Hin*dIII fragment of the 9G8 cDNA clone 3 (see above), extending from the *Xho*I site 51 bp upstream of the ATG codon to the *Hin*dIII cloning site of pBluescript, was inserted between the corresponding sites of pTL2(glo). Capped RNA were transcribed from linearized plasmids by T7 RNA polymerase incorporating [^{35}S]CTP at low specific activity.

In vitro translation was performed using reticulocyte lysate or wheat germ extract minus methionine in the presence of 1 μg of each RNA and of unlabeled or 20 μCi of ^{35}S-labeled methionine in a 50 μl reaction mixture according to the manufacturer's instructions (Promega). After incubation at 30 or 25°C for 1 h, RNA was digested by micrococcal nuclease and an equal volume of 2 × SDS loading buffer was added (Laemmli, 1970).

Protein analysis and immunoblotting

Various protein preparations were analyzed on an 11.5% SDS−polyacrylamide gel as described by Laemmli (1970). The separated proteins were visualized either by Coomassie blue staining, fluorography or immunoblotting. For immunoblots, the proteins were electrotransferred to nitrocellulose and the blots were blocked in 1 × PBS containing 3% BSA for 1 h. Antibody treatment was by a 16 h incubation at 4°C in the presence of anti-9G8 hybridoma culture supernatant (dilution 1:10), followed by successive 1 h incubations at room temperature with rabbit anti-mouse IgG + IgM (1 mg/ml; Jackson Immuno Research) diluted 1:2000 and with ^{125}I−protein A (Amersham) diluted 1:1000. Each dilution was performed in 1 × PBS + Tween 0.1%, and the incubations were followed by a 20 min wash in this buffer.

Expression of recombinant 9G8 with baculovirus system

To insert the 9G8 coding sequence into the non-fusion transfer vector pVL1392 (Invitrogen), a PCR-generated fragment was synthesized from a full-length 9G8 cDNA, using an upstream primer including the sequence CGGCCGCCATGGCT, to create an *Eag*I site upstream from the ATG codon and change the second amino acid from serine to alanine to improve the translation efficiency. The downstream primer included *Eco*RV, *Hin*dIII and *Bam*HI cloning sites. The *Eag*I−*Bam*HI fragment was cloned into the *Not*I and *Bam*HI sites of pVL1392. Preparation and purification of recombinant virus expressing the 9G8 protein was as described by Summers and Smith (1987). Sf9 cells were infected with 2−5 p.f.u./cell and the infected cells were recovered after 2 days. These were used to prepare crude cytoplasmic and nuclear fractions, according to a protocol similar to that of Dignam *et al.* (1983), except that the cytoplasmic fraction obtained after cell lysis was immediately dialyzed against buffer D and then centrifuged at 13 000 r.p.m. for 4 min in a Sorvall-HB4 rotor. The cytoplasmic fractions containing recombinant 9G8 or estrogen receptor (ER, a gift from H.Gronemeyer), used as a control, were precipitated with 60% saturated ammonium sulfate, and the precipitated proteins were dissolved in buffer D and dialyzed extensively against the same buffer. Nuclear fractions containing recombinant 9G8 were also used to immunopurify 9G8 under stringent conditions, as described above.

Computer analysis

Sequence analysis was performed using the University of Wisconsin GCG package (Devereux *et al.*, 1984). 9G8 peptidic sequence was compared with entries in the PIR-Protein (release 36.0) and SWISS-PROT (release 25.0) databases by the FASTA screening program.

Acknowledgements

We are grateful to Dr N.Foulkes for critical reading of the manuscript. We thank G.Hildwein for excellent technical assistance and the cell culture group for growing human cells. Thanks to A.Staub, J.M.Garnier, F.Ruffenach, P.Eberling and G.Eisenmann for technical help, B.Boulay and J.M.Lafontaine for help with photography, and Y.Lutz who participated in the isolation of the 9G8 hybridoma and produced the antibodies. We are also grateful to Drs A.Krainer (Cold Spring Harbor Laboratory), J.Manley (Columbia University), B.Perbal (Université Paris VI), H.Gronemeyer and P.Kastner (our Laboratory) for gifts of cDNA clones and T.Maniatis (Harvard University) for the gift of SC35 antibody. This work was supported by funds from the Institut National de la Santé et de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Centre Hospitalier Universitaire Régional and the Association pour la Recherche sur le Cancer.

References

Ast,G., Goldblatt,D., Daniel,O., Sperling,J. and Sperling,R. (1991) *EMBO J.*, **10**, 425−432.

Ayane,M., Preuss,U., Köhler,G. and Nielsen,P.J. (1991) *Nucleic Acids Res.*, **19**, 1273−1278.

Behrens,S.-E., Galisson,F., Legrain,P. and Lührmann,R. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 8229−8233.

Bennett,M. and Reed,R. (1993) *Science*, **262**, 105−108.

Bennett,M., Michaud,S., Kingston,J. and Reed,R. (1992) *Genes Dev.*, **6**, 1986−2000.

Bentley,R.C. and Keene,J.D. (1991) *Mol. Cell. Biol.*, **11**, 1829−1839.

Brosi,R., Gröning,K., Behrens,S.-E., Lührmann,R. and Krämer,A. (1993a) *Science*, **262**, 102−105.

Brosi,R., Hauri,H.-P. and Krämer,A. (1993b) *J. Biol Chem.*, **268**, 17640−17646.

Caceres,J.F. and Krainer,A.R. (1993) *EMBO J.*, **12**, 4715−4726.

Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387−395.

Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475−1489.

Dingwall,C. and Laskey,R.A. (1991) *TIBS*, **16**, 478–481.
Eperon,I.C., Ireland,D.C., Smith,R.A., Mayeda,A. and Krainer,A.R. (1993) *EMBO J.*, **12**, 3607–3617.
Frank,D. and Guthrie,C. (1992) *Genes Dev.*, **6**, 2112–2124.
Fu,X.-D. (1993) *Nature*, **365**, 82–85.
Fu,X.-D. and Maniatis,T. (1990) *Nature*, **343**, 437–441.
Fu,X.-D. and Maniatis,T. (1992a) *Proc. Natl Acad. Sci. USA*, **89**, 1725–1729.
Fu,X.-D. and Maniatis,T. (1992b) *Science*, **256**, 535–538.
Fu,X.-D., Mayeda,A., Maniatis,T. and Krainer,A.R. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 11224–11228.
Gattoni,R., Chebli,K., Himmelspach,M. and Stévenin,J. (1991) *Genes Dev.*, **5**, 1847–1858.
Ge,H. and Manley,J.L. (1990) *Cell*, **62**, 25–34.
Ge,H., Zuo,P. and Manley,J.L. (1991) *Cell*, **66**, 373–382.
Gentz,R., Chen,C.-H. and Rosen,L.A. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 821–824.
Green,M.R. (1991) *Annu. Rev. Cell Biol.*, **7**, 559–599.
Green,L.M. and Berg,J.M. (1989) *Proc. Natl Acad Sci. USA*, **86**, 4047–4051.
Guthrie,C. (1991) *Science*, **253**, 157–163.
Kenan,D.J., Query,C.C. and Keene,J.D. (1991) *TIBS*, **16**, 214–220.
Kim,Y.-J., Zuo,P., Manley,J.L. and Baker,B.S. (1992) *Genes Dev.*, **6**, 2569–2579.
Krainer,A.R., Conway,G.C. and Kozak,D. (1990a) *Genes Dev.*, **4**, 1158–1171.
Krainer,A.R., Conway,G.C. and Kozak,D. (1990b) *Cell*, **62**, 35–42.
Krainer,A.R., Mayeda,A., Kozak,D. and Binns,G. (1991) *Cell*, **66**, 383–394.
Krämer,A. and Utans,U. (1991) *EMBO J.*, **10**, 1503–1509.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Legrain,P. and Chapon,C. (1993) *Science*, **262**, 108–110.
Legrain,P. and Choulika,A. (1990) *EMBO J.*, **9**, 2775–2781.
Li,H. and Bingham,P.M. (1991) *Cell*, **67**, 335–342.
Lutz,Y., Jacob,M. and Fuchs,J.-P. (1988) *Exp. Cell. Res.*, **175**, 109–124.
Mattaj,I.W. (1993) *Cell*, **73**, 837–840.
Mayeda,A., Zahler,A.M., Krainer,A.R. and Roth,M.B. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 1301–1304.
McKeown,M. (1993) *Curr. Opin. Cell Biol.*, **5**, 448–454.
Moore,M.J., Query,C.C. and Sharp,P.A., (1992) In Gesteland,R. and Atkins,J. (eds), *RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 1–30.
Patton,J.G., Porro,E.B., Galceran,J., Tempst,P. and Nadal-Ginard,B. (1993) *Genes Dev.*, **7**, 393–406.
Perez,A., Kastner,P., Sethi,S., Lutz,Y., Reibel,C. and Chambon,P. (1993) *EMBO J.*, **12**, 3171–3182.
Rajavashisth,T.B., Taylor,A.K., Andalibi,A., Svenson,K.L. and Lusis,A.J. (1989) *Science*, **245**, 640–643.
Roth,M.B., Zahler,A.M. and Stolk,J.A. (1991) *J. Cell Biol.*, **115**, 587–596.
Ruby,S.W. and Abelson,J. (1991) *TIG*, **7**, 79–85.
Ruskin,B., Zamore,P.D. and Green,M.R. (1988) *Cell*, **52**, 207–219.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning, A Laboratory Manual, 2nd edn*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Scherly,D., Boelens,W., Dathan,N.A., van Venrooij,W.J. and Mattaj,I.W. (1990) *Nature*, **345**, 502–506.
Schmitt,P., Gattoni,R., Keohavong,P. and Stévenin,J. (1987) *Cell*, **50**, 31–39.
Stévenin,J., Gattoni,R., Keohavong,P. and Jacob,M. (1982) *J. Mol. Biol.*, **155**, 185–205.
Summers,M.D. and Smith,G.E. (1987) *Texas Agricultural Experiment Sation Bulletin*, **1555**.
Tazi,J., Kornstädt,U., Rossi,F., Jeanteur,P., Cathala,G., Brunel,C. and Lührmann,R. (1993) *Nature*, **363**, 283–286.
Tian,M. and Maniatis,T. (1993) *Cell*, **74**, 105–114.
Vellard,M., Soret,J., Sureau,A. and Perbal,B. (1991) *C. R. Acad. Sci. Paris*, **313**, 591–597.
Vellard,M., Sureau,A., Soret,J., Martinerie,C. and Perbal,B. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 2511–2515.
Wu,J.Y. and Maniatis,T. (1993) *Cell*, **75**, 1061–1070.
Zahler,A.M., Lane,W.S., Stolk,J.A. and Roth,M.B. (1992) *Genes Dev.*, **6**, 837–847.
Zahler,A.M., Neugebauer,K.M., Lane,W.S. and Roth,M.B. (1993a) *Science*, **260**, 219–222.
Zahler,A.M., Neugebauer,K.M., Stolk,J.A. and Roth,M.B. (1993b) *Mol. Cell. Biol.*, **13**, 4023–4028.
Zamore,P.D., Patton,J.G. and Green,M.R. (1992) *Nature*, **355**, 609–614.
Zillmann,M., Zapp,M.L. and Berget,S.M. (1988) *Mol. Cell. Biol.*, **8**, 814–821.
Zuo,P. and Manley,J.L. (1993) *EMBO J.*, **12**, 4727–4737.

Received on February 1, 1994; revised on March 11, 1994