Recognition of exonic splicing enhancer sequences by the Drosophila splicing repressor RSF1
Emmanuel Labourier, Eric Allemand, Stéphanie Brand, Maggy Fostier, Jamal Tazi, Henri-Marc Bourbon

To cite this version:

HAL Id: hal-02462726
https://hal.archives-ouvertes.fr/hal-02462726
Submitted on 31 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Recognition of exonic splicing enhancer sequences by the *Drosophila* splicing repressor RSF1

Emmanuel Labourier, Eric Allemand, Stéphanie Brand\(^1\), Maggy Fostier\(^1,+\), Jamal Tazi* and Henri-Marc Bourbon\(^1\)

Institut de Génétique Moléculaire, UMR5535 du CNRS, 1919 Route de Mende, F34293 Montpellier Cedex 5, France and \(^1\)Centre de Biologie du Développement, UMR5547 du CNRS, Université Paul Sabatier, 118 Route de Narbonne, F31062 Toulouse, France

Received January 27, 1999; Revised March 19, 1999; Accepted April 18, 1999

**ABSTRACT**

The *Drosophila* repressor splicing factor 1 (RSF1) comprises an N-terminal RNA-binding region and a C-terminal domain rich in glycine, arginine and serine residues, termed the GRS domain. Recently, RSF1 has been shown to antagonize splicing factors of the serine/arginine-rich (SR) family and it is, therefore, expected to play a role in processing of a subset of *Drosophila* pre-mRNAs through specific interactions with RNA. To investigate the RNA-binding specificity of RSF1, we isolated RSF1-binding RNAs using an *in vitro* selection approach. We have identified two RNA target motifs recognized by RSF1, designated A (CAACGAC-GA) and B (AAACGGCGCG)-type sequences. We show here that the A-type cognate sequence behaves as an SR protein-dependent exonic splicing enhancer. Namely, three copies of the A-type ligand bind SR proteins, stimulate the efficiency of splicing of reporter pre-mRNAs several fold and lead to inclusion of a short internal exon both *in vitro* and *in vivo*. However, three copies of a B-type ligand were much less active. The finding that RSF1 acts as a potent repressor of pre-mRNA splicing *in vitro* led us to propose that the equilibrium between a limited number of structurally-related general splicing activators or repressors, competing for common or promiscuous binding sites, may be a major determinant of the underlying mechanisms controlling many alternative pre-mRNA processing events.

**INTRODUCTION**

The accurate excision of intervening sequences from RNA polymerase II transcripts is crucial for the expression of most metazoan genes. Recently, progress has been made in the identification of both *cis*-acting elements required for regulation and *trans*-acting factors that interact with such elements (1,2). The *trans*-acting pre-mRNA splicing factors include the U1, U2, U4/U6 and U5 small ribonucleoprotein particles (snRNPs) and non-snRNP proteins. These components assemble onto the pre-mRNA in a dynamic fashion to form a large ribonucleoprotein complex, the so-called spliceosome, where intron excision occurs by two transesterification reactions (3). The earliest detectable metazoan pre-spliceosome complex (E) is formed in an ATP-independent manner (4). It contains the non-snRNP splicing factor U2AF (U2 snRNP auxiliary factor) which comprises two subunits (U2AF 35 kDa and 65 kDa), U1 snRNP and several other proteins (5,6). The assembly of this complex is a major control point for the initial recognition and pairing of splice sites (7) and is therefore thought to be an important step in the regulation of alternative splicing (6). Among the proteins that contribute to the formation of the E complex, there are members of the serine/arginine-rich (SR) protein family that are known to influence the splice site choice (8–11). These proteins also bind a class of purine-rich splicing enhancers known as exon splicing elements (ESEs), that have been demonstrated to play a role in both alternative and constitutive splice site selection in several experimental systems (12–22).

SR proteins share characteristic structural features with several metazoan splicing factors involved in splice site selection. They all contain RNA recognition and arginine-serine (RS)-rich sequence motifs (8–10) that are essential for their function as splicing factors (23–25). The RS domain is responsible for specific protein–protein interactions between RS domain-containing proteins (26–28), interactions which are thought to constitute a bridge between 5′ and 3′ splice sites during splice site selection (26–28). Such interactions promote the binding of U1 snRNP to the 5′ splice site and U2 snRNP auxiliary factor (U2AF\(^{65}\)) to the 3′ splice site at the earliest stages of the spliceosome assembly (6,11). The SR protein family now includes nine identified members: SRp20 (X16 or RBP1), SRp30a, SRp30b, SRp30c, 9G8, SRp40, SRp54, SRp55 (B52) and SRp75.

Recently, we provided evidence that the ability of SR proteins to promote early splicing complexes and U1 snRNP binding to the 5′ splice site can be selectively counteracted by a splicing repressor called RSF1 (29). Like SR proteins, RSF1 has a modular structure consisting of a single N-terminal RNA recognition motif (RRM) domain and a C-terminal domain.
enriched for glycine, arginine and serine residues (GRS domain) (30). The GRS domain mediates specific interactions with itself and with members of the SR protein family (29), whereas the RRM domain is required for RNA binding. In this article, we show that RSF1 cognate sequences selected from a pool of randomized sequences have ESE properties, suggesting that RSF1 suppresses pre-mRNA splicing by interfering with normal bridging and recruitment activities of SR proteins.

MATERIALS AND METHODS

Oligonucleotides

The sequences of the synthetic oligonucleotides (Isoprime SA, Toulouse) used in this study as cloning adaptors or PCR primers are (name, sequence is given 5′→3′):

- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB4T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB2T1, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
- HB1T2, CTACCCCTGACCCAGAGCTTGTCTTTAGCCG.
cross-linked samples with proteinase K was performed with 2 µg of proteinase K and 0.1% SDS for 30 min at 42°C.

The splicing reactions were done under standard conditions for 1.5 h in a total volume of 20 µl containing 50 fmol labeled pre-mRNA, 6 µl of HeLa nuclear extract (35) complemented with buffer D. In Figure 4A, cold competitor RNAs were incubated in HeLa NE for 5 min at room temperature before addition of radiolabeled pre-mRNAs. Splicing products were analyzed by electrophoresis on denaturing 6% (Fig. 4 A) or 7% (Fig. 5 C and D) polyacrylamide gels and revealed by autoradiography.

Transfection and RT–PCR

HindIII–BamHI fragments from pDUP51 T, 3A, 3B and 3S were subcloned between HindIII and BamHI sites of pJ6Ω (36) to obtain pJ6DUP51 T, 3A, 3B and 3S expression vectors. Monolayer Drosophila S2 cells or HeLa cells were grown on 3 cm diameter dishes (Nunc) to 70–80% confluence and transfected, S2 cells or HeLa cells were grown on 3 cm diameter dishes (Nunc) to 70–80% confluence and transfected, resulting in a large molar excess of a pool of 64-base synthetic RNA molecules containing a randomized region of 25 bases flanked by constant regions for primer annealing. After 5 cycles of selection–amplification, cDNA fragments corresponding to selected RNAs were cloned and the internal sequences of 40 independent clones were determined. The vast majority (i.e. 37 sequences) of inserts contained purine-rich sequences from which two partially-related consensus sequences were derived and designated A (YAWCGACRR) and B (AAWCGCGGYG) (Y is C or U, R is A or G and W is A or U; see Fig. 2). Several clones had more than one copy of these nonamer motifs, and in some cases contained both A and B sequences. It is possible that the latter clones resulted from cooperative binding of RSF1 homodimers, given that RSF1 could interact with itself through protein–protein interaction mediated by the GRS domain (29).

Sequence alignment allowed us to design best-guessed high-affinity RSF1 binding sites: CAACGACGA for A-type sequences and AAACGCGCG for B-type sequences. To assess the affinities and specificities of binding of these ‘winner’ sequences to RSF1, we developed a band shift assay. Radiolabeled 47 nt RNA probes, containing three tandemly-repeated A- or B-type nonamers (termed 3A and 3B, respectively), were mixed with increasing amounts of purified GST–RSF1 (Fig. 1B, lane 2) and mixes were analyzed by native polyacrylamide gel electrophoresis. As shown in Figure 3A, stable complexes were formed with both 3A (panel I, lanes 1–3) and 3B (panel II, lanes 1–3) RNAs but not with a control RNA, indicating that RSF1 could interact with itself through protein–protein interaction mediated by the GRS domain (29).

RESULTS

In order to study the RNA-binding specificity of RSF1, we performed an iterative in vitro selection (SELEX) from a pool of random sequences (37,38). Full-length RSF1 protein (Fig. 1B, lane 2) immobilized on glutathione–agarose beads was used as a selection matrix. In vitro selection was carried out using a large molar excess of a pool of 64-base synthetic RNA molecules containing a randomized region of 25 bases flanked by constant regions for primer annealing. After 5 cycles of selection–amplification, cDNA fragments corresponding to selected RNAs were cloned and the internal sequences of 40 independent clones were determined. The vast majority (i.e. 37 sequences) of inserts contained purine-rich sequences from which two partially-related consensus sequences were derived and designated A (YAWCGACRR) and B (AAWCGCGGYG) (Y is C or U, R is A or G and W is A or U; see Fig. 2). Several clones had more than one copy of these nonamer motifs, and in some cases contained both A and B sequences. It is possible that the latter clones resulted from cooperative binding of RSF1 homodimers, given that RSF1 could interact with itself through protein–protein interaction mediated by the GRS domain (29).

Sequence alignment allowed us to design best-guessed high-affinity RSF1 binding sites: CAACGACGA for A-type sequences and AAACGCGCG for B-type sequences. To assess the affinities and specificities of binding of these ‘winner’ sequences to RSF1, we developed a band shift assay. Radiolabeled 47 nt RNA probes, containing three tandemly-repeated A- or B-type nonamers (termed 3A and 3B, respectively), were mixed with increasing amounts of purified GST–RSF1 (Fig. 1B, lane 2) and mixes were analyzed by native polyacrylamide gel electrophoresis. As shown in Figure 3A, stable complexes were formed with both 3A (panel I, lanes 1–3) and 3B (panel II, lanes 1–3) RNAs but not with a control RNA, indicating that RSF1 could interact with itself through protein–protein interaction mediated by the GRS domain (29).

Sequence alignment allowed us to design best-guessed high-affinity RSF1 binding sites: CAACGACGA for A-type sequences and AAACGCGCG for B-type sequences. To assess the affinities and specificities of binding of these ‘winner’ sequences to RSF1, we developed a band shift assay. Radiolabeled 47 nt RNA probes, containing three tandemly-repeated A- or B-type nonamers (termed 3A and 3B, respectively), were mixed with increasing amounts of purified GST–RSF1 (Fig. 1B, lane 2) and mixes were analyzed by native polyacrylamide gel electrophoresis. As shown in Figure 3A, stable complexes were formed with both 3A (panel I, lanes 1–3) and 3B (panel II, lanes 1–3) RNAs but not with a control RNA, indicating that RSF1 could interact with itself through protein–protein interaction mediated by the GRS domain (29).
Figure 2. In vitro selection–amplification of high-affinity RNA target sequences for GST–RSF1. (A) The sequences of individual clones after five cycles are shown distributed among two sets according to the presence of one or two motifs resembling the nonamer YAWCGACRR (R = A or G, W = A or T, Y = C or T) or the related nonamer AAWCGCGYR. These motifs are referred to as A or B nonamers, respectively, and are represented in bold type characters. Nucleotides shown in lower case characters belong to the flanking constant regions. Numbers between brackets behind the name of each clone indicate the number of mismatches (up to three were allowed) between the nonamer motifs present in these sequences and the deduced consensus (in which each indicated nucleotide was found in at least half of the A- or B-type nonamers). (B) RSF1 and a selected range of SR proteins share highly related RRM. The RRM domain of RSF1 is compared with homologous domains of human SRp20 (DDBJ/EMBL/GenBank accession no. 338484), human SRp55 (DDBJ/EMBL/GenBank accession no. 3929380), human SRp30b (DDBJ/EMBL/GenBank accession no. 3929379), human SRp30c (DDBJ/EMBL/GenBank accession no. 730826), and human SC35 (DDBJ/EMBL/GenBank accession no. 266992). The amino acid sequences of the RRMs of RSF1 and SR proteins are aligned using BESTFIT program. Identical matches are boxed in reverse type and functionally related matches are boxed in grey. Gaps are introduced in the amino acid sequences in order to obtain an optimal alignment and are shown with horizontal bars. Pairwise percent amino acid identities versus similarities to RSF1 are indicated between brackets in the right-hand column.

whereas the control sequence did not (Fig. 3B). These data provide further evidence that RSF1 binds specifically to these consensus selected sequences.

Given that RSF1 has a modular structure with an N-terminal RRM domain and C-terminal GRS domain, both of which are capable of binding homopolymers (data not shown), we decided to determine the contribution of individual domains for the binding of RSF1 to cognate RNAs. RSF1–RNA interaction probably involves the RRM domain of RSF1, because a truncated version of RSF1 in which the RRM domain was selectively removed (GST–RSF1ΔN, Fig. 1B, lane 4) failed to bind 3A (Fig. 3A, panel I, lanes 7–9) or 3B (Fig. 3A, panel II, lanes 7–9) sequence. In contrast, a RSF1 mutant lacking the GRS domain (GST–RSF1ΔC, Fig. 1B, lane 3) efficiently bound to these sequences (Fig. 3A, panels I and II, lanes 4–6, respectively), implying that the GRS domain does not contribute to RSF1–RNA interaction. Critical comparison of the band shift patterns obtained with GST–RSF1 and GST–RSF1ΔC showed that the latter formed three complexes of different mobilities and the former only one. This result could be explained if the 3A or 3B sequence binds three molecules of GST–RSF1ΔC simultaneously (Fig. 3A, panels I and II, lanes 4–6, respectively) but only one molecule of GST–RSF1. These data validate the SELEX experiments and further indicate that the RRM per se, with very few neighboring residues, is necessary and sufficient to interact with high affinity to single nonamer sequences.

The similarity of RSF1 binding sites with purine-rich ESEs, the fact that RSF1 and a range of SR proteins share closely related
RSF1 specifically interacts through its RRM domain with 3A and 3B purine-rich sequences. (A) Mobility shift assay of GST–RSF1. Radiolabeled RNA containing three copies of the winner A- or B-type RSF1 high-affinity binding sequence (3A and 3B RNA, panels I and II, respectively) was incubated with increasing amounts (0.5, 2 or 8 pmol) of the recombinant purified protein indicated. Complexes were separated from the free probe on a 8% non-denaturing polyacrylamide gel. No shift was observed with a control probe (T RNA) derived from β-globin exon 2 pre-mRNA (panel III). A rough estimate of the affinity of the binding of GST–RSF1 to 3B or 3A RNA by measuring the percentage of complexed probes gives a calculated equilibrium dissociation constant ($K_d$) of $\sim 10^{-8}$ M. (B) Competition assay with 2 pmol of GST–RSF1 and 3A RNA as radiolabeled probe. Each competitor RNA (T, 3A or 3B) was used at 20- (lanes 2, 4 and 6) or 100-fold (lanes 3, 5 and 7) molar excess over the probe.

To identify the factor(s) that interact(s) with GST–RSF1 target sequences, we used UV cross-linking assays. Radiolabeled probes corresponding to the 3A or 3B RNA were cross-linked in the presence of HeLa nuclear extract, and labeled proteins were analyzed by electrophoresis on SDS–polyacrylamide gels and visualized by autoradiography. The 3A RNA transferred label predominantly to protein species with apparent molecular weights of ~20, 30, 40, 55, 75 and 130 kDa (Fig. 4 A, lanes 1). Allowing for the 130 kDa protein which was only seen in nuclear extract, protein bands with identical mobility also appeared when cross-linking experiments were performed with either purified SR proteins (Fig. 4 B, panel II, lane 1) or nuclear extract supplemented with purified SR proteins (Fig. 4 B, panel II, lane 6). This result indicates that the 20, 30, 40, 55 and 75 kDa bands are probably SRp20, SRp30, SRp40, SRp55 and SRp75, respectively. Although the amount of cross-linking of these proteins correlated with their abundance in the extract, indicating that the 3A RNA has no preference for binding any of the SR proteins, a very strong signal was observed at the level of SRp30. Since at least four SR proteins are expected to have the same electrophoretic behavior at 30 kDa, it is possible that one or several SRp30 species make strong contact with 3A RNA. A likely candidate could be 9G8 protein from human, because a recent SELEX performed with this protein resulted in purine-rich sequences that resemble the 3A RNA consisting mainly of a repetition of three GAC triplets. Accordingly, comparison of the RRM domain of RSF1 with those of SR proteins revealed that the highest homology score (56% identity and 68% similarity) was detected over the RRM of 9G8 from human (Fig. 2 B). The cross-linking of the 3A RNA to SR proteins was likely to be specific, because it was not competed out with a large excess of...
Figure 4. Splicing of \(\beta\)-globin pre-mRNA in vitro is inhibited by competing amounts of 3A but not 3B RNA. (A) Aliquots of 50 fmol of \(\beta\)-globin pre-mRNA (lane CTL) were incubated in HeLa nuclear extract under splicing conditions without (lane 1) or with 0.01, 0.05, 0.1, 0.5, 1 or 5 pmol of cold 3A or 3B competitor RNA (lanes 2–7 and 8–13, respectively). Splicing products (shown on the right of the panel) were analyzed on a 6% denaturing polyacrylamide gel. (B) UV cross-linking of SR proteins to 3A and 3B RNAs. Four micrograms of SR proteins purified from HeLa cells (lane 1), 0.5 \(\mu\)g of GST–RSF1\(\Delta\)N (lane 2), 0.5 \(\mu\)g of GST–RSF1 (lane 3), 100 \(\mu\)g of total proteins from HeLa nuclear extract (lanes 4–8) supplemented with 4 \(\mu\)g of SR proteins (lane 6), 2 \(\mu\)g of GST–RSF1\(\Delta\)N (lane 7) or 2 \(\mu\)g of GST–RSF1 (lane 8) were incubated for 10 min under splicing conditions but without polyvinyl alcohol, in the absence of any probe (panel I) or with 100 fmol of radiolabeled 3A or 3B RNA (panels II and III, respectively). After irradiation with UV light, reaction mixtures analyzed on a 10% SDS–polyacrylamide gel without prior treatment with RNase, were either stained with Coomassie Blue (panel I) or dried and autoradiographed (panels II and III) to reveal cross-linked proteins. Total SR proteins from HeLa cells were purified according to Zahler et al. (8). (C) Competition assay with 3A RNA as radiolabeled probe. UV cross-linking experiments were performed as in Figure 4B with 100 \(\mu\)g of total proteins from HeLa nuclear extract. Each competitor RNA (3S, 3A or 3B) was used at 10- (lanes 2, 4 and 6) or 50-fold (lanes 3, 5 and 7) molar excess over the probe. Only cross-linked SRp30 is shown.

nuclear proteins (Fig. 4B, panel II, lanes 4 and 6), whereas it was effectively competed by increasing amounts of an unlabeled 3A RNA (Fig. 4C, lanes 4 and 5). Interestingly, cross-linking to SR proteins was no longer detected when the nuclear extract was supplemented with purified GST–RSF1 (Fig. 4B, panel II, lane 8). Instead, two strong signals were produced by purified GST–RSF1 (Fig. 4B, panel II, compare lanes 3 and 8). The concentration of GST–RSF1 required to achieve this was almost identical to that of SR proteins in nuclear extract (quantitation of the amount of SR proteins in the extract was made by comparison with known amounts of purified SRp30), suggesting that the 3A RNA displays higher affinity for GST–RSF1 than for SR proteins. However, GST–RSF1 yields much less cross-linked product than SR proteins in HeLa extract (Fig. 4B, panel II, compare lanes 4 and 8). Given that GST–RSF1 also gives very poor cross-linking when purified proteins were used (compare...
Figure 5. 3A and 3B sequences act as ESEs in vitro. (A) Nucleotide sequences of T, 3A, 3B and 3S RNAs. (B) Schematic representation of the model wild-type human β-globin pre-mRNA and its derivatives containing the T, 3A, 3B or 3S sequence. Sequences derived from β-globin exon 1 or exon 2 are boxed in white and grey, respectively, and black boxes represent non-β-globin sequences (3A, 3B and 3S). Numbers on top indicate the length of the corresponding exonic and intronic regions and the angled lines represent the splicing events. (C) 3A and 3B sequences enhance splicing of a β-globin intron in vitro. Splicing reactions were performed under standard conditions with 30% HeLa nuclear extract and 50 fmol of the radiolabeled pre-mRNA indicated. Splicing products (depicted on either side of the panel) were analyzed on a 7% denaturing polyacrylamide gel and revealed by autoradiography. (D) 3A and 3B sequences promote exon inclusion in vitro. Reactions were performed as in Figure 5C with the indicated model three-exons β-globin pre-mRNAs. Splicing intermediates and products are depicted on either side of the panel.

lanes 1 and 3), it is possible that amino acids of GST–RSF1 contacting the 3A RNA are moderately reactive after excitation of the nucleic bases by UV light. The failure of GST–RSF1ΔN to compete with cross-linking of SR proteins to 3A RNA (Fig. 4B, panel II, lane 7) makes it very likely that the competition effect was mediated by RSF1's RRM domain. Consistent with this, purified GST–RSF1ΔN did not show any cross-linking to 3A RNA (Fig. 4B, panel II, lane 2).

Although 3B RNA was efficiently cross-linked to purified individual SR proteins (Fig. 4B, panel III, lane 1), a large excess of cold 3B RNA failed to compete with the cross-linking of SR proteins to 3A RNA in nuclear extract (Fig. 4C, lanes 6 and 7). Labeled 3B RNA produced no signal by SR proteins in nuclear extract (Fig. 4B, panel III, lane 4) but, instead, a smear plus a strong signal at the level of high-molecular-weight protein species. Neither purified SR proteins (Fig. 4B, panel III, lane 6) nor purified RSF1 (Fig. 4B, panel III, lane 8) could compete efficiently with the cross-linking of these proteins, indicating that the 3B RNA may display higher affinity for unknown masking proteins present in the nuclear extract. Among those the 130 kDa protein was a likely candidate. Failure of SR proteins to bind the 3B RNA in nuclear extract offers a plausible explanation of why splicing of β-globin pre-mRNA was not affected by a large excess of this RNA.

Purine-rich sequences that bind SR proteins have been shown to act as cis-acting regulatory elements for splicing (13,14,40,41). To better understand the meaning of the binding of SR proteins to the 3A RNA, we asked whether this site could function as a splicing enhancer. We used a β-globin derivative with a short second exon (Fig. 5B, β-T) as a model substrate.
Though it has the wild-type β-globin splice sites, this substrate was very poorly spliced in nuclear extract containing a low amount of SR proteins (Fig. 5C, lane 2). However, splicing efficiency could be altered if a splicing enhancer was introduced downstream from the 3′ splice site. Substitution in the second exon of an optimal binding site for SF2/ASF (Fig. 5B, β-3S), previously shown to constitute a powerful splicing enhancer in vitro (14), in place of β-globin sequences enhanced splicing dramatically (Fig. 5C, compare lanes 2, 5 and 7). An identical enhancing effect was also observed with the 3A sequence (Fig. 5C, lane 3), indicating that the latter can function as a strong splicing enhancer sequence when inserted within an exon. However, the 3B sequence, which did not inhibit splicing in trans, was less efficient at stimulating splicing in cis (Fig. 5C, lane 4). It is significant that 3S and 3A have comparable splicing enhancing activities (Fig. 5C, compare lanes 3 and 5), suggesting that they could bind trans-acting factors with similar affinity. Consistent with this view, cross-linking of isolated copies of the 3A sequence to SR proteins in nuclear extract was competed to the same extent with increasing amounts of unlabeled 3A or 3S competitor RNA (Fig. 4C, lanes 2–5).

To determine whether the GST–RSF1 winner sequences might be required for the regulation of alternative splicing, we used model pre-mRNAs containing two introns and three exons, derived from the human β-globin gene. These substrates have been successfully used to demonstrate ESE-dependent inclusion of a small internal exon (42). Therefore, four nucleotides of the first exon adjacent to the 5′ splice site, the first intron and the second exon of β-globin were inserted downstream of β-T, β-3A, β-3B or β-3S, generating four constructs each containing a 51 nt middle exon flanked by two identical introns, termed DUP51 T, DUP51 3A, DUP51 3B and DUP51 3S, respectively (Fig. 5B). Our constructs differ from those described by Dominski and Kole (42) as the middle exon contained more sequences from the second exon of β-globin. Importantly all four splice sites, as well as their immediately adjacent sequences, were the same in all the pre-mRNAs studied. Thus, splice site usage should reflect solely the influence of the introduced sequences. Splicing was assayed in vitro in HeLa nuclear extracts using 32P-labeled SP6 transcripts and in vivo during transient expression in transfected cells of the same constructs driven by a β-actin promoter. As previously observed for single-intron containing substrates, both DUP51 T and DUP51 3B were very poorly spliced in nuclear extract containing low levels of SR proteins (Fig. 5D, lanes 1 and 2). Interestingly, the patterns of splicing for these transcripts were characterized by the existence of slowly-migrating large superlariat structures which represent intermediate stages of alternative splice products. To demonstrate in vitro, PCR reactions were performed in the presence of two additional primers designed against human GAPDH mRNA. The DNA bands corresponding to unspliced pre-mRNA (651 bp), spliced mRNA (391 bp), skipped mRNA (340 bp) or GAPDH mRNA (228 bp) are marked on the right. (B) Same experiments using total RNA purified from transfected Drosophila S2 cells. As a control, PCR with two oligonucleotides specific for copia mRNA are also shown for each experiment. The DNA bands corresponding to β-globin pre-mRNA, its splicing products and copia mRNA are marked on the right.

to the superlariat containing the middle exon was detected. The number of RNA bands generated during splicing of the two transcripts was identical and correlated well with the electrophoretic properties of the intermediates and final products predicted for the regular splicing pathway of a two-intronic pre-mRNA. Their structures are depicted on the right in Figure 5D.

Patterns of splicing in vitro and in vivo may differ significantly and splice site selection in vitro may be affected by the concentration of the nuclear splicing factors in the extract. We therefore performed in vivo studies in cultured cells to ascertain the splicing pattern of the tested pre-mRNAs. The various DUP51 constructs, inserted within a suitable expression vector, were transfected in HeLa cells or Drosophila S2 cells and total RNAs were prepared 24 h post-transfection to avoid reduction of mRNA levels owing to cell death. The mRNA products of each transfected construct were examined by RT–PCR and normalized to GAPDH mRNA in HeLa cells or to copia retrotransposon mRNA in S2 cells. As shown in Figure 6, internal exons containing either a control β-globin sequence (T) or 3B were skipped (Fig. 6A and B, lanes 3–6) and only the 3B sequence allowed a very low level of exon inclusion in HeLa cells (Fig. 6A,
lanes 5 and 6) but not S2 cells (Fig. 6B, lanes 5 and 6). In sharp contrast, both the 3A and 3S sequences conferred exclusive inclusion of the middle exon in HeLa cells (Fig. 6A, lanes 7–10) and half of these exons were skipped in S2 cells (Fig. 6B, lanes 7–10), consistent with the role of these sequences as powerful and functional ESEs in living cells. Failure to observe exclusive inclusion of exons containing the 3S or 3A sequence following transfection of constructs in S2 cells was attributable to lower levels of SR proteins. Quantitation by western blot analysis revealed that extracts from S2 cells contained at least six times less SR proteins than extract from HeLa cells (data not shown). Thus, the presence in S2 cells of RSF1 and/or other as yet unidentified SR protein antagonists at a concentration that may easily displace the association of SR proteins with exon containing 3A, 3B or 3S sequences, could inhibit exon inclusion during splicing of these reporter pre-mRNAs. This, together with results described by Labourier et al. (29), is consistent with the idea that RSF1 acts as a sequence-specific pre-mRNA splicing repressor in vivo.

**DISCUSSION**

In this paper we have presented several lines of evidence suggesting that representative RSF1 RNA ligands could serve as exon recognition signals that promote exon inclusion. Two purine-rich sequences, A and B, were identified from a random sequence pool as high-affinity binding sites for RSF1. Mobility shift data obtained with full-length and truncated versions of recombinant RSF1 were consistent with the results of the SELEX experiment and demonstrated that its RRM domain was essential and sufficient to mediate specific interaction with these sequences. The finding that RSF1 bound much less efficiently SF2/ASF-selected sequences (data not shown), which were also purine-rich (14), makes it unlikely that RSF1 binds any polyurine-rich sequences but in fact recognizes short distinct sequences. However, there are differences in the behavior of RSF1-selected sequences. Three copies of the A-type sequence display specific cross-linking to SR proteins in HeLa nuclear extract, efficiently stimulate splicing of reporter pre-mRNAs and promote exon inclusion. Significantly, both in vitro and in transfected cells the reiterated A-type sequence acts in a manner indistinguishable from similarly arranged SF2/ASF high-affinity binding sites when similarly positioned within a model pre-mRNA. In sharp contrast three copies of the B-type sequence only had slight effects on these splicing events. Since purified SR proteins do bind the B-type sequence, we assume that in HeLa nuclear extract there are factors that bind the B-type sequence with higher affinity than SR proteins and thereby antagonize the function of SR proteins. These factors could be related to RSF1 because complementing HeLa nuclear extract with purified RSF1 displaces the binding of SR proteins to A-type sequences and impedes splicing of reporter pre-mRNAs (29). Experiments are currently underway to identify the human homolog of RSF1 in HeLa cell nuclear extract depleted of SR proteins, which could specifically interact with 3A or 3B targets.

Although additional work is needed to establish whether the sequences described here occur naturally in pre-mRNAs subject to alternative splicing, it is striking that several splicing silencer sequences (SSEs), including HIV regulatory sequences, contain motifs closely related to those selected by GST–RSF1 (43,44). Interestingly, in some cases the SSE sequences were found to be juxtaposed to ESEs, forming bipartite splicing regulatory elements that may control the overall efficiency of proximal splice site usage (18,45,46). Therefore, the binding of RSF1 to ESEs shown here could be a way to mediate splicing repression to prevent activation of cryptic splice sites within exonic sequences. In agreement with this hypothesis, it has recently been shown that RSFc, a protein in the dipteran Chironomus tentans, highly homologous to RSF1, binds extensively to newly transcribed BR1 and BR2 pre-mRNAs, which contain a huge exon of 30–35 kb but much less to the BR3 pre-mRNA, which largely consists of intron sequences (L. Wieslander, personal communication). Thus, while ESE elements are required to assure that constitutive splicing does take place, an appropriate balance of splicing repressors and splicing activators, expected to bind these sequences, controls splice site usage of suboptimal splicing signals.

How might RSF1 mediate splicing repression? Binding of this protein may interfere directly with the binding of SR proteins and/or their interactions with constitutive splicing factors. Several lines of evidence support a model for initial splice-site recognition in which multiple protein–RNA and protein–protein interactions between SR proteins bound to the exon and the 5′ and 3′ splice sites led to the formation of a stable complex (6,26–28). SR proteins have been shown to enhance interactions between the U1 snRNP and the 5′ splice site (27,47,48), interactions between U2AF and the 3′ splice site as well as bridging the 5′ and 3′ splice sites (27,28). They are also expected to bind ESE sequences to stimulate splicing of the upstream intron or inclusion of an internal exon (5,12,13,40,49–51). Accordingly, ESE sequences have been found in many cellular and viral pre-mRNAs which are subject to alternative splicing (13,15–21) as well as constitutively spliced pre-mRNAs (22,50,51). The finding that one such sequence also serves as a high-affinity binding site for RSF1, a splicing repressor, provides a framework for understanding the mechanism by which splicing repression can be mediated. The binding of RSF1 to ESEs will prevent recruitment of constitutive splicing factors (U1 snRNP and U2AF) to splice sites, and as a result will abrogate the formation of stable enhancer complexes. This view is consistent with the finding that RSF1 induces a dose-sensitive inhibition of splicing for several reporter pre-mRNAs, an inhibition that occurs at the level of early splicing complexes formation (29). The results of UV cross-linking assays with HeLa nuclear extracts (Fig. 5B) clearly show that the binding of SR proteins to an enhancer sequence can be displaced by addition of recombinant RSF1 in the extract. The amount of RSF1 required to achieve this displacement was comparable with the amount of SR proteins in the extract, implying that the binding of both factors to this enhancer sequence is of the same strength. The fact that RSF1 interacts directly, through its GRS domain, with the RS domain of SR proteins (29) makes it likely that protein–protein interactions are also involved in splicing repression. These protein–protein interactions could, indeed, interfere with those occurring between SR proteins and constitutive factors. At least in vitro, we have shown that RSF1 prevents cooperative interaction between SF2/ASF and U1 snRNP 70 kDa protein and thus prevents the stable binding of U1 snRNP at the 5′ splice site (29).

In vitro splicing studies revealed that complementation of HeLa nuclear extract with recombinant GST–RSF1 inhibits splicing of reporter pre-mRNAs, whether they contain the A-type sequences or not (29), indicating that RSF1 inhibition of splicing does not require high-affinity RSF1 binding sites. Therefore, it is possible that, in vitro, RSF1 might act either by binding RNA with low
affinity and/or by binding free SR proteins. In keeping with this possibility, recent studies employing in vitro selection for functional splicing enhancers that bind specific SR proteins (12,52) revealed that both SELEX-isolated high-affinity consensus binding sites (10,14,39) and more degenerate (and possibly lower-affinity) binding sites (12,52) can function as exonic splicing enhancers. Thus, some reporter pre-mRNAs, used for in vitro splicing studies, might contain potential targets for RSF1 which have escaped our notice, because the iterative protocol is designed to yield the highest affinity sequences for purified RSF1, but not target sequences involving interaction of RSF1 with other splicing components. These potential targets could also be enhancer sequences, since it has been demonstrated recently that the second exon of β-globin pre-mRNA, a constitutively spliced pre-mRNA, harbors multiple distinct splicing enhancers (22).

Numerous studies contributed to the view that SR proteins can modulate splice site selection in a concentration-dependent manner (9,10). One imaginable consequence of this is that a wide-variety of pre-mRNAs can be controlled by a small set of evolutionarily conserved antagonistic sequence-specific splicing factors. Consistent with this view, targeted expression in vitro splicing studies, might contain potential targets for RSF1 binding sites (10,14,39) and more degenerate (and possibly functional splicing enhancers that bind specific SR proteins.

ACKNOWLEDGEMENTS

We thank A. Krainer for providing pSP64HΔ6 construct, H. Land for pβ6 plasmid and P. Bouvet for the random oligonucleotide pool. H.-M.B. acknowledges F. Amalric and A. Vincent for their interest and logical support in the initial part of this work. We thank D. Cribbs for support, scientific interest and critical review of the manuscript. A special thanks to A.-M. Duprat for her constant support. We also thank Y. De Preval for synthesizing a number of oligonucleotides. This work was supported by an ATIP grant from the CNRS (to H.-M.B) and a grant from the CNRS and the ARC (to J.T.). E.L., E.A. and S.B. were supported by graduate fellowships from the Ministère de l’Enseignement Supérieur et de la Recherche (MERS); E.L. and S.B. also benefited from graduate training fellowships from the ARC.

REFERENCES