

## The *Drosophila* Suppressor of sable Protein Binds to RNA and Associates with a Subset of Polytene Chromosome Bands

MICHAEL V. MURRAY,<sup>1</sup>† MICHAEL A. TURNAGE,<sup>1</sup> KELLY J. WILLIAMSON,<sup>2</sup>  
WAYNE R. STEINHAUER,<sup>3</sup> AND LILLIE L. SEARLES<sup>1,2\*</sup>

*Curriculum in Genetics and Molecular Biology*<sup>1</sup> and *Department of Biology*,<sup>2</sup> *University of North Carolina, Chapel Hill, North Carolina 27599*, and *Laboratory of Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27632*<sup>3</sup>

Received 6 September 1996/Returned for modification 27 October 1996/Accepted 23 December 1996

**Mutations of the *Drosophila melanogaster* suppressor of sable [*su(s)*] gene, which encodes a 150-kDa nuclear protein [Su(s)], increase the accumulation of specific transcripts in a manner that is not well understood but that appears to involve pre-mRNA processing. Here, we report biochemical analysis of purified, recombinant Su(s) [rSu(s)] expressed in baculovirus and in *Escherichia coli* as maltose binding protein (MBP) fusions and immunocytochemical analysis of endogenous Su(s). This work has shown that purified, baculovirus-expressed rSu(s) binds to RNA in vitro with a high affinity and limited specificity. Systematic evolution of ligands by exponential enrichment was used to identify preferred RNA targets of rSu(s), and a large proportion of RNAs isolated contain a full or partial match to the consensus sequence UCAGUAGUCU, which was confirmed to be a high-affinity rSu(s) binding site. An MBP-Su(s) fusion protein containing the N-terminal third of Su(s) binds RNAs containing this sequence with a higher specificity than full-length, baculovirus-expressed rSu(s). The consensus sequence resembles both a cryptic 5' splice site and a sequence that is found near the 5' end of some *Drosophila* transcripts. Immunolocalization studies showed that endogenous Su(s) is distributed in a reticulated pattern in *Drosophila* embryo and salivary gland nuclei. In salivary gland cells, Su(s) is found both in the nucleoplasm and in association with a subset of polytene chromosome bands. Considering these and previous results, we propose two models to explain how *su(s)* mutations affect nuclear pre-mRNA processing.**

The *Drosophila melanogaster* suppressor of sable [*su(s)*] gene is one of a group of recessive suppressors that, when mutated, modify phenotypes associated with transposon insertion mutations at particular genes (reviewed in reference 38). Upon further examination at the molecular level, it has become apparent that these suppressor genes encode proteins that function more generally in regulating specific aspects of transcription and pre-mRNA processing. For example, *suppressor of white-apricot* [*su(w<sup>a</sup>)*] encodes an alternative splicing factor (47), and the protein encoded by *suppressor of forked* [*su(f)*] is homologous to a human polyadenylation factor subunit (40). The *su(s)* gene encodes a 150-kDa nuclear protein [Su(s)] with relatively little homology to other known proteins. Voelker et al. (43) identified a region within *su(s)* encoding 77 amino acids that is 28.6% identical to a region in the *Drosophila* U1 70K protein (23). This region of U1 70K is thought to be an Arg-Ser domain, which is found in numerous splicing factors (2, 11). Although the corresponding region of Su(s) contains two Arg-rich clusters, the characteristic Arg-Ser dipeptides are virtually absent from Su(s). Thus, the observed similarity between Su(s) and U1 70K may or may not be functionally significant. Voelker et al. (43) also identified a region of Su(s) with weak homology to an RNA recognition motif (RRM). More recently, this region of *su(s)* was reexamined, and using a refined definition of this motif, Birney et al. (2) concluded that it lacks sufficient homology to be classified as an RRM. Thus, the

primary sequence analysis of *su(s)* did not strongly indicate a biochemical function for the encoded protein.

Loss-of-function *su(s)* mutations suppress certain mutant alleles of *vermillion* (*v*) *sable*, *speck*, *purple* (*pr*), *singed*, and *yellow* (*y*) (20). Molecular analyses of suppressible *v*, *y*, and *pr* mutants indicate that *su(s)* mutations affect the processing of transcripts containing transposon 412 or P insertions (9, 10, 12, 19) within transcribed sequences near the 5' end of these genes. The picture that has emerged from our analysis of the interactions between *su(s)* and mutant derivatives of *v* is as follows. The suppressible *v* alleles contain identical insertions of the 7.5-kb retrotransposon 412 in the first *v* exon (36). The 412 sequences are incorporated into the mutant *v* pre-mRNA, and during splicing, cryptic 5' and 3' splice sites near the ends of 412 are used to splice all but a few bases of 412 sequences from the RNA, as though it were an intron (9). The accumulation of these mutant transcripts is exceedingly low in wild-type flies and is elevated by *su(s)* mutations (9, 30). The in vivo analysis of several derivatives of a suppressible *v* allele revealed that splicing of the 412 intron with the cryptic sites is somewhat inefficient, as indicated by the accumulation of RNAs retaining the 412 intron as well as fully spliced transcripts (10). Additionally, this work showed that *su(s)* mutations increase the stability of 412-containing transcripts without altering the efficiency of splicing the 412 intron. The finding that this same effect on transcript accumulation was produced in a *su(s)* wild-type background when one of the cryptic 5' splice sites was improved to a consensus site suggested that the higher stability of these transcripts could be the result of efficient splicing complex assembly at the 5' splice site of the 412 intron. However, from this analysis, it was not possible to establish whether *su(s)* mutations directly affect 5' splicing complex assembly or a different aspect of pre-mRNA metabolism.

In this report, we examine some of the biochemical proper-

\* Corresponding author. Mailing address: 507 Fordham Hall, Department of Biology, University of North Carolina, Chapel Hill, NC 27599. Phone: (919) 966-4989. Fax: (919) 962-1625. E-mail: lsearles@email.unc.edu.

† Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.

ties of purified recombinant Su(s) [rSu(s)], expressed in baculovirus and as *Escherichia coli* maltose binding protein (MBP) fusions, and the intracellular localization of endogenous Su(s). These experiments demonstrate that rSu(s) binds to RNA in vitro and preferentially binds a sequence that resembles both a cryptic 5' splice site and a sequence that is found near the 5' end of certain transcripts. Furthermore, Su(s) colocalizes with RNA processing components in salivary gland nuclei and appears to be associated at high levels with a relatively small number of transcripts. These results are consistent with possible roles of Su(s) in influencing complex assembly on cryptic 5' splice sites or nuclear RNA export.

## MATERIALS AND METHODS

**Cloning of *su(s)* into a baculovirus vector.** Prior to cloning of the *su(s)* cDNA into a baculovirus expression vector, an adapter encoding the amino acid sequence Tyr-His<sub>6</sub> was ligated to a 0.7-kb *SacI/DraI su(s)* genomic fragment (43) containing the 3' coding region of *su(s)*. This 3' *su(s)* fragment and a 3.2-kb *Clal/SacI* cDNA fragment containing the 5' coding sequence were ligated together into the baculovirus expression vector pVL1392 (Invitrogen). The resulting plasmid contains the entire coding sequence of *su(s)* with the His<sub>6</sub> tag, seven bases of 5' untranslated sequence from *su(s)*, and none of the 3' untranslated *su(s)* sequence. A transfection module (Invitrogen) was used to make recombinant baculovirus, and *Spodoptera frugiperda* Sf9 cells were infected with recombinant baculovirus as previously described (39).

**Purification of baculovirus-expressed rSu(s).** Sf9 insect cells were maintained in Grace's medium supplemented with 10% fetal calf serum, antibiotic-antimycotic (Gibco/BRL), and 2 mM L-glutamine at 26°C. To overexpress rSu(s), 250 ml of Sf9 cells ( $6 \times 10^5$  cells/ml) was infected with recombinant baculovirus in a spinner flask. After 3 days, cells were harvested by centrifugation at  $2,600 \times g_{av}$  for 10 min. The cell pellet was resuspended in 10 ml of phosphate-buffered saline (PBS) and recentrifuged. Cells were lysed by resuspending the pellet in 10 ml of buffer B1 (20 mM HEPES [pH 7.6], 2% Triton X-100, 1 mM  $\beta$ -mercaptoethanol, 0.2 mM phenylmethylsulfonyl fluoride, 0.1 mM leupeptin, 0.01 mM pepstatin, 5  $\mu$ g of aprotinin per ml). After being incubated on ice for 10 min with intermittent vortexing, nuclei were recovered by centrifugation at  $714 \times g_{av}$  for 10 min. The nuclear pellet was resuspended in 10 ml of buffer B2 (20 mM HEPES [pH 7.6], 1 M NaCl, 1 mM  $\beta$ -mercaptoethanol, 0.2 mM phenylmethylsulfonyl fluoride, 0.1 mM leupeptin, 0.01 mM pepstatin, 5  $\mu$ g of aprotinin per ml) and centrifuged at  $28,400 \times g_{av}$  for 15 min. The resulting supernatant, which contained rSu(s) in a relatively impure state, was discarded. The pellet was resuspended in buffer B2 and sonicated for seven 5-s pulses, each followed by a 1-min rest on ice, with a Branson Sonifier 250 (microtip at setting 2.5). This sonicate was centrifuged at  $28,400 \times g_{av}$  for 15 min, and the supernatant, containing the majority of rSu(s), was retained. To the nuclear sonicate were added CaCl<sub>2</sub> to 1 mM and micrococcal nuclease to 0.2 mg/ml, and the sample was incubated at 30°C for 20 min or until the majority of the nucleic acid was digested to less than 250 bp. After addition of EGTA to 5 mM, the sample was applied to a Sephacryl S-300 (Pharmacia) column (1.5 by 61.5 cm) equilibrated in 20 mM Tris-Cl (pH 7.6)–1 M NaCl–1 mM  $\beta$ -mercaptoethanol. The micrococcal nuclease and digested nucleotides eluted in the included volume of the column, whereas rSu(s) eluted in the void volume. Fractions containing rSu(s) protein were pooled and dialyzed against 20 mM Tris (pH 7.6)–1 M NaCl–1 mM  $\beta$ -mercaptoethanol–50% (vol/vol) glycerol. Aliquots of the dialysate were stored at –20°C. Protein concentration was determined by the Bradford assay (3).

**Generation and purification of MBP-Su(s) fusion proteins.** Clones of *su(s)* cDNA were cleaved with convenient restriction enzymes, and fragments encoding various portions of Su(s) were adapted with *EcoRI* linkers and cloned into the *EcoRI* site of pMal-c, an MBP expression vector (New England Biolabs). These clones produced fusion proteins containing Su(s) amino acids 1 to 360 [mSu(s)1-360], amino acids 1-434 [mSu(s)1-434], amino acids 361-700 [mSu(s)361-700], amino acids 701-1007 [mSu(s)701-1007], and amino acids 1008-1321 [mSu(s)1008-1321]. Cultures of *E. coli* transformant cells expressing the MBP-Su(s) fusions were grown and the fusion proteins were isolated as suggested by the manufacturer, with the following modification. The cleared cell lysates were adjusted to 1 mM CaCl<sub>2</sub> and 10 mM MgCl<sub>2</sub>; 40 U of protease-free RNase A (Sigma) was added, and the lysates were incubated on ice for 1 h. Proteins were purified by affinity chromatography on amylose-agarose columns as suggested by the manufacturer. Coomassie blue staining of sodium dodecyl sulfate (SDS)-polyacrylamide gels indicated that the fusion protein was the predominant protein in each of the preparations, with purity estimated to be approximately 30%. Protein concentrations were determined by Bradford assays.

**Nitrocellulose filter binding assays.** RNA binding activity was assayed as described previously (35), with modifications. The 100- $\mu$ l reaction mixtures contained 25 mM HEPES (pH 7.6), 100 mM NaCl, and approximately 6 to 9  $\mu$ M of RNA ( $9 \times 10^4$  cpm/pmol) and 0.05 to 50  $\mu$ g of rSu(s) or MBP-Su(s) fusion per ml, diluted immediately before use in 10 mM HEPES (pH 7.6)–500 mM NaCl–1 mM  $\beta$ -mercaptoethanol. Reaction mixtures were assembled and incubated on

ice for 1 to 60 min or at room temperature for 10 min. After addition of 900  $\mu$ l of ice-cold buffer NCB (10 mM Tris-Cl [pH 8], 60 mM KCl, 1 mM EDTA), each sample was filtered through a 24-mm-pore-size nitrocellulose filter (Millipore) presoaked in chilled buffer NCB. Samples were filtered at a rate of 5 ml/min and washed once with 5 ml of ice-cold NCB. Filters were then dried and counted by liquid scintillation counting. The background, i.e., counts per minute bound in the absence of protein, was typically less than 2% of the input. Under these experimental conditions, the dissociation constant ( $K_d$ ) is equal to the protein concentration that results in 50% of the maximal RNA binding. SigmaPlot (Jandel Scientific) was used to curve fit the data used for  $K_d$  determinations.

In the binding experiments with SELEX (systematic evolution of ligands by exponential enrichment) RNAs, templates for in vitro transcription were PCR amplified from plasmid clones in reactions primed with SLX1 and SLX2 (see below). After 10 amplification cycles, the resulting DNA was purified by gel electrophoresis. Synthetic RNAs were synthesized and subsequently purified by denaturing polyacrylamide gel electrophoresis (PAGE) as previously described (46). After being used in RNA binding experiments for  $K_d$  determination, the RNAs were reexamined by electrophoresis on 8 M urea–12% polyacrylamide gels to ensure that they remained intact. By densitometry and PhosphorImager analysis, the purity of these RNAs was typically better than 90%.

**Northwestern blots.** The procedure was done essentially as previously described (34), with the following modifications. The blotting buffer also contained 5  $\mu$ g of sonicated salmon sperm DNA per ml, 5  $\mu$ g of yeast tRNA per ml, and  $1.5 \times 10^6$  cpm *fushi tarazu* (*ftz*) RNA ( $9 \times 10^4$  cpm/pmol). The filter was stained to visualize proteins by incubation for 1 to 2 min in 50% methanol–0.1% Coomassie blue R-250 followed by destaining in 50% methanol–10% acetic acid.

**SELEX.** The selection and amplification of high-affinity RNAs were based on the method of Tuerk and Gold (42). The DNA oligonucleotides used were SLX1 (5'-TAATACGACTCACTATAGGAGCTGCAGACAACCAACC), SLX2 (5'-GTCTCGAGGTTTCTTAGGTTG), and SLX3 (5'-GTCTCGAGGTTTCTTAG GTGN<sub>20</sub>GGTTGGTTGTCTGCAGCTCTATAGTGAGTCGTATTA, where N<sub>20</sub> refers to the 20 randomized nucleotides). To synthesize the starting pool of RNA, 3  $\mu$ g of SLX1, which contains T7 RNA polymerase promoter sequences, and 6.4  $\mu$ g of SLX3, the template for synthesis of the starting RNA pool, were mixed, heated to 95°C, and annealed by slowly cooling to room temperature. RNA was synthesized as described by Milligan et al. (28) or with a Megashortscript kit (Ambion). As a tracer, 10 to 50  $\mu$ Ci of [<sup>32</sup>P]CTP (800 Ci/mmol) was added to these reactions. To select RNA-protein complexes, a 500- $\mu$ l sample containing 50  $\mu$ g (2.5 pmol) of RNA was added to 24  $\mu$ g (approximately 0.1 pmol) of rSu(s) in 10 mM HEPES (pH 7.6)–2 to 3 U of RNasin (Promega)–100 mM NaCl. After being diluted with 4.5 ml of ice-cold NCB, the sample was filtered through a nitrocellulose membrane as described above. The filter was washed four times with 5 ml of NCB and dried, and the bound RNA was quantitated by Cerenkov counting. RNA was eluted from the filter as previously described (41). To make cDNA, 200 ng of SLX2 was mixed with 400 ng of selected RNA in 20  $\mu$ l of 40 mM piperazine-*N,N'*-bis(2-ethanesulfonic acid) (PIPES; pH 6.5)–0.4 M NaCl–1 mM EDTA, incubated for 5 min at 85°C, and then incubated for 15 min at 42°C for annealing; subsequently, cDNA synthesis was performed as described previously (9). The cDNAs were amplified by adding 3 to 9  $\mu$ l of the cDNA synthesis reaction to a 300- $\mu$ l PCR mixture containing SLX1 and SLX2 as primers. After 10 cycles of amplification, the resulting DNA was purified on a 2% low-melting-point agarose gel. The products of the PCR were cloned by digestion of the DNA with *PstI* and *XhoI* and ligation into pBluescript KS(–). Sequencing was performed according to the Sequenase sequencing kit protocol (U.S. Biochemicals).

**Modification interference analysis.** Modification interference assays were performed as previously described (37), with the following modifications. Approximately 0.3 pmol of SELEX 8-5 RNA, 5' end labeled with <sup>32</sup>P, was modified with hydrazine or diethyl pyrocarbonate (DEPC) and added to a sixfold molar excess of rSu(s). The RNA binding reaction mixtures contained a 2,500-fold molar excess of tRNA relative to the labeled RNA to reduce nonspecific interactions of rSu(s) with SELEX 8-5. Bound RNAs were isolated on nitrocellulose filters and cleaved with aniline before being resolved on a 10% polyacrylamide–8 M urea gel. A single experiment contained four replicate samples of both free and the bound modified RNAs. Each band in the RNA ladders was quantitated with a PhosphorImager.

**Anti-Su(s) antibody production and purification.** Polyclonal antibodies were raised against a fusion protein that contained amino acids 648 to 808 of Su(s). To generate the appropriate gene fusion, a *PvuI/BamHI* restriction *su(s)* fragment including nucleotides (nt) 5832 to 6315 (43) was made blunt ended and ligated into *SmaI*-linearized pWR590 and pATH10 expression vectors (6, 13). The resulting *lacZ*- and *trpE-su(s)*648-808 fusion constructs were expressed in *E. coli* MV1184, and the fusion proteins were partially purified by insoluble aggregation (32, 45). The *LacZ*-Su(s) fusion was further purified by SDS-PAGE on a Bio-Rad Prep-Cell. Subsequently, a 100  $\mu$ g sample of purified *lacZ*-Su(s) fusion protein, emulsified in an equal volume of Freund's incomplete adjuvant, was injected subcutaneously into female New Zealand White rabbits. Rabbits were boosted at 1-month intervals with protein as described above. Harvested anti-serum was purified on a column containing the TrpE-Su(s) fusion coupled to Affi-Gel 10 (Bio-Rad). Loadings, washes, and elutions were done as previously described (33).

**Immunocytochemistry.** Embryos were collected and fixed in preparation for antibody staining as described by Ashburner (1). After rehydration, samples were blocked for several hours in blocking buffer (PBS, 2% bovine serum albumin [BSA], 10% normal goat serum [NGS], 0.1% Triton X-100) and incubated overnight with affinity-purified anti-Su(s) antibodies diluted 1:1,000 in blocking buffer. Embryos were washed at least six times for 30 min each with PBT (PBS, 0.2% BSA, 0.1% Triton X-100) prior to an overnight incubation with a 1:2,000 dilution of horseradish peroxidase-conjugated goat anti-rabbit antibodies (Vector), precleared by incubation with fixed and blocked embryos. Embryos were washed as described above prior to incubation with diaminobenzidine (0.5 mg/ml) and 0.003% hydrogen peroxide in PBT. After being washed twice in PBT, twice in PBS, once in ethanol, and finally in methanol, samples were mounted in methyl salicylate and photographed with Nomarski (differential interference contrast) or phase-contrast optics.

Salivary gland squashes were prepared and immunostained as described by Ashburner (1) except that slides were blocked with 5% NGS in PBS for 30 min prior to a 2-h incubation with primary antibody. A 1:500 dilution of affinity-purified, anti-Su(s) antibodies, precleared with fixed and blocked *su(s)*<sup>R39</sup> null mutant salivary glands, or a 1:100 dilution of a mouse anti-U1 70K monoclonal antibody (44) (provided by Arno Greenleaf) was used. Lissamine rhodamine sulfonyl chloride (LRSC)-conjugated goat anti-rabbit and dichlorotriazinyl amino fluorescein-conjugated goat anti-mouse antibodies (2 µg/ml; Jackson ImmunoResearch Laboratories, Inc.) were used as secondary antibodies for Su(s) and U1 70K antibodies, respectively. Salivary gland chromosomes were stained with 4',6-diamidino-2-phenylindole (DAPI) at 200 ng/ml in PBS for 5 min. Squashes were mounted with fluorescent mounting medium (90% glycerol in PBS containing 4% *n*-propyl gallate) and examined by epifluorescence or laser scanning confocal microscopy. DAPI-stained squashes were photographed with P800 color slide film. Individual channel confocal images were pseudo-colored and combined with Photoshop image processing software.

The salivary gland whole mounts were prepared by using the dissection and fixation procedure described for salivary gland squashes (1) except that the acetic acid treatment was omitted. Glands were washed three times with wash buffer (50 mM Tris [pH 7.5], 150 mM NaCl, 0.5% Nonidet P-40) and blocked for several hours in wash buffer containing 5% NGS and 1% BSA. Primary antibody incubations with cleared anti-Su(s) polyclonal antibodies (1:500) and either the anti-U1 70K (1:5) or antihistone (MAB1276; 1:30 Chemicon) monoclonal antibody were carried out overnight in the same buffer. After at least six washes for 30 min each, the glands were incubated overnight with secondary antibody. Glands were washed as described above with wash buffer, washed twice with PBS, suspended in fluorescent mounting medium for at least 1 h prior to being mounted, and examined by confocal microscopy.

## RESULTS

**Overexpression and purification of rSu(s) protein.** To obtain large quantities of purified rSu(s) for biochemical analysis, a *su(s)* cDNA was cloned into a baculovirus expression vector (see Materials and Methods). Whole-cell and fractionated extracts, prepared from baculovirus-infected cells, were examined by SDS-PAGE (Fig. 1). A prominent band, with an apparent molecular mass of 170 kDa, was observed in extracts prepared from cells containing the *su(s)* cDNA (Fig. 1, lane 1). Western blot analysis confirmed that this band, which was not observed in extracts prepared from mock-infected cells, was rSu(s) (data not shown). The vast majority of rSu(s) was associated with the nuclear fraction (Fig. 1, lane 3). Extraction of the nuclear pellet with 1 M NaCl solubilized most of the nuclear proteins (Fig. 1, lane 4), but a large proportion of rSu(s) remained associated with the nuclear pellet and was solubilized by sonication (Fig. 1, lane 5). This fraction was treated with micrococcal nuclease to remove contaminating nucleic acids and then passed through a Sephacryl S-300 column to separate the digested nucleic acids and micrococcal nuclease from rSu(s). The resulting preparation (Fig. 1, lane 7) was determined to have <5% nucleic acid contamination, based on the  $A_{260}/A_{280}$  ratio, and to be about 70% pure rSu(s), based on densitometric analysis of a Coomassie blue-stained gel. The major contaminant of this fraction was a 39-kDa polypeptide. Since this protein did not react with anti-Su(s) antibodies (data not shown), it is either an unrelated protein or a proteolytic fragment of rSu(s) that cannot be detected with existing antibodies.

In the process of subcloning the *su(s)* cDNA for baculovirus

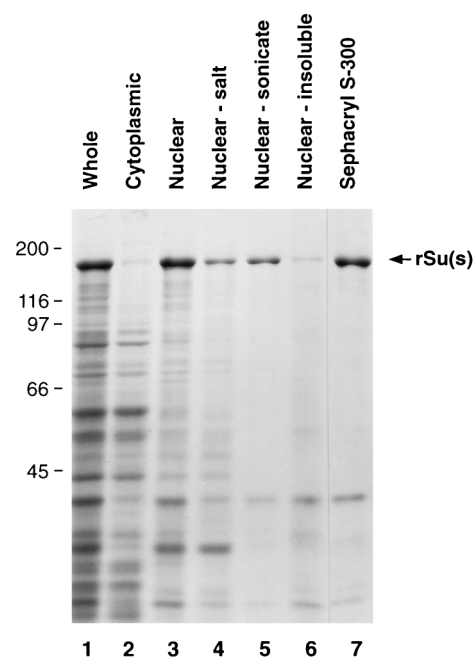


FIG. 1. Baculovirus expression and purification of recombinant Su(s). Protein extracts were prepared from baculovirus-infected Sf9 cells expressing a *su(s)* cDNA clone, and 0.2% of each sample was analyzed on an SDS-9% polyacrylamide gel that was stained with Coomassie blue. Whole-cell extract from infected cells (lane 1) was divided into cytoplasmic (lane 2) and nuclear (lane 3) fractions. Most nuclear proteins and some rSu(s) were solubilized by extraction of nuclei with buffer containing 1 M NaCl (lane 4). A large proportion of rSu(s) was associated with the pellet from the salt extraction and was solubilized by sonication (lane 5). The remainder of the proteins remained insoluble (lane 6). The nuclear sonicate was treated with micrococcal nuclease to digest nucleic acids and passed through a Sephacryl S-300 column. Recombinant Su(s) eluted in the void volume. A 7.5-µg sample of the final protein preparation is shown in lane 7. Sizes and positions of molecular mass standards are indicated in kilodaltons on the left.

expression, a His<sub>6</sub> tag was added to the carboxy terminus of the baculovirus-expressed rSu(s) (see Materials and Methods) so that the protein could be affinity purified on a nickel affinity column (Ni-nitrilotriacetic acid). However, repeated attempts to bind rSu(s) protein to this column under a variety of conditions were unsuccessful. Only under denaturing conditions did rSu(s) bind to the column and elute with imidazole, a specific competitor of the His<sub>6</sub> tag. Thus, the His<sub>6</sub> tag was present but inaccessible in the native protein. Since a good purification was achieved without denaturation, rSu(s) protein was purified as described above without the aid of Ni-nitrilotriacetic acid chromatography.

**Binding of rSu(s) to RNA.** As a first step in the biochemical analysis, we examined whether rSu(s) is capable of binding to RNA. One of the methods used to examine rSu(s) RNA binding activity was the RNA overlay-protein blot, or Northwestern blot, assay (34). In this experiment, polypeptides in the purified rSu(s) protein preparation were resolved by SDS-PAGE and transferred to a nitrocellulose filter. After the proteins were renatured, the blot was incubated with radioactively labeled, 473-nt *ftz* RNA, which we and others have used as a substrate for in vitro splicing reactions (see, for example, reference 31). The *ftz* RNA bound to a 170-kDa polypeptide migrating at the position of rSu(s) (Fig. 2A, lane 1). Since the *ftz* RNA binding activity in the purified rSu(s) preparation also immunoprecipitated with anti-Su(s) antibodies (data not shown), we concluded that the RNA binding activity observed in the North-

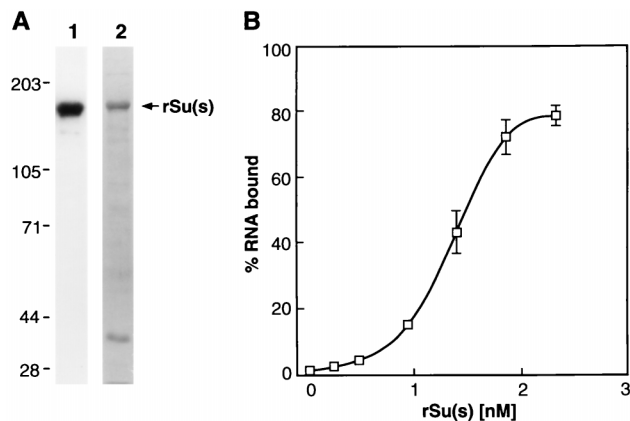


FIG. 2. RNA binding activity of rSu(s). (A) Northwestern blot of the partially purified rSu(s) fraction. A 5- $\mu$ g sample of the purified rSu(s) sample was fractionated an SDS-9% polyacrylamide gel and transferred to a nitrocellulose filter. The filter was probed with radiolabeled *ftz* RNA (see Materials and Methods) (lane 1) and stained with Coomassie blue (lane 2). The sizes and positions of molecular weight standards in kilodaltons are indicated on the left. (B) Nitrocellulose filter binding assay. The affinity of rSu(s) for a 473-nt *ftz* RNA was assayed by incubating various amounts of rSu(s) with a constant amount (6 pM) of  $^{32}$ P-labeled *ftz* RNA, and the bound RNA was recovered by filtration through a nitrocellulose membrane. Under these experimental conditions, the  $K_d$  equals the concentration of rSu(s) that produces half of maximal binding (approximately 80% of the input). Each point is corrected for background (typically less than 2% of the input) and is the average of three binding reactions.

western blot is due to rSu(s) rather than a comigrating contaminant. Other polypeptides in this fraction, including the major 39-kDa contaminant, have no apparent RNA binding activity.

A nitrocellulose filter binding assay (see Materials and Methods) was used to measure the affinity of rSu(s) for *ftz* RNA (Fig. 2B). In this experiment, a constant amount of radiolabeled *ftz* RNA was incubated with various amounts of rSu(s), and the RNA-protein complexes were isolated by filtration through nitrocellulose membranes. Purified rSu(s) bound to *ftz* RNA with a  $K_d$  of 1.4 nM, a value within the range observed for proteins with high affinities for RNA (4). Next we examined the ability of ribohomopolymers to compete with *ftz* RNA for binding to rSu(s) (Fig. 3A). If the observed rSu(s) binding activity were due to nonspecific, e.g., electrostatic, interactions, then the four ribohomopolymers would be expected to inhibit the binding of rSu(s) to *ftz* RNA with similar efficiencies. We quantitated the amount of *ftz* RNA bound in reaction mixtures containing fixed amounts of rSu(s) and labeled *ftz* RNA but various amounts of a given ribohomopolymer. This analysis showed that poly(U) and poly(G) were effective inhibitors of rSu(s) binding to *ftz* RNA when the concentration of ribohomopolymer was 100-fold greater than the concentration of *ftz* RNA. However, poly(C) and poly(A) did not inhibit *ftz* RNA binding even when present at a 1,000-fold excess. These results suggest that rSu(s) is capable of discriminating between different RNA substrates and also eliminate the possibility that the RNA was nonspecifically trapped on the filter.

The ability of rSu(s) to distinguish between RNA and DNA was tested by measuring the ability of salmon sperm DNA to compete for rSu(s) binding to *ftz* RNA (Fig. 3B). Both double-stranded and single-stranded DNA significantly inhibited rSu(s) binding to *ftz* RNA when the ratios of DNA to *ftz* RNA were 1,000 and 10,000, respectively. However, even at the highest DNA concentration tested, rSu(s) bound a substantial

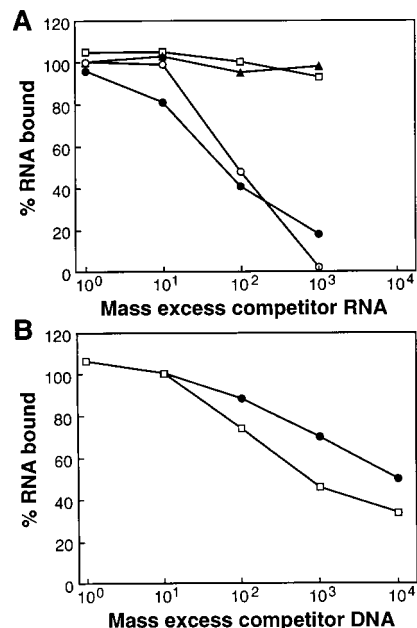


FIG. 3. Ribohomopolymer and DNA inhibition of rSu(s) RNA binding to *ftz* RNA. A constant amount of rSu(s) (1.3 nM) was added to a premixed sample of  $^{32}$ P-labeled *ftz* RNA (6 pM) and various amounts of unlabeled competitor, and the *ftz* RNA-rSu(s) complexes were isolated by filtration through nitrocellulose. The vertical axes of the graphs indicate the percentage of *ftz* RNA bound relative to a control binding reaction lacking competitor. (A) Competition by the following ribohomopolymers: poly(A) (□), poly(C) (▲), poly(G) (○), and poly(U) (●). (B) Competition by double-stranded DNA (□) and single-stranded DNA (●).

portion (30 to 50%) of the *ftz* RNA. Thus, rSu(s) appears to have a higher affinity for *ftz* RNA than DNA.

**Identification of high-affinity RNAs.** Because rSu(s) bound to a variety of different RNAs tested (unpublished observations and data presented below), the SELEX technique (42) was used to identify high-affinity RNA targets of rSu(s). This approach provides a means for the selective enrichment of high-affinity ligands in a randomized pool of RNAs by repeated rounds of isolating RNA-protein complexes and amplification of the bound RNAs. Synthetic oligonucleotides were used as the templates for in vitro synthesis of a pool of RNAs, 59 nt in length and containing random sequences at the central 20 positions (see Materials and Methods). A 25-fold molar excess of the randomized RNA pool was incubated with rSu(s), and bound RNAs were isolated on a nitrocellulose filter. The bound RNAs were eluted and amplified by using reverse transcription-PCR. After eight rounds of selection and amplification, the ability of rSu(s) to bind the selected RNA pool (pool 8) was compared to ability to bind to the starting RNA pool (pool 0) by the nitrocellulose filter binding assay (Fig. 4). The protein exhibited a significantly higher affinity for the RNAs in pool 8 ( $K_d$  of  $3.2 \pm 4.6$  nM) than the RNAs in pool 0 ( $K_d$  of  $24.8 \pm 3.3$  nM). Therefore, the selection-amplification procedure successfully enriched the RNA population with higher-affinity rSu(s) binding sites.

Approximately 20 individual round 8 RNAs were cloned and sequenced. Of these clones, five contain a perfect match to the consensus sequence UCAGUAGUCU, and five other clones contain at least a 6-of-10 match to this sequence (Fig. 5A).  $K_d$  values were determined for most of the individual round 8 RNAs (Fig. 5A), and a selected sample of the binding curves is shown in Fig. 5B. The RNAs with perfect or nearly perfect

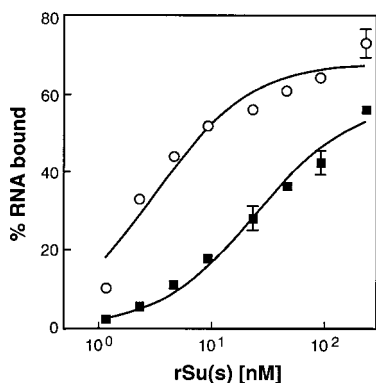


FIG. 4. Binding of rSu(s) to SELEX RNA pools. The randomized pool of RNA was subjected to eight rounds of SELEX with rSu(s). To verify that the selection procedure enriched for higher-affinity RNAs, the nitrocellulose filter binding assay was used to compare the affinity of rSu(s) for a radioactively labeled pool of the selected RNAs (pool 8 [○]) to its affinity for the starting RNA population (pool 0 [■]). The binding reaction mixtures contained the <sup>32</sup>P-labeled RNA pools at 9 pM and 50 μg/ml (approximately a 3 × 10<sup>5</sup> mass excess) of poly(C) to minimize nonspecific binding. Each point is the average of two binding reactions. The data were fitted to a curve by using a nonlinear regression program (see Materials and Methods).

matches to the consensus sequence have the lowest  $K_d$  values (2 to 7 nM). In general, RNAs that contain partial matches to the consensus also have low  $K_d$  values. However, two RNAs, 8-27 and 8-28, have low  $K_d$  values but do not contain the consensus sequence. Another feature apparent from inspection of the round 8 RNA clone sequences is the abundance of GU dinucleotides. The RNAs with  $K_d$  values of about 10 nM or less typically contain two to five GU dinucleotides in a 20-nt sequence. However, RNAs such as 8-21 contain several GU dinucleotides but bind relatively poorly to rSu(s). Thus, the SELEX consensus sequence or GU dinucleotides may contribute to the high affinity of rSu(s) for an RNA, but these features cannot be the only determinants of rSu(s) binding *in vitro*.

For comparison to the selected RNAs, 20 individual RNAs from the starting pool were also cloned and sequenced. Of the 20 round 0 clones sequenced, none had the sequence features, i.e., the SELEX consensus sequence and abundance of GU dinucleotides, found in the round 8 RNAs (data not shown). The round 8 clones, therefore, do not reflect a sequence bias in the starting pool.  $K_d$  determinations were made for five round 0 RNAs (Fig. 5A). Although several of these RNAs have  $K_d$  values that are in the same range as some of the round 8 RNAs, the  $K_d$  values for the round 0 RNAs are generally higher than the values obtained for the round 8 RNAs. The average of the round 0 RNA  $K_d$  measurements (25 nM) is very close to the  $K_d$  of 24.8 nM seen for the pool 0.

**Footprinting analysis of rSu(s) binding to a SELEX RNA.** To confirm that the SELEX consensus sequence represents an rSu(s) binding site, modification interference footprinting was used to map nucleotides on SELEX 8-5 RNA that interact with rSu(s). This RNA contains a 9-of-10 match to the SELEX consensus sequence and is bound by rSu(s) with a  $K_d$  of 2 nM (Fig. 5). SELEX 8-5 RNA, 5' end labeled with <sup>32</sup>P, was treated with DEPC or hydrazine under conditions where guanosine, adenosine, and uridine residues are modified. After incubation of modified 8-5 RNA with rSu(s), RNA-protein complexes were isolated on nitrocellulose filters. The bound RNA was eluted from the filters and treated with aniline, which cleaves RNA at the site of modification, and the products of the cleavage reaction were resolved on denaturing polyacrylamide

### A Round 8 RNAs

Consensus	UCAGUAGUCU	$K_d$
8-5	AGUCUCAGUAGUCAUUGGUG	2.1 ± 0.5 nM
8-40	GUUUCAGUAGUCUJUUGGUG	3.8 ± 1.3 nM
8-32	GUUUCAGUAGUCUJGGGUGUA	2.4 ± 0.8 nM
8-4	CAGUCUCAGUAGUCUJAGUGU	2.8 ± 0.5 nM
8-37	CAGUCUCAGUAGUCUJAGGCG	3.5 ± 0.7 nM
8-17	UUAUCAGUAGUCUJGGGUGUA	7.2 ± 1.8 nM
8-14	<u>UUCGUAGUGU</u> AAAGUAUCAA	10.0 ± 1.5 nM
8-13	GGCC <u>GUAGUG</u> GAGACAA	8.5 ± 0.6 nM
8-15	GGUJGGC <u>UAGUG</u> UCUUUGUA	6.2 ± 0.6 nM
8-21	AGUCUAAUAGGAGUCGUUAC	32.0 ± 4.1 nM
8-31	UGGGCCAGGUUCUAGUGA	14.0 ± 2.4 nM
8-38	AGCCUUUGCCCC <u>CAGU</u> UUUG	14.0 ± 1.0 nM
8-16	GGAAUUGGGAGC <u>CAGUCA</u> U	9.0 ± 0.8 nM
8-10	GGCGGAACCUUACU <u>ACGU</u> Aca	9.0 ± 1.1 nM
8-27	AUGUAGGGAAAUUUUAACG	4.0 ± 0.4 nM
8-28	CGGUCUCAUUUACU <u>UGUGUA</u> ca	2.1 ± 0.5 nM

### Round 0 RNAs

0-2	UCUGUACUCCAGCCUCCAG	64 ± 6.5 nM
0-3	CCUGAUAGGAUGUGUCUUUC	22 ± 2.0 nM
0-6	AUUUAUACGAUUUUCACUCG	16 ± 0.8 nM
0-8	CUCCCCUUUACUACUUAGUC	12 ± 1.3 nM
0-19	AGCAGUAAAAGUAAUCCGUC	13 ± 1.2 nM

### B

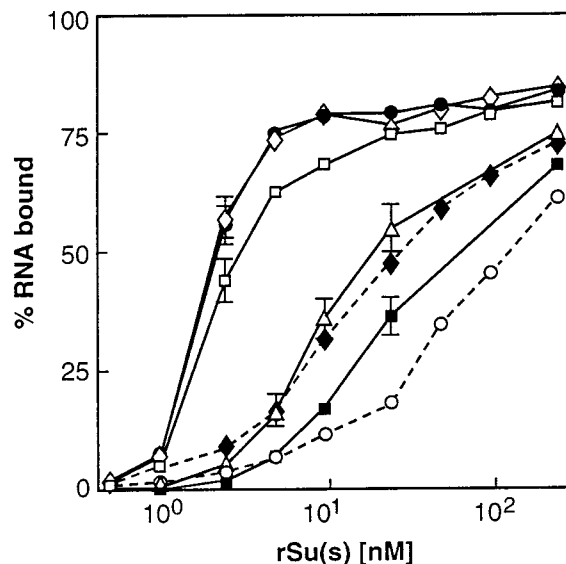


FIG. 5. Analysis of SELEX RNA cDNAs. (A) Sequences of individual RNAs cloned from the round 8 and round 0 pools. Nucleotides derived from the randomized portion of the RNA are shown in uppercase letters, and lowercase letters denote nucleotides from the nonrandomized portion of the RNA. A consensus sequence was derived from alignment of the round 8 clone sequences. Bases that conform to the consensus are underlined and boldfaced. The  $K_d$  values were determined for individual, gel-purified RNAs and with the same preparation of rSu(s), and the data were fitted to a curve by using a nonlinear regression program. For the 8-4 RNA, the  $K_d$  was determined in three separate experiments with different preparations of RNA, and no significant differences in  $K_d$  values were observed. (B) Binding of rSu(s) to several individual round 8 and round 0 RNAs. Nitrocellulose filter binding assays contained various amounts of rSu(s), <sup>32</sup>P-labeled SELEX RNA (9 pM), and 50 μg of poly(C) per ml to minimize nonspecific RNA binding. Each point is the average of two binding reactions. Symbols: □, 8-4 RNA; ◇, 8-5 RNA; ●, fitz RNA; △, 8-31 RNA; ■, 8-21 RNA; ◆, 0-6 RNA; ○, 0-2 RNA. The dotted lines indicate the binding curves for the two RNAs isolated from the starting pool.

gels. The banding patterns produced by bound RNA (Fig. 6, lanes 3 to 6) were compared to the patterns produced by modified RNA that had not been incubated with rSu(s) (Fig. 6, lanes 1, 2, 7, and 8). Several bands in the rSu(s)-bound RNA sample were significantly reduced in intensity but none were

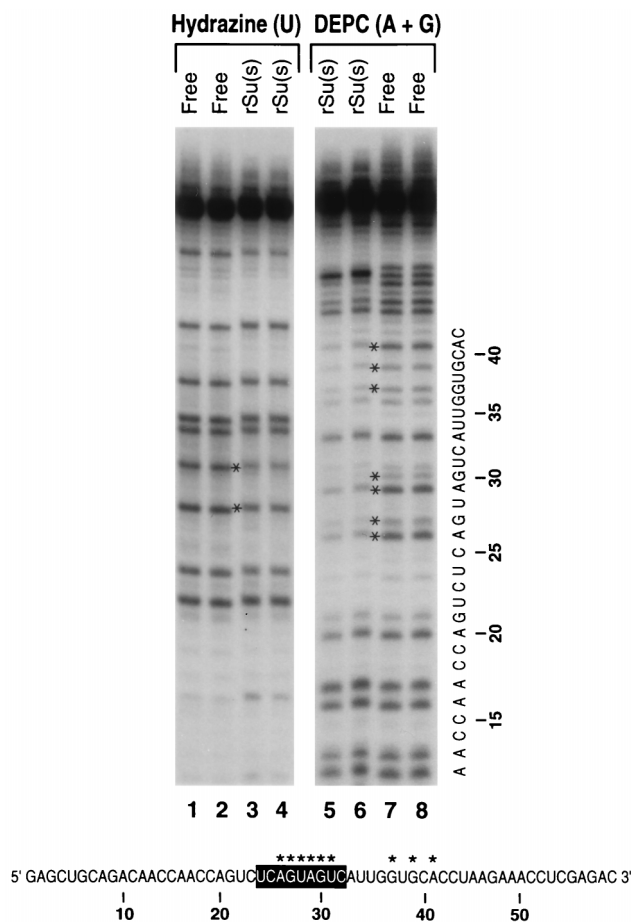


FIG. 6. Modification interference RNA footprinting of rSu(s) on SELEX 8-5 RNA. 5'-end-labeled SELEX 8-5 RNA was modified with hydrazine or DEPC and incubated in the presence or absence of a sixfold molar excess of rSu(s). Subsequently, bound RNA, isolated on nitrocellulose filters, and RNA that had not been incubated with protein were subjected to aniline cleavage followed by electrophoresis on a 10% polyacrylamide-8 M urea gel. The RNA ladders derived from hydrazine-modified RNA (lanes 3 and 4) and DEPC-modified RNA (lanes 5 and 6) that was bound by rSu(s) were compared to the aniline cleavage pattern produced from free hydrazine-modified RNA (lanes 1 and 2) and DEPC-modified RNA (lanes 7 and 8). The nucleotide sequence of SELEX 8-5 RNA is shown below the autoradiograph. The SELEX consensus sequence is in the black box, and the asterisks indicate the bases that were significantly reduced in intensity upon rSu(s) binding. The experiment shown contained four replicate samples of both free and the bound RNAs, and the intensity of each band in these samples was quantitated with a PhosphorImager. For the DEPC-modified RNAs, the bases marked with asterisks are ones whose modification reduced the intensity of the bound RNA band between two- to fourfold relative to the level of free 8-5 RNA. Upon hydrazine modification, the uridines marked with asterisks represent a 1.7- to 2.7-fold reduction in intensity of the bands of bound compared to free 8-5 RNA. Standard deviations indicate that these are significant differences.

completely absent; this result suggests that modification of a single nucleotide can reduce, but not eliminate, rSu(s) binding to SELEX 8-5 RNA. DEPC modification of nucleotides A26, G27, A29, and G30 led to significantly reduced binding by rSu(s) (Fig. 6; compare lanes 5 and 6 to lanes 7 and 8). Nucleotides U28 and U31, when modified with hydrazine, also interfered with rSu(s) binding (Fig. 6; compare lanes 1 and 2 to lanes 3 and 4). These nucleotides (A26 to U31) are located in the SELEX consensus sequence (U24 to C32 [Fig. 6]). The footprinting analysis also indicates that rSu(s) interacts with bases G37, G39, and A41, located outside of the consensus sequence within a G/U-rich region (Fig. 6, lanes 5 and 6).

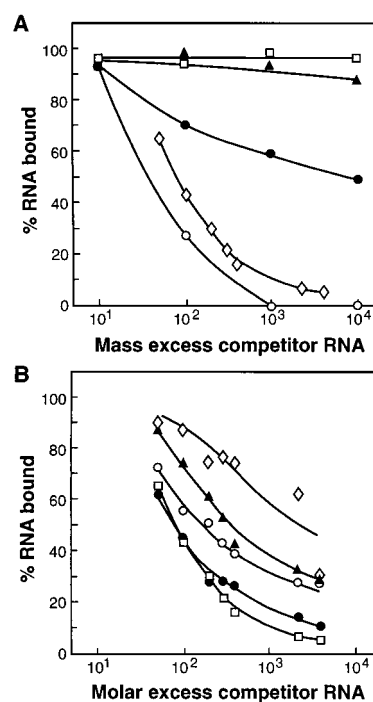


FIG. 7. Competitive nitrocellulose filter binding assay measuring the specific binding of rSu(s) to SELEX 8-5 RNA. Increasing concentrations of various cold competitor RNAs were combined with 17 pM of  $^{32}$ P-labeled SELEX 8-5 RNA. The RNA mixture was then added to binding reactions mixtures containing 0.4  $\mu$ g of rSu(s) per ml, a 160-fold molar excess of rSu(s) over labeled SELEX 8-5 RNA. Under these conditions, 50% of the  $^{32}$ P-labeled 8-5 RNA was bound by rSu(s) in the absence of competitor RNA. The data in both panels represent averages of four independent competition experiments. (A) Competition between ribohomopolymers [poly(A) ( $\square$ ), poly(C) ( $\blacktriangle$ ), poly(G) ( $\circ$ ), and poly(U) ( $\bullet$ ) and  $^{32}$ P-labeled 8-5 RNA.  $\diamond$ , cold 8-5 RNA. (B) Competition between heterogeneous RNA sequences (0-2 RNA [ $\diamond$ ], tRNA [ $\blacktriangle$ ], pool 0 RNA [ $\circ$ ], 8-28 RNA [ $\bullet$ ], and 8-5 RNA [ $\square$ ]) and  $^{32}$ P-labeled SELEX 8-5 RNA. Statistical analysis revealed that with the exception of SELEX 8-28, there is less than a 0.1% chance that the binding curves are identical to the curve produced by competition with SELEX 8-5 RNA. The concentration of cold competitor required to reduce binding to labeled SELEX 8-5 RNA by 50% was determined for each experiment, with standard deviations ranging between 6 and 52%.

Attempts to map the rSu(s) binding site on SELEX 8-21 and 8-28 RNAs, which have a partial match to the consensus or lack it entirely, were unsuccessful. Our results indicate that heating of these RNA samples during the modification procedure significantly reduced binding (data not shown). Thus, RNA structure could be an important component of rSu(s) binding to these particular RNAs.

**Specificity of rSu(s) binding to SELEX 8-5 RNA.** Competition experiments were performed to examine the binding specificity of rSu(s) for an RNA containing the SELEX consensus sequence versus various homopolymeric and heterogeneous RNA sequences (Fig. 7). The ability of a cold (unlabeled) competitor RNA to prevent rSu(s) from binding to the labeled SELEX 8-5 RNA reflects the ability of rSu(s) to discriminate between the 8-5 RNA binding site and sequences found in the competitor RNAs. The ribohomopolymers poly(A) and poly(C) did not compete with 8-5 RNA for rSu(s) binding, even when present at a 10,000-fold mass excess over labeled 8-5 RNA (Fig. 7A). At a 10,000-fold mass excess, poly(U) inhibits binding of SELEX 8-5 RNA by approximately 50%. On the other hand, poly(G) is a strong competitor, inhibiting rSu(s) binding as effectively as cold 8-5 RNA; i.e., binding is inhibited by 50% at about a 100-fold mass excess (Fig. 7A). These results are

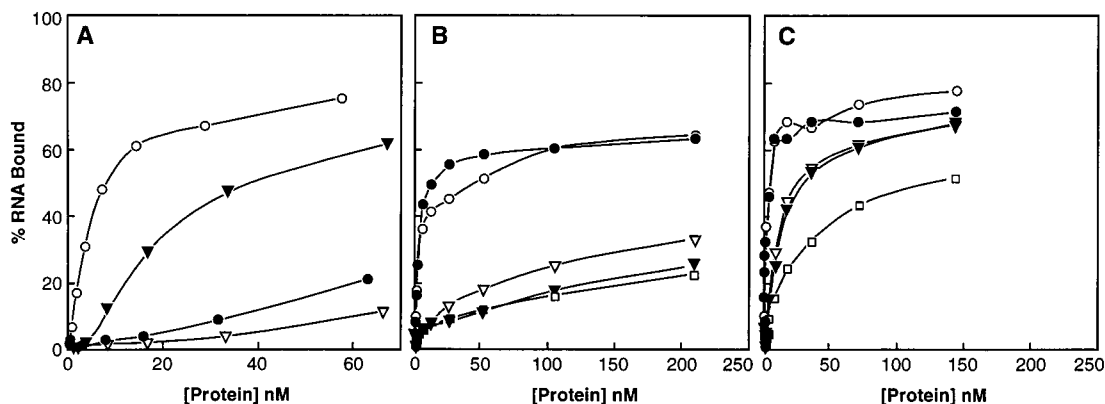


FIG. 8. RNA binding activities of MBP-Su(s) fusion proteins. The nitrocellulose filter binding assay was used to analyze affinity-purified preparations of MBP-Su(s) fusion proteins to determine the region of Su(s) required for RNA binding and for recognition of the SELEX consensus sequence. (A) Fusion proteins containing the amino acid regions of Su(s) indicated by the numbers (see Materials and Methods) were analyzed in RNA binding reactions performed with 6 pM of  $^{32}$ P-labeled *ftz* RNA. Symbols:  $\circ$ , mSu(s)1-360;  $\bullet$ , mSu(s)361-700;  $\triangle$ , mSu(s)701-1007;  $\blacktriangle$ , mSu(s)1008-1321. (B and C) Binding of fusion protein mSu(s)1-434 (B) or full-length rSu(s), purified from baculovirus (C), to several  $^{32}$ P-labeled SELEX RNAs (9 pM). Each curve represents the mean of three independent RNA binding experiments. Symbols:  $\circ$ , 8-5 RNA;  $\bullet$ , 8-32 RNA;  $\triangle$ , 8-28 RNA;  $\blacktriangle$ , 8-27 RNA;  $\square$ , pool 0 RNA.

qualitatively similar to those of the ribohomopolymer competition analysis of rSu(s) binding to *ftz* RNA (Fig. 3), although a larger amount of poly(U) is needed to inhibit binding to the SELEX 8-5 RNA.

Figure 7B illustrates the results of experiments that examined the ability of heterogeneous RNA sequences to compete for rSu(s) binding to SELEX 8-5 RNA. SELEX RNA 8-28, which lacks the consensus sequence but binds with a high affinity (Fig. 5), is as effective a competitor as cold 8-5 RNA. The randomized population of RNAs in the starting pool of the SELEX experiment (pool 0 RNA) and total yeast tRNA were also tested as competitors, and rSu(s) shows a small preference for 8-5 RNA over both of these RNA samples. The 0-2 RNA, isolated from pool 0, is bound by rSu(s) with about a 30-fold-lower affinity than 8-5 (Fig. 5). Consistent with this relatively high  $K_d$  value is the observation that 0-2 RNA is a less effective competitor than the other RNAs tested. Whereas about a 100-fold molar excess of cold 8-5 RNA reduced the binding of rSu(s) to labeled 8-5 RNA by 50%, a 2,400-fold molar excess of cold 0-2 RNA was required for a similar reduction in 8-5 RNA binding (Fig. 7B). Thus, the specificity of rSu(s) for heterogeneous RNA sequences is somewhat low *in vitro*, but the observed binding specificity is within the range observed for other RNA binding proteins (see Discussion). Furthermore, it appears from this analysis that  $K_d$  value comparisons are as useful as competition experiments in providing an indication of the rSu(s) binding specificity for heterogeneous RNA sequences.

**Delineation of the Su(s) RNA binding region.** To define the approximate region(s) of Su(s) responsible for RNA binding, portions of Su(s) were expressed as fusions with *E. coli* MBP, and after affinity purification, the fusion proteins were tested for RNA binding activity (Fig. 8A). Fusion protein mSu(s)1-360, containing the N-terminal 360 amino acids of Su(s), bound *ftz* RNA with a similar affinity as full-length rSu(s). A lower-affinity binding to *ftz* RNA was observed with a fusion protein containing C-terminal 313 amino acids [mSu(s)1008-1321], and no RNA binding was observed with two other fusion proteins containing Su(s) amino acids 361 to 700 or 701 to 1007. The binding specificity of a fusion protein containing the high-affinity, N-terminal RNA binding region was examined in assays using several SELEX RNAs (Fig. 8B). The fusion protein used in this analysis [mSu(s)1-434] extends 74 amino acids

further than the N-terminal fusion protein described in Fig. 8A and exhibits a similar affinity for *ftz* RNA (data not shown). The mSu(s)1-434 fusion protein bound two SELEX consensus-containing RNAs, 8-5 and 8-32, with a much higher affinity ( $K_d$  of 3 to 5 nM) than two SELEX nonconsensus RNAs, 8-27 and 8-28, and the pool 0 RNA. Baculovirus-expressed rSu(s) (Fig. 8C) showed less of a difference in its affinity for these same RNAs, as described earlier. Finding a high-affinity RNA binding activity in mSu(s) fusion proteins containing the N-terminal region confirms that binding RNA *in vitro* is a property of Su(s). Furthermore, the N-terminal mSu(s) fusion binds RNAs containing the SELEX consensus sequence with a substantially greater specificity than baculovirus-expressed rSu(s). The high level of nonspecific RNA binding activity observed with the partially purified, baculovirus-expressed protein could be a property of the full-length protein, possibly involving amino acids in the C-terminal region, or it could be due to a contaminating activity in the preparation.

**Immunocytochemical analysis of *Drosophila* Su(s) protein.** Affinity-purified anti-Su(s) polyclonal antibodies were used to examine the localization of endogenous Su(s) in whole mounts of embryos and larval salivary glands. Embryo nuclei stain with these antibodies (Fig. 9A), and rather than being uniformly distributed, Su(s) appears to be distributed in a reticulated network (Fig. 9C). The absence of staining in samples prepared from the *su(s)<sup>R39</sup>* null mutant (Fig. 9B) confirms that the observed pattern is specific for Su(s). Su(s) is also unevenly distributed in the nuclei of third-instar larval salivary glands (Fig. 9D and E). From the examination of a sample stained with both antihistone and anti-Su(s) antibodies, it appears that the staining of salivary gland nuclei is primarily in the nucleoplasm, and it is absent from the nucleolus and, largely, from chromosomes (Fig. 10A and B). A monoclonal antibody that recognizes the U1 70K protein (44) produced the same staining pattern in salivary gland nuclei as the anti-Su(s) antibodies (compare Fig. 10C and D). This same pattern was observed for U2 snRNP protein B and snRNAs (48).

Antibodies that recognize proteins involved in RNA processing typically stain polytene chromosomes due to the association of proteins with nascent transcripts. For example, anti-hnRNP antibodies stain a large number of polytene chromosome bands (26). With the anti-Su(s) antibodies, fewer than 20 sites on polytene chromosomes stained strongly, and a much larger

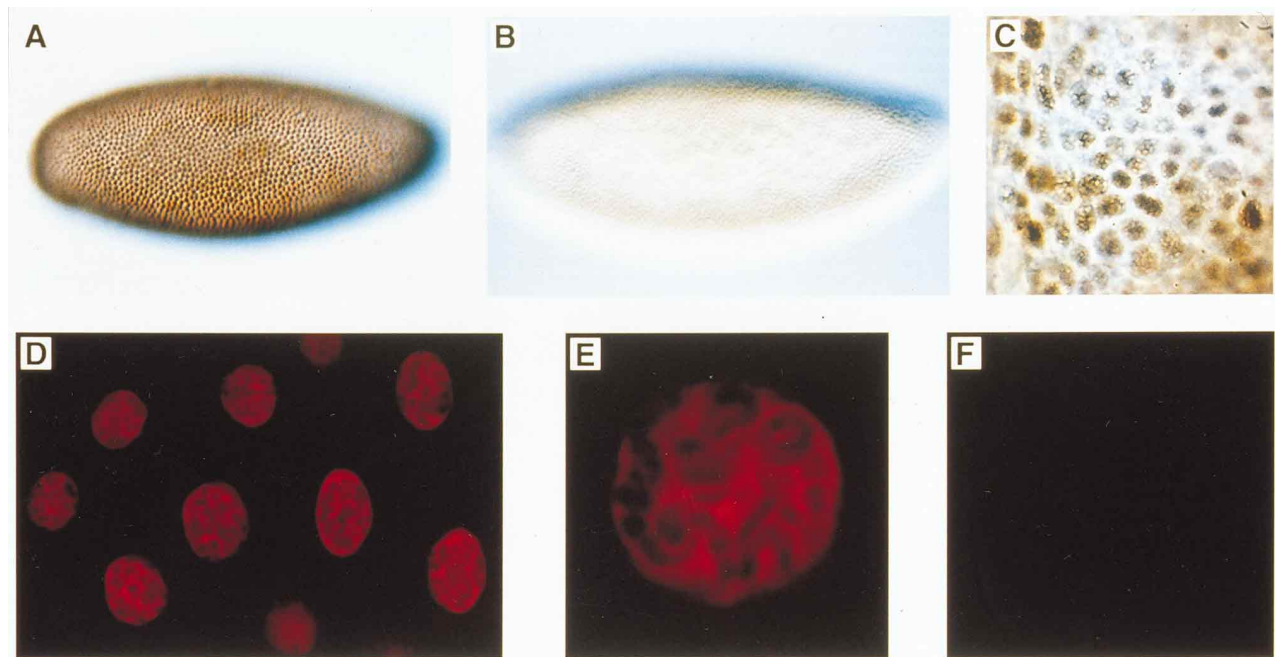


FIG. 9. Immunolocalization of Su(s) in embryo and salivary gland nuclei. (A and B) Differential interference contrast images of stage 4 Oregon R (A) and *su(s)*<sup>R39</sup> null mutant (B) embryos indirectly immunostained with anti-Su(s) polyclonal antibodies. Strong nuclear staining is evident in the wild-type but not the mutant embryo. (C) Higher-magnification phase-contrast image of a stage 8 Oregon R embryo shows the reticulated distribution of Su(s) in embryonic nuclei. (D and E) Confocal images of salivary gland whole mounts, stained by indirect immunofluorescence with the anti-Su(s) antibodies, reveal a similar distribution of Su(s) in salivary gland nuclei. (F) An individual salivary gland nucleus from a *su(s)* null mutant larva that was treated identically and imaged at the same brightness and contrast as in panel E.

number of bands stained weakly (Fig. 9E and F). Similar analysis of chromosome squashes prepared from the null mutant *su(s)*<sup>R39</sup> showed that both the strong and weak signals are due to Su(s) (data not shown). Most chromosomal puffs do not stain strongly; thus, the strength of the Su(s) signal is not correlated with the most actively transcribed genes. Analysis of chromosomal squashes doubly stained with anti-Su(s) and anti-U1 70K antibodies (Fig. 10G) revealed that many of the bands that stain with anti-Su(s) antibodies also stain with the anti-U1 70K antibody; however, some bands stain only with anti-U1 70K and others stain only with anti-Su(s). Therefore, Su(s) is localized, for the most part, at what appear to be sites containing pre-mRNAs that are being processed. Furthermore, finding a relatively high level of Su(s) associated with a limited number of sites indicates that Su(s) has preferred pre-mRNA targets in vivo.

## DISCUSSION

Our finding that rSu(s) binds RNA in vitro is consistent with previous results indicating that *su(s)* mutations affect pre-mRNA processing (9, 10, 12). The  $K_d$  values of baculovirus-expressed rSu(s) for various small RNAs ranged from 2 to 60 nM. Full-length rSu(s) generally exhibited a 5- to 10-fold preference for a SELEX consensus RNA versus other heterologous RNA sequences in vitro. This general range of specificity and affinity has been observed for other RNA binding proteins such as hnRNP proteins, SR proteins, and the *Drosophila* splicing regulators, Tra and Tra-2 (4, 21, 25). Recombinant Su(s) exhibited greater variation in its affinity for various small RNAs at lower protein concentrations than at higher protein concentrations. This observation would be consistent with the protein having a dual RNA binding function: a higher-affinity,

specific mode of binding and a somewhat lower-affinity, non-specific RNA-binding activity. Consistent with this idea are the results of our analysis of the MBP-Su(s) fusion proteins. These experiments showed that a domain capable of specifically recognizing RNAs containing the SELEX consensus sequence is located in the N-terminal region of Su(s), and a lower-affinity RNA binding domain is in the C-terminal region. The RRM-like motif of Su(s) (43) is contained within the C-terminal region and could be responsible for the lower-affinity RNA binding activity. We are performing experiments to define precisely the RNA binding domain(s) in the N-terminal region. The finding that a region of Su(s) specifically recognizes SELEX consensus RNAs increases the likelihood that the consensus sequence is relevant to Su(s) function.

Several conclusions can be drawn from the immunolocalization analysis of Su(s). First, the majority of Su(s) in salivary gland cells is found in the nucleoplasm. The experiments of Zachar et al. (48) indicate that inefficiently spliced pre-mRNAs and splicing factors accumulate in the nucleoplasm; however, we do not know whether this pool of Su(s) is bound to RNA or free. Second, it appears that Su(s) associates with some pre-mRNAs at an early stage, i.e., as they are being transcribed. Third, the finding that Su(s) associates with a relatively small number of polytene chromosome sites suggests that Su(s) is not involved in the metabolism of all pre-mRNAs in this cell type. While this observation suggests a higher level of specificity than would be predicted from our in vitro RNA binding experiments with baculovirus-expressed rSu(s), our analysis of MBP-Su(s) fusions is consistent with a more specific mode of RNA recognition. In addition, the binding specificity in vivo could be affected by parameters such as the intracellular concentration of Su(s) or interactions between Su(s) and other proteins. For example, cooperative binding was observed to



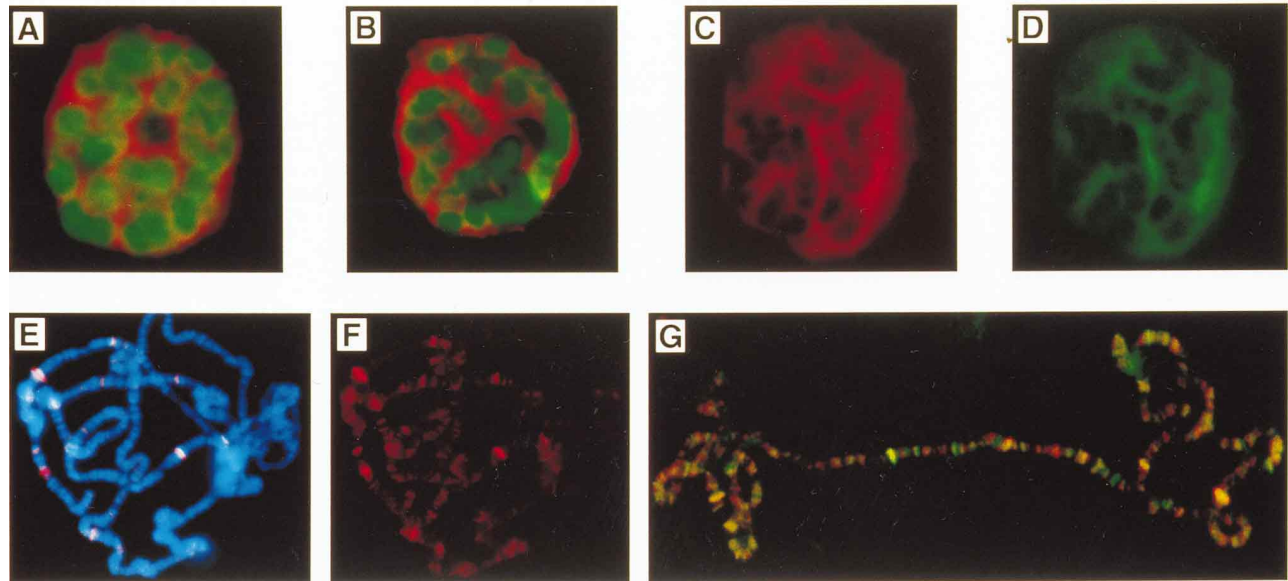


FIG. 10. Extrachromosomal and chromosomal localization of Su(s) in salivary gland nuclei. (A and B) Whole mounts of polytene nuclei indirectly immunostained for Su(s) (red) and histone H1 (green). The images are composites of pseudocolored channel images (rhodamine and fluorescein) from a confocal microscope. Su(s) is predominantly extrachromosomal and is excluded from the nucleolus. A single nucleus was indirectly immunostained with both the Su(s) antibodies (red) (C) and a U1 70K monoclonal antibody (green) (D). Su(s) and U1 70K proteins show identical localization patterns. (E) Double exposure of a salivary gland polytene chromosome spread that was indirectly immunostained for Su(s) (red) and counterstained with DAPI (light blue). The Su(s) antibodies stain a relatively small number of DNA bands. (F) The same chromosome spread in E showing only Su(s) immunolocalization illustrates that in addition to the few strongly staining bands, a much larger number of chromosomal bands appear to stain weakly for Su(s). (G) A polytene chromosome spread indirectly immunostained for Su(s) (red) and U1 70K (green) shows that the banding patterns of these two proteins overlap but are not identical. The image is a composite of pseudocolored channel images (rhodamine and fluorescein) from a confocal microscope.

increase the specific recognition of the *doublesex* splicing enhancer by Tra, Tra-2, and SC-35 (21).

The three classes of *su(s)*-suppressible alleles that have been examined at the molecular level thus far contain transposon insertions near the beginning of the transcript that are either recognized as the first intron or contained within it. We and others have obtained evidence that in *su(s)* mutants these transposon-containing pre-mRNAs are stabilized, and consequently higher levels of spliced transcripts are produced. Our data (10) indicate that the higher stability of the  $\nu$  mutant pre-mRNAs in *su(s)* mutants could result from an increased stability of splicing complexes on the cryptic, transposon 412-associated, 5' splice sites. According to this view, a function of Su(s) would be to prevent recognition of cryptic 5' splice sites. However, an alternative explanation could be that *su(s)* mutations result in these mutant RNAs being retained in the nucleus for a longer time, thereby allowing more time for the splicing out of transposon sequences prior to RNA export. If transported to the cytoplasm prior to the transposon sequences being removed, these RNAs, which are predicted to be poorly translated, would likely be unstable in the cytoplasm. Thus, an alternative function of Su(s) could be to promote export of certain RNAs from the nucleus.

The results of the SELEX experiment are consistent with both models of Su(s) function. Nucleotides 2 through 8 of the SELEX consensus sequence UCAGUAGUCU resemble the 5' splice site consensus CAG/GU(A/G)AGU, and this suggests the possibility that Su(s) binds directly to certain 5' splice sites and affects their recognition. The occurrence of a high proportion of GU dinucleotides in the selected RNAs is also consistent with Su(s) binding to sequences resembling 5' splice sites. The emerging view from recent studies is that multiple interactions are involved in 5' splice site selection (reviewed in

reference 14), and perhaps Su(s) is one of the proteins involved in this process. The SELEX sequence UCAGUAGUCU also resembles a sequence that is commonly found near the 5' ends of many *Drosophila* transcripts. Cherbas and Cherbas (5) found that in about 25% of arthropod transcripts, the sequence TCAGT, which matches the first five bases of the SELEX consensus, preferentially occurs within the first 10 nt of the RNA. The  $\nu$  and metallothionein/ $\nu^k$  long terminal repeat transcripts affected by *su(s)* mutations contain the sequence UCAGU within this region in a context that matches the SELEX consensus at 6 and 7, respectively, of 10 positions (24). A similar sequence is found near the 5' ends of *pr* and *y* transcripts, whose levels are elevated in *su(s)* mutants (19, 24). Thus, Su(s) may bind to a site near the 5' end of an RNA. The 5' cap site has been shown to be important for efficient splicing of the first intron (17, 29). If the effect of *su(s)* mutations is specific for the first intron as Kim et al. (19) suggested, this could involve Su(s) binding near the cap site and influencing splicing of the first intron of these transcripts. Su(s) might also interact with factors that bind to the 5' cap and promote RNA transport (15, 16, 18). For example, Su(s) might normally facilitate transport of these mutant RNAs out of the nucleus before splicing complexes assemble on the cryptic splice sites. This model is analogous to the function of human immunodeficiency virus Rev, which promotes the transport of a class of incompletely spliced human immunodeficiency virus RNA species (7, 8, 22, 27).

#### ACKNOWLEDGMENTS

We thank Bob Voelker for *su(s)* clones and mutants, and we greatly appreciate his overall contribution to the analysis of *su(s)*. We thank Arno Greenleaf for the U1 70K monoclonal antibody, Gustavo Maroni

and Mark Peifer for helpful comments on the manuscript, and Susan Whitfield for illustrations.

This work was supported by grants from the National Science Foundation (MCB-9205425 and MCB-9506301) to L. L. Searles.

#### REFERENCES

- Ashburner, M. 1989. *Drosophila*: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Birney, E., S. Kumar, and A. R. Krainer. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**:5803–5816.
- Bradford, M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248–254.
- Burd, C. G., and G. Dreyfuss. 1994. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**:615–621.
- Cherbas, L., and P. Cherbas. 1993. The arthropod initiator: the capsid protein plays an important role in transcription. *Insect Biochem. Mol. Biol.* **23**:81–93.
- Dieckmann, C. L., and A. Tzagoloff. 1985. Assembly of the mitochondrial membrane system. *J. Biol. Chem.* **260**:1513–1520.
- Fischer, U., J. Huber, W. C. Boelens, I. W. Mattaj, and R. Lührmann. 1995. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82**:475–483.
- Fischer, U., S. Meyer, M. Teufel, C. Heckel, R. Lührmann, and G. Rautmann. 1994. Evidence that HIV-1 Rev directly promotes the nuclear export of unspliced RNA. *EMBO J.* **13**:4105–4112.
- Fridell, R. A., A.-M. Pret, and L. L. Searles. 1990. A retrotransposon 412 insertion within an exon of the *Drosophila melanogaster vermilion* gene is spliced from the precursor RNA. *Genes Dev.* **4**:559–566.
- Fridell, R. A., and L. L. Searles. 1994. Evidence for a role of the *Drosophila melanogaster suppressor of sable* gene in the pre-mRNA splicing pathway. *Mol. Cell. Biol.* **14**:859–867.
- Fu, X. D. 1995. The superfamily of arginine/serine-rich splicing factors. *RNA* **1**:663–680.
- Geyer, P. K., A. J. Chien, V. G. Corces, and M. M. Green. 1991. Mutations in the *su(s)* gene affect RNA processing in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **88**:7116–7120.
- Guo, L., P. P. Stepien, J. Y. Tso, R. Brousseau, S. Narang, D. Y. Thomas, and R. Wu. 1984. Construction of expression vectors for fused proinsulin production in *Escherichia coli*. *Gene* **29**:251–254.
- Horowitz, D. S., and A. R. Krainer. 1994. Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. *Trends Genet.* **10**:100–106.
- Izaurrealde, E., and I. W. Mattaj. 1995. RNA export. *Cell* **81**:153–159.
- Izaurrealde, E., J. Lewis, C. Gamberi, A. Jarmolowski, C. McGuigan, and I. W. Mattaj. 1995. A cap-binding protein complex mediating U snRNA export. *Nature* **376**:709–712.
- Izaurrealde, E., J. Lewis, C. McGuigan, M. Jankowski, E. Darzynkiewicz, and I. W. Mattaj. 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**:657–668.
- Jarmolowski, A., W. C. Boelens, E. Izaurrealde, and I. W. Mattaj. 1994. Nuclear export of different classes of RNA is mediated by specific factors. *J. Cell Biol.* **124**:627–635.
- Kim, N., J. Kim, D. Park, C. Rosen, D. Dorsett, and J. Yim. 1996. Structure and expression of wild-type and suppressible alleles of the *Drosophila purple* gene. *Genetics* **142**:1157–1168.
- Lindsley, D. L., and G. G. Zimm. 1992. The genome of *Drosophila melanogaster*. Academic Press, Inc., San Diego, Calif.
- Lynch, K. W., and T. Maniatis. 1995. Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes Dev.* **9**:284–293.
- Malim, M. H., and B. R. Cullen. 1993. Rev and the fate of pre-mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes. *Mol. Cell. Biol.* **13**:6180–6189.
- Mancebo, R., P. Lo, and S. Mount. 1990. Structure and expression of the U1 small nuclear ribonucleoprotein particle 70K protein. *Mol. Cell. Biol.* **10**:2492–2502.
- Maroni, G. 1993. An atlas of *Drosophila* genes. Oxford University Press, New York, N.Y.
- Mattaj, I. W. 1993. RNA recognition: a family matter. *Cell* **73**:837–840.
- Matunis, M. J., E. L. Matunis, and G. Dreyfuss. 1992. Isolation of hnRNP complexes from *Drosophila melanogaster*. *J. Cell. Biol.* **116**:245–255.
- Meyer, B. E., and M. H. Malim. 1994. The HIV Rev trans-activator shuttles between the nucleus and the cytoplasm. *Genes Dev.* **8**:1538–1547.
- Milligan, J. F., D. R. Groebe, G. W. Witherell, and O. C. Uhlenbeck. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**:8783–8798.
- Ohno, M., H. Sakamoto, and Y. Shimura. 1987. Preferential excision of the 5' proximal intron from mRNA precursors with 2 introns as mediated by the cap structure. *Proc. Natl. Acad. Sci. USA* **84**:5187–5191.
- Pret, A.-M., and L. L. Searles. 1991. Splicing of retrotransposon insertions from transcripts of the *Drosophila melanogaster vermilion* gene in a revertant. *Genetics* **129**:1137–1145.
- Rio, D. C. 1988. Accurate and efficient pre-mRNA splicing in *Drosophila* cell-free extracts. *Proc. Natl. Acad. Sci. USA* **85**:2904–2908.
- Rio, D. C., F. A. Laski, and G. M. Rubin. 1986. Identification and immunocytochemical analysis of biologically active *Drosophila* P element transposase. *Cell* **44**:21–32.
- Robbins, A., W. S. Dynan, A. Greenleaf, and R. Tjian. 1984. Affinity-purified antibody as a probe of RNA polymerase II subunit structure. *J. Mol. Appl. Genet.* **2**:343–353.
- Robbins, S. G., M. F. Frana, J. J. McGowan, J. F. Boyle, and K. V. Holmes. 1986. RNA-binding proteins of coronavirus MHV: detection of monomeric and multimeric N protein with an RNA overlay-protein blot assay. *Virology* **150**:402–410.
- Sachs, A. B., and R. D. Kornberg. 1990. Purification and characterization of polyadenylate-binding protein. *Methods Enzymol.* **181**:332–352.
- Searles, L. L., R. S. Ruth, A.-M. Pret, R. A. Fridell, and A. J. Ali. 1990. Structure and transcription of the *Drosophila melanogaster vermilion* gene and several mutant alleles. *Mol. Cell. Biol.* **10**:1423–1431.
- Siebel, C. W., R. Kanaar, and D. C. Rio. 1994. Regulation of tissue specific P-element pre-mRNA splicing requires the RNA-binding protein PSI. *Genes Dev.* **8**:1713–1725.
- Smith, P. A., and V. G. Corces. 1991. *Drosophila* transposable elements: mechanisms of mutagenesis and interactions with the host genome. *Adv. Genet.* **29**:229–300.
- Summers, M. D., and G. E. Smith. 1988. A manual of methods for baculovirus vectors and insect cell culture procedures. *Tex. Agric. Exp. Stn. Bull.* **1555**:1–57.
- Takagaki, Y., and J. L. Manley. 1994. A polyadenylation factor subunit is the human homologue of the *Drosophila suppressor of forked* protein. *Nature* **372**:471–474.
- Tuerk, C., S. Eddy, D. Parma, and L. Gold. 1990. Autogenous translational operator recognized by bacteriophage T4 DNA polymerase. *J. Mol. Biol.* **213**:749–761.
- Tuerk, C., and L. Gold. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to T4 DNA polymerase. *Science* **249**:505–510.
- Voelker, R. A., W. Gibson, J. P. Graves, J. F. Sterling, and M. T. Eisenberg. 1991. The *Drosophila suppressor of sable* gene encodes a polypeptide with regions similar to those of RNA-binding proteins. *Mol. Cell. Biol.* **11**:894–905.
- Weeks, J. R., S. E. Hardin, J. Shen, J. M. Lee, and A. L. Greenleaf. 1993. Locus-specific variation in phosphorylation state of RNA polymerase II in vivo: correlations with gene activity and transcript processing. *Genes Dev.* **7**:2329–2344.
- Williams, D. C., R. M. Van Frank, W. L. Muth, and J. P. Burnett. 1982. Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. *Science* **215**:687–689.
- Witherell, G. W., H.-N. Wu, and O. C. Uhlenbeck. 1990. Cooperative binding of R17 coat protein to RNA. *Biochemistry* **29**:11051–11057.
- Zachar, Z., T.-B. Chou, and P. M. Bingham. 1987. Evidence that a regulatory gene autoregulates splicing of its transcript. *EMBO J.* **6**:4105–4111.
- Zachar, Z., J. Kramer, I. P. Mims, and P. M. Bingham. 1993. Evidence for channeled diffusion of pre-mRNAs during nuclear RNA transport in metazoans. *J. Cell Biol.* **121**:729–742.