J. Biochem. 121, 739-745 (1997)

Polycistronic Expression and RNA-Binding Specificity of the *C. elegans* Homologue of the Spliceosome-Associated Protein SAP49[1]

Yoshimi Tanaka, Atsunori Ohta, Kazuhiro Terashima, and Hiroshi Sakamoto[2]

Department of Biology, Faculty of Science, Kobe University, 1-1 Rokkodaicho, Nada-ku, Kobe 657

Received for publication, December 5, 1996

Splicing of mRNA precursors (pre-mRNAs) occurs in a multimolecular complex, termed spliceosome, which is comprised of pre-mRNA, small nuclear ribonucleoprotein particles (snRNPs), and other protein factors including spliceosome-associated proteins (SAPs). SAP49 is thought to be a subunit of the essential splicing factor SF3b and is involved in U2 snRNP function in mammalian cells. We have isolated a *Caenorhabditis elegans* cDNA encoding an RNA-binding protein with two RNA recognition motifs (RRMs) which shows extensive similarity to the human SAP49. The primary transcript for this *C. elegans* SAP49 homologue (cSAP49) seems to contain at least two additional cistrons and can be processed into three different mature mRNAs by *trans*-splicing. The cSAP49 mRNA, like other mRNAs in the same polycistronic unit, is expressed in most of the developmental stages, consistent with its putative essential function for mRNA splicing. By means of an *in vitro* RNA selection system, we demonstrate that cSAP49 possesses specific RNA-binding activity which resides in its second RRM.

Key words: *C. elegans*, *in vitro* selection, polycistronic expression, RNA-binding specificity, spliceosome-associated protein (SAP).

In eukaryotes, most gene expression undergoes modification by splicing, with precise removal of introns from pre-mRNAs, to give rise to mature mRNAs. Splicing occurs sequentially in multimolecular complexes which consist of pre-mRNA, small nuclear ribonucleoprotein particles (snRNPs), and many other factors (for review, see Refs. *1* and *2*). Assembly of these splicing complexes requires base-pairing interactions between RNA components (snRNAs) of snRNPs and several *cis*-acting elements on the pre-mRNA.

One of the essential interactions for the formation of the splicing complexes is a contact between U2 small nuclear RNA (snRNA) and the branch point sequence (BPS) just upstream of the 3′ splice site. This interaction specifies the branch point adenosine, leading to the formation of a lariat intermediate in the first step of the splicing reaction (*3*). In mammals, interaction of U2 snRNA with the BPS requires U2-specific spliceosome-associated proteins (SAPs) (*4*). The U2-specific SAPs 61, 62, and 114 together form the splicing factor SF3a, and SAPs 49, 130, 145, and 155 are subunits of splicing factor SF3b (*4-6*). Both SF3a and SF3b are required for the early splicing complex (A complex) assembly (*5, 6*). It has been shown by crosslinking experiments, that most of these SAPs are located near the BPS region in a specific order in the A complex, but their localization seems to be sequence-independent (*7-10*). In addition, it has not yet been determined whether individual SAPs have specific RNA-binding activity *per se*, although some of these contain known RNA-binding motifs (*9, 11, 12*).

Here we report the isolation and characterization of an unusual polycistronic *Caenorhabditis elegans* cDNA which possibly originates from incomplete *trans*-splicing of the primary transcript. Among the three open reading frames (ORFs) in the cDNA, the third was found to encode a 388 amino acid protein containing two RNA recognition motifs (RRMs) and a proline-rich domain, showing extensive similarity with the human SAP49. By an *in vitro* selection and UV-crosslinking analysis using bacterially-expressed fusion proteins, we show that this *C. elegans* SAP49 possesses specific RNA-binding activity, and that this activity is mediated by the second RRM.

MATERIALS AND METHODS

Oligonucleotides—The oligonucleotide primers used in this work are as follows.

YT1: 5′-CGATATCGTAAGCTTNA(C/T)(A/G)AAN(G/C)C(A/G)(A/T)ANCC-3′
YT2: 5′-AACAGCTATGACCATG-3′
PF1207: 5′-GGGAATTCAGATGTCAGCAGGGCCG-3′
PFrev: 5′-GTGGATCCGAATTCGGTACCGGGCCCCCC-CTCG-3′
PF1207-303R: 5′-CCGGAATTCCATATTCTTTTCATG A-3′
PF1207-337F: 5′-CGCGGATCCGTTGGTGCAAATATTT-3′
PGX2T-5′R: 5′-GAATTGGGGATCCACGCGGAACCAG-3′

[1] The cDNA sequence described in this paper has appeared in the Genbank database with the accession no. U24189. This work was supported in part by research grants from the Ministry of Education, Science, Sports and Culture of Japan, the Asahi Glass Foundation, the Inamori Foundation, and the Senri Life Science Foundation to H.S., and by the JSPS Research Fellowship for Young Scientists to Y.T.
[2] To whom correspondence should be addressed. Tel & Fax: +81-78-803-0554, E-mail: sakamoto@inherit.biosig.kobe-u.ac.jp

PGX2T-3′F: 5′-GATGGGAATTCATCGTGACTGACTG-3′

Screening and Sequencing—PCR was performed using the lysate from 10^6 pfu of λZAPII cDNA clones as template and the primers YT1 (for binding to RRM region) and YT2 (for binding to the invariable region of the λZAPII phage vector) by 30 cycles (94°C, 1 min; 50°C, 1 min; 72°C, 3 min). Amplified fragments were cloned into pUC118 vector and subjected to sequence analysis. A PCR fragment, isolated in this way, was used as a probe for hybridization to obtain the full-length cDNA encoding cSAP49 from the same cDNA library. Sequencing was performed using ABI autosequencer model 373A.

Northern Blot Analysis—Isolation of worms at specific developmental stages was performed according to the method by Sulston and Hodgkin (*13*). Total RNA from worms was isolated using QuickPrep total RNA extraction kit (Pharmacia). The RNAs (~8 μg) were electrophoresed on a 1% agarose gel containing 2% formaldehyde, and then subjected to Northern blot analysis (*14*). The labeled ORF1/ORF2-specific and ORF3(cSAP49)-specific probes were prepared from the YT1207 cDNA fragments spanning the nucleotide regions 1-1646 and 1843-3181, respectively.

Preparation of Fusion Proteins—To prepare GST-RRM1+2 fusion protein, an *Eco*RI sites was introduced just upstream of the cSAP49 initiation codon by PCR using primers PF1207 and PFrev, and the amplified fragment was blunt-ended by Klenow fragment after digestion with *Eco*RI and *Eco*RV, then cloned into the *Sma*I site of pGEX2T plasmid vector (Pharmacia). To construct GST-RRM1.5, the same PCR fragment was digested with *Eco*RI and *Hin*dIII, blunt-ended, and cloned into the *Sma*I site of pGEX2T. The plasmid for GST-RRM1+2, pGEX-RRM1+2 was subjected to PCR-based mutagenesis using two sets of primers, PF1207-303R & PGX2T-3′F and PF1207-337F & PGX2T-5′ to generate the constructs for the following fusions, GST-RRM1 and GST-RRM2, respectively. All GST fusion proteins were prepared as described previously (*15*).

In Vitro Selection of Ligand RNAs and Binding Analysis—*In vitro* selection using GST-RRM1+2 was performed as described previously (*15*). In brief, each RNA in the random pool was designed to have a 25 consecutive random sequence between the primer binding sites for amplification by RT-PCR. At selection, we used the affinity of the GST portion of the fusion proteins for glutathione-Sepharose to recover the RNAs that were bound by the proteins. The resultant RNAs were reverse transcribed and cloned into the *Bam*HI site of pUC118. To prepare the probe RNAs for *in vitro* binding analysis, *Bam*HI fragments from the plasmid clones were purified and used as templates for *in vitro* transcription with T7 RNA polymerase in the presence of [α-^{32}P]GTP. The labeled transcripts were purified by gel electrophoresis. The standard binding reaction mixture (10 μl) consisted of 10 kcpm probe RNA, 50 ng yeast RNA, and 100 ng of purified fusion protein in binding buffer (20 mM HEPES-NaOH pH 7.9, 200 mM KCl, 1 mM DTT, 0.1 mM PMSF, 0.1% Triton X-100, 5% glycerol). The reaction mixture was incubated at 20°C for 20 min followed by UV light irradiation and RNase digestion, and then analyzed by SDS-polyacrylamide gel electrophoresis (PAGE) as described previously (*16*). The efficiency of label transfer to the fusion proteins was analyzed

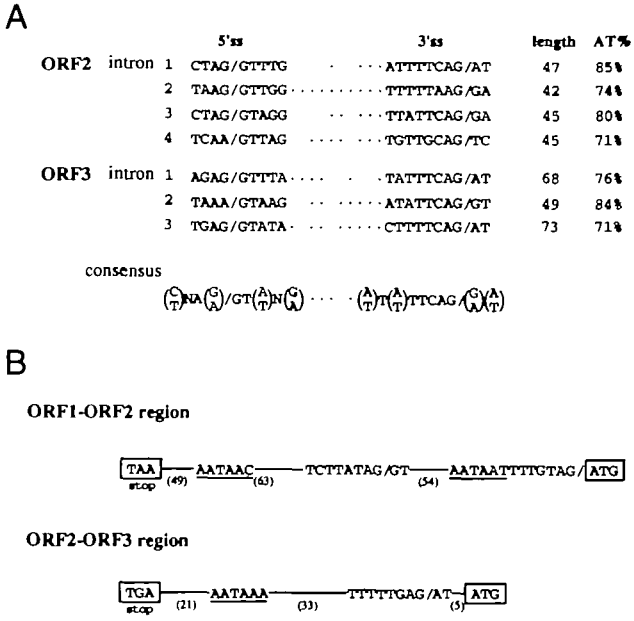

Fig. 2. **Sequences for *cis*-splice sites and putative *trans*-splice sites of the cSAP49 gene cluster.** (A) *Cis*-splice donor site (5′ss) and acceptor site (3′ss) sequences and their consensus. The numbers on the right show the length in nucleotide and AT content of each intron. (B) Putative polyadenylation signals (underlined) and *trans*-splice acceptor site (slashed) in the intergenic region. Stop and initiation codons are boxed.

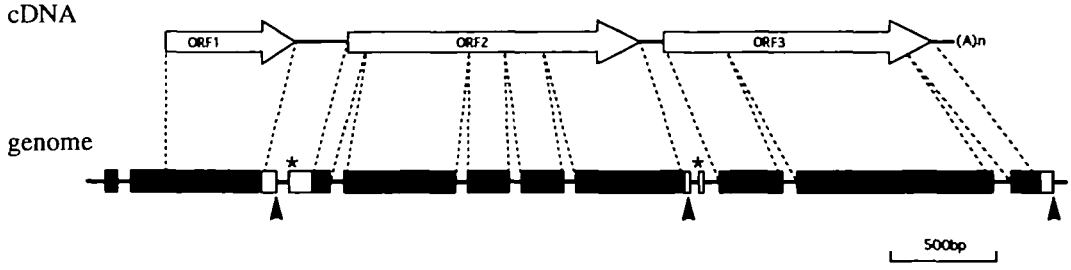

Fig. 1. **Schematic representation of the YT1207 cDNA and the corresponding genomic region.** Three ORFs within the cDNA are shown by open arrows. By comparing the sequence of YT1207 cDNA with that of the genomic cosmid clone CEC08B11 in the database, the protein-coding regions of the genome are shown by shaded boxes. Arrowheads and asterisks show the possible polyadenylation sites and *trans*-splice sites, respectively.

by densitometry using a Fuji BAS 2000 Image Analyzer. A negative control RNA used in the *in vitro* binding analysis contained the sequence 5′-UCAAUAGUGCAUUGAGAC-CCGGGUA-3′.

RESULTS

Nucleotide Sequence of the cDNA Encoding cSAP49— We attempted to identify cDNAs encoding novel RNA-binding proteins from a *C. elegans* mixed-stage cDNA library by degenerate PCR. A PCR fragment that seemed to encode a novel RRM, was then used as a probe to obtain a full-length cDNA from the same library. 4×10^5 pfu were screened, resulting in a positive clone (YT1207) which was isolated and subjected to sequence analysis. The 3207 nucleotide cDNA surprisingly contained three ORFs (*1-3*), the third of which corresponded to the probe fragment used in the screening (Fig. 1). The unusual structure of this cDNA seemed to result from incomplete RNA processing, since many mRNAs are believed to become mature by *trans*-splicing of polycistronic primary transcripts in *C. elegans* (*17, 18*). This side-by-side arrangement of potential genes represented by the three ORFs was confirmed by Southern blot analysis of the *C. elegans* genome using ORF-specific probes (data not shown). The result is also consistent with the sequence data of a *C. elegans* cosmid clone in the database (CEC08B11; accession no. Z46676). This appears to suggest that the genomic locus contains at least three cistrons in a single transcription unit.

Comparison of the cDNA sequence with the corresponding genomic sequence revealed that the genomic regions for ORF2 and ORF3 contain four and three introns, respectively (Fig. 1 and Fig. 2A). It should be noted that the second and fourth introns for ORF2 were newly identified in this study, although other introns were predicted previously by the genome project analysis. The putative cleavage/polyadenylation sites and the internal *trans*-splice sites, both of which would separate three ORFs to generate individual mRNAs, are located downstream of both ORF1 and ORF2 (Fig. 1 and Fig. 2B). Based on the genomic sequence information and the absence of a stop codon in the 5′ region of the cDNA, ORF1 likely extends further upstream. The corresponding gene contains one putative intron.

The protein deduced from ORF1 (107 amino acids) did not show any noticeable homology to known proteins and seemed to represent a partial protein sequence as mentioned above, and ORF2 encodes a 418 amino acid actin-related protein as described in the *C. elegans* database. The protein deduced from ORF3 contains two predicted RRM-type RNA-binding domains in the amino-terminal half and a proline-rich domain near the carboxy-terminus (Fig. 3). As judged by the position of a possible *trans*-splice site, translation initiates from the second methionine codon within ORF3, producing a 388 amino acid protein. Database

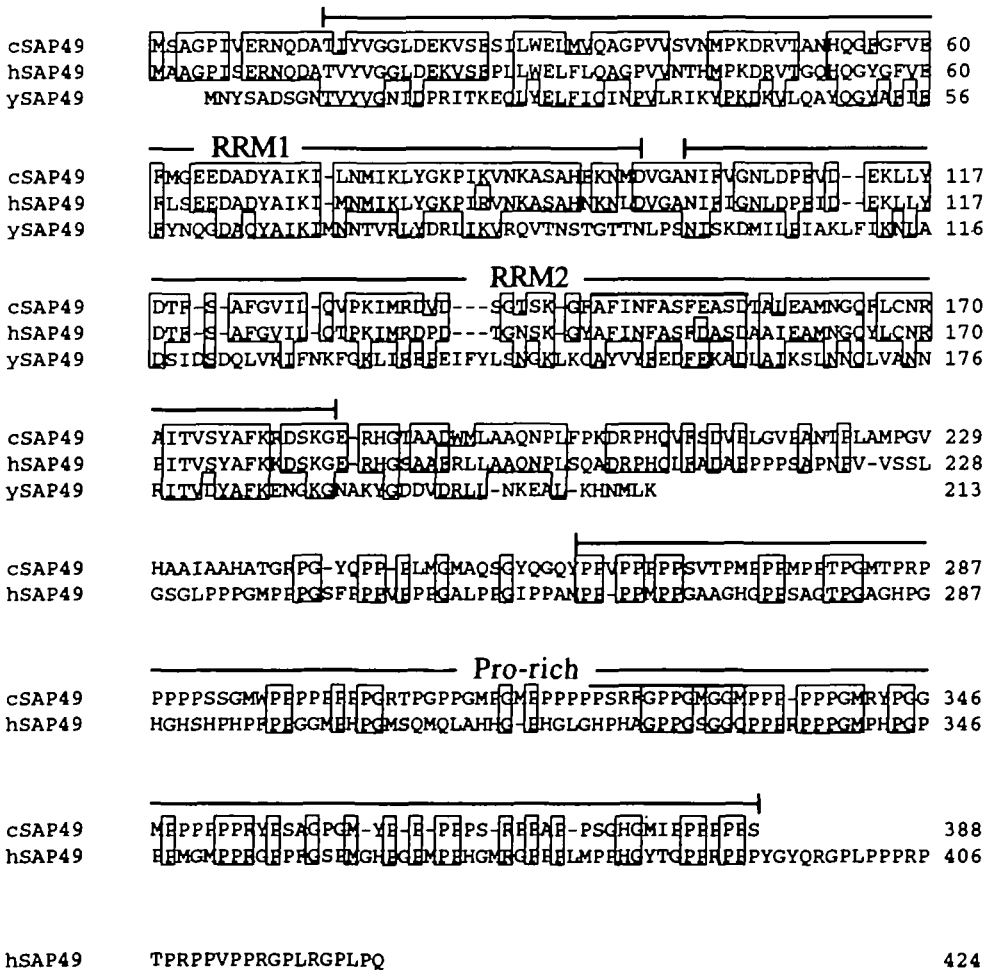

Fig. 3. **Comparison of the amino acid sequence of cSAP49 with those of the human (hSAP49) and yeast (ySAP49) counterparts.** Two putative RRM regions (RRM1 and RRM2) and proline-rich region (Pro-rich) are indicated. Identical amino acids are boxed. Amino acid residue numbers are shown on the right side.

searches have revealed that the protein shows extensive similarity (56% identity) to the human spliceosome-associated protein SAP49 (hSAP49) (*9*). In particular, the predicted RRM regions of the two proteins exhibit 82% identity. In addition, the proline-rich carboxy-terminal region is well conserved among these proteins. Thus, we conclude that the *C. elegans* protein encoded by ORF3 is a homologue of hSAP49 and was designated cSAP49. Interestingly, the putative yeast homologue (ySAP49) in the database shares moderate similarity with both hSAP49 and cSAP49 within the RRM regions, but lacks the carboxy-terminal proline-rich domain (*19*).

Developmental Expression Pattern of cSAP49 mRNA—If cSAP49 functions, like its human counterpart, as a subunit of splicing factor SF3b in *C. elegans*, it would be expected to be expressed constitutively in all cells at all developmental stages, since splicing is required for the production of almost all mRNAs. To examine the expression of cSAP49, total RNA was isolated from worms at various developmental stages and subjected to Northern blot analysis (Fig. 4). Using a cSAP49-specific probe spanning the ORF3 region, a 1.6 kb transcript slightly larger than the ORF3 was detected at all the developmental stages examined (from egg to young adult stages). Thus, the transcript seemed to be a mature polyadenylated mRNA for ORF3. This constitutive expression pattern of cSAP49 mRNA is consistent with its putative housekeeping role *in vivo*.

In addition, when a probe specific for both ORF1 and ORF2 was used, three distinct transcripts (1.0, 1.6, and 2.6 kb) were detected at all developmental stages. Of these bands, as judged by their sizes, the 1.0 and 1.6 kb transcripts appear to be the mature mRNAs for ORF1 and ORF2, respectively. The other very weak band of 2.6 kb seemed to correspond to an intermediate RNA containing ORF1 and/or ORF2 regions. Taken together, these results indicate that the primary transcript which gave rise to the YT1207 cDNA may be processed into at least three distinct mature mRNAs by *trans*-splicing.

In Vitro Selection of RNAs with Affinity for cSAP49—It was not known whether the human SAP49 possesses the specific RNA-binding ability by itself. To address if cSAP49 has this ability, we took advantage of the *in vitro* selection system as described previously (*15, 20*). The system requires the use of glutathione S-transferase (GST) fusion proteins for affinity selection, so we attempted to make a GST fusion with full-length cSAP49, but for unknown reasons, we were unable to obtain a successful clone from *Escherichia coli*. However, we were successful in producing a truncated GST fusion containing the two RRMs of cSAP49 (amino acids 1-256) which lacked most of the proline-rich region in the carboxy-terminus. This GST-RRM1+2 fusion was analyzed by the *in vitro* selection method.

After five rounds of *in vitro* selection, RNAs bound by GST-RRM1+2 were reverse-transcribed, and 34 clones were subjected to sequence analysis. The selection yielded four classes of sequences, as represented by clones #22, #61, #215, #216 in Fig. 5A. The outstanding feature of these RNA sequences is that all contained either the octamer

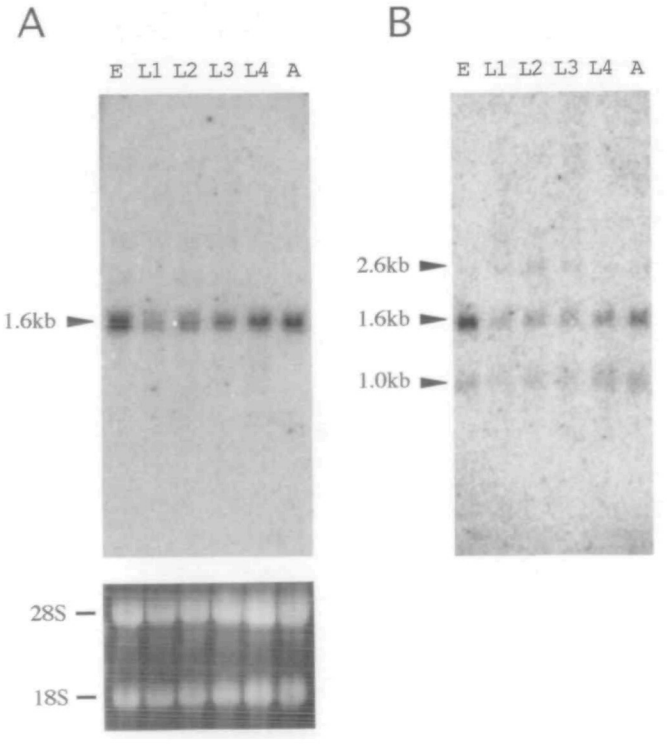

Fig. 4. **Developmental expression patterns of the mRNAs corresponding to the three ORFs of the YT1207 cDNA.** Total RNA at each developmental stage was isolated from synchronized cultures and subjected to Northern blot analysis. E, egg; L1, 1st larval stage; L2, 2nd larval stage; L3, 3rd larval stage; L4, 4th larval stage; A, young adult stage. (A) The 1.6 kb mRNA for ORF3 encoding the cSAP49. The ethidium bromide-stained small (18S) and large (28S) ribosomal RNAs for the same gel are shown below as loading controls. (B) Two distinct mRNAs for ORF1 (1.0 kb) and ORF2 (1.6 kb) were observed as well as a faint band of 2.6 kb RNA using an ORF1-ORF2 specific probe in the same blot of the panel A.

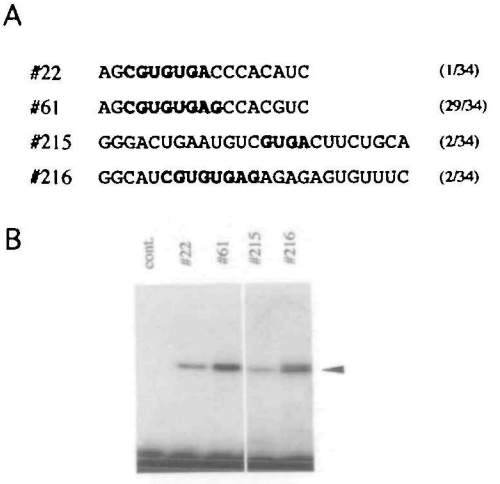

Fig. 5. ***In vitro* RNA selection and binding to the cSAP49.** (A) Sequences of the cSAP49-selected RNAs. The octamer motif and the similar sequences are boldfaced. The frequency of appearance of each RNA is shown on the right. (B) *In vitro* binding of cSAP49 to the selected RNAs. GST-RRM1+2 was incubated with the RNAs indicated above, UV-irradiated followed by RNase treatment, electrophoresed on a 12% SDS-polyacrylamide gel, and autoradiographed. Cont, a negative control RNA containing a 25-mer sequence unrelated to those of the selected RNAs.

motif, 5′-CGUGUGAG-3′, or a close derivative. The sequence of clone #61 was the most frequently selected (29/34). Clone #22 (1/34) is identical to clone #61 except for two base-substitutions (G10C and G15A) and clone #216 (2/34) contains the common octamer motif observed in the clones #61. Clone #215 (2/34) seems to be somehow different from others, but contains a part of the octamer motif (GUGA) observed in #61 and #216. The consistent appearance of such a common motif in most of the clones selected here strongly suggests that cSAP49 possesses a specific RNA-binding activity.

In Vitro Binding of cSAP49 to the Selected RNAs—To confirm the results of our *in vitro* selection analysis, we directly tested the ability of cSAP49 to bind to the four RNAs selected *in vitro* (clones #22, #61, #215, and #216) by means of UV-crosslinking assays (Fig. 5B). The RNAs from clones #61 and #216, which contain the octamer motif, were efficiently bound by GST-RRM1+2, whereas no binding to the negative control RNA containing an unrelated 25-mer sequence could be detected. Interestingly, the binding to clone #22 RNA, despite its close similarity to clone #61, was reduced by ~75% compared with that of #61, suggesting the importance of the last G residue in the octamer motif for efficient cSAP49-binding. As expected from the absence of half of the octamer motif, clone #215 RNA showed very weak affinity (~10%) as compared with that of #61. These results demonstrated that cSAP49 has a specific RNA-binding activity.

Determination of the Region Responsible for the Specific Binding—To define the region capable of mediating the specific RNA-binding ability, we made several deletion mutants of GST-RRM1+2 and analyzed their affinity for the #61 RNA (Fig. 6). GST-RRM1 (amino acids 1-97) and GST-RRM1.5 (amino acids 1-154), which contained the first RRM alone and the first RRM plus half of the second RRM, respectively, were not able to bind RNA. In contrast, GST-RRM2, which contains the second RRM alone (amino acids 98-256) bound just as efficiently as GST-RRM1+2. The binding profile of GST-RRM2 to the other RNAs (#22, #215, and #216) was found to be essentially identical to that of GST-RRM1+2 (data not shown). These results indicated that the second RRM is necessary and sufficient for cSAP49 RNA-binding ability to the RNAs selected in this study.

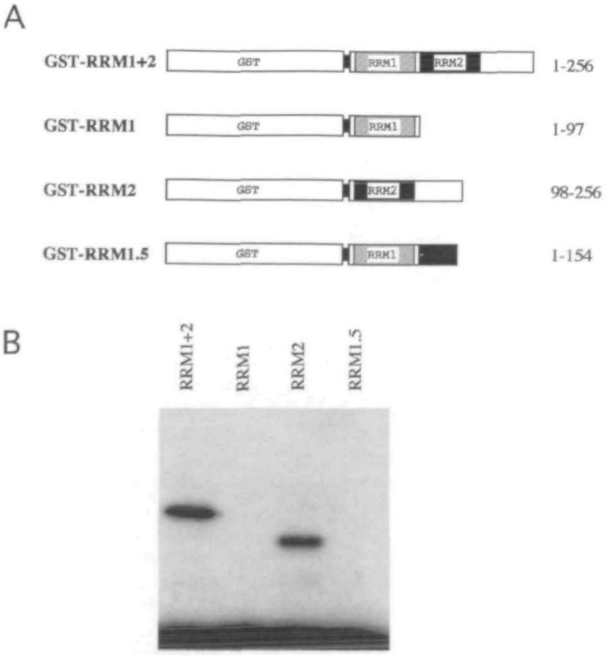

Fig. 6. **Determination of the functional binding domain of cSAP49.** (A) Schematic representation of the GST fusion proteins. The cSAP49 region of each GST fusion protein is shown as amino acid residue numbers on the right. (B) *In vitro* binding of each GST fusion indicated above to the #61 RNA. Experimental procedure is as described in Fig. 5.

DISCUSSION

Polycistronic Transcription and Trans-Splicing in C. elegans—Unlike in most other multicellular eukaryotes, a considerable number of mRNAs in *C. elegans* are produced through *trans*-splicing using the spliced leader (SL) RNAs, SL1 and SL2 (for review, see Refs. *18* and *21*). Recent studies on *C. elegans* mRNA expression indicate that some genes are transcribed polycistronically as in the case for bacterial operons (*17, 22-27*). According to an estimate, as many as one quarter of the genes in the genome form clusters and undergo polycistronic expression (*17*). Those primary transcripts are believed to be processed into monocistronic mature mRNAs by cleavage/polyadenylation and *trans*-splicing, as well as by *cis*-splicing, although direct evidence has been shown in only one case, due to the difficulty in detecting the short-lived polycistronic precursors in cells (*17, 18, 22*). Fortuitously, we obtained a polycistronic *C. elegans* cDNA containing three distinct ORFs, which appears to be a product of incomplete RNA processing. This finding provides direct evidence for the polycistronic organization of the *C. elegans* genome, and implies that *cis*-splicing precedes *trans*-splicing in the organism as all the *cis*-introns of each ORF in the cDNA have already been removed. As evidenced by the possible presence of a precursor which has yet to undergo *trans*-splicing between ORF1 and ORF2 in Northern blot analysis, it seems that the *trans*-splicing step separating each cistron in this gene cluster occurs relatively slowly as is the case with the gene cluster headed by *mai-1* (*22*).

The major difference between introns which are removed by *trans*-splicing and *cis*-splicing is thought to be simply that the former lacks a 5′ exon with a donor splice site, because *trans*- and *cis*-splice acceptor sites share the same consensus sequences and can be interchangeable (*28-30*). Our data on the splice sites (Fig. 2) supports the above hypothesis in that all 7 *cis*-introns identified are relatively short in length (42 to 73 nt), AT-rich (71 to 85%), and their donor and acceptor sites are well conserved. In the case of *trans*-splicing at the internal region between ORF2 and ORF3, similar AT-richness is observed within the 41-nucleotide region between the putative ORF2 polyadenylation signal (AATAAA) and the possible splice acceptor site (TTTGAG/AT) which lies 7 nucleotides upstream of the initiation codon of ORF3. This satisfies in part the requirements observed for *cis*-introns, although no evident donor splice site can be found in this region. It will be of importance to examine whether internal *trans*-splicing of the precursor RNA containing the cSAP49 coding region occurs in conjunction with SL2 RNA as suggested by Zorio *et al.* (*17*).

Functional Domains of SAP49 Implicated by Phylogenetic Conservation—Extensive similarity between cSAP49 and hSAP49 can be seen throughout the proteins. In particular, the first and second RRMs show 80 and 84% identity, respectively. The high degree of conservation in the RRM regions between humans and *C. elegans* and the developmentally constitutive expression of cSAP49 mRNA in *C. elegans* suggest an important role for SAP49 *in vivo*. The mouse SAP49 mRNA is also expressed in almost all tissues, albeit at varied levels among tissues (our unpublished observations). These results further support the idea that SAP49 is a basic component of an essential splicing factor, as has been suggested (*9, 10*).

Our *in vitro* selection was performed using a fusion protein consisting of the two highly conserved RRMs of cSAP49, however all of the selected RNAs were found to be targets for the second RRM, suggesting that the first RRM may have very weak or no RNA-binding activity. Use of GST-RRM1.5 (which contains only a partial second RRM) for *in vitro* selection, did not result in the detection of any common sequence motifs (data not shown). Taking both the phylogenetic conservation of the first RRM and previous observations of direct interaction of SAP49 with another SF3b subunit, SAP145 (*9*) into consideration, it is likely that the first RRM functions as a protein-protein interaction domain with SAP145 rather than as an RNA-binding domain. Interestingly, the carboxy-terminal proline-rich domain which is conserved between human and *C. elegans* SAP49, is not found in the putative yeast homologue (*19*). A similar proline-rich domain is also lacked in the yeast homologue of the human SAP62 (*11*). Such differences may imply a function for the proline-rich domain in mediating more complex interactions which occur in multicellular organisms.

What Is the Target RNA of SAP49 In Vivo?—Most of SF3a/SF3b subunits including SAP49 can be crosslinked in a specific order to sites upstream of the pre-mRNA BPS region in the splicing complex A (*9, 10*). As their localization seems to be sequence-independent, the possibility remained that some of the crosslinks in the splicing complex may be attributed to indirect interaction with the pre-mRNA. Our data provides the first evidence that one of the SF3b subunits does have intrinsic RNA-binding activity, suggesting that SAP49 may bind directly to the BPS region of pre-mRNA in the splicing complex. However, we could not find any significant homology between the cSAP49-selected sequences with the sequence of the BPS region. Under the conditions employed in our *in vitro* binding system, cSAP49 could not bind to the RNA sequences which have been reported for crosslinking of hSAP49 (data not shown). This does not rule out the possibility that SAP49 binding specificity may undergo modification *via* interaction with SAP145 and/or with other SF3b subunits, and thus become capable of binding to the BPS region.

Another clue to the function of cSAP49 is the sequence similarity of the *in vitro* selected octamer motif, CGUGUGAG, with splice site sequences. The octamer contains GU and AG, which are highly conserved in the 5′ and 3′ ends of introns, respectively. It has been well accepted that GU dinucleotides at the 5′ end of introns are recognized by U1 snRNA, but little is known about factors which recognize AG dinucleotides at the 3′ end, although the same U1 snRNP is involved in recognition of the AG dinucleotide in fission yeast (*31*). Thus, SAP49 might bind to either or both splice sites to tether the 5′- and 3′-ends of the intron in close proximity with U1 snRNP, and stabilize U2 snRNP near the 3′ splice site region in the splicing complex. Alternatively, the target of the cSAP49 second RRM might be snRNAs, which play essential roles in mRNA splicing, rather than pre-mRNA. It will be of great interest to examine whether cSAP49 interacts with SAP145 as is the case in humans and whether it directly binds to pre-mRNA and/or snRNAs in the presence or absence of other SF3b subunits.

We are grateful to Y. Kohara and Y. Andachi for providing the *C. elegans* cDNA library and experimental advice, U. Kikkawa and S. Kuroda for advice on DNA sequencing, and E. Sakashita for advice on *in vitro* RNA selection. We also thank R. Yu for critical reading of the manuscript and helpful comments.

REFERENCES

1. Moore, M.J., Query, C.C., and Sharp, P.A. (1993) Splicing of precursors to mRNA by the spliceosome in *The RNA World* (Gesteland, R.F. and Atkins, J.F., eds.) pp. 303-357, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
2. Baserga, S.J. and Steitz, J.A. (1993) The diverse world of small ribonucleoproteins in *The RNA World* (Gesteland, R.F. and Atkins, J.F., eds.) pp. 359-381, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
3. Query, C.C., Moore, M.J., and Sharp, P.A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* **8**, 587-597
4. Hodges, P.E. and Beggs, J.D. (1994) RNA splicing. U2 fulfils a commitment. *Curr. Biol.* **4**, 264-267
5. Brosi, R., Groning, K., Behrens, S.E., Luhrmann, R., and Kramer, A. (1993) Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science* **262**, 102-105
6. Brosi, R., Hauri, H.P., and Kramer, A. (1993) Separation of splicing factor SF3 into two components and purification of SF3a activity. *J. Biol. Chem.* **268**, 17640-17646
7. Gozani, O., Patton, J.G., and Reed, R. (1994) A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. *EMBO J.* **13**, 3356-3367
8. Staknis, D. and Reed, R. (1994) Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. *Mol. Cell. Biol.* **14**, 2994-3005
9. Champion-Arnaud, P. and Reed, R. (1994) The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes Dev.* **8**, 1974-1983
10. Gozani, O., Feld, R., and Reed, R. (1996) Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. *Genes Dev.* **10**, 233-243
11. Bennett, M. and Reed, R. (1993) Correspondence between a mammalian spliceosome component and an essential yeast splicing factor. *Science* **262**, 105-108
12. Chiara, M.D., Champion-Arnaud, P., Buvoli, M., Nadal-Ginard, B., and Reed, R. (1994) Specific protein-protein interactions between the essential mammalian spliceosome-associated proteins SAP 61 and SAP 114. *Proc. Natl. Acad. Sci. USA* **91**, 6403-6407
13. Sulston, J. and Hodgkin, J. (1988) Methods in *The Nematode Caenorhabditis elegans* (Wood, W.B., ed.) pp. 587-606, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
14. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
15. Sakashita, E. and Sakamoto, H. (1994) Characterization of RNA

binding specificity of the *Drosophila sex-lethal* protein by *in vitro* ligand selection. *Nucleic Acids Res.* **22**, 4082-4086
16. Sakamoto, H., Inoue, K., Higuchi, I., Ono, Y., and Shimura, Y. (1992) Control of *Drosophila Sex-lethal* pre-mRNA splicing by its own female-specific product. *Nucleic Acids Res.* **20**, 5533-5540
17. Zorio, D.A., Cheng, N.N., Blumenthal, T., and Spieth, J. (1994) Operons as a common form of chromosomal organization in *C. elegans*. *Nature* **372**, 270-272
18. Blumenthal, T. (1995) *Trans*-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends Genet.* **11**, 132-136
19. Wells, S.E., Neville, M., Haynes, M., Wang, J., Igel, H., and Ares, M., Jr. (1996) CUS1, a suppressor of cold-sensitive U2 snRNA mutations, is a novel yeast splicing factor homologous to human SAP 145. *Genes Dev.* **10**, 220-232
20. Abe, R., Yamamoto, K., and Sakamoto, H. (1996) Target specificity of neuronal RNA-binding protein, Mel-N1: direct binding to the 3′ untranslated region of its own mRNA. *Nucleic Acids Res.* **24**, 2011-2016
21. Nilsen, T.W. (1993) *Trans*-splicing of nematode premessenger RNA. *Annu. Rev. Microbiol.* **47**, 413-440
22. Spieth, J., Brooke, G., Kuersten, S., Lea, K., and Blumenthal, T. (1993) Operons in *C. elegans*: polycistronic mRNA precursors are processed by *trans*-splicing of SL2 to downstream coding regions. *Cell* **73**, 521-532
23. Morgan, W.R. and Greenwald, I. (1993) Two novel transmembrane protein tyrosine kinases expressed during *Caenorhabditis elegans* hypodermal development. *Mol. Cell. Biol.* **13**, 7133-7143
24. Hengartner, M.O. and Horvitz, H.R. (1994) *C. elegans* cell survival gene *ced-9* encodes a functional homolog of the mammalian proto-oncogene *bcl-2*. *Cell* **76**, 665-676
25. Clark, S.G., Lu, X., and Horvitz, H.R. (1994) The *Caenorhabditis elegans* locus *lin-15*, a negative regulator of a tyrosine kinase signaling pathway, encodes two different proteins. *Genetics* **137**, 987-997
26. Huang, L.S., Tzou, P., and Sternberg, P.W. (1994) The *lin-15* locus encodes two negative regulators of *Caenorhabditis elegans* vulval development. *Mol. Biol. Cell* **5**, 395-411
27. Land, M., Islas-Trejo, A., and Rubin, C.S. (1994) Origin, properties, and regulated expression of multiple mRNAs encoded by the protein kinase C1 gene of *Caenorhabditis elegans*. *J. Biol. Chem.* **269**, 14820-14827
28. Conrad, R., Thomas, J., Spieth, J., and Blumenthal, T. (1991) Insertion of part of an intron into the 5′ untranslated region of a *Caenorhabditis elegans* gene converts it into a *trans*-spliced gene. *Mol. Cell. Biol.* **11**, 1921-1926
29. Conrad, R., Liou, R.F., and Blumenthal, T. (1993) Conversion of a *trans*-spliced *C. elegans* gene into a conventional gene by introduction of a splice donor site. *EMBO J.* **12**, 1249-1255
30. Conrad, R., Lea, K., and Blumenthal, T. (1995) SL1 *trans*-splicing specified by AU-rich synthetic RNA inserted at the 5′ end of *Caenorhabditis elegans* pre-mRNA. *RNA* **1**, 164-170
31. Reich, C.I., VanHoy, R.W., Porter, G.L., and Wise, J.A. (1992) Mutations at the 3′ splice site can be suppressed by compensatory base changes in U1 snRNA in fission yeast. *Cell* **69**, 1159-1169