RNA (1996), 2:969–981. Cambridge University Press. Printed in the USA.
Copyright © 1996 RNA Society.

Structural and functional conservation of the *Drosophila doublesex* splicing enhancer repeat elements

KLEMENS J. HERTEL, KRISTEN W. LYNCH, EDWARD C. HSIAO, ERIC H.-T. LIU, and TOM MANIATIS
Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

ABSTRACT

We have compared the RNA sequences and secondary structures of the *Drosophila melanogaster* and *Drosophila virilis doublesex* (*dsx*) splicing enhancers. The sequences of the two splicing enhancers are highly divergent except for the presence of nearly identical 13-nt repeat elements (six in *D. melanogaster* and four in *D. virilis*) and a stretch of nucleotides at the 5′ and 3′ ends of the enhancers. In vitro RNA structure probing of the two enhancers revealed that the 13-nt repeats are predominantly single-stranded. Thus, both the primary sequences and single-stranded nature of the repeats are conserved between the two species. The significance of the primary sequence conservation was demonstrated by showing that the two enhancers are functionally interchangeable in Tra-/Tra2-dependent in vitro splicing. In addition, inhibition of splicing enhancer activity by antisense oligonucleotides complementary to the repeats demonstrated the importance of the conserved single-stranded structure of the repeats. In vitro binding studies revealed that Tra2 interacts with each of the *D. melanogaster* repeat elements, except for repeat 2, with affinities that are indistinguishable, whereas Tra binds nonspecifically to the enhancer. Taken together, these observations indicate that the organization of sequences within the *dsx* splicing enhancers of *D. melanogaster* and *D. virilis* results in a structure in which each of the repeat elements is single-stranded and therefore accessible for specific recognition by the RNA-binding domain of Tra2.

Keywords: phylogeny; regulated splicing; RNA/protein interactions; RNA structure

INTRODUCTION

Sex-specific alternative splicing of the *Drosophila melanogaster doublesex* (*dsx*) pre-mRNA requires the regulatory proteins Transformer (Tra) and Transformer 2 (Tra2), and a Tra- and Tra2-dependent splicing enhancer (the *doublesex* repeat element *dsx*RE) that is located 300-nt downstream of the female-specific 3′ splice site (for review see Baker, 1989; Maniatis, 1991). Tra is produced exclusively in females by the sex-specific splicing of Tra pre-mRNA, whereas Tra2 is expressed in both males and females (Boggs et al., 1987; Amrein et al., 1988). As shown in Figure 1, *dsx* pre-mRNA contains six exons: three common exons (exons 1–3), a female-specific exon (exon 4), and two male-specific exons (exons 5 and 6). In males, exon 3 is joined to exon 5 to produce an mRNA containing exons 1, 2, 3, 5, and 6. In females, exon 3 is joined to exon 4 to produce an mRNA containing exons 1, 2, 3, and 4 (Burtis & Baker, 1989). The female-specific 3′ splice site in intron 3 deviates significantly from the consensus 3′ splice site, so it is not recognized by the splicing machinery in the absence of Tra and Tra2, or in the absence of the *dsx*RE (Burtis & Baker, 1989; Hedley & Maniatis, 1991; Hoshijima et al., 1991; Ryner & Baker, 1991; Tian & Maniatis, 1992, 1993; Zuo & Maniatis, 1996).

The *dsx*RE is a 270-nt regulatory element that contains six 13-nt repeat sequences, and a purine-rich element (PRE) located between repeats 5 and 6 (Burtis & Baker, 1989; Lynch & Maniatis, 1995). The presence of both types of elements is required for efficient Tra/Tra2-dependent use of the female-specific 3′ splice site. Tra and Tra2 bind to the *dsx*RE and facilitate the recruitment of splicing factors to the weak female-specific 3′ splice site (Hedley & Maniatis, 1991; Inoue et al., 1992; Tian & Maniatis, 1992, 1993; Zuo & Maniatis, 1996). Thus, in females, the splicing machinery is preferen-

Reprint requests to: Tom Maniatis, Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

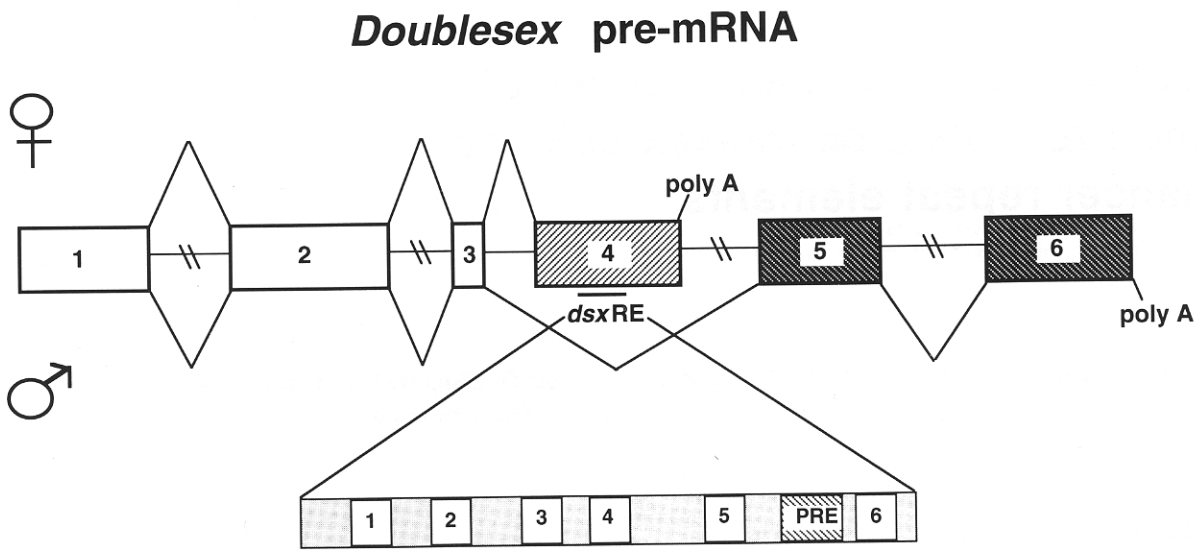

FIGURE 1. Sex-specific alternative splicing pattern of *dsx* pre-mRNA and the *dsx* splicing enhancer (*dsx*RE). Top: Open boxes represent the common exons 1, 2, and 3, the light hatched box represents the female-specific exon 4, and the dark hatched boxes represent the male-specific exons 5 and 6. The sex-specific splicing pattern is illustrated by the lines above (female) and below (male) the pre-mRNA. Sites of cleavage and polyadenylation are labeled poly A. Bottom: Enlargement shows the organization of the *D. melanogaster dsx*RE comprised of six 13-nt repeat elements (boxes 1–6) and a PRE located between repeats 5 and 6.

tially directed to an intrinsically weak splice site recognition signal.

In vitro splicing experiments have shown that, in addition to Tra and Tra2, one or more members of the SR (serine/arginine) family of general splicing factors are required for *dsx*RE-dependent recognition of the female-specific 3′ splice site (Tian & Maniatis, 1993). UV-crosslinking experiments and the characterization of affinity-purified enhancer complexes have shown that Tra and Tra2 recruit SR proteins to the *dsx*RE to form a multicomponent splicing enhancer complex (Tian & Maniatis, 1992, 1993). Both Tra and Tra2 contain SR domains, and are therefore considered members of the superfamily of SR-containing splicing factors (for review see Fu, 1995). Although Tra2 and SR proteins contain an RNA recognition motif (RRM) (Bandziulis et al., 1989), Tra is lacking a known RNA-binding domain. Binding studies in HeLa cell nuclear extracts or with purified recombinant Tra, Tra2, and SR proteins revealed that Tra2 binds with significant specificity to the *dsx*RE, whereas Tra exhibits low or no specificity (Hedley & Maniatis, 1991; Inoue et al., 1992; Tian & Maniatis, 1992; Lynch & Maniatis, 1995, 1996). In addition, these studies showed that Tra, Tra2, and SR proteins bind cooperatively to the *dsx*RE. Recent UV-crosslinking experiments have shown that Tra and Tra2 recruit a specific SR protein to the *dsx* repeats (Lynch & Maniatis, 1996). This SR protein binds to the 5′ ends of the repeats, whereas Tra2 binds to the middle and 3′ regions. Tra is an essential component of this heterotrimeric complex, but does not appear to directly contact RNA.

These observations, in conjunction with studies showing that these proteins interact with each other through their SR domains (Wu & Maniatis, 1993; Amrein et al., 1994), suggest that a complex involving multiple specific protein–protein and protein–RNA interactions is assembled on the *dsx*RE.

Tra2 and SR proteins contact RNA through their RRMs, but relatively little is known about the structure of the binding sites in the *dsx*RE. Recently, the three-dimensional structure of a complex between the RNA-binding protein U1A and its binding site in U1 snRNA was determined (Oubridge et al., 1994). U1A, which contains an RRM similar to that present in SR proteins and Tra2, specifically interacts with the single-stranded region of a hairpin structure formed by U1 snRNA. The single-stranded region provides a surface for extensive interactions between the protein and exposed nucleotides. These results and other studies suggest that interactions with single-stranded RNA might be a general mode of recognition for the RRM domain/RNA complex (Nagai et al., 1995).

Although the *dsx*RE is the only regulated splicing enhancer thus far characterized, a number of constitutive splicing enhancers have been described (Sun et al., 1993; Watakabe et al., 1993; Dominski & Kole, 1994; von Oers et al., 1994; Ramchatesingh et al., 1995). Most of these elements are short (approximately 10 nt) purine-rich sequences, but a few are pyrimidine-rich. None thus far reported resemble the *dsx* repeat sequences. Both types of enhancers require SR proteins for their activities, but only the *dsx*RE requires Tra and Tra2. The primary difference between constitutive en-

hancers and the *dsx*RE is that the former can function only within 100 nt of the affected 3′ splice site, whereas the latter can function at least 1,000 nt away (Tian & Maniatis, 1994). This ability to function at a distance may be due, in part, to the complex organization of the *dsx*RE, and to Tra and Tra2, which may function to promote enhancer complex assembly and stability. Consistent with both of these possibilities is the observation that individual *dsx* repeats or the PRE function as constitutive (Tra- and Tra2-independent) splicing enhancers in vitro when located within 100 nt of the female-specific 3′ splice site (Lynch & Maniatis, 1995). Although individual repeats can function as Tra- and Tra2-dependent enhancers at a distance (Hoshijima et al., 1991), maximal splicing efficiency in vitro requires the combination of multiple repeats, the PRE and Tra and Tra2 (Tian & Maniatis, 1994; Lynch & Maniatis, 1995).

The unique organization of the *dsx*RE suggests the interesting possibility that this arrangement of sequences results in the formation of a specific secondary and tertiary structure that is required for optimal Tra- and Tra2-dependent splicing. To investigate this possibility, we have compared the sequence of the *D. melanogaster dsx*RE with the corresponding sequence from the distantly related *D. virilis*. In addition, we have conducted chemical and enzymatic RNA probing experiments to investigate the secondary structure of the *dsx*REs from both species. This structural information was then used to optimize computer-assisted RNA-folding predictions. These studies revealed significant phylogenetic conservation of the primary sequence of the repeat elements, and the secondary structure of the complete element.

We also conducted RNase footprinting experiments with purified Tra and Tra2 proteins and both *dsx*REs to investigate specific protein–RNA interactions in the complex. Based on these assays, we found that Tra2 binds to each of the repeat elements with affinities that are indistinguishable, whereas Tra binds nonspecifically to the *dsx*RE. We conclude that the organization of the *dsx*RE results in the formation of an RNA secondary structure in which each of the repeats is present as single-stranded RNA that is recognized specifically by the RRMs of both Tra2 and SR proteins.

RESULTS

Comparison of the *D. melanogaster dsx*RE and the corresponding region in the *D. virilis dsx* pre-mRNA

The female-specific fourth exon of *D. virilis* was identified from a genomic DNA library (Newfeld et al., 1991) using the PCR-amplified third intron of *D. virilis* as the hybridization probe (intron sequence from Burtis & Baker, 1989). The DNA sequence of this exon was determined and compared to the corresponding region of the *D. melanogaster dsx* gene. An alignment of the two sequences revealed several highly conserved regions between the two species (Fig. 2). The first 100 nt of the fourth exon are 90% identical between the two species, encoding only 1 different amino acid of the 30 translated amino acids (98% homology on the amino acid level). Because this region contains the coding sequence for the carboxy terminus of the female-specific *dsx* protein, this high degree of sequence conservation is not unexpected. Previously, an interspecific nucleotide sequence comparison of the *Drosophila* hsp82 gene demonstrated that the coding regions of the distantly related *D. melanogaster* and *D. virilis* species are 90% homologous at the DNA level and 97–99% identical at the amino acid level (Blackman & Meselson, 1986). In contrast, little or no sequence conservation was observed in the intron or the nontranslated exon 1 sequences of hsp82. Consistent with this observation, the conservation of noncoding sequences of exon 4 in *dsx* is weak, except at the 5′ and 3′ ends of the *dsx*RE and the repeat sequences. As shown in Figure 2, there are two regions of approximately 30 nt at the 5′ and 3′ ends of the *dsx*RE that are highly conserved. Although these regions in the *D. melanogaster dsx*RE are not required for maximal levels of Tra- and Tra2-dependent splicing in vitro (K.W. Lynch & T. Maniatis, unpubl.), they may be required for a regulatory function in vivo.

By introducing gaps into both *dsx*RE sequences, it is possible to align the 13-nt repeat sequences of the *D. melanogaster dsx*RE with nearly identical sequences in the *D. virilis* exon 4. In agreement with a recent study (Heinrichs & Baker, 1995), six repeat elements are present in the *D. melanogaster dsx*RE, whereas only four such repeats were found in the enhancer region of *D. virilis*. In addition, the nucleotide composition of the 13-nt repeat elements varies slightly in *D. melanogaster* (Burtis & Baker, 1989), whereas all of the repeats in *D. virilis* are identical to each other and to the predominant repeat sequence in *D. melanogaster*. Thus, the repeat sequences are highly conserved between the two species, but the surrounding sequences are nearly random, suggesting that specific sequences are not required for *dsx*RE function outside of the repeats.

The *D. virilis dsx*RE does not contain a sequence that is identical to the *D. melanogaster* PRE. However, there is a purine-rich sequence located immediately downstream of the fourth repeat in the *D. virilis dsx*RE. The factors that bind to this purine-rich sequence are similar to the factors associated with the *D. melanogaster* PRE, and both purine-rich sequences are sufficient to act as a constitutive enhancer when in close proximity to the weak 3′ splice site (K.W. Lynch & T. Maniatis, unpubl.). Thus, the function, but not the exact sequence, of the PRE may be conserved between the two species.

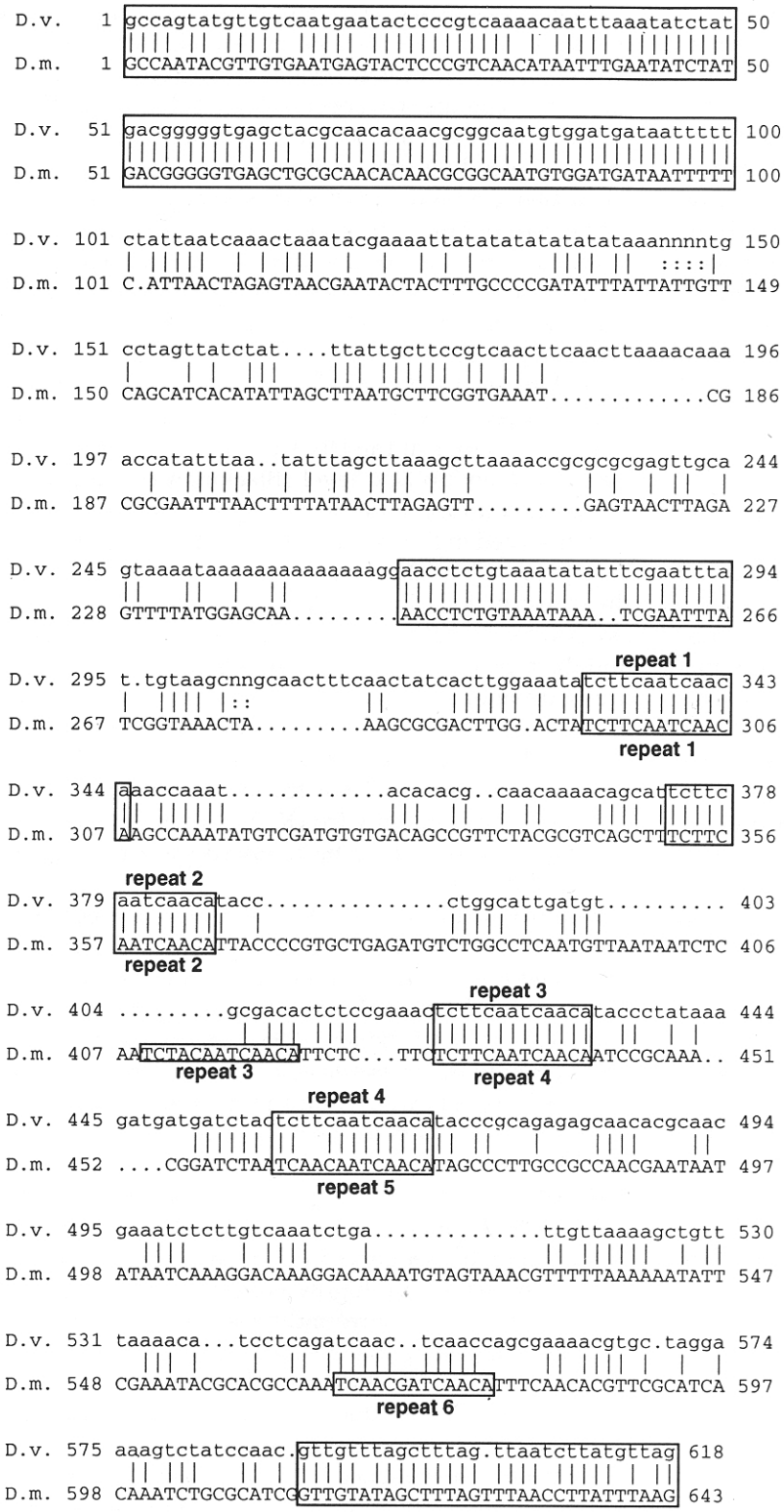

FIGURE 2. Sequence alignment of the female-specific fourth exon from *D. melanogaster* and the distantly related *D. virilis*. Shown are the first 650 nt of the fourth exon, including the *dsx*RE. Vertical bars indicate identical nucleotides. Alignment was accomplished by using the Sequence Analysis Software Package, version 7.2, from the Genetics Computer Group. The 13-nt repeat elements identified in each species are labeled and boxed. Other boxed regions are regions of high sequence homology.

Comparison of the *D. melanogaster* and *D. virilis* splicing enhancer activities in HeLa cell nuclear extracts

To investigate the biological significance of the sequence conservation between the *D. melanogaster* and *D. virilis dsx*REs, we compared their activities in an in vitro splicing assay. To accomplish this, a pre-mRNA substrate (D2) was constructed in which the *dsx*RE of *D. melanogaster* was replaced by the enhancer region of *D. virilis* containing only four repeats. Consistent with earlier studies, splicing of substrate D1 was observed

only in the presence of Tra and Tra2 (Fig. 3). Similarly, the in vitro splicing activity of the *D. virilis dsx*RE required *D. melanogaster* Tra and Tra2, and the concentrations of these proteins required for maximal splicing efficiency are indistinguishable from the concentrations required for substrate D1. Thus, the *dsx*RE from both species can substitute functionally for each other in in vitro splicing assays.

The structural analysis of the *dsx* enhancer described below was conducted with RNAs containing portions of the fourth exon, but lacking the adjacent intron. It

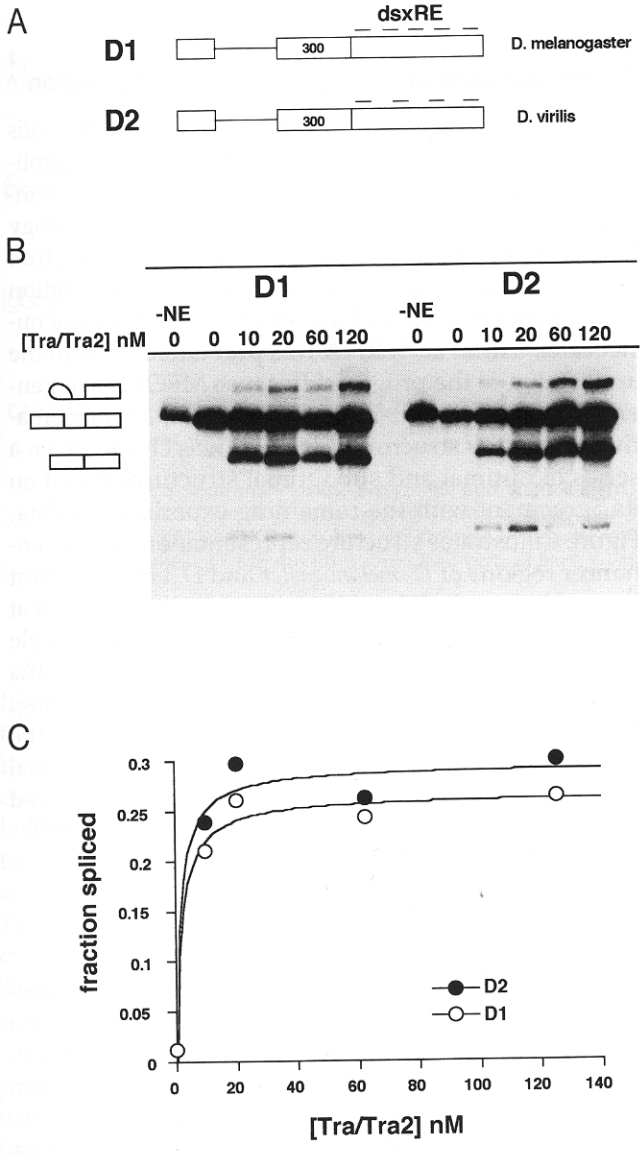

FIGURE 3. *dsx*RE from *D. virilis* can functionally substitute for the *dsx*RE in *D. melanogaster*. **A:** Both oligonucleotides D1 and D2 share part of exon 3 and the entire regulated intron derived from *D. melanogaster*. Substrate D2 contains the *D. virilis dsx*RE in place of the *D. melanogaster dsx*RE. **B:** Splicing efficiencies of D1 and D2 are compared as a function of Tra/Tra2 concentration. **C:** Quantitation of the data in B. At the Tra/Tra2 concentrations used, the splicing efficiency of D1 (open circles) is indistinguishable from the splicing efficiency observed for D2 (closed circles).

is therefore important to demonstrate that the isolated *dsx*RE can form a specific regulatory complex. In fact, several lines of evidence show that *dsx*RE RNA fragments are capable of specifically binding to or competing for Tra/Tra2. For example, an isolated *dsx*RE containing all six repeat elements can specifically inhibit the splicing of D1 (Tian & Maniatis, 1993) and it interacts specifically with Tra and Tra2 in nuclear extracts (Tian & Maniatis, 1992). Similar competition experiments conducted with isolated enhancer elements used here are in agreement with the results of Tian and Maniatis (1993). The titration of enhancer competitor RNA reduced the splicing efficiency of D1 dramatically, whereas the presence of a nonspecific competitor at the same concentration had no significant effect (data not shown). In addition, the splicing efficiency of the Tra/Tra2-independent β-globin pre-mRNA was unaffected by either specific or nonspecific competitor RNA. Thus, the isolated *dsx*REs inhibit *dsx* splicing specifically by competing for *trans*-acting factors that are not components of the basic spliceosome, but are essential for the Tra/Tra2-dependent recruitment of the spliceosome to the *dsx* pre-mRNA. It is therefore reasonable to argue that enhancer elements in the absence of any splice sites are capable of binding to or competing for regulatory factors required for female-specific splicing of *dsx* pre-mRNA.

Secondary structure analysis of the *D. melanogaster* and *D. virilis dsx*REs

Direct enzymatic and chemical in vitro structure probing was used to identify regions within the enhancer elements that are single stranded or involved in secondary or higher-order interactions (Knapp, 1989; Krol & Carbon, 1989). The digestion pattern of the *D. melanogaster dsx*RE RNA revealed that the 13-nt repeat elements are predominantly in a single-stranded configuration (Fig. 4). Our analysis also identified several base paired regions with different sensitivities to RNase V1, an enzyme that specifically recognizes nucleotides involved in base pairing. Similarly, absence of RNase T1 digestion is indicative of guanosine residues that are not accessible for modification. These nucleotides are thought to be in the immediate vicinity of or directly involved in higher-order structural arrangements.

Three different permutations of the *D. melanogaster* enhancer region (R1-6, R2-5PRE, R2-6) were used to address whether the nucleotides 5′ or 3′ from the repeat elements influence the folding pattern. Neither the removal of the first nor the last repeat element of the *dsx*RE changed the digestion pattern (data not shown). Thus, the in vitro folding of the enhancer element does not depend on interactions between the 5′ and 3′ ends of the RNA.

In order to substantiate the secondary structure information of the *D. melanogaster* enhancer region ob-

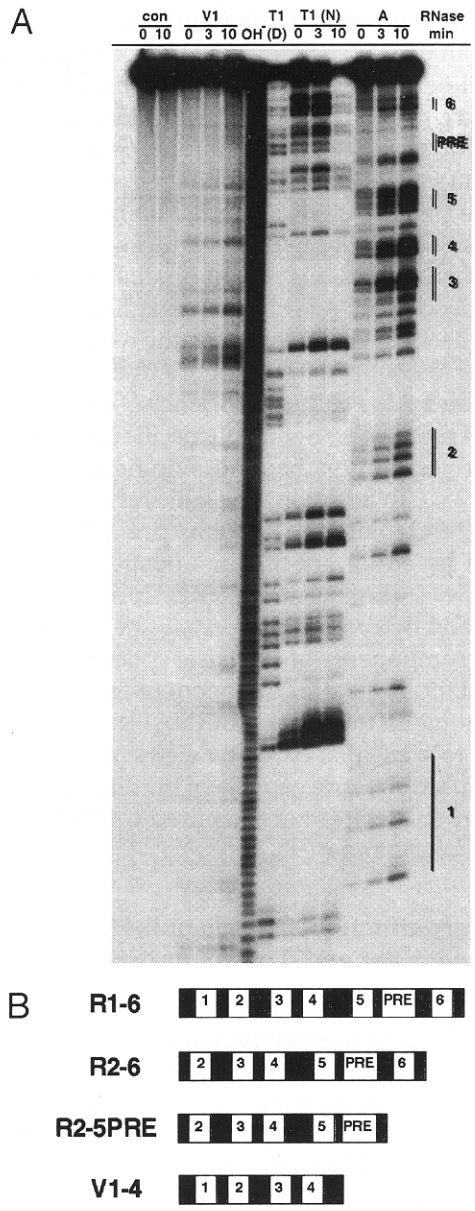

FIGURE 4. In vitro structure probing of the *dsx*RE. **A:** RNase digestion pattern of R1-6. Lanes: con, no RNase added; V1, RNase VI digestion; OH^-, alkaline hydrolysis ladder; T1 (D), RNase T1 digestion at denaturing conditions; T1(N), RNase T1 digestion at assay conditions; A, RNase A digestion. Locations of repeats 1–6 and the PRE are indicated on the right side. **B:** Summary of the various oligonucleotides used for the *dsx*RE structure probing. R1-6 contains all repeats present in the *dsx*RE of *D. melanogaster* and the PRE; R2-5PRE contains repeats 2–5 and the PRE; R2-6 contains repeats 2–6, including the PRE, and V1-4 contains the *dsx*RE from *D. virilis* containing all four repeat elements.

tained by nuclease digestion, the RNA was treated with the single-stranded chemical modifier DMS followed by primer extension (Krol & Carbon, 1989). DMS methylates adenine and cytosine residues at the N-1 and N-3 positions, respectively, with some preference for adenine. Residues that interact with other nucleotides through N-1- or N-3-mediated hydrogen bonding are not accessible for the chemical modification. The results of a series of DMS methylation experiments are summarized in Figure 5A. As observed in the RNase digestion experiments, the 13-nt repeat elements are very accessible to chemical modification and are therefore in a predominantly single-stranded conformation. By contrast, the PRE appears to be in a predominantly base-paired configuration.

Similar experiments were conducted with the *D. virilis dsx*RE. As with the *D. melanogaster dsx*RE, the 13-nt repeat elements of the *D. virilis* RNA are predominantly in single-stranded regions. Only one of the four repeats appears to be involved in some secondary structure (Fig. 5B).

Computer-assisted folding of the enhancer region

The enhancer regions of *D. melanogaster* and *D. virilis* were folded using the MFOLD and PLOTFOLD application programs (version 7.2) from the Genetics Computer Group, University of Wisconsin Biotechnology Center. With the availability of biochemical structure data collected in the RNase and chemical modification experiments described above, the folding of several nucleotides within each RNA was prevented prior to the application of the program. Because MFOLD can generate and analyze suboptimal structures, a representative secondary structure depiction was chosen from a series of optimal and suboptimal structures based on its agreement with the remaining experimental data. Figure 6 illustrates structure representations for the enhancer regions of *D. melanogaster* and *D. virilis* that best fit the biochemical data. With the exception of repeat 2 in each enhancer, all of the 13-nt repeats are single stranded. In the representations, approximately 40% of each RNA is involved in Watson–Crick base pairing. This is a relatively low percentage compared to the well-defined RNA structures within the ribosomal RNAs (Noller, 1984), but similar to the extent observed for U2 snRNA in *Tetrahymena* (Zaug & Cech, 1995).

Tra/Tra2-dependent in vitro footprinting of the enhancer region

Previous studies have shown that Tra2 can bind specifically to the *dsx*RE (Hedley & Maniatis, 1991) or a short oligonucleotide comprised of two repeat elements (Inoue et al., 1992), but the site of this interaction is not known. Therefore, we conducted in vitro footprinting studies to identify Tra2 binding sites within the *dsx*RE. An initial binding specificity screen was established to evaluate the binding specificity of Tra or Tra2 to the enhancer probe. In this screen, a mixture of 5′ end-labeled oligonucleotides cleaved at a single residue by alkaline hydrolysis was mixed with increasing concentrations of Tra or Tra2, and the bound molecules were recovered by retention on nitrocellulose

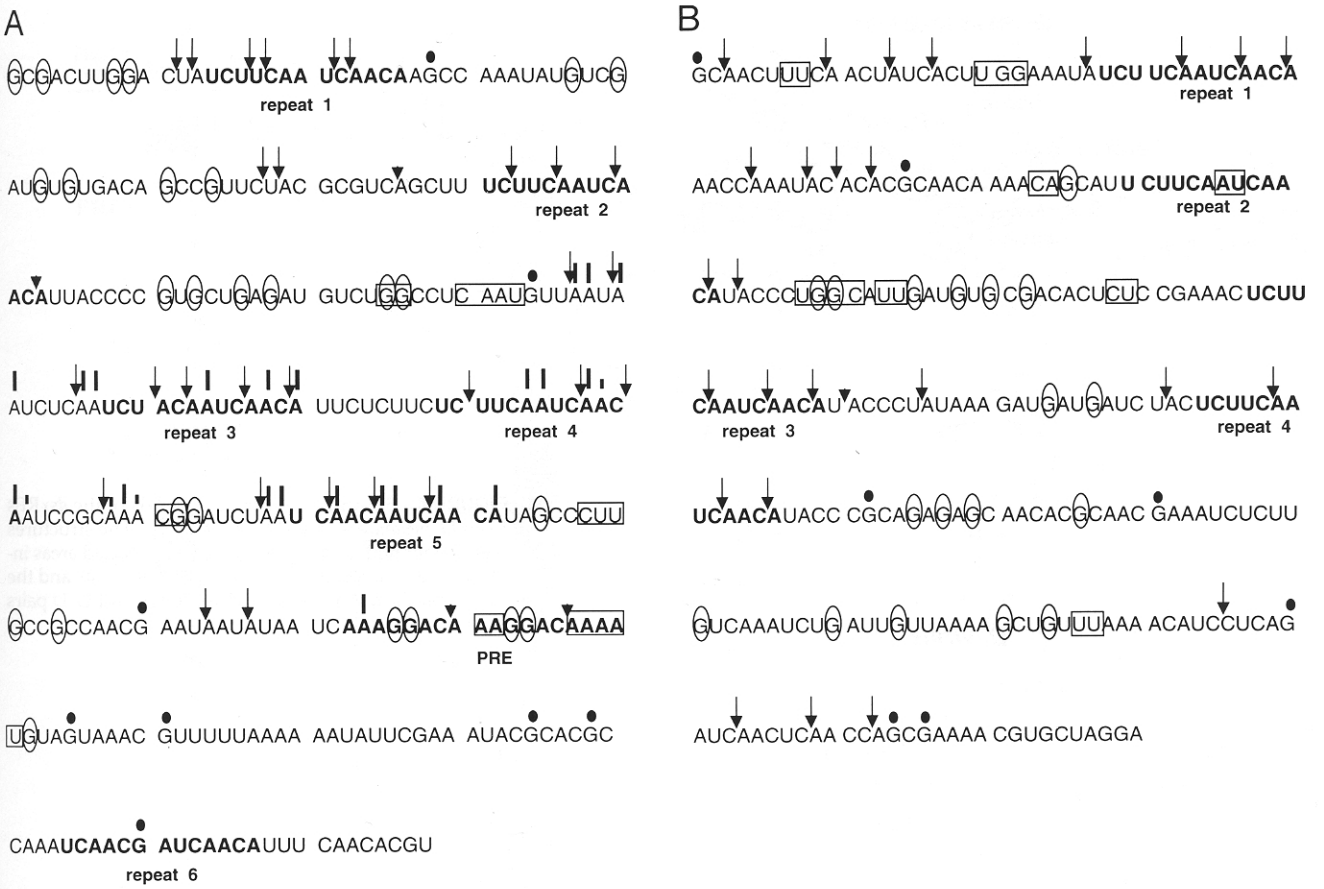

FIGURE 5. Summary of the enzymatic and chemical structure probing of the *dsx*RE from (**A**) *D. melanogaster* and from (**B**) *D. virilis*. Symbols: long arrows, strong RNase A cut; small arrows, weak RNase A cut; boxed nucleotides, moderate to strong RNase V1 cut; solid dots, strong T1 cuts; open ovals, G residues protected from T1 digestion; long vertical bars, strong sites of DMS modification; small vertical bars, weak sites of DMS modification. Repeat elements in each *dsx*RE are in bold.

filters (Gott et al., 1993). Analysis of the oligonucleotides recovered from the filters indicated that retention by Tra binding is very efficient for all truncated RNAs, even for those that contain only *dsx* unrelated polylinker sequences. In contrast, only those RNAs containing one or more complete 13-nt repeat elements were retained on the filter by Tra2 (data not shown). Thus, consistent with previous filter-binding and UV-crosslinking studies (Lynch & Maniatis, 1995, 1996), Tra interacts with RNA with little or no specificity, whereas Tra2 binds specifically to the repeat elements.

To determine whether Tra2 can protect specific regions of the *dsx*RE from RNAse digestion, the R2-5PRE was subjected to RNase A digestion in the presence of increasing concentrations of Tra2. As shown in Figure 7, all of the repeat elements present in R2-5PRE remain accessible to RNase A digestion at Tra2 concentrations lower than 20 nM. Surprisingly, all of the repeat elements except repeat 2 are protected to similar degrees at concentrations of ≥100 nM Tra2. This observation correlates well with the observed binding affinity of $K_d = 50$ nM for Tra2 measured under identical conditions by a nitrocellulose filter-binding assay (data not shown). The absence of selective binding of Tra2 to individual repeats indicates that the 13-nt repeat elements represent multiple Tra2 binding sites within the enhancer region. These binding sites, except repeat 2, are occupied at similar Tra2 concentrations. Similarly, Tra2 binds to the PRE at approximately the same concentration. Interestingly, our RNA structural probing data (Fig. 4) and computer-assisted folding analysis predict that the PRE is primarily in a double-stranded configuration. However, with increasing concentrations of Tra2, part of the PRE sequence becomes more susceptible to RNase A cleavage and part of the region is protected from this cleavage (Fig. 7). Thus, Tra2 binding appears to induce a conformational change in the PRE RNA.

A series of modification interference experiments were performed to identify whether specific nucleotides within the enhancer region are essential for Tra or Tra2 binding. The protocol for modification interference requires conditions to achieve a single modification per RNA molecule and relies subsequently on the

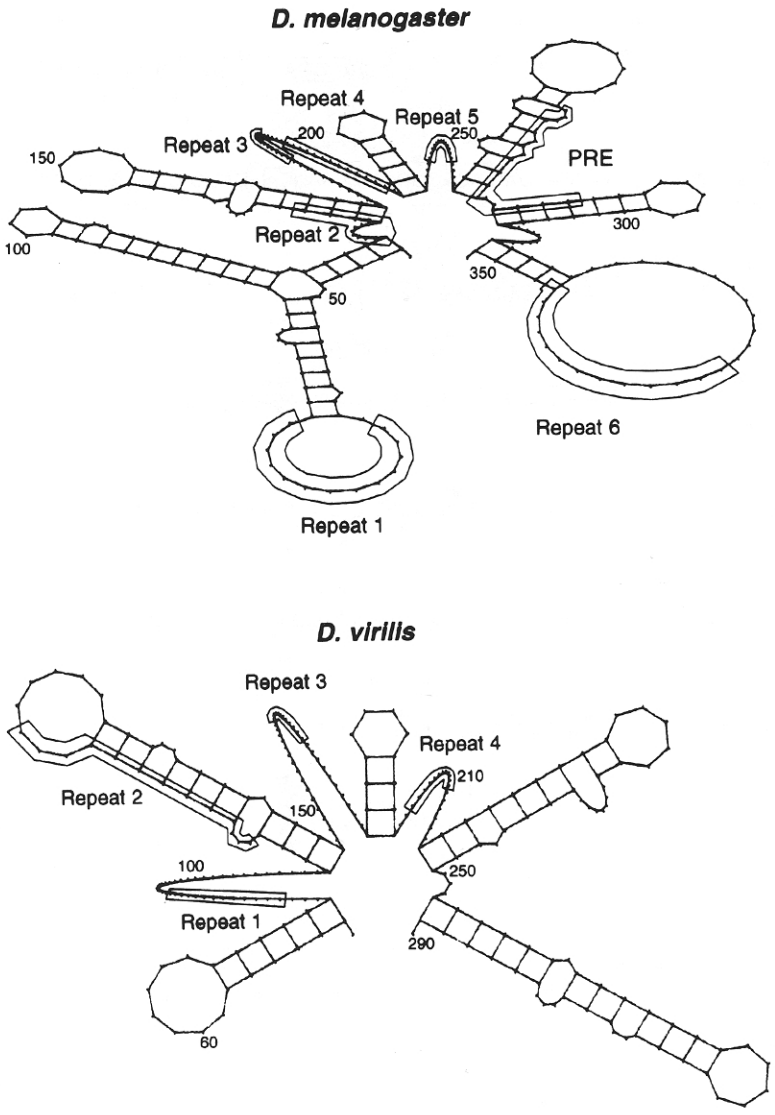

FIGURE 6. Secondary structure models of the *dsx*REs from *D. melanogaster* and from *D. virilis*. The structures were generated as described in the text. Boxed areas indicate the position of the 13-nt repeat elements and the PRE. Standard Watson–Crick base pairs and G·U pairs are indicated by bars.

separation of free RNA from the bound species by a nitrocellulose filter-binding assay (Conway & Wickens, 1989). The comparison of RNA molecules selected by Tra or Tra2 binding with the initial RNA pool did not lead to the identification of nucleotides essential for the interaction of Tra or Tra2 with the enhancer (data not shown). This data supports the above conclusion that the enhancer region contains not one high-affinity site for the interaction with Tra2, but several. Although Tra2 is capable of binding to the *dsx*RE in the absence of additional proteins, we note that the specificity of Tra2 dramatically increases in the presence of SR proteins (Lynch & Maniatis, 1995) and in nuclear extracts (Lynch & Maniatis, 1996) under splicing conditions. Thus, the experiments presented here show only that each of the repeats can bind to Tra2 specifically. They do not address the nature of the RNA–protein complex assembled on a functional *dsx*RE.

Antisense inhibition of *dsx* pre-mRNA splicing

To determine whether the single-stranded configuration of the 13-nt repeats in the *D. melanogaster dsx*RE is required for enhancer function, we conducted experiments to determine whether the repeats can function in a double-stranded configuration. An antisense oligonucleotide complementary to the 13-nt repeat was annealed to the *dsx*RE and the effects on pre-mRNA splicing were examined. As shown in Figure 8, the presence of the antisense oligonucleotide has a dramatic effect on the splicing efficiency of substrate D2. In the absence of antisense oligonucleotide, approximately 25% of the substrate is spliced. In contrast, no spliced products are detected above the degradation levels regardless of whether the substrate was preincubated with the antisense oligonucleotide at conditions identical to those used in the secondary structure anal-

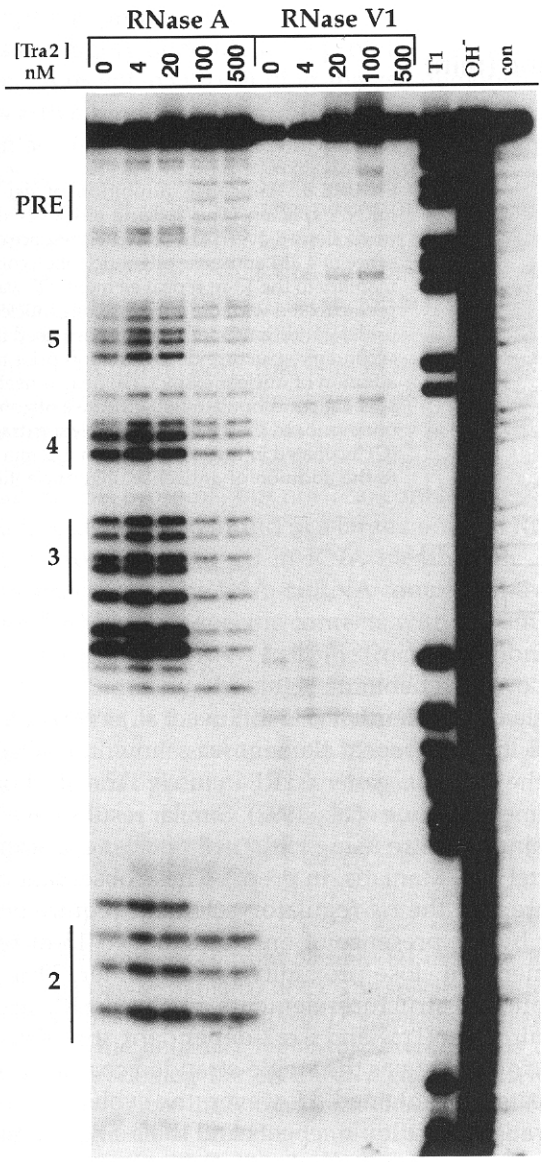

FIGURE 7. RNase protection pattern of R2-5PRE as a function of increasing concentrations of Tra2. Lanes: T1, RNase T1 digestion at denaturing conditions; OH^-, alkaline hydrolysis; con, no RNase added. Eight microliters of 5′-labeled R2-5PRE in reaction buffer was allowed to equilibrate with increasing concentrations of Tra2 at 30 °C for 20 min. Each binding reaction was then adjusted with either 2 μL of a 0.02 U/mL solution of RNase A for 8 min or with 2 μL of a freshly prepared 3.5 U/mL solution of RNase V1 for 10 min at 30 °C. Positions of the 13-nt repeat elements and the PRE are indicated on the left side.

ysis (Fig. 8A), heat annealed prior to the addition of nuclear extract (Fig. 8B), or whether the antisense oligonucleotide was added shortly after the incubation of D2 in nuclear extract (Fig. 8C). Thus, the *dsx* enhancer presents the repeats in a single-stranded configuration in the absence (Fig. 8A) and in the presence of nuclear extract (Fig. 8C). Similar results were obtained in antisense experiments using D1 as the substrate with the exception of a significantly reduced splicing efficiency for heat-treated D1 in the absence of antisense (data not shown). In control experiments, the presence of the antisense oligonucleotide at concentrations that inhibited *dsx* pre-mRNA splicing did not affect the splicing of β-globin pre-mRNA (data not shown). We conclude that the single-stranded character of the 13-nt repeats is essential for splicing enhancer activity. Thus, the conservation of both the sequence and the structure of the 13-nt repeats are essential for *dsx* splicing enhancer activity.

DISCUSSION

On the basis of results presented here, we propose that the *dsx*RE adopts a secondary structure that optimizes interactions between individual repeat elements and the RNA-binding domains of Tra2 and SR proteins. In all cases, except repeat 2, the repeats are present in a single-stranded configuration. This structural difference correlates with the observation that repeat 2 is not protected as efficiently from RNase A digestion at Tra2 concentrations that maximally protect the other repeats. Thus, the single-stranded character of the repeats may be essential for the formation of a stable multiprotein complex consisting of Tra2, Tra, SR proteins, and possibly other nuclear proteins (Tian & Maniatis, 1993). This stable enhancer complex can then facilitate the recruitment of general splicing factors to the upstream female-specific 3′ splice site (Zuo & Maniatis, 1996).

A comparison of the *dsx*REs from *D. melanogaster* and *D. virilis* revealed conserved sequences at the 5′ and 3′ ends of the element, and conserved repeat sequences that are separated by highly divergent RNA sequences. Although attempts to identify a function for the conserved 5′ and 3′ sequences by in vitro splicing assays have thus far failed (K.W. Lynch & T. Maniatis, unpubl.), it seems likely that this conservation is important in flies. The conservation of the repeat sequences is almost certainly due to a conserved recognition by Tra and Tra2, because we have shown that the *D. melanogaster* and *D. virilis dsx*REs are functionally interchangeable, and both require *D. melanogaster* Tra and Tra2 for their functions. In addition, a recent study demonstrated that the *D. virilis* Tra homologue can partially rescue the Tra mutant phenotype in transgenic *D. melanogaster* (O'Neil & Belote, 1992). Although the primary sequence and the detailed secondary structures of the inter-repeat sequences are poorly conserved, our data indicate that the repeats in both species are maintained as single-stranded RNA regions. Thus, the inter-repeat structure may have evolved to maintain the repeats in this configuration. The importance of maintaining the repeats in a single-stranded configuration was demonstrated by showing that a 13-nt RNA complementary to the repeats inhibits its enhancer-dependent splicing. Thus, the repeat ele-

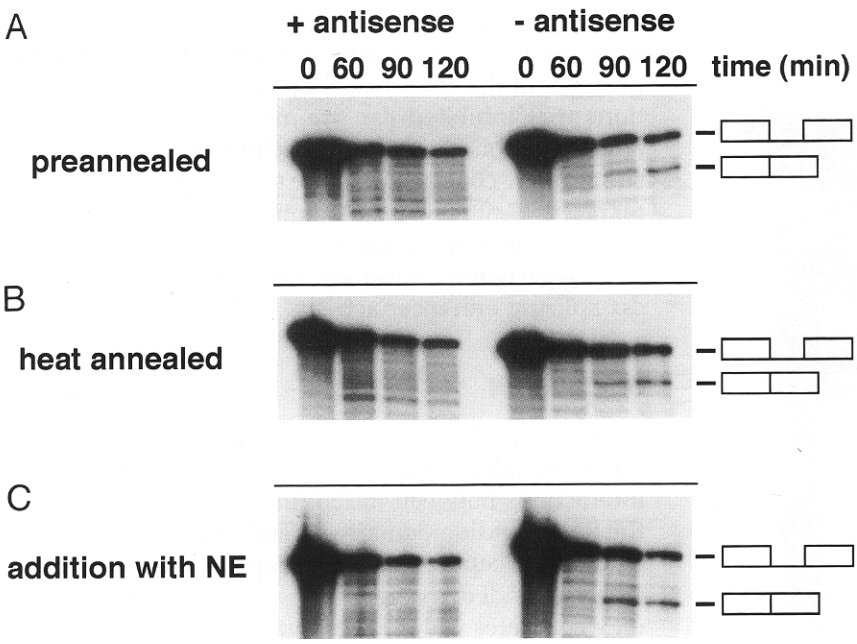

FIGURE 8. Antisense inhibition of *dsx* pre-mRNA splicing. The splicing efficiency of D2 was followed over 120 min in the presence or absence of 4 μM antisense oligonucleotide complementary to the 13-nt repeat element. D2 was (**A**) preincubated with the antisense oligonucleotide under conditions identical to those used in the secondary structure determination prior to the addition of nuclear extract, (**B**) heat annealed at splicing conditions to the antisense oligonucleotide prior to the addition of nuclear extract, or (**C**) incubated in nuclear extract for 1–2 min prior to the addition of antisense oligonucleotide.

ments within the *dsx* enhancer are only recognizable by Tra/Tra2 and SR proteins when presented in a single-stranded configuration.

Comparison of the *D. melanogaster* and *D. virilis* sequences revealed no conservation of the PRE identified in the *dsx*RE from *D. melanogaster* (Lynch & Maniatis, 1995). However, a distinct purine-rich sequence is found in the *D. virilis dsx*RE, and this sequence is capable of functioning as a constitutive splicing enhancer (K.W. Lynch & T. Maniatis, unpubl.). Surprisingly, our RNA structural probing data and the computer-assisted folding analysis indicate that the PRE in *D. melanogaster* is primarily in a duplex configuration, even though Tra, Tra2, and SR proteins crosslink specifically to this sequence in nuclear extracts (Lynch & Maniatis, 1996). RNA footprinting data (Fig. 7) suggests that Tra2 binding induces a conformational change in the PRE.

Although the secondary structure representations of the *dsx*RE, which are shown in Figure 6, are consistent with the biochemical data, it is important to note that each provides only one of many possible and very similar configurations. There are several architectural similarities in the *dsx*RE structures obtained. Repeats 3 and 4 in *D. virilis* are flanked by extended hairpins and separated by a four-base pair hairpin. In *D. melanogaster*, the structural format resembles that observed in *D. virilis*, with the exception that an additional 13-nt repeat element is included. Thus, repeats 3 and 5 are flanked by extended helical regions and repeats 4 and 5 are separated by a small but stable hairpin. In both proposed structures, repeat 2 is involved in the formation of the hairpin that flanks repeat 3.

There is little information to speculate on the functional importance of the proposed secondary structure elements flanking each repeat element. Previously, a single repeat element (Hoshijima et al., 1991) or a synthetic tandem repeat element was shown to substitute for the *D. melanogaster dsx*RE in in vivo transfection experiments (Inoue et al., 1992). Similar results have been obtained in vitro using HeLa cell nuclear extracts (K.J. Hertel & T. Maniatis, in prep.). These observations indicate that the *cis*-regulatory element requirement is met by the presence of only one or two 13-nt repeat elements in close proximity to each other. Thus, the secondary structure elements in the *dsx*RE are not required for Tra- and Tra2-dependent stimulation of splicing as long as the single-stranded character of the repeats is maintained. However, the evolutionary conservation of multiple repeats and their single-stranded nature argue strongly that both are required for the fine-tuned regulation of *dsx* pre-mRNA splicing in flies. For example, five of six or three of four repeat elements in *D. melanogaster* and in *D. virilis*, respectively, are in single-stranded configuration. Thus, the inter-repeat structure elements might influence indirectly the efficiency of splice site activation by maintaining the accessibility of the repeat elements. In flies, where the levels of Tra and Tra2 are likely to be less than those generated in transfection or in vitro experiments, this arrangement may be essential for the controlled function of the *dsx*RE. By contrast, the high levels of Tra and Tra2 used in in vitro experiments or produced in cotransfection experiments are sufficient to observe splice site activation with only a single repeat element.

In addition to the structural analysis of the *dsx*RE, the footprinting and terminal truncation results have shown that the 13-nt repeat elements can act as binding sites for Tra2. The data are therefore consistent with the model of a direct interaction of Tra2 with the 13-nt repeat element. Other lines of evidence indicate

that specific protein–RNA interactions within the repeats are highly dependent on protein–protein interactions. A recent in vitro binding analysis demonstrated cooperative binding of Tra, SR proteins, and Tra2 to the intact *dsx*RE (Lynch & Maniatis, 1995). When assayed in nuclear extracts, efficient binding of Tra2 to the 13-nt repeat is highly dependent on the presence of Tra (Lynch & Maniatis, 1996). Given these observations, it is very likely that the number of repeat sequences in the *dsx*RE and their context-dependent accessibility to Tra, Tra2, and SR proteins results in protein–RNA interactions that lead to the formation of an enhancer complex capable of promoting 3′ splice site recognition at a distance. This property is not shared with simple constitutive enhancer elements.

Tra2 and SR proteins, but not Tra, contain the RRM RNA-binding domain found in a large family of RNA-binding proteins involved in RNA metabolism. The crystal structure of the U1A–snRNA complex showed that the RRM makes specific contacts with the single-stranded region of an RNA hairpin structure (Oubridge et al., 1994). Similarly, the single-stranded nature of the repeat elements of the *dsx*RE is consistent with the possibility that they are recognized by the RRM of Tra2 and possibly SR proteins.

MATERIALS AND METHODS

RNA

D1 RNA was synthesized from plasmid D1 using T7 RNA polymerase (Tian & Maniatis, 1992). The splicing substrate, D2, in which the enhancer region of *D. melanogaster* was substituted with the enhancer region of *D. virilis*, was constructed by subcloning the enhancer region (inclusive of repeat 1 to just before the 3′ conserved region) into a PCR-generated *Eco*R I site of D1 located just upstream of the first repeat sequence (Lynch & Maniatis, 1995).

The probes R2-5PRE, R2-6, and D6 were generated as described previously (Lynch & Maniatis, 1995). The construct encoding R1-6 was made by cloning a fragment containing the T7 transcription start site and repeat 1 into the *Mlu* I site of R2-6; RNA was then synthesized by in vitro transcription. V1-4 was transcribed from a construct in which the *D. virilis* enhancer PCR fragment from D2 was cloned into the *Eco*R I site downstream of the T7 promoter in SP72.

Splicing substrates were labeled uniformly with [^{32}P]UTP. 5′-End-labeling of the oligonucleotides R2-5PRE, R2-6, R1-6, and V1-4 synthesized by T7 RNA polymerase was accomplished by removing 5′ triphosphates with calf intestine phosphatase followed by reaction with [γ-^{32}P]ATP and T4 polynucleotide kinase. Oligonucleotide concentrations were determined from specific activities for radiolabeled RNAs, assuming a residue extinction coefficient of $8.5 \times 10^3\ M^{-1}\ cm^{-1}$ at 260 nm for nonradioactive RNA.

In vitro splicing reactions

In vitro splicing reactions were generally conducted as described in Tian and Maniatis (1992). The RNA antisense experiments were conducted using identical conditions except for the presence of 4 μM antisense oligonucleotide and approximately 50 nM poly I-C. The presence of poly I-C was required to reduce the activity of a double-stranded deaminase (dsRad) activity present in HeLa cell nuclear extracts (Yang et al., 1995). Control experiments with *dsx* pre-mRNA and β-globin pre-mRNA established that the presence of poly I-C did not affect splicing efficiency significantly. The splicing substrate was incubated under splicing conditions with the antisense RNA oligonucleotide for 30 min at 30 °C with or without a prior heat anneal step (95 °C for 1.5 min in the absence of $MgCl_2$, then add $MgCl_2$). The splicing reaction was then initiated by the addition of nuclear extract, poly I-C, and Tra/Tra2. The final concentrations were 30% (v/v) nuclear extract, 4 μM antisense RNA, 50 nM poly I-C, 50 nM Tra, and 50 nM Tra2 in a volume of 50 μL. In another experiment, the substrate was incubated with nuclear extract for 1–2 min prior to the addition of poly I-C, antisense RNA, and Tra/Tra2.

Recombinant proteins

Recombinant Tra and Tra2 were expressed in baculovirus and purified as described in Tian and Maniatis (1992).

Cloning and sequencing of the *D. virilis doublesex* female-specific exon

The female-specific exon of *D. virilis doublesex* gene was isolated from a *D. virilis* genomic library constructed in EMBL3, which was kindly provided by Stuart Newfeld. The library was screened through successive rounds of high-stringency hybridization to a DNA probe that contained the sequence of the third intron of the *D. virilis doublesex* gene. The probe was isolated by PCR from *D. virilis* genomic DNA using primers based on the published *D. virilis doublesex* intron sequence (Burtis & Baker, 1989). After a positive clone was identified, the phage DNA was isolated, digested with various restriction enzymes, and analyzed by Southern blot to determine the minimal fragment that contained the intron sequence. A 1.8-kDa *Bst*Y I fragment, which hybridized strongly to the intron probe, was then subcloned into SP73 and sequenced using the T7 and SP6 priming sites.

Enzymatic and chemical structure probing

All reactions were conducted at 30 °C in a buffer containing 72 mM KCl, 12 mM HEPES, pH 7.9, 3.2 mM $MgCl_2$, 1 mM ATP, 20 mM creatine phosphate, and 4% glycerol. These conditions were chosen to mimic those used in the splicing reaction. For the enzymatic probing of the enhancer RNAs, the RNases A, T1, and V1 were used. RNase A and T1 are single-stranded and nucleotide-specific RNases leaving 3′-phosphate products. RNase A cleavage is pyrimidine-specific with a preference for CpN bonds (Knapp, 1989). RNase T1 recognizes GpN bonds. Both enzymes remain active in EDTA. RNase V1 was used to determine which portions of the RNA are found base paired at the conditions used. V1 requires the presence of Mg^{2+} for activity. In a typical structure probing experiment, a trace amount of 5′ end-labeled RNA in 10 μL reaction buffer was incubated with either 0.02 U/mL RNase A, 0.4 U/mL T1, or 0.07 U/mL V1.

Time points (3 μL) were taken at appropriate time intervals, mixed with a formamide buffer containing 10 mM EDTA, 0.02% bromophenol blue, and 0.02% xylene cyanol and immediately frozen to −70 °C until all time points were collected. Each time point was then subjected to 6% PAGE. For all RNA probes tested, the presence of carrier tRNA or the addition of a denaturing/renaturing step prior to the digestion did not result in an altered susceptibility to the RNases used.

In addition to the enzymatic probing, chemical base modification assayed by reverse transcription was used to examine the secondary structure of the *D. melanogaster* enhancer region. The DMS treatment was conducted according to Zaug and Cech (1995). The modified RNAs were annealed to a 20-nt [γ-^{32}P]-end-labeled DNA primer, ATTTTGTCCTTTGT CCTTTG, corresponding to sequences between the fifth and sixth repeats. The annealed primer/RNA complex was then extended with SuperScript reverse transcriptase (Gibco BRL). Typically, 0.5 μg of R2-5PRE in 50 μL of reaction buffer (90 nM R2-5PRE) was incubated for varying times with 1–3 μL of a 30% (v/v) DMS/ethanol mix. Each time point was quenched with 0.5× volume of 0.75 M sodium acetate and 0.5 M β-mercaptoethanol. After ethanol precipitation and resuspension, approximately 0.1 μg of the modified RNA in 10 μL (80 nM R2-5PRE) was annealed to a fourfold molar excess of 5′-end-labeled DNA primer. Each of the 10-μL extension reactions used 2 μL of the annealed primer mixture and were supplemented with 450 mM dNTPs and 100 U of SuperScript reverse transcriptase. After 1 h at 42 °C, reactions were terminated by the addition of an equal volume of formamide buffer, heated to 95 °C for 2 min, and then subjected to 6% PAGE.

Computer analysis

The primary sequence data from the enhancer regions of *D. melanogaster* and *D. virilis* were aligned by computer to determine primary sequence conservation. They were then analyzed independently with the MFOLD and PLOTFOLD application programs of the Sequence Analysis Software Package (version 7.2) from the Genetics Computer Group, University of Wisconsin Biotechnology Center. Version 7.2 makes use of an RNA-folding algorithm developed by Zuker (Jaeger et al., 1989) and incorporates updated bond energies. The secondary structure data accumulated in experiments described above was used to restrict the folding of nucleotides that are predominantly in a single-strand conformation. MFOLD generates optimal and suboptimal structures. These were then analyzed for agreement with the remaining experimental data.

ACKNOWLEDGMENTS

We thank Stuart Newfeld for providing the *D. virilis* genomic library and members of our laboratory for discussion and comments on the manuscript. This investigation was supported by a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research (K.J.H.) and grant GM 42231 from the National Institutes of Health (T.M.).

Received May 13, 1996; returned for revision June 21, 1996; revised manuscript received August 2, 1996

REFERENCES

Amrein H, Gorman M, Nöthiger R. 1988. The sex-determining gene *tra-2* of *Drosophila* encodes a putative RNA binding protein. *Cell 55*:1025–1035.

Amrein H, Hedley ML, Maniatis T. 1994. The role of specific protein–RNA and protein–protein interactions in positive and negative control of pre-mRNA splicing by *Transformer 2*. *Cell 76*:735–746.

Baker BS. 1989. Sex in flies: The splice of life. *Nature 340*:521–524.

Bandziulis RJ, Swanson MS, Dreyfuss G. 1989. RNA-binding proteins as developmental regulators. *Genes & Dev 3*:431–437.

Blackman RK, Meselson M. 1986. Interspecific nucleotide sequence comparisons used to identify regulatory and structural features of the *Drosophila* hsp82 gene. *J Mol Biol 188*:499–515.

Boggs RT, Gregor P, Idriss S, Belote JM, McKeown M. 1987. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the *transformer* gene. *Cell 50*:739–747.

Burtis KC, Baker BS. 1989. *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternative spliced mRNAs encoding related sex-specific polypeptides. *Cell 56*:997–1010.

Conway L, Wickens M. 1989. Modification interference analysis of reactions using RNA substrates. *Methods Enzymol 180*:369–379.

Dominski Z, Kole R. 1994. Identification of exon sequences involved in splice site selection. *J Biol Chem 269*:23590–23596.

Fu XD. 1995. The superfamily of arginine/serine-rich splicing factors. *RNA 1*:663–680.

Gott JM, Pan T, LeCuyer KA, Uhlenbeck OC. 1993. Using circular permutation analysis to redefine the R17 coat protein binding site. *Biochemistry 32*:13399–13404.

Hedley ML, Maniatis T. 1991. Sex-specific splicing and polyadenylation of *dsx* pre-mRNA requires a sequence that binds specifically to *tra-2* protein in vitro. *Cell 65*:579–586.

Heinrichs V, Baker BS. 1995. The *Drosophila* SR protein RBP1 contributes to the regulation of *doublesex* alternative splicing by recognizing RBP1 RNA target sequences. *EMBO J 16*:3987–4000.

Hoshijima K, Inoue K, Higuchi I, Sakamoto H, Shimura Y. 1991. Control of *doublesex* alternative splicing by *transformer* and *transformer-2* in *Drosophila*. *Science 252*:833–836.

Inoue K, Hoshijima K, Higuchi I, Sakamoto H, Shimura Y. 1992. Binding of the *Drosophila* Transformer and Transformer-2 proteins to the regulatory elements of *doublesex* primary transcripts for sex-specific RNA processing. *Proc Natl Acad Sci USA 89*:8092–8096.

Jaeger JA, Turner DH, Zuker M. 1989. Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci USA 86*:7706–7710.

Knapp G. 1989. Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol 180*:192–212.

Krol A, Carbon P. 1989. A guide for probing native small nuclear RNA and ribonucleoprotein structures. *Methods Enzymol 180*:212–227.

Lynch KW, Maniatis T. 1995. Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes & Dev 9*:284–293.

Lynch KW, Maniatis T. 1996. Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila doublesex* splicing enhancer. *Genes & Dev 10*:2089–2101.

Maniatis T. 1991. Mechanisms of alternative pre-mRNA splicing. *Science 251*:33–34.

Nagai K, Oubridge C, Ito N, Avis J, Evans P. 1995. The RNP domain: A sequence-specific RNA binding domain involved in processing and transport of RNA. *Trends Biochem Sci 20*:235–240.

Newfeld SJ, Smoller DA, Yedvobnick B. 1991. Interspecific comparison of the unusually repetitive *Drosophila* locus *mastermind*. *J Mol Evol 32*:415–420.

Noller HF. 1984. Structure of ribosomal RNA. *Annu Rev Biochem 53*:119–162.

O'Neil MT, Belote JM. 1992. Interspecific comparison of the *transformer* gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. *Genetics 131*:113–128.

Oubridge C, Ito N, Evans PR, Teo C-H, Nagai K. 1994. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature 372*:432–438.

Ramchatesingh J, Zahler AM, Neugebauer KM, Roth MB, Cooper TA. 1995. A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Mol Cell Biol 9*:4898–4907.

Ryner LC, Baker BS. 1991. Regulation of doublesex pre-mRNA pro-

cessing occurs by 3′-splice site activation. *Genes & Dev 5*:2071–2085.

Sun Q, Mayeda A, Hampson RK, Krainer AR, Rottmann FM. 1993. General splicing factors SF2/ASF promote alternative splicing by binding to an exonic splicing enhancer. *Genes & Dev 7*:2598–2608.

Tian M, Maniatis T. 1992. Positive control of pre-mRNA splicing in vitro. *Science 256*:237–240.

Tian M, Maniatis T. 1993. A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. *Cell 74*:105–114.

Tian M, Maniatis T. 1994. A splicing enhancer exhibits both constitutive and regulated activities. *Genes & Dev 8*:1703–1712.

van Oers CCM, Adema GJ, Zandberg H, Moen TC, Bass PD. 1994. Two different sequence elements within exon 4 are necessary for calcitonin-specific splicing of the human calcitonin/calcitonin gene-related peptide I pre-mRNA. *Mol Cell Biol 14*:951–960.

Watakabe A, Tanaka K, Shimura Y. 1993. The role of exon sequences in splice site selection. *Genes & Dev 7*:407–418.

Wu JY, Maniatis T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell 75*:1061–1070.

Yang JH, Sklar P, Axel R, Maniatis T. 1995. Editing of glutamate receptor subunit B pre-mRNA in vitro by site specific deamination. *Nature 374*:77–81.

Zaug AJ, Cech TR. 1995. Analysis of the structure of *Tetrahymena* nuclear RNAs in vivo: Telomerase RNA, the self-splicing rRNA intron and U2 snRNA. *RNA 1*:363–374.

Zuo P, Maniatis T. 1996. The splicing factor U2AF^{35} mediates critical protein–protein interactions in constitutive and enhancer-dependent splicing. *Genes & Dev 10*:1356–1368.