MOLECULAR AND CELLULAR BIOLOGY, July 1991, p. 3584–3602
0270-7306/91/073584-19$02.00/0

Vol. 11, No. 7

The Complex Set of Late Transcripts from the *Drosophila* Sex Determination Gene Sex-lethal Encodes Multiple Related Polypeptides

MARK E. SAMUELS,* PAUL SCHEDL, AND THOMAS W. CLINE†

Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544

Received 19 December 1990/Accepted 25 April 1991

Sex-lethal (*Sxl*), a key sex determination gene in *Drosophila melanogaster*, is known to express a set of three early transcripts arising during early embryogenesis and a set of seven late transcripts occurring from midembryogenesis through adulthood. Among the late transcripts, male-specific mRNAs were distinguished from their female counterparts by the presence of an extra exon interrupting an otherwise long open reading frame (ORF). We have now analyzed the structures of the late *Sxl* transcripts by cDNA sequencing, Northern (RNA) blotting, primer extension, and RNase protection. The late transcripts appear to use a common 5′ end but differ at their 3′ ends by the use of alternative polyadenylation sites. Two of these sites lack canonical AATAAA sequences, and their use correlates in females with the presence of a functional germ line, suggesting possible tissue-specific polyadenylation. Besides the presence of the male-specific exon, no additional sex-specific splicing events were detected, although a number of non-sex-specific splicing variants were observed. In females, the various forms of late *Sxl* transcript potentially encode up to six slightly different polypeptides. All of the protein-coding differences occur outside the previously defined ribonucleoprotein motifs. One class of *Sxl* mRNAs also includes a second long ORF in the same frame as the first ORF but separated from it by a single ochre codon. The function of this second ORF is unknown. Significant amounts of apparently partially processed *Sxl* RNAs were observed, consistent with the hypothesis that the regulated *Sxl* splices occur relatively slowly.

Sexual development in fruit flies is one of the best-characterized systems for the study of determination in a metazoan organism. In *Drosophila melanogaster*, the initial signal for specification of the male or female state is the ratio of X chromosomes to autosomes. This signal is recognized and interpreted by an array of genes that act in sequence making the decisions ultimately leading to sexual differentiation (3, 19, 20, 29, 36, 57). A considerable amount is now known regarding the molecular nature of some of these regulatory genes (2, 4, 28).

Central in the pathway leading to sex determination lies the gene Sex-lethal (*Sxl*). In female flies, *Sxl* carries out a variety of related functions. It governs the phenotypic development of sexual dimorphism (18, 53) and is essential for female viability through its involvement in dosage compensation (17, 23, 37). *Sxl* also positively regulates its own activity in a feedback loop believed to be critical for maintaining the determined state throughout development (18). Finally, *Sxl* is required for normal development of the female germ line and may possess different female-specific somatic and germ line activities (48, 50, 54).

Although *Sxl* expresses a suite of functions in females, genetic analysis suggests that it is not active or necessary in males (16, 18, 38, 50). However, the gene is transcribed in males, and the transcripts generated in adult males and females are of different sizes (7, 38, 51). This results from the occurrence in male RNAs of an extra exon not present in female RNAs. The extra exon interrupts an otherwise long open reading frame (ORF), apparently eliminating protein-coding potential and thereby generating inert transcripts in males. The exclusion of the male exon from female transcripts is associated with the expression of full-length *Sxl* polypeptides in female flies (11).

A mechanism for *Sxl* action has been hypothesized on the basis of a comparison of *Sxl* with a gene downstream of it in the sex determination hierarchy: transformer (*tra*). Like *Sxl*, *tra* is transcribed in both sexes but is differentially spliced (10, 14, 40). Sex-specific alternative splicing of *tra* depends on *Sxl* function (41, 44), suggesting that the regulation of *tra* expression by *Sxl* is mechanistically similar to *Sxl* autoregulation. Furthermore, a conserved sequence element has been identified at regulated splice acceptor sites in the *tra* and *Sxl* genes, mutation of which in the case of *tra* leads to the abrogation of *Sxl* control (55). The polypeptide encoded by a female *Sxl* cDNA possesses two copies of a 90-amino-acid ribonucleoprotein (RNP) motif, shared by a number of known RNA-binding proteins (6, 58). More recently, *Sxl* protein was shown to bind in vitro to the conserved RNA sequence element mentioned above (30). This RNA-binding capability, together with its genetic effects on sex-specific RNA processing, lead to the postulate that *Sxl* is a *trans*-acting splicing factor that represses the use of particular acceptor sites. This model is attractive, as it unifies *Sxl*'s autoregulatory and somatic differentiation activities through a common biochemical process, but its validity for *Sxl*'s regulation of dosage compensation awaits molecular analysis of the downstream genes involved in that process (8, 59).

Preliminary characterization of the *Sxl* transcript pattern revealed considerable complexity (51). What are the structural relationships among the various transcripts, and what are the consequences for protein-coding potential? An answer to this question is essential for understanding how *Sxl*

* Corresponding author.
† Present address: Department of Molecular and Cell Biology, Division of Genetics, University of California at Berkeley, Berkeley, CA 94720.

carries out its functions and specific mutations interfere with those functions and for interpreting information on sequence conservation for *Sxl* among various species.

The present report greatly refines our understanding of the structure of the late mRNA species which are present throughout most of development. The late transcript pattern is shown to be even more complex than previously thought. The 5′ and 3′ ends and precise exon structure of these mRNAs are described. There do not appear to be significant developmental stage-specific differences in RNA splicing, but there may be tissue-specific differences in polyadenylation. The transcripts encode a variety of structurally related proteins. One simplifying aspect of the results reported here is the discovery that despite the considerable complexity in the *Sxl* RNA splicing pattern, the only sex-specific splicing difference found so far is that involving the translation-terminating exon 3 reported previously.

MATERIALS AND METHODS

Fly stocks. Flies were raised at 25°C on a standard cornmeal dextrose-yeast medium. Sterile adult females were obtained as heterozygotes carrying the dominant female sterile allele ovo^{D1} (13). *Sxl* deficiencies *RJ2* and *7B0* are described by Salz et al. (50).

cDNA isolation and DNA sequence analysis. A 1.5- to 5-h embryonic cDNA library in λgt10 (23a, 33) was screened with nick-translated probes from *Sxl* regions h1 and j (see Fig. 1), and recombinant phage were purified essentially according to the method of Sambrook et al. (52). cDNA inserts were removed as *Eco*RI fragments and cloned into pBluescript KS (Stratagene). None of these cDNAs was full length; only one, EF5a21, is explicitly discussed in the text. Subsequently, a 12- to 24-h embryonic cDNA library in plasmid pNB40 (12) was screened with combined probes from *Sxl* regions h1 and f3. Because of uneven library amplification, multiple isolates of many recombinant cDNAs were obtained. Independent inserts were recloned as *Eco*RI fragments in pBluescript KS. For complete sequencing of MS3 and MS11, nested deletions were made with *Exo*III (25). Dideoxy sequencing was performed by the method of Sanger (52) with Sequenase (USB), and sequence assembly was carried out according to the method of Staden (56). Like the previously sequenced cDNA cF1, both MS3 and MS11 contained an additional G residue, not encoded in the genomic sequence, at the 5′ end. This extra G appears to be diagnostic of full-length cDNAs from the Brown library (11a) and for clarity is omitted from the sequences shown in Fig. 2 and 3. In comparing the sequences of the various *Sxl* cDNAs, cM1, as shown in Fig. 2 of Bell et al. (7), should have a G rather than the published A residue at the position equivalent to cF1 bp 768 (6a). Despite the occurrence of both male and female type transcripts in mixed late embryo populations, cDNAs obtained from the Brown library were all either female or unclassifiable because of 5′ truncations. The absence of male cDNAs may have resulted from uneven library amplification.

RNA analysis. Poly(A)$^+$ RNA was prepared essentially as described elsewhere (46), except that frozen embryos, larvae, or adults were disrupted in a Brinkmann homogenizer. All preadult RNA preparations are from mixed-sex populations.

For Northern (RNA) blot analysis, generally 2.5 to 10 μg of poly(A)$^+$ RNA was loaded per lane. RNA was fractionated on 1% agarose-formaldehyde gels according to the method of Sambrook et al. (52) and after partial alkaline hydrolysis was transferred either to nitrocellulose (Schleicher & Schuell) or to Zetabind nylon (AMF-Cuno). Nitrocellulose membranes were hybridized as described previously (7), with omission of dextran sulfate from hybridization mixtures. After hybridization, membranes were washed at 65°C with 0.2× SSPE (1× SSPE is 0.18 M NaCl, 10 mM $NaPO_4$, and 1 mM EDTA [pH 7.4])–0.1% sodium dodecyl sulfate (SDS). RNA was UV cross-linked to nylon membranes and hybridized by a modification of the procedure of Amasino (1). Hybridization was in 0.125 M $NaHPO_4$ (pH 7.0)–0.25 M NaCl–7% SDS–1 mM EDTA–10% polyethylene glycol 8000 at 65°C. Washes were as described by Amasino (1) or as described above for nitrocellulose. Hybridization probes from gel-purified *Sxl* genomic fragments were prepared by nick translation or by random priming with *Escherichia coli* Klenow fragment (Boehringer) according to the manufacturer's recommendation. For normalization of mRNA concentrations in the different poly(A)$^+$ RNA fractions, the blot shown in Fig. 4A was reprobed with a fragment containing the ribosomal protein 49 gene (45). Sizes of the various *Sxl* transcripts are given on the basis of the sequences of the representative cDNAs and on RNase protections mapping the 5′ and 3′ ends.

RNase protections were performed according to the method of Melton et al. (42) with the following modifications. High-specific-activity uniformly labeled antisense riboprobes were synthesized with SP6, T3, or T7 RNA polymerase as suggested by the manufacturer (Promega). Riboprobes were purified by spin column chromatography; 1 to 2% eluate was used directly in hybridizations. Nonradioactive control sense strand RNAs from *Sxl* cDNAs or genomic subclones were synthesized as suggested by the manufacturer (Promega). Following hybridization of radioactive probes with synthetic control RNA or poly(A)$^+$ RNA, reactions were quenched with RNases A and T1 at dilutions optimized for each different probe (see figure legends). RNase mix (1×) was 40 μg of RNase A per ml and 2 μg of RNase T1 per ml. RNase treatment was generally for 60 min at room temperature; in the case of probe k1a, a shorter, 10-min incubation was necessary to prevent overdigestion. RNases were inactivated with protease treatment, and samples were extracted, ethanol precipitated, resuspended, and electrophoresed on denaturing urea–Tris-borate-EDTA–polyacrylamide gels. Molecular weight markers consisted of end-labeled *Hae*III-cut φX174 or *Msp*I-cut pBR322.

Primer extensions employed a polynucleotide kinase-labeled oligonucleotide with the sequence CACTGCGCGACT TTCCTTCTTC, complementary to the downstream end of exon 1, and were performed as described elsewhere (22) with 5 μg of poly(A)$^+$ RNA and avian myeloblastosis virus reverse transcriptase (USB). Extension products were resolved by electrophoresis on denaturing urea–Tris-borate-EDTA–8% polyacrylamide gels. To control for reverse transcriptase artifacts, extensions were also performed on synthetic plus-strand RNA from genomic subclone d2 (−223 to +242); with this synthetic RNA, no strong stops were observed in the region of interest (see Fig. 5A).

For autoradiography, X-ray films were generally preflashed (except for the autoradiogram shown in Fig. 5A) and exposed with an intensifying screen.

In vitro translations. Translatable mRNAs were synthesized from *Sxl* cDNAs with SP6 RNA polymerase as described previously (32), including the cap analog GpppG in the transcription reactions. Capped RNAs were translated in treated reticulocyte lysate (Promega) as recommended by the manufacturer. Typically, 8- to 10-μl reactions were

performed with 5 μCi of [^{35}S]methionine and empirically determined optimal concentrations of synthetic mRNA. Where noted, partially purified yeast suppressor tRNAs (a kind gift of M. Bienz, J. Smith, and J. Kohli) were included in translations. Products were resolved on SDS–10% polyacrylamide gels, and gels were dried and autoradiographed. Markers were in vitro-translated brome mosaic virus proteins (Promega) or ^{14}C-methylated proteins (Amersham).

Nucleotide sequence accession numbers. The nucleotide sequences described here have been entered in the GenBank data base and assigned accession numbers M59447 (pMS3) and M59448 (pMS11).

RESULTS

Isolation of *Sxl* cDNA clones. The initial analysis of *Sxl* gene expression showed that the locus encodes at least 10 distinct transcripts whose patterns of expression vary according to the sex and developmental stage of the animals (51). In adult females four transcripts were detected, falling into three size classes: a large RNA of approximately 4 kb, two middle RNAs of about 3 kb, and a small RNA of 2 kb. Expression of the two middle RNAs differed in that one was absent from sterile female flies while the other was retained; these RNAs were therefore termed the mid-gld (germ line-dependent [missing in sterile females]) and the mid-gli (germ line-independent) transcripts. In adult males three RNAs were detected: large, middle, and small. Each of these RNAs was approximately 200 nucleotides (nt) larger than the equivalent female RNA, because of the presence of an extra exon (7). It remained unclear whether the inclusion of this male-specific exon represented the only *Sxl* RNA processing difference between the two sexes.

The adult-type transcripts were shown to occur throughout much of development. They were thus termed the late transcripts to distinguish them from another set of RNAs observed uniquely during early embryogenesis (2 to 4 h). The early transcripts also fell into three size classes: large, middle, and small, each approximately 200 nt smaller than the analogous adult female RNA.

The numerous *Sxl* transcripts overlap extensively. Although some differences in sequence content among these transcripts were detected by probing Northern blots with particular genomic DNA segments (51), this level of analysis was insufficiently detailed to distinguish exact exonic structure. To examine the differences among the transcripts at higher resolution, we have resorted to an extensive cDNA analysis. A 12- to 24-h embryo library (12), expected to include both male and female late transcripts, was screened and provided 10 cDNAs longer than 1.5 kb. Two of the largest clones, MS3 and MS11, were fully sequenced, and their exon structures were determined by comparison with cDNAs cM1 and cF1 and with *Sxl* genomic sequences (14a). These sequenced cDNAs are presented schematically in Fig. 1, together with the deduced structures of the eight other Brown library cDNAs and one earlier Goldschmidt-Clermont library cDNA. A genomic restriction map of the *Sxl* locus is provided at the top of the figure, and various genomic subclones used for Northern blot and RNase protection analysis are indicated.

Structure of clone MS3 and related cDNAs. MS3 is 3,094 bp in length (see Fig. 2). Its 5′ end is exactly the same as that of the previously sequenced female-specific cDNA cF1. At its 3′ end is a short poly(A) tail, preceded 35 bp upstream by a canonical AATAAA polyadenylation signal. MS3 contains seven exons, corresponding to numbers 1, 2, 4, 5, 6, 7, and 8 of Bell et al. (7) (Fig. 1). Exon number 3, the male-specific exon, is not present in MS3, which is therefore a female-specific cDNA. On the basis of its size and sequence content, MS3 appears to represent a full-length version of the mid-gli transcript (see Fig. 9). The cDNA contains a single long ORF, potentially encoding a 354-amino-acid polypeptide, assuming that translation initiates at the first available methionine of the long ORF (bp 426). The predicted sequence of this polypeptide is identical to that of cF1, including two RNP motifs conserved in a variety of RNA-binding proteins (encoded in MS3 exons 5 to 8).

Clones MS21, MS24, MS29, MS31, and MS59, together with cF1 and cM1, resemble truncated versions of MS3 (Fig. 1). All are shorter than MS3, with 3′ ends occurring at the positions indicated in Fig. 2; many of these 3′ ends occur at long runs of A residues in the genomic sequence and may have resulted from internal priming by oligo(dT). Most of these cDNAs are also truncated at the 5′ end, at the approximate positions shown in Fig. 1. MS13 and MS50, also truncated at the 3′ end, are peculiar in that they contain unspliced intron sequences. These two clones probably represent incompletely processed intermediates.

Structures of very large clones MS11 and MS16. MS11 and MS16 contain sequences from regions j and k and hence are clearly distinct from the cDNAs described above (Fig. 1). These regions are specific in females to the 4-kb large and 3-kb mid-gld transcripts (51). However, both in total size (5 kb) and in the presence of k3 sequences (see below), the cDNAs differ from these transcripts. They appear instead to be very large, similar to the large transcript but containing about 1 kb of additional sequence at the 3′ end.

MS11 is 5,163 bp long (see Fig. 3), with a 5′ end identical to that of MS3. Like MS3 it contains exons 1, 2, 4, 5, 6, and 7 and lacks male exon 3. Therefore, MS11 has the potential to encode a polypeptide containing the two RNP motifs. However, MS11 differs critically from the MS3-type cDNAs in exon 8. A splice at bp 1395 of MS11 truncates exon 8, eliminating 23 amino acids present in MS3 and generating thereby a 43-amino-acid alternative C terminus encoded in exons 9 and 10. Another alternative splice, into exon 5 of MS11, eliminates the first 8 amino acids of coding potential present in exon 5 of MS3. These two forms of exon 5 were previously noted in cF1 and cM1 (7).

In addition to the ORF containing the two RNP consensus motifs, a second long ORF of 355 amino acids occurs in MS11 following the UAA translation terminator at bp 1524 (Fig. 3). This second long ORF is in the same frame as the first ORF and is henceforth designated ORF2. The first ORF is therefore called ORF1 or the RNP-ORF. ORF2 is not obviously related to ORF1, although it does in general match the *Drosophila* codon bias. ORF2 is rich in glutamine, serine, and asparagine and shows similarity to many *opa* repeat-containing genes (62) in the GenBank and EMBL data bases.

Besides the different C terminus, MS11 has two internal amino acid differences from MS3 at the positions noted in Fig. 2 and 3. Oddly, male cDNA cM1 resembles MS3 at the first site and MS11 at the second. Clone cF1 is identical to MS3 at both positions.

Clone MS16 differs slightly from MS11 in that it has the same 5′ and 3′ ends but skips exon 9, splicing exon 8 directly into exon 10 with the same junctions as MS11 and preserving the translational reading frame. Thus, MS16 should lack the 22 amino acids encoded by exon 9 but have the same C terminus as MS11, as well as ORF2 in exon 10. In addition to MS11 and MS16, six early embryonic cDNAs and one larval

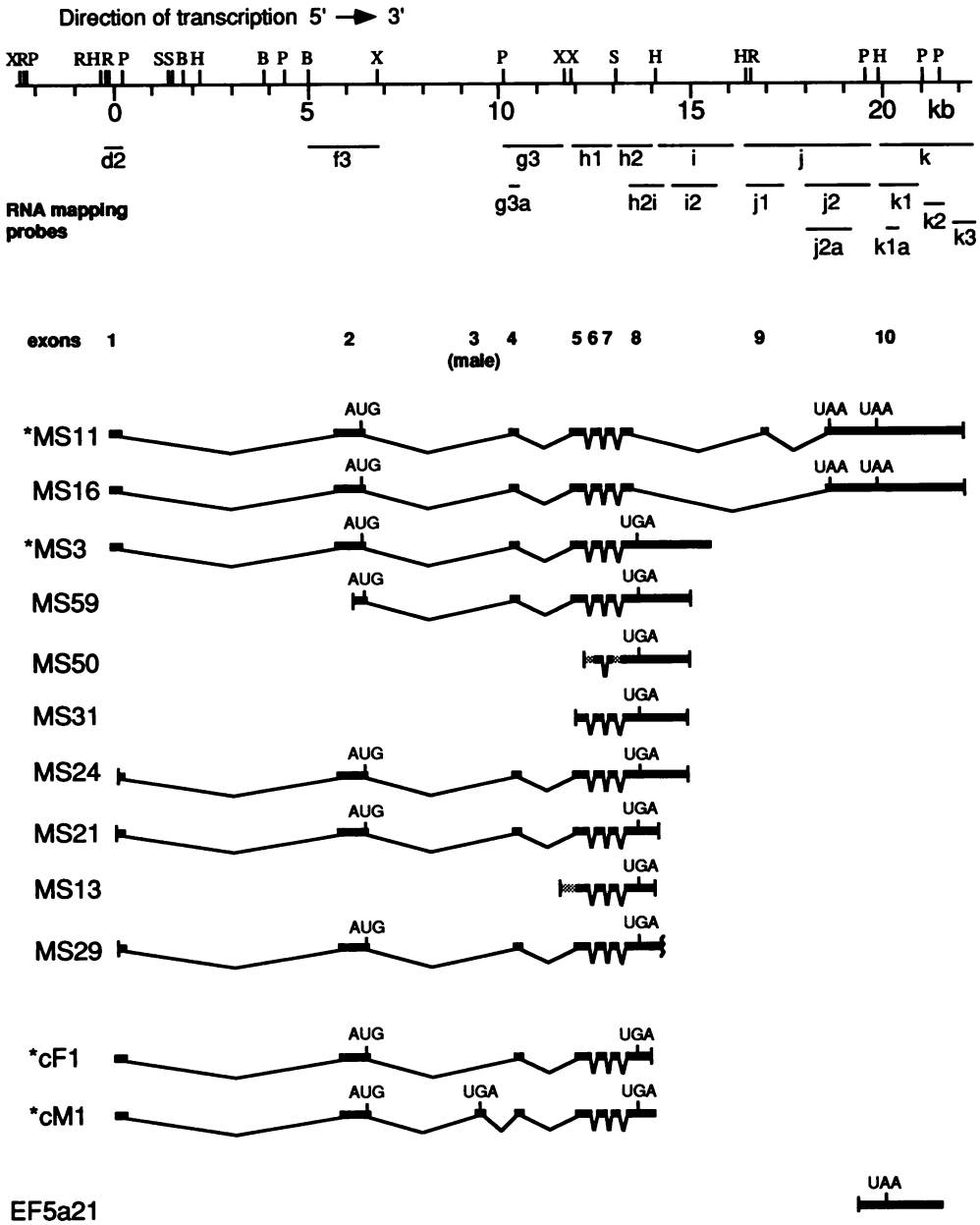


FIG. 1. Exon structures of *Sxl* cDNAs. A 12- to 24-h embryonic cDNA library was screened for *Sxl* clones. The 10 longest inserts (MS3 to MS59) were mapped by restriction analysis, Southern blotting with genomic *Sxl* probes, 5'- and 3'-end sequencing, and RNase protection. The structures of previously sequenced adult cDNAs cF1 and cM1 are provided for comparison. Partial cDNA EF5a21 was isolated from the Goldschmidt-Clermont library as described in the text. Asterisks denote cDNAs which have been completely sequenced. Ends of cDNAs which are believed to represent truncations of *Sxl* transcripts (either 5' or 3') are indicated with vertical lines. Putative translational initiation and termination codons are indicated; these are inferred for partially sequenced cDNAs on the basis of exon mapping. Exon-intron junctions were determined by comparison with genomic *Sxl* sequences (for reasons of clarity, not all exons are shown to scale). The presence of microexon 4 was not rigorously demonstrated for partially sequenced cDNAs. Exon 3, the male-specific exon, was apparently absent from all the late embryonic cDNAs. Retained intervening sequences in MS13 and MS50 are shown by stippled boxes. *Sxl* genomic subclones used for hybridization probes are shown at the top of the figure. The *Sxl* nucleotide numbering system has been revised so that map position 0, previously corresponding to the site of insertion of a P element, now corresponds to the 5' end of the late *Sxl* transcription unit. In addition, the numbering system now increases positively in the direction of transcription (51). Restriction sites are designated as follows: B, *Bam*HI; H, *Hin*dIII; P, *Pst*I; R, *Eco*RI; S, *Sal*I; X, *Xho*I.

cDNA contain the same splice out of exon 8 (data not shown). Of all nine cDNAs in this class, four contain the exon structure 8-9-10 while the remaining five have the structure 8-10.

Developmental RNA analysis. To extend the previously published analyses of *Sxl* transcription, Northern blots were performed with a series of probes (Fig. 1) specific to different *Sxl* transcripts. In this way the expression patterns of all

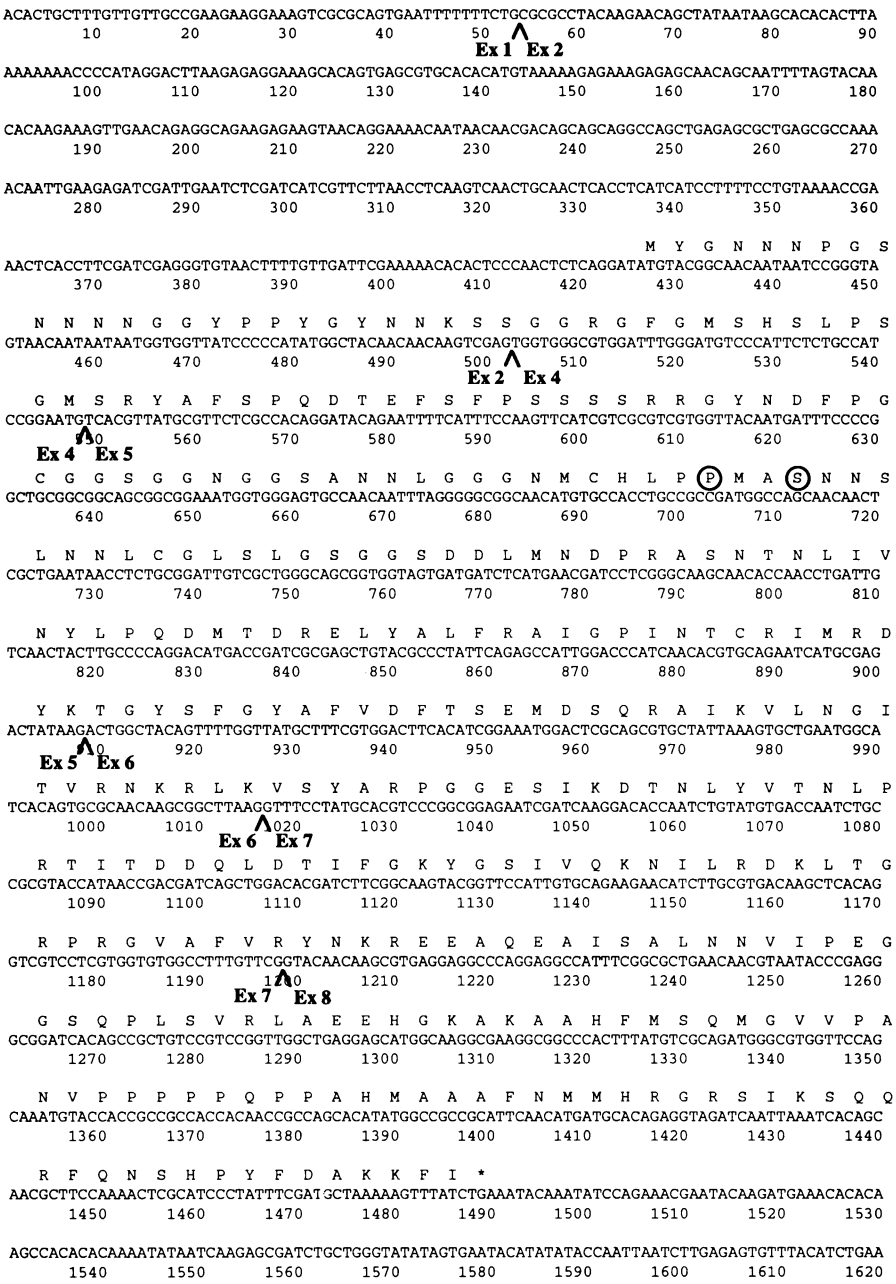


FIG. 2. DNA sequence of cDNA MS3. The junctions of exons 1 to 8 (Ex 1 to Ex 8) are indicated by arrowheads. Positions of 3′ termini of related cDNAs (see text) are marked; potential canonical and noncanonical polyadenylation signals are underlined (see Discussion). The 3′ end of MS29 was not determined precisely but is probably similar to that of MS21. The amino acid sequence of the presumptive MS3 translation product is shown, with polymorphic amino acids circled. The two RNP motifs as defined by Bell et al. (7) are encoded between bp 756 and 1019 and bp 1020 and 1274. The T at bp 731 of MS3 is a C in other cDNAs; this third position change does not introduce a difference in coding potential.

seven late transcripts were determined from embryogenesis through adulthood (Fig. 4).

Fig. 4A shows the developmental profile of all *Sxl* transcripts, visualized with general probe h1. The 10 previously identified *Sxl* transcripts are indicated by arrows on either side of the figure: on the right, arrows are placed at the positions of the late transcripts; the left arrows point out the early transcripts. The salient features observed by Salz et al. (51) were confirmed, notably, the occurrence of the early class of transcripts between 2 and 4 h of embryonic development and the subsequent appearance of the late male and female transcripts during midembryogenesis. Note that the 3.3-kb male RNA is initially less abundant than its 3.1-kb female counterpart (5 to 12 h) but that the two transcripts eventually achieve similar levels. A large RNA, not previously defined, of unknown sex specificity is transiently visible in late embryos (arrow marked "very large"). Although the overall level of *Sxl* transcription is considerably

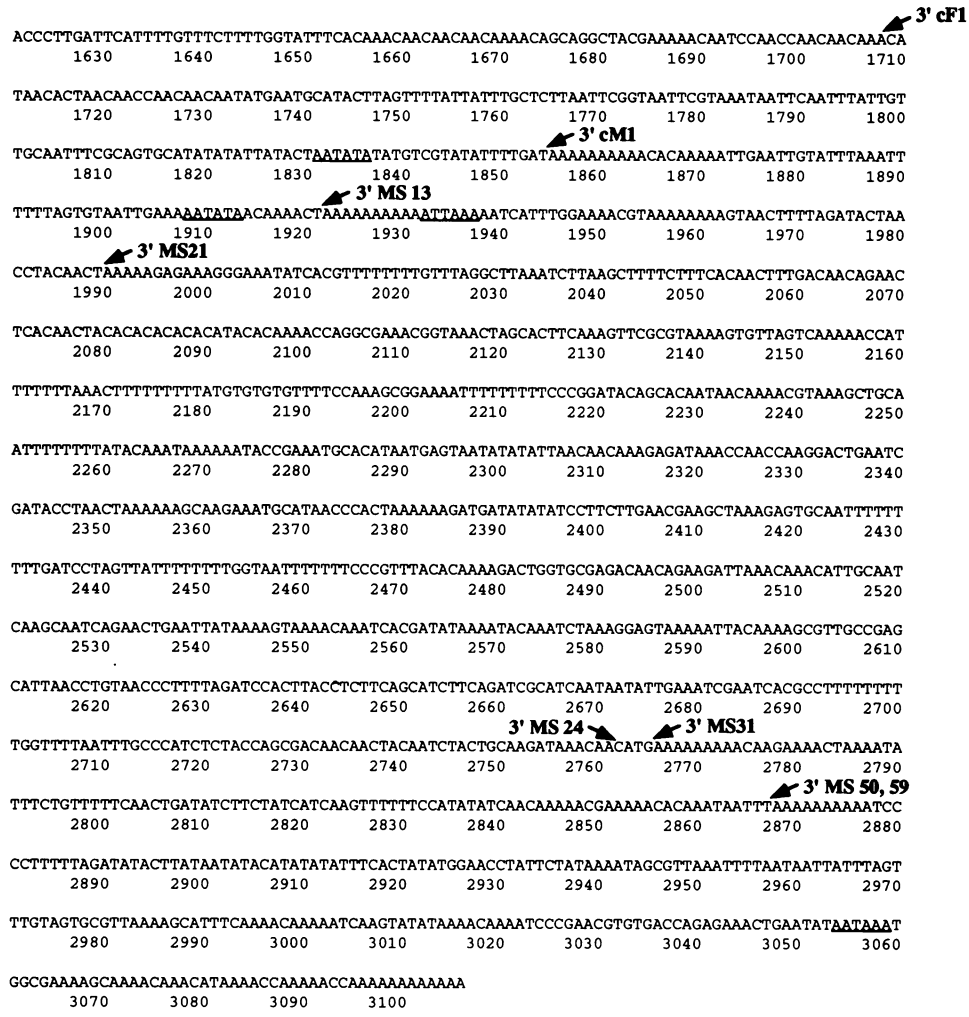


FIG. 2—*Continued.*

reduced during larval and pupal stages, there is no significant qualitative change in the late transcript pattern until adulthood. At this time the large (4.3-kb) male transcript, not detectable in preadult stages in this experiment, is observed. The small (1.9-kb) transcript reappears in females, and comparison of fertile and sterile females shows the expected germ line dependence of the 3.1- and 1.9-kb RNAs.

The relative contributions of the two different middle transcripts were determined with probes specific for these RNAs. As shown in Fig. 4B, mid-gli-type transcripts visualized with probe i appear throughout development. As noted above, the 3.3-kb male transcript appears less abundant than the 3.1-kb mid-gli female transcript in 5- to 12-h embryos, but at all later stages these transcripts are expressed at approximately equal levels. In contrast to these results, middle-size transcripts of the gld type, visualized with probe k1, occur uniquely in fertile adult females (Fig. 4C, arrow on right marked "3.1 fem.-gld"). The blots in Figs. 4B and C together indicate that the middle transcripts observed with general probe h1 result exclusively from gli-type RNA during preadult development and in adult males.

Probe k1 also anneals to the large RNAs at all stages of development, as well as to the very large RNA in late embryos (Fig. 4C). Additional Northern blots demonstrate that the very large RNA, but not the large RNAs, contains sequences as far distal as region k3 (data not shown).

5′-end analysis. As noted above, several *Sxl* cDNAs begin at the same site, the A residue numbered +1 in MS3 and MS11, while none has been identified carrying sequences from further upstream. That this +1 position is the 5′ end of the late *Sxl* transcription unit was verified by primer extension and RNase protection.

Figure 5A shows the results of primer extensions. In late embryo, adult male, and adult female RNAs a prominent extension product was observed, corresponding to the +1 A of the cDNAs (41 nt, asterisks on left). Some larger extension products were also detected in these lanes. However, since the larger products were seen with RNA from male flies deleted for the *Sxl* transcription unit (Sxl^{fP7BO} and Sxl^{fPRJ2}), they probably resulted from weak cross hybridization of the primer to non-*Sxl* mRNAs (data not shown).

RNase protection experiments confirmed the 5′-end mapping. Synthetic MS11 control RNA yielded a protected product consistent with the expected size of 53 nt (Fig. 5B, arrow marked "exon 1"). Protection products identical to the control were observed with late embryonic and adult female RNAs. Male RNA yielded similar results, and as

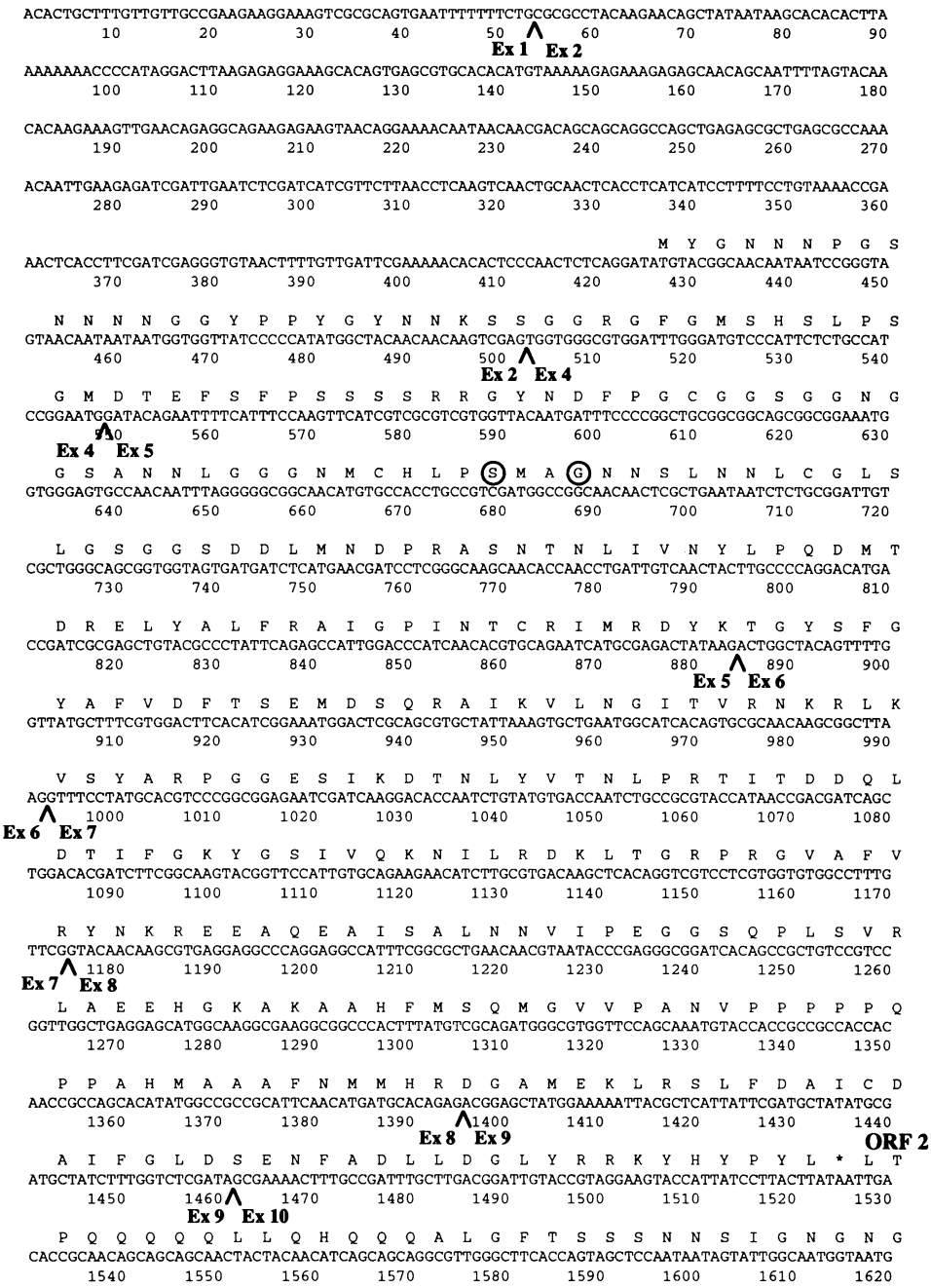


FIG. 3. DNA sequence of cDNA MS11. The junctions of exons 1 to 10 (Ex 1 to Ex 10) are indicated by arrowheads. The position of the 3′ terminus of related cDNA EF5a21 is marked; potential canonical and noncanonical polyadenylation signals are underlined (see Discussion). The amino acid sequence of the presumptive translation product encoded by ORF1 is shown, with polymorphic amino acids circled. The two RNP motifs are encoded between bp 735 and 995 and bp 996 and 1253. The hypothetical translation product of ORF2, encoded from bp 1527 to 2591, is also shown. Because mechanisms for utilization of ORF2 remain speculative (see Discussion), translation is shown from the beginning of the ORF rather than from the first available methionine.

expected the 53-nt product was absent in RNA from *Sxl*-deficient males (data not shown).

In the cDNA analyses several different versions of the exon1-exon 2 splice were observed (cf. cF1 and cM1), with one of three different exon 1 donor sites joined to one of two exon 2 acceptor sites. However, no evidence of the other exon 1-spliced forms was found in RNase protections such as that shown in Fig. 5B (data not shown), although all three donors show a 7 of 9 match to the consensus sequence (43).

In addition to the anticipated 53-nt band, an unexpected band of about 250 nt was observed in late embryonic and adult female RNAs (Fig. 5B, arrow marked "unspliced RNA"). This band corresponds in size to the product expected from an *Sxl* precursor transcript. On the basis of

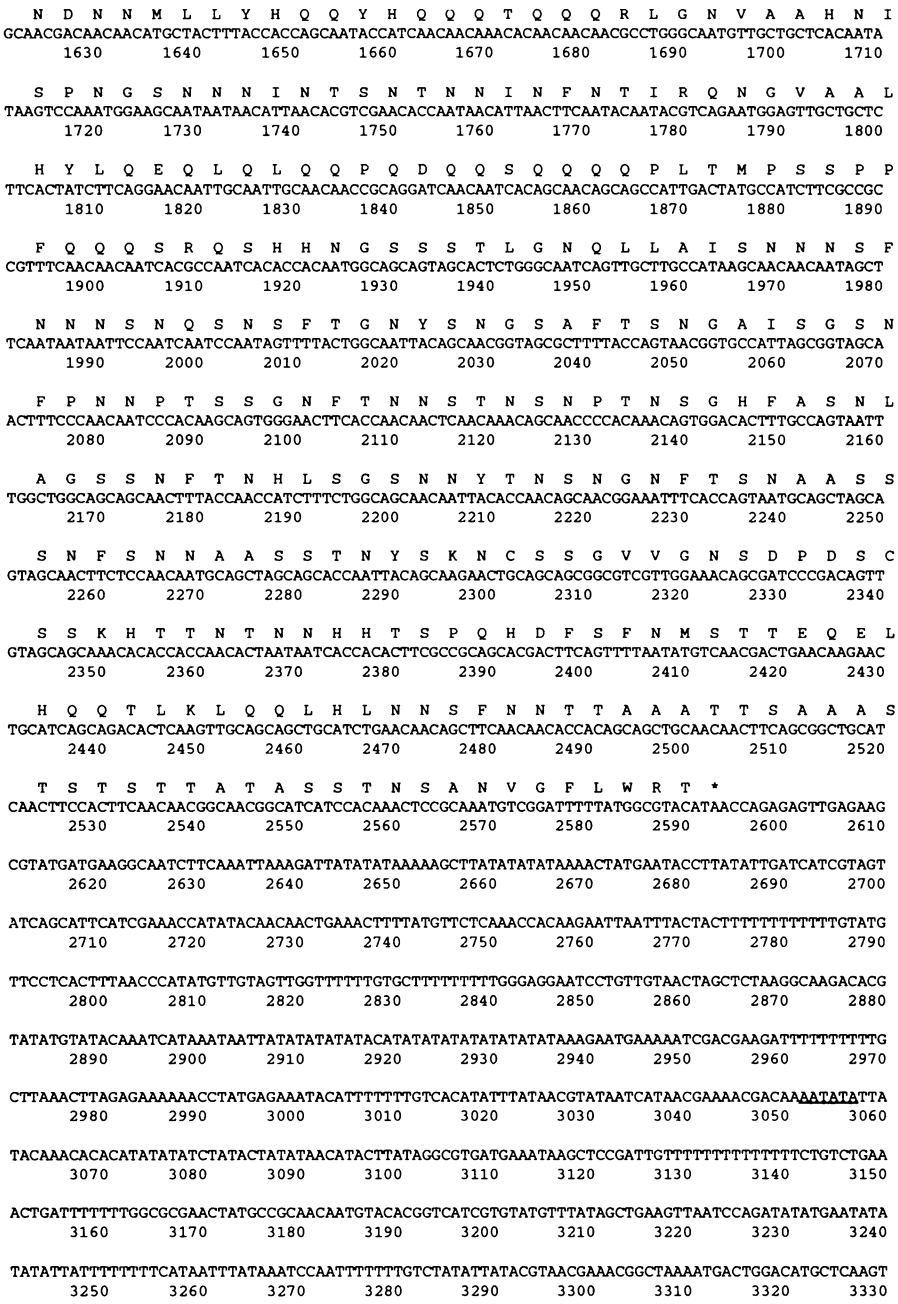


FIG. 3—*Continued.*

this interpretation, the results imply that a significant fraction of the total protected RNA from this part of the gene resides in the unspliced state. In addition to the 250-nt band, small amounts of another unexpected product were observed (arrow marked "readthrough RNA"). This product was about 500 nt, approximately as expected for full protection of the probe, and must result from RNA beginning further upstream and reading through region d2. Such RNA could represent a minor alternative 5′ end of *Sxl*, or else it could represent run-on from the next transcription unit upstream. The upstream transcription unit lies close to *Sxl* (less than 1 kb upstream of position +1) and is transcribed in the same orientation (49a).

3′-end analysis. The 3′ ends of the various *Sxl* transcripts were mapped by Northern blotting and RNase protection. Together these experiments identified four different 3′ ends, summarized schematically in Fig. 9. Two of these ends occur throughout development, while the other two are stage or tissue specific.

The large RNA 3′ ends were mapped to a site corresponding to bp 4116 of MS11 (Fig. 3). This was based on the size of the major RNase protection product of probe k2 (Fig. 6A) and the recovery of partial cDNA EF5a21 whose 3′ end was at this position. Some fully protected k2 probe can also seen in Fig. 6A (arrow marked "readthrough RNA"), probably because of the very large RNA noted above.

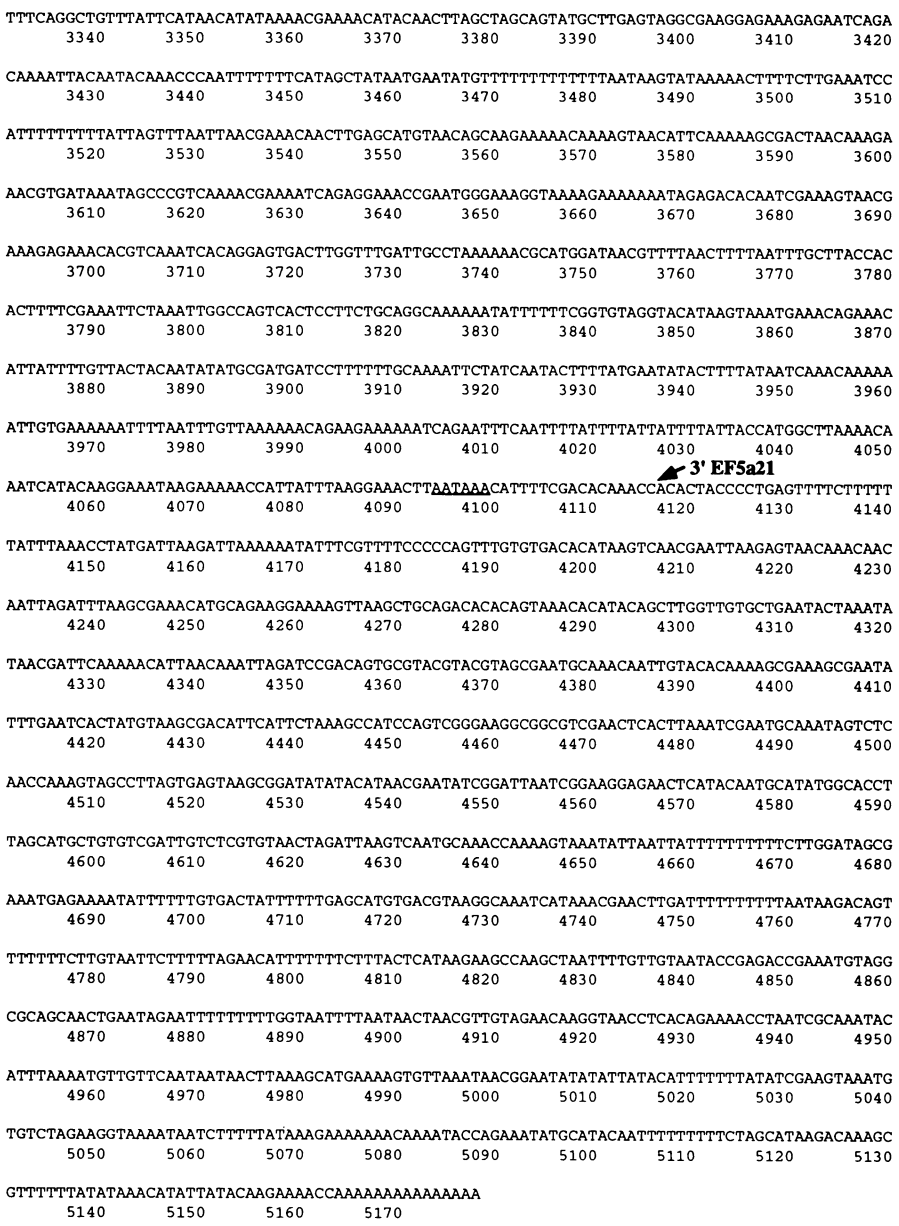


FIG. 3—*Continued.*

RNase protections with probe i2 confirmed that the 3′ end of MS3 coincides with the bona fide end of the 3.1-kb mid-gli female, 3.3-kb male, and 2.9-kb early embryonic transcripts (data not shown). Similarly, experiments with probe k1a placed the 3′ end of the 3.1-kb mid-gld female transcript at a position corresponding to bp 3100 of cDNA MS11 (data not shown). No canonical AATAAA occurs near the indicated mid-gld 3′ end.

To map the 3′ end(s) of the 1.9-kb small female transcript, RNase protections were performed with probe h2i (Fig. 6B). The ends could not be directly identified in protections with female RNA, because of the presence of additional bands resulting from mid-gli transcripts (MS3 control lane). Because the 1.9-kb RNA shows significant germ line dependence in adult females, protection products of this transcript were expected to show decreased abundance in RNA from sterile females relative to that of fertile ones. Two bands having the predicted specificity can be seen in the figure. Small 3′ end 1 lies between the ends of MS13 and MS21 (Fig. 2), while small 3′ end 2 lies slightly upstream of the end of MS13, at a site possibly corresponding to the end of cM1. As with the mid-gld 3′ end, no AATAAA sequences are observed in this region. In accordance with the Northern blot data, bands corresponding to the small female 3′ ends were present to some extent in male and embryo RNAs.

Confirmation of the complete exon structure of *Sxl* transcripts. The structures of the internal exons of *Sxl* transcripts were all confirmed by RNase protection. Besides the occurrence of the male-specific exon, several other alternative splices were noted upon comparison of the cDNAs, including the use of three different exon 1 splice donors as described above (cf. MS3, MS24, and cF1), two exon 2

acceptors (cf. MS3 and cM1), two exon 5 acceptors (cf. MS3 and MS11), the alternative use of exon 10 (cf. MS3 and MS16), and the optional presence of exon 9 (cf. MS11 and MS16). The possible sex, tissue, and stage specificities of these heterogeneities were examined.

Protections with a probe from region f3 confirmed the structure of exon 2. As seen in Fig. 7A, protection products corresponding to both splice acceptors were detected in roughly equal proportions in all natural RNA samples. In this experiment the exon 2 protection products from 2- to 4-h embryos must have derived from late-type transcripts present in that RNA preparation. The early transcripts in 2- to 4-h embryos lack both exons 1 and 2, substituting a different 5′ exon (30a).

The structure of exon 4 was confirmed by RNase protection with probe g3a, although the determination revealed a significant artifact of the technique. As seen in Fig. 7B, the 46-nt exon band defined by the MS11 control was present in early and late embryos and adult females. In adult male RNA a slightly smaller band, indicated in the figure with an open dot, was observed instead. A reconstruction experiment with synthetic plus-strand RNA derived from a male cDNA showed that this smaller band was the normal product of male RNA containing the 46-nt exon 4 (data not shown). The size difference of the products protected by female versus male RNA probably resulted from the different exons immediately upstream of exon 4 in the two sexes.

Besides the products expected for fully processed exon 4, two additional unexpected bands were observed. The first was about 150 nt (arrow marked "unspliced RNA"), a size consistent with protection by unspliced RNA containing intron sequences on both sides of exon 4. The second band was about 130 nt (arrow marked "partially spliced RNA"). This size is consistent with protection of exon 4 plus upstream intron sequences and is the expected result for an RNA having made the exon 4-exon 5 downstream splice but no upstream splice into exon 4. The reciprocal partial-splice product, that having made the upstream splice to exon 4 but not the downstream splice, was not observed.

Exons 5, 6, and 7 are all encoded in region h1. RNase protections with h1 antisense riboprobe demonstrated the presence of all three exons in poly(A)$^+$ RNA at all stages (Fig. 7C). The use of two acceptor sites for exon 5, situated 24 bp apart, was noted at all developmental stages examined and was confirmed not to be sex specific.

The different versions of exon 8, typified by cDNAs MS3 and MS11, were verified by RNase protections with probes h2, h2i, and i2 (data not shown). Protections with probe j1 verified the presence of the small 66-nt exon 9 in embryos and adults (data not shown). The relative proportion of RNAs including exon 9 in vivo could not be determined from the protection data. No sex or stage specificity was noted in the use of exons 8 and 9.

The structure of exon 10 was confirmed by RNase protections with probes j2 and j2a (data not shown) together with probes k1 and k2 (see above) and was shown to be invariant in the two sexes. These experiments confirmed that the only difference between cDNA MS11 and the bona fide large female transcript is the greater length of the 3′ nontranslated region in the cDNA. It was particularly noteworthy that no additional splice acceptor sites for exon 10 were detected at any developmental stage. Such sites, if they occurred downstream of the first UAA stop codon (at bp 1524 relative to the MS11 sequence), could allow the generation of a longer *Sxl* alternate protein fusing the RNP-ORF to some part of ORF2.

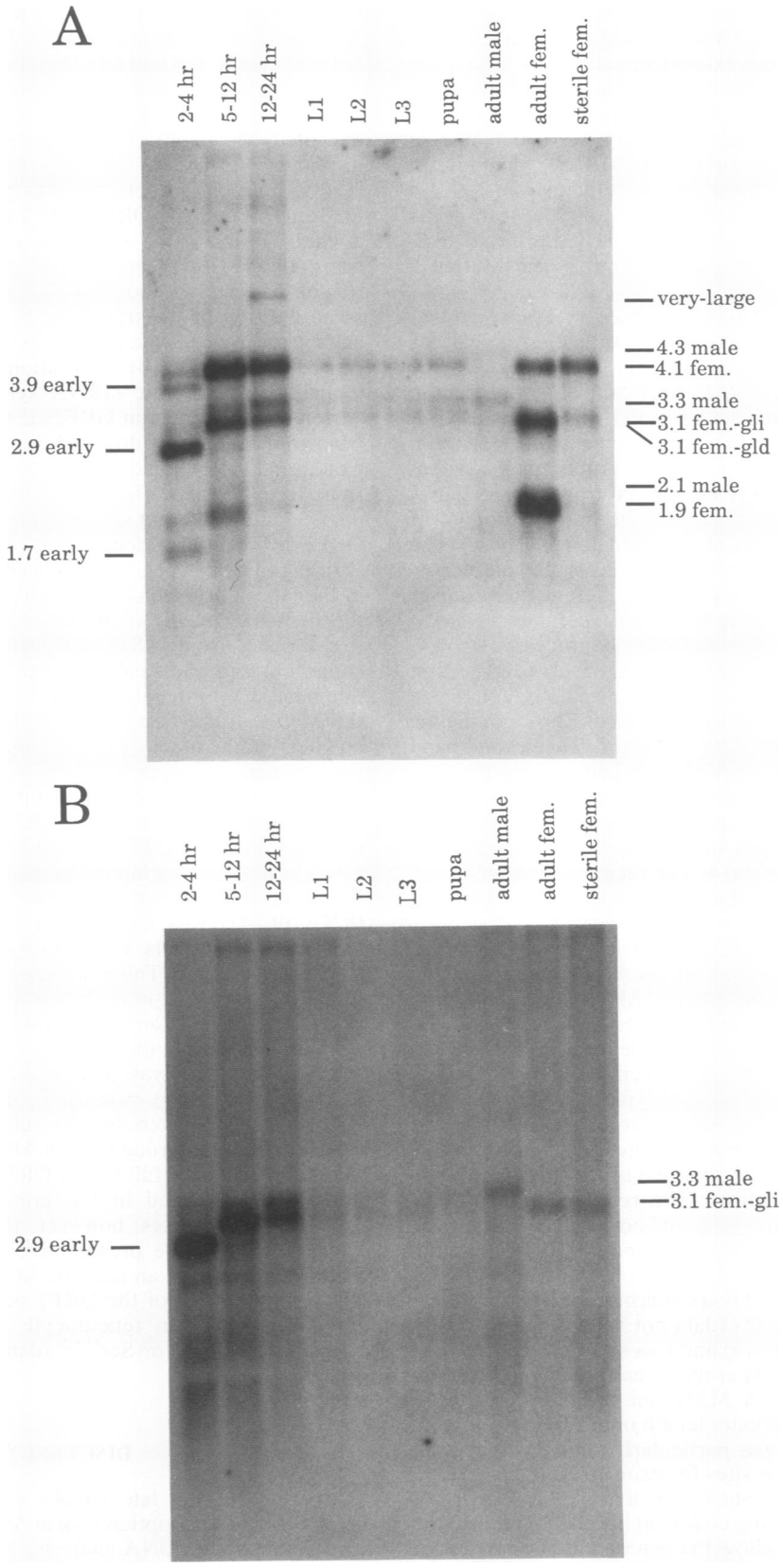
In vitro translation of cDNAs and UAA suppression. The predicted coding potential of cDNAs was tested by programming reticulocyte lysates with capped mRNA synthesized in vitro. In each case the cDNA-derived mRNA yielded a prominent translation product of the expected size, 38 kDa for MS3 and MS16 and 40 kDa for MS11 (Fig. 8, left panel). Surprisingly, in these in vitro translation experiments a band of about 70 kDa was frequently observed in addition to the expected 38- to 40-kDa products. This was particularly intriguing since the 70-kDa size was consistent with readthrough of the ochre termination codon in MS11 and MS16 to produce an ORF1-ORF2 fusion protein. However, a band of the same size was occasionally observed in translations of MS3, in which no possible readthrough protein of that size was possible. Therefore, we suspected that some fraction of the translated *Sxl* proteins formed stable aggregates under these electrophoresis conditions. This hypothesis was tested by translations of MS11 in which the template for synthetic mRNA synthesis was truncated at various internal sites in ORF2. If the 70-kDa band derived from translational readthrough into ORF2, then such truncations in ORF2 should result in a decrease in the size of this translation product. However, ORF1 aggregates should be unaffected by different sites of truncation downstream of the ORF1 termination codon.

In Fig. 8, left panel, the 70-kDa band is clearly detectable in the lanes corresponding to MS16 and MS11 linearized at the 3′ ends of the cDNA inserts. The slightly smaller size of this product from MS16 compared with that of MS11 is consistent with the smaller size of the monomeric polypeptide band due to the absence of exon 9. As seen in the next three lanes, truncation of MS11 at bp 2156, 2245, or 2053 had no effect on the mobility of the 70-kDa band, although ochre readthrough products of 63, 66, or 59 kDa, respectively, would be expected. These results clearly imply that this 70-kDa band is not an ORF1-ORF2 fusion polypeptide but contains only the ORF1-encoded polypeptide in an aggregated form (possibly homodimeric on the basis of its size).

In contrast to these data, when a yeast Tyr^{UAA} suppressor tRNA (9, 31) was included in the in vitro translations, readthrough products were observed for MS16 and MS11 (Fig. 8, right panel). This was especially evident for the three truncations of MS11 in ORF2, in which translation products migrating as 60, 62, and 58 kDa were observed, close to the sizes predicted from the amino acid sequence. In these lanes the 70-kDa band was also present, suggesting that the partitioning between monomeric and aggregated forms of ORF1 was unaffected by the relatively small amount of translational readthrough. With MS16 and MS11 truncated at their 3′ ends, the full-length ORF1-ORF2 fusion polypeptides were produced in the presence of the suppressor tRNA. In these cases, however, the readthrough products comigrated with the presumptive monomer aggregate 70-kDa bands, causing an increase in intensity of these bands. Thus, readthrough of the ORF1 ochre codon in MS11 and MS16 can occur in reticulocyte lysates, but only when exogenous Tyr^{UAA} or Ser^{UAA} (data not shown) tRNAs are added.

DISCUSSION

Structure of the late female *Sxl* transcripts. We have analyzed the transcription map of *Sxl* at high resolution using a combination of cDNA analysis, Northern blotting, primer extension, and RNase protection. In this way, the exon structures of the previously described late *Sxl* transcripts


A
2-4 hr
5-12 hr
12-24 hr
L1
L2
L3
pupa
adult male
adult fem.
sterile fem.
very-large
4.3 male
4.1 fem.
3.3 male
3.1 fem.-gli
3.1 fem.-gld
2.1 male
1.9 fem.
3.9 early
2.9 early
1.7 early
B
2-4 hr
5-12 hr
12-24 hr
L1
L2
L3
pupa
adult male
adult fem.
sterile fem.
3.3 male
3.1 fem.-gli
2.9 early

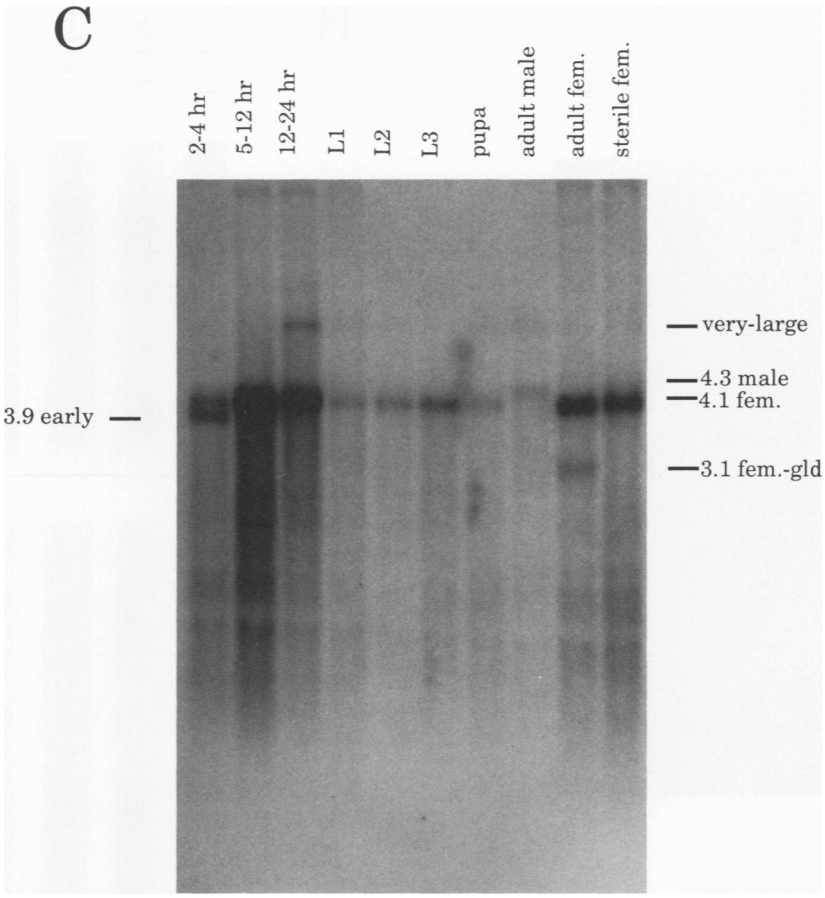


FIG. 4. Developmental Northern blot analysis of *Sxl*. Poly(A)$^+$ RNAs from the indicated stages were used at 2.5 μg per lane (A) or 10 μg per lane (B and C). Hybridization probes were derived from the *Sxl* genomic subclones shown in Fig. 1: probe h1 (A), probe i (B), and probe k1 (C). The *Sxl* transcripts visualized in panels B and C were aligned with the various size classes defined in panel A by hybridizing strips from a single gel loaded with adult female RNA with each of the probes (data not shown). Sizes are marked in kilobases.

(51), four female and three male, have been deduced. These structures are presented schematically in Fig. 9.

The four female-specific transcripts fall into two structural classes, depending on the choice of 3′ exon. The first class includes the 4.1-kb large and the 3.1-kb female gld transcripts. These two RNAs terminate with exon 10 and are expected to have the same coding potential, typified by cDNA MS11. Exon 9 is present in some fraction of these transcripts. Comparison of RNAs from fertile and sterile females strongly suggests that the large and mid-gld female transcripts differ only in their 3′ ends, which lie about 1 kb apart. This is sufficient to account for the relative sizes of these RNAs as seen on Northern blots. There are, however, some caveats to this conclusion. Although germ line-dependent internal exons were not observed in the RNase protections of Fig. 7, the entire *Sxl* gene has not been systematically scanned for such exons. Moreover, no cDNAs representing the mid-gld transcript have been isolated. Thus, the mid-gld structure shown in Fig. 9 must be considered tentative.

The second female transcript class is composed of the 3.1-kb female mid-gli and the 1.9-kb small RNAs, which differ from RNAs of the first group by terminating in exon 8. cDNA MS3 typifies the second group in coding potential. As in the case of the large versus mid-gld transcripts, the 1-kb size difference between the mid-gli and small RNAs appears to result from a truncation of the 3′ exon. Conceivably, the choice of exon 8 or exon 10 as the 3′ terminal exon could be made mechanistically at the level of either splicing or polyadenylation.

The two transcripts whose high levels in adult females depend on the presence of a functional germ line (the 3.1-kb female mid-gld and the 1.9-kb small transcripts) use polyadenylation sites internal to those of the germ line-independent transcripts. The germ line-dependent 3′ ends were mapped to regions k1a and h2i, respectively; no canonical AATAAA sequences (49) occur anywhere near their predicted positions. Instead, the related sequences AATATA and AT TAAA occur near these sites, as indicated in Fig. 2 (bp 1831, 1908, and 1932) and Fig. 3 (bp 3052). Variations in the AATAAA signal are relatively rare but well documented in eukaryotes (39, 63). An intriguing possibility now arises that such a variant is involved in tissue-specific polyadenylation in the female germ line. Use of such an alternative polyadenylation signal seems not to be strictly confined to the female germ line, however, since small *Sxl* RNAs are also seen in adult males and at early stages of embryogenesis.

Another *Sxl* transcript apparently produced by variant polyadenylation is the very large RNA. We believe it is similar to the large RNAs, but with a longer 3′ nontranslated region. Its 3′ end lies either in genomic region k3 or slightly

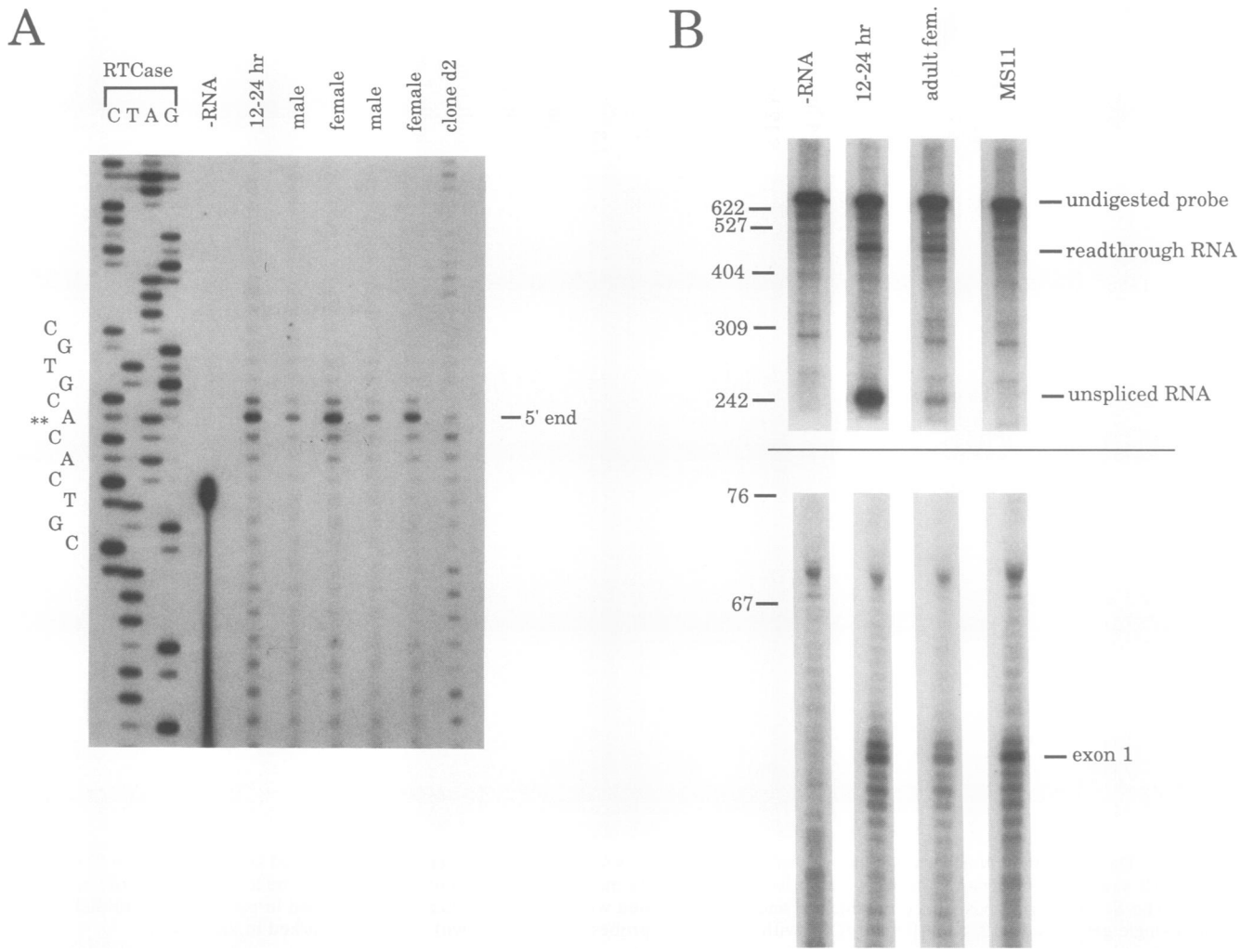


FIG. 5. 5'-end analysis of *Sxl*. (A) Primer extension. Primer extensions were performed as described in Materials and Methods, with an end-labeled oligonucleotide complementary to *Sxl* exon 1 sequences. The oligonucleotide was annealed to poly(A)$^+$ RNA from 12- to 24-h embryos, adult males (two different preparations), or adult females (two preparations). In the lane marked "clone d2," the oligonucleotide was annealed to sense strand synthetic RNA transcribed from *Sxl* subclone d2 containing exon 1 plus surrounding genomic sequences. In separate extension reactions with d2 RNA, dideoxynucleotides were included to generate the sequencing ladder used to align the in vivo RNA extension products. Asterisks indicate a product corresponding to the +1 A of the cDNAs. (B) RNase protection. A uniformly labeled d2 antisense RNA probe was mock hybridized or hybridized to 10 μg of 12- to 24-h embryonic mRNA or 15 μg of adult female mRNA. Nonhybridized RNA was removed by digestion with RNases at 1/10× (see Materials and Methods). In the positive control lane (MS11), the d2 antisense probe was annealed to nonradioactive synthetic plus-strand RNA transcribed from cDNA MS11.

downstream, and it is probably this RNA which gave rise to cDNA clones of the MS11 and MS16 types. Although most abundant in late embryos, the very large RNA was faintly detectable in adults and appeared to be the same size in both sexes (data not shown). Possibly the expected 200-nt size difference was below the limit of Northern blot resolution in this RNA size range.

In this study the fine structures of the late *Sxl* transcripts were compared in late embryos and in adults, with the result that no differences were observed at the 5' end or in internal splice junctions. Midembryonic, larval, and pupal stages were not examined in the same detail, but *Sxl* RNAs from these stages show the same sizes and hybridization patterns on Northern blots as the late embryonic and adult transcripts. We therefore conclude that the structures of the predominant *Sxl* late transcripts remain essentially constant throughout the life cycle, although the occurrence of additional minor transcripts cannot be ruled out.

The late male transcripts and the extent of sex-specific splicing. As was predicted from earlier experiments, the late male RNAs closely resemble the late female RNAs in fine structure. Bell et al. previously showed that all of the male transcript size classes include exon 3 in a sex-specific fashion (7). There were other splicing differences between the male and female cDNAs described by Bell et al., which are shown here not to be sex, tissue, or stage specific. Indeed, one of the most significant simplifying results in the present study is that no other exclusively sex-specific splicing event is observed besides that involving the translation-terminating exon 3. The absence in males of an equivalent to the mid-gld female transcript as discussed above probably reflects tissue rather than sex specificity per se.

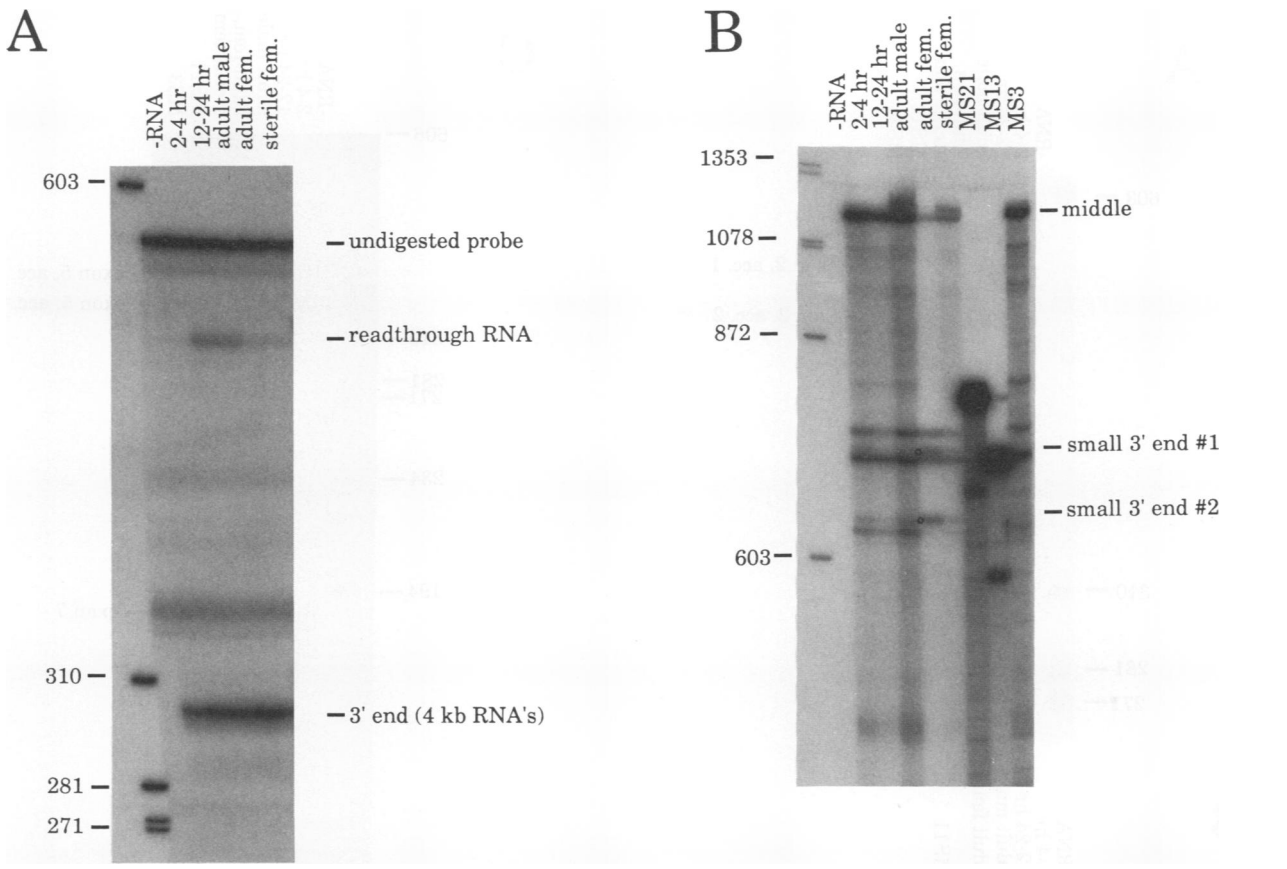


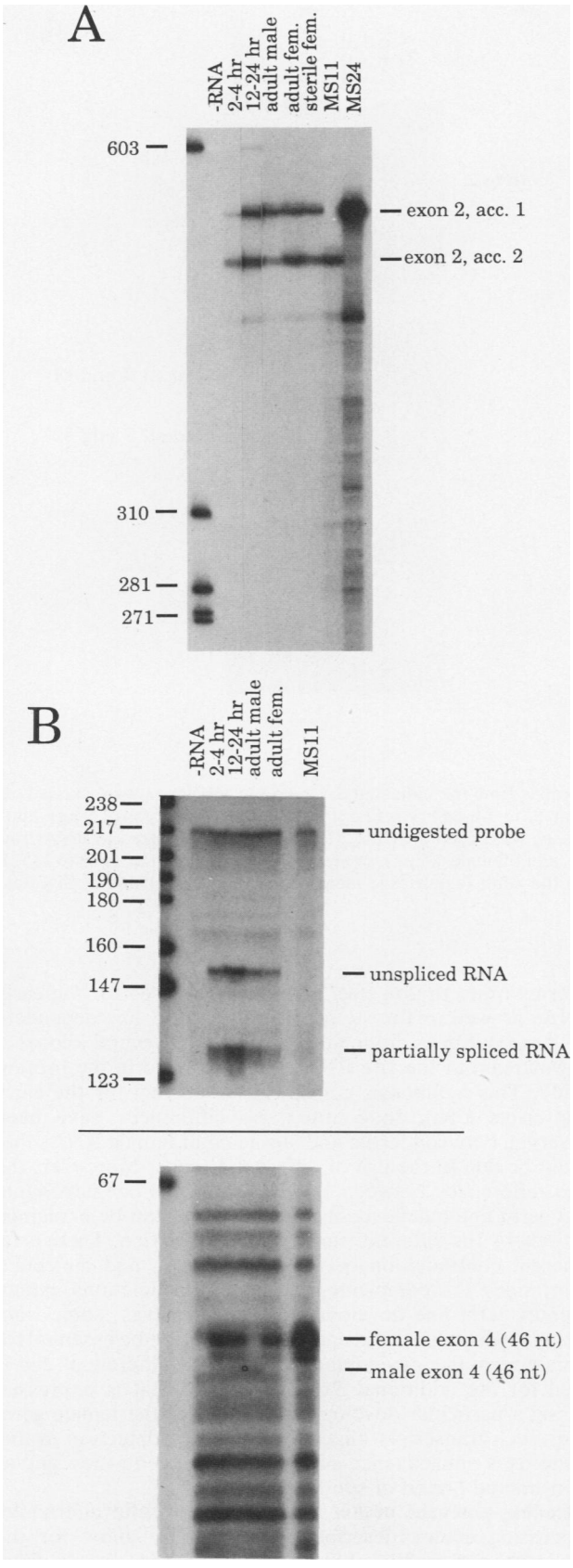
FIG. 6. 3'-end analysis of *Sxl*. RNase protections with antisense RNA probes from the indicated *Sxl* genomic subclones were carried out as described in Materials and Methods. (A) Probe k2 hybridized to RNA from 2- to 4-h embryos (2.5 μg), 12- to 24-h embryos (2.5 μg), adult males (10 μg), adult females (2.5 μg), or sterile adult females (2.5 μg). RNases were used at 1/10×. (B) Probe h2i hybridized to RNA from 2- to 4-h embryos (2.5 μg), 12- to 24-h embryos (2.5 μg), adult males (20 μg), adult females (5 μg), or sterile adult females (15 μg) or to MS21, MS13, or MS3 synthetic RNA. RNases were used at 1/90×. Open dots in the adult female lane indicate small 3' ends 1 and 2. Sizes are indicated on the left in base pairs.

It was previously noted that the late male transcripts appear somewhat later during embryonic development than the late female transcripts (51). This observation was confirmed in the present study. The different rates of appearance of the late transcripts in the two sexes could reflect a transcriptional bias for the late promoter when it is first activated at about 5 h or else differential stability of the male and female transcripts. Analysis of relative transcription rates in males versus females is further complicated by the differences in gene dosage in the two sexes. Changes in the ratios of the large and middle transcripts within each sex were also noted; these may reflect modifications of the general RNA processing machinery during development and in females may indicate differential requirements for the various *Sxl* protein products (see below).

The 5' end of *Sxl*. One major 5' end was mapped by primer extension and RNase protection. Although the mapping has a resolution of plus or minus approximately 1 nt, the known preference of RNA polymerase II for initiation at adenosine (5) supports the hypothesis that transcription of *Sxl* actually begins at the +1 A residue as shown in Fig. 5 rather than at one of the adjacent C's. However, the possibility of several neighboring initiation sites cannot be excluded.

On the basis of the ability of *Sxl* 5' partial deletions to complement certain mutant alleles, Salz et al. postulated that internal transcription start sites might be used for the early RNAs as well as for the adult female germ line-dependent RNAs (51). Such additional sites would lie several kilobases downstream of the late RNA 5' end mapped in the present study. This hypothesis appears to be correct for the early transcripts (30b). In contrast, no differences have been observed between fertile and sterile adult female RNAs that could be due to the use of different 5' ends. Moreover, the size differences between the various germ line-dependent and germ line-independent *Sxl* transcripts can be explained entirely by the different sites of polyadenylation. There is no inherent contradiction between the genetic and molecular data; since the complementing 5' partial deletions cannot support germ line development by themselves, some transcript(s) which have the standard 5' end must be essential for female germ line development. Possibly a different 5' end is used for the additional *Sxl* transcript(s) that is expressed during a particular developmental stage in the female germ line. Such transcripts might have escaped detection in this study as a consequence of their low concentration and/or their limited period of synthesis.

Coding potential of *Sxl* late transcripts. The alternative processing events described in this study allow for the generation of multiple *Sxl* polypeptides in female flies. Altogether, six different forms of *Sxl*-encoded protein are



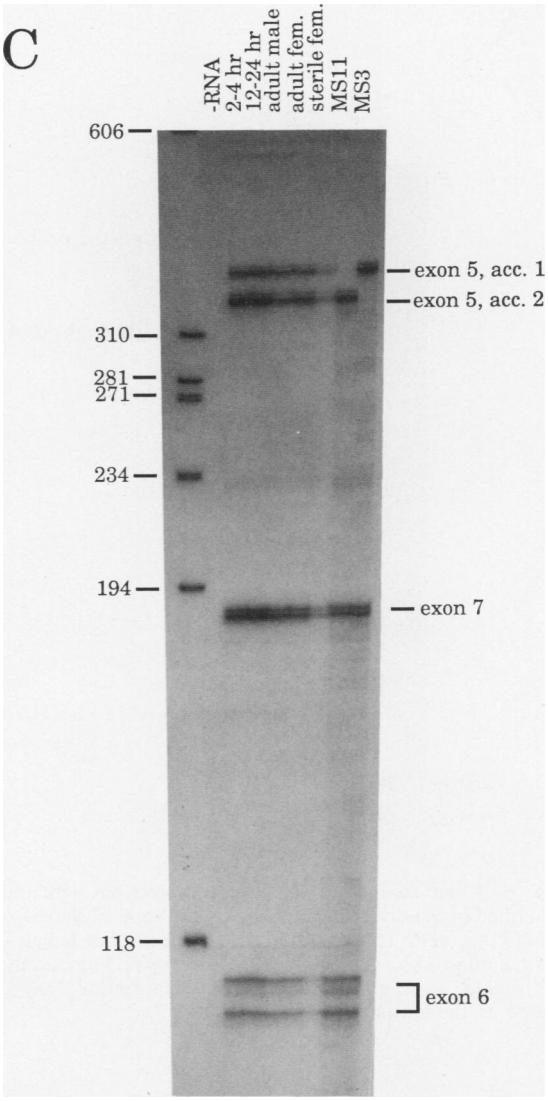


FIG. 7. Mapping internal *Sxl* exon junctions. RNase protections with antisense RNA probes from the indicated *Sxl* genomic subclones were carried out as described in Materials and Methods. (A) Probe f3 hybridized to RNA as in Fig. 6A, with 5 μg of RNA from sterile adult females and MS11 and MS24 synthetic RNAs. RNases were used at 1/30×. (B) Probe g3a hybridized to RNA from 2- to 4-h embryos (10 μg), 12- to 24-h embryos (10 μg), adult males (20 μg), or adult females (10 μg) or MS11. RNases were used at 1/10×. An open dot in the male RNA lane indicates the position of the exon 4 protection product. (C) Probe h1 hybridized to RNA as in Fig. 6A. RNases were used at 1×. The multiple bands observed for exon 6 result from variable extents of RNase digestion and not from true splice site heterogeneity as in the case of exon 5. Sizes on the left are in base pairs.

possible, assuming that all combinations of these processing variants arise. Because exons were mapped individually, it remains unknown whether in fact all combinations do arise in vivo. However, three of the six possible forms occur in female cDNAs cF1, MS3, and MS11. Moreover, multiple forms of *Sxl* protein have been observed on Western immunoblots of female *Drosophila* extracts. Some of these species comigrate with in vitro translation products and show distinct tissue specificity (11).

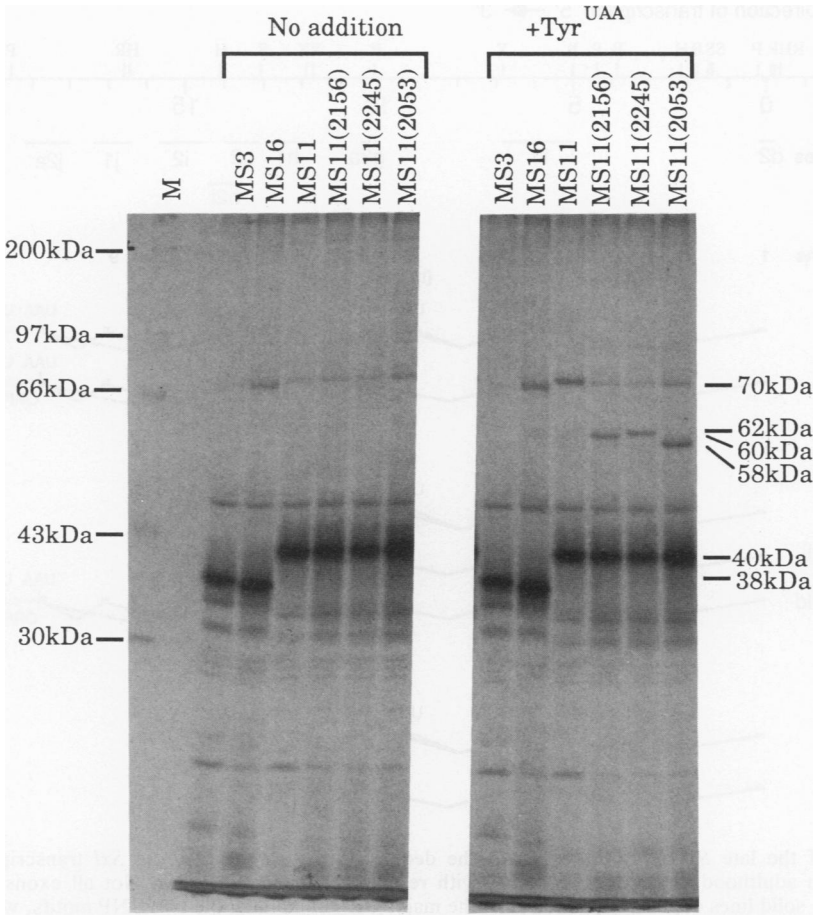


FIG. 8. Translation of synthetic *Sxl* mRNAs and effects of adding suppressor tRNA. Capped synthetic sense strand mRNAs were generated from *Sxl* cDNAs and translated in reticulocyte lysates as described in Materials and Methods. In lanes marked MS3, MS16, and MS11, the cDNAs were linearized at the 3′ ends of the inserts to create the transcription templates. In lanes marked MS11(2156), MS11(2245), and MS11(2053), cDNA MS11 was truncated within ORF2 at the indicated positions to create the transcription templates. In the left panel, only tRNAs endogenous to the reticulocyte lysate were present; in the right panel, partially purified yeast ochre suppressor $tRNA^{Tyr}$ was included in the translations.

The different predicted *Sxl* polypeptides potentially yield a considerable range in physical properties. As shown in Fig. 9, they range in size from 344 amino acids (37,393 Da) to 374 amino acids (40,743 Da). Furthermore, the predicted pIs span almost a full pH unit, from 9.96 to 10.88. These properties could readily be envisioned to modulate the biological specificity of *Sxl* proteins. On the other hand, none of the coding differences occurs within the two 90-amino-acid RNP motifs which have been postulated to constitute RNA-binding sites. Hence, if the differences alter RNA-binding specificity in vivo, the effects may involve long-range tertiary folding effects, sites of modification, or subunit interactions (whether homotypic or heterotypic).

As mentioned in the introduction, *Sxl* has three distinguishable functions in female somatic tissue. These include *Sxl* autoregulation, regulation of sexual differentiation via *tra*, and regulation of dosage compensation. It was speculated that the different forms of *Sxl* protein might be specialized for these different activities. It now appears that such specialization cannot be absolute, since a single *Sxl* cDNA encoding one particular form of the RNP ORF has been shown to possess all three somatic functions to some extent (6b). The differential distribution of *Sxl* polypeptides in various tissues could indicate more subtle functional differences among the proteins. In addition, the role of *Sxl* in the female germ line, where it may have a different set of targets than in somatic tissue, remains to be elucidated.

ORF2 and translation readthrough. A remarkable feature of *Sxl* is the presence of a second long ORF in one class of transcripts. This ORF2 occurs in exon 10 of the large and mid-gld female RNAs, in the same frame as ORF1 and separated from it by a single UAA stop codon. If used in fusion with ORF1, ORF2 would contribute to synthesis of a major variant form of *Sxl* gene product, one which might have significantly different biological and biochemical properties from the ORF1 forms described above. Situations in which separate ORFs are combined in a regulated fashion are found in a number of biologically interesting systems (60). These systems suggest several ways in which the protein-coding potential of ORF2 could be used. First, translation might continue through the UAA termination codon ending ORF1, directly into ORF2. Numerous examples of such stop codon readthrough are known (15, 21, 24, 35, 64). Although readthrough of *Sxl* ORF1-2 was not observed in vitro except in the presence of exogenous suppressor tRNAs, regulated readthrough in *D. melanogas-*

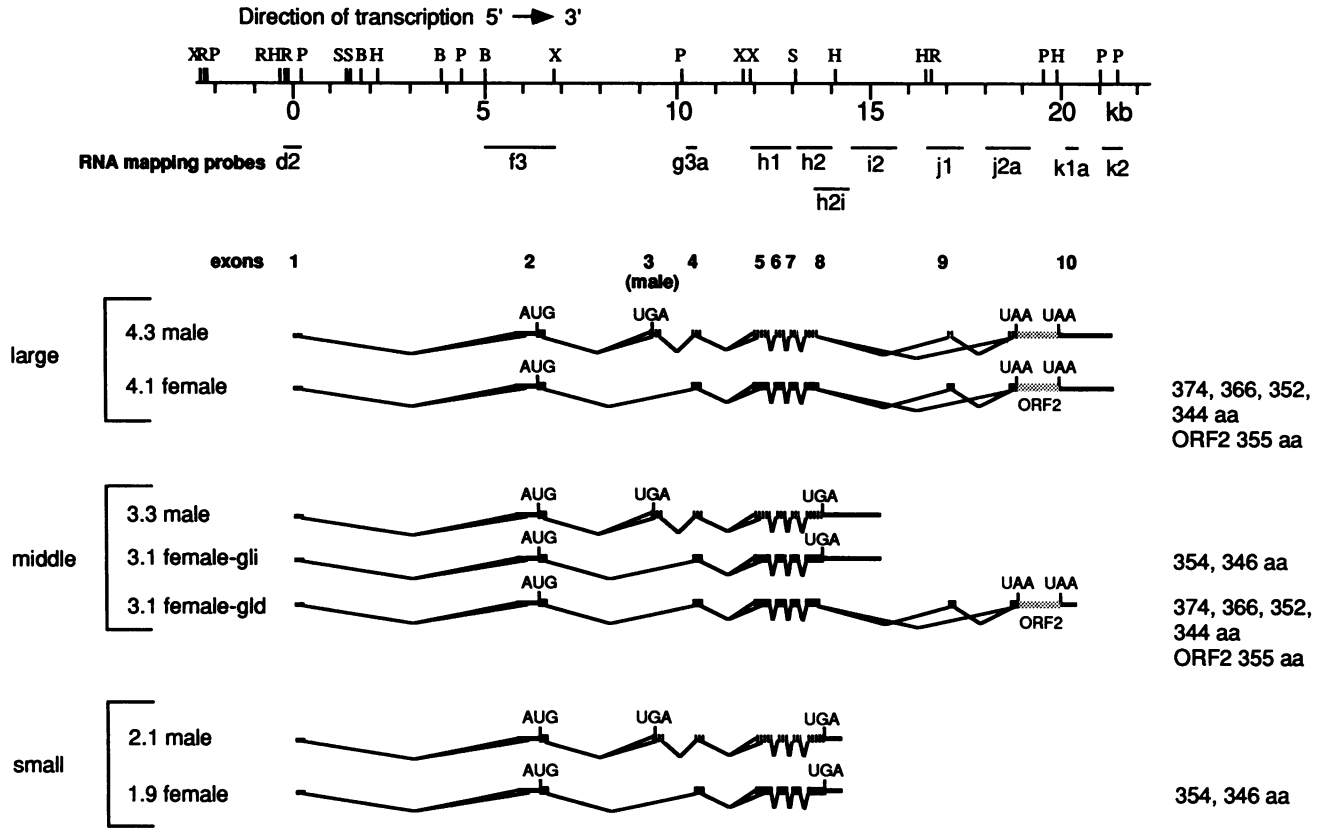


FIG. 9. Organization of the late *Sxl* transcription unit. The deduced structures of the late *Sxl* transcripts, those synthesized from midembryogenesis through adulthood, are shown, together with relevant genomic subclones. Not all exons are drawn to scale. In the female-specific transcripts, solid lines depict various forms of the major ORF containing the two RNP motifs, with translation start and stop signals as indicated. ORF2, encoded in exon 10, is shown by stippled boxes. Sizes in amino acids (aa) of polypeptides potentially encoded by each size class of female-specific transcript are given on the right, assuming that all possible combinations of RNA processing variants occur in each size class. In the male-specific transcripts, regions lying between the termination codon in exon 3 and the termination codons of the female RNP-ORFs are shown in vertical hatch marks. Uncertainties in the structure of the 3.1-kb mid-gld female transcript are discussed in the text. See Fig. 1 for an explanation of the numbering system.

ter could conceivably take place in a particular tissue or at a particular stage of development with factors not present in the heterologous translation system. Another possibility is that alternative splicing could generate a message fusing ORF1 directly to ORF2, bypassing the stop codon completely. However, RNase protections with probe j2a did not detect such a splice at any stage of embryonic, larval, or adult development.

ORF2 might conceivably be read as an independent translation unit, either by reinitiation of translation at a methionine codon following termination at the UAA stop codon or by true internal initiation by ribosomes bypassing ORF1. There are several examples of translation reinitiation in eukaryotic systems, in which the first ORF may be quite short (26, 27, 34) or very long (47, 61). As a final alternative, *Sxl* ORF2 might be encoded independently by another transcription unit beginning either in exon 10 itself or slightly upstream. This seems unlikely, since Northern blot data provide no evidence for such an RNA. In addition, deletions of *Sxl* which encompass ORF2 are wholly male viable (50); hence, ORF2 cannot be part of an independent essential gene. Conceivably, an independently translated ORF2 product might play a role in *Sxl* expression or function in females.

Amino acid polymorphism in *Sxl*. In comparing the various cDNAs with each other and with genomic *Sxl* sequences, nonconservative amino acid variants at two nearby positions in exon 5 were discovered (Fig. 2 and 3). No other amino acid variants were observed at any positions in ORF1. The fact that three of the four possible combinations of these two variants occur in a small sample of cDNAs derived from several independent libraries suggests that these represent long-established polymorphisms. The possible functional significance of these polymorphisms remains to be determined.

Early *Sxl* transcripts. As discussed above, *Sxl* expresses a set of transcripts uniquely early (2 to 4 h) during embryogenesis (51). Although the present study does not focus on these early transcripts, we can conclude that these early RNAs do not contain differently spliced forms of those exons which are shared with the late RNAs and that the large, middle, and small early RNAs probably do correspond in exons 4 to 10 to the large, mid-gli, and small late female RNA species. These exons include the bulk of protein-coding potential, including the two RNP motifs.

Unspliced *Sxl* RNA. An unexpected result of the RNase protection experiments was the observation of appreciable amounts of apparently unspliced *Sxl* RNA with some probes. One interpretation of these data is that splicing of

some regions of *Sxl* is relatively slow, allowing the accumulation of processing intermediates to high levels. This is especially intriguing with respect to exon 4, since splicing of this exon constitutes a part of the critical regulatory mechanism of *Sxl* expression (7, 51). The results presented in Fig. 7 suggest that the sex-specific upstream splice of exon 2 or 3 into exon 4 is slower than the constitutive downstream exon 4-exon 5 splice. Possibly an inverse correlation exists for *Sxl* between the efficiency of intervening sequence excision and accessibility of a particular splice to biological regulation.

ACKNOWLEDGMENTS

We thank H. Salz and E. Maine for providing many of the *Sxl* DNAs employed in this study and for introducing us to the complexities of *Sxl* gene expression. We also thank L. Bell, M. Bernstein, J. Erickson, and J. Horabin for valuable discussions and D. Bopp, L. Keyes, and G. Calhoun for communicating results prior to publication. P. Sziber and G. Calhoun provided excellent technical assistance.

This work was supported by grants to M.E.S., P.S., and T.W.C. from the National Institutes of Health, to P.S. from the American Cancer Society, and to M.E.S. from the Life Sciences Research Foundation (Burroughs Wellcome Fund fellowship).

REFERENCES

1. **Amasino, R. M.** 1986. Acceleration of nucleic acid hybridization rate by polyethylene glycol. Anal. Biochem. **152**:304–307.
2. **Baker, B. S.** 1989. Sex in flies: the splice of life. Nature (London) **340**:521–524.
3. **Baker, B. S., and J. M. Belote.** 1983. Sex determination and dosage compensation in *Drosophila melanogaster*. Annu. Rev. Genet. **17**:345–393.
4. **Baker, B. S., R. N. Nagoshi, and K. C. Burtis.** 1987. Molecular genetic aspects of sex-determination in *Drosophila*. Bioessays **6**:66–70.
5. **Baker, C. C., and E. B. Ziff.** 1981. Promoters and heterogeneous 5′ termini of the messenger RNA's of adenovirus serotype 2. J. Mol. Biol. **149**:189–221.
6. **Bandziulis, R. J., M. S. Swanson, and G. Dreyfuss.** 1989. RNA-binding proteins as developmental regulators. Genes Dev. **3**:431–437.

6a. **Bell, L.** Personal communication.

6b. **Bell, L. R., J. I. Horabin, P. Schedl, and T. W. Cline.** 1991. Positive autoregulation of *Sex-lethal* by alternative splicing maintains the female determined state in Drosophila. Cell **65**: 229–239.

7. **Bell, L. R., E. M. Maine, P. Schedl, and T. W. Cline.** 1988. *Sex-lethal*, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. Cell **55**:1037–1046.
8. **Belote, J. M.** 1983. Male-specific lethal mutations of *Drosophila melanogaster*. II. Parameters of gene action during male development. Genetics **105**:881–896.
9. **Bienz, M., E. Kubli, J. Kohli, S. de Henau, and H. Grosjean.** 1980. Nonsense suppression in eukaryotes: the use of the Xenopus oocyte as an *in vivo* assay system. Nucleic Acids Res. **8**:5169–5178.
10. **Boggs, R. T., P. Gregor, S. Idriss, J. M. Belote, and M. McKeown.** 1987. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the *transformer* gene. Cell **50**:739–747.
11. **Bopp, D., L. R. Bell, T. W. Cline, and P. Schedl.** 1991. Developmental distribution of female-specific *Sex-lethal* proteins in *Drosophila melanogaster*. Genes Dev. **5**:403–415.

11a. **Brown, N.** Personal communication.

12. **Brown, N. H., and F. C. Kafatos.** 1988. Functional cDNA libraries from *Drosophila* embryos. J. Mol. Biol. **203**:425–437.
13. **Busson, B., M. Gans, K. Komitopoulou, and M. Masson.** 1983. Genetic analysis of three dominant female sterile mutations located on the X-chromosome of *Drosophila melanogaster*. Genetics **105**:309–325.
14. **Butler, B., V. Pirrotta, I. Irminger-Finger, and R. Nöthiger.** 1986. The sex-determining gene *tra* of *Drosophila*: molecular cloning and transformation studies. EMBO J. **5**:3607–3613.

14a. **Calhoun, G.** Personal communication.

15. **Chambers, I., J. Frampton, P. Goldfarb, N. Affara, W. McBain, and P. R. Harrison.** 1986. The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by a 'termination' codon, TGA. EMBO J. **5**:1221–1227.
16. **Cline, T. W.** 1979. A male-specific lethal mutation in *Drosophila melanogaster* that transforms sex. Dev. Biol. **72**:266–275.
17. **Cline, T. W.** 1983. The interaction between *daughterless* and *Sex-lethal* in triploids: a lethal sex-transforming maternal effect linking sex-determination and dosage compensation in *Drosophila melanogaster*. Dev. Biol. **95**:260–274.
18. **Cline, T. W.** 1984. Autoregulatory functioning of a *Drosophila* gene product that establishes and maintains the sexually determined state. Genetics **107**:231–277.
19. **Cline, T. W.** 1985. Primary events in the determination of sex in *Drosophila melanogaster*, p. 301–327. *In* H. O. Halvorson and A. Monroy (ed.), Origin and evolution of sex. Alan R. Liss, Inc., New York.
20. **Cline, T. W.** 1988. Evidence that *sisterless-a* and *sisterless-b* are two of several discrete "numerator elements" of the X/A sex determination signal in *Drosophila* that switch *Sxl* between two alternative stable expression states. Genetics **119**:829–862.
21. **Feng, Y.-X., D. L. Hatfield, A. Rein, and J. G. Levin.** 1989. Translational readthrough of the murine leukemia virus *gag* gene amber codon does not require virus-induced alteration of tRNA. J. Virol. **63**:2405–2410.
22. **Geliebter, J., R. A. Zeff, R. W. Melvold, and S. G. Nathenson.** 1986. Mitotic recombination in germ cells generated two major histocompatibility complex mutant genes shown to be identical by RNA sequence analysis: K^{bm9} and K^{bm6}. Proc. Natl. Acad. Sci. USA **83**:3371–3375.
23. **Gergen, J. P.** 1987. Dosage compensation in *Drosophila*: evidence that *daughterless* and *Sex-lethal* control X chromosome activity at the blastoderm stage of embryogenesis. Genetics **117**:477–485.

23a. **Goldschmidt-Clermont, M.** Unpublished data.

24. **Hatfield, D.** 1985. Suppression of termination codons in higher eukaryotes. Trends Biochem. Sci. **10**:201–204.
25. **Henikoff, S.** 1987. Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol. **155**:156–165.
26. **Hinnebusch, A. G.** 1984. Evidence for translational regulation of the activator of general amino acid control in yeast. Proc. Natl. Acad. Sci. USA **81**:6442–6446.
27. **Hinnebusch, A. G.** 1988. Novel mechanisms of translational control in *Saccharomyces cerevisiae*. Trends Genet. **4**:169–174.
28. **Hodgkin, J.** 1989. *Drosophila* sex determination: a cascade of regulated splicing. Cell **56**:905–906.
29. **Hodgkin, J.** 1990. Sex determination compared in *Drosophila* and *Caenorhabditis*. Nature (London) **344**:721–728.
30. **Inoue, K., K. Hoshijima, H. Sakamoto, and Y. Shimura.** 1990. Binding of the *Drosophila Sex-lethal* gene product to the alternative splice site of *transformer* primary transcript. Nature (London) **344**:461–463.

30a. **Keyes, L.** Personal communication.

30b. **Keyes, L.** Unpublished data.

31. **Kohli, J., T. Kwong, F. Altruda, D. Soll, and G. Wahl.** 1979. Characterization of a UGA-suppressing serine tRNA from *Schizosaccharomyces pombe* with the help of a new *in vitro* assay system for eukaryotic suppressor tRNA's. J. Biol. Chem. **254**:1546–1551.
32. **Konarska, M. M., R. A. Padgett, and P. A. Sharp.** 1984. Recognition of cap structure in splicing *in vitro* of mRNA precursors. Cell **38**:731–736.
33. **Kornfeld, K., R. B. Saint, P. A. Beachy, P. J. Harte, D. A. Beattie, and D. S. Hogness.** 1989. Structure and expression of a family of *Ultrabithorax* mRNA's generated by alternative splicing and polyadenylation in *Drosophila*. Genes Dev. **3**:243–258.
34. **Kozak, M.** 1986. Bifunctional messenger RNA's in eukaryotes. Cell **47**:481–483.
35. **Li, G., and C. M. Rice.** 1989. Mutagenesis of the in-frame opal

termination codon preceding nsP4 of Sindbis virus: studies of translational readthrough and its effect on virus replication. J. Virol. **63**:1326–1337.
36. **Lucchesi, J. C., and E. Manning.** 1987. Gene dosage compensation in *D. melanogaster*. Adv. Genet. **24**:371–429.
37. **Lucchesi, J. C., and T. Skripsky.** 1981. The link between dosage compensation and sex differentiation in *Drosphila melanogaster*. Chromosoma **82**:217–227.
38. **Maine, E. M., H. K. Salz, P. Schedl, and T. W. Cline.** 1985. *Sex-lethal*, a link between sex determination and sexual differentiation in *Drosophila melanogaster*. Cold Spring Harbor Symp. Quant. Biol. **50**:595–604.
39. **Manley, J. L.** 1988. Polyadenylation of mRNA precursors. Biochim. Biophys. Acta **950**:1–12.
40. **McKeown, M., J. M. Belote, and B. S. Baker.** 1987. A molecular analysis of *transformer*, a gene in *Drosophila melanogaster* that controls female sexual differentiation. Cell **48**:489–499.
41. **McKeown, M., J. M. Belote, and R. T. Boggs.** 1988. Ectopic expression of the female *transformer* gene product leads to female differentiation of chromosomally male *Drosophila*. Cell **53**:887–895.
42. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
43. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
44. **Nagoshi, R. N., M. McKeown, K. C. Burtis, J. M. Belote, and B. S. Baker.** 1988. The control of alternative splicing at genes regulating sexual differentiation in *D. melanogaster*. Cell **53**: 229–236.
45. **O'Connel, P., and M. Rosbash.** 1984. Sequence, structure, and codon preference of the *Drosophila* ribosomal protein 49 gene. Nucleic Acids Res. **12**:5495–5513.
46. **O'Hare, K., R. Levis, and G. M. Rubin.** 1983. Transcription of the *white* locus in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **80**:6917–6921.
47. **Peabody, D. S., and P. Berg.** 1986. Termination-reinitiation occurs in the translation of mammalian cell mRNAs. Mol. Cell. Biol. **6**:2695–2703.
48. **Perrimon, N., D. Mohler, L. Engstrom, and A. P. Mahowald.** 1986. X-linked female-sterile loci in *Drosophila melanogaster*. Genetics **113**:695–712.
49. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–214.
49a. **Salz, H.** Personal communication.
50. **Salz, H. K., T. W. Cline, and P. Schedl.** 1987. Functional changes associated with structural alterations induced by mobilization of a P element inserted in the *Sex-lethal* gene of *Drosophila*. Genetics **117**:221–231.
51. **Salz, H. K., E. M. Maine, L. N. Keyes, M. E. Samuels, T. W. Cline, and P. Schedl.** 1989. The *Drosophila* female-specific sex-determination gene, *Sex-lethal*, has stage-, tissue-, and sex-specific RNAs suggesting multiple modes of regulation. Genes Dev. **3**:708–719.
52. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
53. **Sanchez, L., and R. Nöthiger.** 1982. Clonal analysis of *Sex-lethal*, a gene needed for female sexual development in *Drosophila melanogaster*. Wilhelm Roux's Arch. Dev. Biol. **191**:211–214.
54. **Schupbach, T.** 1985. Normal female germ cell differentiation requires the female X chromosome to autosome ratio and expression of *Sex-lethal* in *Drosophila melanogaster*. Genetics **109**:529–548.
55. **Sosnowski, B. A., J. M. Belote, and M. McKeown.** 1989. Sex-specific alternative splicing of RNA from the *transformer* gene results from sequence-dependent splice site blockage. Cell **58**:449–459.
56. **Staden, R.** 1986. The current status and portability of our sequence handling software. Nucleic Acids Res. **14**:217–231.
57. **Steinmann-Zwicky, M., H. Amrein, and R. Nöthiger.** 1990. Genetic control of sex-determination in *Drosophila*. Adv. Genet. **27**:189–237.
58. **Swanson, M. S., T. Y. Nakagawa, K. LeVan, and G. Dreyfuss.** 1987. Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. Mol. Cell. Biol. **7**:1731–1739.
59. **Uenoyama, T., A. Fukunaga, and K. Ioshi.** 1982. Studies on the sex-specific lethals of *Drosophila melanogaster*. V. Sex transformation caused by interactions between a female-specific lethal, Sxl^{f*1}, and the male-specific lethals *mle(3)132*, $msl\text{-}2^{27}$, and *mle*. Genetics **102**:233–243.
60. **Valle, R. P. C., and M.-D. Morch.** 1988. Stop making sense, or regulation at the level of termination in eukaryotic protein synthesis. FEBS Lett. **235**:1–15.
61. **Wang, F., L. Petti, D. Braun, S. Seung, and E. Kieff.** 1987. A bicistronic Epstein-Barr virus mRNA encodes two nuclear proteins in latently infected, growth-transformed lymphocytes. J. Virol. **61**:945–954.
62. **Wharton, K. A., B. Yedvobnick, V. G. Finnerty, and S. Artavanis-Tsakonas.** *opa*: a novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in *D. melanogaster*. Cell **40**:55–62.
63. **Wickens, M., and P. Stephenson.** 1984. Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent messenger RNA 3′ end formation. Science **226**:1045–1051.
64. **Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan.** 1985. Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA **82**:1618–1622.