MOLECULAR AND CELLULAR BIOLOGY, Feb. 1991, p. 894–905
0270-7306/91/020894-12$02.00/0

The *Drosophila* Suppressor of Sable Gene Encodes a Polypeptide with Regions Similar to Those of RNA-Binding Proteins

ROBERT A. VOELKER,* WILLIE GIBSON, JOAN P. GRAVES, JOAN F. STERLING, AND MARCIA T. EISENBERG

Laboratory of Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

Received 13 July 1990/Accepted 14 November 1990

The nucleotide sequence of the *Drosophila melanogaster* suppressor of sable [*su(s)*] gene has been determined. Comparison of genomic and cDNA sequences indicates that an ~7,860-nucleotide primary transcript is processed into an ~5-kb message, expressed during all stages of the life cycle, that contains an open reading frame capable of encoding a 1,322-amino-acid protein of ~150 kDa. The putative protein contains an RNA recognition motif-like region and a highly charged arginine-, lysine-, serine-, and aspartic or glutamic acid-rich region that is similar to a region contained in several RNA-processing proteins. In vitro translation of in vitro-transcribed RNA from a complete cDNA yields a product whose size agrees with the size predicted by the open reading frame. Antisera against *su(s)* fusion proteins recognize the in vitro-translated protein and detect a protein of identical size in the nuclear fractions from tissue culture cells and embryos. The protein is also present in smaller amounts in cytoplasmic fractions of embryos. That the *su(s)* protein has regions similar in structure to RNA-processing proteins is consistent with its known role in affecting the transcript levels of those alleles that it suppresses.

The factors that effect genetic regulation are likely to include a category of regulators that influences the fate and stability of unspliced and/or spliced message. Here we describe a protein whose structure suggests that it may be an RNA-binding protein and whose function indicates that it may function in determining the level or stability of mRNAs (14, 16a).

In *Drosophila melanogaster*, altered, reduced, or loss-of-function mutations at the suppressor of sable [*su(s)*] locus suppress mutations at second-site loci that are caused by 412 or P-element insertions. Suppressible mutations at the vermilion (*v*) locus are caused by insertions of the mobile element 412; the suppressible alleles of the uncloned purple and speck loci are probably also caused by 412 insertions (32, 38). Studies of *v* suggest that suppression occurs pretranslationally: the ~10 to 20% of the wild-type amount of *v* message in organisms with suppressed genotypes parallels the ~10 to 20% of the wild-type amount of tryptophan oxygenase (the *v* translation product) observed in organisms with the same genotypes, whereas in organisms with unsuppressed genotypes, the levels of *v* message and tryptophan oxygenase are virtually undetectable (31–33). Thus, suppression appears to influence the amount of processed transcript produced by the suppressible alleles. Learning how this increase in message level is effected is one of the goals of our investigations. This information may provide insights into how the *su(s)* protein functions in the processing of transcripts that are not interrupted by mobile-element insertions.

In this paper we present the results of our molecular analysis of the *su(s)* locus. The nucleotide sequence of the genomic segment that encodes the *su(s)* function has been determined, and by using cDNA analysis, we have deduced the structure of the coding region. The *su(s)* protein as we imagine it has structural properties that suggest that it might be involved in RNA metabolism because it contains regions of similarity to proteins that are known to be involved in RNA processing. By cell fractionation and immunodetection procedures, we show that the *su(s)* protein occurs primarily in the nucleus of *Drosophila* cells from cell culture and embryos.

MATERIALS AND METHODS

Genomic DNA sequence determination. Previous work indicated that the ~8-kb *Hin*dIII segment of genomic DNA (Fig. 1A) is necessary and sufficient to give $su(s)^+$ function (37). The *Eco*RI fragments of this region plus several hundred bases on each end were subcloned into M13mp18 or M13mp19, and both strands were sequenced by the Sanger dideoxy method, as shown in Fig. 1B. Nested deletions were prepared (7) and generally allowed the complete sequence of both strands to be determined. Gaps in the sequence were filled in by using synthetic oligonucleotide primers to obtain the sequence information necessary for overlap. Synthetic primers were also prepared to sequence across the three *Eco*RI sites to ensure that sequence information was not overlooked because of closely situated *Eco*RI sites. Initially, sequencing was carried out with the Klenow fragment and ^{32}P or ^{35}S. However, early in the project, a change was made to Sequenase (U.S. Biochemical). Generally, only the dGTP set of reactions was run, but compression ambiguities were resolved by using both the dGTP and dITP sets of reactions to sequence both strands.

Construction and screening of cDNA library. Because screens of existing cDNA libraries failed to yield *su(s)* cDNA clones longer than ~1 kb, a cDNA library was constructed by Stratagene, Inc., for obtaining longer clones. Doubly oligo(dT)-selected poly(A)$^+$ RNA was prepared from $y^2\ w^{bf}$ [$=su(s)^+$] embryos and supplied to Stratagene, Inc., for the synthesis of the cDNA library as follows. First-strand synthesis was carried out using methyl mercuric hydroxide and avian myeloblastosis virus reverse transcriptase. After second-strand synthesis with reverse transcriptase, methyla-

* Corresponding author.

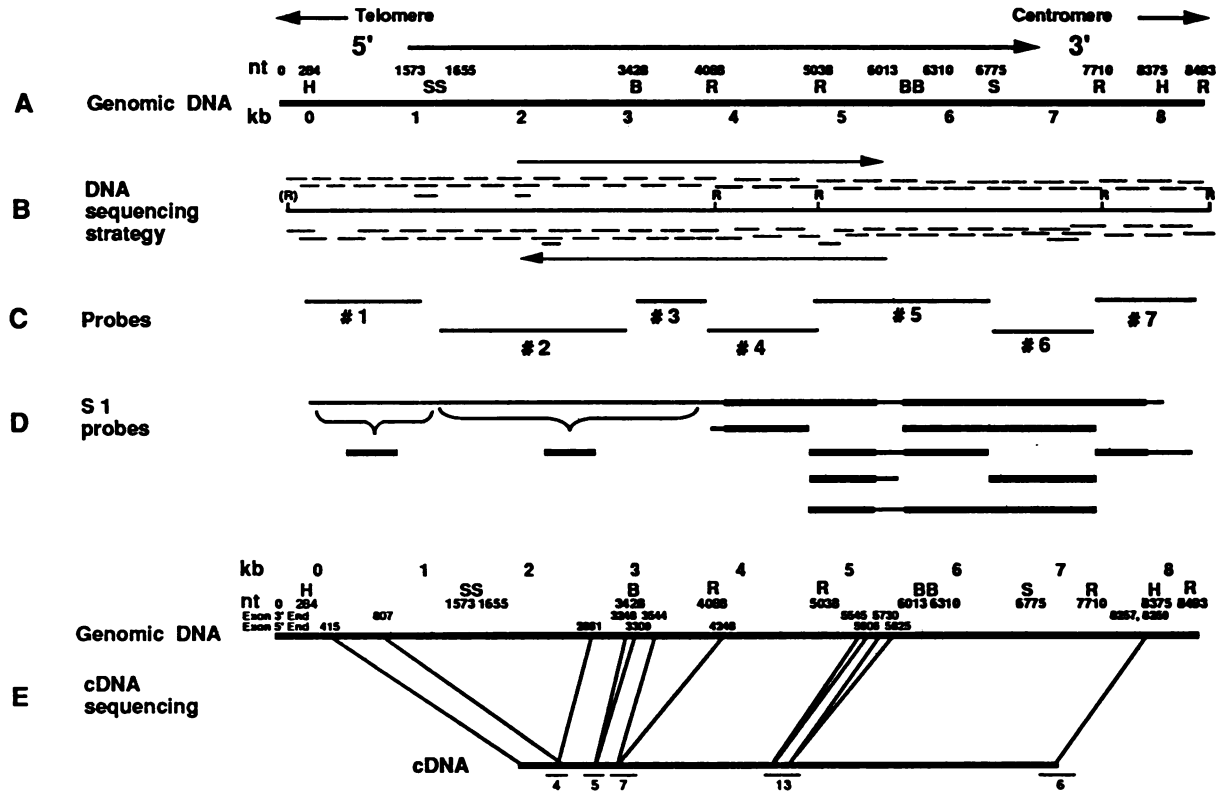
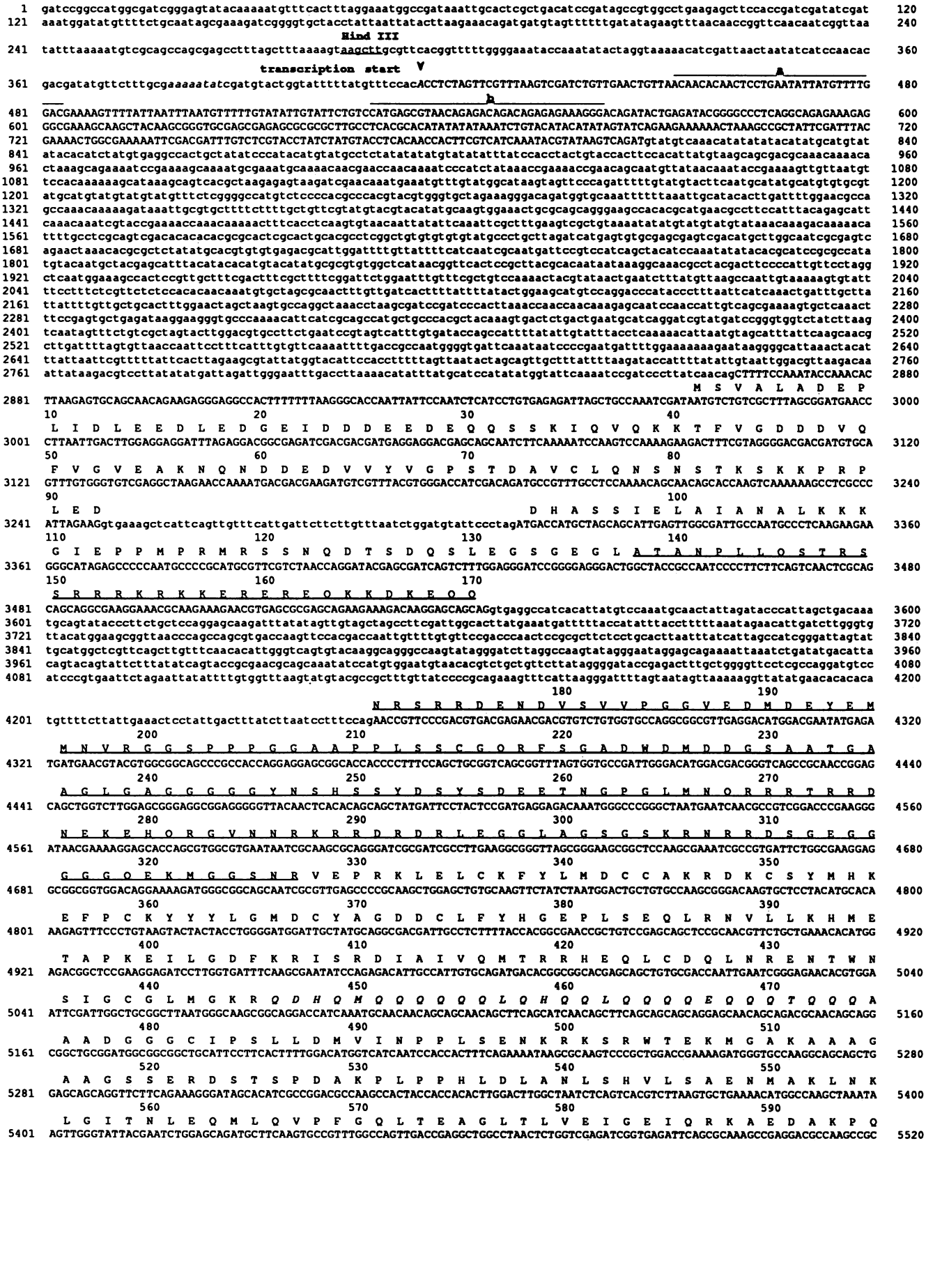


FIG. 1. Molecular characterization of the *su(s)* region. (A) Molecular structure of genomic *su(s)* region. The kilobase scale, with 0 indicating the site of the P-element insertion that was used to clone the gene by transposon tagging, and transcription direction are as previously described (5). The nucleotide numbers indicate the locations of the restriction sites in the DNA. The 0 to +8.0 *Hin*dIII fragment, when reintroduced by P-element transformation of embryos, was shown to rescue *su(s)* mutations (37). Restriction enzymes: B, *Bam*HI; H, *Hin*dIII; R, *Eco*RI; S, *Sal*I. (B) Sequencing of genomic *su(s)* region. The four *Eco*RI fragments [(R) indicates a λEMBL4 linker-derived site] were cloned into M13 in both orientations and sequenced after a nested deletion series was generated (7). The short lines above and below the genomic DNA indicate the lengths of the sequences that were integrated by computer analysis (University of Wisconsin Genetics Computer Group Software). Oligonucleotide primers were synthesized to sequence across the *Eco*RI sites in a clone that contained the 0 to +8.0 *Hin*dIII fragment. (C) Probes. Fragments used as radioactively labeled probes as indicated in text. (D) Results of S1 mapping. The continuous thick and thin lines represent the cloned genomic fragments (using the restriction sites shown in panel A as ends) that were annealed as unlabeled probes to wild-type poly(A)$^+$ RNA. After treatment with S1 nuclease and separation on an alkaline agarose gel, the DNAs were blotted to nitrocellulose filters and probed with the radioactively labeled probes shown in panel C. The thick portions of the lines indicate the portions that were protected, indicating that they are exons. The exons at the left were located within the region spanned by the braces. (E) Location of splice sites by cDNA sequencing. The relationship between genomic DNA and cDNA sequences is shown, indicating the pattern of splicing of the primary transcript to yield the mature message. The numbers under the cDNA line indicate the number of independent cDNA clones that were sequenced across each splice site. Restriction sites are abbreviated as in panel A.

tion, and blunting of the ends with S1 nuclease, *Eco*RI linkers were ligated to the cDNA. The cDNAs were digested with *Eco*RI, the excess linkers were removed, and the cDNAs were ligated into λ_{ZAP} (Stratagene) that had been cut with *Eco*RI and dephosphorylated. The library was amplified and screened.

Approximately 10^9 plaques were plated. Duplicate sets of nitrocellulose (Schleicher & Schuell) lifts were prepared. The filters were initially probed with a ^{32}P-labeled (Boehringer Mannheim random primer labeling kit) mixture of probes 4 and 7 (Fig. 1C), which yielded ~150 positive plaques. The filters were stripped, and one set of filters was reprobed with probe 4 and the other set was reprobed with probe 7. The filters were again stripped, and one set was reprobed with probe 1 and the other was reprobed with probe 6. The results of these screenings yielded a subset of clones that contained inserts from across the entire gene. Twenty-seven clones were plaque purified, and the pBSK− vector and insert were excised from λ_{ZAP} by the recommended procedure (Stratagene). These were analyzed by restriction mapping and DNA sequencing from the primer sites flanking each end of the polylinker. They represented 19 different cDNA clones with the following recovery frequencies: 1 clone recovered four times, 5 clones recovered twice, and 13 clones recovered once each. DNA sequencing revealed that 10 of the 19 clones contained mixed inserts [e.g., the insert consisted of a partial *su(s)* cDNA that had ligated directly to another cDNA of unknown origin that had not been separated by an *Eco*RI linker]. This apparently resulted from too low an effective concentration of linkers when they were ligated to the cDNAs. This did not present a serious problem, because the cDNA sequences were being compared with the already completely determined genomic sequence.

Primer extension. Two synthetic oligonucleotides were used in the primer extension analysis to prime approximately


Hind III
transcription start
M S V A L A D E P
L I D L E E D L E D G E I D D D E E D E Q Q S S K I Q V Q K K T F V G D D D V Q
F V G V E A K N Q N D D E D V V Y V G P S T D A V C L Q N S N S T K S K K P R P
L E D D H A S S I E L A I A N A L K K K
G I E P P M P R M R S S N Q D T S D Q S L E G S G E G L A T A N P L L Q S T R S
S R R R K R K K E R E R E Q K K D K E Q Q
N R S R R D E N D V S V V P G G V E D M D E Y E M
M N V R G G S P P P P G G A A P P L S S C G Q R F S G A D W D M D D G S A A T G A
A G L G A G G G G G G Y N S H S S Y D S Y S D E E T N G P G L M N Q R R R T R R D
N E K E H Q R G V N N R K R R D R D R L E G G L A G S G S K R N R R D S G E G G
G G G Q E K M G G S N R V E P R K L E L C K F Y L M D C C A K R D K C S Y M H K
E F P C K Y Y Y L G M D C Y A G D D C L F Y H G E P L S E Q L R N V L L K H M E
T A P K E I L G D F K R I S R D I A I V Q M T R R H E Q L C D Q L N R E N T W N
S I G C G L M G K R Q D H Q M Q Q Q Q Q Q Q L Q H Q Q L Q Q Q Q Q E Q Q Q T Q Q Q A
A A D G G G C I P S L L D M V I N P P L S E N K R K S R W T E K M G A K A A A G
A A G S S E R D S T S P D A K P L P P H L D L A N L S H V L S A E N M A K L N K
L G I T N L E Q M L Q V P F G Q L T E A G L T L V E I G E I Q R K A E D A K P Q

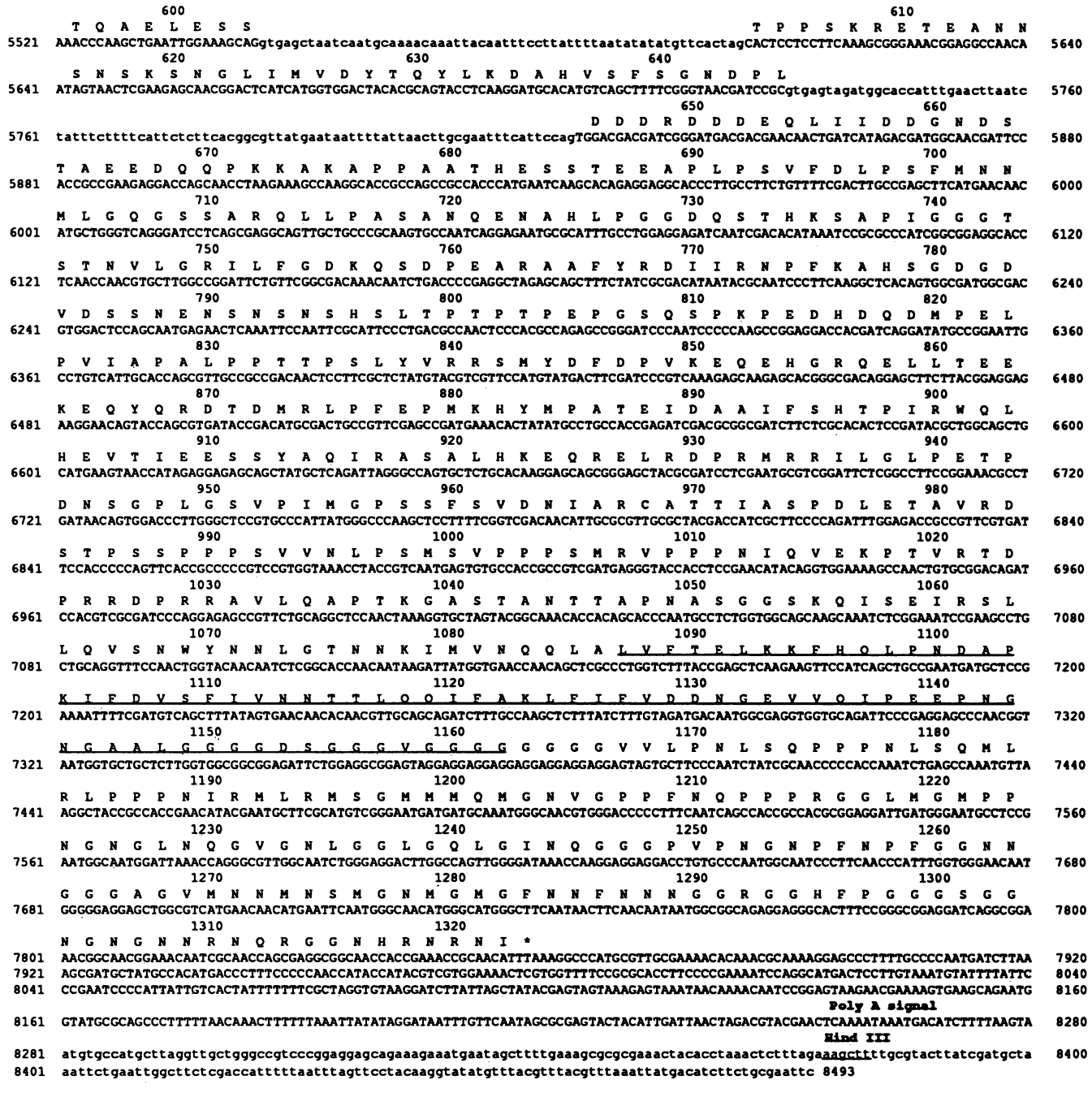


FIG. 2. Nucleotide sequence of genomic *su(s)* region, the ~8.5-kb genomic fragment shown in Fig. 1A. Capital letters in the nucleotide sequence indicate exons, while lowercase letters indicate nontranscribed or intron sequences. The italicized nucleotide A-415 (indicated by the arrowhead) denotes the transcription start site, and the *aaaaatat* at a-379 is the TATA-like box. **a** and **b** above the lines indicate oligonucleotide primers used to localize the transcription start site. The poly(A) signal is at 8259 to 8264. Single-letter amino acid designations denote the putative protein with translation initiating at the first ATG in the large ORF. The underlined amino acids (138 to 327 and 1087 to 1162) show similarity to several poly(A)$^+$ RNA-processing proteins. The italicized amino acids (446 to 474) denote the opa-like sequence.

140 and 80 nucleotides from the 5′ end of the mRNA (Fig. 2). The oligonucleotides were end labeled with T4 polynucleotide kinase (Promega) and [γ-^{32}P]dATP (NEN; 6,000 Ci/mmol) as described previously (8). Poly(A)$^+$ adult male or embryonic RNA (10 μg) was used as a template according to published procedures (4). The primer extension products were analyzed on 8 M urea–6% acrylamide gels and subsequently autoradiographed. Sequencing reactions using the same end-labeled oligonucleotide used for the primer extension were done with the appropriate DNA templates and run to provide size markers.

S1 nuclease analysis. Initial S1 protection mapping experiments were carried out (25) with few modifications. Single-stranded DNA recovered from subclones appropriately sub-

cloned into M13mp18 or M13mp19 was used as protector DNA. y^2 w^{bf} [=*su(s)*$^+$] embryonic poly(A)$^+$ RNA (5 to 10 μg) was incubated with 30 ng of single-stranded phage DNA for several hours to allow annealing. After S1 digestion, the samples were phenol extracted, ethanol precipitated, and suspended in loading dye. Electrophoresis was performed on alkaline denaturing gels, and the DNA was blotted to nitrocellulose filters. Probes were ^{32}P labeled by nick translation.

To localize the transcription start site by S1 mapping, an *Eco*RI-*Apa*I fragment from the 5′ end of *su(s)* (nucleotides [nT] 1 to 583 of Fig. 2; the *Eco*RI site is derived from the linker of the λEMBL vector in which the clone was recovered) was subcloned into the *Eco*RI-*Apa*I sites of pBlueScript SK− (Stratagene). Template DNA was prepared by a modified alkaline lysis procedure (2) as described in the Promega 1988–1989 catalog and applications guide. Synthetic transcripts were synthesized using T7 polymerase (Stratagene) as described by the supplier's specifications in the presence of [α-^{32}P]CTP (Amersham; 400 Ci/mmol). S1 nuclease analysis was performed as described elsewhere (6) with 10 μg of y^2 w^{bf} [=*su(s)*$^+$] poly(A)$^+$ adult male or embryonic RNA. S1-resistant fragments were analyzed on 8 M urea–6% acrylamide gels by autoradiography. Sequencing reactions performed as described above were run to provide size markers.

Analysis of RNA expression. RNA for determination of *su(s)* expression throughout the life cycle was prepared according to previously described procedures (5). Embryos were collected from overnight egg depositions of y^2 w^{bf} females, and aliquots were placed in half-pint milk bottles containing cornmeal-molasses-agar medium. At 24-h intervals through day 14, the contents of each bottle were harvested and total nucleic acid was prepared. Poly(A)$^+$ RNA was prepared from this total nucleic acid according to described procedures (5). Northern (RNA) blots were run with 10 μg of poly(A)$^+$ RNA per lane. The *Drosophila ras* gene, which is expressed nearly constantly throughout the life cycle (27), was used as a loading standard. Then an identical Northern blot was run and probed with a probe from within *su(s)* that hybridized only with the *su(s)* message.

Preparation and immunoblotting of *su(s)* protein. Antisera were produced against three separate regions of the open reading frame (ORF). In the preparative and analytic procedures, proteins were separated on 7.5% acrylamide-sodium dodecyl sulfate (SDS) denaturing gels. Blotting to nitrocellulose or Immobilon (Millipore) membranes was carried out by using a BioRad Trans-Blot Cell. Secondary antibody detections utilized the Vectastain ABC kits with either horseradish peroxidase- or alkaline phosphatase-based color indicator systems. Fusion protein 1 was prepared prior to knowledge of the translation reading frame. The *Sma*I-*Hin*cII fragment (nt 4526 to 5452; amino acids ~262 to 578 of Fig. 2) was gel purified, and its ends were randomized by treatment with *Bal* 31S (IBI) and blunted with T4 DNA polymerase (New England BioLabs). It was subcloned into pJG200 (16) into which a *Bam*HI-*Sma*I adaptor had been introduced, and a subclone that exhibited readthrough of the entire insert was selected. The insert was cut out with *Bam*HI and ligated into pWR590 (19), a β-galactosidase fusion protein expression vector. For fusion proteins 2 and 3, the *Bam*HI-*Sal*I (nt 6310 to 6775; amino acids 807 to 963) and *Sal*I-*Bgl*II (nt 6775 to 7247; amino acids 962 to 1121) fragments, respectively, were subcloned into pWR590-1. Fusion proteins of these constructs were expressed and recovered (23). Approximately 400 μg was recovered by cutting out the band from an acrylamide gel that had been stained with 0.25 M KCl. The fusion protein-acrylamide mixture was cut into small pieces and further pulverized by passage back and forth 5 to 10 times between two syringes that were connected by a 13-gauge steel needle. One milliliter of Freund's complete adjuvant (initial immunization) or Freund's incomplete adjuvant (booster injections) was added, and the mixture was passed 10 to 15 times between two syringes connected by a 20-gauge needle. Female New Zealand White rabbits were then immunized by intramuscular injections of 0.5 ml into each hip and an interscapular injection of 1.0 ml. Booster shots were given at approximately 3-week intervals. Blood was drawn and serum was cleared by standard procedures.

To determine whether the antisera contained antibodies directed against the *su(s)* portion of the fusion protein, the DNA inserts from the vectors described above were inserted into pATH vectors (9) with appropriate reading frames. The antisera against fusion proteins 1 and 3 recognized the respective *trpE-su(s)* fusions but not the TrpE protein without an *su(s)* insertion. The antiserum against fusion protein 2 recognized the TrpE protein alone as well as its respective *trpE-su(s)* fusion protein.

In vitro translation of synthetic RNAs made from cDNAs. From cDNA and genomic sequence analysis, the putative translation start was deduced to be the ATG at 2975, which is 7 nt downstream from the *Cla*I site (ATCGAT) at nt 2968. A complete cDNA was assembled and cloned into the *Cla*I site of BlueScript SK− (Stratagene). Synthetic RNAs were produced from both strands according to recommended specifications (Stratagene) and in vitro translated in a rabbit reticulocyte system (Promega) according to recommended procedures, with ^{35}S (Amersham) as the radioactive label. Proteins were analyzed on denaturing SDS-polyacrylamide gels.

Nucleotide sequence accession number. The EMBL, GenBank, and DDBJ accession number for the nucleotide sequence discussed here is M57889.

RESULTS

Sequence of genomic *su(s)* gene. Genetic transformation experiments have demonstrated that an ~8-kb *Hin*dIII fragment (0 to +8.0 in Fig. 1A) is sufficient to complement the suppression, cold-sensitive male sterility, and reduced viability associated with *su(s)* mutations (37). The nucleotide sequence of this fragment plus several hundred base pairs on either side has been determined (Fig. 1B), and the results are shown in Fig. 2. The insertion point of *W20*, the P-element insertion mutation that was used to clone the gene by transposon tagging, was arbitrarily set as the origin for genomic reference (5); that site is 3′ to nt 446 in the DNA sequence shown in Fig. 2 (36).

Location of introns and exons by S1 protection mapping. The direction of transcription of the *su(s)* message is from left to right, as shown in Fig. 1A (5). S1 protection mapping was used to determine the general features of how the 0 to +8.0 genomic region gives rise to the ~5-kb mature message. The 8-kb *Hin*dIII fragment and various of its subregions were used as probes for S1 analysis (Materials and Methods). The results of those studies are summarized in Fig. 1D. First, when the entire ~8.0-kb *Hin*dIII fragment was used as a probe, four protected fragments (exons) of approximately 2.5, 1.35, 0.4, and 0.4 kb were observed (top line of Fig. 1D). Probing with subfragments (Fig. 1C) of the 8-kb region allowed the positioning of the ends of the two

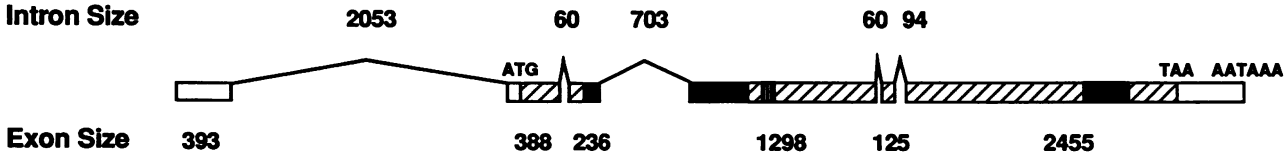


FIG. 3. Structure of mature *su(s)* mRNA. The sizes (in nucleotides) of the introns and exons are as indicated. Symbols: □, 5′ untranslated leader and 3′ untranslated trailer; ▨, translated region; ▩, RRM; ■, highly charged region; and ▥, opa sequence.

largest exons as shown (second to fifth lines of Fig. 1D). Those results indicated that the ~2.5-kb exon was 3′-most and that the ~1.35-kb exon was separated from it by several hundred nucleotides. The use of subfragments of the 0 to +4 interval as probes (Fig. 1C) indicated that the smaller exons arise from that interval, but the determination of the precise origins of the smaller exons came from cDNA sequencing results presented below.

In summary, the S1 protection results indicated the sizes and approximate placement of the ends of the two largest exons. This information allowed determination of the specific locations of the splice junctions by sequencing the ends of these large exons in the cDNAs. In addition, these results indicated that the 0 to +4 genomic interval consists mostly of introns but contains a few exons ≤400 nt long.

Poly(A) addition and splicing pattern as determined from cDNA sequence information. The manner in which the primary transcript is processed to give rise to the mature message is shown in Fig. 1E and 3. Transcription starts at A-415 (Fig. 2; see below also) and produces a primary transcript ~7,860 nt long. Poly(A) addition occurs following the AATAAA signal at nt 8259. The six cDNAs that contained oligo(dT)-primed 3′ ends were polyadenylated following either nt 8277 (four clones) or nt 8279 (two clones).

cDNAs were sequenced completely in the 5′ region of the gene, because S1 mapping data had yielded little information about the origin of the two ~400-nt exons (Fig. 1D and E). The number of different clones that were sequenced across each of the splice junctions are as follows: nt 807 to 2861, four clones; nt 3248 to 3309, five clones; and nt 3544 to 4248, seven clones. Thirteen different cDNA clones were sequenced across the splice between the two largest exons; they indicated the presence of a small 125-nt exon between the two large exons that had not been detected in S1 mapping. For all five splice junctions, all cDNA sequences were identical, indicating that there is probably no alternative splicing involved in the processing of the *su(s)* transcript. Thus, the cDNA sequence results defined six exons (5′ to 3′) of lengths 393, 388, 236, 1,298, 125, and 2,455 nt, separated, respectively, by five introns of lengths 2,053, 60, 703, 60, and 94 nt (Fig. 3 and Table 1). All introns are bounded by 5′ GT donor and 3′ AG acceptor consensus splice junction signals (Table 1). The 507-nt nontranslated leader arises from the splicing of the 2,053-nt intron 114 nt upstream from the probable translation start site (Fig. 2 and 3). The large ORF includes parts of five exons and is followed by an ~420-nt noncoding trailer (Fig. 2 and 3).

Localization of transcription start site. The location of the transcription start site could be bracketed between nt 284 and 459 by two pieces of information. First, the rescue of *su(s)* mutations by the nt 284 to 8375 *Hin*dIII fragment (37) led us to expect the transcription start site to lie downstream of the nt 284 *Hin*dIII site (Fig. 2). Second, the farthest 5′ extension of any cDNA was nt 459. However, because the method of cDNA library construction that creates blunt ends with S1 nuclease results in the loss of 5′ information, it was expected that the transcription start site would lie upstream of nt 459.

Results of S1 protection experiments (not shown) using three different fragments as protectors (nt 1 to 584, 387 to 584, and 1 to 794 in Fig. 2) indicated that the transcription start site lies between nt 415 and 430, although no experiment consistently gave a single distinct protected fragment. We suspect that this lack of a distinct protected fragment is because the region just downstream of the transcription start site (nt 467 to 522) is composed of 85% AT base pairs, and such AT-rich regions are susceptible to S1 nuclease digestion (4).

Primer extension experiments to precisely localize the transcription start site were performed by using two primers (shown in Fig. 2). The results are shown in Fig. 4. Many extensions of primer a (left panel) stopped at A-421 and some

TABLE 1. Intron and exon sizes and splice junctions of *su(s)* transcript

Exon no.	Exon size	5′ Splice junction[a]	Intron no.[b]	Intron size	3′ Splice junction[c]
1	393	TAAGTCAGAT.*gtatgtcaaa*			
			1	2,053	*ttatcaacag*.CTTTTCCAAA
2	388	CCATTAGAAG.*gtgaaagctc*			
			2	60	*tattccctag*.ATGACCATGC
3	236	GGAGCAGCAG.*gtgaggccat*			
			3	703	*tcctttccag*.AACCGTTCCC
4	1,298	GGAAAAGCAG.*gtgagctaat*			
			4	60	*tgttcactag*.CACTCCTCCT
5	125	AACGATCCGC.*gtgagtagat*			
			5	94	*ttcattccag*.TGGACGACGA
6	2,455				

[a] Italics indicate consensus splicing donor signal.
[b] Intron 1 separates exons 1 and 2, etc.
[c] Italics indicate consensus splicing acceptor signal.

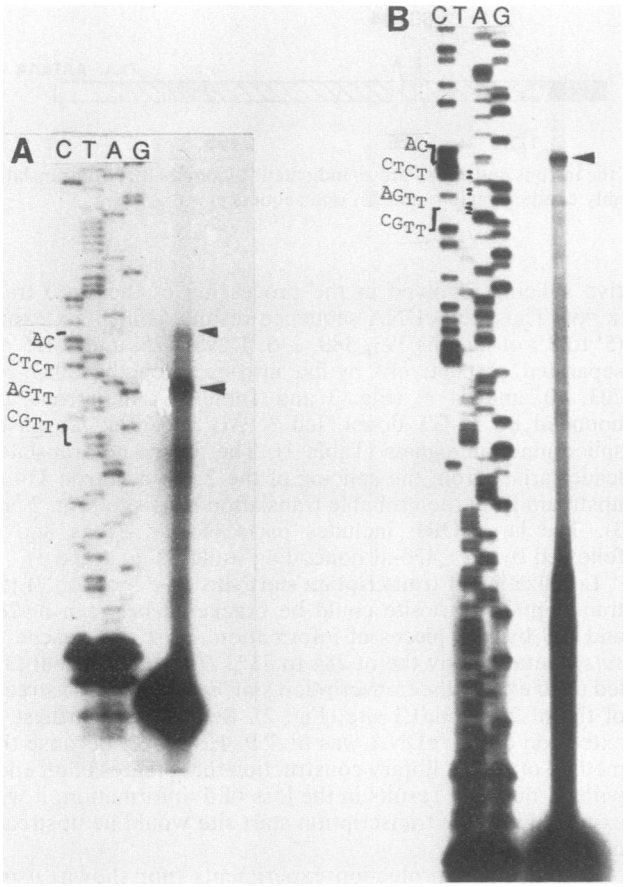


FIG. 4. Localization of transcription start site by primer extension. In both experiments the same primer was used to generate the DNA sequence size standard and to prime the mRNA. Appropriate controls lacking RNA were run (results not shown). (A) Extension of primer a (Fig. 2) yielded an apparent strong stop at A-421, although many molecules extended to A-415 and none extended beyond that point. Identical results were obtained with and without methyl mercuric hydroxide. (B) Extension of primer b (Fig. 2) exhibited a single strong extension stop at A-415, and there was no extension beyond that point. The very faint signals in the T lane of the DNA sequencing ladder are indicated by the hand-drawn hatch marks. That no extension occurred beyond A-415 indicates that it is the transcription start site.

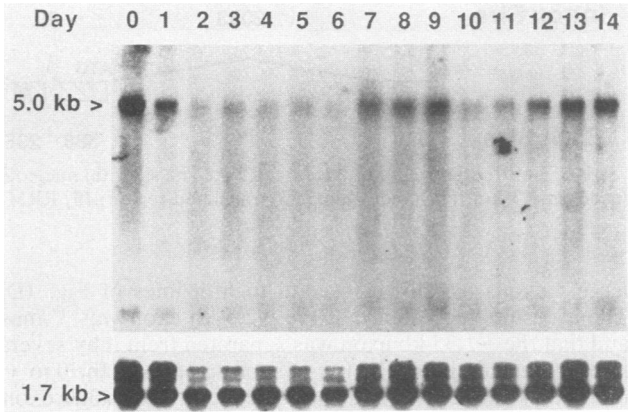


FIG. 5. Profile of wild-type *su(s)* mRNA expression. Poly(A)$^+$ RNAs were prepared from y^2 w^{bf}[=*su(s)*$^+$], as previously described (5). Overnight collections were made and reared at 25°C, and RNA preparations were made at 24-h intervals thereafter until day 14. The collections represent stages of the life cycle as follows: day 0, embryos; days 1 to 5, larvae; days 6 and 7, larvae and pupae; days 8 to 12, pupae; days 13 and 14, adults. Poly(A)$^+$ RNA (5 μg per lane) was loaded on formaldehyde denaturing gels. An identical filter (bottom) was probed with *Drosophila ras* DNA (27) to ensure that comparable amounts of RNA were loaded. Probe 5 (Fig. 1C) was used for the filter shown; identical results were obtained when probes 1, 2, and 3 were used. The faint band at ~1.7 kb is probably residual rRNA in the poly(A)$^+$ RNA.

extended to A-415, but none extended beyond that. Primer b (right panel) gave a single strong signal suggesting A-415 as the start site. Experiments with each primer were carried out at least twice, and the results were consistent. Moreover, the experiments with primer a were done both with and without methyl mercuric hydroxide, which relaxes mRNA secondary structure, and the results were identical.

Thus, the primer extension results suggest that the transcription start is close to A-415. The heptanucleotide start A-415 CCTCTA contains only three matches to the ATCAG/TTC/T *Drosophila* consensus transcription start site (20). However, a better match of the consensus can be produced by beginning at C-417 and deleting T-420, giving C-417TC AGTT, in which case six of the seven nucleotides match. The tetranucleotide AGTT has been observed as a transcription start signal in other *Drosophila* genes: two of the *Drosophila* heat shock genes and the *Drosophila E74* gene (35) begin transcription with AGTT, and a number of *Drosophila* genes contain AGTT as the last 4 nt of the consensus heptanucleotide signal (20).

While no TATA box per se is found ~30 nt upstream of the putative transcription start site, the sequence AAAA ATAT is found 34 nt (nt 379) upstream of the transcription start site; it is preceded and followed, respectively, by three and two GC base pairs. A similar AT-rich sequence nested in a GC-rich region has been identified as the functional TATA equivalent for the *Drosophila E74* gene (35).

Temporal profile of *su(s)* transcription. Northern analysis indicated that the *su(s)* message is produced at all stages of development (Fig. 5). The presence of a single 5-kb message is consistent with the identical splicing pattern observed in all cDNAs. As a loading control, an identically prepared filter was probed with the *Drosophila ras* gene (bottom panel), because *ras* is expressed at nearly constant levels throughout the life cycle (27). Examination of the *ras* panel indicates that the amounts of poly(A)$^+$ RNA loaded varied somewhat from stage to stage. In general, the darker bands in the *su(s)* panel parallel the darker bands in the *ras* panel. Thus the *su(s)* message may be expressed at a nearly constant level throughout the life cycle.

Size and cellular location of *su(s)* protein. The *su(s)* mRNA (Fig. 3) contains an ORF that could encode a protein of 1,322 amino acids (molecular mass of ~145 kDa; pI of 5.30), assuming that translation starts at the first AUG in the ORF (nt 2975 in Fig. 2).

In order to study the *su(s)* protein by immunodetection procedures, polyclonal antisera were produced in rabbits against β-galactosidase fusion proteins containing three different portions of the *su(s)* protein (Fig. 2): protein 1, amino acids 262 to 578; protein 2, amino acids 807 to 963; and protein 3, amino acids 962 to 1121. The abilities of the three antisera to detect the *su(s)* portions of the fusion proteins were verified as described in Materials and Methods.

Synthetic RNA from a cDNA containing the complete

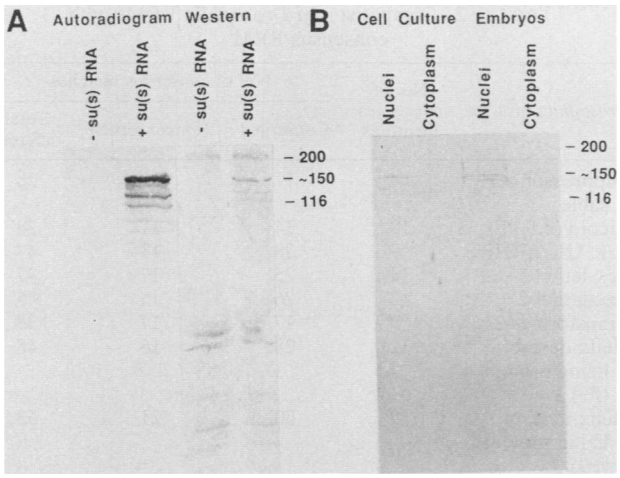


FIG. 6. Identification and localization of *su(s)* protein. (A) In vitro translation and immunodetection of *su(s)* protein. A riboprobe runoff RNA from a complete *su(s)* cDNA was translated in vitro in a rabbit reticulocyte translation system with [^{35}S]methionine as the label. The product was electrophoresed on a denaturing SDS-polyacrylamide gel and blotted to nitrocellulose. The left two lanes show the autoradiogram. The $-su(s)$ RNA lane contains no signal. The $+su(s)$ RNA lane shows a prominent band of molecular weight ~150,000, although several fainter, smaller bands are also visible. A similarly prepared filter was probed with anti-*su(s)* antiserum. The $-su(s)$ RNA control lane again contains no detectable signal. The $+su(s)$ RNA lane exhibits the same array of bands seen in the autoradiogram. These results indicate that the in vitro translation product is in the size range predicted by the ORF and that the translation product is detectable by antisera made against a portion of the *su(s)* ORF. (B) Cellular localization of in vivo-translated *su(s)* protein. Nuclear and cytoplasmic fractions were prepared from S2/M3 cells and from $su(s)^+$ embryos, electrophoresed on SDS-denaturing polyacrylamide gels, blotted to nitrocellulose, and probed with the same anti-*su(s)* antiserum used in the experiments of panel A. The nuclear fraction of the S2/M3 cells contains a band of the same size as detected in the in vitro translation experiments of panel A, but the cytoplasmic fraction does not. The $su(s)^+$ embryo nuclear fraction also contains a prominent band of the same size as the in vitro translation product, and the protein is also present to a lesser extent in the cytoplasmic fraction. Whether the doublet band indicates two different primary translation products or posttranslational modification or degradation is not known.

ORF was translated, and the product was analyzed on a denaturing SDS-polyacrylamide gel. The $+su(s)$ RNA lane of the autoradiogram (Fig. 6A) shows a strong band at ~150 kDa and several smaller bands in the 100- to 120-kDa range. When antiserum 1 was used to probe a protein immunoblot of the same in vitro-translated *su(s)* polypeptide (Fig. 6A), a parallel pattern was observed. Several smaller bands, perhaps resulting from incomplete translation or degradation, were also observed. Antisera 2 and 3 gave similar results (data not shown). The observed size of the protein agrees with the predicted size of ~145 kDa, if translation of the *su(s)* mRNA begins with the first AUG in the ORF.

Cells of the S2/M3 Drosophila cell line were fractionated into nuclear and cytoplasmic components (10), and the proteins were separated on a denaturing SDS-polyacrylamide gel and analyzed by protein immunoblotting. Antiserum 1 detected a protein of the same size as the largest in vitro translation product, and that protein was localized primarily in the nuclear fraction (Fig. 6B).

The same apparent ~150-kDa protein was abundantly detected in a nuclear extract prepared from embryos (0 to 24 h), and traces were also detected in the cytoplasmic fraction (Fig. 6B). The slightly smaller of the ~150-kDa doublet bands in the nuclear fraction is frequently observed, but its origin is unknown. The presence of *su(s)* protein in embryos is consistent with the presence of *su(s)* mRNA during the 0- to 24-h embryo stage.

Thus, polyclonal antisera made against three different regions of the ORF all apparently detect the same ~150-kDa protein, whether synthesized in vitro or recovered from *Drosophila* embryos or tissue culture cells. In vivo, the protein is primarily located in the nucleus.

***su(s)* protein has regions structurally similar to RNA-binding proteins.** The most interesting features of the *su(s)* protein structure are two regions of similarity to proteins that are involved in mRNA processing. One similarity involves an 80- to 90-amino-acid motif known as the ribonucleoprotein particle consensus sequence-type (RNP-CS) RNA-binding domain (1, 11), or RNA recognition motif (RRM) (29), that has been identified in some RNA or single-stranded nucleic acid-binding proteins. In Fig. 7, amino acid residues 1087 to 1162 of the *su(s)* protein are compared with residues in 10 other *Drosophila* RRM-containing proteins and with a consensus RRM that has been derived from a comparison of nearly 40 RRMs (22a, 24, 29, 30).

Qualitatively, the *su(s)* protein contains specific amino acids at strategic locations which implicate this region as some type of RRM. In the consensus octamer, the most highly conserved region of the RRM, the *su(s)* sequence fits the consensus at five of the eight amino acids, and the leucine and isoleucine residues are found at least once in the same respective positions of other octamers. Two highly conserved phenylalanine residues (F-3 and F-45) that have been shown to cross-link to oligodeoxynucleotides (26) when UV irradiated are both present in the *su(s)* protein, as are all the other aromatic residues except at position 48 in the octamer. In general, the fit of the *su(s)* protein to the consensus sequence appears best from the amino-terminal end to the octamer, while the fit to the consensus sequence is poorest at the carboxyl end of the *su(s)* RRM; because the carboxyl end of the *su(s)* RRM is very glycine rich, it is possible that the glycines that fit the consensus sequence are fortuitous.

Because of the great diversity in RRMs, it is difficult to quantitatively assess significant similarity between them. In Table 2 we have attempted to quantify the similarities between the various *Drosophila* RRMs and the consensus RRM by showing the number of identities and conservative amino acid replacements (34). Most RRMs have between 36 and 53 total matches with the consensus sequence. Only the *su(s)* and *bicoid* proteins, with 28 total matches each, have poorer fits. The *su(s)* protein fits the consensus sequence at 17 of the 32 conserved residues, whereas the *bicoid* protein fits the consensus sequence at only 14 conserved residues. On the other hand, the *bicoid* protein has matches at 10 of the 16 residues occupied by specific amino acids compared with only four matches for the *su(s)* protein. Additionally, there are 22 cases in which the amino acid occupying a specific position of the *su(s)* protein is a non-consensus sequence residue but is found at the same position in at least one other RRM (22a, 24, 29). Thus, while the RRM of the *su(s)* protein does not fit the consensus sequence as well as do the RRMs of most other *Drosophila* proteins, it appears to contain similarities at enough critical positions within the motif to include it as a member of the RRM-containing protein family.

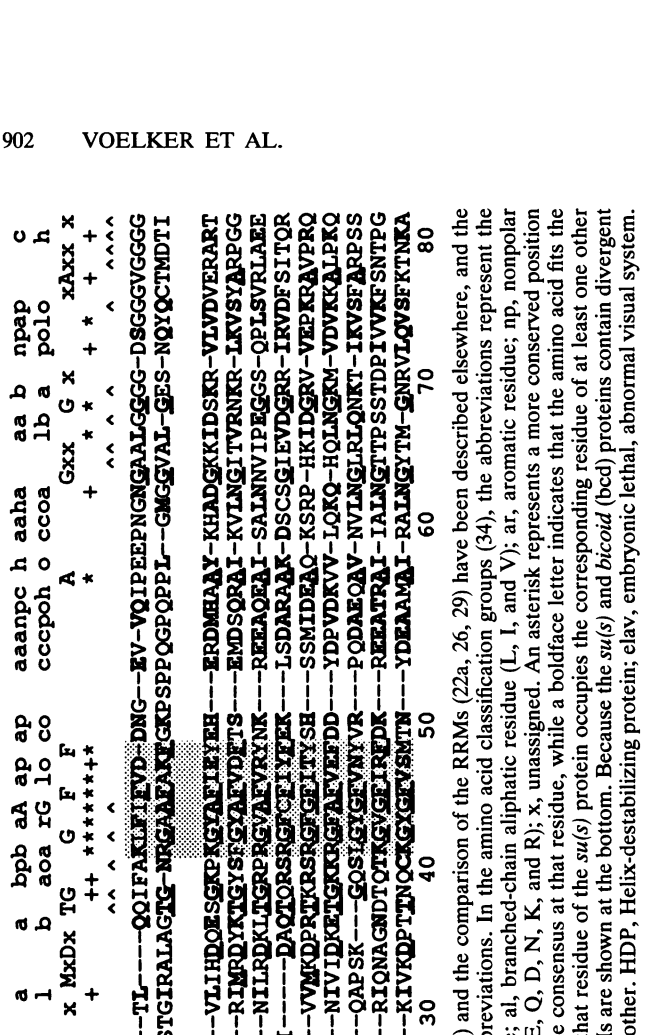


FIG. 7. Comparison of consensus and other RRMs to *su(s)* RRM. The consensus RRM (29) and the comparison of the RRMs (22a, 26, 29) have been described elsewhere, and the *su(s)* protein sequence is from this paper. Capital letters indicate the standard amino acid abbreviations. In the amino acid classification groups (34), the abbreviations represent the following: ac, acidic residue; ba, basic residue; ch, charged residue; ho, hydrophobic residue; al, branched-chain aliphatic residue (L, I, and V); ar, aromatic residue; np, nonpolar residue; po, polar residue; aa, amides or acids (E, Q, D, and N); ab, amides, acids, or bases (E, Q, D, N, K, and R); x, unassigned. An asterisk represents a more conserved position than a plus sign. An underlined boldface letter indicates a match to a specific amino acid in the consensus at that residue, while a boldface letter indicates that the amino acid fits the consensus group at that residue. A caret (^) indicates that the specific amino acid occupying that residue of the *su(s)* protein occupies the corresponding residue of at least one other RRM-containing protein (22a, 24, 29, 30). The 10 *Drosophila* proteins that contain typical RRMs are shown at the bottom. Because the *su(s)* and *bicoid* (bcd) proteins contain divergent RRMs, they are grouped at the top for comparison with the consensus sequence and with each other. HDP, Helix-destabilizing protein; elav, embryonic lethal, abnormal visual system.

TABLE 2. Comparison of *Drosophila* RRMs with consensus RRM

Drosophila protein	No. of consensus identities	No. of consensus matches		
		Conserved	Nonconserved	Total different
Suppressor of sable	4	17	10	28
Bicoid (2.6 kb)	10	14	11	28
70K U1 snRNP	9	28	17	47
Sex-lethal 1	14	25	17	47
Sex-lethal 2	12	26	15	45
Transformer-2	9	17	17	38
Helix-destabilizing protein (P9) 1	12	28	16	48
Helix-destabilizing protein (P9) 2	9	30	21	53
Elav[a] 1	7	19	16	36
Elav 2	9	22	16	40
Elav 3	12	26	11	43

[a] Elav, Embryonic lethal, abnormal visual system.

A second region of the *su(s)* protein is similar to highly charged regions of the human, *Xenopus*, and *Drosophila* 70,000-molecular-weight (70K) U1 small nuclear RNPs (snRNPs) (12, 24, 29), to the *Drosophila* suppressor of white-apricot protein (6), and to the *Drosophila* transformer and transformer-2 proteins (3, 18). To illustrate, the alignment to maximize homology (TFASTA; 28) of the regions of the *su(s)* and human 70K U1 snRNPs is shown in Fig. 8. While only occasionally is there colinearity of identical amino acids, the overall contents of the similar regions are comparable. As seen in Table 3, the identity between the *su(s)* protein and the other proteins ranges between 15.6 and 28.6%. However, when the conservative replacements are added, the total similarity in the regions of comparison ranges between 69 and 77%. The regions of similarity are rich in arginine, lysine, serine, aspartic acid, and glutamic acid. This regional similarity of amino acid content without colinear identity suggests that the regions of similarity might comprise hydrophilic domains of the respective proteins that have generally similar functions.

Because the putative *su(s)* protein is primarily located in the nucleus, its amino acid sequence was examined also for the presence of structural features found in other nuclear proteins. It does not contain regions associated with DNA binding such as a homeobox (15) or zinc fingers (13). Amino acids 446 to 474 appear to comprise an opa or opa-like sequence. The opa sequence, first reported from the *Drosophila* Notch locus (39), consists of a run of ~30 glutamines interspersed with other amino acids. In the *su(s)* protein, a run of 29 amino acids consists of glutamines interspersed with 8 other amino acids (2 each of histidine and leucine and 1 each of aspartic acid, glutamic acid, methionine, and threonine). The functional role of opa remains unknown.

DISCUSSION

The results of this investigation indicate that the *Drosophila* suppressor of sable gene produces a protein of ~150 kDa that is located primarily in the nucleus. The apparent molecular weight of the protein on SDS-polyacrylamide denaturing gels is consistent with the predicted size if translation initiates at the first AUG in the ORF. However,

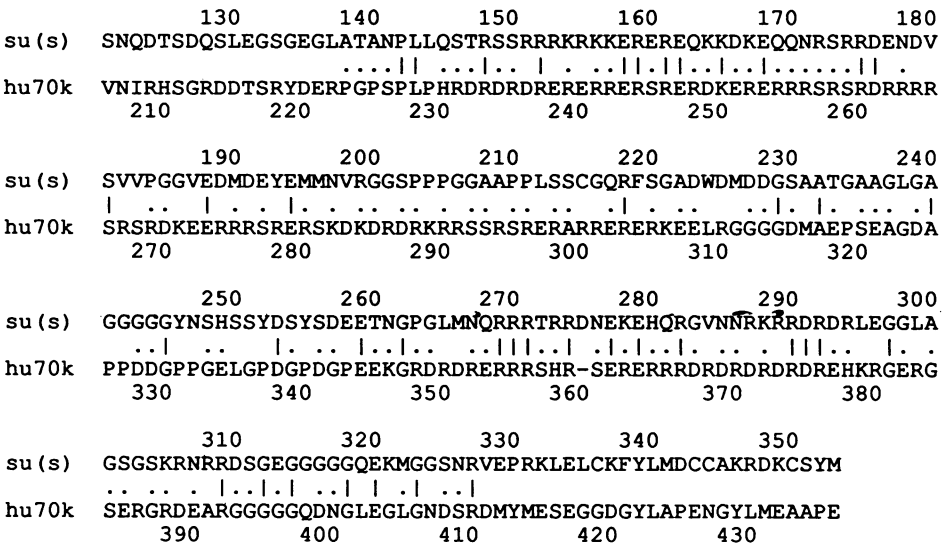


FIG. 8. Similarity of highly charged regions of *su(s)* protein and human 70K U1 snRNP. Comparison of proteins by the FASTP program (28) identified the similarity illustrated. Vertical lines between amino acids indicate identities, and single dots indicate conservative replacements. Within a 190-amino-acid region there is 21.6% identity and 50% conservative substitutions.

it should be noted that the apparent size judged by migration might be distorted, because the highly charged region [i.e., the region of similarity between the *su(s)* and 70K proteins] causes the 70K protein to have a larger apparent molecular weight (70,000) than the molecular weight predicted from the cDNA ORF (54,000) (29).

The *su(s)* protein was detected only in the nuclear fraction of *Drosophila* tissue culture cells. In 0- to 24-h embryos, the protein was abundant in the nuclear fraction, but traces were also observed in the cytoplasmic fraction. Whether this trace in the cytoplasmic fraction is really of cytoplasmic origin or resulted from leakage from the nuclei into the syncytium present in early embryonic stages remains unknown. However, preliminary experiments suggest that the *su(s)* message but not the protein may be present in unfertilized eggs. If this observation is confirmed, it suggests that the cytoplasmic traces could be early translation products that are destined for but have not yet been sequestered in the nuclei.

How is the *su(s)* protein involved in suppression? The present study has shown that the putative *su(s)* protein has two regions of similarity to RNA-binding and -processing proteins: an RRM and a highly charged region. While RRMs have been found in about 40 proteins (22a, 29), the presence of both an RRM and a highly charged region within the same protein has been reported only in human, *Xenopus*, and *Drosophila* 70K U1 snRNPs (24, 29) and in the *Drosophila* transformer-2 protein (18). The 70K U1 snRNP is part of the RNP complex (splicesosome) which splices introns from primary transcripts during transcript processing, and the transformer-2 protein regulates the splicing of transcripts involved in *Drosophila* sex determination. While it has been shown that the RRM of the 70K protein is essential for binding to U1 RNA (29), the substrates of RRMs found in many other proteins remain unknown. The function of the highly charged ~190-amino-acid region is not yet known. It may bind directly to RNA, interact with other proteins, or be a part of a ribonucleoprotein complex. Nevertheless, the existence of these regions of similarity between the *su(s)* protein and other known RNA-processing proteins suggests that the *su(s)* product is involved in some aspect of mRNA metabolism.

TABLE 3. Comparison of *su(s)* protein with other RNA-processing proteins

Protein	Region of comparison (no. of amino acids)	% Identity	% Conservative replacements	% Similarity
Human 70K	190	21.6	50.0	71.6
Xenopus 70K	285	17.9	51.2	69.1
Drosophila 70K	77	28.6	42.8	71.4
Drosophila transformer	66	18.2	59.1	77.3
Drosophila suppressor of white-apricot	109	15.6	56.0	71.6

It might be argued that the similarity between the two regions of the *su(s)* protein and their counterparts in other proteins is insufficient evidence of similarity of function. Two other members of the RRM-containing family of proteins contain divergent forms: the human UP2 protein completely lacks the portion of the RRM carboxyl to the consensus octamer (29), and the Drosophila *bicoid* protein contains a variant form of the RRM and octamer (30). The variation in form of the highly charged region in the different proteins is extensive enough that it is difficult to define a prototype. Until the diverse forms of the RRM and the highly charged region are completely known, it appears reasonable to include the *su(s)* protein as a potential member of the group of proteins containing these domains.

How is suppression effected? Several other observations also implicate the *su(s)* protein in some aspect of mRNA processing or stability. *su(s)* mutations suppress some mobile-element-caused mutations at the following loci: vermilion (*v*), purple (*pr*), speck (*sp*), yellow (*y*), and singed (*sn*). The suppressible *v* alleles are caused by insertions of the retrotransposon 412 (32, 38), and it is probable that the suppressible *pr* and *sp* alleles are also caused by 412 insertions (32). The suppressible alleles at the *y* (16a, 17) and *sn* (32a) loci are caused by P-element insertions. Irrespective of

the mobile-element inserted, the common feature of *su(s)*-suppressible alleles is that the mobile-element insertion occurs in 5′ transcribed but nontranslated regions. For both *v* and *y*, pseudo-wild-type, presumably translatable, messages are produced by splicing the mobile-element sequences from the mutant primary transcripts by using cryptic splice sites within 412 (14) or by using a donor site within the P element and an acceptor site within the P element or just 3′ to the P-element insertion site but still in the 5′ nontranslated exon (16a). Analyses of cDNAs from both the *v* and *y* loci indicate that this splicing occurs in both nonsuppressed [*su(s)*$^+$] and suppressed [*su(s)* mutant] genotypes. However, in *su(s)* mutant flies, higher levels of the pseudo-wild-type message are found. At *y* the levels of both mutant (P-element-containing) and pseudo-wild-type messages coordinately increase in *su(s)* mutant genotypes (the ratio between the mutant and pseudo-wild-type messages does not change) (16a); at *v* the level of pseudo-wild-type message increases, but the effect on the level of mutant (e.g., 412-containing) transcript is unknown (14, 31). At *v* the increase in the level of pseudo-wild-type message from barely detectable levels to 10 to 20% of the wild-type amount (31) parallels the increase in tryptophan oxygenase (the *v* translation product) (33), indicating that the pseudo-wild-type message is probably translatable.

How does the *su(s)* protein affect transcript stability? That *su(s)*-suppressible alleles contain mobile-element insertions in their 5′ transcribed but nontranslated leader sequences suggests that the *su(s)* protein interacts with the transcripts rather than on the process of transcription. Perhaps one function of the *su(s)* protein is to monitor the 5′ leaders of transcripts prior to splicing. If certain leader signals are not found or are interrupted, the message is marked for degradation and does not enter the splicing pool. The wild-type level of wild-type *su(s)* protein accomplishes this monitoring at an optimal, but less than 100%, efficiency. However, the reduced efficiency of either a mutant protein or a reduced amount of wild-type protein allows some of the mutant (mobile-element-containing) transcript to enter the pool of transcripts being spliced. Once the mobile-element sequences are spliced from the transcript, it is no longer detectable by the *su(s)* protein as aberrant and persists to become a pseudo-wild-type message that is presumably translatable. In addition to the predicted effects that mutant *su(s)* alleles might exhibit, this model predicts that higher-than-normal levels of wild-type *su(s)* protein might act to enhance the mutant effect exhibited by suppressible alleles, because the mutant transcripts might be even more efficiently detected and subsequently removed. Such effects have indeed been reported, in that additional doses of *su(s)*$^+$ act as genetic enhancers of suppressible purple mutations (22).

The model described above may also explain the phenotypic effects of *su(s)* mutations. Although *su(s)* is apparently not lethal mutable, mutations caused by base substitution-causing mutagens (ethyl nitrosourea, ethyl methanesulfonate, diepoxybutane, and X and γ rays) are more deleterious than mutations caused by the insertion of mobile elements into transcribed but non-protein-coding regions (36, 37). If the role of the *su(s)* protein is to identify putatively nonfunctional transcription products for degradation, then it might be expected that a reduction in the amount of qualitatively wild-type *su(s)* protein would lead to a decreased efficiency of the transcript processing and translation processes. While this decreased efficiency might be somewhat deleterious, it probably would not be strongly deleterious: a slight reduction in viability is in fact observed in those mutations caused by mobile-element insertions into non-protein-coding regions. On the other hand, it might be expected that mutant *su(s)* proteins produced by base substitutions in protein-coding sequences would sometimes function aberrantly, perhaps even removing some normal transcripts; this could significantly reduce viability, as has been observed in some *su(s)* alleles induced by base substitution-causing mutagens (21, 37). We are currently examining *su(s)* mutations induced by mutagens that cause primarily base substitutions (ethyl nitrosourea, ethyl methanesulfonate, and irradiation) to identify domains in which critical amino acid substitutions occur.

Determining whether the model described above is correct and whether the *su(s)* protein has a role in the processing of at least some of the myriad of primary transcripts that are not interrupted by mobile-element insertions is the goal of future investigations.

ACKNOWLEDGMENTS

The contributions of the following undergraduate students who assisted with various phases of the project were much appreciated: Terri Maness, Scott Carpenter, Inga Oleksy, Sharon Lingle, Michael Murphy, and Sandy Volrath. The assistance of Bill Quattlebaum in use of the computer in DNA sequence analysis is gratefully acknowledged. We also thank Burke Judd, Lillie Searles, and Michael Simmons for their comments on, suggestions for, and criticisms of the manuscript.

REFERENCES

1. **Bandziulis, R., M. Swanson, and G. Dreyfuss.** 1989. RNA-binding proteins as developmental regulators. Genes Dev. **3**:431–437.
2. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1522.
3. **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.
4. **Calzone, F. J., R. J. Britten, and E. H. Davidson.** 1987. Mapping of gene transcripts by nuclease protection assays and cDNA primer extension. Methods Enzymol. **152**:611–632.
5. **Chang, D.-Y., B. Wisely, S.-M. Huang, and R. A. Voelker.** 1986. Molecular cloning of *suppressor of sable*, a *Drosophila melanogaster* transposon-mediated suppressor. Mol. Cell. Biol. **6**:1520–1528.
6. **Chou, T.-B., Z. Zachar, and P. M. Bingham.** 1987. Developmental expression of a regulatory gene is programmed at the level of splicing. EMBO J. **6**:4095–4104.
7. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. Plasmid **13**:31–40.
8. **Davis, L. D., M. D. Dibner, and F. Battey.** 1986. Basic methods in molecular biology. Elsevier Science Publishing, Inc., New York.
9. **Dieckmann, C. L., and A. Tzagologg.** 1985. Assembly of the mitochondrial membrane system. J. Biol. Chem. **260**:1513–1520.
10. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11**:1475–1489.
11. **Dreyfuss, G., M. S. Swanson, and S. Pinol-Roma.** 1988. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. Trends Biochem. Sci. **13**:86–91.
12. **Etzerodt, M., R. Vignali, G. Ciliberto, D. Scherly, I. W. Mattaj, and L. Philipson.** 1988. Structure and expression of a *Xenopus* gene encoding an snRNP protein (U1 70K). EMBO J. **7**:4311–4321.

13. **Evans, R. M., and S. M. Hollenberg.** 1988. Zinc fingers: gilt by association. Cell **52**:1–3.
14. **Fridell, R. A., A.-M. Pret, and L. L. Searles.** 1990. A retrotransposon *412* insertion within an exon of the *Drosophila melanogaster vermilion* gene is spliced from the precursor. Genes Dev. **4**:559–566.
15. **Gehring, W. J., and Y. Hiromi.** 1986. Homeotic genes and the homeobox. Annu. Rev. Genet. **20**:147–173.
16. **Germino, J., and D. Bastia.** 1984. Rapid purification of a cloned product by genetic fusion and site specific proteolysis. Proc. Natl. Acad. Sci. USA **81**:4692–4696.
16a. **Geyer, P., and V. Corces.** Personal communication.
17. **Geyer, P. K., K. L. Richardson, V. G. Corces, and M. M. Green.** 1988. Genetic instability in Drosophila melanogaster: *P element* mutagenesis by gene conversion. Proc. Natl. Acad. Sci. USA **85**:6455–6459.
18. **Goralski, T. J., J.-E. Edström, and B. S. Baker.** 1989. The sex determination locus *transformer-2* of Drosophila encodes a polypeptide with similarity to RNA binding proteins. Cell **56**: 1011–1018.
19. **Guo, L.-H., P. P. Stepien, J. Y. Tso, R. Brousseau, S. Narang, D. Thomas, and R. Wu.** 1984. Synthesis of human insulin gene. VIII. Construction of expression vectors for fused proinsulin production in *Escherichia coli*. Gene **29**:251–254.
20. **Hultmark, D., R. L. Klemenz, and W. J. Gehring.** 1986. Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. Cell **44**:429–438.
21. **Jacobson, K. B., E. H. Grell, J. J. Yim, and A. L. Gardner.** 1982. Mechanism of suppression in *Drosophila melanogaster*. VIII. Comparison of *su(s)* alleles for ability to suppress the mutants *purple*, *vermilion*, and *speck*. Genet. Res. **40**:19–32.
22. **Jacobson, K. B., J. J. Yim, E. H. Grell, and C. R. Wobbe.** 1982. Mechanism of suppression in Drosophila: evidence for a macromolecule produced by the $su(s)^+$ locus that inhibits sepiapterin synthase. Cell **30**:817–823.
22a. **Keene, J.** Personal communication.
23. **Mahaffey, J. W., and T. C. Kaufman.** 1987. Distribution of the *Sex combs reduced* gene products in *Drosophila melanogaster*. Genetics **117**:51–60.
24. **Mancebo, R., P. Lo, and S. M. Mount.** 1990. Structure and expression of the *Drosophila melanogaster* gene for the U1 small nuclear ribonucleoprotein particle 70K protein. Mol. Cell. Biol. **10**:2492–2502.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Merrill, B. M., K. L. Stone, F. Cobianchi, S. H. Wilson, and K. R. Williams.** 1988. Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 heterogeneous nuclear ribonucleoprotein. J. Biol. Chem. **263**:3307–3313.
27. **Mozer, B., R. Marlor, S. Parkhurst, and V. Corces.** 1985. Characterization and developmental expression of a *Drosophila ras* oncogene. Mol. Cell. Biol. **5**:885–889.
28. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
29. **Query, C. C., R. C. Bentley, and J. D. Keene.** 1989. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70k U1 snRNP protein. Cell **57**:89–101.
30. **Rebagliati, M.** 1989. An RNA recognition motif in the *bicoid* protein. Cell **58**:231–232.
31. **Searles, L. L., R. S. Ruth, A.-M. Pret, R. A. Fridell, and A. J. Ali.** Structure and transcription of the *Drosophila melanogaster* vermilion gene and several mutant alleles. Mol. Cell. Biol. **10**:1423–1431. in press
32. **Searles, L. L., and R. A. Voelker.** 1986. Molecular characterization of the *vermilion* locus and its suppressible alleles. Proc. Natl. Acad. Sci. USA **83**:404–408.
32a. **Simmons, M., and K. O'Hare.** Personal communication.
33. **Tartof, K.** 1969. Interacting gene systems. I. The regulation of tryptophan pyrrolase by the *vermilion-suppressor of vermilion* system in Drosophila. Genetics **62**:781–795.
34. **Taylor, W. R.** 1987. Protein structure prediction, p. 313. *In* M. J. Bishop and C. J. Rawlings (ed.), Nucleic acid and protein sequence analysis: a practical approach. IRL Press, Washington, D.C.
35. **Thummel, C. S.** 1989. The Drosophila *E74* promoter contains essential sequences downstream from the start site of transcription. Genes Dev. **3**:782–792.
36. **Voelker, R. A., J. Graves, W. Gibson, and M. Eisenberg.** 1990. Mobile element insertions causing mutations in the *Drosophila suppressor of sable* locus occur in DNase I hypersensitive subregions of 5′-transcribed nontranslated sequences. Genetics **126**:1071–1082.
37. **Voelker, R. A., S.-M. Huang, G. B. Wisely, J. F. Sterling, S. P. Bainbridge, and K. Hiraizumi.** 1989. Molecular and genetic organization of the *suppressor of sable* and *M(1)1B* region of *Drosophila melanogaster*. Genetics **122**:625–642.
38. **Walker, A. R., A. J. Howells, and R. G. Tearle.** 1986. Cloning and characterization of the *vermilion* gene of *Drosophila melanogaster*. Mol. Gen. Genet. **202**:102–107.
39. **Wharton, K. A., B. Yedvobnick, V. Finnerty, and S. Artavanis-Tsakonas.** 1985. opa: a novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in *D. melanogaster*. Cell **40**:55–62.