Cell, Vol. 29, 1041–1051, July 1982, Copyright © 1982 by MIT

An Expandable Gene That Encodes a Drosophila Glue Protein Is Not Expressed in Variants Lacking Remote Upstream Sequences

Marc A. T. Muskavitch* and David S. Hogness†
Department of Biochemistry
Stanford University Medical Center
Stanford, California 94305

Summary

The Drosophila melanogaster gene *Sgs4* encodes one of the glue polypeptides, sgs-4, synthesized in the larval salivary gland. We have examined the structure and expression of *Sgs4* in five strains that produce abundant amounts of sgs-4 and its mRNA and in four that do not. The nonproducers include three Japanese strains that accumulate trace amounts of mRNA and one strain, BER-1, that contains no detectable *Sgs4* RNA. *Sgs4* carries a tandem array of repeated 21 bp elements within its coding sequence. The number of elements per array varies, causing considerable differences in the lengths of *Sgs4* and its mRNA among the strains. These differences in length are not correlated with differences in mRNA abundance; rather, the low or zero abundance in nonproducers correlates with the loss of DNA upstream from the gene. The Japanese nonproducers carry a 52 bp deletion 305 bp upstream from the 5′ end of *Sgs4,* and BER-1 carries a 95 bp deletion 392 bp upstream. Curiously, each deletion encompasses one or more of the salivary-gland-specific DNAase I–hypersensitive sites which are known to flank the *Sgs4* gene.

Introduction

The salivary glands of Drosophila melanogaster larvae synthesize large amounts of a small set of polypeptides that are secreted at the end of larval life to form a glue which attaches the pupal case to dry surfaces during metamorphosis to the adult (Korge, 1975, 1977a; Beckendorf and Kafatos, 1976; reviewed by Berendes and Ashburner, 1978). The genes encoding several of these glue polypeptides have been genetically mapped to chromosomal positions occupied by members of the set of "intermolt" puffs that are active in the salivary gland polytene chromosomes during glue polypeptide synthesis (Korge, 1975, 1977b; Akam et al., 1978; Velissariou and Ashburner, 1980).

We have previously reported on the molecular organization of the intermolt puff site at 3C11-12 in the X chromosome, and on the *Sgs4* glue polypeptide gene whose transcription generates that puff (Muskavitch and Hogness, 1980). Three results from this initial characterization of *Sgs4* provide the basis for the experiments reported here. Two derive from an examination of the *Sgs4* mRNA in nine D. melanogaster strains, five of which produce an abundance of the sgs4 glue polypeptide encoded by *Sgs4* (producers), and four of which do not (nonproducers). First, the relative amounts of *Sgs4* mRNA closely parallel the relative amounts of sgs4 in the different strains; second, the length of the *Sgs4* mRNA is strain-specific, ranging from 0.95 kb to 2.7 kb. Finally, restriction site mapping of the cloned *Sgs4* gene from Oregon-R (our reference producer strain) suggested that almost half of this gene consists of a tandemly repeated unit of approximately 21 base pairs.

We speculated that the variation in mRNA length resulted from a gene that could expand or contract by changing the number of repeat units within such a tandem array (Muskavitch and Hogness, 1980). We show that *Sgs4* does contain tandemly repeated units, each coding for a sequence of seven amino acids, and that the number of these units per array varies among the strains so as to account for the different lengths of their mRNAs.

By contrast, no correlation is evident between the lengths and the amounts of the mRNAs in the different strains, suggesting that the differences in amount do not result from differences in the stability of the mRNAs arising from differences in their structure. This, and our inability to detect longer *Sgs4* transcripts that might serve as precursors to the mRNAs, led us to suppose that differences in the amount of mRNA (and hence in the amount of sgs4) among the strains are determined primarily by differences in the rate of transcription of *Sgs4* (Muskavitch and Hogness, 1980). An examination of the DNA sequences flanking the 5′ end of the *Sgs4* gene in both producers and nonproducers showed that the nonproducers carry small deletions 300 to 400 base pairs upstream from the mRNA start site. This finding is made all the more remarkable by the fact that the deletions encompass salivary-gland-specific DNAase I–hypersensitive sites (Shermoen and Beckendorf, 1982).

Results

The Boundaries of the mRNA-Homologous Sequences in *Sgs4*

We previously isolated a set of overlapping cloned DNA segments that define a 60 kb region centered on the *Sgs4* gene in the 3C11-12 intermolt puff site (Muskavitch and Hogness, 1980). Figure 1 shows a restriction map for Oregon-R DNA that was abstracted from the previous studies to include only that part of the region and those restriction endonucleases that are relevant to the present experiments. Our previous studies also determined that the DNA sequences homologous to the 0.95 kb *Sgs4* mRNA of Oregon-R are confined to the D and E Hind III fragments given in Figure 1a, and localized the 5′ boundary of the mRNA-homologous region to within ~50 bp of the Eco RI site shown in Figure 1b.

* Present address: The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138.
† To whom reprint requests should be addressed.

This mRNA-homologous region has been further defined by S1 protection experiments (Berk and Sharp, 1977) and by reverse transcription of the *Sgs4* mRNA from specific cloned DNA primers. Figure 2a shows the results of an S1 protection experiment defining the mRNA start point. The 0.31 kb Mbo II–Dde I fragment encompassing the putative start point at or near the Eco RI site was 5′-end-labeled at the Dde I site and hybridized with poly(A)$^+$ RNA from Oregon-R third instar larvae in the intermolt puff phase. Removal of single-stranded regions from the hybridization products by S1 nuclease digestion, fol-

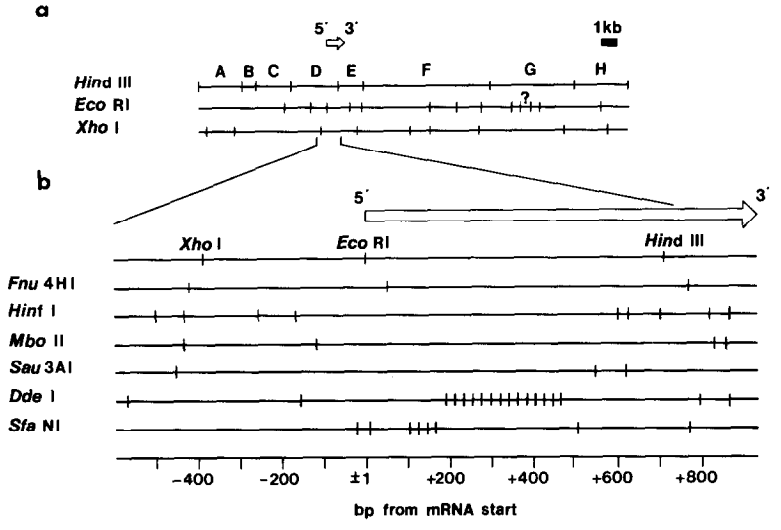

Figure 1. Restriction Map of the Chromosomal DNA in the *Sgs4* Region

The maps in (a) and (b) are abstracted from Figures 1 and 2, respectively, of Muskavitch and Hogness (1980), except that the boundaries of the mRNA-homologous region indicated by open arrows were determined as described in Results. The leftmost of the two Eco RI sites within the Hind III fragment D is present in Canton-S but not Oregon-R; otherwise, the sites given in (a) apply to both strains. The question mark indicates that an Eco RI site is present in one of the two locations to either side of it. The maps in (b) are of the Oregon-R chromosomal DNA. The number of Dde I sites in the +100 to +500 bp region was determined by subtracting the 197 bp length of the Eco RI–Dde I fragment, and the 228 bp length of the Dde I–Hind III fragment, from the 700 bp length of the Eco RI–Hind III fragment, dividing the result by the 21 bp length of Dde I fragments from the region and adding 1. The resulting value of 14 Dde I sites is accurate to only ±1 because the Eco RI–Hind III fragment length was determined by gel electrophoresis rather than from the nucleotide sequences used to determine the other lengths (Figure 3).

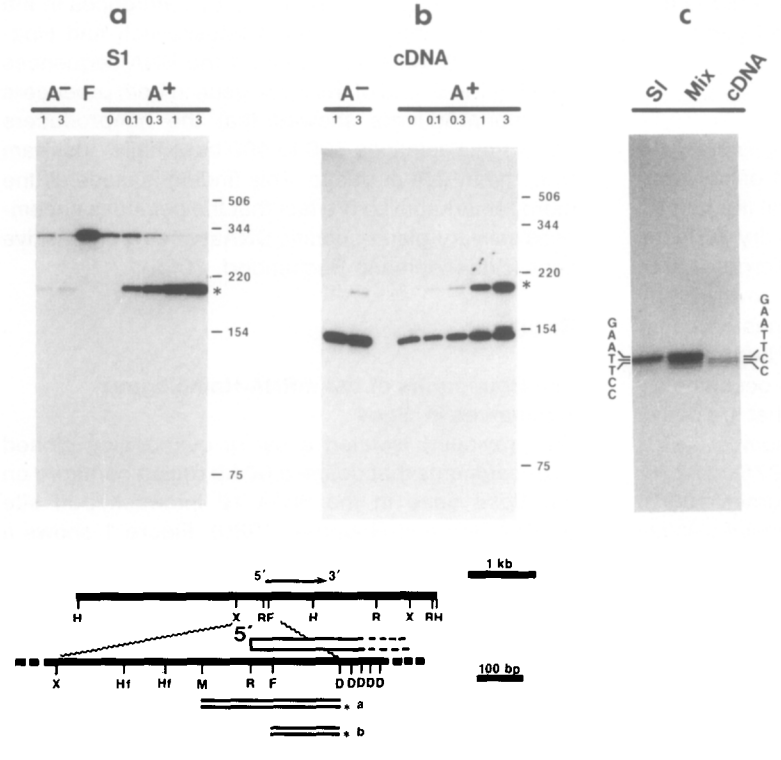

Figure 2. Definition of the 5′ End of the mRNA-Homologous Region by S1 Protection and cDNA Elongation Experiments

See Results for general description of these experiments and Experimental Procedures for other details. The end-labeled fragments used for the S1 protection and cDNA elongation experiments are designated a and b, respectively, in the diagram at the bottom of the figure, where the restriction site designations are: D, DdeI; F, Fnu 4HI; H, Hind III; Hf, HInf I; M, Mbo I; R, Eco RI; and X, Xho I. The autoradiograms given in (a) and (b) result from the labeled ends of fragments a and b, respectively. The amount of poly(A)$^-$ and poly(A)$^+$ RNAs used in each hybridization experiment is given (in micrograms) by the numbers under A$^-$ and A$^+$, respectively, at the tops of the autoradiograms, where F denotes a control consisting of fragment a run alone and without S1 treatment to reveal any fragments arising solely from this probe. Numbers to the right of the (a) and (b) autoradiograms give the positions of length standards (in bases) consisting of end-labeled Hinf I fragments of pBR322 (Sutcliffe, 1978) run in a parallel lane. The asterisk just below the 220 base marker indicates the position of the 196-base strand from the Eco RI–Dde I fragment that overlaps fragments a and b, and was also 5′-end-labeled with T4 polynucleotide kinase at the Dde I site. Electrophoresis was on 8% acrylamide and bisacrylamide (30:0.8) and 8 M urea gels in 40 mM Tris-acetate (pH 8.0), 2 mM EDTA. Autoradiogram (c) shows the S1 and cDNA products from (a) and (b) run separately or together (Mix) in a thin sequencing gel (Sanger and Coulson, 1977), and is described in the text.

lowed by denaturation and electrophoresis in a denaturing gel, yielded the two types of autoradiographic bands seen in the figure. One type increases in intensity as the amount of poly(A)$^+$ RNA is increased and represents that part of the end-labeled DNA strand that is protected from digestion by RNA hybridization. It is 196 nucleotides long and places the mRNA start point within the Eco RI site; this protected strand comigrates with that from the Eco RI–Dde I fragment (data not shown). The other type decreases in intensity with increasing poly(A)$^+$ RNA and represents the Mbo II–Dde I fragment probe, which is protected by reassociation of the complementary DNA strands. The virtual absence of the 196 nucleotide product from the poly(A)$^-$ RNA control is expected, given that this fraction contains little or no *Sgs4* RNA (Muskavitch and Hogness, 1980).

Similar S1 protection experiments were carried out with probes consisting of either the Hind III fragment D, which contains the mRNA start point, or the adjacent Hind III–Eco RI fragment, which contains the 3′ boundary of the mRNA-homologous region, both labeled at the Hind III site they share. The first yielded a 0.70 kb protected product, the second a 0.22 kb product (data not shown). These results place the mRNA start point at the Eco RI site shown in Figure 1b and demonstrate that the mRNA-homologous region in Oregon-R extends without interruption 0.92 kb downstream to its right—a length that is consistent with the 0.95 kb length of the Oregon-R mRNA, given its polyadenylation.

The above experiments demonstrate that *Sgs4* either lacks introns or restricts them to positions within a few dozen base pairs of its ends. Figure 2b shows the results obtained from a reverse transcription experiment designed to test for such a hypothetical intron near the 5′ end. In this case, the *Sgs4* mRNA was hybridized with the complementary, 5′-end-labeled DNA strand from the Fnu 4HI–Dde I fragment shown in Figure 2 (fragment b); the 3′ end of the DNA strand in the resulting hybrid was used as a primer for reverse transcriptase, and the DNA strand thereby extended to the 5′ end of the RNA template (Experimental Procedures). The length of this cDNA should therefore be the same as that of the S1-protected fragment in Figure 2a if there is no terminal intron, and should be greater if there is. Figures 2b and 2c show that the lengths of this cDNA and of the protected fragment differ by only a few nucleotides—too little to allow for an intron at the 5′ end.

Figure 2c shows the results obtained when the products from the S1 protection experiments of Figure 2a and the cDNA elongation experiments of Figure 2b were examined by electrophoresis in a sequencing gel and compared with an adjacent sequence ladder (not shown) originating from the same Dde I site used to end-label the S1 and cDNA products (Experimental Procedures). The S1 protection product yields a single major band at a position corresponding to the first T of the Eco RI sequence (5′-GAATTC-3′) at the mRNA start point in the mRNA-equivalent strand. This T is marked on the nucleotide sequence of the 5′ half of the *Sgs4* gene given in Figure 3, and is assigned the +1 position, which we take to be the mRNA start point. The two faint bands observed to either side of the major S1 product correspond to the A and T at positions −1 and +2. The cDNA elongation product yields two major bands corresponding to the T and C at positions +2 and +3, and a faint band corresponding to the C at +4. Given that we know neither why two bands are obtained nor why they differ by one and two nucleotides from the S1 product, there is an element of arbitrariness in which nucleotide among the three at +1 and +3 is assigned the mRNA start point.

Sgs4 Contains a Tandemly Repeated Sequence

Figure 3 shows the nucleotide sequence of the 5′ half of *Sgs4* in Oregon-R and in two other strains: Canton-S, another producer, and Hikone-R, one of the four nonproducers. The initial methionine codon begins 13 bp from the mRNA start point and is followed by 21 codons that yield the same sequence of amino acids for each strain. These codons are followed by a tandem array of repeated elements, each coding for a sequence of seven amino acids. The nucleotide sequence of each element in an array has not been determined; rather, we have determined the sequences of seven or eight contiguous elements at each end of the array in each of the three strains. The canonical nucleotide sequence for the 46 elements examined is:

ACA TGC AAA ACT GAG CCA CCC.

Of the two possible reading frames (the third being eliminated by a TGA termination codon), only that indicated above is in phase with an upstream methionine codon. In this frame the canonical element translates to:

Thr–Cys–Lys–Thr–Glu–Pro–Pro.

We have defined the element by the circular permutation that is determined by the 3′ rather than the 5′ end of the array, because deviations from canonical sequence are concentrated at the 5′ end. Thus each array is divisible into a large 3′ subset, including elements that are either canonical or differ only by a neutral Arg-for-Lys replacement at the third position, and a small 5′ subset, in which every element contains at least one significant replacement. This division can be illustrated with the Oregon-R array, where the 3′ subset extends from position 6 to the position designated N, inclusive. Figure 3 shows that each of the ten elements of known sequence in this subset contains a Dde I site that is absent from all other elements of the Oregon-R array shown in the figure. Figure 1

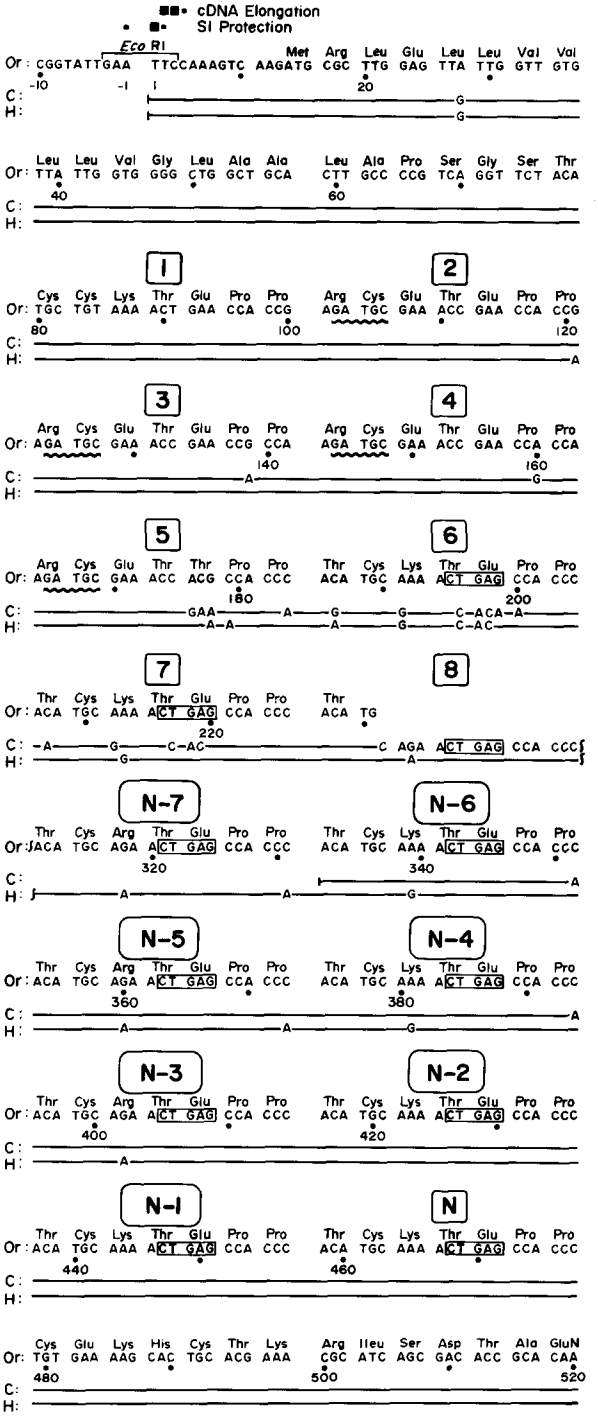

Figure 3. Sequences Downstream from the 5′ End of *Sgs4* in Oregon-R, Canton-S and Hikone-R

The nucleotides in the 5′-to-3′ strand of Oregon-R are given opposite Or, and the amino acids are those translated from this sequence. The Canton-S sequences are given in the lines opposite the C, and the Hikone-R sequences are given opposite H; a continuous line indicates that the sequence is identical with that in Oregon-R. Differences from the Oregon-R sequence are designated by the bases that interrupt these lines except where the Oregon-R sequence has not been determined (+232 to +247), in which case the Canton-S nucleotide sequence is given. Where no line or sequence is given, the sequence has not been determined with sufficient reliability to warrant publication (Experimental Procedures). Positions corresponding to the major products of the S1 protection and cDNA elongation experiments of Figure 2 are indicated by large black squares, minor products by small squares. The N repeated elements in a tandem array are identified by the enclosed numbers that indicate the position of the element within the array. N = 19 for the Oregon-R array (see text), and the numbers for the nucleotide positions in and downstream from the N-7 element (that is, ≧+311) apply only to Oregon-R; in Canton-S N is 21, and in Hikone-R N is 31 (Table 1), from which numbers the appropriate nucleotide position numbers can be calculated. All of the elements in the Oregon-R array except the first carry either a Sfa NI site (wavy underline) or a Dde I site (boxed), as is also indicated in Figure 1.

shows that the Oregon-R array contains a total of 14 Dde I elements. Evidently, the four Dde I elements whose sequences have not been determined occupy positions 8 through 11, and the element occupying position 12 corresponds to N-7; hence N = 19 in Oregon-R. We assign all of the 14 contiguous Dde I elements to the 3′ subset, assuming that the elements in positions 8 through 11 exhibit the canonical, or essentially conincal, structures defined for the other 10.

The 5′ subset exhibits the curious characteristic that the number of significant replacements per element decreases upstream from the boundary between the two subsets. Thus, in Oregon-R, the element in position 5 contains three replacements (Arg-for-Thr at the first position, Glu-for-Lys at the third position and Thr-for-Glu at the fifth position). The elements at positions 2 through 4 each contain the first two of these replacements, and the first element contains only a Cys-for-Thr replacement in the first position. The arrays in the other two strains are similarly organized, except that the 5′ subset contains 7 elements in Canton-S and 6 in Hikone-R, and the 3′ subset contains 14 elements in Canton-S and 25 in Hikone-R (Figure 3 and next section).

The Length of *Sgs4* and Its mRNA Varies According to the Number of Tandemly Repeated Elements

We have used restriction site mapping of genomic DNA to examine *Sgs4* and its flanking sequences in nine different strains in a search for structural variations that correlate with the differences in the lengths and amounts of *Sgs4* mRNA that have been observed among these strains (Muskavitch and Hogness, 1980). Five of the strains are producers that can be divided into two groups: one group, including Oregon-R, Chieti V and Urbana-S, accumulates as much mRNA as the reference Oregon-R strain; the other, including Canton-S and D323, accumulates ~40% as much. Similarly, the four nonproducers can be divided into those that accumulate trace amounts of *Sgs4* RNA (Seto, Hikone-R and Kochi-R) and BER-1, in which such RNA is not detectable (Table 1).

Figure 4 shows that the length of the Hind III frag-

Table 1. Length Correlations between the *Sgs4* mRNA and the Eco RI–Hind III Fragment Containing the *Sgs4* Tandem Array

Strain[a]	(Length$_i$ − Length$_{Or}$)[b] mRNA (bases)	Fragment (base pairs)	Elements per Array[c]
Producers			
High (100%)			
Oregon-R	0	0	19
Chieti V	50	70	22
Urbana-S	130	160	27
Low (~40%)			
Canton-S	30	40	21
D323	330	320	34
Nonproducers			
Trace (~2%)			
Seto	150	190	28
Hikone-R	280	260	31
Kochi-R	1780	1730	101
Zero (<0.2%)			
BER-1	ND	60	22

[a] The percentages given in parentheses are the amount of mRNA found in the indicated strains relative to that found in Oregon-R (Muskavitch and Hogness, 1980).

[b] Values are the length of the mRNA or the Eco RI–Hind III fragment in the indicated strain (length$_i$) minus that in Oregon-R (length$_{Or}$), the Oregon-R lengths being 950 bases and 700 base pairs, respectively. mRNA lengths are the average of four or five determinations carried out as described by Muskavitch and Hogness (1980). Fragment lengths are the average of two determinations carried out as in Figure 5c, with λ (Philippsen et al., 1978) and pBR322 (Sutcliffe, 1978) length standards.

[c] Calculated by dividing the value in the preceding column by 21 and adding the result to 19, the number of elements in the Oregon-R array (see text).

ND: not determined; strain BER-1 produces no detectable transcript.

ment D, which contains the 5′ end and tandem array of *Sgs4*, is different in each strain, whereas the length of the adjacent Hind III fragment E, which carries the 3′ end, is the same in all nine strains. Figure 5c shows that the length of the Eco RI–Hind III fragment, which extends from the 5′ end through the tandem array to the Hind III site common to fragments D and E, is highly variable among the nine strains. Table 1 gives these lengths as the difference between the fragment length for a given strain and that for Oregon-R, which, at 700 bp, is the shortest. Table 1 also shows the difference in *Sgs4* mRNA lengths between a given strain and Oregon-R, whose 950 base mRNA is also the shortest. The differences in fragment lengths among the strains correlate strongly with the differences in mRNA lengths, indicating that strain variation in fragment and in mRNA lengths derive from the same source, which we think is a variation among the strains in the number of elements per tandem array.

This is certainly the case for differences in the lengths of the Eco RI–Hind III fragments from Oregon-R, Canton-S and Hikone-R. Sequence analysis shows that each strain carries the same number of nucleotide residues in the region between the Eco RI site and the

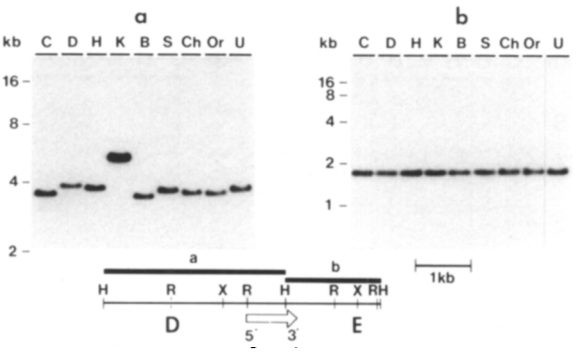

Figure 4. Interstrain Variation in Hind III Fragments D and E

Equal amounts (2 μg) of total adult DNA from each of nine strains were digested with Hind III, electrophoretically fractionated in agarose gels of 0.7% for (a) and 1.0% for (b), blot-transferred to nitrocellulose and hybridized with ^{32}P-labeled Oregon-R Hind III fragments D and E (indicated in the diagram by black bars a and b, respectively), with results shown in autoradiograms (a) and (b), respectively. Length standards, in kilobases, are shown to the left of each autoradiogram, and strains used for each track are indicated at its top, where C designates Canton-S; D, D323; H, Hikone-R; K, Kochi-R; B, BER-1; S, Seto; Ch, Chieti V; Or, Oregon-R; and U, Urbana-S. Restriction site designations are the same as in Figure 2.

5′ end of the array, and in the region between the 3′ end of the array and the Hind III site; indeed, the nucleotide sequence of the 299 residues in the two regions is conserved among the three strains with 98% identity (Figure 3 and Discussion). Differences in the lengths of the Eco RI–Hind III fragments among the three strains must, then, result from differences in the number of elements per array.

Evidence for the generality of this conclusion derives from the analysis of Dde I digests of the *Sgs4* Eco RI–Hind III fragment from cloned segments of eight of the nine strains, Kochi-R not having been analyzed. Gel electrophoresis of each digest yields three bands, two of which are invariant and correspond to the 21 bp Dde I fragments from contiguous elements containing Dde I sites (3′ subset) and to the 228 bp fragment extending from the Dde I site in the Nth element to the Hind III site. All eight strains are therefore invariant with respect to the number of *Sgs4* nucleotides located downstream from the tandem array. The third band corresponds to the DNA extending from the Eco RI site at the 5′ end of the gene to the first Dde I site in the array, and should therefore include the elements lacking Dde I sites (5′ subset). The eight strains divide into three groups according to the length of the fragments in this band, each group containing one of the three *Sgs4* genes whose nucleotide sequence is known.

The Oregon-R group exhibits an Eco RI–Dde I fragment of 197 bp and includes Chieti V, Urbana-S, Seto and BER-1. Given the lengths of the Eco RI–Hind III fragment (L_T), of the Eco RI–Dde I fragment ($L_{5'}$), and of the Dde I–Hind III fragment ($L_{3'}$), we can calculate the number, $n_{3'}$, of contiguous Dde I–containing elements (3′ subset) from the equation:

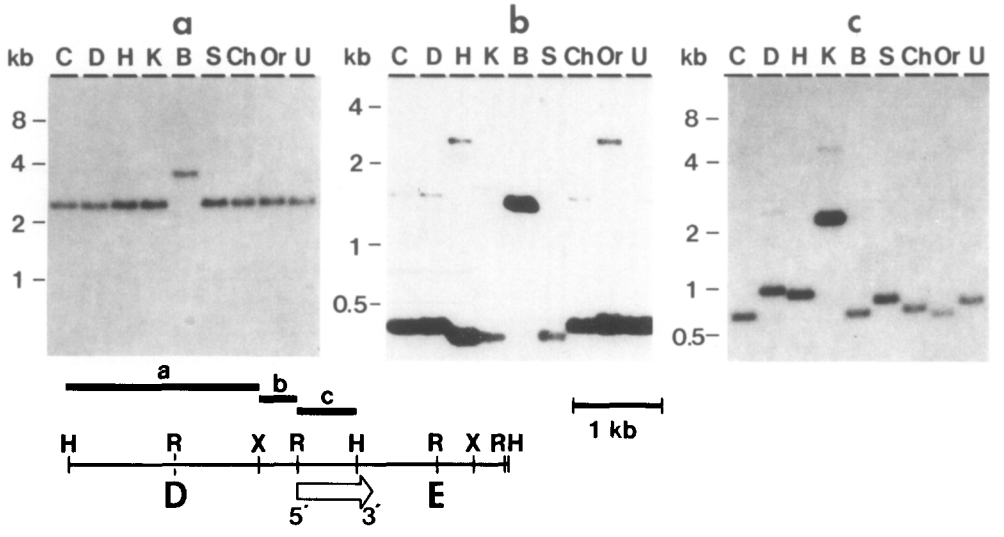

Figure 5. Interstrain Variation in Fragments Upstream and Downstream from the *Sgs4* 5′ Terminus

Equal amounts (2 μg) of total adult DNA, autoradiograms (a) and (c), or cloned Hind III fragment D, autoradiogram (b), from each of nine strains were digested with pairs of enzymes—(a) Hind III and Xho I; (b) Xho I and Eco RI; (c) Eco RI and Hind III—then electrophoretically fractionated in agarose gels of 1.0% for (a) and (c) or 0.8% for (b), blot-transferred to nitrocellulose and hybridized with the ^{32}P-labeled Oregon-R restriction fragments indicated by the black bars in the diagram below the autoradiograms. (a) 2.4 kb Hind III–Xho I fragment indicated by bar a; (b) 0.39 kb Xho I–Eco Ri fragment indicated by bar b; (c) 0.70 kb Eco RI–Hind III fragment indicated by bar c. Length standards and strain designations are as in Figure 4. The Hind III fragment D was cloned in pBR322 (Bolivar et al., 1977) for all strains except Kochi-R, where λ590 (Murray et al., 1977) was used as the cloning vector (Experimental Procedures). Faint bands of longer fragments observed in some tracks are due to probe contamination. In separate experiments, we have compared digests of the cloned segments from each of the eight strains with equivalent digests of their total adult chromosomal DNA and detected no differences. This is true even with respect to the Eco RI–Sau I fragments that encompass the tandem arrays (Figure 1), indicating that the arrays are stable to cloning.

$n_{3'} = 1 + [(L_T - L_{5'} - L_{3'})/21]$.

The resulting values are 14 for Oregon-R, 17 for Chieti V, 22 for Urbana-S, 23 for Seto and 17 for BER-1. If the Eco RI–Dde I fragment contains a 5′ subset of five elements, as is the case in Oregon-R, then the number of elements per array will total 5 more than the above values, and will equal those given in Table 1.

The Canton-S group includes D323 and yields an Eco RI–Dde I fragment of 239 bp, from which we calculate $n_{3'}$ values of 14 for Canton-S and 27 for D323. These values are 7 less than those in Table 1 because this Eco RI–Dde I fragment, at least in Canton-S, contains a 5′ subset of 7 elements. The third group contains only Hikone-R; here the Eco RI–Dde I fragment is 218 bp long, yields a value for $n_{3'}$ of 25 and contains a 5′ subset of 6.

Nonproducers Lack DNA Sequences Upstream from the mRNA Start Point

Figures 5a and 5b show that the DNA upstream from the mRNA start point yields restriction fragments whose lengths in the four nonproducers differ from those in the producers. The three nonproducers that yield a trace of *Sgs4* mRNA (Seto, Hikone-R and Kochi-R) each exhibit Xho I–Eco RI fragments immediately upstream from the mRNA start site that are shorter by ~50 bp than the equivalent 390 bp fragments of the five producers (Figure 5b). The adjacent Hind III–Xho I fragments do not, by contrast, differ among these Japanese nonproducers and the producers (Figure 5a). Similar analyses of the restriction fragments formed from cleavage at the Eco RI site at position ±1 bp and at the two Hinf I sites, at −163 bp and −251 bp (Figure 1), show that all nine strains are invariant with respect to the lengths of such fragments (data not shown). Hence, the ~50 bp deleted in each of the three Japanese nonproducers must be in the ~140 bp region between the Hinf I site at −251 bp and the Xho I site at −390 bp.

Figures 5a and 5b also show that the BER-1 fragments detected in the respective Hind III–Xho I and Xho I–Eco RI digests are abnormally long. This is because the BER-1 genomic DNA lacks the Xho I site shown on the restriction map given in Figure 5. Consequently, the BER-1 band shown in Figure 5a does not result from a Hind III–Xho I fragment, as in the other strains, but rather represents the same Hind III fragment shown in Figure 4a. Similarly, the BER-1 band in Figure 5b represents a fragment formed by cleavage at the two Eco RI sites that straddle the missing Xho I site. The observation that this 1.25 kb Eco RI fragment is 0.1 kb shorter than the equivalent fragment from a producer strain, such as Canton-S (data not shown), indicates that BER-1 has suffered a deletion of ~100 bp and suggests that this deletion eliminates the Xho I site.

These propositions are validated by the data in Figure 6. The black bar extending from −392 to −486 bp, inclusive, indicates the 95 bp region deleted

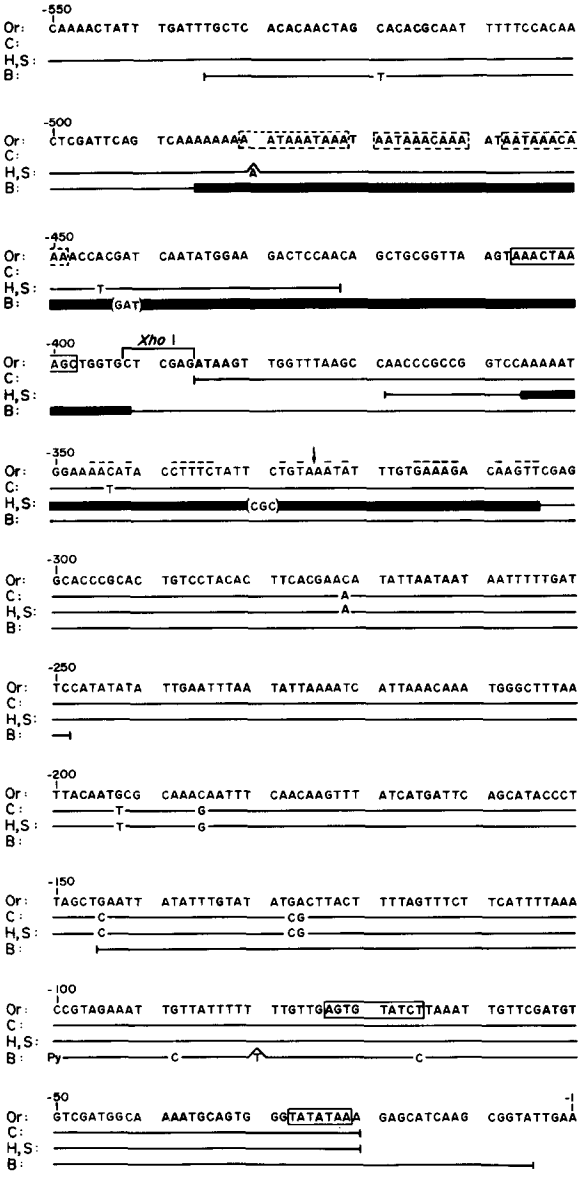

Figure 6. Sequences Upstream from the 5′ End of *Sgs4* in Oregon-R, Canton-S, Hikone-R, Seto and BER-1

The strain designations are those given in Figures 3–5, and the sequences are represented as in Figure 3, except that the boxed regions have the different meanings given below, a base under an inverted V represents an extra residue not present in the Oregon-R sequence, thick black bars indicate that the strain does not contain the Oregon-R sequence above the bar, and parenthetic sequences within a bar are those that replace the missing Oregon-R sequence. Because the parenthetic GAT triplet in the BER-1 black bar is also present in Oregon-R, one can also imagine that BER-1 is related to Oregon-R by two simple deletions: −392 to −440, and −444 to −486. The H,S line is a composite of the sequence for Hikone-R from −21 to −363 and of the sequence for Seto from −264 to −369 and from −423 to −550, with the Oregon-R sequence numbers; where the sequences overlap (−264 to −363), they are identical, including the boundaries of the deleted Oregon-R sequence and the substituted CGC triplet. The boxed sequence at −22 to −29 is the Goldberg-Hogness (Goldberg, 1979) or TATA box (Breathnach and Chambon, 1981), and that at −66 to −74 is similar to a second homology region, 5′-GGPyCAATCT-3′, noticed at this approximate position upstream from the mRNA start point of several eucaryotic genes in BER-1—or more precisely, the 95 bp sequence in Oregon-R that is replaced in BER-1 by the 3 bp (GAT) indicated within the black bar. One of the six base pairs in the Xho I site at −387 to −392 is included within the BER-1 deletion, which thereby inactivates this site.

The Hikone-R and Seto deletions are indicated by the black bar extending from −305 to −356 bp, inclusive; these 52 bp of Oregon-R are replaced in the two Japanese nonproducers by 3 bp (CGC). The boundaries of the deletion in the third Japanese nonproducer, Kochi-R, are within ±2 bp of those of the other two, the precision being less in this case because the sequencing gels exhibit ambiguities in the critical region.

Discussion

The mRNA-Homologous Sequences of *Sgs4*

The sgs4 polypeptide is, as are the other components of the salivary gland glue, a secretory protein. After accumulation in secretory vesicles during the latter half of the larval third instar (Lane et al., 1972), it is secreted into the gland lumen and expectorated at pupariation. Conceptual translation of the nucleotide sequence in the mRNA-homologous region shown in Figure 3 reveals an amino-terminal region of 22 amino acids that precedes the tandem array and exhibits the properties of a signal peptide (Wickner, 1980). Of the 22 residues, 17 are hydrophobic, and the rest are near either the initiator methionine or the expected cleavage site for this putative signal peptide.

Curiosities exist on both sides of the sequence coding for this putative signal peptide. Upstream, the 5′ noncoding leader sequence of the mRNA contains only 13 nucleotides (Figure 3). The precedents for such short leaders derive from RNA viruses, where lengths of 9 nucleotides in brome mosaic virus mRNAs (Dasgupta et al., 1975), 10 in vesicular stomatitis virus mRNAs (Rose, 1978) and 13 in reovirus mRNAs (Kozak, 1978) have been reported. However, the *Sgs4* leader is the shortest we are of aware of for nonviral eucaryotic genes. The short leader fits both the scanning model for initiation of translation (Kozak, 1978, 1980), in which ribosomes bind to the 5′ end and advance to the first AUG, and the model of Hagenbüchle et al. (1978), in which ribosomal binding involves base-pairing of sequences in the leader with sequences near the 3′ end of 18S rRNA. With respect to the second of these notions, the 5′-UUCCAAAG-3′ sequence at the 5′ end of the *Sgs4* mRNA can base-

(Breathnach and Chambon, 1981). The boxed sequence at −398 to −407 indicates elements of the sequence that Shermoen and Beckendorf (1982) have noted a few hundred base pairs upstream from the mRNA start point of certain D. melanogaster genes. The three dashed boxes enclose repeated AT-rich decamers within the region deleted in BER-1. A weak dyad symmetry within the Oregon-R DNA deleted in Hikone-R and Seto is indicated by the 11 overlined bases to each side of the arrow at −325/−326.

pair at six of its eight positions with the conserved 3′-AAGGCGUC-5′ sequence located 7 bases from the 3′ end of the 18S rRNA (Hagenbüchle et al., 1978). The tandem array of 21 base elements immediately downstream from the signal peptide coding sequence is interesting with respect both to the structure of the sgs4 glue polypeptide and to the stability of what Ycas (1972) has called periodic genes—that is, genes that exhibit short domain repetition, as is the case for the genes that encode collagen (Bornstein and Sage, 1980), fibroin (Suzuki and Brown, 1972) and other structural proteins, now including sgs4.

The 19 elements in the Oregon-R *Sgs4* account for 43% of the ~920 bases (exclusive of the poly(A) tail) in its mRNA, and an estimated 56% of its coding sequences. (This estimate derives from the data in Figure 3 and a preliminary sequence analysis of the +521 to +907 region in Oregon-R [Muskavitch, 1980], and the equivalent, essentially identical region in Canton-S and Hikone-R, that yields a termination codon at +726 ± 1 in phase with the codons shown in Figure 3, and an AATAAA, Proudfoot-Brownlee [1976] box at +887 to +892.) The amino acids in the repeated domain therefore dominate the composition of sgs4, particularly in the other strains, where they account for an even greater fraction of the coding sequences, reaching an estimated 70% in D323, the producer with the longest array (Table 1). Proline and threonine, which are represented twice per canonical element, will therefore each comprise 16% to 20% or more of sgs4. This glue polypeptide is glycosylated (Beckendorf and Kafatos, 1976; Korge 1977a), and we assume that the threonine residues act as glycosylation receptors. The proline residues may also play a role in such reactions (Aubert et al., 1976), although one expects their structural function to be more extensive, perhaps to provide structures that also facilitate the crosslinking of sgs4 to other polypeptides in the glue via the cysteine found in each element. It may be useful to imagine the peptide element as a target for such reactions, whose efficiency is increased by increasing the size of the target via tandem repetition of the element. One might then expect other glue polypeptides, and consequently their genes, to contain tandemly repeated elements. The gene, *Sgs3,* that encodes the sgs3 glue polypeptide has been identified and cloned by Meyerowitz and Hogness (1982), and analysis of its sequence has revealed a tandem array of repeated 15 bp elements among its coding sequences (K. Burtis and D. S. Hogness, unpublished experiments). By contrast, sequence analysis of three other putative glue polypeptide genes has not revealed such tandem arrays (E. M. Meyerowitz, personal communication; G. Guild, K. Burtis and D. S. Hogness, unpublished experiments).

The interstrain variation in the number of elements per array indicates that the arrays are also targets for genetic rearrangement reactions. Although our results do not distinguish among the possible rearrangement mechanisms (including, for example, unequal crossover between sister chromatids), they do indicate that the changes are primarily in the 3′ subset, and that they can occur over relatively short evolutionary times. With regard to these times, we found that three different stocks of Hikone-R from laboratories in Sweden (Umea culture collection), England (M. Ashburner) and California (S. K. Beckendorf) exhibit two *Sgs4* lengths that differ by at least three repeat elements; yet each stock presumably derives from the same isolate collected in 1952 at Hikone, Japan. Given that each of the three stocks exhibits only one *Sgs4* length, as is the case for all the strains we have examined, it appears unlikely that this difference derives from a polymorphism in the 1952 stock. Similarly, Hikone-R, Seto and Kochi-R exhibit significantly different numbers of elements per array (Table 1), yet appear to carry X chromosomes of recent common ancestry, as judged by the identical deletions in Hikone-R and Seto (Figure 6) and probably in Kochi-R.

Correlations between Expression and Genetic Structure

Developmental profiles of the amount of mRNA at a given stage were carried out for the nonproducers, as described for the Oregon-R producer by Muskavitch and Hogness (1980). No *Sgs4* RNA was detected in BER-1 at any larval, pupal or adult stage tested, while the Japanese nonproducers yielded the trace amounts of *Sgs4* mRNA indicated in Table 1 only during the last half of larval third instar—the period of mRNA production in producers (Muskavitch and Hogness, 1980). It is as if a door for transcription of *Sgs4* is open in the Japanese nonproducers at the same developmental times that it is open for producers, but open only a crack; and in BER-1, it never opens. We think the nonproducers are defective in transcription, rather than in RNA processing or stability, because *Sgs4* does not contain introns, we have not been able to detect *Sgs4* transcripts longer than the mRNAs (Muskavitch and Hogness, 1980) and no consistent difference between producers and nonproducers has been detected in the mRNA-homologous sequences. We note that the "zero" nonproducer, BER-1, and the "high" producer, Chieti V, both contain 22 elements in their arrays, and these appear to be distributed identically between the 3′ subset (17 elements) and the 5′ subset (5 elements). Similarly, the range of elements per array exhibited by producers is 19 to 34, and this encompasses all values exhibited by the nonproducers, with the sole exception of Kochi-R. Furthermore, no significant difference between producers and nonproducers is apparent in the mRNA-homologous sequences upstream or downstream from the tandem arrays (Figure 3 and Muskavitch, 1980).

Given that the nonproducers are defective in tran-

scription, one might expect from results obtained with other eucaryotic structural genes (reviewed by Breathnach and Chambon, 1981) that nonproducers and producers would exhibit differences in DNA sequences upstream from the mRNA start point. Figure 6 shows that both kinds of strains exhibit the same Goldberg-Hogness (Goldberg, 1979),or TATA box, which appears to be required for accurate initiation of transcription (Breathnach and Chambon, 1981). The 5′-TATATAA-3′ sequence at −28 to −22 conforms nicely to the consensus sequences (5′-TATA$^{A}_{T}A^{A}_{T}$-3′) and positions (3′-terminal A or T at −24 ± 4) derived from 60 examples of TATA boxes reviewed by Breathnach and Chambon (1981). A second box in Figure 6 outlines a sequence at −66 to −74 that exhibits a loose match with the consensus sequence (5′-GG$^{C}_{T}$CAATCT-3′) for a second region of homology at this approximate position (Breathnach and Chambon, 1981). Although the results of McKnight et al. (1981) indicate that the transcription of the gene coding for herpes thymidine kinase requires sequences at this approximate position upstream from the mRNA start site, they give little indication of whether the substitution in BER-1 of a C for the 3′-terminal T of the consensus sequence would be sufficient to eliminate *Sgs4* transcription. In any case, the Hikone-R sequence in and around this region is identical with that in the Oregon-R and Canton-S producers.

The only consistent differences in genetic structure that we have detected between producers and nonproducers are the deletions in the nonproducers of sequences present in the producers a few hundred base pairs upstream from the mRNA start point. We therefore propose that the developmentally controlled and efficient initiation of *Sgs4* transcription depends upon the sequences deleted in BER-1 and the Japanese nonproducers (Figure 6). More specifically, we imagine that some part of the 95 bp between −392 and −486 that are deleted in BER-1 is required for the developmentally specific regulation of *Sgs4*; that is, if these sequences are missing, the chromatin in and surrounding *Sgs4* cannot be differentiated with respect to cell type, and is closed to transcription in all cell types. By contrast, we imagine that the 52 bp between −305 and −356, which are deleted in the Japanese nonproducers, are not required for this developmentally specific regulation, because the trace amounts of *Sgs4* mRNA in these strains are normally regulated with respect to developmental time. We suppose that they are instead primarily involved in the efficiency of transcription initiation—perhaps in transmitting to the initiation site the effects of the regulatory interactions with the DNA of the BER-1 deletion. This distinction between the sequences deleted in BER-1 and in the Japanese nonproducers is compatible with the findings of Shermoen and Beckendorf (1982) concerning the effects of the two deletions on the five salivary-gland-specific DNAase I–hypersensitive sites located at or upstream from the 5′ end of *Sgs4*. Whereas the Hikone-R deletion eliminates only that site within the deleted DNA, the BER-1 deletion eliminates not only the two sites within this deleted DNA but also the other three sites.

We note without favoritism sequences within the two deletions which pique the curiosity because of symmetry or repetition, either within the region or at other loci. For example, the AT-rich region at −449 to −481 within the BER-1 deletion contains three direct repeats of the decamer, 5′-AATAAA$^{C}_{T}$AAA-3′, as indicated in Figure 6 by the three dashed boxes. One of the two DNAase I–hypersensitive sites within the BER-1 deleted DNA is located in the region of these repeated decamers. The other overlaps the boxed sequence at −398 to −407, which Shermoen and Beckendorf (1982) have emphasized because of this overlap and because they have noted similar sequences 157 to 420 bp upstream from several other D. melanogaster genes. Finally, we note a dyad symmetry about the arrow associated with the DNA deleted in Hikone-R and Seto. Curiously, the high producer (Oregon-R) and low producer (Canton-S) differ in this region by a single replacement at −344, which destroys one of the elements of the dyad (Figure 6).

In conclusion, we note that the two deletions discussed here appear to represent another case in a rapidly growing list of genes whose expression is affected by sequences quite far removed from the gene. Particularly impressive are the "enhancing" (Banerji et al., 1981) or "augmenting" (M. Fromm and P. Berg, personal communication) effects of the 72 bp repeated sequence in SV40, because these effects have been generalized to several genes by recombination of enhancer and gene in vitro and testing of the new combination in vivo. The functions that we have postulated for the DNA deleted in the BER-1 and Japanese nonproducers are now testable by this means, given the recent discovery of whole-organism, germline transformation in D. melanogaster by G. Rubin and A. Spradling (personal communication).

Experimental Procedures

Drosophila Strains

The origins of the strains used in this work have been given previously (Muskavitch and Hogness, 1980).

Nucleic Acid Isolation

Procedures for the isolation of D. melanogaster embryonic DNA, λ phage DNA, plasmid DNA and total organismal RNA have been described elsewhere (Muskavitch and Hogness, 1980). Poly(A)$^{+}$ RNA was isolated following the procedure of Meyerowitz and Hogness (1981). DNA fragments fractionated on agarose or polyacrylamide gels were prepared by electroelution (Jeppesen, 1980).

Genomic DNA from Drosophila adults was isolated by a technique developed by R. Lifton. One to two grams of Drosophila adults were frozen in liquid nitrogen and ground in a mortar and pestle chilled with liquid nitrogen. The resulting brown powder was placed in a 14 ml Dounce homogenizer (Kontes) with 10 ml homogenization buffer (0.5 M NaCl, 0.2 M sucrose, 0.01 M EDTA, 0.03 M Tris-HCl [pH

8.0]), and suspended by short bursts of vortexing. Cells were disrupted by repeated strokes until the pestle moved easily, followed by two to four more strokes. This suspension was filtered, under suction, through a 3-400-37 Nitex screen supported by Nitex 183 on a Buchner funnel. The homogenizer and screen were then rinsed with 2–3 ml of homogenization buffer. The filtrate was centrifuged at 9000 rpm for 10 min at 4°C in an HS-4 rotor to pellet nuclei. The supernatant was decanted, and the nuclear pellet was suspended in 1.5 ml lysis buffer (0.5 M EDTA, 10 mM Tris-HCl [pH 9.5]). Five milliliters of lysis buffer containing 2% (w/v) Sarkosyl (Chemical Additives Company) heated to 65°C was added, and the mixture was incubated at 50°C for 7 min. We then added 2.5 ml of 0.2 mg/ml proteinase K (EM Reagents) in water, and incubated the solution at 50°C for 4 hr. Following the incubation the solution was cooled, and 0.95 g cesium chloride was added per milliliter of solution. This mixture was centrifuged in a Ti50 rotor at 45,000 rpm for 3 days (19°C). Each tube was fractionated with a model 341 syringe pump (Sage Instruments) with flow rate set between 0.5 and 1.5 ml per minute. The most viscous fractions collected, which contain chromosomal DNA, were then pooled and dialyzed against four changes of 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA (TE8) at 4°C, and stored at 4°C.

Recombinant DNA

Genomic libraries were constructed in λ phage vectors, with either adult or embryonic DNA, by the method described in Davis et al. (1980). Hind III fragments were cloned in λ590 (Murray et al., 1977), and SacI fragments in λSep6 (Meyerowitz and Hogness, 1982). Recombinants were identified and isolated by the plaque filter hybridization method of Benton and Davis (1977). Hybridization probes were labeled by nick translation (Rigby et al., 1977). Recombinant phages were carried through two or more cycles of single plaque purification before being amplified into permanent stocks. Genomic segments of particular interest were subcloned into plasmid vectors, essentially as described in Davis et al. (1980).

Nucleic Acid Blots and Hybridization

DNA (Southern, 1975) and RNA (Alwine et al., 1977) transfers were carried out by standard procedures, with minor modifications, following fractionation of restricted or denatured nucleic acid on horizontal agarose gels in 40 mM Tris-acetate (pH 8.0) and 2 mM EDTA (McDonell et al., 1977). Hybridization probes were labeled by nick translation (Rigby et al., 1977). Conditions and procedures for hybridization, washing and autoradiography of nucleic acid blots have been described elsewhere (Muskavitch and Hogness, 1980).

DNA End-Labeling

Labeling of the 5′ termini of fragments with γ-^{32}P-ATP by use of T4 polynucleotide kinase (P-L Biochemicals) was performed following the method of Chaconas and Van de Sande (1980), except that calf intestinal alkaline phosphatase activity was eliminated by either phenol extraction or heating at 70°C for 30 min. Labeling of 3′ termini by filling of restriction sites with a single labeled deoxynucleoside triphosphate was performed under standard nick translation conditions (Rigby et al., 1977), with addition of only those cold and labeled deoxynucleoside triphosphates required to fill to the desired nucleotide. Incubation was for 10 min at 37°C, followed by 10 min at 70°C to eliminate DNAase I and DNA polymerase I activities. Desired fragments were isolated by electroelution following gel electrophoresis.

S1 Nuclease Mapping

This procedure was based on that of Berk and Sharp (1977). Probe fragments were end-labeled with T4 polynucleotide kinase or by filling in restriction sites. Probe, polyadenylated RNA and carrier (50 μg E. coli tRNA) were combined, precipitated from 0.1 M sodium acetate, 70% (v/v) ethanol, rinsed with 70% (v/v) ethanol and dried. Each pellet was resuspended in 30 μl deionized formamide, and the solution was adjusted to contain 400 mM NaCl, 40 mM PIPES (pH 6.4), 1 mM EDTA. The hybridization mixtures were heated at 70°C for 10 min, then incubated at 50°C for 4 hr. The mixtures were placed on ice, and 300 μl of an ice-chilled solution containing 30 mM Na acetate (pH 4.6), 50 mM NaCl, 1 mM Zn acetate, 40 μg/ml denatured calf thymus DNA and 60 U S1 nuclease (Sigma) were added to the mixture. The mixture was vortexed and incubated at 37°C for 30 min. The mixture was then placed on ice and precipitated, following addition of 900 μl of 95% (v/v) ethanol, as described above. Precipitated material was resuspended in 80% (v/v) formamide, 40 mM Tris-acetate (pH 8.0), EDTA and 0.1% (w/v) each of bromophenol blue and xylene cyanol. Samples were heated at 90°C for 1 min, quickly chilled in an ice-water bath and loaded on the desired gel system.

cDNA Elongation Experiments

A constant amount of end-labeled primer fragment was combined with various amounts of RNA, then precipitated, rinsed, hybridized and reprecipitated as described for S1 nuclease mapping. Each pellet was resuspended in 15 μl H_2O and adjusted to contain 50 mM Tris-HCl (pH 8.3), 5 mM $MgCl_2$, 140 mM KCl, 3 mM 2-mercaptoethanol, 100 μg/ml actinomycin D and 450 μM of each deoxyribonucleoside triphosphate in a total volume of 30 μl. AMV reverse transcriptase (1 μl; J. Beard) was added, and the reaction was incubated at 37°C for 90 min. Products were precipitated, rinsed and dried, following the addition of 5 μg E. coli tRNA carrier, as described for S1 mapping. Samples were prepared as for S1 mapping.

Sequence Determination and Strategy

Sequences were determined by the method of Maxam and Gilbert (1977), with minor modifications. Sequencing gels (0.5 mm × 30 cm × 40 cm) were run according to the technique of Sanger and Coulson (1978). Combined analyses of sequences from strains BER-1, Canton-S, Hikone-R, Oregon-R and Seto allowed determination of sequences from both strands over virtually all of the region between positions −550 and −1, inclusive. The remainder of the region was analyzed in two or three strains on the same strand. Deletions were localized by duplicate sequencing runs (BER-1 and Hikone-R) or single sequencing runs (Seto and Kochi-R) on the same strand. Sequences between positions +1 and +247, inclusive, as well as positions +311 and +520, inclusive, were determined from one strand in each strain. Sequences for the region +521 to +907 have been reported by Muskavitch, 1980 (in this Ph.D. thesis, the numbers for the equivalent region are +571 to +957).

Acknowledgments

We thank William McGinnis, Antony Shermoen and Steven Beckendorf for communication of their results. Kenneth Burtis and Katsuya Shigesada provided useful suggestions on sequencing techniques. Karen Muskavitch, Spyridon Artavanis-Tsakonas, Jeremy Nathans and Enno Krebbers were of immeasurable help during the completion of this work. M. A. T. M. is grateful to Susan, Robert and Emma Saint for the gracious hospitality extended to him while finishing this manuscript. This work was supported by grants from the National Institutes of Health and the National Science Foundation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 8, 1982; revised May 4, 1982

References

Akam, M. E., Roberts, D. B., Richards, G. P. and Ashburner, M. (1978). Drosophila: the genetics of two major larval proteins. Cell *13*, 215–225.

Alwine, J. C., Kemp, D. J. and Stark, G. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl paper and hybridization with DNA probes. Proc. Nat. Acad. Sci. USA *74*, 5350–5354.

Aubert, J.-P., Biserte, G. and Locheux-Lefebvre, M.-H. (1976). Carbohydrate–peptide linkage in glycoproteins. Arch. Biochem. Biophys. Acta *175*, 410–418.

Banerji, J., Rusconi, S. and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Beckendorf, S. K. and Kafatos, F. C. (1976). Differentiation in the salivary glands of Drosophila melanogaster: characterization of the glue proteins and their developmental appearance. Cell *9*, 365–373.

Benton, W. D. and Davis, R. W. (1977). Screening λgt recombinant clones by hybridization to single plaques in situ. Science *196*, 180–182.

Berendes, H. D. and Ashburner, M. (1978). The salivary glands. In The Genetics and Biology of Drosophila, *2B*, M. Ashburner and T. R. Wright, eds. (New York: Academic Press), pp. 453–498.

Berk, A. J. and Sharp, P. A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease–digested hybrids. Cell *12*: 721–732.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heynecker, H. L., Boyer, H. W., Crossa, J. H. and Falkow, S. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene *2*, 95–113.

Bornstein, P. and Sage, H. (1980). Structurally distinct collagen types. Ann. Rev. Biochem. *49*, 957–1003.

Breathnach, R. and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Ann. Rev. Biochem. *50*, 349–383.

Chaconas, G. and Van de Sande, J. H. (1980). 5′-^{32}P-labelling of RNA and DNA restriction fragments. Meth. Enzymol. *65*, 75–85.

Dasgupta, R., Shih, D. S., Saris, C. and Kaesberg, P. (1975). Nucleotide sequence of a viral RNA fragment that binds to eucaryotic ribosomes. Nature *256*, 624–628.

Davis, R. W., Botstein, D. and Roth, J. R. (1980). A Manual for Genetic Engineering: Advanced Bacterial Genetics. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Goldberg, M. L. (1979). Sequence analysis of Drosophila histone genes. Ph.D. thesis, Stanford University.

Hagenbüchle, O., Santer, M., Steitz, J. A. and Mans, R. J. (1978). Conservation of the primary structure at the 3′ end of the 18S rRNA from eucaryotic cells. Cell *13*, 551–563.

Jeppesen, P. G. N. (1980). Separation and isolation of DNA fragments using linear polyacrylamide gradient gel electrophoresis. Meth. Enzymol. *65*, 305–319.

Korge, G. (1975). Chromosome puff activity and protein synthesis in larval salivary glands of Drosophila melanogaster. Proc. Nat. Acad. Sci. USA *72*, 4550–4554.

Korge, G. (1977a). Larval saliva in Drosophila melanogaster: production, composition, and relationship to chromosome puffs. Dev. Biol. *58*, 339–355.

Korge, G. (1977b). Direct correlation between a chromosome puff and the synthesis of a larval protein in Drosophila melanogaster. Chromosoma *62*, 155–174.

Kozak, M. (1978). How do eucaryotic ribosomes select initiation regions in messenger RNA. Cell *15*, 1109–1123.

Kozak, M. (1980). Evaluation of the "scanning model" for initiation of protein synthesis in eukaryotes. Cell *22*, 7–8.

Lane, N. J., Carter, Y. R. and Ashburner, M. (1972). Puffs and salivary gland function: the fine structure of the larval and prepupal salivary glands of Drosophila melanogaster. Wilhelm Roux' Archiv. *169*, 216–238.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA *74*, 560–564.

McDonell, M. W., Simon, M. N. and Studier, F. W. (1977). Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. *110*, 119–146.

McKnight, S. L., Gavis, E. R., Kingsbury, R. and Axel, R. (1981). Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell *25*, 385–398.

Meyerowitz, E. M. and Hogness, D. S. (1982). Molecular organization of a Drosophila puff site that responds to ecdysone. Cell *28*, 165–176.

Murray, N. E., Brammar, W. J. and Murray, K. (1977). Lambdoid phages that simplify the recovery of in vitro recombinants. Mol. Gen. Genet. *150*, 53–61.

Muskavitch, M. A. T. (1980). Molecular analysis of a gene in a developmentally regulated puff of Drosophila melanogaster. Ph.D. thesis, Stanford University.

Muskavitch, M. A. T. and Hogness, D. S. (1980). Molecular analysis of a gene in a developmentally regulated puff of Drosophila melanogaster. Proc. Nat. Acad. Sci. *77*, 7362–7366.

Philippsen, P., Kramer, R. A. and Davis, R. W. (1978). Cloning of the yeast ribosomal DNA repeat unit in SstI and HindIII lambda vectors using genetic and physical size selections. J. Mol. Biol. *123*, 371–386.

Proudfoot, N. J. and Brownlee, G. G. (1976). 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature *263*, 211–214.

Rigby, P. W., Dieckmann, M., Rhodes, C. and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. *113*, 237–251.

Rose, J. K. (1978). Complete sequences of the ribosome recognition sites in VSV mRNAs: recognition by 40S and 80S complexes. Cell *14*, 345–353.

Sanger, F. and Coulson, A. R. (1978). The use of thin acrylamide gels for DNA sequencing. FEBS Letters *87*, 107–110.

Shermoen, A. W. and Beckendorf, S. K. (1982). A complex of interacting DNAase I–hypersensitive sites near the Drosophila glue protein gene, *Sgs4*. Cell *29*, 601–607.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–517.

Sutcliffe, J. G. (1978). pBR322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs long. Nucl. Acids Res. *5*, 2721–2728.

Suzuki, Y. and Brown, D. D. (1972). Isolation and identification of the messenger RNA for silk fibroin from Bombyx mori. J. Mol. Biol. *63*, 409–429.

Velissariou, V. and Ashburner, M. (1980). The secretory proteins of the larval salivary gland of Drosophila melanogaster. Chromosoma *77*, 13–27.

Wickner, W. (1980). Assembly of proteins into membranes. Science *210*, 861–868.

Ycas, M. (1972). De novo origin of periodic proteins. J. Mol. Evol. *2*, 17–27.