The suffix sequence is involved in processing the 3' ends of different mRNAs in *Drosophila melanogaster*

Nickolai A.Tchurikov, Alexander K.Ebrallidze and Georgii P.Georgiev

Institute of Molecular Biology, USSR Academy of Sciences, Vavilov str. 32, 117984 Moscow B-334, USSR

Communicated by G.P.Georgiev

We describe here the nucleotide sequences of several genomic and mRNA copies of the suffix, a short dispersed actively transcribed repeat located at the 3' ends of many different genes of *Drosophila melanogaster*. Only one strand of the suffix is transcribed. The patterns of suffix-containing mRNAs vary during development. The five randomly selected genomic copies of the suffix are 265 bp long and quite conservative in their sequence. The non-transcribed strand is terminated with oligo(A) preceded by AATAAA sequence. No repetitive flanking sequences can be detected. The three other genomic copies selected by hybridization with suffix-containing cDNA clones are less conservative, especially in the 5' part. In particular, they contain short insertions carrying a polyadenylation signal AATAAA at exactly the same position of the suffix. Comparison of genomic and cDNA clones shows that mRNAs are polyadenylated at the last nucleotide of these insertions. The cDNA clones include the same part of the suffix, from the 39th to 112th nucleotide. Thus, a segment of the suffix forms the last exon for both genes. In one case, the beginning of the last intron coincides with the beginning of the suffix, creating a very unusual donor splicing site. We conclude that the suffix sequence is directly involved in the formation of the last splicing site and 3'-end maturation of mRNA, at least in the case of the two genes analysed.

Key words: transcribed repeat/developmental regulation/splicing/3'-mRNA maturation/*Drosophila*

Introduction

Short transcribed repetitive elements have been found in a variety of eukaryotes (for reviews, see Davidson and Posakony, 1982, and references therein). The functions of these elements are not known yet, although it is often speculated that they are involved in the regulation of gene expression (Georgiev, 1969; Britten and Davidson, 1969; Constantini *et al.*, 1980; Murphy *et al.*, 1983). It has been shown that the pattern of mRNAs partly composed of repeated sequences changes during development (Scheller *et al.*, 1978; Zuker and Lodish, 1981).

Earlier, we described a short actively transcribed element denoted as suffix which is repeated 300–400 times in the *Drosophila melanogaster* genome and occurs at the 3' end of many different mRNAs (Tchurikov *et al.*, 1982). In particular, the suffix was found to constitute the 3'-end exon of gene H55 located in the *cut* locus. Thereafter, some other short dispersed transcribed repeats were described in *D. melanogaster*, such as homeobox and opa (McGinnis *et al.*, 1984; Wharton *et al.*, 1985). These repeats were found to contain an open reading frame and to constitute certain parts of coding gene sequences (Gehring, 1985;

Wharton *et al.*, 1985). They occur in many eukaryotic organisms. Their functions in the cell are as yet unknown. The homeobox-containing genes play a role in the control of development (Gehring, 1985). The homeo domain may be involved in DNA or chromatin binding (McGinnis *et al.*, 1984; Shepherd *et al.*, 1984).

Another possible function of short transcribed repeats is to control the earlier steps of gene expression like transcription, pre-mRNA processing, mRNA maturation, its transport from the nucleus to cytoplasm, or translation.

To understand the role of the suffix, we have sequenced several genomic and mRNA copies of the suffix together with the flanking sequences. The data obtained provide evidence that the suffix is responsible for the maturation of the 3' domain in mRNA including the formation of the last exon, 3' end cleavage and polyadenylation at least in two different mRNAs. A part of the suffix ~80 bp in length constitutes separate 3' end exons in the corresponding genes.

Results

Nucleotide sequence of five randomly selected genomic copies of the suffix

We have screened a number of genomic clones containing the suffix by colony hybridization of *D. melanogaster* genomic DNA-bank in pBR322 plasmid with a suffix-containing segment of clone H55.

Four randomly selected clones and a suffix-containing segment of clone H55 were sequenced using the modified Maxam and Gilbert (Chuvpilo and Kravchenko, 1985) technique (Figure 1).

Fig. 1. The genomic copy of the suffix. The lines above the sequence indicate the putative polyadenylation signal and TG-rich segment. The dotted line indicates the AATAAA sequence in suffix-complementary strand. The deviations of individual clones from the consensus sequence as well as two examples of flanking sequences are presented.

Fig. 2. Developmental regulation of transcripts complementary to the suffix. About 5 μ g of different RNA preparations [total, nuclear and cytoplasmic poly(A)⁺ and poly(A)⁻ RNAs] were separated on formaldehyde-agarose gels and transferred onto nitrocellulose filters as described by Thomas (1980). The Northern blots were hybridized with ³²P-labelled probe prepared from an M13 subclone containing a minus strand of *Xba*I-*Hind*III fragment of H55. λ *Hind*III fragments were used as mol. wt standards.

One can see that the region of homology in all the five clones extends to 265 nucleotides, i.e. the length of the suffix is 265 nucleotides.

Analysis of the sequence allows several conclusions to be drawn. First, the suffix is extremely homogeneous. The sequence of individual copies deviates from the consensus one in only one or two nucleotide changes, i.e. the homology is >99%. Second, there are no nucleotide duplications at the ends of the suffix copies analysed. This makes the suffix different from the known mobile elements of *D. melanogaster* (see also Figures 6 and 7). Third, the transcribed strand of the suffix (Tchurikov *et al.*, 1982; see also below) begins with an oligo(dT) stretch 16 nucleotides long. This corresponds to oligo(dA) in the complementary strand preceded there by an AATAAA sequence, 13 nucleotides upstream. Analysis of the mRNA strand allows the determination of the region which may serve as a polyadenylation signal. This is an AATAAT sequence (positioned at 148 bp) followed downstream by a TG-rich segment (162–171 bp). Finally, it should be noted that the suffix contains a number of stop codons in all three reading frames. Thus, as a whole, it cannot serve as a protein-coding sequence.

All transcribed copies of the suffix are read in the same direction
Since the direction of suffix transcription was known only in the case of H55, further experiments were made using Northern blots

of RNA preparations at different stages of *Drosophila* development, which were probed with strands of the suffix cloned in the M13 phage.

Northern blot autoradiographs (Figure 2) reveal that the different poly(A)⁺ RNA preparations, nuclear and cytoplasmic, obtained from embryos, larvae, pupae and culture cells always hybridize with only one strand of the suffix. No signal can be detected upon hybridization with the other strand (data not shown). As judged from the Northern blot analysis, only polyadenylated RNA molecules in both the nuclei and cytoplasm contain a suffix.

The distribution of the label in the Northern blot shows a high heterogeneity of suffix-containing RNAs. However, besides a smear, one can see from three to four more prominent bands corresponding to abundant mRNA species. The size of abundant mRNAs is from 1 to 8 kb. The comparison of the hybridization patterns in Figure 2 indicates that suffix-containing mRNAs are present throughout development. However, embryos and larvae display different patterns of expression. The same is true of pupae and imago whereas changes between larvae and pupae are small.

Structure of the suffix present in two mRNAs (cDNA clones)

The earlier data suggested that the suffix formed a separate 3'-end exon in H55 gene and possibly in some other genes (Tchurikov *et al.*, 1982). To study how the suffix is represented in mRNAs,

Fig. 3. The cDNA regions containing the suffix. The sequences of the cDNA clones C208 (A) and C363 (B) are shown. The direction of transcription (5' to 3') is from left to right. Underlined is the region of the suffix. The brackets above the sequence indicate the small insertions in the suffix region. The sequences from the 3' polyadenylation signal to the 3' polyadenylation site are in boxes. The asterisks indicate single deviations from the consensus suffix sequence. The arrows indicate the position of small (4 and 5 bp) deletions in the suffix region of cDNA clone C363 (B).

we prepared a cDNA library in pUC9 plasmid using embryonic poly(A)⁺ RNA as a template for reverse transcriptase and screened it for suffix-hybridizing clones.

Several positives were obtained and the insertions from two of them (C208 and C363) were sequenced. We found both cloned cDNAs contained a long poly(A) tail at one end indicating that they represented polyadenylated mRNAs (Figure 3). The poly(A) tail is preceded by an AATAAA sequence located 16 and 22 nucleotides upstream, which is typical of usual mRNAs.

No complete suffix sequence could be detected in either C208 or C363 RNAs. Only a part of the suffix is present in the 3' region of mRNAs studied, and this part is identical in both clones. In clone C208, a segment of the suffix from 39 to 112 nucleotides is located just before an AATAAA sequence. The latter is not revealed in the consensus suffix (see Figure 1). Even this rather short segment of the suffix differs from the consensus one by an insertion of short foreign sequences (up to 8 bp long) at four different sites.

The structure of clone C363 is even more complicated. The same segment of the suffix (39–112 bp) is split into three parts in the 3'-end area of the mRNA and interrupted with two non-suffix segments of 146 and 43 nucleotides in length. Short oligonucleotide insertions into the suffix sequences are seen too. An

AATAAA sequence is located 3' beyond the end of the suffix segment.

Hence the segment recovered in the mRNAs is 73 nucleotides long, which is less than one-third of the total suffix length.

A number of stop codons occur in the part of clone C363 preceding the suffix sequences, which indicates the latter to be certainly located in the 3' non-translated region. On the contrary, one of the frames in clone C208 contains a stop codon within the suffix only (positioned at nucleotide 266, see Figure 3A). Of course, it does not rule out the possibility that a stop codon is present upstream from the cloned part of C208 mRNA. The unique parts of both clones were used in Northern hybridization experiments. Only one mRNA band was detected in both cases (Figure 4). Since unique parts of the genes corresponding to two mRNAs have different genomic distributions (see below), we may conclude that these mRNAs with similar sizes (~1.5 kb) are different. The signals are rather strong to suggest that these mRNAs correspond to actively transcribed genes. However, they do not correspond to the major suffix-containing bands.

Comparison of related genomic and mRNA (cDNA) suffix-containing sequences

To elucidate the suffix structure in genes corresponding to cDNA clones, we have isolated genomic λ -clones containing C208 and C363 sequences with non-repetitive parts of cDNA clones.

Preliminary Southern blot hybridization has shown that the gene encoding C208 mRNA is represented in the Oregon RC line by two copies or two variants while that coding for C363 mRNA is represented by only one (Figure 5). In the *cl^{C20}*-line of *D. melanogaster*, the gene C208 is represented by one variant. The DNA fragments from genomic clones were sequenced (Figures 6 and 7). Two copies of gene 208 (AG208 and CG208) are very similar and both contain a suffix. However, their suffixes differ since one of them contains short oligonucleotide insertions. These insertions coincide with those in the cDNA clone and therefore the corresponding genomic copy, namely CG208, is the one responsible for C208 mRNA formation.

First, one can see that CG208 contains practically a full-length copy of the canonical suffix. It only possesses a shorter oligo(dT) stretch at the 5' end and lacks the six last nucleotides at the 3' end. The suffix is interrupted with several oligonucleotide insertions, the largest one starting with an AATAAA signal (see also the C208 cDNA clone, Figure 3). The 3' part of the suffix in CG208 (and AG208) is closer to the consensus suffix sequence than is its 5' part. Comparison of C208 (cDNA clone) and CG208 (gene) sequences reveals that the first 25 nucleotides of the suffix form a small intron in the gene. The 3' end of mRNA exactly coincides with the 3' end of the insertion starting with AATAAA. Thus, over half of the suffix is located beyond the mRNA 3' end.

Similar results, differing in detail, were obtained with gene CG363 (Figure 7). The suffix in the genomic CG363 clone begins with a slightly shortened oligo(dT) stretch. It can be concluded by comparing C363 and CG363 sequences that the beginning of the suffix is spliced out in mRNA in such a way that the acceptor site coincides with that in gene CG208. However, the donor site in gene CG363 is located far upstream within the single copy part of the gene. Actually it is not detectable in the sequenced area of the genomic clone.

The 5' part of the genomic clone which follows the donor splicing site is identical to that in cDNA and contains the same insertions into the suffix. The last insertion starting with an AATAAA signal is located in the same position as in gene CG208. Similarly, the last nucleotide of the insertion in gene CG363 is a polyadenyl-

Fig. 4. Northern blotting analysis of DNA sequences of cDNA clones C208 (A) and C363 (B) located upstream of the suffix fragment. About 5 μ g of RNAs from embryos, larvae, pupae and flies were electrophoresed on formaldehyde-agarose gel, transferred onto nitrocellulose filters (Thomas, 1980), and hybridized to 32 P-labelled probes of 121 bp *AluI* fragment of the clone C208 (77–198 bp) and 59 bp by *MspI*–*HaeIII* fragment of the clone C363 (323–382 bp).

ation site which causes the 3' part of the suffix to be lost from the cDNA clone. In the genomic clone, however, the perfect 3' part of the suffix, 140 nucleotides long, follows the 22-bp AAT-AAA-containing insertion.

Thus, in both the genes analysed, the suffix provides the gene with an acceptor-splicing site and with a polyadenylation signal, and in one case, with a donor-splicing site. The polyadenylation signals result from a short AATAAA-containing insertion at the same site of the suffix sequence (see Figures 6–8).

Discussion

The suffix supplies genes with splicing sites and forms the 3' exon

The sequencing data presented here and summarized in Figure 8 confirm the major conclusion from the previous paper (Tchurikov *et al.*, 1982) based on hybridization analysis, that the suffix sequence is located at the 3' ends in different genes and possibly forms the last exon. At least in the two genes studied here, the last exon is really formed by the suffix. In both cases, the same sequence within the suffix is used as a 3'-splicing site. Its structure is rather close to the consensus one: $\frac{T}{C}N\frac{C}{T}AG\downarrow$ for the consensus and $TTTA\frac{G}{T}\downarrow$ for the copies of the suffix in genes CG363 and CG208.

The 5'-splicing site is located within a non-suffix sequence in one case. It nearly coincides with the beginning of the suffix in the other where it is unusual: $\underline{GGC}\downarrow TTTATT$ versus the consensus sequence $\frac{C}{A}AG\downarrow GT\frac{A}{G}AGT$ (Mount, 1982). Moreover, the intron in CG208 is unusually short (only 28 nucleotides). A special splicing mechanism may operate in this case.

Several other repetitive sequences are located near the 3' ends of certain genes. The homeobox and possibly opa being coding

Fig. 5. Southern blotting analysis of DNA sequences of cDNA clones C208 (A) and C363 (B). About 5 μ g of DNA from Oregon RC or *ctJc20* embryos were digested with *EcoRI*, electrophoresed on a 0.8% agarose gel, and hybridized to 32 P-labelled DNA probes of *AluI* (77–198 bp) fragment of the clone C208 and *MspI*–*HaeIII* (323–382) fragment of the clone C363.

Fig. 6. Nucleotide sequence of genes AG208 (A) and CG208 (B). The direction of transcription (5' to 3') is from left to right. The region of the suffix is underlined. Dotted line indicates the sequence of cDNA clone C208. Brackets above the sequence indicate small insertions in the suffix. The splice junctions of the intron lie at positions 227 and 253 bp. The sequence from AATAAA signal to poly(A) site is boxed. Schematic diagram showing the relations of sequenced regions of the gene CG208 and cDNA clone C208. The hatched and solid bars in the gene correspond to the suffix and insertion, respectively. The open arrows indicate the position of small (13 and 6 bp) deletions in the suffix.

sequences are totally incorporated into mature mRNA (McGinnis *et al.*, 1984; Wharton *et al.*, 1985). A short sequence ~200 bp long containing an oligo(A) stretch and several polyadenylation signals at one end was studied recently by O'Hare in one of the *white* mutants of *D. melanogaster* (personal communication). Direct sequence comparison revealed no similarity between this sequence and the suffix. In some cases, a mouse B2 sequence, though non-translated, is also present as a whole at the 3' end of mRNAs (Ryskov *et al.*, 1983). It is noteworthy that B2 as well as B1 and Alu repeats contain sequences which are reminiscent of both donor and acceptor-splicing sites, but no data are available on their functioning. Thus, the suffix still remains the only repetitive sequence supplying genes with the functional splicing site(s).

The suffix supplies genes with signals for 3'-end cleavage and polyadenylation

The processing of a primary transcript is an important step in mRNA formation. Besides capping and splicing, it includes the formation of a specific polyadenylated 3' terminus, i.e. RNA cleavage at the polyadenylation site followed by poly(A) addition. Only such mature mRNA is transported into the cytoplasm.

The only stringently conserved sequence important for the 3' end recognition is an AATAAA hexanucleotide although minor specific nucleotide substitutions are tolerated (Wickens and Stephenson, 1984). It is found between 11 and 30 nucleotides upstream from the polyadenylation site of nearly all eukaryotic mRNAs (Proudfoot and Brownlee, 1976). Deletions or point mutations within this oligonucleotide severely reduce or abolish correct polyadenylation (Fitzgerald and Shenk, 1981; Montell *et al.*, 1983; Higgs *et al.*, 1983).

Although an AATAAA sequence is necessary for the maturation of the mRNA 3' end, it is not sufficient: the sequence re-

mains unrecognized in some coding regions and sometimes in 3' non-coding regions between two polyadenylation signals which are utilized (Manley, 1983; Fitzgerald and Shenk, 1981; Tosi *et al.*, 1981). It has been shown that 35 nucleotides distal to the site of poly(A) addition suffice for the maturation of the mRNA 3' end (McDevitt *et al.*, 1984; Gil and Proudfoot, 1984; McLauchlan *et al.*, 1985). A conserved G+T-rich sequence was found in this region located downstream from mRNA 3' termini (Gil and Proudfoot, 1984; McLauchlan *et al.*, 1985). Deletions of this G+T-rich sequence downstream from the 3' poly(A) site dramatically reduce normal enzyme activity (McLauchlan *et al.*, 1985). DNA sequences involved in 3' cleavage are also located in the region upstream of the AATAAA signal (Wickens and Stephenson, 1984).

Thus, the general conclusion may be drawn that dozens of nucleotides before the polyadenylation signal and ~40 nucleotides beyond the polyadenylation site are crucial for selecting the exact site at which an mRNA precursor will be cleaved. Both these regions are represented within a genuine suffix sequence in genes CG208 and CG363 (Figures 7 and 8). Therefore the suffix sequence around an AATAAA-containing insertion should be directly involved in mRNA 3' end maturation. The downstream area contains an oligo(dT) stretch and a TGT-element characteristic of such a region in several other genes (for reviews, see Birnstiel *et al.*, 1985).

The question arises which part of the suffix forms the 3' end in the absence of an AATAAA-containing insertion, for example, in gene H55. This is now under investigation. Possibly this is an AATAAT sequence (positioned at 148 bp) followed by TGGGTTGGTA (Figure 8) which is close to a consensus GT-rich motif (TGTGTTGGAA) found downstream from the polyadenylation site in several other genes (Taya *et al.*, 1982; Gil and Proudfoot, 1984).

10 20 30 40 50 60 70
AAGCTAAGTAACGTAGCGAGGCACTACCCCTTCTCTGTGTCCGACCCCTCAAGCGCGACAGCTGAAGC
30 90 100 110 120 130 140
TATATCTTCTGACAGTTTGGCGAATTTCTGGCAGTATTGCCGCCCTTCTGCCGCACTTTTGTCTG
150 160 170 180 190 200 210
ATATGCTGATACCGCACTTCCGAGCCTCTGCGCCTCTGTGGACGCCCGACGACGCCCGCCGCGC
12 220 230 240 250 260 270 280
CGCTGTTTTTATTTTACCTTTTATTTAGTGAATCGTCTTAAGAAGCAACAGATCCACAGGCTGAGTT
290 300 310 320 330 340 350
CCACACAGCAGATCAAGGAGCAGGAGGCCATCCAGCGCCCAAGGAGTTCCTCGAGAACCTGCACA
360 370 380 390 400 410 420
AGTGAGCGGTGATCCACCACCGGAGAGGCGACCGACGACGATGAACGAGATAAGATTAAAAATGC
430 440 450 460 470 480 490
ATTTCACTCTCTTTTATGTTCCCAAGTCTCACTACCACTTACAGTAACCAAAATATATATGT
500 510 520 530 540 550 560
ATTGAAAGTGGTAAATAACCATCCATATTTGTATGCTTTTTCACACTGCTCTATGTTAAACGGC
570 580 590 600 610 620 630
CCTAATAATTATTCGCTGGGTGGTAGGCCCTTTCGCCGGAGACGGGAACGCGTGGTAGCTGGATTAG
640 650 660 670 680 690 700
ACCTCGCGCTAGGTGGTGGGTGCGTGGCAAGCTAGTTTCATTTTTGCGTCCGCGCGACAGGCGC
710 720 730 740 750 760 770
TCCGTTCCGCTGCTGTTTCAGACGACCGGCGCATCTGCTCGAAGCTTTTGGCCCATGGCAGAAATTTTC
780 790 800 810 820
TGCCGCCGTTGCCGAGGAAGCTGCACGACCGGCACTTACAGACGCGTTTCG

Fig. 7. Nucleotide sequence of gene CG363. The direction of transcription (5' to 3') is from left to right. The region of the suffix is underlined. Brackets indicate the smaller insertions in the suffix. The region from 3' poly(A) signal to 3' poly(A) site is in the box. Dotted line indicates the sequence of cDNA clone C363. The 3' splice junction lies at position 242 bp. The scheme shows the relations of sequenced regions of the gene and cDNA clone. The designations are the same as in Figure 6B.

The nature and origin of the suffix

Concerning the nature and origin of the suffix, our data create more questions than answers. Usually dispersed repetitive sequences are now considered as mobile elements (Georgiev *et al.*, 1984). The complementary (minus) strand of the suffix, an anti-suffix, is similar to the mRNA 3' end which is characteristic of retroposons or sequences transposed by reverse transcription. Like a typical retroposon such as mouse B2-element (Krayev *et al.*, 1982), the anti-suffix contains ATAAA signals and an oligo(A) tail 13 nucleotides downstream.

However, an anti-suffix does not possess RNA polymerase III initiation and termination signals which make it possible for a B2 repeat to be transcribed into small B2 RNA by RNA polymerase III (Krayev *et al.*, 1982). In fact, no anti-suffix sequence was found among both nuclear and cytoplasmic transcripts (see Figure 4). Furthermore, in contrast to the B2 sequence and other retroposons, the suffix is not flanked by direct repeats.

Nevertheless, one cannot exclude the possibility that the anti-suffix sequence is transcribed at some particular stage of development. It should also be pointed out that not all copies of B-type (Alu-like) repeats were found to be flanked by direct repeats. Probably in some cases the cleavage with staggered ends is dispensable for retroposon insertion.

Fig. 8. Summarized data on the suffix structure. The mRNA strand of the suffix is presented. The direction of transcription (5' to 3') is from left to right. Small arrows indicate the beginning and the end of suffix copies present in CG208 and CG363 genes as well as splice sites. The large arrow indicates the position of AATAAA-containing insertions. The open triangles indicate the positions and sizes (in bp) of insertions. The asterisks indicate deviations from the consensus suffix sequence. The brackets indicate the small deletions in the suffix regions in genes studied. The designations above and below the sequence correspond to CG208 and CG363, respectively. The part of the suffix represented in mRNA is underlined. The putative 3' poly(A) signal is in the box. TG regions are underlined.

Another possible explanation for the absence of target duplication is the heterogeneity of suffix-containing transcripts used for retroposition which might start outside the suffix itself.

Whatever the real mechanism is, we suggest that the anti-suffix sequence was used for retroposition of the suffix. Thereafter several copies of the suffix were inserted into appropriate genomic sites to form the 3' ends of corresponding genes. Such constructions survived in the course of natural selection and the suffix emerged at the 3' ends of several different genes. It seems much easier to select by hybridization the perfect copies of the suffix from a genomic bank than split ones. Nevertheless, the extremely high homogeneity of five randomly selected copies of the suffix is a very intriguing feature. It suggests either a very recent suffix amplification or the existence of an efficient correction, for example, through gene conversion. It is not clear whether these homogeneous suffix sequences are functional, except for the copy isolated from clone H55 where the suffix is certainly a part of the gene (Tchurikov *et al.*, 1982).

It is even more difficult to explain the emergence of two different short insertions at exactly the same site of two different copies of the suffix present in the functional genes. They both begin with AATAAA and contain a common ATA^{ACTTTG} sequence (Figure 8).

The important question is whether the suffix location differs among *D. melanogaster* strains as in the case of *Drosophila* mobile genetic elements. This question awaits further experiments.

Materials and methods

Isolation of nucleic acids

DNA and RNA preparations were isolated as described by Tchurikov *et al.* (1982). Total RNA was prepared using a guanidium thiocyanate extraction procedure (Chirgwin *et al.*, 1979).

Recombinant DNA

The bank of *Drosophila* DNA in plasmid pBR322 was prepared as described (Tchurikov *et al.*, 1982).

The cDNA library in pUC9 plasmid was obtained as follows. Poly(A)⁺ RNA from 0–24 h Oregon RC embryos was primed with oligo(dT) and the first strand of cDNA was made with reverse transcriptase. RNA was removed by alkaline

hydrolysis and cDNA was separated on a Sephadex G-100 column. The second strand was synthesized with DNA polymerase I and, after S1 treatment, the dsDNA was used in the reaction with the Klenow fragment of DNA polymerase I to obtain perfect blunt ends. Equimolar amounts of dsDNA and pUC9 vector treated with *Sma*I endonuclease were ligated, incubated at 65°C for 5 min, treated with *Sma*I enzyme, and used for transformation of JM103 *Escherichia coli* cells.

A library in λ EMBL-4 vector was prepared from Oregon RC DNA partially digested with *Sau*3A restriction endonuclease. After electrophoresis in 0.3% agarose gel, a number of DNA zones were excised and the DNA was eluted electrophoretically. After a concentration procedure with *n*-butanol, the DNA solution was extracted with phenol–chloroform and precipitated with ethanol. The final pellet was washed with 70% ethanol and resuspended in 200 μ l 1 mM Tris pH 7.4–0.1 mM EDTA (TE). After dialysis for 2 h against TE, the aliquots of DNA samples were used for control electrophoresis. The genomic DNA preparation of selected size was mixed with λ -arms. λ -Arms were eluted from the agarose gel and after electroelution and purification were used for ligation. Fifty to one hundred microlitres of the mixture containing $\sim 3 \mu$ g λ -arm and 2 μ g *Drosophila* DNA in buffer containing 0.07 mM ATP, 0.1 M NaCl, 0.05 M Tris-HCl buffer, pH 7.4, 8 mM MgCl₂, 8 mM β -mercaptoethanol and 50 μ g/ml bovine serum albumin were incubated at 65°C for 5 min and cooled to room temperature for 1 h. Then 5 units of T4 DNA ligase was added and the mixture was incubated overnight at 4°C.

Southern and Northern blotting and hybridization

Agarose gels (0.8–3%) in Tris-acetate or Tris-borate buffers were used throughout the study. DNA fragments were denatured in 0.2 M NaOH, 0.6 M NaCl for 1 h, neutralized in 1 M Tris-HCl (pH 7.5), 0.6 M NaCl for 1 h, and blotted in 20 $\times$ SSC onto nitrocellulose sheets (Schleicher and Schuell). The filters were rinsed in 2 $\times$ SSC and vacuum baked at 80°C.

RNA was analysed in formaldehyde 1.3% agarose gels as described by Maniatis *et al.* (1982). RNA was blotted onto nitrocellulose using the procedure of Thomas (1980) without any pre-treatment of the gel.

Single-stranded DNA hybridization probes were prepared according to Hu and Messing (1982).

DNA sequencing

All DNA sequences were obtained using the chemical degradation method of Maxam and Gilbert (1980) modified by Chuvpilo and Kravchenko (1985), a simplified solid-phase technique for DNA sequencing which is much faster and less laborious than the original method.

Acknowledgements

We thank our colleagues from the Laboratory of Nucleic Acid Biosynthesis for help during this work. We are particularly indebted to Dr M. Ivanova for her valuable advice, to Dr N. I. Barbakar and Dr R. Hechumyan for assistance, to Dr L. G. Polukarova for propagation of *Drosophila* culture cells, and to I. N. Strizhak for typing the manuscript. We are grateful to Dr J. I. Yoder for reading the manuscript.

References

- Birnsteil, M.L., Busslinger, M. and Strub, K. (1985) *Cell*, **41**, 349–359.
- Britten, R.J. and Davidson, E.H. (1969) *Science*, **165**, 349–353.
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry*, **18**, 5294–5299.
- Chuvpilo, S.A. and Kravchenko, V.V. (1985) *FEBS Lett.*, **179**, 34–36.
- Constantini, F.D., Britten, R.J. and Davidson, E.H. (1980) *Nature*, **287**, 111–117.
- Davidson, E.H. and Posakony, J.W. (1982) *Nature*, **297**, 633–635.
- Fitzgerald, M. and Shenk, T. (1981) *Cell*, **24**, 251–260.
- Gehring, W.J. (1985) *Cell*, **40**, 3–5.
- Georgiev, G.P. (1969) *J. Theor. Biol.*, **25**, 473–480.
- Georgiev, G.P., Ilyin, Y.V., Chmelianskaite, V.G., Ryskov, A.P., Kramerov, D.A., Skryabin, K.G., Krayev, A.S., Lukanidin, E.M. and Grigoryan, M.S. (1981) *Cold Spring Harbor Symp. Quant. Biol.*, **46**, 641–653.
- Gil, A. and Proudfoot, N.J. (1984) *Nature*, **312**, 473–474.
- Higgs, D.R., Goodbourn, S.E.Y., Lamb, J., Clegg, J.B., Weatherall, D.J. and Proudfoot, N.J. (1983) *Nature*, **306**, 398–400.
- Hu, N., Messing, J. (1982) *Gene*, **17**, 271–277.
- Krayev, A.S., Markushcheva, T.V., Kramerov, D.A., Ryskov, A.P., Skryabin, K.G., Bayev, A.A. and Georgiev, G.P. (1982) *Nucleic Acids Res.*, **10**, 7461–7475.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, NY.
- Manley, J.L. (1983) *Cell*, **33**, 595–605.
- Maxam, A.M. and Gilbert, W. (1979) *Methods Enzymol.*, **65**, 499–560.
- McDevitt, M.A., Imperiale, M.J., Ali, H. and Novins, J.R. (1984) *Cell*, **37**, 993–999.
- McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984) *Nature*, **308**, 428–433.
- McLauchlan, J., Gaffney, D., Whitton, J.L. and Clements, J.B. (1985) *Nucleic Acids Res.*, **13**, 1347–1368.

- Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1983) *Nature*, **305**, 600–605.
- Mount, S.M. (1982) *Nucleic Acids Res.*, **10**, 457–472.
- Murphy, D., Brickell, P.M., Latchman, D.S., Willison, K. and Rigby, P.W.J. (1983) *Cell*, **35**, 865–871.
- Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature*, **263**, 211–214.
- Ryskov, A.P., Ivanov, P.L., Kramerov, D.A. and Georgiev, G.P. (1983) *Nucleic Acids Res.*, **11**, 6541–6558.
- Scheller, R.H., Constantini, F.D., Kozlowski, M.R., Britten, R.J. and Davidson, E.H. (1978) *Cell*, **15**, 189–203.
- Shepherd, J.C.W., McGinnis, W., Carrasco, A.E., De Roberts, E.M. and Gehring, W.J. (1984) *Nature*, **310**, 70–71.
- Taya, Y., Devos, R., Tavernier, J., Cheroutre, H., Engler, G. and Fiers, W. (1982) *EMBO J.*, **1**, 955–958.
- Tchurikov, N.A., Naumova, A.K., Zelentsova, E.S. and Georgiev, G.P. (1982) *Cell*, **28**, 365–373.
- Thomas, P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201–5205.
- Tosi, M., Young, R.A., Hagelbuechle, O. and Schibler, U. (1981) *Nucleic Acids Res.*, **9**, 2313–2323.
- Wharton, K.A., Yedvobnick, B., Finnerty, V.G. and Artavanis-Tsakonas, S. (1985) *Cell*, **40**, 55–62.
- Wickens, M. and Stephenson, P. (1984) *Science*, **226**, 1045–1051.
- Zuker, C. and Lodish, H.F. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 5386–5390.

Received on 6 May 1986