The EMBO Journal vol.3 no.5 pp.1003 – 1013, 1984

Drosophila maternal and embryo mRNAs transcribed from a single transcription unit use alternate combinations of exons

Alain Vincent, Peter O'Connell[1], Mark R. Gray and Michael Rosbash*

Department of Biology, Brandeis University, Waltham, MA 02254, USA

[1]Present address: Howard Hughes Medical Institute, 732 Wintrobe Building, University of Utah Medical School, Salt Lake City, UT 84132, USA

*To whom reprint requests should be sent

Communicated by V. Pirrotta

We have investigated the organization and transcription of several genes in *Drosophila melanogaster* which are clustered on an 18-kb cloned DNA fragment (c25) that maps at 99D on the cytogenetic map. Multiple mRNAs, transcribed from genes which lie adjacent to a ribosomal protein (rp49) gene, are present during oogenesis, embryogenesis, or both. At least five mRNAs are transcribed from one of these genes (EH8); three are zygotic transcripts, of which two are blastoderm stage-specific, whereas two others accumulate during oogenesis and are therefore maternal mRNAs. The complex transcription pattern of this gene indicates that alternate usage of protein-coding exons results in the production of different mRNAs with different coding capabilities during oogenesis and embryogenesis. The EH8 transcription unit is framed by genes actively expressed in the adult male fly; thus, the blastoderm stage-specific promoter may be silent although within a region of transcriptionally active chromatin.

Key words: Drosophila/embryogenesis/gene organization/transcription unit

Introduction

The correspondence between the number of genetic and chromomeric units in the *Drosophila* genome [a chromomeric unit corresponds to a band plus an adjacent interband in polytene chromosomes (Hogness *et al.*, 1975)], currently referred to as the one band – one gene hypothesis (reviewed by Lefevre, 1974), is still a controversial issue. A one-to-one correlation between complementation groups and chromomeric units has been described for several regions of the *Drosophila* genome, based on genetic and cytogenetic data. However, little is known about the number or density of genes that are detectable by transcript mapping (see discussion and references in Spierer *et al.*, 1983, Hall *et al.*, 1983). Data relevant to this question for many regions of the *Drosophila* genome will shed light on the one band – one gene hypothesis.

We have begun to study this problem using the recombinant phage c25, which contains the gene coding for the ribosomal protein 49 (rp49) (Vaslet *et al.*, 1980). c25 maps at 99D on the cytogenetic map, within a haploinsufficient region of the third chromosome (Lindsley *et al.*, 1972). This makes it very difficult at present to isolate newly induced mutations mapping within genes in c25. The observation of multiple R-loop structures formed between c25 DNA and *Drosophila* embryonic RNA suggested, however, that several genes transcribed predominantly during oogenesis and embryogenesis are clustered within this 18-kb DNA segment. As the rp49 gene is expressed throughout the entire fly life cycle (O'Connell and Rosbash, in preparation), we were interested in locating the adjacent embryo-oocyte genes relative to the rp49 gene. The results of this study indicate that a single gene or gene region (EH8) lies adjacent to the rp49 gene and gives rise to five distinct mRNAs with developmental profiles different from each other as well as from rp49 mRNA. Remarkably, two of the EH8 transcripts are blastoderm stage-specific, yet share sequences with other EH8 transcripts which are present during oogenesis, during later embryogenesis, or both.

It is known from the studies of developmental mutants that embryonic development requires the expression of a number of zygotic genes whose products are not provided, or are insufficiently provided, maternally (Nusslein-Volhard and Weischaus, 1980). Nevertheless, the molecular contribution of zygotic gene expression to early embryogenesis is largely unknown. Although recombinant DNA technology has permitted the isolation of a few genes expressed early in embryogenesis (Scherer *et al.*, 1981; Sina and Pellegrini, 1982; Lengyel *et al.*, 1983), none have been characterized in detail. Our results suggest that the division of transcripts into maternal and zygotic mRNA may prove more difficult than previously thought, since, as found here in the case of the EH8 gene, some transcripts present only during embryogenesis share sequences with maternal transcripts.

Results

Several transcription units are clustered in c25

We previously reported the isolation of a recombinant phage, designated c25, which contains the gene coding for ribosomal protein 49 (rp49); c25 maps on the right arm of the third chromosome at 99D on the cytogenetic map (Vaslet *et al.*, 1980). The position and direction of transcription of rp49 and an adjacent gene (gene 1) are shown in Figure 1A with a restriction map of c25 and its plasmid subclones, pDH4, pDH8, and pDHR4.2 (containing fragments H4, H8, and HR 4.2, respectively) (see Wong *et al.*, 1981, for details of the mapping of gene 1 and rp49).

Electron microscopic analysis of R-loop structures formed between c25 DNA and poly(A)$^+$ RNA from *Drosophila* embryos revealed several additional regions transcribed during embryogenesis (not shown). To map in more detail these transcripts, RNA blot hybridization experiments were undertaken. For the initial Northern blots, RNA from syncytial cleavage stage (0 – 1 h) and blastoderm stage (2 – 4 h) embryos was used (Figure 1B). As previously reported, only the 0.6-kb rp49 mRNA and the 1.7-kb gene 1 mRNA are visible

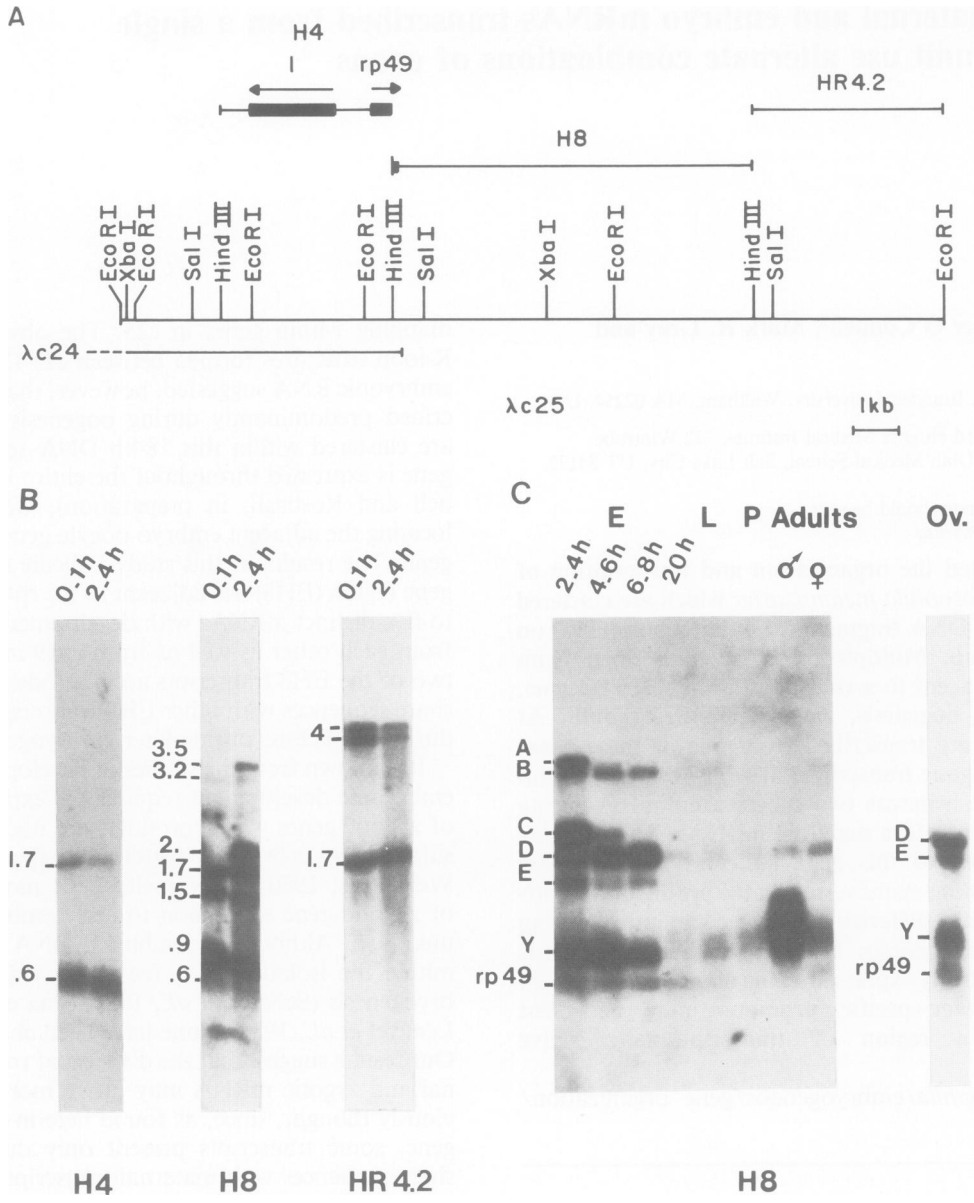


Fig. 1. Several transcription units are clustered in an 18-kb *Drosophila* DNA fragment. **(A)** Restriction map of the recombinant phage c25. The 4-kb *Hind*III-*Hind*III (H4), 8-kb *Hind*III-*Hind*III (H8), and the 4.2-kb *Hind*III-*Eco*RI (HR 4.2) fragments were subcloned in pBR322 or pUc8. The right end of the overlapping phage c24 is indicated. Gene 1 and rp49 are represented by black boxes and have been previously described (Wong *et al.*, 1981); the arrows denote the direction of transcription. The 3′ non-translated region of rp49 extends into H8 as indicated by the small black box at the left end of H8. **(B)** Identification of mRNA complementary to c25 subclones. 1 μg of poly(A)$^+$ RNA from the pre-blastoderm (0−1 h) or blastoderm (2−4 h) stage embryos was electrophoresed on formaldehyde agarose gels and blotted to nitrocellulose. The filter was successively hybridized to nick-translated pDH4, pDH8 and pDHR4.2. The lengths of the transcripts detected by each probe are indicated to the left of each panel. Size standards were yeast 28S and 18S ribosomal RNAs. **(C)** Developmental pattern of mRNA transcribed from H8. 1 μg of poly(A)$^+$ RNA from different developmental stages (E, embryos; L, larvae; P, pupae) or 5 μg of unfractionated RNA from ovaries of virgin females (Ov.) were electrophoresed and transferred to a nitrocellulose filter as in **(A)**. Hybridization was with nick-translated plasmid pDH8. The transcripts complementary to H8 were designated A to E and Y as noted on the left of the panel.

using the left-most c25 subclone pDH4 as a probe; these two RNAs are present both in pre-blastoderm and in blastoderm poly(A)$^+$ RNA. The pattern of blastoderm poly(A)$^+$ RNA complementary to the pDH8 probe is, however, much more complex. It consists of seven distinct mRNAs: the mRNA complementary to rp49, and six other mRNAs of lengths 3.5 kb, 3.2 kb, 2 kb, 1.7 kb, 1.5 kb and 0.9 kb, designated A through E and Y, respectively. Of these six RNA species, B, D, E and Y are also detected in pre-blastoderm poly(A)$^+$ RNA, while A and C are not. The right-most c25 subclone, pDHR4.2, detects three embryonic RNAs, one of which co-migrates with an RNA detected by pDH8 (mRNA D). Altogether, these RNA blots indicate that at least 11 transcripts complementary to c25 DNA are present during early embryogenesis.

We chose to focus on the seven transcripts complementary to H8 for several reasons. Firstly, the A and C transcripts are fairly abundant at the blastoderm stage but are absent from pre-blastoderm poly(A)$^+$ RNA. This suggests that they are early zygotic gene products first transcribed at the stage when cell determination occurs. To date, very few genes transcribed at blastoderm and not during oogenesis have been identified (Scherer *et al.*, 1981; Lengyel *et al.*, 1983), and none have been characterized in detail. Secondly, the sum of the sizes of the H8 transcripts (13.4 kb) exceeds the size of H8 DNA (8 kb), which is entirely single copy (data not shown). Thus,

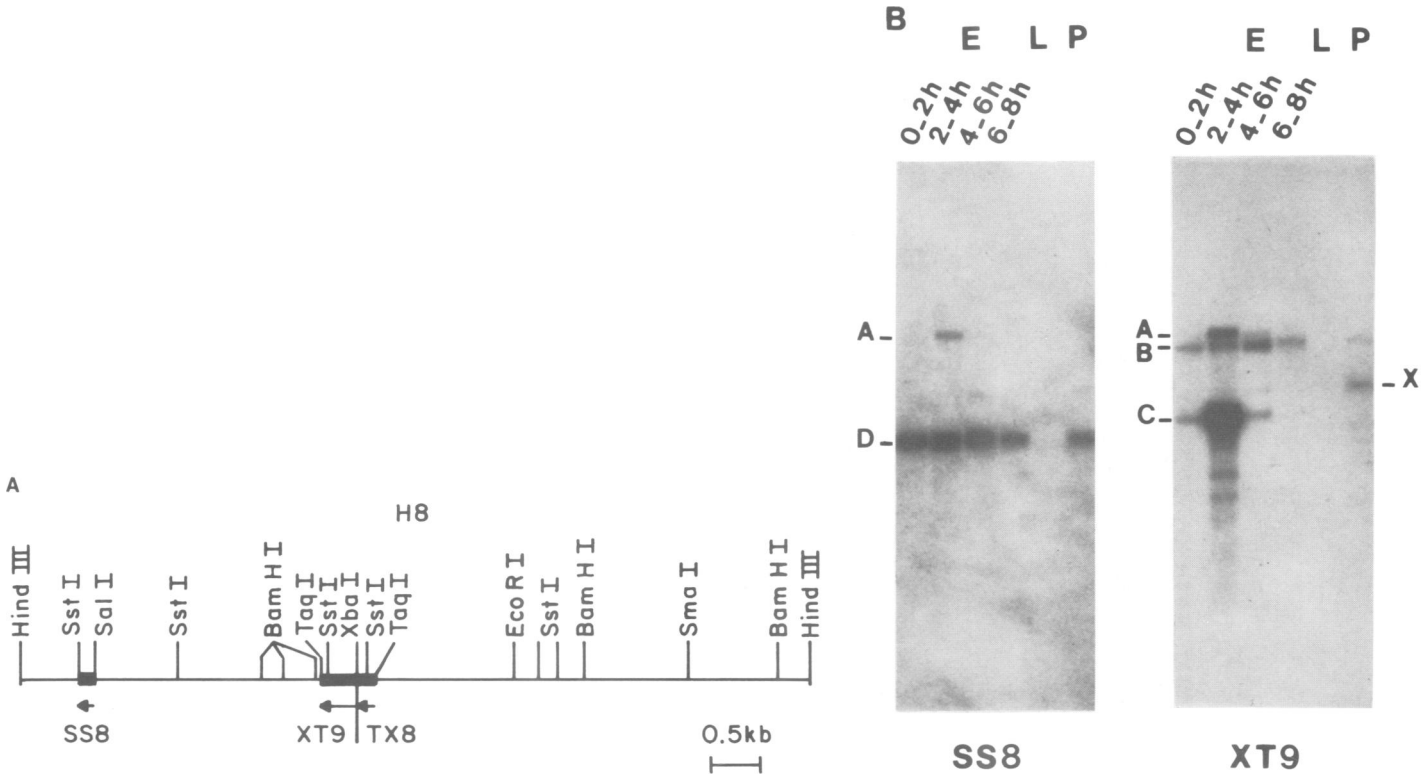


Fig. 2. Two small fragments in H8 are each transcribed into several distinct mRNAs. **(A)** Map of restriction enzyme cleavage sites of pDH8. The positions of the 3 H8 fragments inserted into M13 mp8 or mp9 (SS8, TX8 and XT9) are indicated by black boxes. The arrows indicate the direction of transcription of the mRNA complementary to these inserts. **(B)** 1 μg of poly(A)$^+$ RNA from different developmental stages was electrophoresed and blotted to nitrocellulose as in Figure 1. The filter was successively hybridized to radioactive SS8 and XT9 phage DNAs. The result obtained using TX8 was identical to that of XT9 (data not shown). The mRNAs complementary to each phage are designated by letters on the left of each panel. Letter X on the right refers to an mRNA detected in RNA isolated from 4–6 h old embryos (as a faint band on the original autoradiogram) and pupae.

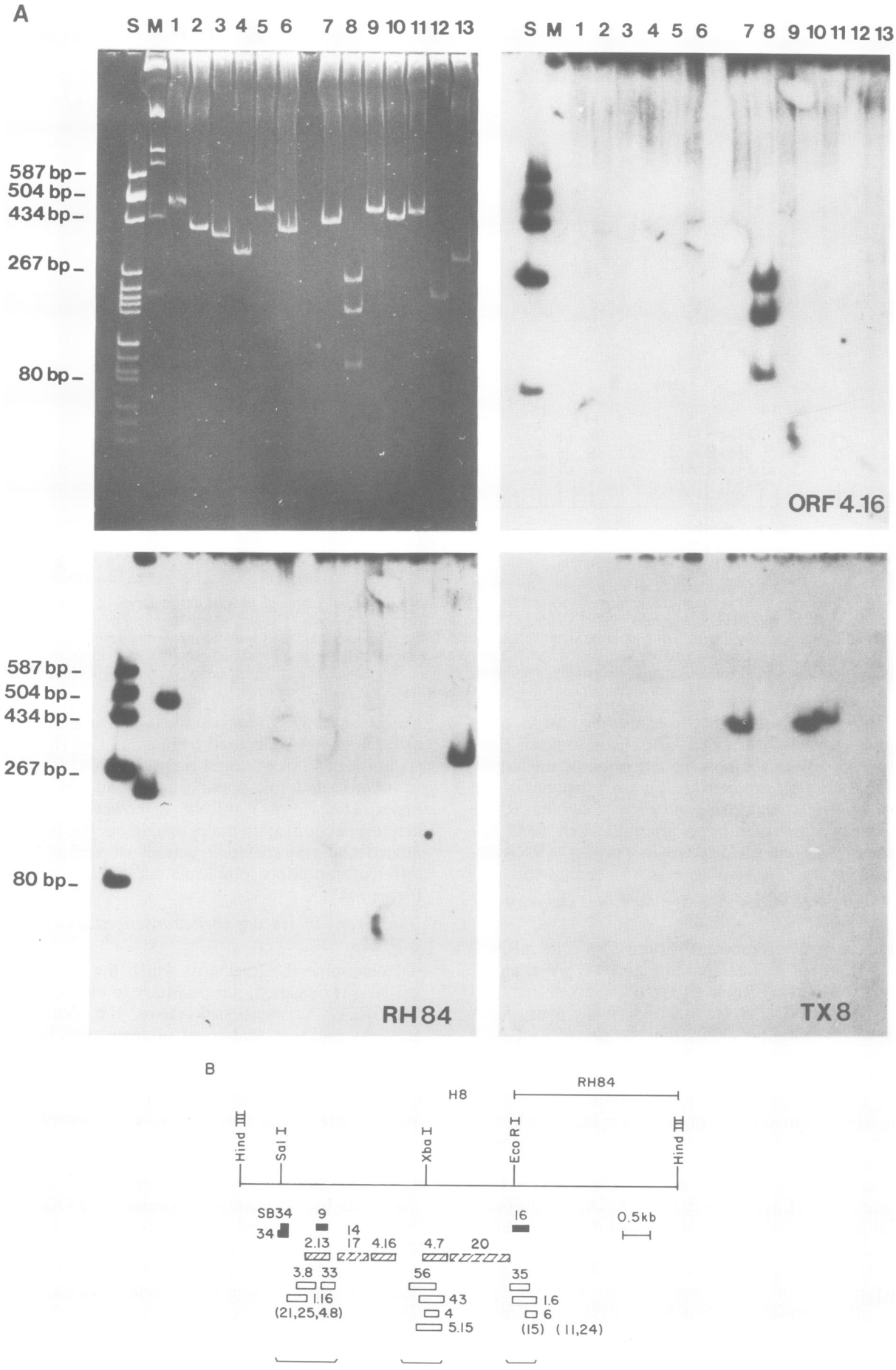
some of these mRNAs are partly transcribed from other DNA segments beyond the limits of c25, or some of these mRNAs overlap. With the possible exception of mRNA D, none of the transcripts are detected by any subclones of c25 DNA flanking H8, suggesting strongly that the RNAs overlap. Thirdly, these transcripts are most likely *bona fide* mRNAs since they are all detected in poly(A)$^+$ RNA extracted from embryo polyribosomes (data not shown).

mRNAs transcribed from H8 accumulate in a stage-specific manner

The characteristic developmental pattern of all the mRNAs transcribed from H8 is illustrated in Figure 1C. A, B and C are not detected in ovary RNA. (It should be noted that total RNA and not poly(A)$^+$ RNA was used in the ovary RNA lane. This excludes the possibility that the presence of A, B and C in embryo poly(A)$^+$ RNA is due to an adenylation event rather than *de novo* transcription.) A and C are blastoderm stage (2–4 h)-specific transcripts. Indeed, they are not visible in RNA from oocytes or from pre-blastoderm embryos and are almost undetectable after 4 h of embryogenesis. B and E are present in syncytial cleavage, blastoderm and gastrula stages RNA, but are undetectable in late embryo RNA. Since mRNA E is present in ovary RNA, it is probably a maternal transcript which is accumulated during oogenesis and which disappears during embryogenesis. B is not detected in oocytes but is visible in 0–1 h embryos (Figure 1B) prior to the major onset of zygotic transcription. D is the most abundant mRNA transcribed from H8 in RNA from unstaged embryos; it is also present at relatively low levels in late embryos and at later stages of the fly life cycle. D and Y are almost certainly synthesized during oogenesis, since they are detected in RNA prepared from ovaries of virgin females. Y is abundant in ovaries and early embryos, rare between late embryonic and mid-pupal stages, and abundant in adult males. The size of Y mRNA is, however, larger and more heterogeneous in adult males than in oocytes or embryos. The limited and very different periods of accumulation for the different transcripts complementary to H8 are summarized in Figure 6B.

Two exons in H8 are each transcribed into several distinct mRNAs

To determine the level(s) at which the expression of these mRNAs is regulated, it is necessary to map the transcripts on H8 DNA. As mentioned above, the total size of these mRNAs indicates that some of them may be overlapping. Also, preliminary Northern blot analysis with isolated fragments suggested that more than one mRNA is complementary to small subregions of H8. To determine whether mRNAs are indeed overlapping, small restriction fragments from H8 were subcloned into the M13 phage (Messing and Vieira, 1982) and used as strand-specific probes for RNA blots. The *Sal*I-*Sst*I (103 bp), *Taq*I-*Xba*I (180 bp) and *Xba*I-*Taq*I (365 bp) fragments, designated SS8, TX8 and XT9, respectively, were cloned in M13mp8 and mp9 vectors (Figure 2A). Southern blot analysis under standard stringency conditions confirmed that these sequences were single copy in c25 and in the *Drosophila* genome (data not shown). Developmental Northern blots were hybridized to each of these small labelled subclones. The blots show that SS8 (in M13mp8) and XT9 (in M13mp9) are transcribed in one and


A
S M 1 2 3 4 5 6 7 8 9 10 11 12 13
587 bp
504 bp
434 bp
267 bp
80 bp
ORF 4.16
RH84
TX8
B
H8
RH84
Hind III
Sal I
Xba I
Eco R I
Hind III
0.5kb
SB34
34
9
14
2.13
17
4.16
4.7
20
16
3.8
33
56
35
1.16
43
1.6
(21,25,4.8)
4
6
5.15
(15)
(11,24)

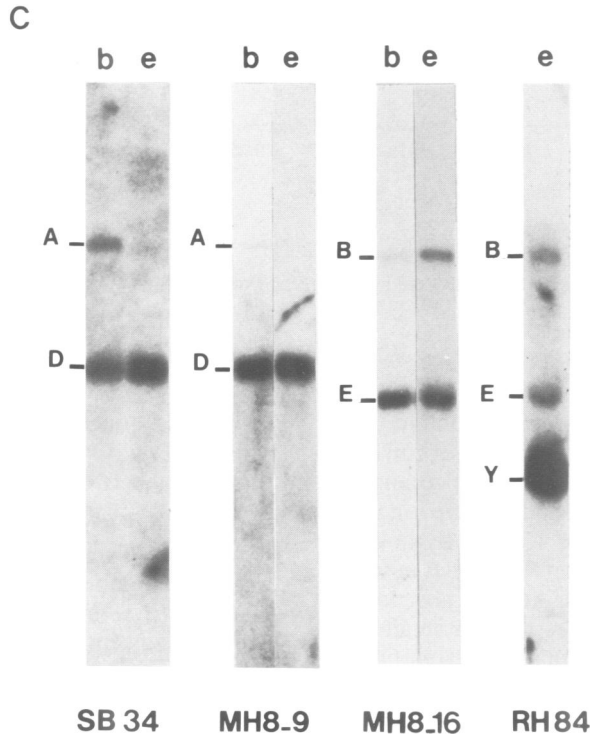


Fig. 3. The positions of open reading frame (ORF) DNA fragments suggest that there are two or three main coding exons in mRNAs A–E. (**A**) Open reading frame DNA fragments were mapped on H8 using a modification of the Southern blot method. The plasmids containing ORFs were digested with *Bam*HI and electrophoresed on polyacrylamide gels (**left upper panel**); the DNA was denatured in the gel by a 15 min incubation in 0.4 M NaOH, 1 M NaCl prior to transfer (for 12 h at 100 mA in 20 mM Tris-HCl pH 7.4, 1 mM EDTA) to nitrocellulose by electroblotting. The filter was used for successive hybridizations to labelled phage or plasmid DNA probes as indicated on each panel (ORF 4.16, RH84, TX8). S; *Hae*III-cut pBR322 fragments used as mol. wt. standards; M: *Sst*I-digested pDH8 plasmid DNA; 1–13: ORF clones 1.6, 1.16, 21, 25, 2.13, 3.8, 4.7, 4.16, 4.8, 5.15, 56, 33 and 16. (**B**) The position (±0.2 kb) of ORF DNA fragments mapped on H8 is indicated. Each ORF fragment is indicated by a numbered box. The black boxes refer to fragments subcloned in M13mp8. The hatched boxes refer to the additional clones used as probes for Northern blots. Three of these clones (ORF 14, 17 and 20) map in the region indicated by the boxes with dotted lines, but have not been mapped with precision. Numbers in parentheses refer to additional ORF clones mapping in the designated region. Brackets indicate the positions of the three clusters of ORF clones. The 3-kb *Hind*III-*Eco*RI fragment RH84 was independently cloned in the plasmid pUc8. (**C**) Hybridization of H8 ORF clones with embryonic RNA. 1 μg of poly(A)$^+$ RNA extracted from blastoderm stage (b) or 0–16 h embryos (e) was electrophoresed and transfered to nitrocellulose as in Figure 1. Several duplicate filters were used two or three times for hybridization. The letters on the left of each panel identify the mRNA complementary to the probe indicated on the bottom (see Figure 4B).

the same direction, opposite to that of the rp49 gene (Figure 2A, B and Figure 1). The results of all experiments using TX8, in M13mp8 (not shown), are identical to those using XT9.

In each case more than one mRNA species is detected with these short probes (Figure 2B), indicating that several mRNAs are transcribed from the same H8 DNA fragments: A is complementary to both SS8 and XT9, B and C are complementary to XT9 but not to SS8, D is complementary to SS8 but not to XT9. (An additional very low abundance mRNA, designated X on the figure, is also detected in 4–6 h embryos and pupae. This mRNA species is not considered further due to a lack of relevant data.) The results suggest that the mRNAs A, B, C and D might be, at least in part, the result of alternative RNA processing. The DNA sequence of the XT9 and TX8 fragments and surrounding DNA contains a single open reading frame of at least 540 bp in the vicinity of the *Xba*I site (not shown). The DNA sequence on each side of the *Sal*I site in SS8 contains one open reading frame of 270 bp in the direction of transcription of mRNAs A and D (see Figure 5). These and other data shown below suggest that each small region (SS8 and TX8 + XT9) is part of a protein-coding exon and is shared by more than one transcript.

Mapping of open reading frame DNA fragments in H8

Since the data indicate that several mRNAs are transcribed from the same DNA fragments, the systematic approach of open reading frame DNA (ORF) cloning (Gray *et al.*, 1982) was used to locate protein-coding regions and their complementary mRNAs. To that end, randomly sheared H8 DNA fragments were cloned by insertion into the cloning and expression vector pMR100 (Gray *et al.*, 1982). Two size classes of DNA fragments (~250-bp and 425-bp average length) were used. Large ORF DNA fragments are a better indication of protein-coding DNA than small fragments; nevertheless, since the size of protein coding exons in H8 was *a priori* unknown, we also cloned smaller fragments in pMR 100 to identify and isolate potential short exons. ORF DNA clones, selected from each size class, were mapped on H8 DNA using blots of H8 DNA cut with various restriction enzymes (data not shown) and blots of the ORF fragments probed with labelled H8 restriction fragments and other ORF fragments (Figure 3A).

The sizes of some cloned ORF fragments are shown in Figure 3A and Table I. Their distribution on H8 DNA shows the clustering of many clones at three different locations (indicated by brackets in Figure 3B); of 25 clones, 9, 6 and 5 map in the vicinity of the *Sal*I, *Xba*I and *Eco*RI unique sites, respectively. The positions of the two left-most ORF clone clusters correlate well with the positions of exons as indicated by the results of the Northern blot experiments using the H8 restriction fragment probes (Figure 2) and with the positions of the sequenced open reading frames surrounding the *Xba*I

Table I. Northern blot analysis using various H8 subclones

H8 subclones[a]	RH84	ORF16 MH8−16[b]	ORF20	ORF4−7	TX8[b]	XT9[b]	ORF4−16	ORF17	ORF14	ORF9 MH8−9[b]	ORF2−13	SB34[b]	SS8[b]
Insert size (kb)	3.0	0.27	0.20	0.43	0.18	0.37	0.45	0.27	0.28	0.18	0.45	0.14	0.10
Complementary[c] mRNA	B,E,Y	B,E	B,E	A,B,C		A,B,C		A,B,C		(A)[d],D		A,D	A,D

[a]H8 subclones are listed from the 5′ to the 3′ end of the EH8 gene in the direction of mRNAs A−E: clones with partly overlapping sequences are represented in a single column.
[b]The underlined clones are M13 recombinant phages.
[c]The mRNAs detected by each of the listed clones used as probes on Northern blots are represented by letters (A to E and Y) in each column.
[d]Parentheses indicate a relatively faint signal.

and *Sal*I sites; both of these bracketed regions contain ORF clones with inserts >0.4 kb, reinforcing the assertion that SS8 and XT9 fragments are part of large protein-coding exons. The mapping of ORF clones suggests that there is a third H8 exon mapping to the right of the R1 site (right-most bracket in Figure 3B). The lack of many ORF clones at intermediate positions on H8 may be due to the absence of protein-coding exons at these locations or to their small size, leading to an under-representation of this DNA among the selected ORF clones. It is also possible that some open reading frames are not identified by this procedure, perhaps because their polypeptide sequences are not stable in the bacterial host cell.

Hybridization of H8 ORF clones with embryonic RNA

To determine further which mRNAs are transcribed from which regions of H8, selected ORF clones were used as small probes for Northern blots of embryo poly(A)$^+$ RNA. Most ORF clones were nick-translated directly. The inserts from ORFs 9, 16, 34 and the insert from ORF 34 cut with *Sal*I, were subcloned in both orientations in the single-strand M13mp8 vector to generate short (140−270 nucleotides) strand-specific probes of high sensitivity (respectively referred as to MH8-9, MH8-16, MH8-34 and SB34). Some of the Northern blot hybridizations are shown in Figure 3C. The results with all clones are summarized in Table I.

ORF clones 16 (or the single-stranded equivalent MH8-16) and 20 detect RNA transcripts B and E, suggesting that these two mRNAs share an exon. Only one orientation of the single-strand probes made from MH8-16 detects both the B and E transcripts, indicating that both are transcribed in the same direction as mRNAs A, C and D. The *Eco*RI-*Hind*III 3-kb fragment cloned in the pUc8 plasmid vector (designated RH84 in Figure 3B) detects only transcripts B, E and Y, indicating that the 5′ ends of mRNAs A, C, and D are not transcribed from H8 sequences to the right of the RI site on the restriction map. This fragment was also cloned into the pEMBL9 (+) vector (Dente *et al.*, 1983) and used as a single-stranded probe. The strand anti-complementary to mRNA A to E is complementary to mRNA Y (not shown). Thus, mRNAs A to E and Y are transcribed in opposite directions from independent transcription units, henceforth designated genes EH8 and Y.

ORF clones 4.7, 4.16 and 17, which are distributed over a 2 kb region around the *Xba*I site in the center of H8, detect mRNAs A, B and C. The data suggest that a major protein-coding exon shared by these three mRNAs might be as large as 2 kb. As observed with SS9 (Figure 2B), ORF 34 (or the derivative M13 phage SB34) detects mRNAs A and D, indicating an exon shared by only these two mRNAs. ORFs 2-13 and 9 (or the M13 phage MH8-9) hybridize strongly to mRNA D but weakly to mRNA A, suggesting only limited homology between the two ORF fragments and mRNA A.

Mapping the ends of the exons

The ends of some of the exons were mapped using mRNA to protect end-labelled DNA fragments from digestion by S1 nuclease (Berk and Sharp, 1978) (Figure 4). The position of a splice site at approximately the position of ORF 9 was identified by an S1 nuclease mapping experiment using a 5′ end-labelled *Sal*I-*Hind*III restriction fragment (see Figure 4A). As assayed on denaturing gels, two major fragments with lengths of 740 and 1300 nucleotides are protected from S1 nuclease digestion by hybridization with embryonic mRNA. A third weaker band, 900 nucleotides long, is also observed, suggesting that another minor splicing event might take place within this exon. The more intense 1300 nucleotide band probably corresponds to protection by the more abundant mRNA D. This interpretation is strengthened by the increased intensity of this band in the experiment using RNA enriched in smaller mRNA by fractionation on a sucrose gradient (Figure 4B, lanes 2 and 3). An intense band of 1300 nucleotides is visible when the S1-resistant hybrids are electrophoresed in non-denaturing conditions (Figure 4B). The correspondence between the sizes observed on denaturing and non-denaturing gels suggests that mRNA D is not transcribed from upstream sequences in H8. The 740-bp band (or possibly the 900-bp band) probably represents a fragment protected by mRNA A. These S1 data, and the blotting data for the left-hand end of H8 (i.e., the fact that probes to the right of the *Sal*I site hybridize well to D but poorly to A), are best accounted for by a splice site of mRNA A mapping within the coding sequence of mRNA D (see Figure 6A).

S1 mapping experiments were also done with a fragment 5′-labelled at the *Xba*I site, a position which is complementary to mRNAs A, B and C (see Figure 2B). Embryonic RNA protects three 5′ *Xba*I-labelled fragments, with sizes of 1500, 550 and 450 nucleotides, respectively (Figure 4B, lane 5). Based on the relative intensities of these bands, the relative amounts of the three RNAs (Figure 2B), and their respective sizes, these bands are probably due to mRNAs A, B and C respectively. Additional S1 experiments, using a 3′-labelled *Xba*I-*Hind*III fragment, indicate that the exon shared by mRNAs A, B and C extends without interruption for ~1.3 kb downstream from the *Xba*I-site (data not shown).

With a fragment 5′-labelled at the *Eco*RI site (a position complementary to mRNAs B and E), a single fragment of 300 nucleotides is protected from S1 nuclease digestion by embryo RNA or by embryo RNA enriched in small RNA (e.g., mRNAs D, E, Y; Figure 4B, lanes 7 and 8). A single 500-bp

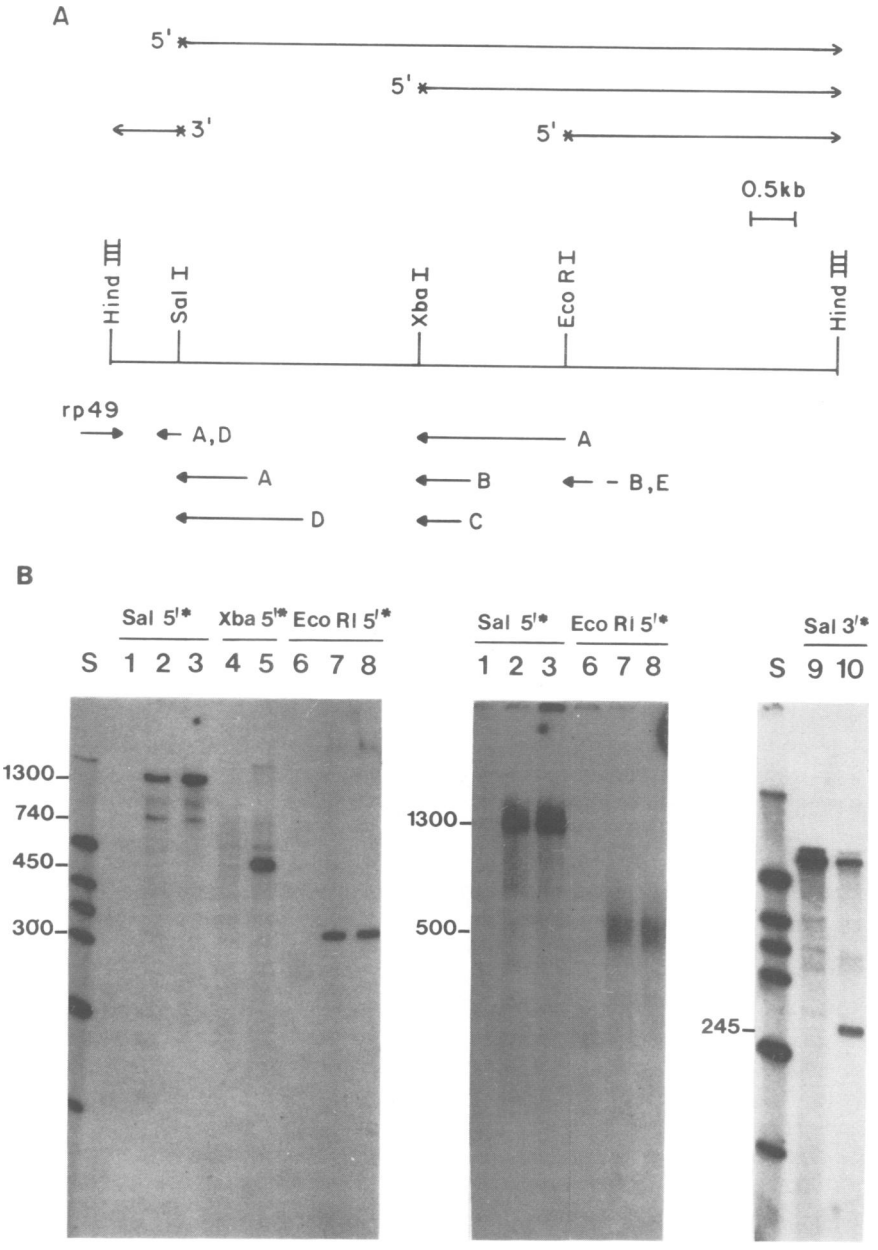


Fig. 4. S1 nuclease mapping of mRNA transcribed from H8. **(A)** The diagram indicates the strategy for mapping the positions of exons in H8 transcripts and summarizes the protected DNA segments. The positions of ^{32}P-labelled ends are indicated by 5'* or 3'* for each DNA fragment. Solid lines with arrows indicate the length and direction of transcription of individual RNA segments as deduced from the protection of DNA from S1 digestion (Figure 6B) and the Northern blot data summarized in Table I. **(B)** The protected DNA fragments were analysed on 4.5% acrylamide gels containing 7.5 M urea (**left** and **right panels**) or without urea (**central panel**); their sizes are indicated on the left (in nucleotides). *Hinf*I-digested pBR322 was used as size standards (S). **Lanes 1, 4, 6, 9**: no mRNA; **lanes 2, 5, 7, 10**: embryonic (0–16 h) poly(A)$^+$ RNA; **lanes 3** and **8**: embryonic poly(A)$^+$ RNA preparation enriched in the smaller mRNAs D and E (as compared to the longer A and B mRNAs) by size fractionation.

band is observed upon electrophoresis in non-denaturing conditions, suggesting that an intervening sequence is located 300 bp upstream from the *Eco*RI site. The fact that only a single band is visible is consistent with the notion that the exon in the vicinity of the *Eco*RI site may begin with a single 5' junction (used by both RNAs), 300 bp upstream of the *Eco*RI site. Experiments using a 3'-labelled *Eco*RI-*Hind*III fragment indicate that this exon ends at two different positions, 220 bp and 900 bp downstream from the *Eco*RI site (data not shown).

The putative 3' end of the exon shared by mRNAs A and D was identified by S1 nuclease mapping using the *Sal*I-*Hind*III fragment 3'-labelled at the *Sal*I site (Figure 4B, lanes 9 and 10). A single fragment of 245 nucleotides was protected by RNA. Of the RNAs complementary to H8, only rp49 mRNA is detected by Northern blot hybridization with cloned DNA extending 16 kb to the left of H8 (phage c24, Figure 1); it is likely, therefore, that this 245-bp fragment is the 3'-terminal part of the 3'-terminal exon of mRNAs A and D.

The sequence of the 3' end of the EH8 gene complex

The results from the S1 nuclease mapping experiments described above indicate that mRNA D is transcribed from a 1.55-kb (1.3 + 0.25) exon in H8. The size of the RNA, measured by agarose gel electrophoresis, is 1.7 kb. Considering the average length of the poly(A) tail of mRNA to be 50–150 nucleotides, it appears that mRNA D may be entirely

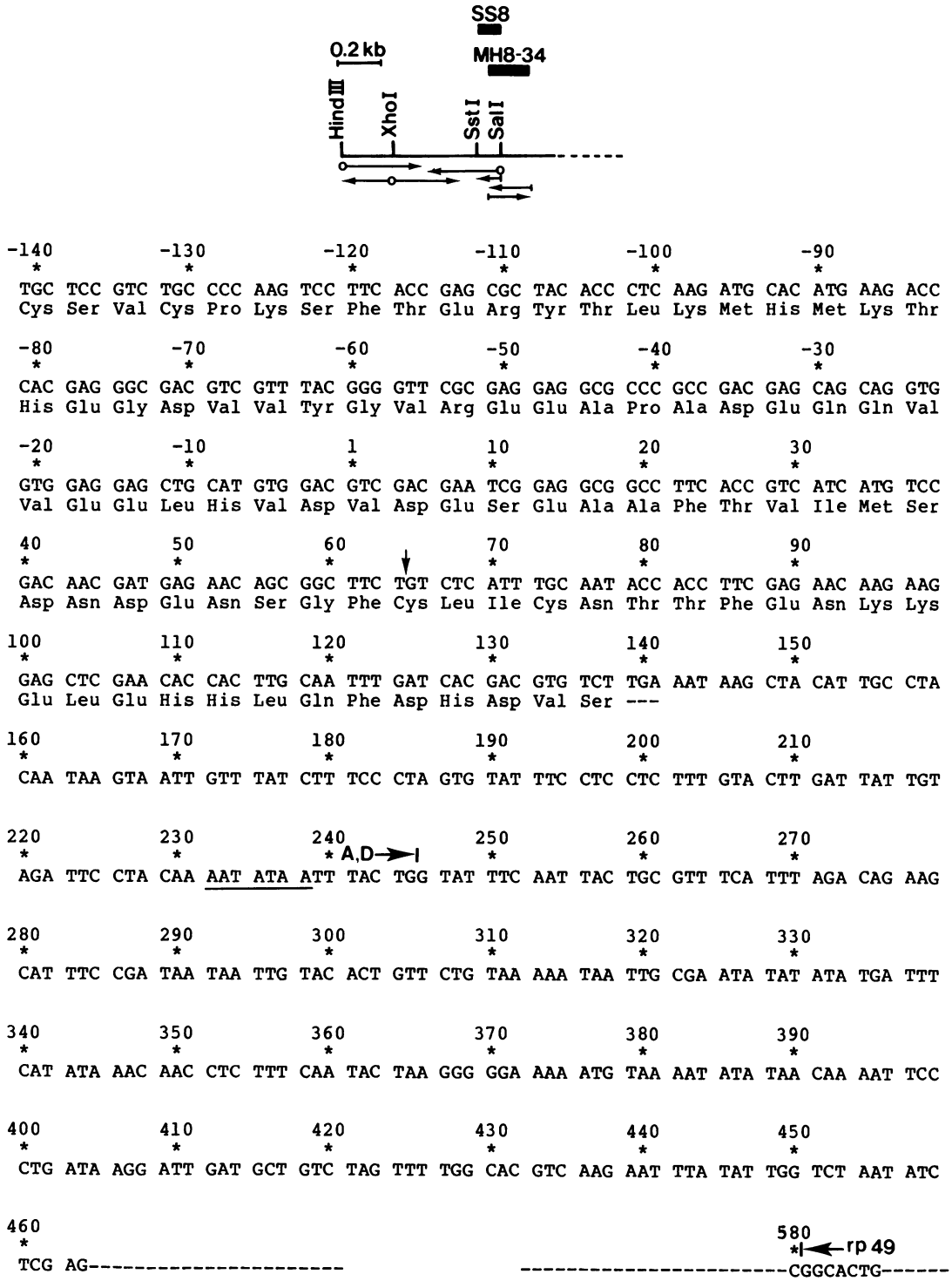


Fig. 5. DNA sequence of the 3′ end of mRNAs A and D. The strategy for sequencing this region is indicated by a graph on the top left of the figure; fragments marked ○→ were sequenced using the Forward-Backward method (Sief *et al.*, 1980); fragments marked ⊢→ were sequenced using the dideoxy terminator method (Sanger *et al.*, 1977). The sequence presented arbitrarily ends at the *Xho*I site (X). The sequence shown is that of the transcribed strand beginning at position −141. Position 1 is at the H8 *Sal*I restriction site (see Figure 4A). The open reading frame that is most probably used ends at the TGA at position 139. The arrow at position 245 indicates the end of mRNAs A and D as indicated by an S1 nuclease mapping experiment (Figure 4). The putative polyadenylation signal AATATAA at position 230 is underlined. The ORF 34 fragment starts at nucleotide −141 and ends at nucleotide 65 (indicated by an arrow). Transcription of the rp49 gene ends at position 580 on the opposite strand (O'Connell and Rosbash, in preparation).

transcribed from the left-hand end of H8 DNA. This and additional Northern blot data (not shown) suggest further that the 1.7-kb mRNA complementary to the upstream region (HR4.2, Figure 1B) is distinct from mRNA D. To obtain more information on the 3′ ends of the A and D mRNAs, the DNA sequence around the *Sal*I site was determined. The sequence of a 604-bp fragment (sequence SBXHO), starting 141 nucleotides to the right of the *Sal*I site at position −141 and ending at position 464, shortly beyond the *Xho*I site, is shown (Figure 5). Two different open reading frames extend beyond the *Sal*I site, ending respectively, at positions 48 and 139 (TGA) of the sequence (the S1 experiment in Figure 4 in-

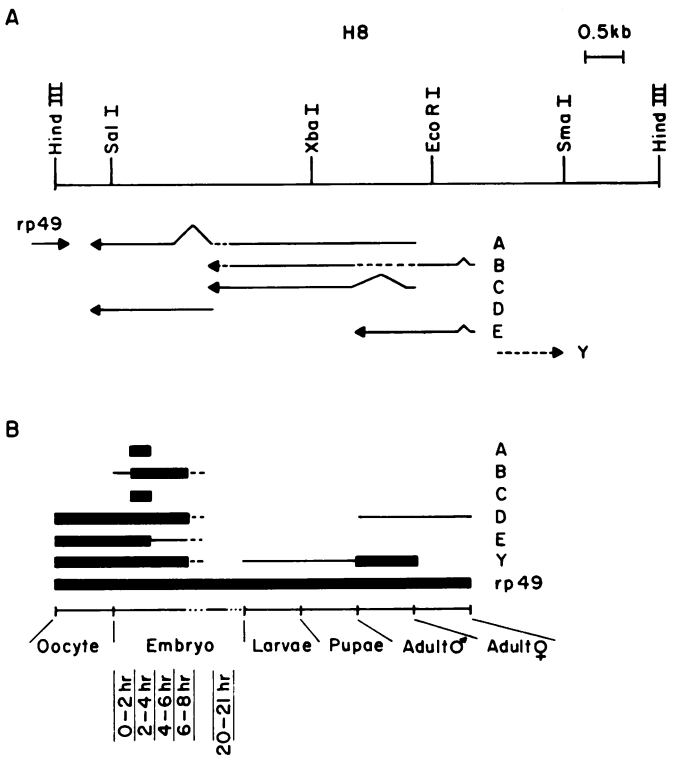


Fig. 6. Summary of the organization and time of expression of the mRNAs transcribed from H8. (**A**) The arrangement of mRNAs transcribed from H8 was deduced from electron micrographs of R-loop structures, Northern blot results and S1 nuclease mapping data. The solid lines indicate mapped exons; the dashed lines indicate regions for which no subclones were obtained (see Figure 3B); the bent lines indicate probable introns. (**B**) Time of accumulation of each mRNA transcribed from H8. Each stage examined is listed on the abscissa (distance is arbitrary with respect to developmental times). Thick and thin lines indicate a relatively strong or a weak signal, respectively, on a developmental Northern blot (Figure 1).

dicates that A and D end at position 245 ± 5). Although both open reading frames might be used *in vivo*, previous analysis of codon frequencies in *Drosophila* (O'Connell and Rosbash, in preparation) predicts that at least the open reading frame which ends at the TGA at position 139 (the one used in Figure 5) is used. Downstream 91 bp from this TGA, at position 230, is the sequence AATATAA, similar to the consensus polyadenylation signal (Fitzgerald and Shenk, 1981); this is located 15 ± 5 bp from the 3′ end of the mRNA as defined by the S1 protection experiments (Figure 4). The rp49 gene is transcribed on the opposite strand and the 3′ end of rp49 mRNA maps to position 580 on this sequence (O'Connell and Rosbash, in preparation). If the EH8 gene was to extend downstream of H8 DNA, the rp49 gene would map within an EH8 intron. Combined with the fact that we detect no EH8 transcript complementary to phage c24 (Figure 1), the data presented here require that the EH8 gene ends between positions 139 and 580 as shown in Figure 5. While transcripts A and D terminate at position 245, it is likely that transcripts B, C, and E use alternative polyadenylation sites located upstream, since these transcripts are not detected by probes made from short fragments on either side of the *Sal*I site, SS8 and SB34 (Figures 2 and 3), or by phage c24 as previously indicated. A summary of the results of the experiments which have mapped the various EH8 transcripts is shown in Figure 6A.

Discussion

The organization of the EH8 gene transcripts

The existence of at least five EH8 mRNAs (A–E), differing in sequence and time of accumulation, makes the organization of this *Drosophila* gene unique among those described to date. We tentatively propose that these mRNAs are generated through the use of alternate transcription start sites, alternate RNA processing of multiple protein-coding exons, or both. These exons have been defined by several criteria. As an example we consider here the DNA in the vinicity of the *Xba*I site in H8 (Figure 6A). Three different mRNAs (A, B and C) are detected when short DNA fragments which surround the *Xba*I site are used as strand-specific probes on Northern blots of RNA (Figure 2A, B). Although this result could be explained by the presence of multiple mini-exons, each transcribed into a single mRNA species, the small sizes of the probes make this unlikely. Also, the DNA fragment including the *Xba*I site contains a single open reading frame of at least 540 bp in the transcribed direction (data not shown). All three mRNAs are similarly detected by additional cloned fragments (see Table I) which distribute over 1.5 kb of DNA. It is likely, therefore, that mRNAs A, B and C come from the same exon (~2 kb) in this central region. Similar criteria were used to define other protein-coding exons shared downstream by mRNAs A and D, and upstream by mRNAs B and E. The presence of a short additional exon 5′ to the *Xba*I site in mRNA C is suggested by the analysis of S1-resistant hybrids on non-denaturing gels (not shown). A definitive interpretation of the S1 mapping data and a definition of the proteins encoded by the EH8 mRNAs will require a rigorous analysis of full-length cDNA clones. Nevertheless, the results to date suggest strongly that each EH8 mRNA is composed of a different combination of protein-coding exons.

Particularly interesting is mRNA B. This mRNA shares a 5′ exon with mRNA E. Unlike mRNA E, however, it is also complementary to the exon mapping in the vicinity of the *Xba*I site (Table I). Also unlike mRNA E, mRNA B is not detectable in ovaries. It is, however, detectable before blastoderm (Figures 1C and 6), and perhaps represents one of the rare mRNAs transcribed in pre-blastoderm embryos (McKnight and Miller, 1976). It is possible that mRNAs E and B are alternate processing products from a single primary transcript and that the transition from E to B occurs at fertilization. Alternate promoter usage (Benyajati *et al.*, 1983; Schibler *et al.*, 1983) and alternate processing of transcripts (Hagenbuchle *et al.*, 1981; Rozek and Davidson, 1983) have been reported to contribute to the developmental control or tissue-specific expression of a single gene. We propose the existence of three separate promoters: one for the transcription of mRNAs B and E, one for mRNA D, and one for mRNAs A and C (Figure 6A). Thus, both transcriptional and post-transcriptional regulatory steps are used to control the production of the various EH8 mRNAs.

Gene autonomous regulation

The times of accumulation of the EH8 and rp49 gene transcripts argue strongly that their expression is not coordinate. S1 nuclease mapping of EH8 transcripts (this report) and rp49 transcripts (O'Connell and Rosbash, in preparation) show that they are transcribed in opposite directions and that their 3′ ends are separated by only 335 bp of DNA. As described above, it is almost certain that this short, AT-rich

interval contains the 3′ end of the EH8 transcription unit. Unfortunately, the position of the 5′ end of EH8 is not known with equal certainty. The fact that none of the mRNAS transcribed from EH8 are detected by Northern blot hybridization with a 4.2-kb DNA fragment mapping to the right of H8, suggests that these mRNAs are not transcribed from DNA upstream of H8. Furthermore, a different gene (Y) maps to the right of the EH8 gene within H8. The EH8 and Y genes are transcribed in opposite directions, indicating that they are two separate genetic units. Another gene (gene 1), transcribed during oogenesis and possibly during embryogenesis, is located 1 kb to the left of the rp49 gene and transcribed in a direction opposite to that of the rp49 gene (Wong *et al.*, 1981). From the developmental patterns of transcripts from each of the 4 genes mapped within c25 (gene 1, rp49, EH8 and Y), we conclude that the regulation of expression of each of these clustered genes is autonomous.

Blastoderm-specific mRNA

One of the most interesting features of the EH8 gene is that two of its transcripts (A and C) are blastoderm-specific. A considerable amount of attention has focussed on genes expressed at this period of embryogenesis in *Drosophila*, as cell determination is generally thought to take place between nuclear migration (1.5 h) and gastrulation (4 h) (Wieschaus and Gehring, 1976; Nusslein-Volhard and Weischaus, 1980). The determination of blastoderm cells appears to be a function of their position but the molecular nature of the relevant positional information is completely unknown. Intense zygotic transcription starts at ~2 h following syngamy (Zalokar, 1976; Anderson and Lengyel, 1979), and there is strong genetic evidence that it contributes to the establishment of the segmentation pattern of the embryo (Nusslein-Volhard and Wieschaus, 1980).

Using differential screening of cDNA libraries made from blastoderm poly(A)$^+$ RNA, two laboratories have isolated genes expressed uniquely at the blastoderm stage (Scherer *et al.*, 1981; Lengyel *et al.*, 1983). [Incidentally, of the five clones isolated by Lengyel and colleagues, one contains the EH8 gene (Lengyel *et al.*, 1983 and personal communication)]. The number (five and four, respectively) of clones isolated by each of these groups is lower than expected from the number of moderately abundant proteins synthesized only at early embryogenesis (Trumbly and Jarry, 1983), or from the number of blastoderm-specific mRNA sequences (>44) predicted from RNA-cDNA hybridization analysis (Arthur *et al.*, 1979). The overlapping nature and pattern of expression of the EH8 transcripts offer an explanation for these somewhat contradictory findings, i.e., overlapping transcripts reduce the sensitivity of differential screening for stage-specific genes. (In the case of EH8, we presume that the intense signal from mRNA C at blastoderm is sufficient to allow differential detection.) Thus, blastoderm-specific genes may be somewhat more numerous than indicated by the differential screening done to date. If numerous maternal and zygotic transcripts overlap, this may also require a redefinition of genes whose expression is specific to embryogenesis and bring a new perspective on the gene expression reprogramming events that occur at fertilization.

Materials and methods

Subcloning

The subclones pDH4 and pDH8 were made by insertion of 4-kb and 8-kb c25 *Hind*III fragments (H4 and H8) into the plasmid pBR322 (Wong *et al.*, 1981). The pDHR4.2 subclone is a 4.2-kb *Eco*RI-*Hind*III c25 fragment (HR4.2) inserted into the plasmid pUc8 (Vieira and Messing, 1982). The ORF clones were constructed in the open reading frame cloning and expression vector pMR 100 (Gray *et al.*, 1982), using sonicated H8 DNA. In one set of ligations, BAL 31 nuclease was used to generate blunt-ended fragments; 200–300-bp fragments were selected for cloning. In another, a combination of S1 nuclease digestion and T4 polymerase repairing was used to make blunt ends. In this experiment, 400–500-bp fragments were cloned. Plasmid DNA was prepared by standard CsCl ethidium bromide centrifugation methods or from minilysates as indicated in Gray *et al.* (1982). To analyze the ORF inserts, the DNA extracted from 1 ml of cells was re-suspended in 25 μl of TBE buffer. The insert was cut out by digestion of 5 μl of DNA with 2 units of *Bam*HI in a final volume of 20 μl. The digested DNA was run on a 10% polyacrylamide gel in TBE buffer and electroblotted to a nitrocellulose filter for hybridization. To reclone these inserts into M13mp8 each insert was purified by electroblotting to DEAE cellulose paper, eluted and concentrated by ethanol precipitation (Gray *et al.*, 1982).

Subcloning into M13

SS9 was constructed by inserting a H8 *Sst*I-*Sal*I fragment into M13mp8 digested with *Sma*I and *Sal*I (Messing and Vieira, 1982). XT9 and TX8 were constructed by inserting the *Taq*I-*Xba*I and *Xba*I-*Taq*I fragments into M13mp8 and mp9 digested with *Acc*I and *Sma*I. The fragments were made blunt-ended at the *Sst*I and *Xba*I sites using the Klenow enzyme in conditions described by Wong *et al.* (1981). The open reading frame DNA fragments removed from pMR100 by digestion with *Bam*HI were recloned into M13mp8 cut with the same enzyme. In one case (ORF 34), the insert was recut with *Sal*I and cloned in M13mp8 cut with *Sal*I and *Bam*HI. Each construction was verified by sequencing.

RNA isolation

Drosophila melanogaster (Oregon R strain) was used in this investigation. Staged embryos were collected for 2 h following a 1 h pre-collection and aged at 25°C to give 0–1.5, 2–4, 4–6, 6–8 and 20–21 h old embryos; unstaged embryos were collected over a 16 h period. Embryos were homogenized using a Dounce homogenizer in extraction buffer (30 mM Tris-HCl, 100 mM EDTA pH 7.4, 40 mM NaCl, 0.2% Triton X-100, and 0.5% diethyl pyrocarbonate), modified from Hough-Evans *et al.* (1980). After centrifugation for 15 min at 15 000 *g* at 4°C, the supernatant was made 1% in SDS and RNA was extracted twice with phenol chloroform (20/1) and twice with saturated phenol (pH 7.8). Oocyte RNA was prepared from fully mature ovaries of virgin females conditioned for 48 h on a yeast-rich medium. The hand-dissected ovaries were immediately homogenized in extraction buffer and the RNA isolated as described above.

Larvae and pupae were frozen in liquid nitrogen and ground to a fine powder with a pestle and mortar. The tissue was solubilized in 30 mM Tris-HCl, 10 mM EDTA pH 7.4, containing 1% SDS and 0.5% diethyl pyrocarbonate and extracted three times with equal volumes of phenol chloroform. Adults were directly homogenized in the same buffer using a Dounce homogenizer and the RNA extracted as described above. Poly(A)$^+$ RNA was isolated by oligo(dT) cellulose chromatography according to Marcu *et al.* (1978).

Northern blot RNA analysis

Formaldehyde gel electrophoresis was performed using 1.5% agarose gels (Colot and Rosbash, 1982). Transfer of RNA to nitrocellulose filters according to Thomas (1980) was in 10 x SSC. After baking at 80°C for 2 h *in vacuo*, the filters were pre-hybridized for 2 h at 42°C, hybridized and washed as described in Colot and Rosbash (1982). Exposure to Kodak XAR-5 films was at −70°C with intensifying screens.

Radiolabelling of nucleic acids

Double-stranded DNA (100–200 ng) was labelled by nick translation according to Golden *et al.* (1980) using 50 μCi of [^{32}P]dATP or dCTP (800 Ci/mmol, New England Nuclear). Single-stranded M13 phage DNA was labelled by primer extension (Hu and Messing, 1982) with some modifications. The annealing DNA mixture containing 4 μl of primer (10 ng), 5 μl of template (~400 ng), and 1 μl of 10 x buffer (100 mM Tris-HCl pH 7.9, 600 mM NaCl, 66 mM $MgCl_2$) was boiled for 3 min and cooled at room temperature for 30 min. The DNA synthesis reaction was allowed to proceed for 30 min at 20°C.

DNA sequence analysis

The Forward-Backward method of Sief *et al.* (1980) was used for sequence analysis of restriction fragments labelled at the 5′ or 3′ termini. M13 subclones were sequenced by the dideoxy chain termination method (Sanger *et al.*, 1977; Messing *et al.*, 1981).

Transcriptional mapping with S1 nuclease

For S1 nuclease protection experiments (Berk and Sharp, 1978), 10 μg of

polyadenylated RNA from 0–16 h old embryos were hybridized with purified labelled restriction fragments in 100 mM Pipes pH 7.8, 10 mM EDTA, 400 mM NaCl, 70% formamide. DNA was denatured for 3 min at 75°C and hybridization was for 2.5 h at 50°C; after hybridization, the reaction was diluted with 10 volumes of ice cold S1 buffer and digested with S1 at a concentration of 500 units/ml for 30 min at 30°C. The digestion products were analyzed by electrophoresis on 4.5% polyacrylamide gels.

Enzymes

Restriction enzymes were purchased from New England Biolabs and Bethesda Research Laboratories. DNA polymerase I was from Boehringer-Mannheim and the Klenow fragment from New England Nuclear and Bethesda Research Laboratories.

Acknowledgements

We are grateful to N.Abovich and J.Teem for their helpful suggestions during the course of this work, and to A.R.Campos and A.M.Valles for their assistance with some M13 experiments. We wish to thank W.Bender, H.V. Colot, L.Gritz and J.C.Hall for critical reading of the manuscript, and T.Tishman for preparing it. One of the authors (A.V.) was the recipient of a NIH-CNRS (France) exchange fellowship. This work was supported by grants from the National Institutes of Health to M.R.

References

Anderson,K.V. and Lengyel,J.A. (1979) *Dev. Biol.*, **70**, 217-231.
Arthur,C.G., Weide,C.M., Vincent,W.S. and Goldstein,E.S. (1979) *Exp. Cell Res.*, **121**, 87-94.
Benyajati,C., Spoerel,N., Haymerle,H. and Ashburner,M. (1983) *Cell*, **33**, 125-133.
Berk,A.J. and Sharp,P.A. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 1284-1278.
Colot,H.V. and Rosbash,M. (1982) *Dev. Biol.*, **94**, 79-86.
Dente,L., Cesarini,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645-1655.
Fitzgerald,M. and Shenk,T. (1981) *Cell*, **24**, 251-260.
Golden,L., Schafer,U. and Rosbash,M. (1980) *Cell*, **22**, 835-844.
Gray,M.R., Colot,H.V., Guarente,L. and Rosbash,M. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6598-6602.
Hagenbuchle,O., Tosi,M., Schibler,U., Bovey,R., Wellauer,P.K. and Young,R.A. (1981) *Nature*, **289**, 643-646.
Hall,L.M.C., Mason,P.J. and Spierer,P. (1983) *J. Mol. Biol.*, **169**, 83-96.
Hogness,D.S., Wensink,P.C., Glover,D.M., White,R.L., Finnegan,D.J. and Donelson,J.E. (1975) in Peacock,W.J. and Brock,R.D. (eds.), *The Eukaryote Chromosome*, Australian National University Press, Canberra, pp. 1-33.
Hough-Evans,B.R., Jacobs-Lorena,M., Cummings,M.R., Britten,R.J. and Davidson,E.H. (1980) *Genetics*, **95**, 81-94.
Hu,N.T. and Messing,J. (1982) *Gene*, **17**, 271-277.
Lefevre,G.,Jr. (1974) *Annu. Rev. Genet.*, **8**, 51-62.
Lengyel,J.A., Thomas,S.R., Boyer,P.D., Salas,F., Strecker,T.R., Lee,I., Graham,M.L., Roark,M. and Underwood,E.M. (1984) in Malacinski, G.M. and Klein,W.H. (eds.), *Molecular Aspects of Early Development*, Plenum Press, NY, pp. 219-252.
Lindsley,D.L., Sandler,L., Baker,B.S., Carpenter,A.T.C., Denell,R.E., Hall,J.C., Jacob,P.A., Miklos,G.L.C., Davis,B.K., Gethmann,R.C., Hardy,R.W., Hessler,A., Miller,S.M., Nozawa,H., Perry,D.M. and Gould-Somero, M. (1972) *Genetics*, **71**, 157-184.
Marcu,K., Valbuena,O. and Perry,R.P. (1978) *Biochemistry*, **17**, 1723-1733.
McKnight,S.L. and Miller,O.L. (1976) *Cell*, **8**, 305-319.
Messing,J., Crea,R. and Seeburg,H. (1981) *Nucleic Acids Res.*, **9**, 309-321.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269-276.
Nusslein-Volhard,C. and Wieschaus,E. (1980) *Nature*, **287**, 795-801.
Rozek,C.E. and Davidson,N. (1983) *Cell*, **32**, 23-24.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Scherer,G., Telford,J., Baldari,C. and Pirrotta,V. (1981) *Dev. Biol.*, **86**, 438-447.
Schibler,U., Hagenbuchle,O., Wellauer,P.K. and Pittet,A.C. (1983) *Cell*, **33**, 501-508.
Sief,I., Khoury,G. and Dahr,R. (1980) *Nucleic Acids Res.*, **8**, 2225-2240.
Sina,B.J. and Pellegrini,M. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7351-7355.
Spierer,P., Spierer,A., Bender,W. and Hogness,D.S. (1983) *J. Mol. Biol.*, **168**, 35-50.
Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201-5205.
Trumbly,R.J. and Jarry,B. (1983) *EMBO J.*, **2**, 1281-1290.
Vaslet,C.A., O'Connell,P., Izquierdo,M. and Rosbash,M. (1980) *Nature*, **285**, 674-676.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259-268.
Wieschaus,E. and Gehring,W. (1976) *Dev. Biol.*, **50**, 249-263.
Wong,Y.C., O'Connell,P., Rosbash,M. and Elgin,S.C.R. (1981) *Nucleic Acids Res.*, **9**, 6749-6762.
Zalokar,M. (1976) *Dev. Biol.*, **49**, 425-437.

Received on 31 January 1984