Volume 13 Number 19 1985 Nucleic Acids Research

Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes

Janice A.Fischer and Tom Maniatis

Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA

Received 22 July 1985; Accepted 9 September 1985

ABSTRACT

The D. melanogaster Adh gene is transcribed from two different promoters; a proximal (larval) promoter is active during late embryonic and larval stages, and a distal (adult) promoter is active primarily in third instar larvae and in adult flies (1). Genetic analyses suggest that several species of the mulleri subgroup (distant relatives of D. melanogaster) have two closely-linked Adh genes, Adh-1 and Adh-2, each of which expresses a different ADH protein (2). The temporal pattern of expression of Adh-1 and Adh-2 is similar to the expression of D. melanogaster Adh from the proximal and distal promoters (2,3,4). We are interested in the molecular basis for the pattern of Adh expression in the mulleri subgroup species and in the mechanism of the switch in Adh promoter utilization. For these reasons, we have studied the structure and transcription of the Adh locus of D. mulleri, a species of the mulleri subgroup. We show that the ADH-1 and ADH-2 proteins are expressed from two distinct genes separated by 2 kilobase pairs, and that Adh-1 and Adh-2 are transcribed in the expected temporal pattern. In addition, we find a pseudogene 1.2 kb upstream from Adh-2, which is transcribed in a temporal pattern similar to Adh-2.

INTRODUCTION

The Drosophila alcohol dehydrogenase (Adh) gene is particularly suited for an analysis of the molecular mechanisms involved in temporal and tissue specific gene expression. In addition to displaying characteristic patterns of expression in a limited number of tissues (5), the Drosophila melanogaster Adh gene is transcribed from two different promoters during different stages of development. The proximal (larval) promoter is active during late embryonic and larval development, while the distal (adult) promoter is active primarily in third instar larvae and in adult flies (1). P-element transformation experiments have shown that an 11.8 kb genomic DNA fragment bearing the Adh gene carries all of the DNA sequences necessary for correct stage and tissue-specific gene expression, and for the correct temporal switch in promoter utilization (6). Thus, germ line transformation can be used to identify the DNA sequences required for the larval to adult promoter switch in D. melanogaster.

An alternative approach to the analysis of temporal switching of Adh gene expression is suggested by the observation that members of the mulleri subgroup (distant relatives of D. melanogaster) produce two different ADH proteins, ADH-1 (larval) and ADH-2 (adult), in a temporal pattern similar to the expression from the D. melanogaster Adh larval and adult promoters (2,3,4). Genetic analysis of electrophoretic variants of ADH-1 and ADH-2 in D. buzzatii (a species of the mulleri subgroup) provided evidence for the hypothesis that ADH-1 and ADH-2 are encoded by two closely-linked genes that are differentially expressed during development (2). However, other possibilites such as alternate splicing of a single transcript to produce the two proteins could not be ruled out. Whatever the processes involved, a comparison of the structure and transcriptional regulation of the Adh genes in these distantly related species could lead to an understanding of the mechanism and biological significance of the temporal switch in Adh promoter selection.

Towards this goal, we have cloned and sequenced the Adh locus of Drosophila mulleri, a species of the mulleri subgroup. We show that the ADH-1 and ADH-2 proteins of D. mulleri are encoded by two distinct genes separated from each other by 2 kilobases of DNA. In addition, we detect an Adh pseudogene 1.2 kb upstream from the Adh-2 gene. A transcriptional analysis reveals that Adh-1 and Adh-2 are each transcribed from a distinct expected temporal pattern, and that the pseudogene is transcribed in a temporal pattern similar to Adh-2.

MATERIALS AND METHODS

Construction of a D. mulleri genomic library

D. mulleri flies (stock #15081-1371.8 (3370.1)) were obtained from the Drosophila Species Resource Center at Bowling Green, Ohio, and cultured on standard cornmeal food. Genomic DNA was prepared from D. mulleri adult fly nuclei as previously described (7) and EMBL4 phage (8) DNA was purified by standard methods (9). High molecular weight Sau 3A partial digest fragments of D. mulleri genomic DNA were isolated, ligated to Bam HI-digested phage arms and packaged (9). A library of 80,000 recombinant phage was amplified on strain Q359 (10).

Isolation of Adh clones

The amplified D. mulleri library was plated on strain K802 (11) and screened by hybridization with the nick-translated plasmid pRX8 (6) which contains a 4.5 kb SalI-XhoI fragment bearing the Adh gene from D. melanogaster. Standard procedures were followed for plaque hybridization

(9). Hybridization and washing were performed under low stringency conditions: hybridization was done at 42°C for 18 hours in 5X SSC, 20% formamide, 0.5% SDS, 5X Denhardt's, 100 ug/ml denatured salmon sperm DNA, 10 mM EDTA and with or without 10% dextran sulfate, and the filters were washed at 50-55°C in 1X SSC and 0.5% SDS.

Characterization of D. mulleri Adh clones

Plasmid and phage DNA mini-preps, restriction digestion, Southern and Northern blotting, nick-translation, and subcloning were all carried out according to standard procedures (9). Enzymes were purchased from New England Biolabs and Bethesda Research Laboratories, and ^{32}P-labelled deoxynucleotides were purchased from Amersham.

DNA sequence analysis.

The 2.4 kb and 7.8 kb Eco RI fragments of recombinant phage λD (see Results) were each subcloned, in both orientations, into the Eco RI site of the M13 origin-containing plasmid pSDL13 (12). Nested sets of deletion plasmids were constructed from each of the four resulting plasmids according to the following modification of the procedures described by Frischauf, et. al. (13) and Hong (14). Supercoiled plasmid DNA (10 ug) was digested at room temperature with 1 ng of DNAse I (presoaked on ice for at least 30 minutes in 0.02 M Tris-Cl (pH 7.5), 1.5 mM $MnCl_2$ and 100 ug/ml BSA) in a total volume of 100 ul of 1.5 mM $MnCl_2$ and 20mM Tris-Cl (pH7.5). In this buffer, DNAse I makes double-stranded cleavages; the digestion time was optimized for production of linearized (once-cut) plasmid molecules. The DNAse I digestions were stopped by phenol extraction, followed by phenol/chloroform extraction. Each digest was then run on a 0.6% low melting temperature agarose (Sigma Type VII) gel from which the linear molecules were recovered by melting the agarose and extracting with phenol pre-warmed to 37°C. Following phenol/chloroform extraction and ethanol precipitation, the linearized plasmids were digested to completion with Bam HI, thus removing the DNA between the sequencing primer and the DNAse I cleavage site. The Bam HI and DNAse I ends were blunted with T4 DNA polymerase, and the reaction was stopped by phenol/choroform extraction . The DNA fragments were then size fractionated on a 0.6% low-melting temperature agarose gel run at 50 volts overnight. The gel was cut in intervals corresponding to approximately 500 bp, starting from the size of the linearized plasmid to the size of the vector without insert. Each gel slice was melted and the DNA self-ligated (without removing the agarose) in a total volume of 1.5 ml. A small aliquot of each ligation was used to transform E. coli strain XS127

(12). Several transformants from each ligation were analyzed by restriction digestion of the plasmid DNA. A set of overlapping clones to be sequenced was thus isolated, and 0.1 ml of a saturated culture of XS127 bearing each deletion plasmid was infected with the M13 variant rv-1 as described in Levinson, et. al. (12). The infected cells were grown to saturation at 37° C in 5 ml of L-Broth, and the standard 1.5 ml miniprep procedure was used to make the single-stranded templates (15). Dideoxynucleotide sequencing reactions were carried out essentially according to the standard Sanger sequencing protocol (16) using the M13 reverse sequencing primer, obtained from New England Biolabs. Gel plates were 55 cm long. The sequence of each Eco RI fragment was determined in one direction, reading from 150 to 350 nucleotides from each clone, and each clone overlapping the next by 50 to 200 nucleotides. In regions of uncertainty, the opposite strand (from another set of nested deletions) was also sequenced. An additional clone was made to sequence through the junction of the two Eco RI fragments: a 620 bp Cla I-Pvu II fragment from phage λD (see Results) spanning the Eco RI site was cloned into pSDL13 cut with AccI and Sma I.

RNA Analysis

Total nucleic acid was prepared from larvae by the following modification of a procedure for RNA extraction from Xenopus oocytes in (17). First, the larvae were transferred from a collection plate or vial into a glass homogenizer, and washed free of media with 0.5% Triton X-100. RNA homogenization buffer (50 mM Tris-Cl (pH7.5), 50 mM NaCl, 10 mM EDTA, 0.5% SDS) and 10 ul/ml of 20 mg/ml proteinase K were added to the washed larvae which were then homogenized with a type A pestle. (The volume of homogenization buffer used per organism is not important, as long as the flies or larvae remain relatively easy to homogenize.) The homogenate was incubated at 37°C for 45-90 min., phenol/chloroform extracted twice, and then ethanol precipitated. The precipitate was washed with 100% ethanol, resuspended in 10 mM Tris-Cl (pH7.5) and 0.2% SDS, and spectrophotometrically quantitated. Total nucleic acid from adult flies was made in the same way, except that the flies were added directly to the RNA homogenization buffer after anaesthetization.

Transcript mapping by RNA:RNA hybridization was performed essentially as in Zinn et. al.(18). (The total nucleic acid preparations were used without DNAse treatment.) In order to make uniformly labeled complementary RNA probes specific for transcripts from each Adh gene (Adh-1, Adh-2 and Ψ-Adh,) the following three clones were constructed with DNA from the 2.4 and

7.8 kb Eco RI fragments subcloned into pUC12 (19): SP6-1 was made by cloning a blunted 1.3 kb Eco RI-Nru I fragment (containing 68 bp of the first coding exon of Adh-1 and sequences 5') into the Sma I site of SP64 (20), SP6-2 was made by cloning a 1.3 kb Nru I fragment (containing 68 bp of the first coding exon of Adh-2 and sequences 5') into the Sma I site of SP64, and SP6- was made by cloning a 3.8 kb Bgl II-Eco RI fragment (containing 25 bp of the first intron, the entire first coding exon of the pseudogene and sequences 5') into SP64 Bam HI/Eco RI. Labeled RNA probes were transcribed from these three clones as previously described (20). Prior to transcription, SP6-1 and SP6-2 were linearized with HpaII which resulted in a 335 bp and a 270 bp probe, respectively. SP6-Ψ was linearized with Nhe I, resulting in a 610 bp probe. SP6 RNA polymerase was purchased from Promega-Biotec, New England Nuclear or Boehringer-Mannheim, RNAsin was from Promega-Biotec, and ^{32}P-UTP was purchased from Amersham.

DNA sequence comparisons

Computer programs for restriction site analysis and for homology searches were from International Biotechnologies Incorporated (IBI), written by James Pustell.

RESULTS AND DISCUSSION

Characterization of D. mulleri Adh clones

A D. mulleri genomic library was constructed and approximately 400,000 recombinant phage were plaque-screened at low hybridization stringency with nick-translated pRX8, a plasmid containing the D. melanogaster Adh gene (Materials and Methods). DNA from forty independently picked recombinant phage was isolated and digested with Eco RI, which generates the left and right arms of the phage vector plus one or more fragments from the D. mulleri DNA insert. Examination of the Eco RI digestion patterns made it possible to classify the forty clones into six different types (A-F in Figure 1). The D. mulleri DNA inserts in each of the six classes of phage were then ordered with respect to each other by blotting/cross hybridization experiments to generate the Eco RI restriction map shown in Figure 1. When aligned in this manner the six overlapping clones span approximately 29 kilobases of DNA. The Adh genes were localized by hybridizing blots of the Eco RI-digested phage with nick-translated pRX8. Only the 2.4 kb and the 7.8 kb Eco RI fragments (see Figure 1) hybridized to pRX8 (data not shown). When the same probe or the 2.4 kb Eco RI fragment is hybridized to genomic blots of D. mulleri DNA digested with Eco RI, only the 2.4 and 7.8 kb bands are detected

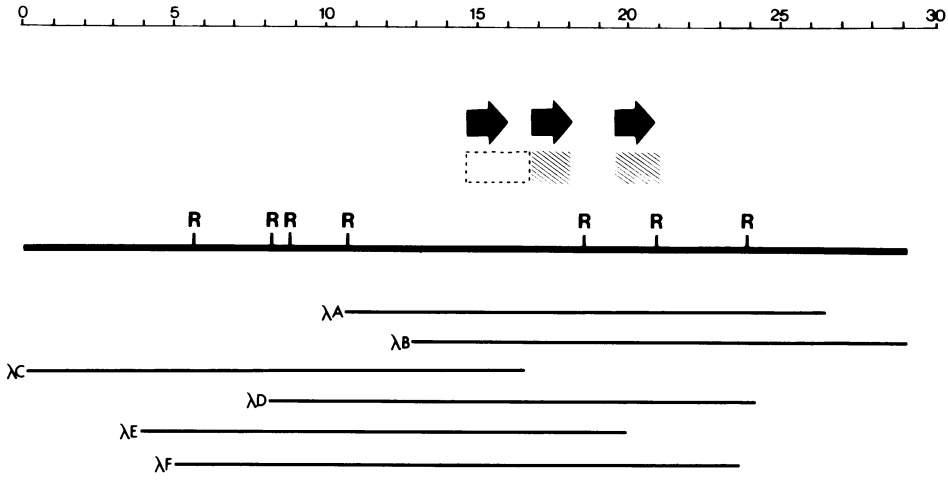


Figure 1: Restriction map of the cloned region of D. mulleri genomic DNA containing the Adh locus. The thick line through the middle of the diagram represents the cloned 29 kb region of D. mulleri genomic DNA. The lines beneath correspond to the six classes of recombinant phage (λA-λF) spanning the cloned region. EcoRI sites (R) are indicated. The shaded boxes above the restriction map indicate the two regions of the recombinant phage DNA which hybridize to the D. melanogaster Adh gene under high stringency conditions, and the box enclosed by broken lines indicates the region of cloned DNA which hybridizes to the D. melanogaster Adh gene only under low stringency conditions. The arrows indicate the 5'-3' orientations of the three regions which hybridize to the Adh gene.

(data not shown), suggesting that all sequences in the D. mulleri genome homologous to Adh are contained within the cloned 2.4 kb and the 7.8 kb Eco RI fragments.

A more precise localization of the Adh sequences within the Eco RI fragments was obtained by constructing a detailed restriction map of each fragment and then hybridizing blots of restriction digests with labeled fragments of pRX8 corresponding to different regions of the D. melanogaster Adh coding region. These experiments also made it possible to determine the direction of transcription of each gene (data not shown). The results of this analysis suggest that the 2.4 kb Eco RI fragment contains one complete Adh coding sequence, and that the 7.8 kb Eco RI fragment contains two complete Adh coding sequences, covering the regions indicated in Figure 1. Moreover, all three genes appear to have the same transcriptional orientation, as indicated in Figure 1. The most upstream of the two Adh genes in the 7.8 kb Eco RI fragment has diverged significantly from the D.

melanogaster Adh gene, since hybridization to this gene is not observed when the wash stringency is increased.

DNA sequence analysis

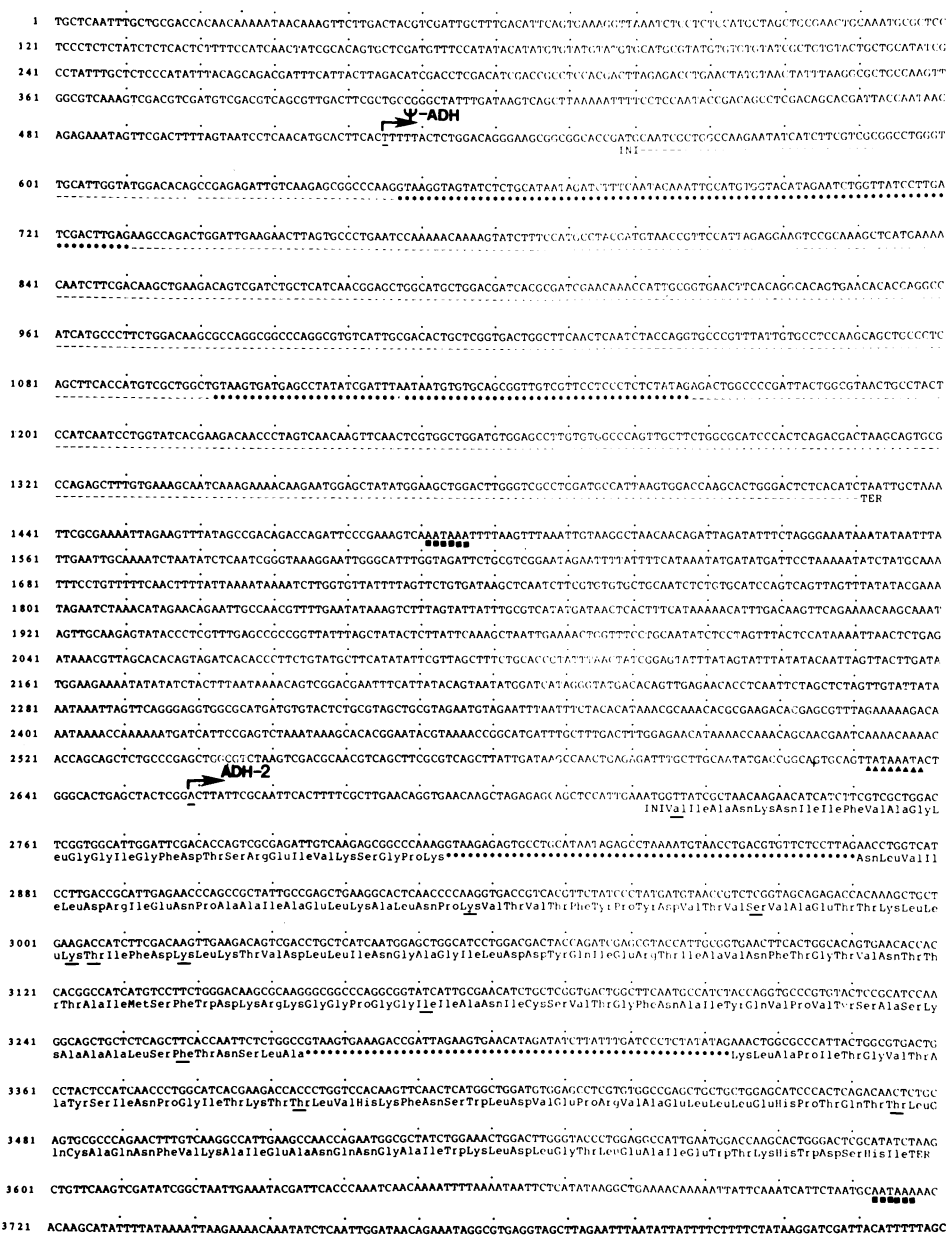
The nucleotide sequence of the entire 2.4 kb Eco RI fragment, and 4.4 kb of the 7.8 kb Eco RI fragment was determined as described in Materials and Methods and is shown in Figure 2. A summary of the structure of the sequenced region is given in Figure 3. Examination of the DNA sequence leads to the following conclusions:

1: Consistent with the blot-hybridization experiments described above, the D. mulleri Adh locus consists of three Adh genes which are all transcribed from the same DNA strand (Figure 2). Conceptual translation of the two downstream Adh genes reveals that they differ by ten amino acids (underlined in Figure 2.) These differences identify the upstream gene as Adh-2 and the downstream gene as Adh-1, since ADH-2 is the more basic protein (2). The region bearing the weakest homology to all regions of the D. melanogaster Adh gene in the blot hybridization experiments is a pseudogene. The sequence of this gene reveals that it has exons and introns (with all splice junction sequences intact) whose sizes and arrangement are similar to those of Adh-1 and Adh-2. However, the gene has accumulated numerous nucleotide substitutions, deletions and additions, so that it no longer codes for an ADH protein.

2: The intron/exon arrangements of Adh-1 and Adh-2 are almost exactly alike, and quite similar to the D. melanogaster Adh gene; the coding region is interrupted by two small introns, at exactly the same locations with respect to the amino acid sequence in both Adh-1 and Adh-2 and in the D. melanogaster Adh gene.

3: The coding region sequences of Adh-1 and Adh-2 are quite conserved; there are only 30 single base differences, 10 of which lead to amino acid substitutions. A recent gene conversion event appears to have occurred between the two genes, since the entire 57 bp first intron of each gene has exactly the same base sequence, while the second intron is not nearly as conserved. It is possible that the converted region includes 62 bp of the first exon (the entire coding region of the first exon downstream of the second codon) as well as the first 34 bases of the second exon, since these sequences also are exactly the same in the two genes.

4: Adh-1 and Adh-2 each appear to have a single TATA-box, at bases 5512 and 2633 in Figure 2, respectively. There is no sequence evidence for a distal promoter and a splice junction donor sequence, resembling the organization


Ψ-ADH
INI
TER
ADH-2

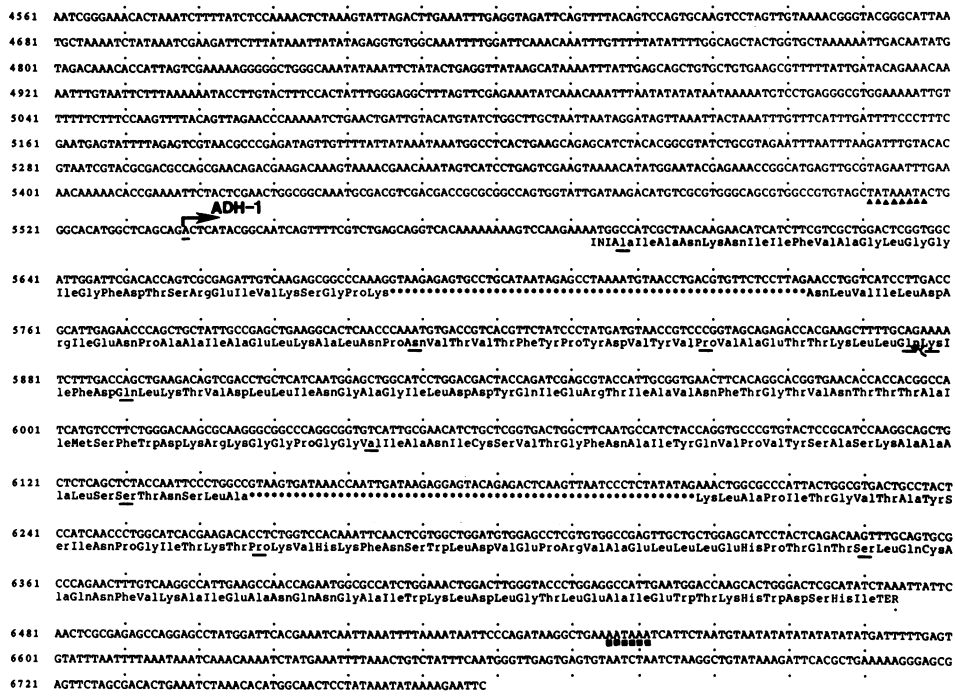


Figure 2: Nucleotide sequence of 6779 bases containing all three D. mulleri Adh genes. The approximate start of transcription of each gene is indicated by an arrow above an underlined nucleotide. Initiation (INI) and termination (TER) codons are indicated, and the TATA-boxes and putative polyadenylation sites (AATAAA) for each gene are underlined. Introns are indicated by asterisks (*) beneath the appropriate nucleotides, and the amino acids are indicated beneath their codons. The amino acids which differ between Adh-1 and Adh-2 are underlined.

of the D. melanogaster Adh gene (1), upstream from either of the two genes.

Transcriptional analysis

The sizes of the D. mulleri Adh transcripts were investigated by RNA blotting, and the correspondence of these transcripts to the cloned DNA was determined by hybridization with complementary SP6 RNA probes specific to each of the three Adh genes. Northern blots of D. mulleri larval and adult RNA hybridized with either nick-translated pRX8, or the 2.4 kb Eco RI fragment containing the Adh-1 gene, showed the same single band in both RNA types (data not shown) which, as expected from the DNA sequence of Adh-1 and Adh-2, is similar in size to the larval and adult transcripts of the D. melanogaster Adh gene (1100 and 1150 nucleotides, respectively (1,21)). In

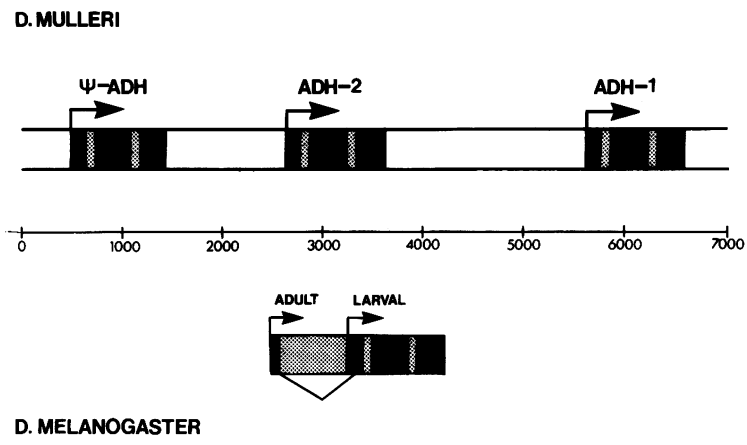


Figure 3: Schematic diagram of the D. mulleri Adh locus and the D. melanogaster Adh gene. The three D. mulleri Adh genes are indicated by rectangles; the black regions represent exons, the dotted regions represent introns and the direction of transcription of each gene is indicated by an arrow. The numbers on the scale correspond to the nucleotide numbers in the DNA sequence in Figure 2, and refer to the D. mulleri locus above the scale. For comparative purposes, the D. melanogaster Adh gene is pictured beneath the scale. The arrows indicate the adult and larval transcript starts (as labeled) and the lines beneath indicate the boundaries of the intron in the leader sequence of the adult transcript.

order to determine whether there is a transcript corresponding to each of the three Adh genes, three SP6 RNA probes were made, each complementary to all or part of the first coding exon and several hundred base pairs 5' to the presumptive initiation codon (ATG) of each gene. The construction of the probes is described in Materials and Methods, and they are diagrammed in detail in Figure 4. RNA from D. mulleri third instar larvae and adults was assayed with each probe. As shown in Figure 4, each of the three probes protects an RNA fragment of an appropriate size in one or both RNA preparations. Thus, Adh-1, Adh-2 and the pseudogene are transcribed.

As diagrammed in Figure 4, the sizes of the RNA fragments protected by each of the probes indicate either the 5' end of the transcript, or a splice acceptor site. It is likely that the lengths of the RNA fragments protected by the Adh-1 and Adh-2 probes correspond to the 5' ends of the transcripts, because there is a TATA-box in the DNA sequence starting at 28 bp 5' to the end of each transcript mapped by the probes. As shown in Figure 4, the Adh-1-specific probe (SP6-1) protects an approximately 122 nucleotide fragment of RNA, thus mapping the 5' end of the Adh-1 transcript to about 54 nu-

cleotides upstream of the ATG of Adh-1. The Adh-2-specific probe (SP6-2) protects an RNA fragment of approximately 130 nucleotides, thus mapping the 5' end of the Adh-2 transcript to about 62 nucleotides upstream of the ATG of Adh-2. As also shown in Figure 4, the pseudogene-specific probe (SP6-ψ) protects a 120 nucleotide RNA fragment. Assuming that the first intron (predicted by the DNA sequence) of the pseudogene primary transcript is removed by splicing, the end of the pseudogene transcript thus mapped is about 32 nucleotides upstream of the ATG. Unlike the mapped ends of the Adh-1 and Adh-2 transcripts, there is no TATA-box sequence at an appropriate distance upstream from the pseudogene transcript end (nor is there a splice acceptor sequence.)

The appearance of the Adh-1 and Adh-2 transcripts shows the expected stage specificity. The Adh-1 probe protects the 122 nucleotide RNA fragment in larval RNA only, and the Adh-2 probe protects the 130 nucleotide RNA fragment in both larval and adult RNA preparations (Figure 4). These results are predicted by ADH protein assays of larval and adult homogenates, which show that both ADH-1 and ADH-2 are present in D. mulleri third instar larvae, but that only ADH-2 is present in adult protein extracts (2,4).

As shown in Figure 4, the pseudogene transcript is present at relatively low levels in adults. This transcript is also detected at even lower levels in late third instar larval RNA preparations (data not shown). RNA polymerase II pseudogenes are a common feature of gene clusters in mammals (22-27), but occur rarely in Drosophila (28). The D. mulleri Adh pseudogene differs from the other pseudogenes examined thus far in that it is transcribed at an easily detectable level. Previous attempts to detect pseudogene transcripts in mammals (29) and in Drosophila (28) have failed.

In summary, RNA mapping experiments show that there is a transcript corresponding to each of the three Adh genes, and that the Adh-1 and Adh-2 transcripts are the size predicted by the DNA sequence. In addition, the Adh-1 and Adh-2 transcripts show the expected temporal specificity, while the pseudogene has a temporal pattern of expression similar to the adjacent gene, Adh-2.

Comparisons of Adh sequences

To identify conserved sequences of possible regulatory significance we compared the DNA sequences 5' to Adh-1, Adh-2 and the pseudogene to each other and to the DNA sequences 5' to the larval and adult transcription start sites of the D. melanogaster Adh gene. One kilobase of DNA sequence 5' to Adh-1 and to Adh-2 and 400 bp 5' to the pseudogene were compared in the

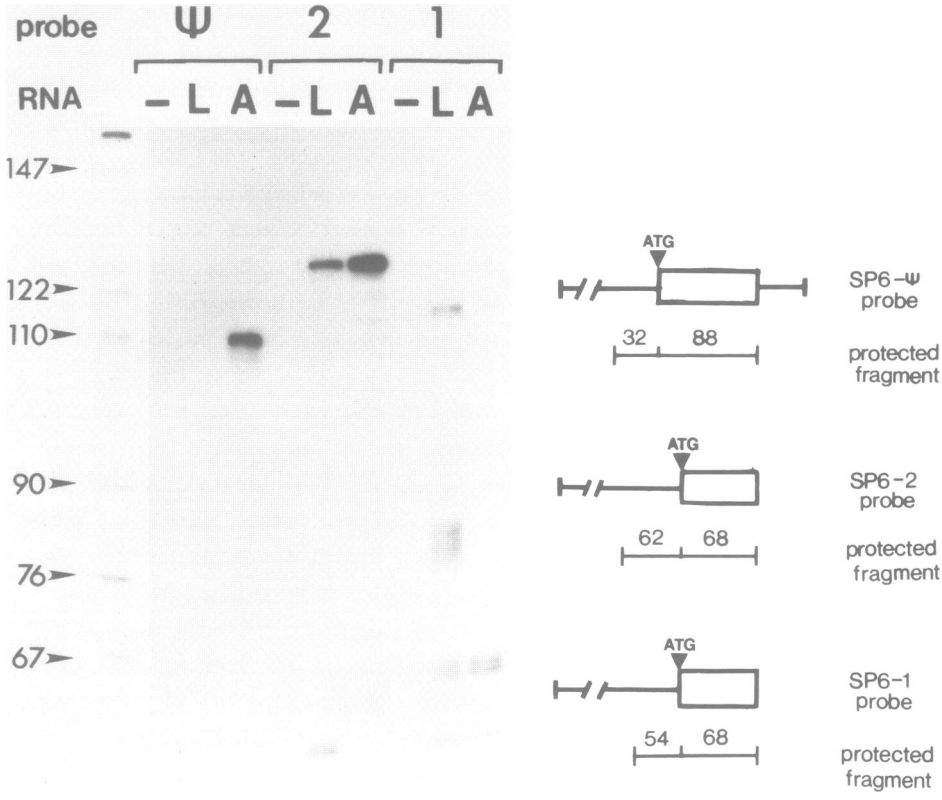


Figure 4: Mapping of transcripts from each D. mulleri Adh gene by RNA:RNA hybridization with SP6 probes. The construction of the three SP6 probes specifically complementary to putative transcripts from Adh-1 (SP6-1), Adh-2 (SP6-2) and Ψ-Adh is described in detail in Materials and Methods. Each probe was hybridized with 20 ug of tRNA (- lanes). The SP6-1 and SP6-2 probes were each hybridized with 5 ug of RNA from both D. mulleri third instar larvae (L lanes) and adult flies (A lanes), and the SP6-Ψ probe was hybridized with 50 ug of larval and adult RNA. None of the three probes protect any RNA fragments in the tRNA controls. The SP6-Ψ probe protects an RNA fragment of approximately 120 nt (88 nt of Ψ-coding sequence and 32 nt 5' to the ATG) in the adult RNA. The SP6-2 probe protects an RNA fragment of approximately 130 nt (68 nt of coding sequence and 62 nt 5' to the ATG) in both larval and adult RNA. The SP6-1 probe protects an RNA fragment of approximately 122 nt (68 nt of coding sequence and 54 nt 5' to the ATG) in the larval RNA. The ladder of bands seen in the larval RNA/SP6-1 hybridzation lane is an artifact due to RNAse A clipping at a run of 9 A residues in the Adh-1 transcript; this ladder does not appear when RNAse T1 alone is used to digest the non-hybridized RNA (data not shown). The apparently low level of Adh-1 transcripts is thus also an artifact of the RNAse clipping to produce the ladder of bands. The smaller protected band seen in the SP6-2/larval RNA hybridization is a result of SP6-2 protection of a 62 nt stretch of the first exon of the Adh-1 transcript. The two small bands seen in the SP6-

1/larval RNA and SP6-1/adult RNA hybridizations are a result of protection by SP6-1 of a 62 nt stretch of the first exon of the Adh-2 transcript, which is separated by a 2 nt mismatch from a 4 nt stretch of RNA bases complementary to the probe. The RNAse sometimes clips at the 2 nt mismatch, and thus both 62 nt and 68 nt protected fragments are seen.

three possible combinations by searching for highly homologous short (9 bp or more) regions of sequence.

Using these screening conditions, the only significant sequence homologies detected between the 5' flanking regions of the D. mulleri Adh genes lie within a region of approximately 400 bp 5' to their transcription start sites. A summary of the results of this sequence comparison is shown is shown in Figure 5. The Adh-1 and Adh-2 5' flanking sequences are, in general, much more homologous to each other than they are to the corresponding sequences upstream from the pseudogene. The TATA-box regions (-28 to -13 from each transcript start) are identical in Adh-1 and Adh-2, while there is no obvious TATA box 5' to the transcription start of the pseudogene; the nearest ATA is at -40 (Figure 2). There is one region of sequence homology shared by all three Adh genes. As shown in Figure 5, this homology spans the regions of -110 to -95, -168 to -129 and -208 to -190 from the pseudogene transcript start, -119 to -70 from the Adh-2 transcript start and -97 to -85 from the Adh-1 transcript start. The pseudogene and Adh-2 sequences in this region are more similar to each other than they are to Adh-1.

To determine whether any DNA sequences 5' to the D. mulleri Adh genes are also present 5' to the D. melanogaster Adh gene, we compared (as above) the regions 5' to the three D. mulleri Adh genes with 1500 bp of sequence of the D. melanogaster Adh gene (30,31). The region of the D. melanogaster sequence compared was from -660 to the distal transcript start, through the adult leader sequence, the adult intron, and the larval leader to the ATG start codon (see Figure 3 and ref.30).

The D. melanogaster Adh 5' flanking sequence was found to contain relatively little homology to the 5' flanking sequences of the D. mulleri Adh genes. In the 5' region of the D. melanogaster Adh gene from -660 to -80 from the adult transcript start, and in the region from the start of the adult transcript through -140 from the larval transcript start (containing most of the adult intron,) a few 9-12 nucleotide sequences were found to appear also in the 5' flanking regions of Adh-1 and Adh-2 (data not shown.) These conserved sequences are not in any particular positions with respect

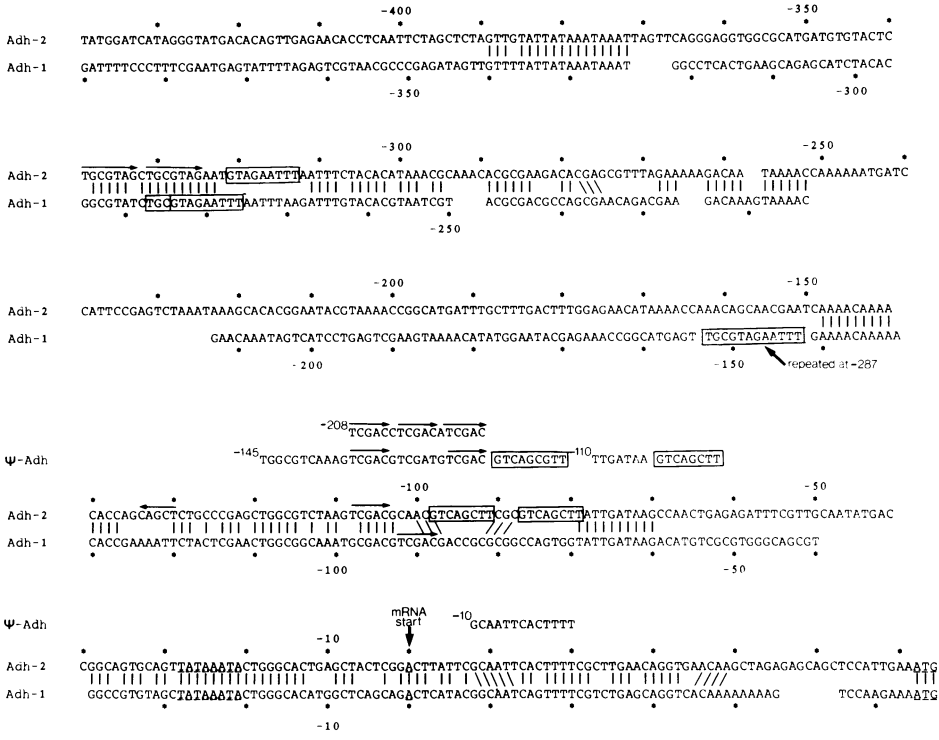


Figure 5: Nucleotide sequence comparison of the 5' flanking regions of the three D. mulleri Adh genes. The 5' flanking sequence (starting at -390) and 5' untranslated region of Adh-1, and the 5' flanking sequence (starting at -439) and 5' untranslated region of Adh-2 are aligned to show sequence homology. Where appropriate, portions of the 5' flanking region of the pseudogene are aligned to show homology with the Adh-1 and Adh-2 sequences. Vertical lines indicate matching bases. Some conserved sequences are boxed to indicate that they are repeated several times. The arrows above other sequences also indicate repeating units of nucleotide sequence.

to the D. melanogaster Adh promoters as compared to their locations with respect to the promoters of Adh-1 and Adh-2.

As shown in Figure 6, one interesting region of conserved sequence is found near both the D. melanogaster promoters and near all three D. mulleri promoters. Heberlein, et. al. (32) have noted that the sequence between -84 and -59 from the D. melanogaster adult transcription start and the sequence between -140 and -113 from the larval transcription start each have the pentanucleotide GC$^{A}_{T}$GC repeated four times. This pentanucleotide repeat is also found four times between -147 and -90 from the start of the pseudogene transcript, twice between -134 and -78 from the Adh-2 transcription start, and

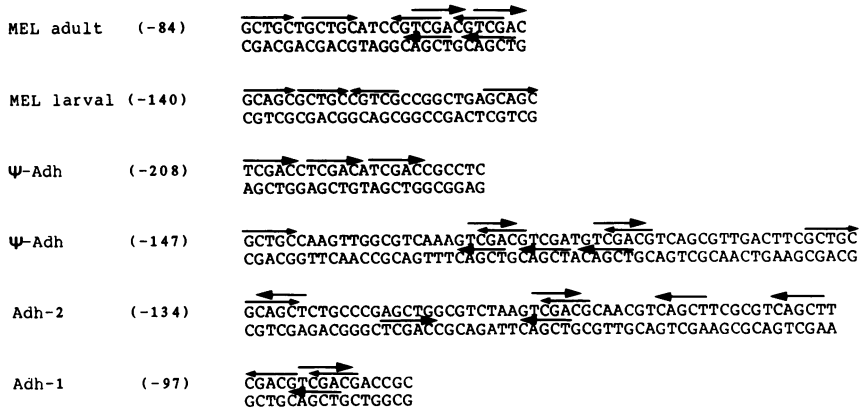


Figure 6: Conserved sequences 5' to the D. melanogaster larval and adult transcripts and the three D. mulleri Adh genes. The numbers on the left indicate the distance of the first nucleotide of each sequence from the start of transcription. Both strands of DNA are pictured. Two repeating sequence motifs are indicated by the arrows. Thin arrows are above the pentanucleotide repeat GC$\overset{A}{T}$GC where it appears in the top strand only. Thicker arrows are above the pentanucleotide repeat TCGAC, on each strand where it appears.

twice between -97 and -85 from the Adh-1 transcription start (Figure 6). With the exception of the region near the D. melanogaster larval promoter, these sequences bear different numbers of copies of the pentanucleotide TCGAC. This pentanucleotide is also repeated three times in the sequence between -208 and -186 from the pseudogene transcription start. Several regulatory elements near Drosophila (33), yeast (34), viral (35,36) and mammalian (37-41) promoters contain short, repeated nucleotide sequences. In the case of the D. melanogaster Adh gene, the repeated sequence elements in Figure 5 may be functionally significant. Heberlein et. al. (32) have shown that the region between -70 and -58 from the distal (adult) promoter of the D. melanogaster Adh gene is necessary for efficient transcription in vitro using Kc cell extracts. They have also identified a protein factor (Adf-1) within these extracts which specifically binds to the region between -85 and -58 from the adult transcript start, and also to the region between -150 and -105 from the larval transcript start.

Another interesting region of sequence conservation with the 5' flanking region of a D. mulleri Adh gene is found at the D. melanogaster adult promoter. As seen in Figure 7A, at -186 to -168 5' to the mapped end of the pseudogene transcript, there is a region homologous to the adult promoter (TATA-box) of the D. melanogaster Adh gene. This sequence is not likely to

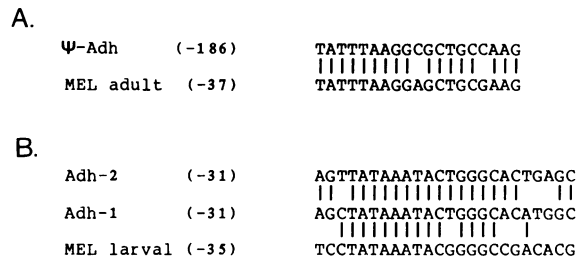


Figure 7: A. Nucleotide sequences conserved between the D. melanogaster adult promoter TATA-box at -37 from the start of the adult transcript, and the sequence starting at -186 from the pseudogene transcript start. B. Nucleotide sequence conservation between the TATA-box regions of Adh-1, Adh-2 and the larval promoter of the D. melanogaster Adh gene.

be part of a functional promoter, because the pseudogene transcript 5' end maps about 186 bp downstream; splicing is not a likely explanation since the appropriate signals are not present. As shown in Figure 7B, both the Adh-1 and Adh-2 TATA-box sequences resemble the TATA-box of the D. melanogaster larval promoter.

In summary, we find that although the 5' flanking regions of Adh-1 and Adh-2 show significant sequence conservation, they show little sequence homology with the DNA sequences near the larval (proximal) and adult (distal) promoters of the D. melanogaster Adh gene. There are several precedents for large blocks of conserved sequences in the 5' flanking regions (40,42-44) or intragenic regulatory sequences (45-48) of homologous genes from similarly divergent species. However, there are other examples of genes from divergent species in which only very small sequence elements are conserved amid largely divergent 5' flanking sequences (49,50). In most of these systems, the conserved sequences have been shown to be functionally significant (33,39-41,51-57). We know that regulatory elements must also be functionally conserved between the D. mulleri and D. melanogaster Adh genes, because Adh-1 and Adh-2 are expressed at normal levels in the proper tissue and stage-specific pattern when the D. mulleri Adh locus in introduced into the D. melanogaster genome via p-element transformation (58).

There are several possible explanations for the paucity of sequence conservation in the 5' flanking regions. P-element transformation experiments with in vitro synthesized mutants of the D. melanogaster Adh gene (59) and the D. mulleri Adh genes (58) indicate that some functionally homologous regulatory elements are in widely different locations with respect to the

promoters they affect in the two species. This could be a consequence of the duplication (or triplication) event which is believed to have resulted in the two-gene Adh locus in the mulleri subgroup. The lack of spatial conservation of the arrangement of the regulatory sequences, combined with the small size of many regulatory elements, makes it difficult to localize conserved elements of possible functional significance solely by inspection of the nucleotide sequences. P-element transformation experiments are in progress to localize and to investigate the properties of the DNA sequence elements involved in the promoter switch and in other aspects of the regulation of Adh expression in D. mulleri.

ACKNOWLEDGEMENTS

We would like to thank Martin Kreitman for his unpublished DNA sequence data, Brian Seed for providing sequencing vectors and bacterial strains prior to their publication, Jim Posakony for suggesting the cause of the SP6 mapping artifact, and David Goldberg for advice and encouragement during the cloning. We are grateful to Paul Macdonald and Gary Struhl for many useful comments on the manuscript. J.A.F. acknowledges support from an NSF Predoctoral Fellowship. This work was supported by a grant from the N.I.H. to T.M.

REFERENCES

1. Benyajati, C., Spoerel, N., Haymerle, H. and Ashburner, M. (1983) Cell 33, 125-133.
2. Oakeshott, J.G., Chambers, G.K., East, P.D., Gibson, J.B., and Barker, J.S.F. (1982) Aust. J. Biol. Sci. 35, 73-84.
3. Batterham, P., Lovett, J.A., Starmer, W.T. and Sullivan, D.T. (1983) Dev.Biol. 96, 346-354.
4. Batterham,P., Chambers, G.K., Starmer, W.T. and Sullivan, D.T. (1984) Evolution 38, 644-657.
5. Ursprung, H., Sofer, W.H. and Burroughs, N. (1970) Wil. Roux. Archiv. 164, 201-208.
6. Goldberg, D.A., Posakony, J.W. and Maniatis, T. (1983) Cell 34, 59-73.
7. Bingham, P.M., Levis, R. and Rubin, G.M. (1981) Cell 25, 693-704.
8. Frischauf, A.M., Lehrach, H., Poustka, A. and Murray, N. (1983) J. Mol. Biol. 170, 827-842.
9. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning-A Laboratory Manual. Cold Spring Harbor.
10. Kam, J., Brenner, S., Barnett, L. and Cesareni, G. (1980) Proc. Nat'l. Acad. Sci. 77, 5172-5176.
11. Wood, W.B. (1966) J. Mol. Biol. 16, 118-122.
12. Levinson, A., Silver, D. and Seed, B. (1984) J. of Mol. and Appl. Gen. 2, 507-517.
13. Frischauf, A.M., Garoff, H. and Lehrach, H. (1980) Nucl. Acids Res. 8, 5541-5549.
14. Hong, G.F. (1982) J. Mol. Biol. 158, 539-549.

15. Schreier, P.H. and Cortese, R. (1974) J. Mol. Biol. 129, 169-172.
16. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat'l. Acad. Sci. USA 74, 560-564.
17. Kreig, P. and Melton, D.A. personal communication
18. Zinn, K., DiMaio, D. and Maniatis, T. (1983) Cell 34, 865-879.
19. Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
20. Melton, D.A., Kreig, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) Nucl. Acids Res. 12, 7035-7056.
21. Goldberg, D.A. (1980) Proc. Nat'l. Acad. Sci. 77, 5794-5798.
22. Little, P. R. F. (1982) Cell 28, 683-684.
23. Huang, H., Crews, S. and Hood, L. (1981) J. Mol. Appl. Genet. 1, 93-101.
24. Steinmetz, M., Moore, K.W., Frelinger, J.G., Sher, B.T., Shen, F.-W., Boyce, E.A. and Hood, L. (1981) Cell 25, 683-692.
25. Lee, M. G.-S., Lewis, S.A., Wilde, C.D. and Cowan, N.J. (1983) Cell 33, 477-487.
26. Karin, M. and Richards, R.I. (1982) Nature 299, 797-802.
27. Lemischka, I. and Sharp, P.A. (1982) Nature 300, 330-335.
28. Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J. and Davidson, N. (1982) Cell 29, 1027-1040.
29. Lacy, E. and Maniatis, T. (1980) Cell 21, 545-553.
30. Kreitman, M. (1983) Nature 304, 412-417.
31. Kreitmen, M. unpublished sequence data, personal communication
32. Heberlein, U., England, B. and Tjian, R. (1985) Cell 41, 965-977.
33. Pelham, H.R.B. and Bienz, M. (1982) EMBO J. 1, 1473-1477.
34. Guarente, L. (1984) Cell 36,. 799-800.
35. Everett, R.D., Baty, D. and Chambon, P. (1983) Nucl. Acids Res. 11, 2447-2464.
36. McKnight, S.L., Kingsbury, R.C., Spence, A. and Smith, M. (1984) Cell 37, 253-262.
37. Carter, A.D., Felber, B.K., Walling, M.J., Jubier, M.F., Schmidt, C.J. and Hamer, D.H. (1984) Proc. Nat'l. Acad. Sci. USA 81, 7392-7396.
38. Stuart, G.W., Searle, P.F., Chen, H.Y., Brinster, R.L. and Palmiter, R.D. (1984) Proc. Nat'l. Acad. Sci. USA 81, 7318-7322.
39. Goodbourn, S., Zinn, K. and Maniatis, T. (1985) Cell 41, 509-520.
40. Fujita, T., Ohno, S., Yasumitsu, H. and Taniguchi, T. (1985) Cell 41, 489-496.
41. Ryals, J., Dierks, P., Rags, H. and Weissman, C. Cell 41, 497-507.
42. Blackman, R.K. and Meselson, M. (1985) J. Mol. Biol., submitted
43. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 653-668.
44. Kelly, R.A. and Pitha, P.M. (1985) Nucl. Acids Res. 13, 825-839.
45. Mills, F.C., Fisher, L.M., Kuroda, R., Ford, A.M. and Gould, H.J. (1983) Nature 306, 809-812.
46. Rabbitts, T.H., Forster, A., Baer, R. and Hamlyn, P.H. (1983) Nature 306, 806-809.
47. Hayday, A.C., Gillies, S.D., Saito, H., Wood, C., Wiman, K., Hayward, W.S. and Tonegawa, S. (1984) Nature 307, 334-340.
48. Emorine, L., Kuehl, M. Weir, L., Leder, P. and Max, E.E. (1983) Nature 304, 447-449.
49. Parslow, T.G., Blair, D.L., Murphy, W.J. and Granner, D.K. (1984) Proc. Nat'l. Acad. Sci. USA 81, 2650-2654.
50. Falkner, F.G. and Zachau, H.G. (1984) Nature 310, 71-74.

51. Mason, J.O., Williams, G.T. and Neuberger, M.S. (1985) Cell 41, 479-487.
52. Banerji, J., Olson, L. and Schaffner, W. (1983) Cell 33, 729-740.
53. Gillies, S.d, Morrison, S.L., Oi, V.T. and Tonegawa, S. (1983) Cell 33, 717-728.
54. Queen, C. and Baltimore, D. (1983) Cell 33, 741-748.
55. Neuberger, M.S. (1983) EMBO J. 2, 1373-1378.
56. Picard, D. and Schaffner, W. (1984) Nature 307, 80-82.
57. Dudler, R. and Travers, A.A. (1984) Cell 38, 391-398.
58. Fischer, J.A. and Maniatis, T. manuscript in preparation
59. Posakony, J.W. and Maniatis, T. manuscript in preparation