Structure of Two Unlinked Drosophila melanogaster Glyceraldehyde-3-phosphate Dehydrogenase Genes*

J. Yun Tso, Xiao-Hong Sun, and Ray Wu†

From the Section of Biochemistry, Molecular and Cell Biology, Division of Biological Sciences, Wing Hall, Cornell University, Ithaca, New York 14853

Two Drosophila genes that code for the enzyme glyceraldehyde-3-phosphate dehydrogenase (Gapdh) have been isolated and their structures determined by DNA sequence analysis. The two genes, Gapdh-1 and Gapdh-2, are homologous to each other in their coding regions but differ entirely in the 5' and 3' flanking regions. Both genes are functionally expressed in adult flies as determined by Northern blot analysis using gene-specific probes. Gapdh-1 is mapped by in situ hybridization at position 43E-F on the right arm of the second chromosome and Gapdh-2 at position 13F on the left arm of the X chromosome. Transcription initiation sites as well as polyadenylation sites for both Gapdh transcripts have also been determined. Gapdh-1 lacks a sequence homologous to the TATA box in its -30-base pair region that is characteristic of many RNA polymerase II transcribed promoters. In contrast, Gapdh-2 contains a consensus TATA box sequence as well as a CAAT box in its promoter region. Furthermore, a sequence element ATTTGCAT (dc) and non-tandem multiple direct repeats have been found in the −35 to −155-base pair 5' flanking region. Other than the intron located in the 5' noncoding region of Gapdh-2, both genes lack intervening sequences.

The glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase (d-glyceraldehyde-3-phosphate:NAD oxidoreductase (EC 1.2.1.12)) plays an important role in energy metabolism from carbohydrates. The gene for this protein is constitutively expressed in almost all cells, but its expression is highest in cells undergoing a high rate of glycolysis such as skeletal muscle or fermentative yeast. We are interested in its gene structure for two reasons. The first concerns the evolution studies of this highly conserved enzyme on the nucleotide level. Specifically we would like to analyze the pattern of changes that are not evident on the protein level. Parameters such as synonymous nucleotide substitutions, placement of introns, and nucleotide changes in the noncoding regions. Secondly, we are interested in the promoter function of this gene because of the constitutive nature of its high expression. We expected that the promoter of Gapdh-1 (where Gapdh represents glyceraldehyde-3-phosphate dehydrogenase gene) might offer a simple system for studies in transcription by RNA polymerase II. As a prerequisite for these kinds of studies, we isolated and sequenced the genes for glyceraldehyde-3-phosphate dehydrogenase from Drosophila melanogaster. This organism was chosen for its small genome size and the existence of techniques in tissue culture cell transfection (1) and germ-line transformation (2). The latter two techniques are essential in analyzing the in vitro mutated promoter and other regulatory elements, and the polynucleotides were used accordingly in the isolation of several Drosophila genes (3–5). In this article, we report the isolation and DNA sequence analysis of two Drosophila Gapdh genes. In addition, we characterized the transcription unit for both genes to aid future studies in transcriptional regulation.

EXPERIMENTAL PROCEDURES

Materials—[α-32P]dATP (410 Ci/mol) and [α-32P]dCTP (>8000 Ci/mol) were purchased from Amersham Corp. Biotinylated dUTP was from Enzogen Inc. Restriction endonucleases, T4 DNA ligase, Escherichia coli DNA polymerase I and its large fragment, and SI nuclease were purchased from Bethesda Research Laboratories. Avian myeloblastosis virus reverse transcriptase was from Life Sciences. Oligo(dT)-cellulose was from Collaborative Research Inc.

Screening the Drosophila Genomic and cDNA Library—D. melanogaster Canton-S genomic library was prepared by Maniatis et al. (6). It is a Charon 4A random-shear library from embryonic DNA with inserts of about 16 kb and terminated with synthetic EcoRI linkers. Recombinant phages were propagated in E. coli strain DP50 supF (7). One genome equivalent in this library corresponds to about 10,000 recombinant phages. D. melanogaster Oregon-R genomic library was prepared by Roger Hackett at Cornell University. The genomic DNA was partially digested to 15–20 kb with MboI and inserted into the BamHI site of λ EMBL-4 (8). E. coli strain Q559 (8) was used as host cells. D. melanogaster Oregon-R pupae cDNA library was prepared by M. Goldschmidt-Clemont at Stanford University. The library was constructed by inserting EcoRI linker ligated double-stranded cDNA into the EcoRI site of Agt10. E. coli strain Y1088 was used as host for the maintenance of this library. Each phage library was screened by the plaque hybridization technique of Benton and Davis (9) using a specific 32P-labeled DNA fragment as probe.

Nucleic Acids Isolation—Recombinant λ phage DNA was isolated by the plate lysate method of Fritsch (10). Plasmid DNA was prepared by the procedure of Clewell and Helinski (11) and Birnboim and Doly (12). Restricted DNA fragments were extracted from low-melting-point agarose by phenol after melting the gel slices at 65 °C for 10 min. Short DNA fragments (<500 bp) were purified and eluted from the polyacrylamide gels according to Maxam and Gilbert (13). Genomic DNA from D. melanogaster Oregon-R was prepared from embryos as described by Wensink et al. (14). Procedures for total RNA isolation from adult flies were those described by Simon et al. (15). Polyadenylated RNA was enriched by selection on oligo(dT)-cellulose as described by Aviv and Leder (16).

Hybridization to DNA or RNA on nitrocellulose filters was carried out in 0.5% formamide, 5 × SSC, 1 × Denhardt's solution (17), 0.2% (w/v) SDS, and 0.1 mg/ml sheared and denatured salmon sperm DNA. Heat-denatured radioactive DNA probe was added to the hybridization mixture to 1 × 106 cpm/ml. Hybridization was allowed

* This research was supported by Research Grant GM29179 from the National Institutes of Health, United States Public Health Service. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed.

1 The abbreviations used are: kb, kilobase pair(s); bp, base pair(s); SDS, sodium dodecyl sulfate.
to proceed at either room temperature (low stringency) or at 42°C (high stringency) for 24-48 h. Filters were washed once with 50% formamide, 5 × SSC, and 0.2% SDS and three times with 2 × SSC and 0.2% SDS at the hybridization temperature prior to autoradiography.

**DNA Sequence Analysis**—DNA sequencing was performed by the dideoxynucleotide chain-termination technique of Sanger et al. (18) as modified by Messing (19).

**In Situ Hybridization to the Polyten Chromosome**—Preparation of the biotin-tagged DNA fragment was essentially the nick translation procedure of Rigby et al. (20) except that 14.5 μCi biotinylated dUTP was used instead of dTTP. Polyten chromosomes were prepared from larval salivary gland nuclei as described in Simon et al. (15). In situ hybridization was performed with biotin-tagged probe according to the procedure described by Enzynge Inc.

**S1 Mapping of the Gapdh Transcripts**—S1 nuclease mapping experiments were carried out according to the following procedure. The 32P-labeled probe was synthesized in vitro from a restriction fragment primer complementary to the DNA sequence cloned in M13 mp10 or mp11. In a typical reaction, 2 μg of template was used at a molar ratio of primer to template of 1:1. The mixture was heated at 100°C for 3 min and annealed at 60°C for 30 min in 10 mM Tris-HCl, pH 7.6, 50 mM NaCl, 10 mM MgCl2, and 1 mM dithiothreitol. DNA synthesis was allowed to proceed by adding to the above mixture dGTP, dCTP, dTTP (final concentration, 20 μM), [α-32P]dATP (10 μCi, 410 Ci/m mol), and DNA polymerase large fragment (2 units). After incubation at 25°C for 15 min, the reaction mixture was chased with 2× SSC, 50 μM dUTP for 15 min. After digestion with an appropriate restriction enzyme, the reaction product was ethanol precipitated. The pellet was resuspended in 30% dimethyl sulfoxide, 10 mM Tris-HCl, pH 8.0, 0.03% xylene cyanol, 0.03% bromphenol blue, heated at 100°C for 5 min, and loaded onto a 5% polyacrylamide gel to isolate the single-stranded DNA probe (15). 32P-labeled DNA probes were then hybridized to total Drosophila RNA (10 μg) or poly(A+) RNA (1 μg) in conditions described by Treisman et al. (21). The samples were diluted to 30 μl with S1 digestion buffer containing 250 mM NaCl, 30 mM sodium acetate, pH 4.5, 1.5 mM ZnSO4, 100 μg/ml S1 nuclease, 20 μg/ml denatured calf thymus DNA, and incubated at 25°C for 30 min. The digestion products were ethanol precipitated and analyzed on a 8% polyacrylamide, 7 M urea gel.

**results**

**Isolation of the Drosophila Gapdh-1**—Amino acid sequences of glyceraldehyde-3-phosphate dehydrogenase from various organisms are highly conserved. Blocks of sequence homology containing invariant amino acid residues are observed when several glyceraldehyde-3-phosphate dehydrogenase sequences are aligned. The most conserved region is located at the catalytic site containing the essential Cys-152. A block of 12 invariant amino acid residues is contained in this region in sequence comparison among five organisms (22). It is likely that sequence homology also exists on the DNA level, and it would be possible to use a heterologous Gapdh gene sequence as a probe to isolate the Drosophila Gapdh gene by cross-hybridization. The feasibility of this approach has recently been demonstrated by Musti et al. (23). We had used a combination of full length human and rat Gapdh cDNA (24) as hybridization probes to screen a Chalcon 4A-D. melanogaster Canton-S genomic library for the Gapdh gene in this organism. Fifty thousand recombinant phages, which correspond to about five times the genome size, were screened by the plaque hybridization technique under conditions of low hybridization stringency. Several putative positive signals were observed in our primary screening, but only one positive recombinant phage was isolated upon secondary screening. The isolated phage was designated λDMGAP1. It can cross-hybridize to either human or rat Gapdh cDNA under conditions of low hybridization stringency. To further localize the Gapdh gene, DNA from λDMGAP1 was digested with various restriction enzymes, and the resulting restriction fragments were subjected to Southern blot analysis. One of the hybridizing restriction fragments, a 3.2-kb XbaI-HindIII fragment, was isolated and subcloned into plasmid pUC13 for detailed restriction mapping and sequence analysis.

**Drosophila Gapdh-1 Sequence Analysis**—To ascertain whether λDMGAP1 contains a Gapdh gene, we used the shotgun DNA-sequencing method (25) to locate and sequence the coding region of Gapdh in this recombinant phage. A 3.2-kb XbaI-HindIII fragment derived from λDMGAP1 was digested separately with restriction enzymes Sau3A1 and Ncol, and the resulting fragments were cloned, respectively, into the BamHI site and the AclI site of M13 phage mp10. M13 recombinant phages that hybridized to the rat cDNA probe were then picked and amplified to isolate the single-stranded DNA for sequencing by the dideoxynucleotide chain-termination method. Using this approach, we have determined from nine independent shotgun clones nearly the entire coding sequence of the Drosophila Gapdh gene contained in λDMGAP1. Concurrently the detailed restriction map of the 3.2-kb XbaI-HindIII fragment was determined. The restriction map and the initial shotgun sequence data allowed us to select for specific restriction fragments for further sequencing to extend and complete our analysis to include the flanking regions of the Gapdh gene. A total of about 2.4-kb DNA sequence starting from the XbaI site to the most distal ClaI site was determined. Results of the restriction mapping and the sequencing strategy are summarized in Fig. 1a. More than 98% of the sequence was obtained from both strands of the DNA and with sufficient overlaps to minimize any sequencing mistakes. The coding sequence and the 5' and 3' flanking sequences of this D. melanogaster Gapdh gene, termed Gapdh-1, are shown, respectively, in Figs. 2, 3, and 4. Assuming that the initiating methionine residue is removed after translation, the gene would code for an enzyme of 332 amino acid residues, the sequence of which is homologous to glyceraldehyde-3-phosphate dehydrogenase from other organisms. Interestingly, no intervening sequence was found in the coding region of Gapdh-1.

**Southern Blot Analysis of the Drosophila Genomic DNA**—To determine the number of Gapdh genes in Drosophila, a restriction fragment from the Gapdh-1 coding region was used as a probe to examine the genomic complexity of Gapdh-related sequences. Restriction digests of chromosomal DNA from D. melanogaster Oregon-R strain were electrophoresed on an agarose gel, transferred to nitrocellulose, and hybridized to a 32P-labeled 600-bp BglII-BglII fragment (nucleotides 36-635, Fig. 2) from the coding sequence of Gapdh-1. The result of this experiment is shown in Fig. 5b. For each digest, two restriction fragments were observed to hybridize with the probe. In all cases the lower hybridizing fragment has the molecular weight corresponding to that predicted by the Southern analysis of λDMGAP1. The additional bands seen in the genomic blot, however, could not be due to partial restriction digestion since use of excess enzymes failed to convert them to other forms. A more plausible explanation for the presence of these additional hybridizing fragments is the existence of a second Gapdh gene in the Drosophila genome.

**Isolation of the Drosophila Gapdh-2**—To isolate the second Gapdh gene we rescreened the D. melanogaster Canton-S
genomic library with a probe containing mainly the coding sequence of Gapdh-1 from ADmGAP1 (1.4-kb Xbal-PvuI fragment, Fig. 1) under conditions of low hybridization stringency to isolate as many Gapdh gene-related sequences as possible. After screening 200,000 plaques, we isolated 10 recombinants that hybridized to the Gapdh gene probe. Nine of these isolates contain the same hybridizing EcoRI fragments as in ADmGAP1, and one contains a variant shorter EcoRI fragment of Gapdh-1 generated by random shearing. We were thus unable to isolate the second Gapdh gene in this library.

We next screened a D. melanogaster Oregon-R pupae cDNA library using the same approach and were able to isolate one recombinant phage that hybridizes to the coding sequence of Gapdh-1 after screening about 50,000 plaques. This phage contains a 1.1-kb cDNA insert but lacks many of the unique restriction sites that are in Gapdh-1. DNA sequence analysis of this cloned cDNA insert revealed that it contains a truncated Gapdh-coding sequence that codes for amino acid residues 86-332 and a 3' noncoding sequence of 328 bp before polyadenylation. The coding sequence is similar but distinct from that of Gapdh-1 in many positions, and no sequence homology was observed at the 3' noncoding region. The cDNA clone is thus likely encoded by a second Gapdh gene seen in the Southern blot in Fig. 5b. This conclusion is confirmed by genomic Southern blot analysis using the 3' noncoding sequence of the cDNA clone as a probe. Only one fragment corresponding to the additional band seen previously in Fig. 5b was observed for each digest. Likewise, when the 3' noncoding sequence of Gapdh-1 was used as a probe for genomic blot only one hybridizing band was observed. The 3' noncoding sequences are thus specific for each Gapdh gene (Fig. 5, a and c).

We finally screened a newly constructed D. melanogaster Oregon-R genomic library. Only two recombinant phages that hybridized to the 3' noncoding sequence of the second Gapdh gene cDNA were obtained after screening 100,000 plaques. On the other hand, 15 recombinant phages that hybridized to Gapdh-1 3' noncoding sequences were obtained for the same size screening. The second Gapdh gene, termed Gapdh-2, is thus under-represented in both the Canton-S and Oregon-R genomic libraries. Restriction and Southern analyses of the Gapdh-1-containing phages isolated from the Oregon-R library showed, in most cases, no restriction polymorphism when compared to ADmGAP1, which was isolated from the Canton-S library. One of the recombinant phages containing Gapdh-2 was selected for further studies; it is designated ADmGAP2. A 6-kb EcoRI-EcoRI fragment containing the Gapdh-2 gene was isolated from this phage and subcloned into plasmid pUC13 for further restriction mapping and sequence analysis.

**Drosophila Gapdh-2 Sequence Analysis**—We used the same approach to locate and sequence the Gapdh-2 as described above for Gapdh-1. A total of 2.7 kb of DNA sequence starting from the 5' EcoRI site to a CiaI site distal to the gene was determined. Results of the restriction mapping and the sequencing strategy are summarized in Fig. 1b. The coding sequence of Gapdh-2 is shown along with that of Gapdh-1 in Fig. 2. Only those nucleotides that differ are shown. Again, no intervening sequence is found in the Gapdh-2 coding region, which codes for a protein product of the same chain length as does Gapdh-1. The 5' and 3' flanking sequences for this gene are given in Figs. 6 and 7, respectively. No sequence homology is observed when the flanking sequences of the two genes are compared.

**Northern Blot Analysis of Gapdh mRNA**—To estimate the size and number of the Gapdh transcripts, we analyzed the mRNA by the Northern blot method. Total and poly(A)+ RNA from adult flies were electrophoresed on a formaldehyde gel, blotted on nitrocellulose, and probed with a 32P-labeled fragment. The probe used was a 1.4-kb Xbal-PvuII fragment containing all the coding sequences of Gapdh-1 from ADmGAP1. It hybridized to a RNA species of 1.6 kb in both cases (Fig. 8). This species, however, could be resolved into two bands of similar intensities upon prolonged electrophoresis. Using the 3' noncoding sequences from two genes as specific probes, we confirmed that the upper band is the transcript of Gapdh-2 and the lower band Gapdh-1 (data not shown). Thus transcripts of both genes are present in adult flies. Assuming that the 3' poly(A) tail for mRNA is about 100-200 bases long, the actual transcript for both genes should contain about 1.3-1.5 kilobases of coding and noncoding sequences. The level of Gapdh mRNA is similar to that of actin mRNA based on semiquantitative Northern blot analysis.

**In Situ Hybridization to the Polytene Chromosome**—We
used in situ hybridization techniques to determine the cytological location of the two Gapdh genes in the chromosome squash. The first probe used was a 1.4-kb XbaI-PvuII fragment containing all of the coding region of Gapdh-1. It hybridized specifically to the right arm of the second chromosome at position 43 E-F, and no other hybridizing band was observed elsewhere in the nucleus. The possibility that the two Gapdh genes are linked was ruled out in other experiments using the 3' noncoding sequences from two genes as probes specific for each gene. Whereas the Gapdh-1 noncoding sequence hybridized to the same cytological location as its coding sequence (43 E-F), the noncoding sequence of Gapdh-2 hybridized to the X chromosome at location 13 F (data not shown). The two genes are thus unlinked. The lack of cross-hybridization in the first analysis was probably due to the stringent conditions used.

**Fig. 2.** Nucleotide and deduced amino acid sequences of *Drosophila Gapdh*-1 and Gapdh-2. Both the nucleotide and amino acid sequences are numbered from the ATG codon. Only those nucleotides and deduced amino acids that are different from Gapdh-1 are shown for Gapdh-2.
The translation initiation codon ATG is boxed. The ▽ above each nucleotide indicates sites of transcription initiation as determined by primer extension (see Fig. 9a), and the * under each nucleotide represents sites of transcription initiation as determined by S1 nuclease mapping (see Fig. 9b).

**FIG. 3. Nucleotide sequence of Gapdh-1 5′ flanking region.** The translation initiation codon ATG is boxed. The ▽ above each nucleotide indicates sites of transcription initiation as determined by primer extension, and extended to the 5′ cap site of the transcript and then hybridized to the isolated 5′ cap site. The symbols for restriction enzymes: A, BamHI; S, SstI; E, EcoRI; and X, XbaI.

**FIG. 5. Detection of two Gapdh genes in the Drosophila genome.** Restriction digests of D. melanogaster DNA (5 μg/lane) are electrophoresed on a 0.8% agarose gel, transferred to nitrocellulose paper, and probed with a nick-translated: (a) 690-bp PvuII-ClaI fragment from the 3′ noncoding region of the Gapdh-1; (b) 600-bp BglI-BglII fragment from the coding region of Gapdh-1; (c) 340-bp fragment starting from the Ttfl site (nucleotide number 983, Fig. 3) to the 3′ end of the Gapdh-2 cDNA. Symbols for restriction enzymes: B, BamHI; S, SstI; E, EcoRI; and X, XbaI.

Even in cases where it is known that the transcripts used had only one 5′ end (29), Luse et al. (28) suggested these doublets are artifacts of the reverse transcriptase reaction related to the cap structure. If this is the case, the position of the lower band would correspond to the true mRNA start site. When compared to the sequencing ladder the lower band corresponds to an A residue, 62 bases from the translation initiation codon ATG, on the noncoding strand (Fig. 3). Other than these doublets no other major extension product resulted from the cross-hybridization of the primer to Gapdh-2 mRNA was obtained.

The result of the primer extension experiment was confirmed by S1 nuclease mapping. The Ddel-ClaI fragment was again used, this time to synthesize a 32P-labeled coding strand probe up to the XbaI site in the noncoding region of Gapdh-1 (464 bases). The synthesized probe was hybridized to the Drosophila mRNA, and the unhybridized single-stranded region of the probe was then cut with the S1 nuclease. The digested product was electrophoresed along with the sequencing reaction products for molecular weight standards on an 8% polyacrylamide-urea gel. As shown in Fig. 9b, a cluster of fragments ranging from 124–127 residues long were protected by Gapdh-1 mRNA. The molecular weight of these fragments indicated the mRNA start site is about 62–65 bases from the translation initiation codon ATG, on the noncoding strand (Fig. 3). Other than these doublets no other major extension product resulted from the cross-hybridization of the primer to Gapdh-2 mRNA was obtained.

A similar strategy was used to map the mRNA start site of Gapdh-2. In this case, we used a 109-bp Hpal-ClaI fragment spanning the 5′ noncoding and coding region of Gapdh-2 (nucleotide 1053 in Fig. 6, nucleotide 105 in Fig. 2). This
The translation initiation codon ATG, the TATA box, and the CAAT extension experiment are indicated by tentatively identified as an artifact which resulted from cross-short pentanucleotide repeats in the promoter region are underlined. The 3' noncoding region is bracketed. The experiments gave four discrete cDNA products (Fig. 48, 49, which mapped the primer extension to the cap structure as discussed above. Whether the doublets in Fig. 410, 420 are 5' noncoding species represents one of the true smallest molecular weight species represents one of the true.

The intron in the mRNA start site but rather at the intron junction where mismatch occurs between the spliced mRNA and the probe.

Indeed, a consensus 3' splicing signal is located where the S1 mapping indicated. The sequence at the putative splicing junction is CAG1AU (nucleotides 1036-1040, Fig. 6) and is preceded by a stretch of pyrimidine-rich residues (30, 31). Moreover, an internal signal sequence that is conserved in the introns of Drosophila pre-mRNA, CTAAT, can also be found 26 bases from the putative 3' splicing site (32, 33). The region where the S1 nuclease method mapped thus contains all the characteristics of the 3' intron splicing site. We had searched for the 5' intron splicing signal in a sequence of about 1 kb upstream and were able to locate only one such site. The sequence, CAG1AU (nucleotides 1036-1040, Fig. 6), is located 230 bases from the putative 3' splicing site and agrees well with the consensus 5' splicing signal CAG1AU (30). Since the primer extension experiments indicated the spliced mRNAs have 5' noncoding sequences of 46, 48, and 49 bases long, the transcription start sites are thus located about 28, 30, and 31 bases from the spliced junction (Fig. 6). A consensus TATA box sequence, TATAAT, can be found at position.

FIG. 6. Nucleotide sequence of Gaphd-2 5' flanking region. The translation initiation codon ATG, the TATA box, and the CAAT box are boxed. Transcription initiation sites as determined by S1 nuclease experiment are indicated by v. The intron in the 5' noncoding region is bracketed. The * under each nucleotide represents 3' intron splicing sites as determined by S1 nuclease mapping. The short pentanucleotide repeats in the promoter region are underlined by broken lines, and the 12-bp imperfect direct repeats are underlined by arrows. The wavy lines indicate the dc and imperfect cd sequences.

FIG. 7. Nucleotide sequence of Gaphd-2 3' flanking region. The stop codon: TAA is boxed. The polyadenylation signal AATAAA is underlined. The arrow indicates site of polyadenylation in the Gaphd-2 transcript as determined in the cDNA sequence of Gaphd-2.

FIG. 8. Northern blot analysis of the Gaphd mRNA. Poly(A') RNA (5 µg) and total RNA (20 µg) were electrophoresed on a 1% agarose-formaldehyde gel, transferred to nitrocellulose paper, and probed with a nick-translated 1.4-kb XhoI-PvuII fragment from λDmGAP1. Similar results were obtained when the 3' noncoding sequences from each gene were used as probes.
Fig. 9. Mapping of transcription initiation site for Gapdh-1 by (a) primer extension and (b) S1 nuclease mapping. a, a $^{32}$P-labeled 68-bp DdeI-ClaI fragment from Gapdh-1 was used as a primer (wavy lines) for cDNA synthesis after hybridized to total fly RNA (10 µg). The synthesized cDNA was electrophoresed on an 8% polyacrylamide-urea gel along with sequencing reaction products initiated with the same primer. b, a $^{32}$P-labeled XbaI-ClaI single stranded fragment was used as S1 nuclease mapping probe (open and closed area). Total fly RNA (10 µg) was hybridized to the probe and treated with S1 nuclease (100 units/ml). The RNA-protected portion of the probe (closed area) was electrophoresed on an 8% polyacrylamide-urea gel with the sequencing reaction products mentioned in a as molecular weight markers. The numbers indicated by arrows represent fragment sizes in bp.

Fig. 10. Mapping transcription initiation site for Gapdh-2 by (a) primer extension and (b) S1 nuclease mapping. a, a $^{32}$P-labeled 108-bp HpaI-ClaI fragment from Gapdh-2 was used as primer (wavy line) for cDNA synthesis. b, a $^{32}$P-labeled XhoI-ClaI single-stranded fragment from Gapdh-2 was used as a S1 nuclease mapping probe (open and closed area). Experiment conditions are the same as described in the legend of Fig. 9. The numbers indicated by arrows represent fragment sizes in bp.

Drosophila Genes for Glyceraldehyde-3-phosphate Dehydrogenase

about ~38 from the putative transcription sites. In addition, a CAAT box can also be found at position ~90. Gapdh-2 thus contains promoter structure characteristics of the RNA polymerase II transcribed gene. Examination of the sequence upstream of the transcription initiation site of Gapdh-1, however, failed to reveal any of these features.

S1 nuclease mapping was also performed to map the 3' end polyadenylation site of the Gapdh-1 mRNA. Mapping of the Gapdh-2 mRNA 3' end was not necessary since sequencing of the Gapdh-2 cDNA already indicated that the polyadenylation site was at the base number 331 (Fig. 7) after the stop codon. A coding strand probe covering the sequence between the PvuII and ClaI site (nucleotides 38–728, 690 bases in length, Fig. 4) in the 3' noncoding region of the Gapdh-1 gene was used as a probe for S1 nuclease digestion. The length of the probe protected by mRNA from S1 nuclease digestion was determined to be around 157 bases (Fig. 11). The polyadenylation site of Gapdh-1 mRNA was calculated from this result to be around 213 bases from the stop codon (Fig. 4). Polyadenylation signal sequences, AATAAA, were observed about 62 and 15 bases upstream from the polyadenylation site, but the presence of the unique 3' end of the Gapdh-1 transcript suggests that only one of them is effective as a signal. On the other hand, multiple mRNA species had been reported because of the existence of several of these signals in the 3' noncoding region of rat cytochrome c gene (34). The same signal can be found in the noncoding sequence of Gapdh-2 gene 50 bases away from the polyadenylation site.

**DISCUSSION**

Based on the results presented in this report, D. melanogaster contains two Gapdh genes coding for two fully functional glyceraldehyde-3-phosphate dehydrogenase enzymes. Existence of isoenzymes for this protein in different organisms was investigated by Lebherz and Rutter (35). They reported that in several organisms multiple forms of glyceraldehyde-3-phosphate dehydrogenase can be found and some are expressed in a tissue-specific manner. Except for yeast, however, this phenomenon seems to be restricted to species that had ploidy changes in their evolutionary history. In a polyploidy genome, multiple copies of the same gene could diverge and give rise to different isoenzymes (36). In organisms where polyploidy does not exist, the presence of isoenzymes for glyceraldehyde-3-phosphate dehydrogenase is rare. The insect honeybee, for example, has only one form of glyceraldehyde-3-phosphate dehydrogenase (38). In Drosophila it appears that the two genes have arisen by a duplication event. Translocation of one of the genes to another chromosome could have occurred either concomitantly or subsequently.

The time of divergence for the two genes can be estimated based on the number of amino acid residue substitutions. Glyceraldehyde-3-phosphate dehydrogenase is a highly conserved protein that evolved with a UEP (unit evolutionary period, defined as the time in million years required to change 1% of the sequence information between the two lines) of about 24 (23, 37).

The derived primary structure of the two isoenzymes is homologous to each other (97.5%), having only 8 amino acid residue substitutions for a mature chain length of 332 residues. Based on this comparison, the two genes diverged approximately 60 million years ago, which was long before the speciation of Drosophila (38). It is thus likely that most of the Drosophila species contain two Gapdh genes. The conservation of nucleotides for the two genes in the coding region is much lower (89%). Many of the nucleotide changes can be
attributed to synonymous substitutions that do not result in amino acid replacement. When nucleotide changes are divided into synonymous and replacement substitutions, we found there are 99 for the former and only 10 for the latter. The corrected percentage of divergence at the synonymous sites when calculated according to Perler et al. (39) is 60%, which corresponds to a UEP of 1 or a mutation rate of 5 × 10⁻⁹ substitutions/synonymous site/year. This fast rate of change is identical to the uniform rate determined by Miyata et al. (40) for the synonymous codon substitutions in the mammalian genome.

Immediately 5' and 3' to the coding regions, DNA sequences between the two genes have diverged so substantially that homology can no longer be detected. Because of this lack of homology it is not possible to estimate the unit that was copied in the ancient duplication or translocation event. The drastic changes are most evident in the 5' noncoding regions. Here the two genes differ not only in the structures of the putative promoter but also in the placement of an intron. In this region Gapdh-1 lacks the "TATA" box at the −25- to −30-bp region that is characteristic of many pol II (RNA polymerase II) promoters. Although there are pol II promoters that lack the TATA box (41), these are exceptions rather than the general rule. Based on several lines of evidence (42) the role of the TATA box seems to position the RNA polymerase for accurate initiation. As Gapdh-1 has very discrete transcription initiation, there must be an analogous sequence that serves this function. The Gapdh-2, on the other hand, contains a consensus TATA box sequence, TATATAT, at −33 bp from the transcription initiation site. Farther upstream at −85 bp there is another sequence that has been conserved in some organisms, although it is absent in most if not all Drosophila pol II promoter. The sequence, CCAAT, is sometimes known as the CAAT box. The exact role of this sequence in promoter function is not clear.

Also upstream of the TATA box, there are six non-tandem direct repeats of a short pentanucleotide, TATGAT, in the −55- to −165-bp region of the Gapdh-2 promoter. This arrangement of short direct repeats is similar to that of the SV40 early promoter in the −30- to −100-bp region, although the repeats in that case are G/C rich. Deletions in such regions in the SV40 sequence affect its transcription, indicating that it is important for efficient transcription (43). Two of the short direct repeats in Gapdh-2 fall into tandem but imperfect repeats of 12 bp that are positioned at −100 to −124 bp. In the case of the β-globin promoter, a pair of imperfect repeats at similar positions had been determined to be crucial in gene expression (44). A sequence element ATTTGCGT (de) was also found in −132 in Gapdh-2. This element or its inverted repeat sequence (cd) have been found upstream of human and mouse immunoglobulin variable region genes and they are needed for correct transcription (45). The same elements are present in a chicken ovalbumin gene, sea urchin histone gene cluster (45), and Drosophila homoelogous gene cluster (46). An imperfect homologous cd sequence ATGAGAT is also found in −179-bp region. Neither of these elements nor the non-tandem and tandem repeats can be found in the Gapdh-1 5' flanking region. We do not know whether the drastic differences in the 5' noncoding regions of the two genes reflect differences in gene functions. Developmental profile and tissue specificity studies on their expression will be useful in answering this question. A long-term goal of this investigation will be to identify sequence information that is essential for the Gapdh expression. It should be mentioned that in an independent study, Sullivan et al. (47) have demonstrated the presence of two glyceraldehyde-3-phosphate dehydrogenase isozymes in D. melanogaster, and they too have mapped the two Gapdh genes in locations similar to those mentioned in this report using isolated recombinant phages as probes.

Acknowledgments—We thank Dr. John Lis and his colleagues for their generous supply of materials and helpful discussions, Jeff Simon for performing the in situ hybridization experiments, Dr. David Hogness for providing the cDNA library, Roger Hackett for the genomic library, Dr. David Sullivan for communicating his unpublished results, and Dr. John Lis for valuable criticisms.

REFERENCES
1. Di Nocera, P. P., and Dawid, I. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 7085–7088
2. Spradling, A. C., and Rubin, G. M. (1982) Science 218, 341–347
3. Benyajati, C., and Dray, J. F. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 1701–1705
4. Goldberg, D. A., Posakony, J. W., and Maniatis, T. (1983) Cell 34, 59–73
5. Scholnick, S. B., Morgan, B. A., and Hirsch, J. (1983) Cell 34, 37–45
6. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O’Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A. (1978) Cell 15, 687–701
Drosophila Genes for Glyceraldehyde-3-phosphate Dehydrogenase

7. Leder, P., Tiemeier, D., and Enquist, L. (1977) Science 196, 175-177
8. Karn, J., Brenner, S., Barnett, L., and Cesareni, G. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 5172-5176
9. Benton, W. D., and Davis, R. W. (1977) Science 196, 180-182
10. Fritsch, E. F. (1982) Molecular Cloning: A Laboratory Manual, pp. 371-373, Cold Spring Harbor Laboratory
11. Clewell, D. B., and Helinski, D. R. (1972) J. Bacteriol. 110, 1135-1146
12. Birnboim, H. C., and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523
13. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 32. Keller, E. B., and Noon, W. A. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 7417-7420
14. Wensink, P. C., Finnegan, D. L., Donelson, J. E., and Hogness, D. S. (1974) Cell 3, 237-251
15. Simon, J. A., Sutton, C. A., Lokell, R. B., Glaser, R. L., and Lis, J. T. (1985) Cell 40, 805-817
16. Aviv, H., and Leder, P. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1408-1412
17. Denhardt, D. T. (1966) Biochem. Biophys. Res. Commun. 23, 641-646
18. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467
19. Messing, J. (1983) Methods Enzymol. 101, 20-78
20. Rigby, P. W. J., Dieckmann, M., Rhodes, C., and Berg, P. (1977) J. Mol. Biol. 113, 237-251
21. Triesman, R., Proudfoot, N. J., Shander, M., and Maniatis, T. (1982) Cell 29, 903-911
22. Dayhoff, M. O. (1978) Atlas of Protein Sequence and Structure, Vol. 5, Suppl. 3, pp. 62-63, National Biomedical Research Foundation, Washington D. C.
23. Musti, A. M., Zehner, Z., Bostian, K. A., Paterson, B. M., and Kramer, R. A. (1983) Gene 25, 133-143
24. Tso, J. Y., Sun, X.-H., Kao, T., Reece, K. S., and Wu, R. (1985) Nucleic Acids Res. 13, 2485-2502
25. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A. (1980) J. Mol. Biol. 143, 161-178
26. Ghosh, P. K., Reddy, V. B., Swincoe, J., Lebowitz, P., and Weissman, S. M. (1978) J. Mol. Biol. 126, 813-846
27. Berk, A. J., and Sharp, P. A. (1977) Cell 12, 721-732
28. Luse, D. S., Haynes, J. R., VanLeeuwen, D., Schon, E. A., Cleary, M. L., Shapiro, S. G., Lingrel, J. B., and Roeder, R. G. (1981) Nucleic Acids Res. 9, 4339-4354
29. Proudfoot, N. J., Shander, M. H. M., Gefter, M. L., and Maniatis, T. (1980) Science 209, 1329-1336
30. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-384
31. Mount, S. M. (1982) Nucleic Acids Res. 10, 459-472
32. Keller, E. B., and Noon, W. A. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 7417-7420
33. Langford, C. J., and Gallwitz, D. (1983) Cell 32, 473-482
34. Lebherz, H., and Rutter, W. J. (1967) Science 157, 1198-1199
35. Lebherz, H., and Rutter, W. J. (1967) Science 157, 1198-1199
36. Ferris, S. D., and Whitt, G. S. (1979) J. Mol. Evol. 12, 267-307
37. Hood, L. E., Wilson, J. H., and Wood, W. B. (1975) Molecular Biology of Eucaryotic Cells, p. 279, The Benkamin/Cumming Publishing Co., Menlo Park, CA
38. Throckmorton, L. H. (1975) in Handbook of Genes (King, R. C., ed) Vol. 3, pp. 421-469, Plenum Press, New York
39. Perler, F., Efrat, A., Loames, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980) Cell 20, 555-566
40. Miyata, T., Yasunaga, T., and Nishida, T. (1990) Proc. Natl. Acad. Sci. U. S. A. 88, 81, 7417-7420
41. Nevins, J. R. (1983) Annu. Rev. Biochem. 52, 441-446
42. Parker, C. S., and Topol, J. (1984) Cell 36, 357-369
43. Fromm, M., and Berg, P. (1982) J. Mol. Appl. Genet. 1, 457-481
44. Dierks, P., Van Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J., and Weissmann, C. (1983) Cell 32, 695-706
45. Falkner, F., and Zachau, H. G. (1984) Nature 310, 71-74
46. Tsonis, P. A. (1984) Nature 312, 314
47. Sullivan, D. T., Carroll, W. T., Kanik-Ennulat, C. L., Hitti, Y. S., Lovett, J. A., and Von Kalm, L. (1980) J. Biol. Chem. 260, 4345-4350