Volume 17 Number 21 1989 Nucleic Acids Research

Sequence, structure and evolution of the gene coding for sn-glycerol-3-phosphate dehydrogenase in *Drosophila melanogaster*

Glenn C.Bewley, Julia L.Cook*, Shinichi Kusakabe[2], Terumi Mukai[2], Donna L.Rigby[1] and Geoff K.Chambers [1+]

Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695-7614, [1]Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA and [2]Department of Biology, Kyushu University, Fukuoka 812, Japan

Received July 26, 1989; Revised and Accepted September 27, 1989 EMBL accession no. X14179

ABSTRACT

We present the complete nucleotide and deduced amino acid sequence for the gene encoding Drosophila sn-glycerol-3-phosphate dehydrogenase. A transcription unit of 5kb was identified which is composed of eight protein encoding exons. Three classes of transcripts were shown to differ only in the 3'-end and to code for three protein isoforms each with a different C-terminal amino acid sequence. Each transcript is shown to arise through the differential expression of three isotype-specific exons at the 3'-end of the gene by a developmentally regulated process of 3'-end formation and alternate splicing pathways of the pre-mRNA. In contrast, the 5'-end of the gene is simple in structure and each mRNA is transcribed from the same promoter sequence. A comparison of the organization of the Drosophila and murine genes and the primary amino acid sequence between a total of four species indicates that the GPDH gene-enzyme system is highly conserved and is evolving slowly.

INTRODUCTION

Sn-Glycerol-3-phosphate dehydrogenase (GPDH:NAD^+ 2 oxidoreductase, EC 1.1.1.8) is a cytosolic enzyme present in all eukaryotic organisms. In *Drosophila melanogaster* the enzyme activity consists of a family of three distinct isozymes, designated as GPDH-1, -2, and -3, which exhibit a unique temporal and tissue-specific pattern of expression (4). GPDH-3 is the only enzymatically active form found in the larval and early pupal stages of development, where it is most concentrated in the fat body and malpighian tubule cells. In the late pupa, a developmental switch in isozyme expression occurs with the expression of GPDH-1 which is largely restricted to adult thoracic flight muscle where it participates in the glycerolphosphate cycle. GPDH-2 and GPDH-3 in the adult fly are found predominantly in the internal abdominal tissues.

Genetic studies have unambiguously demonstrated that each Drosophila isozyme is encoded by the same structural gene which maps to a single site on the left arm of the second chromosome (5) and is localized to the cytogenetic interval 26A (6). Protein structural studies have shown that these isozymes display a C-terminal amino acid sequence heterogeneity where GPDH-3 has the same C-terminal sequence as the mouse and rabbit muscle proteins, i.e. Asn-His-Pro-Glu-His-Met-COOH, while GPDH-1 is unique in that this same C-terminal sequence is extended by the amino acids Gln-Asn-Leu-COOH (7). More recently, molecular studies have demonstrated that the *Gpdh* locus in Drosophila expresses three classes of transcripts, each differing in the 3'-untranslated region and coding for a protein with a different C-terminal amino acid sequence (8). Each transcript was shown to arise through the differential expression of three isotype-specific exons at the 3'-end of the gene (8). In contrast, the multiple forms of GPDH found in vertebrate tissues are encoded by separate loci. In the mouse and human the two major isozymes have been

mapped to separate linkage groups (9,10). In the chicken, rat, and rabbit, protein structural studies strongly suggest separate structural genes encoding isozymic forms of GPDH (11–13). It has been suggested that in vertebrates an ancestral gene duplication event gave rise to the tissue-specific isozymic forms (3).

Among the proteins examined for polymorphism in natural populations, GPDH is considered to be one of the least structurally variable enzymes. Within the genus Drosophila this conclusion is based on extensive surveys for intra- and inter-specific variation upon gel electrophoresis (14), cryptic variation in the form of thermo-stability alleles (15–17), and on calculations of sequence divergence based on immunological distance (18). Generally the story that has emerged from these studies is one of evolutionary conservatism. Indeed, *D. melanogaster* is one of only four species out of over 200 surveyed which are polymorphic at the *Gpdh* locus (14), where three allozymes ($GPDH^S$-slow, $GPDH^F$-fast, and $GPDH^{UF}$-ultra fast) have been described from natural populations (19). Population surveys have demonstrated both seasonal flucuations and non-random geographical clines for the two common allozymes, $GPDH^S$ and $GPDH^F$, which has tacitly been assumed to result from some form of balancing selection (20–22). However, in attempts to establish a basis for selection at the *Gpdh* locus no consistent difference in either biochemical properties (19) or levels of expression (17,23) can be associated with any of these allozymes. To date, the basis for selection at the *Gpdh* locus remains obscure.

In this communication we describe the structure of the glycerol-P dehydrogenase gene in *Drosophila melanogaster* which includes both the complete nucleotide and deduced amino acid sequence. In addition we compare the organization of the Drosophila and murine genes and the primary amino acid sequence between a total of four species. These comparisons have confirmed the results of previous studies, utilizing more indirect measurement, which predict that the GPDH gene-enzyme system is evolving very slowly. These results now provide a framework for additional studies on the evolution of gene structure and regulation within the GPDH gene complex.

MATERIALS AND METHODS

Isolation and Characterization of Glycerol-P Dehydrogenase Clones

The overlapping genomic clones λ*Dm59* and λ*Dm60a(c)* were isolated from a genomic library constructed in Charon 4 using DNA isolated from the inbred line Canton S which bears the fast allozyme for GPDH (6). cDNA clones were isolated from both larval and adult libraries constructed in λ*gt10* using DNA from the inbred line Oregon R which bears the fast allozyme (8).

The sequencing strategy for each of the genomic clones involved the subcloning of overlapping PstI and HindIII fragments into M13mp19 and the subsequent generation of a deletion series of subclones from each primary clone utilizing exonuclease III (8). cDNA clones were sequenced utilizing synthetic oligonucleotides as primers.

Purification and Structural Analysis of GPDH Allozymes

The fast and slow allozymes were purified from two inbred lines established from natural populations in Death Valley, California, i.e. line DV1-$GPDH^F$, and DV8A-$GPDH^S$, using affinity chromatographic methods as previously described (19,24).

Structural analysis of each purified protein including comparative tryptic peptide mapping, isolation of peptides and the subsequent determination of amino acid composition and primary amino acid sequence was conducted as previously described (7,24). The results

of this analysis were compared to nucleotide deduced amino acid sequence data to derive the final amino acid sequence of the GPDH protein.

Calculation of Amino Acid Substitution Rates

Amino acid similarities between a total of four sequences were tabulated and the proportion (p) of amino acid differences between each pair was calculated as $p = n_d/n$ where n is the total number of amino acid residues compared and n_d is the number of different amino acids between paired sequences. The average number of amino acid substitutions per site between two polypeptides corrected for multiple changes and the variance were calculated as $d = -\log_e(1-p)$ and $V(d) = p/(1-p)n$ according to Kimura (25) based on the assumption of a Poisson process. Insertions and deletions between polypeptides were excluded from these calculations.

Alignment of nucleotide sequences between the *Drosophila melanogaster* and murine gene was conducted using the algorithm of Needleman/Wunseh (Intelligenetics).

RESULTS

Structure of the Glycerol-P Dehydrogenase Gene in D. melanogaster

The overlapping clones λ*Dm59* and λ*Dm60a(c)* define 22kb of genomic DNA previously shown to map at the distal end of the cytogenetic region 26A on 2L by *in situ* hybridization to polytene chromosomes, and to contain sequences that encode the gene *Gpdh* (Figure 1A). The complete nucleotide sequence of 6kb of this genomic DNA and the deduced amino acid sequence for Drosophila GPDH is illustrated in Figure 2. A transcription unit size of 5kb was identified which is composed of eight protein encoding exons (Figure 1B). A summary of the DNA sequence at the exon-intron boundaries is illustrated in Table 1. In all but one case the sequence at the 5′- and 3′-splice junction for each exon conforms to the published consensus sequence (26,27). The 5′-splice site associated with exon 3 is exceptional in that a GC is found instead of the usual GT dinucleotide. However, the analysis of cDNA clones for both GPDH-3 and GPDH-1 transcripts indicates that in both instances this splice site is functional and that accurate splicing occurs (Figure 1B).

Analysis of cDNA clones and protein sequence data demonstrates that the first five exons of the Drosophila GPDH gene are constitutively spliced while exons 6, 7, and 8 at the 3′-end of the gene are differentially expressed giving rise to three isotype specific transcripts encoding the unique C-terminal end of each isozyme (Figure 1B). The 5′-end of exon 6 encodes thirty-one amino acid residues, the sequence of which is identical to the C-terminal sequence of GPDH-3 (7). In addition, the G residue of the terminal methionyl codon of exon 6 forms the beginning of an internal 5′-splice site which is activated in a tissue specific manner (8), and which overlaps with a TAA termination codon that defines the C-terminal end of the coding region for GPDH-3 transcripts (Figure 1B, and Table 1). Exons 7 and 8 are each preceded by consensus 3′-splice sites and provide the option of two alternate splicing pathways when the internal 5′-splice site associated with exon 6 is activated (Table 1). Splicing of exon 7 to the terminal methionyl codon of exon 6 generates an open reading frame that encodes ten additional amino acid residues, Asp-Thr-Ser-Ile-Met-Pro-Ser-Pro-Lys-Leu-COOH, defining GPDH-2. Alternatively, the splicing of exon 8 to exon 6 generates an open reading frame encoding three additional amino acid residues, Gln-Asn-Leu-COOH, defining GPDH-1 (7,8). Each of these three exons also contain multiple polyadenylation signals, AAUAAA (Figures 1B and 2). Evidence that each pathway of mRNA expression is functional has previously been shown by a

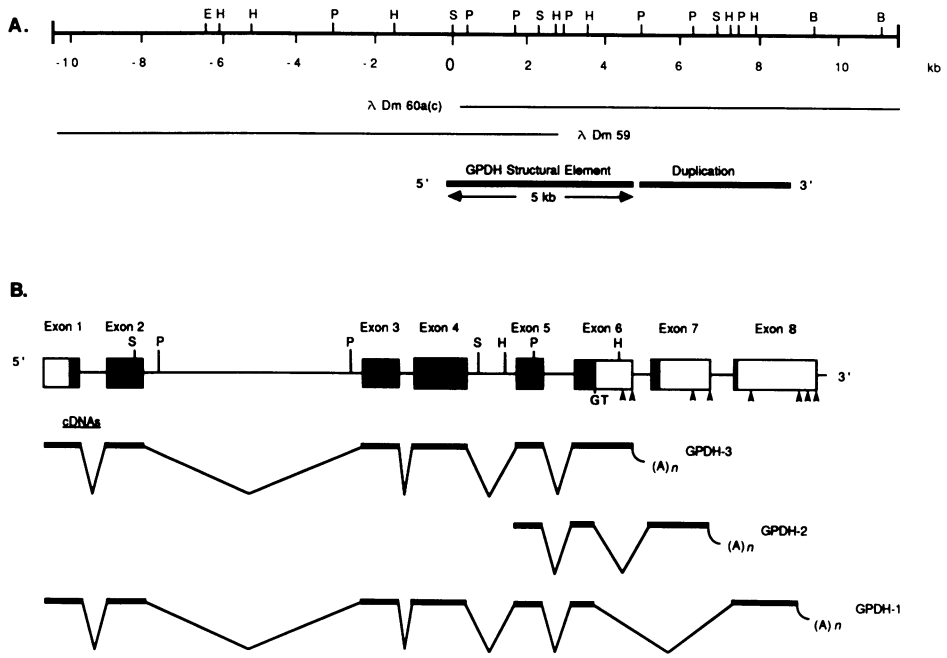


Figure 1. Molecular map summarizing the organization of the gene for glycerol-P dehydrogenase in *D. melanogaster*. Panel A represents a restriction map of approximately 22kb of genomic DNA defined by the overlapping clones λ*Dm60a(c)* and λ*Dm59* where E = EcoRI, H = HindIII, P = PstI, S = SacI, and B = Bam HI. The boldface line illustrates the 5kb GPDH structural element and a partial duplication of this element aligned to the restriction map. The zero kilobase coordinate is defined as a SacI site in exon 2 of the GPDH structural element. Panel B represents an exon diagram of the 5kb GPDH structural element and three isotype-specific transcripts defined by the analysis of cDNA clones. Shaded regions represent protein coding regions of exons and unshaded regions represent untranslated sequences. GT represents the position of an internal 5'-splice site associated with exon 6. The position of potential polyadenylation signals associated with exons 6, 7, and 8 are indicated by arrowheads. Only the 3'-end structure of the GPDH-2 transcript is shown representing the single partial cDNA clone that has been isolated.

sequence analysis of the 3'-end of isotype-specific cDNA clones and by the hybridization pattern of GPDH-specific mRNA using exon-specific oligonucleotides as probes (8).

The structure of the 5'-end of the GPDH gene is illustrated in Figures 1B and 2. A single translation initiation codon is observed that precedes the functional N-terminal amino acid residue Ala. The DNA sequence flanking this initiator codon, AAAT *ATG* G, generally conforms to the consensus sequence for functional initiator codons compiled for Drosophila genes, C/A AA A/C *ATG* (28) and no additional upstream ATG codons are noted within the noncoding leader sequence. Analysis of cDNA clones and primer extension of both larval and adult RNAs (8) have mapped a single transcription start site used in the generation of both GPDH-1 and GPDH-3 mRNAs (Figure 2), that conforms to a well conserved heptanucleotide sequence ATCA G/T T C/T found at the transcription start site for many insect genes (29). The transcriptional start site for GPDH-2 mRNA has not been determined experimentally since we have isolated only one partial cDNA clone which lacks the entire 5'-end and since GPDH-2 transcripts represent only a minor mRNA species (8). However,

there is no evidence from an examination of the genomic sequence data of Figure 2 to suggest that transcription initiation would be different. We therefore suggest that the start site used in the synthesis of each GPDH mRNA is equivalent.

The upstream promoter region contains a single TATA-box sequence, TATATT. A presumptive CAAT-box sequence, AACCAATCT, is located 239bp upstream of the TATA-box that shares a 77.8% sequence identity with the consensus GG C/T CAATCT (30). However, constraints on the distance from a functional TATA sequence question the biological significance of this sequence. A cAMP-regulatory element (CRE) homology, TGGCGCCA, is located 346bp upstream of the TATA sequence and exhibits a 75% identity to the consensus, TGACGTCA (31). Both unmatched bases represent conservative substitutions and it is not unusual for a CRE to differ from the consensus sequence by 25% or more (31). While cAMP has been shown to stimulate GPDH transcription in rodent brain (33), its effect on Drosophila GPDH has not been tested. Also noteworthy is a 34bp CT repeat beginning at position −106, although its significance is unknown.

We have also demonstrated a tandem duplication of the *Gpdh* coding element; however, this duplication does not represent a functional gene copy since it is truncated at the 5′-end and does not contain the first two exon regions. Southern blot analysis of genomic DNA has demonstrated that this duplication is polymorphic in natural populations of Drosophila (34), and in some cases exists as a triplication (35). Preliminary analysis suggests that this duplication may be a recent event since the sequence between the duplicated units has not undergone any nucleotide divergence. A detailed analysis of this duplication will be presented elsewhere (36).

Structural Features of the Primary Amino Acid Sequence

The complete primary amino acid sequence of GPDH for *Drosophila melanogaster* was derived by matching the nucleotide deduced amino acid sequence with that obtained from the sequence of tryptic peptides derived from both $GPDH^F$ and $GPDH^S$ protein. The sequence data from tryptic peptides included 91% of the total amino acid sequence with gaps occurring only for residues 1−20 at the N-terminal end, and at residues 134−138 and 142−147. This consensus sequence aligned to the sequence from three additional species is illustrated in Figure 3. The total number of residues for the Drosophila protein is 349 for GPDH-3 and 352 for GPDH-1 giving a calculated size of 38.4 and 38.8 kDa respectively which is in good agreement with published values determined by SDS-PAGE (19,37,38). The number of amino acid residues for the two mammalian sequences differ from each Drosophila sequence due to insertions-deletions between each sequence and in the case of GPDH-1, due to the three additional C-terminal amino acid residues.

The amino acid sequence can be divided into two functional domains with residues 1−118 defining an NAD-binding domain and residues 119−349 defining a catalytic domain (39). The NAD-binding domain consists of two $\beta\alpha\beta\alpha\beta$ supersecondary structural elements positioned side by side to form a twisted six stranded parallel β-sheet, β_A-β_F, overlain by four linking α -helices, αB, αC, αE and αIF (Figure 5). The four amino acid sequences illustrated in Figure 3 show especially high conservation throughout this domain and the amino acid substitutions which are observed do not substantially alter the predicted positions of the α and β-elements in the model presented for the rabbit enzyme (39). In addition, five residues have been described as invariably conserved in all NAD^+-dependent dehydrogenases (39), which include a glycine at positions 10, 12, 15 and 70 corresponding to the Drosophila sequence and aspartic acid 33. The sequences compared in Figure 3 show strong conservation for these residues with the exception of Asp-33, substituted by

```
AAGCTTCTCCTTTACGACATCGGTAATTTCATCAACTTCGGCGTGCCCAAGAGCTGTATAAAAATAGTGG -825
TTTCTTACAAATAATTATAAGAAAATAACAGCAATGTTCTAAATATTACTGACCACCGAAAATAATTAAG -755
AGTTGTAGAATAGCTAACCACTTGAGGCTGGCCATCTTTTAGTTATGACCGAATAACATTCGATTGTTGA -685
AATCTTCGACGTGTGTAAATTAGATTATTCAGAAAAATAACAGTGACAAAAGGATGACACAGTTGATTAT -615
AAAACATTAGCACTTGTTTTTGGTTGCAATTTGCTAATTTGTTGGCATTCCGAATCAAAACAGGCAAACC -545
CAATACAATTTAATAGTAAAAGGGAAAATAAATAAATAAACTTTGTAAATGTTTATATAAAATAAGAAAA -475
TAAATTCATGAAAAATAGGGGATAGGGGAATGTATGGTTTTCAAATTTTTCACGAATACCGTTAACGAAT -405
GCAAATGCACAATACCAATTTGAATTTGGCGCCACAACTGCTCACTTTGCAGCACTGAACAAGAGAATGC -335
GCAAAGGCGTTATACAGTACTGGTCAATATACCAGGGATGCAACGCTCCAAAGTTCGGCAATTGAACCAA -265
TCTAAAGAGTTTAACTATTTTAGTAAAACTACAGTTTATTTAAATAATTTCCATGGTTAAAACAATACCA -195
GGATGTACTGCGCATTGGAAAAGTGAGCGGATTCCCTGTGCGGAGCTTATCGCACCGCCAACACAAACAT -125
ACACATGCACACGGACTCTCTTCGCTCTCTTCCTCTCTCTCTCTCTCTTCGCAATCGCAACGATTTGTTG -55
                                                    ▼
TTGGCCGCAGTGCTTGTTTTGCTTATATTAGCATGTGTGGAGTGCAAACCAAGAATCAGTAGTCGCCGCG 16
GTTTCGACGTGTCAGTTTGCAGTACAATTACAATTGATTAATCGCGCCAAATCGCGGAGCCAAGTAGTAC 86
                                                          A  D  K  V
TTAGCAAGTAGCAAGAGGAAGCACATCAGCGAGTCCATCGACAACAATAACAAAATATGGCGGATAAAGT 156
 N  V  C  I  V  G  S  G  N  W
AAATGTGTGCATTGTGGGCTCCGGCAACTGGTGAGTGCACTTAATTACCCATCGCCAGATGCACTTACAT 226
AATGACCATCTAATCCACACCCGCACCCAACACCACCCCACCAACATTTTAATCTAAACCTTGCGCCCTT 296
    G  S  A  I  A  K  I  V  G  A  N  A  A  A  L  P  E  F  E  E  R  V
GCAGGGGTTCGGCCATAGCGAAAATCGTGGGCGCCAATGCCGCCGCTCTGCCGGAGTTCGAGGAGCGGGT 366
 A  M  F  V  Y  E  E  L  I  D  G  K  K  L  T  E  I  I  N  E  T  H  E
GACGATGTTCGTCTACGAGGAGCTCATCGATGGCAAGAAGCTGACGGAGATTATCAACGAGACGCACGAG 436
N  V  K  Y  L  K  G  H  K  L  P  P  N  V
AACGTCAAGTACCTGAAAGGACACAAGCTGCCCCCGAATGTGGTGAGTGTTGTTCGCTGGTTGCCAATCA 506
ATGGGGTGGACTTATGTAACGCCGGAGCGGATTGTTAACAAACAGCCGTCCGCTTACATAAAGGCTGCAT 576
ACCTTGATCTTGGCCGTGCGGACCCAAATTATCTAATCGGTCGAATCCGCGTCTATTTTTACACCTCTCT 646
CGCTCTCTCTCTCTCTCTCTGCGCTGCGCTCTGGCTGCAGTCAGATTGCACTCTTCTTGCACTTTTCAGG 716
GCCAGCATCGCACTCTCGCTCTCCGCGAAGAGAGCGCGTCCGGCCCCAGTGCCTTTGACGTAGGTCAAGG 786
CCGGTGGGCGTGCAATGGGCCACCCCCTCGCAGGGCCGCCCAGAATGCGTCAACATTCTATTAGCCGATA 856
ATTGGACAGGCAGGCGATAAGCAGGCGATCCACTTGAAACGGGTGCGTTGCTACACATGTACATGCTCCA 926
CAGGCACTCGAGCTGTTCTGTTTGCCGATTATTGATAAGTCCAACTGGTCCGGGTCGGTTGATAAAGTCC 996
ATCCATAAAGGAGAGTCAGAAGTTCATCCACGACAAAGTTTGCCCCATGCCTTGGAAAAAACCTAGCCAT 1066
GCGTTGGCAGCATGGGGATTGATAAACGTGAAAAACAGAAACCAAACAAAATGTGGGCGACACTTTTCTA 1136
TGCCGCACCGTTTGTTTGCCGAGAGAATTTACACACATATATAAATTTGGCAGACTTTCAGCGACTTTCG 1206
GCGTTTGTTGGAGAGGGGGGAATCAAAACAACAGATGAGCCAAAAGAAAATAGATAAACGAGGCTGAAAT 1276
TCTATAAGCAATGATCGCCACCAATGTCAGATATGCAAAAAGTATGCAAAAGGCAATGAATATGTGTCAG 1346
CAAGCTAATGAGCTTAATGCAACAAAGAACTAATGAAAATATTTGGAAATTCTGTTTATTTACCCGTACA 1416
GAATTACTTTTTCTCCATATGACGACGTAGTGTGGAACGAAATACGTGTCGAATAGCTTAGAAGTCTCAA 1486
TTACCTGAATTGCTAATTGCCAGCTGGTTTTTCTAGTTAAATTTACTTTTACCGATGTCAAATCATAGAT 1556
TGTTACGTTATTGCAAAATATATAACAATTATTTAAATGTCATTGCCCTAGTAAGATCATTAAAATGCAA 1626
GTAGCGCAACATCACACTATACACACTTTATAACTTAACCTTTTAAAATTTCCATTTGCTTGCCATTTTT 1696
GTCTTTTTTGTGAACCCATAATCACTATCGTCGCGCGTAGAAAAAAACAAATACATTAGCAATATGAAAT 1766
CATTTCAAATGCTACAAACAGGTTGGCGAGCCAGCATCCGGGATATGGAGAAATAATACTCTGATGTACA 1836
AAGTCACCCATATCAAGCTCAGCCGATCGATCTGACCAGCCAGTCCGCGATTTAAATTGTATAGACAGCT 1906
GAGGAACGAGCCTCCAAAGGTTGGAGTGGAATACACGTGCTGGGAAAATGATTTCCGTGATCCGTCTGCA 1976
GTTCCCTTCATTTAACCTTCAGCCCGCTTATCGGAGTTATATATTTTTTTTTTGTTTTTTTCCTTATCAC 2046
                                                  V  A  V  P  D  L  V  E  A  A
AGTAATTAGAAAAGTTTACTTTGTGGCCCCTATTCCACAGGTTGCCGTGCCCGACCTGGTTGAGGCCGCC 2116
K  N  A  D  I  L  I  F  V  V  P  H  Q  F  I  P  N  F  C  K  Q  L  L  G
AAGAACGCTGACATCCTGATCTTCGTGGTGCCCCACCAGTTCATTCCCAACTTCTGCAAACAGCTACTGG 2186
  K  I  K  P  N  A  I  A  I  S  L  I  K  G  F  D  K  A  E  G  G  G  I
GCAAGATTAAGCCAAATGCCATCGCTATCTCCCTGATCAAGGGCTTCGACAAGGCCGAGGGCGGCGGCAT 2256
 D  L  I  S  H  I  I  T  R  H  L  K
CGATCTGATCTCGCATATAATCACGCGTCACCTGAAGGCAAGTCAGCGTCGCAAAGGTGACACCGCCAGA 2326
                                  I  P  C  A  V  L  M  G  A  N  L  A  N
ATATCTCAATCCACTTGTGACTTTCCATCGCAGATCCCATGCGCCGTGTTGATGGGCGCAAACCTGGCCA 2396
  E  V  A  E  G  N  F  C  E  T  T  I  G  C  T  D  K  K  Y  G  K  V  L
ACGAGGTGGCTGAGGGCAACTTCTGCGAAACGACAATCGGCTGCACGGACAAGAAGTATGGCAAGGTGCT 2466
 R  D  L  F  Q  A  N  H  F  R  V  V  V  V  D  D  A  D  A  V  E  V  C
GCGCGATCTCTTCCAGGCGAACCACTTCCGCGTGGTGGTCGTGGACGATGCTGATGCCGTGGAGGTTTGC 2536
```

```
                .         .         .         .         .         .
G  A  L  K  N  I  V  A  C  G  A  G  F  V  D  G  L  K  L  G  D  N  T  K
GGTGCCCTCAAGAACATCGTGGCCTGCGGCGCTGGCTTTGTTGATGGCCTCAAGCTGGGCGACAACACCA 2606
 A  A  V  I  R  L  G  L  M  E  M  I  R  F  V  D  V  F  Y  P  G  S  K
AGGCGGCTGTCATCCGGCTGGGCCTCATGGAGATGATTCGCTTCGTGGACGTCTTCTATCCCGGAAGCAA 2676
 L  S  T  F  F  E  S  C  G  V  A  D  L  I  T  T  C  Y  G
GCTGTCCACCTTCTTCGAGAGCTGCGGCGTGGCCGATCTGATCACGACGTGTTACGGTAAGTGCAAAAGT 2746
TCGCCAGGGCATGTGAGTTTGGAGCTCGGTTAAATTGAGAAGGTTATCAGGAACGGGCAACAACGAGCGT 2816
AATAAGTCAACAAACTGCTAGTCAAATGCTATTATTATTTTAAGGCAAAAACAACTCTGCCAGTGACTCA 2886
CATTCAATAGTAATTGCGAATTGCTTTATTAGCCTTTACTAATAGCTGTGTAACAAATTAGTTCTGTTTA 2956
TGACTGTAAGCTTGTTACTTTAAGTTGTTTGTTTTATGTAAGGTACACTTAGGTAGCAATTACAATTTTA 3026
                                    G  R  N  R  R  V  S  E  A  F  V
GAATGTTTGATATAATAACTATGCATCTCTTCCAGGTGGCCGTAACCGCCGCGTCTCTGAGGCGTTTGTG 3096
T  S  G  K  T  I  E  E  L  E  K  E  M  L  N  G  Q  K  L  Q  G  P  P  T
ACTTCTGGAAAAACCATTGAGGAGCTGGAGAAGGAAATGCTCAACGGCCAGAAGCTGCAGGGCCCACCCA 3166
  A  E  E  V  N  Y  M  L  K  N  K  G  L  E  D  K
CTGCCGAGGAGGTCAACTACATGCTGAAGAACAAGGGTCTGGAGGACAAGTACGACTGAGCTTATTATCC 3236
                                                F  P  L  F  T  A  I  H
AATCCAATTCAAATAGTTAACTTAACACTGATCTTGTTTGCTTGCAGATTCCCCCTGTTCACGGCTATTC 3306
  K  I  C  T  N  Q  L  K  P  N  D  L  I  D  C  I  R  N  H  P  E  H  M
ACAAAATATGCACAAATCAGCTCAAGCCTAATGATTTAATTGATTGCATACGCAATCACCCTGAGCATAT 3376
 ***
GTAAGTGTTTCGTTTGCATTGGTTATGGCTCGTTTTCGCCTGTTTATCCCGCTATCCTGATACCCCGCCT 3446
TCCTGAAAATCCCACGAAATGAGCTTTGTATTCTCACCTTTAACGTTTGATTTATTATTGTTTCTTTAAC 3516
TTTTCTTATTTTAAGTTAATCTTTTTCCTATGCTAAGTTGATGCACCTTCTGTTTCTGTTTCCGTTTCTT 3586
TGATTGTTTCCCTATTATTATTTTATTAATACCTAAACTAAGCTTAACTTTAAAATTAAATAAAGTGCTA 3656
AATAATAAAACTCGATGCAAAATTTAAAATTTTTTAAATTAAAAATATTATTGGGAATGTTTAACATTCC 3726
CCAAAGAAATCATTCATAATCCCGATTCACATGCATCAGCGTCATTCGTTTTCAGAGTTATATTCAGCCA 3796
TTTATTTCGTTACATCCCGTAGTCACTACTAATCCATCGAAATAATATTATTTTCCATCCATTACAAATT 3866
                                    D  T  S  I  M  P  S  P  K  L  ***
AGTTTCCTTCTCTTCTTTTTATATTATTTTTATAGGGATACGTCCATCATGCCGTCGCCAAAACTTTAAA 3936
CTGTGTTACACCCAAACAAGATCAGTCAATGTATATTGTAAATACTATATATACCCGCATAGATCAAATC 4006
ACCAGTCATCTTCACCAAGGCCTATAAAGCTGGCATTATATAGATATATGCATAAAGTCAGAAAAATCGT 4076
TCCACGTCAACTGCATTACTTTTTCCTCTAATTCTCGCAAGTGTTTTTCAATGTGTACAGTGCTAAAGTG 4146
TTAGATAATCGAACCAAAACAAAGAGTGTCATAATTGTTGAACCTAATTGGAATTGTTAAAATAAATATA 4216
AATTAAATACAAGTTCTTTGTACATACCAACCAAAAGTGTTCCCAACTCATTTAACATTCAAAATTGGCA 4286
GCTGCAAATCGGCAAAGAGCAGTACGTACAAAAACTCCAAAATTAATAAATTTTGCAATTTTCAGTTAAT 4356
CAAGCACGGTTGTGTTTTATTTTAACGGATGTGTATTAACTAAAACTGATCGAGAAACTTTTGATTATAA 4426
                          Q  N  L  ***
ATACTTTCTTCTTTTATTTCACAGGCAAAATTTGTAAATCCCCTAAAGAAGCTGAGACGTTTTTAACATT 4496
TTTATGCACCACAGAGAATAAATGTATTTTTTCAGCACAACGCCAACTCGATGAATCTTTGGATTCCATT 4566
TTCTAACCAGTTAACTAAAAGGATCGTATAGGTTGTTGCCTTAGTTAGCTTATTGATTTTAATCCAAAAC 4636
CACGCATATCCAACGAACCACGGAAGAAAACTGACAACCAAAGTCTGCTTTAATTTTAAAATTTAGTTGT 4706
TTTCAATTCCAAACTAGCCAAGAAACAAACGCAAGTACCTAACAAATATATTGTAATTTTGTGGCCACAA 4776
GTAGCCGACGGAGCTGGCGAGAGGATCTTAGTATAATTTTAATCGGACTGCGTAGACACTAGAATTAAAT 4846
GATTACCCAAAGGTCGTTGCAAAAATAAAAGCTGAGAATTAACAAGGTTGTTTGTTTATTTTCTATCACC 4916
TAATAAAATGGGATCATTCTTTTCTAAAGATAAGTTTATCTGATGACCAATTTGAAAATAAATTTACATA 4986
CATCCTTACATAGATGGCCGCTAGGTAAAGTTTCTACATGCGCGCCTTTAAGTAATAATTCAACTATTGC 5056
TTACAAGTTTATTTTTTTAGTTGAAAACTTTTTTTTT                                   5093
```

Figure 2. The complete nucleotide sequence of the gene for Drosophila glycerol-P dehydrogenase including the complete structural element and approximately 900 bp of additional sequence extending to the first HindIII site in the 5′-flanking region (see Figure 1). The one-letter amino acid designation indicates the deduced amino acid sequence for protein coding regions of exons. The TATA box, CAAT box, a cAMP-regulatory element (CRE) homology, and potential polyadenylation signal sequences are underlined. The major transcriptional start site is indicated by an arrow head. Translation termination (TAA) codons are indicated by asterisks.

glutamic acid in Drosophila, and Gly-70 which is replaced by proline in all four species. Both substitutions can be considered to be conservative, the latter since both proline and glycine exhibit similar turn properties.

The catalytic domain of GPDH is less well conserved between species and the region

TABLE 1. EXON-INTRON ORGANIZATION OF THE DROSOPHILA GLYCEROL-P DEHYDROGENASE GENE

DNA SEQUENCE OF EXON-INTRON JUNCTIONS		
5'-BOUNDARY ----------	EXON	---------- 3'-BOUNDARY
$(\frac{T}{C})_N$ N $\frac{C}{T}$ AG ↓G	CONSENSUS[1]	$\frac{C}{A}$ A G ↓ GT $\frac{A}{G}$ A G T
	EXON 1	C T G ↓ GT G A G T
T T G C AG ↓G	EXON 2	G T G ↓ GT G A G T
C C A C AG ↓G	EXON 3	A A G ↓ GC A A G T
T C G C AG ↓A	EXON 4	A C G ↓ GT A A G T
T T C C AG ↓G	EXON 5	C A A ↓ GT A C G A
T T G C AG ↓A	EXON 6	T A T ↓ GT * A * A G T[2]
T T A T AG ↓G	EXON 7	* T A * A
T C A C AG ↓G	EXON 8	* T A * A

[1] FROM MOUNT (26), AND SHAPIRO AND SENAPATHY (27).

[2] ASTERISKS DETERMINE POSITION OF TRANSLATION TERMINATION CODONS.

of the protein linking the NAD-binding with the catalytic domain, residues 121 – 140, seems particularly divergent. Overall, the amino acid replacements observed in this domain do not suggest any substantial alterations to the secondary structure assignments based on predictions for the rabbit enzyme. On the basis of tentative structural similarities between the catalytic domains of rabbit GPDH and glyceraldehyde-3-phosphate dehydrogenase, Otto, *et al.* (39) suggests that His-94, Asn-150 and Ser-153, corresponding to positions in the rabbit GPDH sequence, are active site residues important in substrate binding and catalysis. Figure 3 illustrates that both His-95 and Asn-152, the homologous residues for the aligned Drosophila sequence, are preserved in all four species but that Ser-155 is replaced by asparagine in both Drosophila enzymes. However, Asn-155 should be able to bind the substrate phosphate group on L-glycerol-3-phosphate in the same manner proposed for the rabbit enzyme (39).

Amino Acid Substitution Conferring the Electrophoretic Mobility Difference Between Allozymes

Comparative peptide mapping of tryptic peptides derived from $GPDH^F$ and $GPDH^S$ protein show two peptides, ST14 and ST1, from the $GPDH^S$ map which are not present on the map for $GPDH^F$ (data not shown). In turn, the map for $GPDH^F$ shows peptide FT4A which is absent from the $GPDH^S$ map. The sequence of peptide FT4A can be aligned to residues 328 – 343 in Figure 3 and clearly corresponds to the combined sequence of ST14 and ST1 (Figure 4). The replacement of Asn at residue number 336 in $GPDH^F$ with Lys in $GPDH^S$ creates an additional tryptic cleavage site which explains the

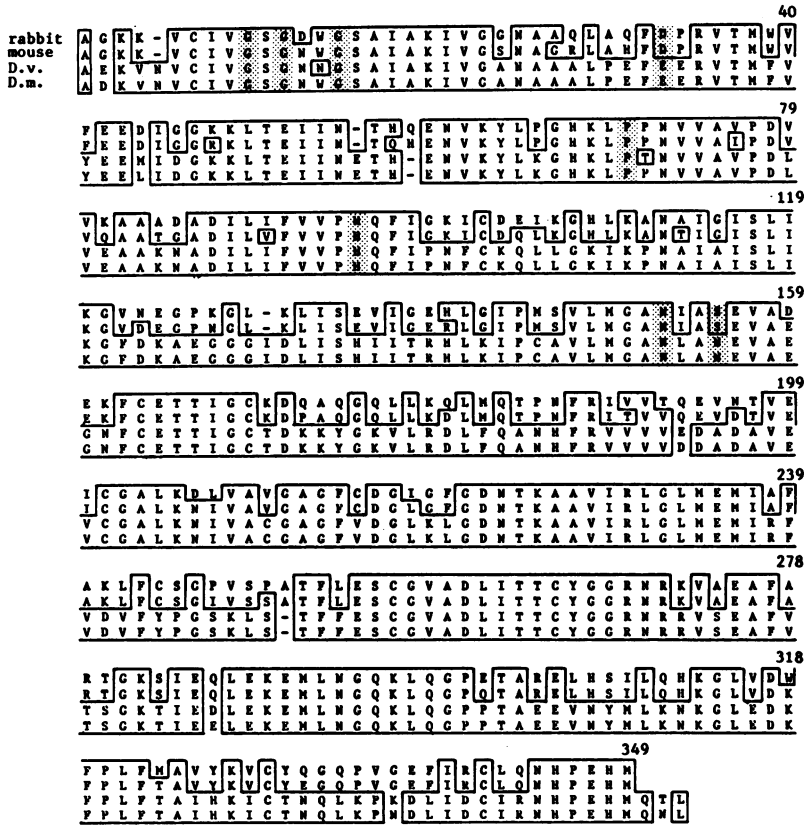


Figure 3. Comparison of the amino acid sequence of glycerol-P dehydrogenase from *D. melanogaster* to the sequence from three other species. The sequence for rabbit and *D. virilis* is derived from the amino acid sequence of overlapping peptides (39,41) while that from the mouse is deduced from nucleotide sequence data (43). All sequences are aligned to *D. melanogaster* and the residue number is for the Drosophila enzyme. Sequences that are boxed are those that share identity with the *D. melanogaster* sequence. Shaded regions represent highly conserved residues common to all NAD dependent dehydrogenases and residues predicted to be important in active site substrate binding (39).

Residue No.	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343
GPDH^F	Ile	Cys	Thr	Asn	Gln	Leu	Lys	Pro	Asn	Asp	Leu	Ile	Asp	Cys	Ile	Arg
								FT4A								
GPDH^S	Ile	Cys	Thr	Asn	Gln	Leu	Lys	Pro	Lys	Asp	Leu	Ile	Asp	Cys	Ile	Arg
				ST14								ST1				

Figure 4. Amino acid sequence of the tryptic peptides FT4A, ST14, and ST1 isolated from comparative tryptic peptide maps of GPDH^F and GPDH^S protein. The arrow indicates an additional tryptic cleavage site generated by the substitution of a Lys in GPDH^S protein for Asn in GPDH^F at residue number 336.

Table 2. Estimation of the Amino Acid Substitution Rate[1] for Glycerol-P Dehydrogenase Between Four Species

	Dm	Dv	mouse	rabbit
Dm		8(0.023)	114(0.329)	120(0.347)
Dv	0.023±0.008		116(0.335)	122(0.353)
mouse	0.400±0.038	0.408±0.038		34(0.098)
rabbit	0.426±0.039	0.435±0.040	0.103±0.018	

[1] Values above the diagonal are the absolute number of amino acid differences between paired species with the value in parenthesis representing the proportion of different amino acids between each sequence. Deletions and insertions were excluded from each computation and the total number of residues used in each comparison was 352 between Drosophila melanogaster (Dm) and Drosophila virilis (Dv), 348 between mouse and rabbit, and 346 between Drosophila and mammals. Values below the diagonal represent the average number of amino acid substitions per site between two species (d) ± the standard error.

occurrence of ST14 and ST1 on GPDH^S maps. A total of 83% of the amino acid sequence between each allozyme was examined and found to be identical except for the substitution at residue 336.

The nucleotide and deduced amino acid sequence from both genomic and cDNA clones isolated from Canton S and Oregon R libraries, each bearing the fast allozyme for GPDH, show that Asn-336 is encoded by AAT. Sequence analysis of genomic DNA isolated from an Ogasawara strain, which bears the slow allozyme, shows Lys substituted as a result of the single base change AAT (Asn) → AAA (Lys) (Kusakabe, S., Baba, H., Koga, A., Bewley, G.C., and Mukai, T., unpublished results). These results suggest that the Asn → Lys substitution confers the electrophoretic mobility difference between these allozymes. An identical substitution has been detected in lines derived from many widely distributed populations and it seems likely that identical alleles, at the level of the amino acid sequence, are segregating in natural populations worldwide (Chambers, G., unpublished results).

Amino Acid Substitutions for Glycerol-P Dehydrogenase Between Species and Calculation of Evolutionary Rates

Alignment of primary amino acid sequence data for GPDH between four species shows a strong sequence conservation with a 98 per cent similarity between *D. melanogaster* and *D. virilus* and a 67 and 65 per cent similarity between Drosophila and the mouse and rabbit sequences respectively (Figure 3). A tabulation of the proportion of amino acid differences observed between each sequence and estimates of the rate of amino acid

substitutions per site between paired species is provided in Table 2. From paleontological evidence the time of divergence between rodents and lagomorphs is estimated at 80 million years and for mammals and arthropods 700 million years (40). These values provide two independent estimates for dating an average age for the Sophophoran (*D. melanogaster*)−Drosophila (*D. virilus*) radiation at 28 million years based on the amino acid substitution rates tabulated in Table 2. This value suggests that the GPDH protein within the genus Drosophila is evolving slowly with a rate of accepted amino acid substitutions of 2.0 per 100 residues per 100 million years.

Intron Placement and Correlations with Functional Domains of the Protein

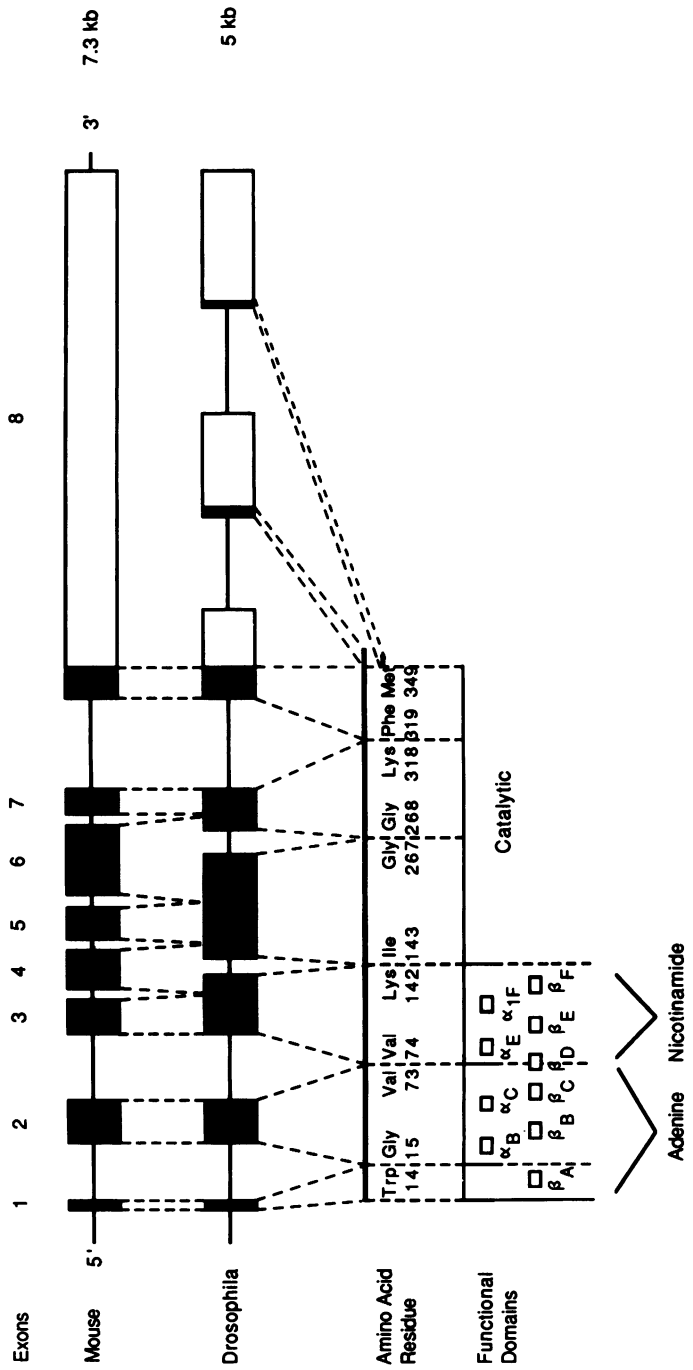
It has been suggested that exons of eukaryotic genes correspond to units which encode functional domains of proteins, and that intron placement is non-random (42). The relative importance of intron placement can be inferred by comparing the organization of functionally related genes between distantly related organisms. Since the structure of the gene encoding murine glycerol-P dehydrogenase, *Gdc-1* on chromosome 15, has been determined (43) we have aligned it to the Drosophila gene *Gpdh* (Figure 5). Although each gene differs in total size, *Gdc-1* is 7.5kb while Drosophila *Gpdh* is 5kb, each contain eight exons and encode a similar number of total amino residues (Figure 3). The principle size difference is due to the length of intron sequences and the non-coding region at the 5′-end of each gene.

Strong sequence and structural conservation is demonstrated within the 5′-end of each gene that corresponds to the region of the NAD-binding domain (Figure 3). The placement of introns 1 and 2 is both identical and highly significant where intron 1 disrupts the β_A from the α_B unit at Trp 14−Gly 15, and intron 2 separates the adenine and nicotinamide mononucleotide binding folds at Val 73−Val 74 (Figure 5). In addition the coding region of exons 1 and 2 between each gene share a 78 and 76% identity respectively at the nucleotide level. While the placement of intron 3 is not strictly conserved between species, occurring between Lys 119−Gly 120 for *Gdc-1* and Lys 142−Ile 143 for *Gpdh*, its position in both genes is close to the boundary that separates the NAD-binding from the catalytic domain and is consistent with that observed for other NAD-dependent dehydrogenases (44).

Less structural similarity is observed within the catalytic domain (Figure 5). Intron placement is more or less random between genes with exons 4−7 of *Gdc-1* corresponding to exons 3−5 of *Gpdh*. The exception is mouse intron 7 which shares an identical placement with Drosophila intron 5 between Lys 318−Phe 319 at the 3′-end of each gene. In addition the 5′-coding region of exon 8, which defines the 31 C-terminal amino acids of the mouse protein, can be aligned to the coding region of Drosophila exon 6, which defines the 31 C-terminal amino acids for GPDH-3, with a 55 per cent nucleotide sequence identity. However, extensive sequence divergence has occurred within the 1.7 kb noncoding region of mouse exon 8 and there is no similarity between this region and the 3′-end of the Drosophila gene beyond the coding region in exon 6.

DISCUSSION

In this study we have determined the complete nucleotide and deduced amino acid sequence, including 900 bp of 5′-flanking sequence, of the gene encoding Drosophila glycerol-P dehydrogenase. We previously demonstrated that the *Gpdh* locus expresses three classes of transcripts that arise through the differential expression of three isotype-specific exons at the 3′-end of the gene and that each transcript codes for an isozyme with a different C-terminal amino acid sequence (8). Furthermore, based on the results of primer extension


Exons
1
2
3
4
5
6
7
8
Mouse
5'
3'
7.3 kb
Drosophila
5 kb
Amino Acid Residue
Trp 14
Gly 15
Val 73
Val 74
Lys 142
Ile 143
Gly 267
Gly 268
Lys 318
Phe 319
Met 349
Functional Domains
Catalytic
Adenine
Nicotinamide

experiments, we predicted that all three mRNAs would be transcribed from the same promoter region (8). The results presented here confirm the complexity at the 3'-end of the Drosophila GPDH gene and demonstrate that each mRNA is transcribed from the same promoter. These features strongly suggest that the 5'-end of the GPDH gene in *D. melanogaster* is simple and is not a determinant in the generation of the three isotype-specific transcripts. Transcript heterogeneity is therefore controlled exclusively through a regulated process of 3'-end formation and alternate splicing pathways at the 3'-end of the pre-mRNA as previously suggested (8).

The primary amino acid sequence of GPDH from *D. melanogaster* shows that the only difference occurring between isozymes is at the C-terminal end of the protein. GPDH-3, encoded by exon 6-specific mRNA, has the same C-terminal end structure as the murine *Gdc-1* and rabbit muscle proteins (Figure 3). It is found predominantly in tissues actively engaged in lipid biosynthesis and in regulating the redox balance, and in this respect is similar in function to the vertebrate counterparts (2,3). GPDH-1 differs in that this same sequence is extended by three amino acids encoded in exon 8-specific mRNAs. GPDH-1 is localized in the indirect flight muscle and serves a specialized function in flight muscle metabolism through its participation in the glycerolphosphate energy shuttle. The importance of this pathway to flight muscle metabolism in insects is underscored by the fact that mutants lacking this enzyme activity are unable to maintain flight (5). Therefore, the complexity observed at the 3'-end of the gene may represent an evolutionary response to natural selection for multiple molecular forms of the enzyme each with divergent and specialized functions. In this sense, it is worth noting that GPDH-1 in *D. virilus* has the same C-terminal structure which suggests that the structure of the gene in *D. virilus* may be similar to that for *D. melanogaster*. It is possible that alternate pathways of RNA processing may be a common strategy for generating multiple forms of GPDH in insects. It is also remarkable that such small differences in primary protein structure can lead to many differences in physicochemical and kinetic parameters between isozymes (4). In this case the C-terminal differences must lead to subtle changes in conformation for each isozyme which can be proposed to (1) alter the net charge and thereby affect the native isoelectric point (45); (2) alter the stability to thermal denaturation (46); and (3) cause slight variation in active site positioning accounting for the observed differences in kinetic and ligand binding properties and in the pH optimum (2).

Comparison of the amino acid sequence of GPDH between four species demonstrates that these sequences have changed slowly over time. Based on this data, we have calculated an average evolutionary rate of 2.0 accepted amino acid substitutions per 100 residues per 100 MY which is within the range previously estimated for the genus Drosophila based on immunological distance (18) and for birds and mammals (3). This represents an overall slow rate of evolution that is somewhat comparable to that for cytochrome C (40). We have also observed that the region of the catalytic domain is changing more rapidly than the region of the NAD-binding domain. Since GPDH serves an important link between carbohydrate and lipid metabolism, distinct rates of evolutionary change within functional

Figure 5. Comparison of the exon-intron structure of the glycerol-P dehydrogenase gene between *D. melanogaster* and the mouse, and the relationship of intron position to functional domains of the protein as predicted by Otto, et al. (39). Shaded regions represent protein coding regions of exons and unshaded regions represent untranslated sequences. Lines represent regions of intervening sequence. Exons are drawn to scale while introns are not drawn to scale. Amino acid residues indicated are for the *D. melanogaster* sequence, and the relative position of regions of secondary structure within the nicotinamide adenine dinucleotide binding domain is indicated.

domains may reflect selective pressure for adaptation to different intracellular mileus and multiple metabolic roles that require divergent catalytic efficiencies.

Isozymic forms of a protein can also be considered a response to evolutionary pressure when an enzyme must operate under the constraints of several metabolic functions. For GPDH two strategies have emerged. The multiple molecular forms of GPDH in Drosophila are shown to be encoded by the same structural gene and to arise through alternate pathways of pre-mRNA processing. In vertebrates an ancestral gene duplication gave rise to the tissue-specific forms (3). Since isozymes produced through duplicate genes can sustain higher rates of accepted mutation in adapting to specialized intracellular environments, it is somewhat remarkable that the similarity between the amino acid sequence of insect and vertebrate GPDH protein is so high. In this sense, the constraints of function on evolutionary change must be substantial in this gene-enzyme system.

A comparison of the overall structural organization between the gene for Drosophila GPDH and murine *Gdc-1* shows a remarkable conservation of structure through the region encoding the NAD-binding domain at the 5'-end. The region representing the catalytic domain is more variable with respect to intron placement which suggests once again that this region of the gene may exhibit a higher rate of evolutionary change than the gene as a whole. The penultimate exon 8 of the mouse gene contains an unusually long 1.7 kb noncoding trailer region that corresponds in size to Drosophila exons 6, 7, and 8. However, no sequence homology is detected between the two genes beyond the coding region of mouse exon 8 and Drosophila exon 6. Ireland, *et al.* (43) suggest that the long noncoding region in the mouse gene may be associated with GPDH mRNA function. An alternate suggestion is that the size may simply be an evolutionary remnant of the complexity of the 3'-end of the Drosophila gene which was lost due to a gene duplication event during vertebrate evolution.

In summary, the high level of amino acid sequence identity between Drosophila and vertebrates, similarity in calculated secondary structural motifs of the protein, and structural similarity of the genes between Drosophila and mouse suggest that the Drosophila enzyme has a common evolutionary origin with vertebrates. Such a relationship can be more clearly defined by comparing the nucleotide sequence and structural divergence of the duplicate genes in vertebrates with the gene in the genus Drosophila.

ACKNOWLEDGMENTS

We thank S. McCord for technical assistance and W. Lane and Dr. D. Andrews of the Harvard Biological Laboratories Microchemistry Facility for peptide analysis and sequencing services. This work was supported in part by NIH Grants GM23617 and AG01739 to G. C. Bewley and GM21179 to R. C. Lewontin. This is paper 12313 of the Journal Series of the North Carolina Agricultural Research Service, Raleigh, NC 27695-7601.

Present addresses: *Ochsner Medical Foundation, Laboratory of Molecular Genetics, New Orleans, LA 70121, USA and +Department of Biochemistry, Victoria University of Wellington, Wellington, New Zealand

REFERENCES

1. Hansford, R.T. and Sacktor, B. (1971) Chem. Zool. 6, 213-248.
2. Bewley, G.C. and Miller, S. (1979) *In*: Isozymes: Current Topics in Biological and Medical Research, Alan R. Liss, New York, Vol. 3, pp. 23-52.
3. Ballas, R.A., Garavelli, J.S. and White, III, H.B. (1984) Evolution 38(3),658-664.
4. Bewley, G.C. (1983) *In*: Isozymes: Current Topics in Biological and Medical Research, Alan R. Liss, New York, Vol. 9, pp. 33-62.

5. O'Brien, S.J. and MacIntyre, R.J. (1972) Genetics 71, 127–138.
6. Cook, J.L., Shaffer, J.B., Bewley, G.C., MacIntyre, R.J. and Wright, D.A. (1986) J. Biol. Chem. 261, 11751–11755.
7. Niesel, D.W., Pan, Y-C.E., Bewley, G.C., Armstrong, F.B. and Li, S.S-L. (1982) J. Biol. Chem. 257, 979–983.
8. Cook, J.L., Bewley, G.C. and Shaffer, J.B. (1988) J. Biol. Chem. 263, 10858–10864.
9. Kozak, L.P., Burkart, D.L. and Hjorth, J.P. (1982) Dev. Genet. 3, 1–6.
10. Hopkinson, D.A., Peter, J. and Harris, H. (1974) Ann. Human Genet. 37, 477–484.
11. White, H. and Kaplan, N.O. (1969) J. Biol. Chem. 244, 6031–6039.
12. Ross, C.R., Curry, S., Wesson A., Schwartz, A.W. and Fondy, T.P. (1971) Arch. Biochem. Biophys. 145, 591–603.
13. Ostro, M.J. and Fondy, T.P. (1977) J. Biol. Chem. 252, 5575–5583.
14. Lakovaara, S. and Keranen, L. (1980) Hereditas 92, 251–258.
15. Bewley, G.C. (1978) Biochem. Genet. 16, 769–775.
16. Milkman, R. (1976) Biochem. Genet. 15, 383–387.
17. Laurie-Ahlberg, C.C. and Bewley, G.C. (1983) Biochem. Genet. 21, 943–961.
18. Collier, G.E. and MacIntyre, R.J. (1977) Proc. Natl. Acad. Sci. USA 74, 684–688.
19. Bewley, G.C., Niesel, D.W. and Wilkins, J.R. (1984) Comp. Biochem. Physiol. 79B, 23–32.
20. Berger, E. (1971) Genetics 67, 121–136.
21. Oakeshott, J.G., Gibson, J.B., Anderson, P.R., Knibb, W.R., Andeson, D.G. and Chambers, G.K. (1982) Evolution 36, 86–96.
22. Oakeshott, J.G., McKechnie, S.W. and Chambers, G.K. (1984) Genetica 63, 21–29.
23. Laurie-Ahlberg, C.C., Maroni, G., Bewley, G.C., Lucchesi, J.C. and Weir, B.S. (1980) Proc. Natl. Acad. Sci. USA 77, 1073–1077.
24. Chambers, G.K., Felton, A.A., Ramshaw, J.A.M., Larson Rigby, D. and Sullivan, D.T. (1985) Biochem. Genet. 23, 801–814.
25. Kimura, M. (1969) Proc. Natl. Acad. Sci. USA 63, 1181–1188.
26. Mount, S.M. (1982) Nucl. Acids. Res. 10, 459–472.
27. Shapiro, M.B. and Senapathy, P. (1987) Nucl. Acids Res. 15, 7155–7174.
28. Cavener, D.R. (1987) Nucl. Acids. Res. 15, 1353–1361.
29. Hultmark, D., Klemenz, R. and Gehring, W.J. (1986) Cell 44, 429–438.
30. O'Connell, P. and Rosbash, M. (1984) Nucl. Acids Res. 12, 5495–5513.
31. Roesler, W.J.D., Vandenbark, G.R. and Hanson, R.W. (1988) J. Biol. Chem. 263, 9063–9066.
32. Comb, M., Birnberg, N.C., Seasholtz, A., Herbert, E. and Goodman, H. (1986) Nature 323, 353–356.
33. Breen, G.A.M., McGinnis, J.F., and de Vellis, J. (1978) J. Biol. Chem. 253, 2554–2562.
34. Koga, A., Kusakabe, S., Tajima, F., Harada, K., Bewley, G.C. and Mukai, T. (1988) Proc. Japan Acad. 64, 9–12.
35. Toshiyuki, T., Kusakabe, S., Koga, A. and Mukai, T. (1989) Proc. Natl. Acad. Sci. USA 86:5000–5004.
36. Kusakabe, S., Baba, H., Koga, A., Bewley, G.C. and Mukai, T. (in preparation).
37. Niesel, D.W., Bewley, G.C., Miller, S.G., Armstrong, F.B. and Lee, C-Y. (1980) J. Biol. Chem. 255, 4073–4080.
38. Collier, G.E., Sullivan, D.T., and MacIntyre, R.J. (1976) Biochem. Biophys. Acta 429, 316–323.
39. Otto, J., Argos, P. and Rossman, M.G. (1980) Eur. J. Biochem. 109, 325–330.
40. Dayhoff, M.D. (1978) Natl. Biomedical Res. Foundation, Washington, D.C. Vol. 5, Supplement 3.
41. Arai, K., Tominaga, H., Yokote, Y. and Narise, S. (1988) Biochem. Biophys. Acta 953, 6–13.
42. Gilbert, W. (1978) Nature 271, 501.
43. Ireland, R.C., Kotarski, M.A., Johnston, L. A., Stadler, U., Birkenmeier, E. and Kozak, L.P. (1986) J. Biol. Chem. 261, 11779–11785.
44. Fukasawa, K.M. and Li, S.S.-L. (1987) Genetics 116,99–105.
45. Wright, D.A. and Shaw, C.R. (1969) Biochem. Genetics 3, 343–353.
46. Bewley, G.C., Rawls, J.M. and Lucchesi, J.C. (1974) J. Insect Physiol. 20, 153–165.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.