Alternative Splicing of Fructose 1,6-Bisphosphate Aldolase Transcripts in Drosophila melanogaster Predicts Three Isozymes*

(Received for publication, August 2, 1991)

Rhonda Shaw-Lee‡, James L. Lissemore‡, David T. Sullivan‡, and Dean R. Tolans§

From the Department of Biology, Boston University, Boston, Massachusetts 02215 and the ‡Department of Biology, Syracuse University, Syracuse, New York 13244

The genes that encode fructose 1,6-bisphosphate aldolase of Drosophila melanogaster have been isolated and characterized. These genes exist in a single copy 8-kilobase pair locus in the Drosophila genome which is located at cytogenetic position 97A-B. The nucleotide sequence and transcript mapping suggest that three overlapping protein isozyme genes may be encoded at this locus. These isozyme genes all share a single promoter, a 5'-untranslated first exon, and two other protein coding exons. The isozyme-specific carboxyl-terminal amino acids are encoded by one of three alternatively utilized fourth exons: 4A, 4B, or 4C by alternative splicing. The transcript containing exon 4C, whose sequence has been reported previously, is abundant throughout development and has a developmental profile similar to other glycolytic gene transcripts; however, it shows developmental specificity in the alternative use of two polyadenylation signals which result in a 2.4-kilobase and a 1.9-kilobase transcript. The transcript containing exon 4B is 1.6 kilobases in size and is most abundant during the larval stages and during the time of eclosion. The transcript containing exon 4A is in low abundance and found only during the adult stage. Sequence comparisons of the alternative fourth exons indicate that the duplication leading to the multiple exons is quite old and preceded the origin of the genus Drosophila.

Fructose 1,6-bisphosphate aldolase (EC 4.1.2.13) is an ubiquitous glycolytic enzyme that catalyzes the reversible aldol cleavage of fructose 1,6-bisphosphate to glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. The class I aldolases of animals and higher plants all consist of four 40,000-dalton subunits (1) with highly conserved amino acid sequences, suggesting that all arose from a common ancestral gene. In vertebrates, three tissue-specific isoforms of aldolase exist (2); aldolase A (muscle and red blood cells), aldolase B (liver, kidney, and intestine), and aldolase C (nerve tissue). The molecular forms and catalytic properties of these vertebrate isozymes have been well defined (3, 4). The classic muscle enzyme, aldolase A, and the liver enzyme, aldolase B, have distinctive properties which reflect their physiological roles in the basic metabolism of the tissues from which they are isolated (5, 6). These isozymes are encoded by three different genes that are located on separate chromosomes (7). Therefore, it is thought that the vertebrate isozymes are coordinately regulated during ontogeny by an obligatory trans-acting mechanism.

Protein sequences of many aldolase isozymes from a number of animal sources have been determined. The primary structures of aldolase A from rabbit (8), rat (9), mouse (10), and human (11) as well as the primary structure of aldolase B from rabbit (12), rat (13), chicken (14), and human (15) and aldolase C from human (16), mouse (17), and rat (18) have all been determined. The major amino acid sequence differences between vertebrate isozymes of the same species are found in the last exon which encodes the 28–31 carboxy-terminal amino acids, depending upon the species (16). In contrast, the aldolase genes for invertebrates have not been well described. The amino acid sequence of aldolase purified from Drosophila melanogaster has been determined (19); however, there are no reports of the existence of isozymes.

In this report we characterize the aldolase gene of D. melanogaster. The Drosophila genome contains a single aldolase locus. As judged from the genomic sequence and the multiple aldolase transcripts, it appears that multiple forms of aldolase with variable carboxy-terminal sequences can be generated. This would result in the production of three isozymes. These isozymes would be generated through differential splicing of three alternative 3'–terminal exons from a single primary transcript. This mode for generation of isozyme diversity is distinctly different from that used in vertebrates and indicates that isozyme function can be modified by changes in this single domain.

EXPERIMENTAL PROCEDURES

Materials—Restriction enzymes were purchased from Bethesda Research Laboratories, New England Biolabs, or Boehringer Mannheim. DNA polymerase I was from Bethesda Research Laboratories. DNA polymerase I/Klenow fragment was from Boehringer Mannheim, T4 DNA ligase was from New England Biolabs. Oligonucleotide primers were synthesized on a Millenium 6500 DNA synthesizer using phosphoramidite chemistry. [α-32P]Deoxyribonucleoside triphosphates were from Amersham Corp. Nitrocellulose BA85 and Nytran paper and filters were from Schleicher & Schuell, and radiography was done with Kodak XAF-5 x-ray film. Genomic clones were isolated from a library of D. melanogaster strain Oregon-R DNA in the vector pEML-4 which was a gift from M. Goldberg of Cornell University.

Blotting Procedures—DNA was prepared from adult D. melanogaster (20) digested with restriction enzymes, electrophoresed, and transferred to Nytran membranes according to the method of Southern (21). RNA was isolated as described previously (22) and analyzed by blotting (23). Blots were probed with random primed deoxyribonucleotide probes generated using the kit supplied by Boehringer

* This work was supported in part by National Institutes of Health Grants GM 26830 (to D. T. S.) and DK 38821 (to D. R. T.). 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 sent: Dept. of Biology, Boston University, 5 Cummington St., Boston, MA 02215.
Mannheim, or 5′-end radiolabeled oligonucleotides using T4 poly-nucleotide kinase (24).

In Situ Hybridization—The cyrogenetic localization of Drosophila aldolase was determined by in situ hybridization of the recombinant λ-Ald DNA to salivary gland chromosomes using Biotin-16-dUTP and a detection kit purchased from Enzo.

Isolation of a Genomic Clone—The rabbit muscle aldolase A clone, pRM223 (8), was used as a probe to screen a Drosophila genomic library constructed in λEMBL-4 (25).

Isolation of cDNA Clones—cDNA clones were isolated from a first instar larval αgt10 cDNA library (29). The library was screened with a 38-mer oligonucleotide primer complementary to the region of the mRNA which encodes amino acids 1–8 in an attempt to isolate full-length cDNA clones. Hybridization and washes were done at 63 °C in 4 × SSC (1 × SSC = 0.15 M sodium chloride, 0.015 M sodium citrate, pH 7.0).

Isolation of DNA Fragments for Sequence Determinations—DNA restriction fragments were isolated either by electrophoretion from agarose gels or from low melting temperature agarose (27). The isolated DNA was cloned into pUC, pSK, or M13 vectors and selected in the appropriate Escherichia coli stains (24). Plasmid DNA was isolated by a procedure described previously (28, 29). Restriction mapping of the HindIII sites in pEA4 was done by 5′-end labeling the large BglII-EcoRI fragment at the EcoRI site and subjecting it to a partial digestion with HindIII (see Fig. 2).

DNA Sequence Determination—Subclones in M13 from either the λ-Ald4 or plasmid subclones were used for DNA sequence analysis (30). A universal primer as well as gene-specific sequencing primers were used (31). The analog, T-deoxycytidine-2′-triphosphate, was substituted for dGTP to eliminate G/C compression (32).

Transcript Mapping—The positions of transcription initiation and polyadenylation were determined by comparing the results of primer extension and RNase protection assays with the genomic sequence. Primer extension using poly(A) tail was conducted as described (33) using an oligonucleotide complementary to nucleotides at positions 2672–2689 (Fig. 3). RNase protection was performed (34) using transcripts synthesized from clones in the vector pSK (Stratagene). RNA probes were verified by gel electrophoresis. The sequence of the 5′-end of the aldolase mRNA was determined using primer extension (8).

RESULTS

The Drosophila Genome Contains a Single Aldolase Region—A group of six recombinant λ clones was isolated from a λEMBL-4 genomic library using a fragment of the rabbit aldolase cDNA clone (8) as a hybridization probe. Each of these recombinant phage originated from the same genomic region as judged by their overlapping restriction maps (data not shown). One of these, λ-Ald4, was selected for further analysis because the aldolase probe hybridized in the middle of the insert DNA of the phage and was judged likely to contain the entire gene. In addition, the genome of Drosophila contains only a single region that is capable of hybridization to aldolase sequence as judged by Southern blots of genomic DNA (Fig. 1A). A single restriction map which corresponded to the λ-Ald4 clone was found in genomic DNA by hybridization to a 2.5-kbp HindIII fragment from λ-Ald4 at low stringency (Fig. 1A). Furthermore, in situ hybridization to salivary chromosomes using λ-Ald4 as a probe (Fig. 1B) revealed a hybridization signal at a single cytogenetic position at region 97A-B, which is consistent with previous genetic mapping of the aldolase gene (35).

Initially, the region of phage λ-Ald4 which contains the aldolase protein coding sequence was localized by Southern blot analysis to a 2.5-kbp HindIII fragment. However, upon localization of the start of transcription and the polyadenylation sites, the gene(s) extended over 8 kbp of this region. The abbreviations used are: kbp, kilobase pairs; kb, kilobase(s); bp, base pair(s).

FIG. 1. Genomic DNA analysis of the single Drosophila aldolase gene. Panel A, Southern blot of Drosophila genomic DNA. 5 μg of DNA was digested with BamHI and EcoRI (lane B/E), SalI and EcoRI (lane s/e), BglII and EcoRI (lane Bg/E). The blot was hybridized to the 2.5-kbp HindIII fragment of the Drosophila genomic clone at 68 °C in 4 × SSC and a final wash at 25 °C in 0.2 × SSC. Size markers were HindIII-digested ADNA. Panel B, in situ hybridization of biotinylated λ-Ald4 DNA to salivary gland chromosomes. The position of the single hybridization site is 97A-B.

Fig. 2 shows the restriction map of λ-Ald4 and the region of the aldolase gene which was sequenced. The gene has six putative exons determined by comparison of the gene sequence (Fig. 3) with the known aldolase protein sequence (19) and mapping of the transcripts. The first exon is 74 bp and encodes most of the 5′-untranslated region of aldolase mRNA. The remaining 21 bp of 5′-untranslated sequence is encoded in exon 2 which is separated from exon 1 by a 3.5-kbp intron. The deduced protein sequence from amino acid 1 to 331 is encoded by exons 2 and 3. Exon 2 (817 bp) encodes amino acids 1–265 and is precisely homologous to exons 2–7 of the vertebrate aldolase genes except that introns are not present (31). Exon 3 (200 bp) is homologous to exon 8 of the vertebrate genes and encodes amino acids 265–331. Exons 2 and 3 are separated by a small 91-bp intron. This intron splits codon 265 in precisely the same fashion as the analogous intron in vertebrates. Inspection of the region in the 3′-direction from this exon 3 revealed an additional three exons, named 4A, 4B, and 4C, which encode the carboxyl-terminal amino acids and respective 3′-untranslated regions. At each putative exon 4-intron boundary there is a plausible 3′-splice site junction and a stop codon after an invariant carboxyl-terminal tyrosine. One of these exons, 4C, would contain the information to encode a protein identical in sequence to that reported for the aldolase purified from Drosophila larvae (19).

Characterization of the 5′-End of the Aldolase Transcripts—The probing of larval mRNA with various subclones of the region upstream of exon 2 located exon 1 on λ-Ald4. The characterization of the 5′-end of aldolase mRNA was determined through primer extension, RNA sequencing, and the analysis of cDNA clones (Fig. 4). Primer extension revealed two extension products of 102 and 150 bases (Fig. 4A) using RNA from either larvae or adults. Sequence determination of the mRNA using reverse transcriptase (Fig. 4B) revealed that there was an intron in the 5′-untranslated region. This also
Fig. 2. Map of the D. melanogaster aldolase gene. The top line depicts the restriction map of the region of the genome encoded in the genomic clone. α-Ald4. Restriction sites are EcoRI (E), SalI (S), HindIII (H), BamHI (B), BgII (Bg), StyI (Sty), and StuI (Stu). The relationship of the genomic clone the right (R) and left (L) arms of the vector is depicted. Genomic subclones are depicted between this map and that of the expanded region containing the sequenced portion of the aldolase gene. Below the restriction map of this region are boxes corresponding to the locations of the exons. The shaded areas correspond to untranslated regions. Below that are lines depicting the various probes used in the text. The arrows at the bottom depict the extent of sequence determination. The mRNA encoding sequence was determined 100% in both directions.

showed that the 102-bp extension product likely was caused by premature termination. cDNA clones encoding this region confirmed both these results (Fig. 4C). The 5′-untranslated sequence of the mRNA is complimentary to the genomic sequence for 21 bp upstream from the start of translation. The remaining 74 nucleotides of 5′-untranslated sequence were located on a 1.4-kbp HindIII subclone of pEA4 using an oligonucleotide probe generated from the RNA sequence. Mapping of the remaining HindIII sites in pEA4 and sequence determination in this region located the start of transcription.

There is no consensus TATA-like sequence upstream from the start of transcription; however, 24 bp upstream there is a TAAAAA sequence.

Characterization of the 3′-End of Aldolase Transcripts and Expression—Analysis of mRNA from different developmental stages of the Drosophila life cycle using a probe containing exons 2-4A detected aldolase transcripts of three different sizes which varied in relative amounts throughout development (Fig. 5A). The sizes of these transcripts were approximately 2.4, 1.9, and 1.6 kb. Aldolase transcripts were abundant at the earliest stages of development, probably as a result of storage in the oocyte from transcription during oogenesis. Aldolase transcripts decreased in abundance during the first hours of embryogenesis and then increased by 12 h. During larval, pupal, and adult stages the amount of aldolase transcript rose, fell, and rose again in a pattern typical of gene products involved in intermediary metabolism.

Determination that each of the three putative fourth exons was present in mRNA was conducted by similar analysis using synthetic oligonucleotides from exons 4B and 4C as probes. Transcripts that hybridize to the oligonucleotide of exon 4B (Fig. 5B) were first detectable during late embryogenesis and were present through larval stages, declined in abundance during pupal stages, and showed a sharp peak close to eclosion. Exon 4C sequences (Fig. 5C) were abundant during the earliest stages of embryogenesis possibly because of synthesis during embryogenesis, declined in amount, and then increased prior to hatching. Exon 4C transcripts were found in all larval and adult stages. The pattern of exon 4C transcription during development is identical to the transcripts of other genes which encode enzymes of glycolysis. Transcripts that hybridized to an exon 4A oligonucleotide could not be detected.

The size classes of the transcripts which hybridize to each of the oligonucleotides was ascertained by reprobing blots with an exon 2-3 probe (data not shown) after initial hybridization with a specific oligonucleotide and marking the position of its hybridization. The exon 4B sequence was confined to the 1.6-kb size class aldolase transcript. Exon 4C sequences were found in the 2.4- and 1.9-kb size classes.

RNase protection analyses were used to map these aldolase transcripts relative to the gene sequence. RNA from various developmental times was used to protect various radiolabeled RNA probes. Probe 1 extended from the EcoRI site at position 2506 to the BamHI site at 5276 in exon 4C and thereby covered the entire aldolase region except the region of exon 1 (see Fig. 2). When RNA from 3-day larvae was used to protect this probe, a set of fragments of 800, 335, 305, 200, and 63 nucleotides was found (Fig. 6). The largest fragment corresponded to the coding fragment of exon 2. The 200-base protected fragment corresponded to exon 3. The 335 and 305 (which always appeared as a doublet) nucleotide fragments were found with probe 1, as well as probe 4, which extends from the HindIII at position 4164 to the same BamHI site at 5276 (see Fig. 2), using 18-h embryonic or 4-day larval RNA. These fragments represent the mature transcripts from the use of two alternatively used polyadenylation sites (positions ~4648 and ~4678 in the sequence, Fig. 3). These two polyadenylation sites each map 26 and 12 bases downstream from consensus polyadenylation signals, respectively (see Fig. 3). Finally, the 63-base fragment seen with both probes 1 and 4 corresponded to the part of exon 4C from 5219 to the BamHI site at the end of the probe.

The 3′-end of the exon 4C was mapped using probe 2, an 1100-nucleotide fragment that extends from the StyI site at position 5323 to the HindIII site at 6413 (see Fig. 2). RNA from 18-h embryos protected fragments of 1100, 650, 530, and 450 nucleotides (Fig. 6). At this developmental stage aldolase was transcribed as both 1.9- and 2.4-kb transcripts (Fig. 5). RNA from 4-day larvae, which lack the larger 2.4-kb transcript, protected just the smaller three fragments. Using this RNA the largest band of 650 bases maps a polyadenylation site very close to a polyadenylation consensus signal found at sequence position 5968. The 1.9-kb transcript, corresponding to expression of exon 4C, seen in 4-day larval RNA, was consistent with use of this site. However, the smaller fragments of 530 and 450 nucleotides were always protected by this probe. The specific origin of these fragments is unclear;

——

3 J. Lissemore and D. Sullivan, unpublished results.
FIG. 3—continued
Fig. 4. Mapping of the 5'-end of aldolase mRNA. Panel A, primer extension; 5 µg of poly(A)+ RNA was hybridized with 32P-labeled primer that was complementary to codons 14–19 of the aldolase mRNA and extended using avian myeloblastosis virus reverse transcriptase. L is larval RNA and A is adult RNA. Panel B, primer extension sequencing as in panel A except using dideoxynucleotide terminators as denoted in each of four reactions. Panel C, nucleotide sequence of the 5'-end of an aldolase cDNA clone. nt, nucleotides in length determined by co-electrophoresis of DNA of known sequence.

Fig. 5. Expression of aldolase during development. Panel A, 2.5 µg of total RNA from each stage of development as denoted at the top was probed (52) with an EcoRI-HindIII fragment extending from sequence position 2515 to 4173. Panel B, as in A except probed with a 5'-end-labeled oligonucleotide from exon 4B (position 4390–4411). Panel C, as in B except an oligonucleotide from exon 4C (position 5258–5281). Panel D, as in A, except probed with a 640-bp fragment of the rp49 gene (53).

However, this region of the aldolase transcript would be very A-T rich, and it is commonly observed that melting and subsequent cleavage of RNA duplexes occurs in A-T-rich regions. Alternatively, these fragments may represent use of nonconsensus polyadenylation signals located upstream of the site at 5968 although this explanation is not consistent with Northern blots which did not reveal 1.6–1.7-kb transcripts (Fig. 5C). The additional full-length fragment of about 1100 nucleotides found using the RNA from 18-h embryos, but not with the 4-day larval RNA, would correspond to the 2.4-kb transcript found in this RNA. The size indicated that there was an additional polyadenylation signal downstream of the HindIII site.

The mapping of the 2.4-kb transcript and interpretation of the results using probe 2 were confirmed using probe 3, which extends from the HindIII site at sequence position 6413 downstream to the EcoRI at the end of the clone (see Fig. 2). A protected fragment of 160 bases was found using 18-h embryo or 13-day adult RNA but not 4-day larval RNA (Fig. 6). This was consistent with the presence of the 2.4-kb aldolase transcript at these times in development (Fig. 5C). The 2.4-kb transcript encoded by exon 4C uses a polyadenylation site which maps 29 bases downstream from the last of a series of four closely spaced polyadenylation signals (position 6542; see Fig. 3). The 1350 bases transcribed from exon 4C, plus exons 1–3, equal approximately 2.4 kb.

The protection of probe 5, which extends from an StuI site at position 3638 to a HindIII site at position 4164 (see Fig. 2), used 5 µg of poly(A)+ RNA from larval and adult flies. This probe protected an 80-nucleotide fragment corresponding to exon 3 from both RNAs as well as less abundant fragments of 280, 225, and 190 nucleotides whose origin is not clear (Fig. 6). However, a 143-base fragment was present only from adult RNA. This fragment corresponded exactly with the expected size of a transcript from the 3'-splice site in exon 4A and the HindIII site of the probe. The 143-nucleotide fragment was present in many fold lower abundance than the 80-nucleotide fragment, indicating that expression of exon 4A was very low compared with the other aldolases and was primarily expressed in adults. This was consistent with the failure to detect exon 4A using an oligonucleotide complementary to this putative exon to probe northern blots. The 3'-polyadenylation site for this transcript would be difficult to map because of the relative levels and overlap with transcripts from exon 4B.

Evolution of the Drosophila Aldolase Isozymes—Alignment of the three carboxyl-terminal protein coding regions found in the Drosophila gene with the comparable regions of the human aldolase A, B, and C isozymes and the cytosolic
envelope of the Drosophila Aldolase Gene

The Drosophila locus for aldolase has been characterized and has been found to have the potential to encode three different aldolase isozymes by differential splicing of alternative 3'-exons. There are likely five aldolase transcripts that result in three different aldolase isozymes expressed at different stages of development. A model showing the structure of each of the putative aldolase transcripts is shown in Fig. 8. The number of nucleotides of exon 1 is 74, in exon 2, 817, and in exon 3, 200. Each transcript would have these 1091 bases in common. The five 3'-ends would vary depending upon choice of splice site and poly(A) addition site. The two transcripts encoding exon 4B are nearly equal at 1.6 kb in size, and those encoding exon 4C are in two widely different size classes, both of whose generation differs depending upon the choice of polyadenylation signals. The transcript encoding exon 4A is relatively rare and could only be detected in poly(A)+ RNA from adults. Its 3'-end has not been determined but could be the same as one of the exon 4B transcripts, which would lead to a 1.7-kb transcript.

In Drosophila there is direct evidence for the existence of an aldolase that contains the carboxyl-terminal amino acids encoded by exon 4C. This isozyme has been purified and sequenced (19). In addition, it seems probable that the proteins whose carboxyl termini would be encoded by exons 4A and 4B are likely to exist. There is direct evidence from RNase protection that transcripts which contain exon 4A and 4B are found in total and poly(A)+ RNA. In addition, sequences encoding exon 4B have been found in cDNA (data not shown). Since transcripts containing exons 4C and 4B are relatively abundant and are present in poly(A)+ RNA, it is unlikely that RNA that hybridizes to exons 4A and 4B probes is found only in mRNA precursors. Each of the fourth exons has an intact reading frame without nonsense codons. The exon 4s are quite divergent from one another, probably having originated via partial gene duplications in the distant past. Comparison of these exon 4s with each other shows 42% overall similarity. The two genes for glyceraldehyde-3-phosphate dehydrogenase in D. melanogaster, another glycolytic enzyme which arose as a result of a gene duplication early in the evolution of the genus Drosophila, shows 97.6% similarity. This suggests that if these aldolase exons and the two glyceraldehyde 3-phosphate dehydrogenase genes are evolving at similar rates the alternative fourth exons of the aldolase gene arose long before the origin of the genus Drosophila.

It is unlikely that the coding potential of each exon 4 would be retained over such a long period if the exons had no coding function. Furthermore, there are eight amino acid positions in the carboxyl-terminal region which are conserved in all aldolase sequences. In addition to these invariant positions (boxed) the proteins encoded by the Drosophila exons 4A, 4B, and 4C share a total of 42% of the same amino acids (shaded). Exon 4C is most like exon 4B, with exon 4A having the least in common between the other two exons, which is consistent with its probable specialized expression in adult tissues. Exon 4C is most similar to the vertebrate muscle isozyme, aldolase A, with 65% amino acid identity.

**DISCUSSION**

The Drosophila locus for aldolase has been characterized and has been found to have the potential to encode three different aldolase isozymes by differential splicing of alternative 3'-exons. There are likely five aldolase transcripts that result in three different aldolase isozymes expressed at different stages of development. A model showing the structure of each of the putative aldolase transcripts is shown in Fig. 8. The number of nucleotides of exon 1 is 74, in exon 2, 817, and in exon 3, 200. Each transcript would have these 1091 bases in common. The five 3'-ends would vary depending upon choice of splice site and poly(A) addition site. The two transcripts encoding exon 4B are nearly equal at 1.6 kb in size, and those encoding exon 4C are in two widely different size classes, both of whose generation differs depending upon the choice of polyadenylation signals. The transcript encoding exon 4A is relatively rare and could only be detected in poly(A)+ RNA from adults. Its 3'-end has not been determined but could be the same as one of the exon 4B transcripts, which would lead to a 1.7-kb transcript.

In Drosophila there is direct evidence for the existence of an aldolase that contains the carboxyl-terminal amino acids encoded by exon 4C. This isozyme has been purified and sequenced (19). In addition, it seems probable that the proteins whose carboxyl termini would be encoded by exons 4A and 4B are likely to exist. There is direct evidence from RNase protection that transcripts which contain exon 4A and 4B are found in total and poly(A)+ RNA. In addition, sequences encoding exon 4B have been found in cDNA (data not shown). Since transcripts containing exons 4C and 4B are relatively abundant and are present in poly(A)+ RNA, it is unlikely that RNA that hybridizes to exons 4A and 4B probes is found only in mRNA precursors. Each of the fourth exons has an intact reading frame without nonsense codons. The exon 4s are quite divergent from one another, probably having originated via partial gene duplications in the distant past. Comparison of these exon 4s with each other shows 42% overall similarity. The two genes for glyceraldehyde-3-phosphate dehydrogenase in D. melanogaster, another glycolytic enzyme which arose as a result of a gene duplication early in the evolution of the genus Drosophila, shows 97.6% similarity. This suggests that if these aldolase exons and the two glyceraldehyde 3-phosphate dehydrogenase genes are evolving at similar rates the alternative fourth exons of the aldolase gene arose long before the origin of the genus Drosophila.

It is unlikely that the coding potential of each exon 4 would be retained over such a long period if the exons had no coding function. Furthermore, there are eight amino acid positions in the carboxyl-terminal region which are conserved in all aldolase sequences, and these eight are conserved in each fourth exon of the Drosophila gene (Fig. 7). In addition, the strong codon bias often noticed in genes of Drosophila is present in the fourth exons of the aldolase genes (36). Taken together this provides strong evolutionary evidence that each of the fourth exons is translated and defines two new aldolase sequences not established previously, although a final confirmation awaits the identification and characterization of these aldolase isozymes in Drosophila.

The generation of isozyme diversity by alternating between carboxyl-terminal sequences is consistent with evidence from the vertebrate aldolase isozymes. This region is thought to be partially responsible for the ability of aldolase isozymes to discriminate between the metabolic substrates, fructose 1,6-bisphosphate and fructose 1-phosphate (37, 38). Indeed, site-directed mutagenesis, partial proteolysis, and domain switching have revealed that this isozyme-specific region modulates discrimination between these two substrates (38). This is an important function that any putative aldolase isozyme must perform. Distinct isozymes in Drosophila which differ only in their carboxyl-terminal domains would add clear evidence for the role of this region in the enzyme as the minimal distinguishing feature between class I aldolase isozymes.

This analysis is the first that indicates the likelihood of

---

2 D. Tolan, unpublished observations.

---


isozymes for this glycolytic enzyme in insects. Including this report, only three Drosophila glycolytic enzymes are known to have isozymic forms. In addition to aldolase, sn-glycerol-3-phosphate dehydrogenase has three isozymes (39), and like aldolase, these differ from each other at their carboxyl termini. The three isozymes of glycerol-3-phosphate dehydrogenase are generated by alternative processing of the transcript of a single gene (40). The third Drosophila glycolytic enzyme that has multiple forms is glyceraldehyde-3-phosphate dehydrogenase. However, in the case of the latter the two isozymes are encoded by two unlinked genes (41).

The mammals and insects have evidently used different routes during their evolution for the generation of three aldolase isozymes. Mammalian aldolases are encoded by genes on different chromosomes. Theories of evolution postulate that vertebrates arose from primitive chordates via two tetraploidization events (42, 43) which have left significant synteny between which vertebrate isozymes evolved. The chromosomal location and divergence of the vertebrate aldolase genes support this general scheme for isozyme evolution (7). Isozyme diversity in insects is not based upon ploidy changes. Examples of the mechanisms used by insects include tandem gene duplication, e.g. Adh (46) and other genes, and alternative processing of the transcript of a single gene, e.g. aldolase, glycerol-3-phosphate dehydrogenase (40), and sex determination genes (47). Isozyme diversity generated by gene duplication and resulting in unlinked genes, e.g. glyceraldehyde-3-phosphate dehydrogenase, has apparently been lost during the evolution of the lineage genes (47). Isozyme diversity generated by gene duplication and resulting in unlinked genes, e.g. glyceraldehyde-3-phosphate dehydrogenase is an uncommon mechanism in insects. In the case of glyceraldehyde-3-phosphate dehydrogenase, this gene duplication may have occurred through a retrotransposition event. Isozyme diversity in mammals has used a variety of mechanisms, including alternative splicing (48), which is particularly common among muscle proteins (49) and includes a glycolytic enzyme, the two M-type isozymes of pyruvate kinase (50) although, in this case further isozyme diversity is created by other pyruvate kinase genes at different chromosomal loci.

This gene is another example which supports the model for the loss of introns during evolution, and in keeping with numerous observations on other Drosophila genes it is apparent that the extent of intron loss has been more extensive in the Drosophila lineage. It seems likely that the primordial aldolase gene had a number of introns, at least five. Have apparently been lost during the evolution of the lineage leading to Drosophila leaving a large exon 2. In addition, the introns downstream are smaller relative to those found in vertebrates (92-bp splitting codon 265 versus 132-411 in vertebrates (9, 14, 16, 31)). The origin of the alternative fourth exons of Drosophila is not clear, but the most reasonable and simplest hypothesis is that they have been generated by duplications of an existing single carboxyl-terminal exon of an aldolase gene. If this is the case, then the introns that separate these exons were at the time of the duplication event new introns which utilized an existing 5′-donor splice site and a new 3′-acceptor site.

Acknowledgments—We thank Mary Tulipani, Peter Beernink, and Joanne Woodward for help with many of these experiments. Oligonucleotides were synthesized at the Boston University DNA Synthesis Facility.

REFERENCES
41. Sullivan, D. T., Carroll, W. T., Kanik-Enmulat, C. L., Hitti, Y.
Alternative Splicing in the Drosophila Aldolase Gene

42. Ohno, S. (1973) Nature 244, 259-262