Structure, Expression and Duplication of Genes Which Encode Phosphoglyceromutase of Drosophila melanogaster

Peter D. Currie¹ and David T. Sullivan²

Department of Biology, Syracuse University, Syracuse, New York 13244-1270

Manuscript received February 5, 1994
Accepted for publication June 8, 1994

ABSTRACT

We report here the isolation and characterization of genes from Drosophila that encode the glycolytic enzyme phosphoglyceromutase (PGLYM). Two genomic regions have been isolated that have potential to encode PGLYM. Their cytogenetic localizations have been determined by in situ hybridization to salivary gland chromosomes. One gene, Pglym78, is found at 78A/B and the other, Pglym87, at 87B4.5 of the Drosophila polyploidy map. Pglym78 transcription follows a developmental pattern similar to other glycolytic genes in Drosophila, i.e., substantial maternal transcript deposited during oogenesis; a decline in abundance in the first half of embryogenesis; a subsequent increase in the second half of embryogenesis which continues throughout larval life; a decline in pupae and a second increase to a plateau in adults. This transcript has been mapped by cDNA and genomic sequence comparison, RNase protection, and primer extension. Using similar analyses transcripts of Pglym87 could not be detected. Pglym78 has two introns which interrupt the coding region, while the Pglym87 gene lacks introns. This and other features support a model of retrotransposition mediated gene duplication for the origin of Pglym87. The apparent absence of a complete, intact coding frame and transcript suggest that Pglym87 is a pseudogene. However, retention of reading frame and codon bias suggests that Pglym87 may retain coding function, or may have been inactivated recently, substantially after the time of duplication, or that the molecular evolution of Pglym87 is unusual. Similarities of the unusual molecular evolution of Pglym87 and other proposed pseudogenes are discussed.

As part of an ongoing study to characterize the control of expression of glycolytic genes of Drosophila melanogaster, a significant subset of these genes has been isolated and characterized (Roselli-Rehuis et al., 1992; Shaw-Lee et al. 1991, 1992; von Kalm et al. 1989; Sullivan et al. 1985). This study details the molecular characterization of an additional member of this set of genes, the gene that encodes phosphoglyceromutase (PGLYM). PGLYM catalyzes the interconversion of 2-phosphoglycerate and 3-phosphoglycerate. PGLYM from insects has not been well characterized. To our knowledge, there have been no reports concerning this enzyme or its gene from Drosophila. In mammals, two genes encode two isozyme subunits, each subunit is 30 kD. The enzyme is a dimer so three isozymes of PGLYM may be found in mammalian tissues. There is a muscle specific subunit, M, and a non-muscle-specific, or brain, subunit B, which is synthesized in many tissues (Omen and Cheung 1974). The homodimer, MM form, is found mainly in mature muscle; the BB form, is found mainly in liver, kidney and brain; and the heterodimer MB form, along with varying proportions of the MM and BB forms, is found mainly in the heart (Omen and Cheung 1974). In addition, PGLYM in humans is developmentally regulated. Early in development fetal muscle contains mainly the nonspecific or brain form. At approximately 80-100 days of gestation, the muscle specific form first appears and becomes the predominant form thereafter (Omen and Cheung 1974; Miranda et al. 1988).

Genes which encode human, rat, rabbit and mouse isozymes have been characterized and comparisons reveal a substantial conservation in nucleotide and amino acid sequence (Urena et al. 1992; Castella-Escala et al. 1990; Yanagawa et al. 1986; LeBoucH et al. 1988). In humans there is a single copy of Pgym-m per haploid genome. This is in contrast to a large family of Pgym-b genes, many of which have been demonstrated to be pseudogenes (Castella-Escala et al. 1990). It has been proposed that multiple functional copies of Pgym-b once existed within the Pgym-b gene family and one evolved to form Pgym-m. This hypothesis is supported by phylogenetetic comparisons which indicate that Pgym-m has diverged from this multigene family relatively recently (Sakoda et al. 1988; LeBoucH et al. 1988).

Transcriptional regulation of muscle specific Pgym-m has been investigated in some detail. Promoter deletion studies have defined a single binding site in the Pgym-m promoter for the MEF-2 protein which is involved in tissue and temporal expression in muscle (Nakatsuji et al. 1992). Band shift assays using promoter fragments demonstrated that protein binding to this site was only

¹ Present address: ICRF Developmental Biology Unit, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.
² To whom correspondence should be addressed.

detected using nuclear proteins derived from muscle extracts. Thus, the mammalian muscle specific isoform is thought to be under direct control of genes that stimulate muscle differentiation (NARATSUJI et al. 1992). Saccharomyces cerevisiae PGLYM has also been well characterized. Both the nucleotide and amino acid sequence have been independently determined (HEINSCH et al. 1991; FOTHERGILL-GILMORE and WATSON 1989). The yeast enzyme is composed of four identical 27-kD subunits encoded by the Gpm1 gene. The crystal structure of the enzyme has been solved (FOTHERGILL-GILMORE and WATSON 1989). One important feature learned from these studies has been the demonstration that two histidine residues are involved in catalysis at the active site of this enzyme. The essential role of these histidines for catalytic activity has been demonstrated further by site-directed mutagenesis (WHITE and FOTHERGILL-GILMORE 1992).

The biochemical properties of PGLYM from the bacteria, Streptomyces coelicolor, the only prokaryote from which the PGLYM encoding gene has been cloned and characterized, indicate that this enzyme has properties similar to the yeast PGLYM (WHITE et al. 1992). The subunit molecular mass is 28 kD and the native mass determined by gel filtration was estimated at 120 kD. It was assumed therefore that the S. coelicolor enzyme is a tetramer, similar in form to the yeast enzyme. The bacterial enzyme possesses the greatest amino acid sequence similarity to the yeast enzyme but also has a high level of amino acid similarity to mammalian PGLYM, reinforcing the view that glycolytic genes have been highly conserved during evolution (WHITE et al. 1992). Results reported here provide the initial characterization of an insect Pglym gene. In D. melanogaster it appears there is one functional gene and a second sequence which may be a pseudogene.

MATERIALS AND METHODS

DNA isolation and preparation: Plasmid DNA, propagated in Escherichia coli strain DH5α, was isolated in small quantities by the method of MORELLE (1988). Large scale preparation of plasmid DNA was performed by as described in MANIATIS et al. (1989). Single-stranded DNA from M13mp18 or M13mp19 sequencing vectors, large scale bacteriophage λ DNA using a plate lysis method or λ lysin in liquid culture were isolated using methods described in MANIATIS et al. (1989). Drosophila genomic DNA was prepared in small scale via the method of LIS et al. (1983). Large scale isolation of Drosophila genomic DNA was performed using a modified method of R. LIFTON (BENDER et al. 1983).

Isolation and radiolabeling of DNA fragments: Agarose gel slices containing DNA restriction endonuclease fragments were visualized and excised from ethidium bromide containing 1% agarose gels and electroeluted according to the dialysis tube method of MANIATIS et al. (1989). A 125–500-ng sample of isolated restriction fragment DNA was random primed utilizing a random priming labeling kit as specified by the manufacturer (Boehringer Mannheim). Unincorporated nucleotides were separated from labeled DNA via passage over a Sephadex G-50 column by a method specified by the manufacturer (Boehringer Mannheim).

Screening of bacteriophage libraries: Genomic clones were isolated from an adult Oregon-R D. melanogaster strain library constructed in the λ vector EMBL-4 (SHAW-LEE et al. 1991). cDNA clones were isolated from an adult Oregon-R strain D. melanogaster library constructed in the vector Agt1 (VON KALM et al. 1989). Then 20,000 pfu of the respective libraries were plated with 400 µl of an overnight culture of the appropriate host strain per large NZCYM plate, and plates were allowed to grow to near confluence. Plates were incubated at 4° for 1 hr to harden top agarose. The libraries were then screened according to the method of BENTON and DAVIS (1977).

Southern and Northern blot analysis: Southern blot transfer was carried out following the method of SOUTHERN (1975) with the adaptations described by the membrane manufacturer (Schleicher & Schuell). Northern blot analysis was adapted from LISSEMORE et al. (1987) using Genescreen membranes according to the manufacturers instructions (New England Nuclear). Modifications included soaking the electrophoresis rig, tray, comb and peristaltic pump recirculation tubing in 0.1% diethyl pyrocatechol overnight before use. We used 200 µl/cm of prehybridization buffer (50% deionized formamide, 5 × SSC, 5 × Denhardt’s solution, 100 µg/ml sonicated denatured salmon sperm DNA, 40 mM sodium phosphate, pH 6.8, 0.5% bovine serum albumin, 1% sodium deoxycholate (SDS), and 10% dextran sulfate) and 100 µl/cm² of hybridization buffer (50% deionized formamide, 5 × SSC, 100 µg/ml sonicated denatured salmon sperm DNA, 40 mM sodium phosphate, pH 6.8, and 10% dextran sulfate). Prehybridization was for 4–6 hr at 42° and hybridization was at 42° for 12–16 hr.

RNase protection: One microgram of linearized SK+ plasmid DNA was used as a template for transcript synthesis using T3 or T7 polymerase and [α-32P]CTP (NEB) utilizing the Ribobasica II transcription kit and a method supplied by the manufacturer (IBI). Two micrograms of IBI RQ1 DNAse I were added to the labeled transcript and incubated at 37° for 30 min. Twenty microliters of 50 mM EDTA were added, and the transcript was purified over a Sephadex G-50 RNA spin column (Boehringer Mannheim). Labeled transcript (20–50,000 cpm) was added to an appropriate amount of total RNA that had been dried under vacuum and resuspended in 30 µl of hybridization buffer (80% formamide, 40 mM PIPES, pH 6.4, 0.4 mM NaCl, 1 mM EDTA) and heated to 85° for 10 min. The sample tubes were then kept at 55° overnight, 300 µl of digestion buffer (0.1 M Tris-Cl pH 8.0, 0.5 M EDTA, 0.5 mM NaCl, 40 µg/ml heat-treated RNase A (Sigma) and 2.87 units/ml RNase T1 (BRL)) were added, and the reaction mixture was incubated at 30° for 30 min. Twenty microliters of 10% SDS and 2.5 µl of 20 mg/ml protease K were then added, and the reaction mixture was incubated at 37° for 30 min. Extractions were performed with the addition of an equal volume of phenol and 0.5 µl of 10 µg/ml yeast RNA was added to the extracted aqueous phase. Reaction products were precipitated by the addition of an equal volume of ethanol, rinsed with 70% ethanol and the pellets dried under vacuum. Pellets were resuspended in loading dye (U. S. Biochemical Corp.) and heated to 85° for 5 min. Protection products were visualized following electrophoresis on a 1 X TBE, 6% acrylamide gel by standard autoradiographic techniques.

Primer extension: Primer extension was performed essentially as described by MANIATIS et al. (1989) with the following modifications: 5–30 µg of poly(A) RNA were used depending on the stage and primer used in each individual experiment; 10–100,000 cpm of the labeled primer were added to the specified amount at poly(A) RNA; annealing temperatures were either...
Drosophila Pglym Genes

Assays of PGLYM activity: The assays for PGLYM activity was modified from the methods of GRISOLIA and CARRERAS (1975). This assay measures formation of phosphoenolpyruvic acid (PEP) by increase in optical density at 240 nm by rate limiting amounts of mutase in an excess of enolase. One gram of adult flies was ground in 10 ml of 50 mM Tris, pH 7.4, 2% of saturation with respect to phenylthiourea, using a motorized mortar and pestle. The homogenate was spun at 10,000 rpm for 15 min at 4°. The supernatant was recovered by filtration through small gauge nylon mesh. Fifty microliters of homogenate were added to 3 pl of 50 mM 3-phosphoglyceric acid (PGA), triosodium salt (Sigma), 10 μmol of MgSO4, 100 μmol of Tris, pH 7.0, and 10 units of enolase in a 1-cm light path quartz cell. The rate of increase of absorbance at 240 nm was measured spectrophotometrically using a Gilford 240 recording spectrophotometer, and the amount of enzyme activity in each sample was determined utilizing a molar extinction coefficient of 1.75 × 10³ for PEP and the fact that 1.5 mmol of 2-PGA formed corresponds to 1 mmol of PEP measured. Protein concentrations were determined using the Bio-Rad microassay with bovine serum albumin as a standard, utilizing the reagent and instructions supplied by the manufacturer. Lethal mutations in the 87B region were obtained from the Umea stock center, Sweden, and from J. GAUZ at the Hungarian Academy of Sciences.

RESULTS

Isolation of the D. melanogaster PGLYM encoding genes: To isolate D. melanogaster Pglym, a full length cDNA which encodes the human brain PGLYM isofrom (TSUJIMO et al. 1989) was used to screen a Drosophila cDNA library prepared using adult mRNA and the vector, Agt11 (VON KÄM. et al. 1989). Five cDNA clones were isolated. All contained an identically sized insert of 1 kb. The clones were sequenced and found to contain a PGLYM encoding open reading frame by comparison to the human PGLYM amino acid sequence. A Drosophila cDNA clone was used to probe a Southern blot of Drosophila genomic DNA. This analysis (Figure 1) revealed multiple regions within the Drosophila genome having various intensities of hybridization. This suggests the likelihood that more than one gene may be present in the genome. A cDNA clone was used to screen a Drosophila genomic library and three genomic clones were isolated. Partial restriction maps of the inserts in these phage were assembled and identify two non-overlapping clones (Figure 2). By comparison to the genomic Southern blot, each clone could be characterized as containing restriction fragments relating to one or the other of the hybridizing regions. This confirmed existence of two Pglym regions within the Drosophila genome. The nucleotide sequence of each was determined according to the strategy indicated in Figure 2. The cytogenetic localization of each gene on the Drosophila polytene map was determined by in situ hybridization to D. melanogaster polytene chromosomes. Using the entire insert of each lambda clone as a probe for hybridization to polytene salivary gland chromosomes, a single site for each was identified (data not shown). One gene is located at 87B4,5 on the right arm of chromosome 3 and

at 25° or at 30° as described. Primers were extended using 40 units of avian myeloblastosis virus reverse transcriptase (Life Sciences, St. Petersburg, Florida). Extension products were separated on 6% acrylamide 3/4 × TBE buffer gradient gels (BIGGIN et al. 1985) and visualized by standard autoradiographic techniques.

Purification of synthetic oligonucleotide primers: Oligonucleotides designed for DNA sequencing were synthesized (Syracuse University DNA and protein core facility) and purified via passage over a NENSORB PREP cartridge (DuPont).

Sequencing strategies: To sequence genomic regions spanning each gene, subclones were isolated that contained regions of bacteriophage clones that were homologous to the cDNA clone used to isolate them. The sequencing strategy for these subclones involved sequencing from either end and using oligonucleotide primers to bridge these sequences. Isolated cDNA clones were subcloned either into pBluescript II SK+ or mp18/19 by subcloning liberated EcoRI fragments of Agt11 inserts or performing BamHI, HindIII directional subcloning. Genomic fragments were cloned into the appropriate polylinker sites of pBluescript II SK+. Single-stranded DNA was sequenced by the dideoxy chain termination method incorporating [α-33P]ATP (SANGER et al. 1977; HONG 1982) and visualized on 6% acrylamide 3/4 × TBE buffered gradient gels (BIRCH et al. 1983). Sequencing of double stranded plasmid templates utilized the following protocol. Two micrograms of plasmid DNA in 20 μl of H2O were denatured by the addition of 2 μl of NaOH, 2 mM EDTA (pH 8.0), and incubation was at 85° for 5 min. The mixture was neutralized by the addition of 3 μl of 3 M sodium acetate (pH 5.2) and 7 μl of H2O. Absolute ethanol (75 μl) was added to the DNA and placed at -70° for 5 min. DNA was recovered by centrifugation for 10 min at 4°. The DNA pellet was washed in 200 μl of 2 M NaOH, 2 mM EDTA (pH 8.0), and placed at 87°C, 5 min. All sequencing utilized the Sequenase version 2.0 DNA sequencing kit as per the manufacturer’s direction (U. S. Biochemical Corp.). Autoradiograms of sequencing gels were read and sequences determined using a digitizer and computer programs from DNASTAR (Madison, Wisconsin). Both Pglym sequences have been submitted to the GenBank database. Pglym78 has been assigned accession no. L27654. Pglym87 has been assigned accession no. L27656.

In situ hybridization to polytene chromosomes: Polytene chromosomes were dissected from late 3rd instar larvae of Oregon R D. melanogaster that had been raised at 18°C on banana media. Dissection of salivary glands, separation and fixation and pretreatment of polytene chromosomes was performed essentially as described in ENGELS et al. (1986). Chromosomes were examined under phase contrast optics and the best chromosomes selected for hybridization. From 250 ng to 1 μg of unrestricted bacteriophage DNA containing the genomic DNA of interest were biotinylated utilizing Bio-11-dUTP (ENZO diagnostics) in a random priming reaction (Boehringer Mannheim), and unincorporated nucleotides were separated by passage over a Sephadex G-50 spin column (Boehringer Mannheim). Twenty micrograms of salmon sperm DNA were added and precipitated with the addition of 2.5 volumes of ethanol. The probe was recovered by centrifugation, resuspended in 75 μl of hybridization buffer (0.6 M NaCl, 50 mM sodium phosphate buffer, pH 6.8, 1 × Denhardt’s solution, 5 mM MgCl2), denatured by boiling for 3 min, and chilled on ice. The probe solution was added to the slides, sealed with a coverslip, and incubated in a moist chamber at 58° for 12–18 hr. The slides were washed three times for 5 min each in 2 × SSC, and the signal was detected using the Detek Hrtp kit as specified by the manufacturer (ENZO Diagnostics).
the other at region 78 A/B on the left arm of chromosome 3. These genes will be called Pglym\textsubscript{87} and Pglym\textsubscript{78}, respectively.

**Sequencing of Pglym\textsubscript{78}:** The PGLYM encoding cDNA used to isolate the genomic clones was demonstrated, by sequencing, to correspond to the transcribed region of Pglym\textsubscript{78}. The sequence of 2290 nucleotides of the Pglym\textsubscript{78} genomic region beginning 632 nucleotides upstream of the translation start ATG and extending 381 nucleotides beyond the translation stop codon and the conceptual translation of the coding region is shown in Figure 3. Pglym\textsubscript{78} contains two introns of 469 and 58 nucleotides, respectively, and three exons of 102, 498 and 166 nucleotides. The promoter region of Pglym\textsubscript{78} does not contain a sequence which is an identifiable TATA box sequence. Exon one, as defined by the 5' extent of the cDNA, was confirmed by primer extension and RNase protection assays (see below). Nucleotide sequencing of the 3' end of the cDNA clone revealed that polyadenylation site occurs 20 bp downstream from the sequence, AATAAA, which matches the consensus polyadenylation signal.

**Transcript analysis of Pglym\textsubscript{78}:** To confirm the transcription unit suggested by the comparison of cDNA and genomic sequences, RNase protection and primer extension analyses were performed on the Pglym\textsubscript{78} transcript (Figure 4). The 5' end of the Pglym\textsubscript{78} transcript was defined using RNase protection and transcripts from the 1.3-kb BglII subclone. The transcripts pro-

![Figure 2](image-url)

**Figure 2.**—Sequencing strategy of the isolated Pglym regions from the genomic lambda clones. Stipled bars represent the BglII bands that hybridized in genomic Southern blot analysis (see Figure 1, lane 5). Enzymes used to construct the maps were, E, EcoRI; Bg, BglII; B, BamHI. Sequencing from restriction sites is denoted with a plain arrow while the oligonucleotide primers used to bridge these sequences are shown as +m. Sequencing of the cDNA and genomic sequence covered both strands of each gene.
Drosophila Pglym Genes

ATACAGCCGACCTGCAATTGAAACACGACGGCCGGCATTGTCGCTGCACTCGGCGCTTCTCCGCAACAAAGGAGGTTGTTCGCAATTACTGATAA-100
AAAGCACTGGAATACCCATCTTCGCTTCTCCGCTGCACTCGGCGCTTCTCCGCAACAAAGGAGGTTGTTCGCAATTACTGATAA-100
TTCTGCTTCTCCGCTGCACTCGGCGCTTCTCCGCAACAAAGGAGGTTGTTCGCAATTACTGATAA-100

This transcript protected fragments of 246,150 and 144 nucleotides. The 246-nucleotide fragment is the same 246-nucleotide fragment protected by the longer transcript (above) indicating that this fragment is from the region 3' to Pglym78. The origin of this fragment is likely from a downstream gene since in the Northern blots described below a Pglym78 transcript that is 246 nucleotides longer at the 3' end would have been resolved in addition to the 1-kb Pglym78 transcript. If this fragment corresponded to a duplicated, alternatively spliced Pglym78 exon it would have been discovered by inspection of the sequence. The bands of 150 and 144 nucleotides represent shortening of the 334-nucleotide protected fragment corresponding to exon 3 and the 3'untranslated region of the transcript as expected. The two bands could reflect alternative polyadenylation sites or some nuclease digestion of the protected molecules in A rich regions. It is unclear if the original 334-nucleotide fragment might be a doublet since a 6-nucleotide difference probably would not be

The 5' end of the Pglym78 gene was analyzed by RNase protection experiments using transcripts from the 3.0-kb BgiII subclone used in sequencing of the 3' end of the probe to the 3' end of Pglym78. The origin of this fragment is Pglym78 transcript that is 246 nucleotides longer at the 3' end would have been resolved in addition to the 1-kb Pglym78 transcript. If this fragment corresponded to a duplicated, alternatively spliced Pglym78 exon it would have been discovered by inspection of the sequence. The bands of 150 and 144 nucleotides represent shortening of the 334-nucleotide protected fragment corresponding to exon 3 and the 3'untranslated region of the transcript as expected. The two bands could reflect alternative polyadenylation sites or some nuclease digestion of the protected molecules in A rich regions. It is unclear if the original 334-nucleotide fragment might be a doublet since a 6-nucleotide difference probably would not be

from embryos and larvae. Three major extension products of 96, 106 and 114 nucleotides were detected. The 5' ends of these extension products correspond directly to the positions of transcript start sites predicted by the RNase protection studies.

The 5' end of the Pglym78 gene was analyzed by RNase protection experiments using transcripts from the 3.0-kb BgiII subclone used in sequencing of the 3' half of this gene. RNase protection using a transcript of the original 334-nucleotide fragment might be a doublet since a 6-nucleotide difference probably would not be

Figure 3.—Sequence and conceptual translation of Pglym78. Nucleotides are numbered with reference to the first nucleotide of the sequenced region and the introns are indicated via a dashed line which breaks the derived amino acid sequence. Splice donor and acceptor sequences are in underlined type. Multiple transcription start sites are denoted by asterisks. Untranslated regions of this gene. RNase protection using a transcript of the 3.0-kb BgiII subclone used in sequencing of the 3' end of the probe to the 3' end of Pglym78. The origin of this fragment is Pglym78 transcript that is 246 nucleotides longer at the 3' end would have been resolved in addition to the 1-kb Pglym78 transcript. If this fragment corresponded to a duplicated, alternatively spliced Pglym78 exon it would have been discovered by inspection of the sequence. The bands of 150 and 144 nucleotides represent shortening of the 334-nucleotide protected fragment corresponding to exon 3 and the 3'untranslated region of the transcript as expected. The two bands could reflect alternative polyadenylation sites or some nuclease digestion of the protected molecules in A rich regions. It is unclear if the original 334-nucleotide fragment might be a doublet since a 6-nucleotide difference probably would not be

...
resolved in the higher molecular weight 334-nucleotide band.

Northern blot analysis using RNA prepared from different developmental stages of the Drosophila life cycle detects a single 1.0-kb transcript which varies in relative amounts throughout development (Figure 5). Pglym78 transcripts are very abundant at the earliest embryonic stages, probably as a result of transcription during oogenesis. They become undetectable as embryogenesis proceeds and an increase in abundance is evident by hour 12 of development. During larval, pupal and adult stages, the amount of Pglym78 transcript rises, falls and rises again in a pattern identical to transcripts from other glycolytic genes of Drosophila (Shaw-Lee et al. 1991, 1992; Roselli-Rehfus et al. 1992).

Nucleotide sequencing of Pglym87: The sequencing strategy for Pglym87 is illustrated in Figure 2. The nucleotide sequence of 1838 nucleotides from Pglym87 and its conceptual translation is given in Figure 6. A putative open reading frame, but lacking the first two codons, which might encode a PGLYM subunit is present and intact through an appropriately positioned translation stop codon. A methionine codon which might serve as an initiating codon does exist, in frame, eight amino acid residues into this reading frame. Translation of an mRNA of this type would yield a PGLYM like molecule lacking the first nine amino acids but otherwise intact. Pglym87 lacks the introns present in Pglym78 which suggests that the Pglym gene duplication occurred by retrotransposition and raises the possibility (discussed below) that Pglym87 is a pseudogene.

Transcript analysis of Pglym87: RNase protection experiments were performed in parallel with and using the same RNA samples as those of Pglym78 (Figure 4).
Drosophila Pglym Genes 359

This experiment failed to detect any transcript from Pglym87 in the RNA from larval and adult stages. To determine if Pglym87 transcripts could be detected at any developmental stage, RNase protections (internally controlled with probes to Pglym78 and ribosomal protein RP49) were performed using RNA isolated from animals which span the complete developmental spectrum, i.e., the same RNA samples as used for the northern blots. These experiments (data not shown) also failed to find any developmental stage at which a Pglym87 transcript could be detected. Attempts to identify a Pglym87 transcript using primer extension analyses were also unsuccessful.

Analysis of the 87B region: Prior to completion of the above analysis which failed to find a Pglym87 transcript, we theorized that a mutation in a PGLYM encoding gene that reduced the total level of PGLYM enzyme significantly might be lethal. GAUSZ et al. (1981) analyzed the 87B cytogenetic region and identified 15 complementation groups represented in a set of 88 alleles. This is an average of 5.9 alleles per complementation group. This suggests that the region may be close to mutationally saturated and raises the possibility that one of the lethal complementation groups might represent Pglym87. Representatives of each complementation group were assayed, as heterozygotes, for significant reduction of PGLYM activity. Results from this analysis are presented in Figure 7. Reduction of PGLYM activity could not be associated with any lethal complementation group. Therefore if Pglym87 has a function, it is likely not a vital function so that mutations in Pglym87 were not detected in this set of complementation groups. In addition, the amount of enzyme encoded by Pglym87 is likely a small fraction of the total phosphoglycerate mutase activity since heterozygous deletions also showed no reduction in activity.
Evolutionary comparison of the Drosophila Pglym genes: Alignment of the putative proteins encoded by the Drosophila genes and the human muscle and brain isoforms, PGAM-M and PGAM-B, respectively (Figure 8) reveals that the potential primary amino acid sequence is highly conserved. The putative Drosophila PGLYM proteins are 73% identical. The first eight amino acids of PGLYM87 are shown in small letters because no putative initiator methionine precedes them, yet they are conserved. Comparison of these amino acid sequences was used to determine whether the Drosophila Pglym duplication represented a duplication having functional similarity to the mammalian Pglym duplication which resulted in genes for different isoforms. PGLYM78 seems to be equally similar to the PGAM-B isoform, 68% and the PGAM-M isoform, 69%. At positions of amino acid substitutions where there can either be a PGA'M or PGA"B type amino acid within the PGLYM78 protein, 12 amino acids are identical to the PGAM-B gene and a further 12 substitutions are PGAM-M type. Putative PGLYM87 has 64% amino acid identity with the human muscle isofrm PGAM-M protein and 68% amino acid identity with general isoform PGAM-B. Analysis of the substitutions of PGLYM87 reveals that eight are of the PGAM-B type and 18 are of the PGAM-M type. As the PGAM-B isoform is the mammalian general isoform and the PGAM-M form is considered tissue restricted isoform this difference may indicate that Pglym87 at some point encoded a protein that was restricted to a specific cell type, e.g., muscle.

Analysis of Drosophila Pglym sequences reveals that each shows a nonrandom distribution of nucleotides in the third codon position (Table 1) with an abundance of C and G and scarcity of A typical of other Drosophila genes (STARMER and SULLIVAN 1989). This nonrandom distribution is reflective of codon utilization bias. Pglym87 has a somewhat less extreme codon bias than does Pglym78 indicating some loss of bias since the duplication.

**TABLE 1**

<table>
<thead>
<tr>
<th>Gene</th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pglym87</td>
<td>29</td>
<td>103</td>
<td>27</td>
<td>95</td>
<td>254</td>
</tr>
<tr>
<td>% of total</td>
<td>11.4</td>
<td>40.6</td>
<td>10.6</td>
<td>37.9</td>
<td></td>
</tr>
<tr>
<td>Pglym78</td>
<td>22</td>
<td>123</td>
<td>14</td>
<td>97</td>
<td>256</td>
</tr>
<tr>
<td>% of total</td>
<td>8.6</td>
<td>48.0</td>
<td>5.5</td>
<td>37.9</td>
<td></td>
</tr>
</tbody>
</table>

**DISCUSSION**

**Molecular genetics of Pglym78:** Pglym78 has properties similar to other Drosophila genes which encode enzymes of the glycolytic pathway. These include a promoter region which has no TATA box but has multiple, closely spaced transcription initiation sites and a transcript expression pattern similar to all of the other Drosophila glycolytic genes tested so far (SHAW-LEE et al. 1991, 1992; ROSELLI-REHFUS et al. 1992). This similarity in temporal regulation of glycolytic genes reinforces the notion that there may be common regulatory control mechanisms exercised on the transcription of these genes.

The Drosophila Pglym78 gene and the human muscle Pglym gene (TSUJIMO et al. 1989) are the only Pglym genes for which intron positions have been determined. Each of these genes has two introns. The first intron of

**Alignment of Drosophila and Human PGLYM s**

<table>
<thead>
<tr>
<th>Gene</th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pglym87</td>
<td>29</td>
<td>103</td>
<td>27</td>
<td>95</td>
<td>254</td>
</tr>
<tr>
<td>% of total</td>
<td>11.4</td>
<td>40.6</td>
<td>10.6</td>
<td>37.9</td>
<td></td>
</tr>
<tr>
<td>Pglym78</td>
<td>22</td>
<td>123</td>
<td>14</td>
<td>97</td>
<td>256</td>
</tr>
<tr>
<td>% of total</td>
<td>8.6</td>
<td>48.0</td>
<td>5.5</td>
<td>37.9</td>
<td></td>
</tr>
</tbody>
</table>

**FIGURE 8.—Alignment of the conceptually translated protein sequence of the Pglym genes of Drosophila and humans. Dashes indicate gaps introduced to maintain alignment.**
Pglym78 interrupts the gene at codon 35 and does not have a counterpart in the human gene. The second intron of Pglym78 interrupts codon 201. The second intron of the human muscle gene interrupts codon 197 of the PGLYM open reading frame. This may represent an intron position conserved in the two genes. It has been suggested that intron positions may ‘slide’ around a conserved position and observations consistent with this interpretation have been made when comparisons between homologous genes that have been separated for a substantial evolutionary period have been made (McKnight et al. 1986; Marchionni and Gilbert 1986).

**Origin of the duplication:** Since Pglym87 lacks the introns present in Pglym78 and appears at least two codons short at the N terminus it seems highly likely that Pglym87 arose through retrotransposition from a Pglym78 transcript. Additional support of this interpretation comes from several features of the Pglym87 region. The intron-less Pglym87 gene is on a separate chromosome arm from the intron containing Pglym78 gene. A stretch of 20 nucleotides, 19 nucleotides 3′ to a putative polyadenylation signal (underlined in Figure 6), contains 53% adenosine residues. This could represent the remnant of a poly(A) tail from the transcript. A similar percentage of adenosine residues 3′ to the consensus polyadenylation signal in human Pgk, another retrotransposed gene, has been used to argue that this region represents the corrected remainder of the reverse transcribed transcript’s poly(A) tail (McCarrey and Thomas 1987).

While evidence presented here indicates that Pglym87 is almost certainly the result of a retrotransposition mediated gene duplication event, examination of the nucleotide sequence around Pglym78 suggests that additional or more complex mechanisms may have played some role in the origin of Pglym87. Sequences immediately 3′ to the putative poly A remnant contain four perfect copies and two corrupted copies of the sequence GGTCTT in adjacent direct repeats. These could encode a poly proline-glutamic acid peptide on the opposite strand. Immediately downstream from this repeat are five perfect copies and one corrupted copy of the repeat GGGTCT which, on the opposite strand would encode a poly proline-arginine peptide. These repeats are similar to the S repeats found within the transposable element Hobo, which is composed of the sequence TGAGGTCTT (Streich et al. 1986). Also, in one open reading frame of the Ac transposable element of maize there is a region of repeated sequence which encodes 10 tandem proline-glutamic acid repeats reminiscent of the hobo S repeats (Muller-Neumann et al. 1984). The presence of these repeat sequences, immediately three prime to the Pglym87 gene that are similar to those found in transposons of the hobo/Activator class, suggests an interesting mechanism for this gene duplication. This class of transposable elements transpose through a DNA intermediate (Calvi et al. 1991). If reverse transcript DNA were to be incorporated into an actively transposing hobo/Ac type element, this could provide a mechanism for integration into different regions of the genome using the usual pathway of the parent transposon.

While retrotranspositions of genes and pseudogenes is found in the genomes of most species and seems particularly frequent in mammalian genomes, only a few examples have been noted in the Drosophila genome. This is despite demonstration of the presence in Drosophila cells of retroviral-like reverse transcriptase enzyme (Heine et al. 1980) and an abundance of transposable elements that are of the retrotransposon class. On the basis of different chromosomal location of the Rh4 opsin gene in Drosophila virilis and D. melanogaster and the absence of introns in the D. virilis gene (Neufeld et al. 1991) have suggested that the Rh4 opsin gene of D. virilis has undergone retrotransposition during its evolution in the D. virilis lineage. In addition, it is likely that retrotransposition has played a significant role during the evolution of the Gapdh gene duplication in the genus Drosophila (Wojtas et al. 1992).

Another example is that of the retrotransposed copies of the alcohol dehydrogenase (Adh) genes of Drosophila yakuba and Drosophila teissieri (Jepps and Ashburner 1991; Jepps et al. 1994). In many respects these retrotransposed Adh genes are similar to Pglym87. Their Adh open reading frames remain intact except for a deletion of the Adh start codon. Each has lost the introns typical of an Adh gene and they occupy chromosomal positions which are not homologous to the Adh position in D. melanogaster. These putative pseudogenes are more similar to one another than to their respective functional Adh genes, even containing similar deletions in 5′ non-coding DNA. The chromosome positions of the two genes in D. teissieri and D. yakuba is conserved which indicates that the two genes have a common origin and that the duplication arose before the species diverged.

**Is Pglym87 a pseudogene?:** Since Pglym87 has its origin through retrotransposition the question that naturally arises is: Is it a pseudogene? Consistent with the view that it is a pseudogene is its origin through retrotransposition and the lack of detectable transcripts. However, we recognize the difficulty in proving that there are no Pglym87 transcripts. These could exist in very low levels, at restricted periods of development or in a very few cells and have gone undetected in our assays. Also consistent with the hypothesis that Pglym87 is a pseudogene is a consideration of its 5′ end. If the methionine encoded by the codon homologous to codon nine of Pglym78 is in fact a translation start in Pglym87, then this would result in the loss of eight relatively conserved amino acids from the N terminus. Furthermore, it is not clear why the sequence encoding the six amino acids preceding the methionine would be conserved as compared to
Pglym78. Another reason to think that Pglym87 might be a pseudogene is that retrotransposed genes have the difficulty of not carrying sufficient information for their correct transcription. In fact, two of the better analyzed retrotransposed genes, human testes specific Pghk (McCarter et al. 1994) and D. virilis Rh4 opsins (Neufeld et al. 1991) both of which are known to function, seem to be reverse transcriptase copies of a transcript substantially longer at the 5′ end than the mature mRNA. In each case copies of sequences required for transcription apparently were included during retrotransposition. Inspection of the sequence 5′ to the PGLYM reading frame of Pglym87 reveals no significant similarity to the sequence 5′ to Pglym78.

In support of Pglym87 being functional are several aspects of its molecular evolution including observations that the PGLYM reading frame is intact and that the gene shows at least some retention of codon utilization bias. In addition, the crystal structure of PGLYM from yeast has been solved and specific amino acids have been implicated in catalysis. These are histidine 8, histidine 181 and serine 11. Histidine 181 has been conserved in all mutases. Serine 11 has been implicated in the catalytic mechanism as a phospho ligand (White and Fothergill-Gilmore 1992). All these amino acids are conserved in the PGLYM78 protein with histidine 8 and 181 corresponding to histidine 12 and 188 respectively and serine 11 is conserved as serine 15. In a hypothetical PGLYM87 protein, the two histidine residues and the serine are also conserved and are located in identical positions as in PGLYM78. Thus, both Drosophila genes possess the capability to encode those amino acid residues shown to be important for PGLYM catalytic activity. If the PGLYM 87 is nonfunctional, then it should not, by the conventional view of pseudogenes, be subject to constraint on its sequence divergence. Conservation of coding frame, codon bias or catalytically relevant specific amino acids would not be expected in a pseudogene. This is clearly the case for mammalian pseudogenes, where the biased base composition seen in the third codon position of functional genes is almost absent and the loss of reading frame is commonly observed (Kimura 1983).

Therefore we are left to chose between two hypotheses regarding the nature of Pglym87, neither of which is in total accord with present views regarding the molecular evolution of products of retrotransposition and pseudogenes. The structural properties suggest Pglym87 is a pseudogene with a peculiar rate and pattern of evolution, while aspects of its molecular evolution suggest it might be a functional gene having puzzling aspects in its structure. In this regard it seems worth noting that in Drosophila species, two other candidates for pseudogenes have been characterized and in each, peculiar evolutionary properties have been noted. The similarity between Pglym87 and the retrotransposed Adh sequences in D. teissieri and D. yakuba sequences are particularly striking. When these sequences were identified, their peculiar pattern of evolution, including retention of reading frame, codon bias and higher rates of substitution at synonymous nucleotide positions was noted (Jeffs and Ashburner 1991). Recently an alternative interpretation for these sequences has been offered in which it is suggested that the retrotransposed Adh sequences captured upstream exons with the resultant evolution of a new gene (Long and Langley 1993). However, one of the principle reasons offered to support the hypothesis that the new gene, called jingwei, is functional is its pattern of molecular evolution. The ultimate answer with respect to jingwei function remains to be verified by molecular analysis of its putative transcription and translation products, but it seems highly improbable that an analogous situation applies to Pglym87. Inspection of the sequence makes no suggestion of captured exons in the 5′ sequence. In another example of a pseudogene in Drosophila, we have recently analyzed the molecular evolution of an Adh sequence that is present in species of the repleta group (Sullivan et al. 1994). This sequence is found in all species of the repleta group and is located immediately upstream of the functional Adh genes. It is not a product of retrotransposition since it retains introns homologous to the functional Adh genes. Analysis of the evolution of this sequence reveals it is diverging more slowly than the neighboring intergenic regions, some codon bias is retained, and the ratio of Ks to Ka ranges from 10 to 14, in pairwise interspecific comparisons using the sequences from seven species. These ratios are only slightly lower than those obtained from equivalent comparisons of the functional Adh genes and clearly not unity, as would be expected for a pseudogene. Yet, in this set of Adh homologous sequences there is no open reading frame of appreciable length common to this set of seven species. Since there are only the two Adh pseudogenes and this report of Pglym, it is unclear at present what the peculiar patterns of molecular evolution of these putative pseudogenes might mean. It seems likely that more extensive analysis of these examples and the discovery of others may reveal features of genome evolution in Drosophila not yet appreciated or understood.

LITERATURE CITED


Castella-Esgola, J.-D., M. O. Czjusz, P. LeBoUcH, V. Joulin, Y. BLOQUET et al., 1990 Isolation and characterization of the gene encoding


Gebbilia, S., and J. Carreras, 1975 Phosphoglycerate mutase from yeast, chicken breast muscle, and kidney (2,3-PGA-dependent).


Communicating editor: V. G. Finnerty