Mol Gen Genet (1992) 235:213–220

MGG

Structure and expression of the phosphoglycerate kinase (*Pgk*) gene of *Drosophila melanogaster*

Linda Roselli-Rehfuss*, Feng Ye, James L. Lissemore and David T. Sullivan

Department of Biology, Syracuse University, 130 College Place, Syracuse, New York 13244, USA

Received February 21, 1992 / Accepted May 19, 1992

Summary. The gene that encodes phosphoglycerate kinase in *Drosophila melanogaster* (*Pgk*) has been isolated and characterized. There is a single copy of *Pgk* in the *Drosophila* genome located at cytogenetic position 23A1-2. Transcripts of *Pgk* are 1.6 kb long and are found during development with a profile similar to the expression pattern of other genes of the glycolytic pathway. There are substantial amounts of maternal transcript in early embryos which decline in abundance until mid-embryogenesis when transcript levels increase; levels remain high, during larval stages, fall during pupariation and rise again at emergence. The nucleotide sequence of the *Pgk* gene reveals two small introns, one of which is at a position identical to the site of an intron found in *Pgk* genes from other organisms. The *Pgk* gene has no TATA box in the region of transcription initiation and has multiple transcription initiation sites that are closely spaced within 110 nucleotides of the translation start site.

Key words: Phosphoglycerate kinase – *Drosophila* – Nucleotide sequence – Glycolysis

Introduction

Phosphoglycerate kinase (PGK; ATP: 3-phosphoglycerate 1-phosphotransferase, EC 2.7.2.3), an enzyme of the glycolytic pathway, catalyzes the reversible conversion of 1,3-diphosphoglycerate to 3-phosphoglycerate, generating one molecule of ATP. PGK from yeast and from several mammals has been extensively characterized (Banks et al. 1979; Huang et al. 1980). X-ray crystallography of horse muscle PGK has demonstrated that this monomeric enzyme is composed of two lobes connected by a waist region. These two protein domains, the C-terminal and N-terminal domains of PGK, may comprise the ATP binding and phosphoglycerate binding regions, respectively (Blake et al. 1972; Bryant et al. 1974). It has been proposed that substrate binding initiates a hinge-bending motion of the two domains which results in the closure of an active site cleft (Banks et al. 1979). Although comparison of amino acid sequences of the yeast and human enzymes show only 65% conservation (Banks et al. 1979; Huang et al. 1980), a domain interchange between human and yeast PGK yields a hybrid enzyme with catalytic properties similar to either parent enzyme (Mas et al. 1986).

Specific human PGK variants have been described, some of which lead to enzyme deficiency, nonspherocytic hemolytic anemia and neurological disturbance (Fujii and Yoshida 1980; Fujii et al. 1980, 1981; Yoshida et al. 1972). In addition to a generally expressed protein which is X-linked, a testis-specific isozyme is found in a number of mammalian species (Pegoraro et al. 1978; Pegoraro and Lee 1978). In the mouse, this isozyme has been mapped to the major histocompatibility complex on chromosome 17 (Eicher et al. 1978), whereas its chromosomal location in humans is linked to the MHC locus on chromosome 6 (Szabo et al. 1984).

Pgk genes have been isolated from various organisms (Ciccarese et al. 1989; Clements and Roberts 1986; Nellmann et al. 1989; Solingen et al. 1988; Vanhanen et al. 1989; Watson et al. 1982). Studies of human and mouse genomes indicate that the mammalian *Pgk* gene family includes two functional loci. The X-linked *Pgk-1* is intron-containing and expressed in all somatic cells and premeiotic cells (McCarrey and Thomas 1987; Michelson et al. 1985a). The intronless mouse *Pgk-2* arose by reverse transcriptase-mediated processing of a transcript from the *Pgk-1* gene (Boer et al. 1987; McCarrey 1987; McCarrey and Thomas 1987) and is expressed only in the meiotic/postmeiotic spermatogenic cells (Adra et al. 1987; Boer et al. 1987) and is linked to the major histocompatibility locus on chromosome 6. In the human genome an X-linked intronless pseudogene (yhPgk-1)

**Present address*: Vollum Institute for Advanced Research, Oregon Health Sciences University, 3181 SW. Sam Jackson Park Rd., Portland, Oregon 97201, USA

Correspondence to: D.T. Sullivan

has also been identified as a retroposon of the functional *Pgk-1* locus (Michelson et al. 1985b). Human genomic *Pgk* clones have been independently isolated (Singer-Sam et al. 1984; Michelson et al. 1985a). The gene contains 10 introns and 11 exons and spans a 23 kb region on the X chromosome. A *Pgk-1* genomic clone from mouse has also been isolated and partially sequenced (Adra et al. 1987). The mouse *Pgk-1* gene is spread over a 16 kb region of the X chromosome and may also contain 10 introns and 11 exons. The human *Pgk* gene has three transcription start points located at 77, 78, and 86 nucleotides upstream of the translation start codon (Singer-Sam et al. 1984). The mouse *Pgk* gene has a single transcription start site located 88 nucleotides 5′ of the ATG codon (Adra et al. 1987).

In *Drosophila*, Chew and Cooper (1973) identified three PGK electrophoretic variants in natural populations. They attributed this variation to three co-dominant alleles of a single *Pgk* gene. Preliminary data have positioned *Pgk* on the left arm of chromosome II, between the genes *al* and *cn*. More precise mapping confirmed the location of *Pgk* on 2L at genetic map position 5.9 (Voelker et al. 1979). Terminal hyperploids have been used to demonstrate the cytogenetic location of *Pgk* to be between *ho* at 22E and *ed* at 24D2-24F1 (Voelker et al. 1979). In this report we describe the isolation, genetic and molecular characterization and expression pattern of the *Pgk* gene of *Drosophila melanogaster*. These studies contribute to our ongoing investigations on the control of expression of the genes that encode the enzymes of glycolysis in *Drosophila*.

Materials and methods

Isolation of Drosophila Pgk *clones. Drosophila* cDNA *Pgk* clones were isolated from a λ-gt 11 cDNA library constructed by von Kalm et al. (1989) using RNA prepared from adults of strain Oregon-R. The library was screened using a 600 bp *Hin*dIII fragment isolated from the coding region of a human *Pgk* cDNA clone. The human clone was a generous gift of A. Yoshida, City of Hope Research Institute, Duarte, Calif. Hybridization was in 4 × SSC at 65° C. Two cDNA clones were obtained and confirmed to be *Pgk* by nucleotide sequencing but neither contained the complete *Pgk* coding sequence. The largest of the cDNA clones was subsequently used to obtain genomic clones by screening a λ EMBL-4 genomic library prepared using strain Oregon-R genomic DNA from adults. Several clones were obtained and one of these, λ-PGK-C, was selected for subsequent studies.

DNA sequencing. A 2.7 kb and a 4.0 kb *Eco*RI fragment and a 2.2 kb *Bam*HI fragment from λ-PGK-C (Fig. 4) were isolated and ligated into *Eco*RI- or *Bam*HI-cut pBluescript II SK$^+$ (Stratagene). In addition, subclones of the 0.6 kb and 2.8 kb *Eco*RI-*Bam*HI and 0.7 kb *Bam*HI fragments were constructed in pBluescript II SK$^+$. Single-stranded phagemid DNAs were rescued by helper phage superinfection according to a rescue protocol by Vieira and Messing (1982). Single-stranded DNA sequencing was performed by the chain-termination sequencing protocol (Sanger et al. 1977) with the US biochemicals Sequenase sequencing kit.

Blotting procedures. DNA was prepared from adult *Drosophila* by the method of Bender et al. (1975), digested with restriction enzymes, electrophoresed in 0.8% agarose gels and transferred to Nytran membranes (Schleicher and Schuell) according to the method of Southern (1975). RNA was isolated according to the method of Skuse and Sullivan (1985). Total RNA (2.5 μg) from each stage was analyzed on Northern blots (Lissemore et al. 1987). Blots were probed with a 1 kb fragment of a *Pgk* cDNA clone made radioactive by random priming using a kit purchased from Boehringer-Mannheim. Developmental stages were timed from the midpoint of a 2 h embryo collection period.

In situ hybridization. The cytogenetic localization of *Drosophila Pgk* was determined by in situ hybridization of the recombinant λ-PGK-C DNA to salivary gland chromosomes using Biotin-16-dUTP and a detection kit (ENZO).

Transcript mapping. RNase protection was done using a Ribobasica kit (IBI) following the procedure of Zinn et al. (1983), with some modifications. Total RNA (5 μg) and 2×10^5 cpm of labeled RNA probe were mixed and hybridized at 55° C overnight. Heat-treated RNase A (12 μg) and 1500 units of BRL RNase T1 was added to the overnight hybridization reaction and incubated at 30° C for 1 h. RNA was phenol extracted and precipitated with 5 μg of yeast tRNA as carrier. The RNA pellet was rinsed with 70% ethanol, dried and resuspended in 4 μl of loading buffer. The samples were run in a 6% acrylamide/8 M urea sequencing gel at 100 W for 1.5–2 h (Calzone et al. 1987). For primer extension, 200 ng of synthetic oligonucleotide primer was end-labelled using 300 μCi of γ-[^{32}P]dATP (7000 Ci/mmol), 10 μl of 50 mM $MgCl_2$, 5 μl of 200 mM mercaptoethanol, and 10 units of T4 polynucleotide kinase in a total volume of 50 μl which was incubated at 37° C for 1 h. The primer was purified on a NENSORB-20 column (NEN), dried, and resuspended in 100 μl of TE (10 mM TRIS-HCl, 1 mM EDTA pH 8.0). Poly(A)$^+$ RNA (5 μg) plus 1×10^6 cpm of labeled primer in 10 μl of 6 × SSC were heated to 70° C for 5 min, and then incubated at 45° C overnight. After hybridization, the samples were ethanol precipitated at −20° C for 1 h. The pellets were resuspended in a final volume of 50 μl including 5 μl of 0.5 M TRIS pH 8.5, 0.1 M $MgCl_2$, 0.4 M KCl, 10 mM DTT, and 1 mM each dNTP, 40 units RNasin, and 30 units of AMV reverse transcriptase. This mixture was incubated at 37° C for 1 h. After incubation, 5 μl of 4 M ammonium acetate, 50 mM EDTA and 4 μl of 10% SDS were added. The samples were extracted with phenol and chloroform and precipitated by ethanol at −20° C for 1 h. The pellets were resuspended in 4 μl of loading dye, heated to 90° C for 5 min, and electrophoresed in a 6% acrylamide/8 M urea sequencing gel (Calzone et al. 1987).

Results and discussion

Identification of a single Pgk *coding region in the* Drosophila *genome*

To determine the copy number of *Pgk* genes, Southern blots of genomic DNA were hybridized with a radiolabelled cDNA clone (Fig. 1). Digestion of genomic DNA with each of nine restriction enzymes results in a single hybridizing band (Fig. 1, lanes B–J), which is consistent with the presence of a single gene. Digestion of genomic DNA with *Bam*HI (Fig. 1, lane A) yields two hybridizing bands. Given the sequence (see below) the two bands found in the *Bam*HI digest are also accounted for by a single gene. A single *Pgk* gene is also indicated by in situ hybridization to salivary gland chromosomes (Fig. 2). The site of hybridization is at cytogenetic position 23A1-2, which is consistent with earlier observations that positioned the *Pgk* locus gene between 22E and 24D2-24F1 (Voelker et al. 1979).

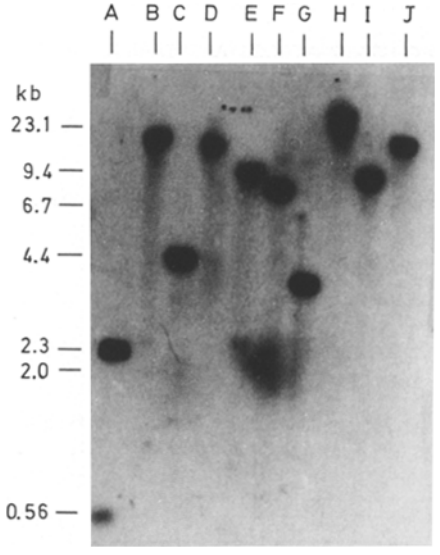

Fig. 1. Southern blot of genomic DNA probed with a 1 kb cDNA fragment. Enzymes used are; A, *Bam*HI; B, *Bgl*II; C, *Eco*RI; D, *Hin*dIII; E, *Pst*I; F, *Sal*I; G, *Sma*I; H, *Sph*I; I, *Sst*II; J, *Xba*I. Size standards are fragments from a *Hin*dIII digest of λ DNA

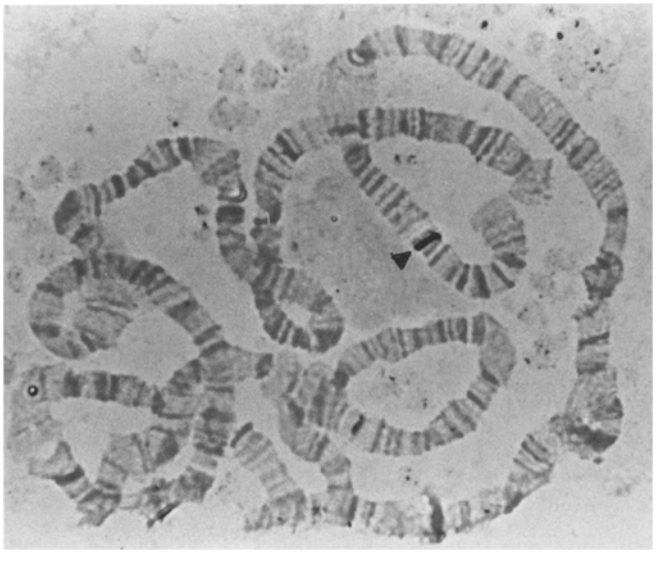

Fig. 2. In situ hybridization of biotinylated λ Pgk-C to salivary gland chromosomes. The *arrow* indicates hybridization to the 23A1-2 region

Developmental expression pattern of Pgk

Expression of *Pgk* during development was studied by Northern blot analysis (Fig. 3). A single zone of hybridization of approximately 1.6 kb is found in RNA samples prepared from all stages of development. In very early embryos, *Pgk* mRNA is abundant but then declines in quantity through the first 6 h. In mid-embryogenesis, at about 12 h, *Pgk* mRNA begins to increase in abundance. The most likely interpretation of this pattern of *Pgk* mRNA abundance is that a substantial amount of mRNA is produced during oogenesis, stored in the oocyte and used to support translation during early development. Transcription of *Pgk*, which is likely to be necessary to support later development, begins at a point prior to 12 h of embryogenesis. *Pgk* mRNA content during the larval and adult stages follows the pattern typical of transcripts for enzymes related to general metabolic pathways. A similar developmental pattern has been observed for the transcripts of other glycolytic genes including *Tpi* (Shaw-Lee et al. 1991), *Aldolase* (Shaw-Lee et al. 1992), and *Gpdh* (Wojtas et al., submitted).

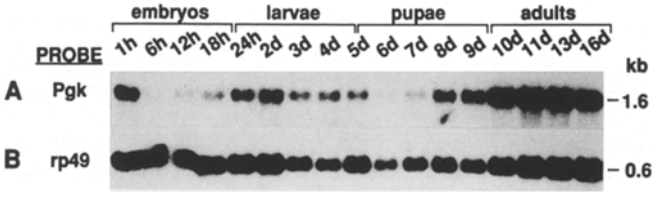

Fig. 3. Northern blot probed with a *Pgk* cDNA clone. Each lane contained 2.5 µg of total RNA from each developmental stage. The blot was reprobed with rp49 as a loading control

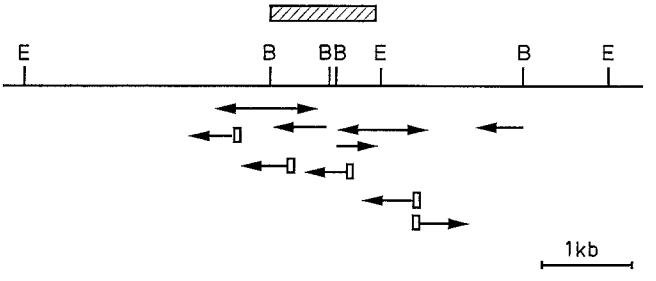

Fig. 4. Partial restriction map of the *Pgk* genomic clone and sequencing strategy diagram. The *boxed-hatched* area is the region where the *Pgk* coding sequence was thought to be located. The direction and extent of the sequence obtained from different subclones are indicated by *horizontal arrows*. *Open boxes* indicate oligonucleotides used as primers. E, *Eco*RI; B, *Bam*HI

Fig. 5. Nucleotide sequence and deduced amino acid sequence of the *Pgk* gene from *Drosophila melanogaster* and its flanking regions. Nucleotides are numbered with reference to the first nucleotide of the sequenced region and the introns are indicated by *asterisks* in the breaks in the derived amino acid sequence. The nucleotides at the splice junctions are *underlined*. EMBL Acc. No. Z14029

Nucleotide sequence of the Pgk *region*

A partial restriction map of λ-PGK-C and the sequencing strategy for the *Pgk* gene is shown in Fig. 4. The 4.0 kb and 2.7 kb *Eco*RI fragments and the 2.2 kb *Bam*HI fragment were isolated and subcloned into pBluescript II SK$^+$ plasmid. Both orientations of the 4.0 kb *Eco*RI and 2.7 kb *Eco*RI fragments and one orientation of the 2.2 kb *Bam*HI fragment were used for sequencing. Subclones of the 0.6 kb and 2.8 kb *Eco*RI-*Bam*HI fragments and the 0.7 kb *Bam*HI fragment were also used for sequencing. Single-stranded DNAs were purified from each of the clones and sequenced by the chain-termination method. Also, five oligonucleotide primers were synthesized and used to sequence from the middle of the inserts as indicated in Fig. 4.

The complete nucleotide sequence of the *Pgk* gene of *D. melanogaster*, along with the deduced primary amino acid sequence of the corresponding protein, is shown in Fig. 5. The sequence includes a coding region of 1248 bp, 925 bp of 5′ noncoding region, 292 bp of 3′ noncoding region and predicts two introns of 64 bp and 58 bp, respectively. The two introns are predicted by comparisons of the *Drosophila* nucleotide sequence with the amino acid sequence of human PGK. The putative introns contain the universal GT and AG junction sequences (Mount 1982). The human and mouse *Pgk-1* genes contain ten introns at conserved positions (Adra et al. 1987; Michelson et al. 1985a). Two introns have been found in *Pgk* genes of three filamentous fungi, *Trichoderma reesei*, *Aspergillus nidulans*, and *Penicillium chrysogenum* (Clements and Roberts 1986; Solingen et al. 1988; Vanhanen et al. 1989). The position of the N-terminal intron of the *Drosophila Pgk* gene is precisely conserved as compared to the human genes and is 1 bp removed from the position of an intron of the three fungal genes. The position of the second intron in the *Drosophila* gene is not conserved as compared to mammalian or fungal *Pgk* genes.

Drosophila Pgk has no discernible TATA box, similar to the human and mouse *Pgk* promoters. At position 636–642 and 666–672 (284 bp and 254 bp upstream of the translation start ATG codon), there are two direct repeats of the sequence CCAACTT. From position 2409 (113 bp downstream of the TAA stop codon), there is a single putative polyadenylation signal, AATAAA. The predicted PGK protein comprises 415 amino acids with a calculated molecular mass of 43952 Daltons and an isoelectric point of 7.71.

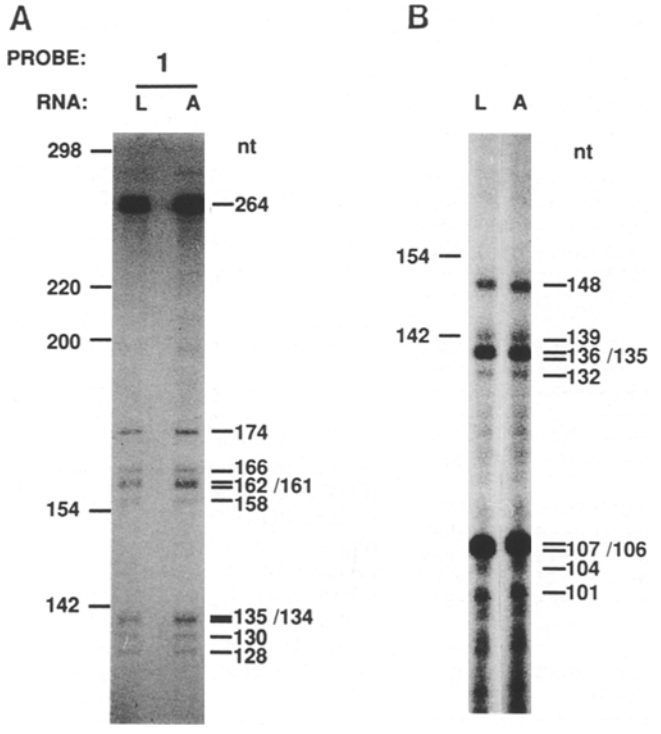

Fig. 6. A RNase protection mapping of the 5′ end of the *Pgk* mRNA from 2-day-old larvae (L) and 11-day-old adults (A). Probe 1 is a T3 RNA polymerase transcript of a subclone in pBluescript II SK$^+$ of the 635 bp *Bam*HI fragment which spans the 5′ end of *Pgk* (positions 674–1309, Fig. 5. **B** Primer extension mapping of the *Pgk* Poly(A)$^+$ RNA from 3rd instar larvae (L) and adults (A). Numbers in the left hand margin are DNA size markers. Numbers in the right hand margin are the calculated sizes of the bands

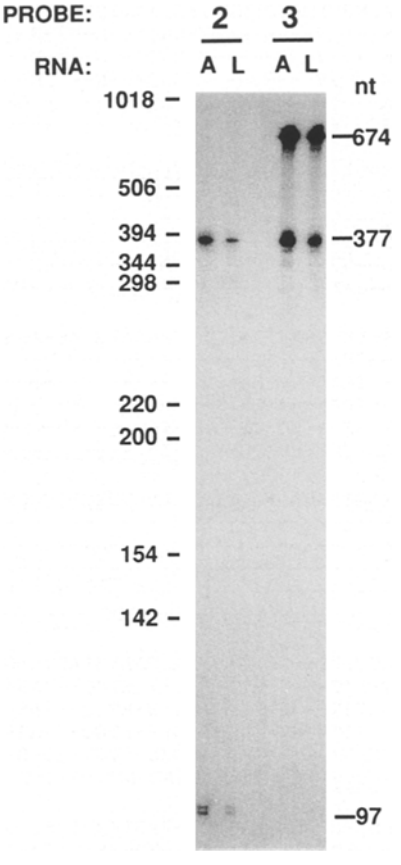

Fig. 7. RNase protection mapping of the 3′ end of the *Pgk* mRNA from 2-day-old larvae (L) and 11-day-old adults (A). Numbers in the left hand margin are DNA size markers. Numbers in the right hand margin are the calculated sizes of the protected fragments. Probe 2 is a T7 RNA polymerase transcript of a subclone, in pBluescript II SK$^+$, of a 2.2 kb fragment which extends downstream from the *Bam*HI site at 1322 (Fig. 5). Probe 3 is a T3 RNA polymerase transcript of a subclone of a 2.7 kb *Eco*RI fragment in pBluescript II SK$^+$ which extends downstream from position 1900 (Fig. 5)

Characterization of the Pgk *transcript*

The 5′ end of the *Pgk* transcript was mapped using a combination of RNase protection and primer extension (Fig. 6). A plasmid containing the 0.7 kb *Bam*HI genomic fragment from the 5′ end was linearized and transcribed using T3 RNA polymerase (probe 1, Fig. 6). *D. melanogaster* total RNA from 2-day-old larvae and 11-day-old adults was used to protect this probe. We expected a 264 nucleotide (nt) protected fragment, representing the distance from the 3′ *Bam*HI site of the 0.7 kb *Bam*HI fragment to the 3′ splice acceptor site of intron 1. Any other protected fragment(s) should represent the distance from the 5′ splice donor site of intron 1 to the transcription start site(s). As shown in Fig. 6A, RNase protection of probe 1 by both larval and adult RNA yielded the expected 264 nt band but also resulted in nine additional bands of the sizes indicated. To elucidate further the positions of transcription initiation, primer extension was performed. A 20 nt oligonucleotide primer which is complementary to the sequence 14–34 nt downstream of the ATG codon was extended and, as shown in Fig. 6B, primer extension of both larval and adult RNAs resulted in nine major bands. The size intervals between members of this set of bands are similar to those produced from the RNase protection. If the two techniques are detecting the same set of transcript ends, the size difference of corresponding bands obtained in RNase protection and primer extension should be 28 nt, i.e. the distance from the 3′ end of the primer to the 5′ end of the donor site of intron 1 (Fig. 6B). The measured size differences between the corresponding bands from RNase protection and primer extension are about 26–28 nt, thus supporting the hypothesis that the multiple bands represent a set of transcripts which have their 5′ ends at different locations within the region 113–67 nt upstream of the ATG start codon.

The same set of multiple 5′ transcript ends detected by primer extension and RNase protection suggests strongly that the *Pgk* gene has multiple transcription start sites. While it is conceivable that the multiple 5′ ends result from exonuclease activity, we have deter-

```
Droso M*AFNKLSIENLDLAGKRVLMRVDFNVPIKEGK*ITSNQRIVAALDSIKLALSKKAKSVVLMSHLG
Human -SLS---TLDK--VK----V---------M-NNQ*--N----K--VP---FC-DNG-----------
Horse -SLS---TLDK-MVK----V---------M-NNQ*--N----K--VP---FC-DDG-----------
Mouse -SLS---TLDK--VK----V---------M-NNQ*--N----K--VP---FC-DNG-----------
Trypa -SLS-----TDV--K-----I--------LD-N-K--------G-IPT--H--DNG--A---------
Yeast -SLSS---VDQ---KD-----FI-------LD*--E-----------PT--YU-ENNPRV---A----

Droso.RPD*GNK*NIKYTLAPVAAELKTLLGQDVIFLSDCVGSEVEAACKDPAPGSVILLENVRFYVEEEG
Human.---*-VPMPD--S-E---V---S---K--L--K----P---K--AN--A--------L--H-----
Horse.---V-*PMPD--S-Q---V---S---K--L--K----P---K--A---A--------L--H-----
Mouse.---*-VPMPD--S-E-------S---K--L--K----P---H--AN--A-T------L--H-----
Trypa.--N*-AVHP*--S-K--VP--EA---KP-T-AR----P---SIVN-AS---------L--HI----
Yeast.--N*-ER*-E--S-----K--QS---K--T--H----P-----V-AS---------L-VHI----

Droso.KGLDASGGKVKADPAKVKEFRASLAKLGDVYVNDAFGTAHRAHSSMMGDGFEQRAAGLLLNKELKY
Human.--K----N----E---IEA-----S----------------------V-VNLP-K-G-F-MK---N-
Horse.--K----N----E---IET-----S----------------------V-VNLP-K-G-F-MK---N-
Mouse.--K----N----E---IDA-----S----------------------V-VNLP-K-G-F-MK---N-
Trypa.SAK-KD-N-T---K---E---KG-TR---I-I--------------V-VDLP-K---F-MK---D-
Yeast.*SAKVD-Q----SKED-QK--HE-SS-A---I-----------V-FDLP-----F--E-----

Droso.FSQALDKPPNPFLAILGGAKVADKIQLIENLLDKVNEMIIGGGMAFTFLKVLNNMKIGGSLFDEEG
Human.-AK--ES-ER-------------------N-M----------------------E--T-------
Horse.-AK--ES-ER-----------------N-M----------------------E--T-------
Mouse.-AK--ES-ER-----------------N-M----------------------E--T-------
Trypa.-AK--ES-QR-----------S------D------DTL-VC-------K-----IP--T-----A-
Yeast.-GK--EH-TR-----------D------DSI---------K---E-TE--D-I--KA-

Droso.SKIVEKLVEKAKKNNVQLHLPVDFVCGDKFAENAAVSEATVEAGIPDGHMGLDVGPKTRELFAAPI
Human.A---KD-MS--E--G-KIT------TA---D---KTGQ---AS---R-W----C--ESSKKY-EAV
Horse.A---KH-MS--E--G-KIT------TA---D-H-KTGQ---AS---R-W----C-TESSKKY-EAV
Mouse.A---KD-MS--E--G-KIT------TA---D---KTGQ---AS---R-W----C-TESSKKY-EAV
Trypa.A-TCPSCA-AKA--VKLV*----YITA---DKD-NTGT--D-S-----W----C-E-SI--YKEA-
Yeast.AE--P--M----RKG-EVV-----ITA-A-SAD-NTKTU-DKE---R-WQ---H--ES-K----TV

Droso.ARAKLIVWNGPPGVFEFPNFANGTKSIMDGVVAATKNGTVSIIGGGDTASCCAKWNTEALVSHVST
Human.T---Q------V----WEA--R---AL--E--K--SR-CIT----------T--------DK------
Horse.----Q------V----WEA--R---AL--E--K--SR-CIT--------T--------DK------
Mouse.G---Q------V----WEA--R---AL--E--K--SR-CIT--------T--------DK------
Trypa.-D--T-L----R-----DK------ATL-R--EGC---KIV--------TVA--YGV-DKL-----
Yeast.-K--T----------EK--R---ALL-E--KSSAR-HTV--------TVAK-YGVTDKI-----

Droso.GGGASLELLEGKTLPGVRALTSA-
Human.------------V----D--SNIL
Horse.------------E----D--SNV-
Mouse.------------V----D--SNV-
Trypa.------------E----V--S-K-
Yeast.------------E----RF-SEKK
```

Fig. 8. Amino acid sequence alignment of phosphoglycerate kinases (PGKs). Extra gaps (*) are included to align PGKs from *D. melanogaster* (Droso), *Saccharomyces cerevisiae* (Yeast) (Watson et al. 1982), *Trichoderma reesei* (Trypa) (Vanhanen et al. 1989), mouse (Mori et al. 1986), horse (Banks et al. 1979) and human (Michelson et al. 1983). (–), identical residues. The 12 β-sheets and adjacent peptides that form the inner loops in the substrate-binding cleft are indicated by *double* and *single underlines*, respectively, according to the structure of horse PGK (Banks et al. 1979). The numbering refers to the mammalian PGKs. The *underlined* residues indicate where the introns are located in *D. melanogaster*, human and *T. reesei* genes

mined the 5′ ends of the *Tpi* and *Ald* transcripts in these same RNA preparations and found that each of those transcripts has a unique end (Shaw-Lee et al. 1991, 1992). Therefore the exonuclease activity would have to be specific for *Pgk*. Two clusters of transcript ends are suggested by these data. In each cluster there is a sequence which is somewhat similar to the consensus *Drosophila* transcription start sequence (Hultmark et al. 1986). Therefore we conclude that the *Drosophila Pgk* gene has multiple (at least two) transcription initiation positions which are utilized similarly at both larval and adult stages.

The 3′ region of the *Pgk* transcript was also mapped by RNase protection using probe 2 and probe 3 (Fig. 7). Probe 2 was transcribed from a plasmid that contained the 2.2 kb *Bam*HI genomic fragment and probe 3 was transcribed from a plasmid that contained the 2.7 kb *Bam*HI genomic fragment. These probes were hybridized to *D. melanogaster* total RNA from 2-day-old larvae and 11-day-old adults. We expected a protected fragment common to both probes, representing the distance from the 3′ splice acceptor site of intron 2 to the end of the transcript. We also expected to protect a 97 nt fragment from probe 2 and a 674 nt fragment from probe 3, representing the distances from the 5′ ends of the probes to the 5′ splice donor site of intron 2. As shown in Fig. 7, the protection by probe 2 yielded the expected 97 nt band and the protection by probe 3 yielded the expected 674 nt band. The two protections yield a common band of 337 nt, which locates the polyadenylation site of *D. melanogaster Pgk* gene at about 138 nt downstream of the translation stop codon TAA and 23 nt downstream of the polyadenylation signal, AATAAA. According to the data from RNase protection and primer extension, the size of the PGK mRNA in *D. melanogaster* should be about 1.5–1.6 kb, consistent with the size obtained using Northern blots (Fig. 3).

Evolutionary conservation of PGK

An alignment of the deduced amino acid sequence of *Drosophila* PGK with mammalian and fungal PGKs is shown in Fig. 8. *Drosophila* PGK has 69.6% sequence identity to the horse enzyme and 62.1% sequence similarity to the yeast enzyme. However, when the pattern of substitutions is examined in light of the proposed tertiary structure (Michelson et al. 1985a), it is evident that there is high sequence conservation in the substrate binding region, which is composed of 12 β-sheets and immediately adjacent peptides that form the inner loops of the molecule. In these regions, the horse and *Drosophila* enzymes are 87.2% identical, having only 18 amino acid differences, many of which are conservative substitutions.

Acknowledgements. This work has been supported by N.I.H. grant GM 26830 to D.T.S. We thank Dr. A. Yoshida for the generous gift of the human *Pgk* clone with which we began our studies.

References

Adra CN, Boer PH, McBurney MW (1987) Cloning and expression of the mouse *Pgk-1* gene and the nucleotide sequence of its promoter. Gene 60:65–74

Banks RD, Blake CCF, Evans PR, Haser R, Rice DW, Hardy GW, Merrett M, Philipps AW (1979) Sequence, structure and activity of phosphoglycerate kinase: a possible hinge-bending enzyme. Nature 279:773–777

Bender W, Spierer P, Hogness DS (1975) Chromosomal walking and jumping to isolate DNA from the *Ace* and *rozy* loci and the bithorax complex in *Drosophila melanogaster*. J Mol Biol 168:17–33

Blake CCF, Evan PR, Scopes RK (1972) Structure of horse-muscle phosphoglycerate kinase at 6 Å resolution. Nature 235:195

Boer PH, Adra CN, Lau F, McBurney MW (1987) The testis-specific phosphoglycerate kinase gene Pgk-2 is a recruited retroposon. Mol Cell Biol 7:3107–3122

Bryant TN, Watson HC, Wendell PL (1974) Structure of yeast phosphoglycerate kinase. Nature 247:14–17

Calzone FJ, Britten RJ, Davidson EH (1987) Mapping of gene transcripts by nuclease protection and cDNA primer extension. Methods Enzymol 152:611–632

Chew GK, Cooper DW (1973) Phosphoglycerate kinase polymorphism in *Drosophila*. Biochem Genet 8:267–270

Ciccarese S, Tommasi S, Vonghia G (1989) Cloning and cDNA sequencing of the rat X-chromosome linked phosphoglycerate kinase. Biochem Biophys Res Commun 165:1337–1344

Clements JM, Roberts CF (1986) Transcription and processing signal in the 3-phosphoglycerate kinase (Pgk) gene from *Aspergillus nidulans*. Gene 44:97–105

Eicher EM, Cherry M, Flaherty L (1978) Autosomal phosphoglycerate kinase linked to mouse major histocompatibility complex. Mol Gen Genet 158:225–228

Fujii H, Yoshida A (1980) Molecular abnormality of phosphoglycerate kinase-Uppsala associated with chronic nonspherocytic hemolytic anemia. Proc Natl Acad Sci USA 77:5461–5465

Fujii H, Krietsch WKG, Yoshida A (1980) A single amino acid substitute (Asp to Asn) in a phosphoglycerate kinase variant (Pgk Munchen) associated with enzyme deficiency. J Biol Chem 255:6421–6423

Fujii H, Chen SH, Akatsuka J, Miwa S, Yoshida A (1981) Use of cultured lymphoblastoid cells for the study of abnormal enzymes: molecular abnormality of a phosphoglycerate kinase variant associated with hemolytic anemia. Proc Natl Acad Sci USA 78:2587–2590

Huang IY, Welch CD, Yoshida A (1980) Complete amino acid sequence of human phosphoglycerate kinase. J Biol Chem 255:6412–6420

Hultmark D, Klemenz R, Gehring WJ (1986) Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. Cell 44:429–438

von Kalm L, Weaver J, DeMarco J, MacIntyre RJ, Sullivan DT (1989) Structural characterization of the α-glycerol-3-phosphate dehydrogenase-encoding gene of *Drosophila melanogaster*. Proc Natl Acad Sci USA 86:5020–5024

Lissemore JL, Colbert JT, Quail PH (1987) Cloning of cDNA for phytochrome from etiolated *Cucurbita* and coordinate photoregulation of two distinct phytochrome transcripts. Plant Mol Biol 8:485–496

Mas MT, Chen CY, Hitzeman RA, Riggs AD (1986) Active human-yeast chimeric phosphoglycerate kinases engineered by domain interchange. Science 233:788–790

McCarrey JR (1987) Nucleotide sequence of the promoter region of a tissue-specific human retroposon: comparison with its housekeeping progenitor. Gene 61:291–298

McCarrey JR, Thomas K (1987) Human testis-specific *Pgk* gene lacks introns and possesses characteristics of a processed gene. Nature 326:501–505

Michelson AM, Markham AF, Orkin SH (1983) Isolation and DNA sequence of a full-length cDNA clone for human X chromosome-encoded phosphoglycerate kinase. Proc Nat Acad Sci USA 80:472–476

Michelson AM, Blake CC, Evans ST, Orkin SH (1985a) Structure of the human phosphoglycerate kinase gene and the intron-mediated evolution and dispersal of the nucleotide-binding domain. Proc Natl Acad Sci USA 82:6965–6969

Michelson AM, Bruns GA, Morton CC, Orkin SH (1985b) The human phosphoglycerate kinase multigene family. J Biol Chem 260:6982–6992

Mori N, Singer-Sam J, Lee C-Y, Riggs AD (1986) The nucleotide sequence of a cDNA clone containing the entire coding sequence for mouse X-chromosome linked phosphoglycerate kinase. Gene 45:275–280

Mount SM (1982) A catalog of splice junction sequences. Nucleic Acids Res 10:459–472

Nellmann LJ, Holm F, Atlung T, Hansen FG (1989) Cloning and characterization of the *E. coli* phosphoglycerate kinase gene. Gene 77:185–191

Pegoraro B, Lee CY (1978) Purification and characterization of two isozymes of 3-phosphoglycerate kinase from the mouse. Biochim Biophys Acta 522:423–433

Pegoraro B, Ansari AA, Lee CY, Erickson RP (1978) Immunological relatedness of two isozymes of 3-phosphoglycerate kinase from the mouse. FEBS Lett 95:371–374

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Shaw-Lee RL, Lissemore JL, Sullivan DT (1991) Structure and expression of the triose phosphate isomerase (*Tpi*) gene of *Drosophila melanogaster*. Mol Gen Genet 230:225–229

Shaw-Lee RL, Lissemore JL, Sullivan DT, Tolan DR (1992) Alternative splicing of the fructose 1, 6, bisphosphate aldolase transcript in *Drosophila melanogaster* predicts three isozymes. J Biol Chem 267:3959–3967

Singer-Sam J, Keith DH, Tani K, Simmer RL, Shively L, Lindsay S, Yoshida A, Riggs AD (1984) Sequence of the promoter region of the gene for human X-linked 3-phosphoglycerate kinase. Gene 32:409–417

Skuse GR, Sullivan DT (1985) Developmentally regulated alternate modes of expression of the *Gpdh* locus of *Drosophila*. EMBO J 4:2275–2280

van Solingen P, Muurling H, Koekman B, Berg JVD (1988) Sequence of the *Penicillium chrysogenum* phosphoglycerate kinase gene. Nucleic Acids Res 16:11823

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517

Szabo P, Grzeschik KH, Siniscalco M (1984) A human autosomal phosphoglycerate kinase locus maps near the HLA cluster. Proc Natl Acad Sci USA 81:3167–3169

Vanhanen S, Lehtovaara L, Knowles J (1989) Isolation and characterization of the 3-phosphoglycerate kinase gene (Pgk) from the filamentous fungus *Trichoderma reesei*. Curr Genet 15:181–186

Vieira J, Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268

Voelker RA, Ohnishi S, Langley C (1979) Genetic and cytogenetic studies of four glycolytic enzymes in *Drosophila melanogaster*: aldolase, triosephosphate isomerase, 3-phosphoglycerate kinase and phosphoglucomutase. Biochem Genet 17:769–783

Watson HC, Walker NPC, Shaw PJ, Bryant TN, Wendell PL, Fothergill LA, Perkins RF, Conroy SC, Dobson MJ, Tuite MF, Kingsman AJ (1982) Sequence and structure of yeast phosphoglycerate kinase. EMBO J 1:1635–1640

Yoshida A, Watanabe S, Chen SH, Giblett ER, Ralcolm LA (1972) Human phosphoglycerate kinase. J Biol Chem 247:446–449

Zinn K, DiMaio D, Maniatis T (1983) Identification of two distinct regulatory regions adjacent to the human β-interferon gene. Cell 34:865–879

Communicated by B.H. Judd