The Dstpk61 locus of Drosophila produces multiple transcripts and protein isoforms, suggesting it is involved in multiple signalling pathways

D Clyde and M Bownes

The University of Edinburgh, Institute of Cell and Molecular Biology, Darwin Building, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, UK

(Requests for offprints should be addressed to M Bownes; Email: Mary.Bownes@ed.ac.uk)

Abstract

The Drosophila gene Dstpk61 encodes a serine threonine protein kinase homologous to human phosphoinositide-dependent protein kinase (PDK1), and also has homologues in S. cerevisiae, S. pombe, C. elegans, A. thaliana, mouse, and sheep. Where its function has been investigated, this kinase is thought to be involved in regulating cell growth and survival in response to extracellular signals such as insulin and growth factors. In Drosophila it produces multiple transcripts, some of which appear to be sex-specific.

In addition to the five Dstpk61 cDNAs we have described previously we report the existence of a further 18 expressed sequence tag (EST) cDNAs, three of which we have fully sequenced. We conclude that Dstpk61 is a complex locus that utilises a combination of alternative promoters, alternative splice sites and alternative polyadenylation sites to produce a vast array of different transcripts. These cDNAs encode at least four different DSTPK61 protein isoforms with variant N-termini. In this paper, we discuss the possible functions of the distinct Dstpk61 transcripts and how they might be differentially regulated. We also discuss the roles that DSTPK61 protein isoforms might play in relation to the protein domains they contain and their potential targets in the cell. Finally, we report the putative structure of the human PDK1 gene based on computer comparisons of available mRNA and genomic sequences. The value of using sequence data from other species for experimental design in mammalian systems is discussed.

Journal of Endocrinology (2000) 167, 391–401

Introduction

Signalling cascades are a crucial means of cell–cell communication in all types of organisms. During development, cells need to communicate with each other in order to form a structured organism, with each cell knowing its identity and role. Many of these pathways form evolutionarily conserved cassettes, where homologous proteins perform equivalent roles in different organisms. For example, the EGFR signalling pathway is conserved in Drosophila and mammals and regulates cell proliferation, differentiation and survival in both systems (Sibilia & Wagner 1995, Van Buskirk & Schupbach 1999).

Signalling cascades are also essential for cells to be able to respond to their environment. Multiple signalling pathways exist, each of which results in different cellular responses. However, some signalling proteins appear to be involved in more than one cascade. Such proteins act as branch points in the signalling hierarchy, so that a single agonist can activate a number of cascades and induce multiple cellular responses or different responses under different conditions. One such protein is human phosphoinositide-dependent kinase (PDK1), which has been shown to phosphorylate protein kinase A (PKA) (Cheng et al. 1998), PKB (Alessi et al. 1997b), PKC (Le Good et al. 1998), serum- and glucocorticoid-inducible kinase (SGK) (Park et al. 1999), and p70S6K (Alessi et al. 1998, Pullen et al. 1998).

PDK1 has been most extensively studied for its ability to phosphorylate PKB in response to insulin (Alessi et al. 1997b). This phosphorylation only occurs in the presence of phosphatidylinositol-4,5-bisphosphate and -trisphosphate (PI(3,4)P2 and PI(3,4,5)P3 respectively), produced by PI 3-K. These membrane lipids interact with the PH domain of PDK1, presumably bringing it into proximity with membrane-bound PKB and allowing phosphorylation to occur (Currie et al. 1999). PKB phosphorylation evokes a number of cellular responses, including: increased glucose uptake and glycolysis; increased glycogen synthesis; increased protein synthesis and a concomitant cell cycle progression from G1 to S phase; and increased cell survival, via the repression of the pro-apoptotic proteins BAD and FKHRL1 (del Peso et al. 1997, Scheid & Duronio 1998, Brunet et al. 1999).
Drosophila homologues have also been identified in the pathway upstream of PKB. CHICO is the Drosophila homologue of IRS1–4 and, when mutated, the resulting flies are viable but about one-third the size of wild-type adults (Bohni et al. 1999). Similarly, the Drosophila homologue of the p110 subunit of PI 3-K, Pd110, results in differences in cell size and number when mis-expressed in imaginal discs (Leevers et al. 1996). Thus, the insulin response pathway in Drosophila also appears to be involved in regulating cell growth and proliferation.

PDK1 has also been shown to phosphorylate PKA, also known as cAMP-dependent protein kinase. Signalling via mammalian PKA is complex – multiple isoforms are present, which respond differently to different extracellular stimuli (Skalhegg & Tasken 1997). Thus, PKA elicits multiple biological effects, not all of which are dependent on cAMP – some responses are thought to act through NFκB (Zhang et al. 1997). In mammals, PKA signalling has effects on metabolism, gene regulation, cell growth and division, differentiation, sperm motility, and ion channel conductivity (Skalhegg & Tasken 1997). Three genes encoding PKA catalytic subunits have been found in Drosophila: PKAc1, PKAc2 and PKAc3, also known as DC0, DC1, and DC2. As in mammals, these genes produce multiple transcripts (Kalderon 1997). Again, this reflects the multiple roles that PKA signalling plays in Drosophila, where it is involved in: segmentation, cell growth, oocyte polarity, learning and memory, and regulation of quantal size at the neuromuscular junction. Interestingly, PKA is thought to exert its effect on segmentation via hedgehog in a cAMP-independent manner (Ohlmeyer & Kalderon 1997).

Many PKC isoforms are also phosphorylated by PDK1 (Le Good et al. 1998, Dong et al. 1999). These different isoforms vary considerably in their time, position and level of expression, reflecting the diverse cellular functions of PKC in cell transformation, growth, differentiation, membrane ruffling, vesicle trafficking, and gene expression (Liu & Heckman 1998). Three PKC encoding genes have been identified in Drosophila: Pkc53E, Pkc98E, and inactivation no afterpotential C (inaC). Pkc98E and Pkc53E are expressed only in the head, and InaC is detected only in the eye. These restricted expression patterns reflect the functions of PKC in Drosophila, where it is thought to be involved mainly in neural activation and learning and memory. However, it is possible that ubiquitously expressed PKC homologues remain to be identified in Drosophila. The complete Drosophila genome sequence is now available and reveals the presence of a few previously unidentified Pke-like genes, although the expression patterns of transcripts from these loci remain to be investigated.

Mammalian p70S6K is activated in response to insulin and mitogen stimulation, and results in the increased translation of 5′ TOP mRNAs. This set of mRNAs encodes mostly proteins of the translational machinery, such as ribosomal proteins. Their upregulation is essential for cell cycle progression from G1 to S phase, and is thus associated with the regulation of cell proliferation and/or differentiation (for review see Peterson & Schreiber 1998). Full activation of p70S6K requires phosphorylation by PDK1 (Alessi et al. 1998, Pullen et al. 1998). A Drosophila homologue of p70S6K exists and is known as Dp70S6K or S6K (Stewart et al. 1996, Watson et al. 1996). Although no functional studies have been reported, it is interesting to note that Dp70S6K encodes two protein isoforms: 3.7 and 5.0 kb transcripts encode an identical 637 amino acid protein, whereas a 2.8 kb transcript encodes a smaller protein that lacks the 51 N-terminal residues of the larger isoform (Watson et al. 1996, Stewart et al. 1996). The mammalian gene also encodes two isoforms, the larger of which is targeted to the nucleus, the other is found in both the nucleus and the cytoplasm. Also, like Dstpk61, Dp70S6K produces sex-enriched transcripts: the 3.7 kb transcript is most abundant in males, whereas the 2.8 kb transcript is most abundant in females (Stewart et al. 1996). The 5.0 kb transcript is expressed at similar levels in both sexes. Thus, it is possible that Dstpk61 and Dp70S6K are involved in a pathway that regulates sex-specific events.

SGK has also been shown to be a substrate of PDK1 (Kobayashi & Cohen 1999, Park et al. 1999). The function of SGK is not clear, but it has been observed to undergo cytoplasmic-nuclear shuttling in synchrony with the cell cycle. This shuttling is differentially regulated depending on the proliferative state of the cell, suggesting that SGK may regulate cell proliferation (Park et al. 1999). No SGK homologue has been reported in Drosophila, and database searches do not identify possible candidate genes. Thus, it is possible that functions carried out by mammalian SGK might be mediated by a different kinase in Drosophila.

It has recently been shown that PDK1 also interacts with sphingosine, enabling it to phosphorylate a different target – PAK1 (King et al. 2000). Thus, interactions between PDK1 and different lipid components of the cell membrane may activate alternative signalling cascades, thereby eliciting different cellular responses. Similarly, it has been shown that phosphorylation of p70S6K by PDK1 is not dependent on the presence of phospholipids (Alessi et al. 1998). Thus, cytoplasmic PDK1 may have yet another set of substrates.

PDK1 homologues have been identified in a number of organisms. S. cerevisiae has two PDK1 genes, PKH1 and PKH2, which are required for cell growth and viability (Casamayor et al. 1999). Similarly, the S. pombe homologue, ksg1, is required for growth, and also for mating and sporulation (Niederberger & Schweingruber 1999). The C. elegans homologue has multiple phenotypic effects, including the induction of diapause at the dauer larval stage and increased lifespan (Paradis et al. 1999).
Biochemical analysis has identified a homologue in *A. thaliana* (AtPDK1) (Deak et al. 1999) and four PDK1 isoforms in sheep brains (Stephens et al. 1998). Perhaps the most extensively studied homologue is the mouse gene, mPDK1. mPDK1 mRNAs appear to be expressed in a wide range of tissues, and produce multiple proteins ranging in size from 60 to 180 kDa. It also appears to encode a testis-specific transcript (Dong et al. 1999).

*Dstpk61* is thought to be the *Drosophila* homologue of PDK1 (Alessi et al. 1997a). We have previously reported the presence of at least four different *Dstpk61* transcripts: a common transcript; a female-specific carcass transcript; an ovary-specific transcript; and a testis-specific transcript (MacDougall et al. 1999). We also reported the sequence of three different cDNAs, one of which (EST40) is thought to represent the female-specific carcass transcript (MacDougall et al. 1999). This paper looks in detail at the different variants of RNA and protein generated by this gene in *Drosophila* and identifies some very intriguing results. This information is used to speculate on how the gene might function in different signalling pathways by producing transcripts and/or proteins that are localised to different regions of the cell. We also report the predicted structure of a mammalian PDK1 gene, and suggest that it too might produce multiple transcripts.

*Drosophila* is a very flexible model system, allowing information to be gathered by a variety of scientific approaches. Thus, a detailed analysis of *Dstpk61* in *Drosophila* should aid the design of informative experiments to further elucidate the role of PDK1 in mammals. The publication of the *Drosophila* and human genomes and the development of expressed sequence tag (EST) databases in both systems should enable analysis of gene structure and the transcripts and proteins generated from a gene in *Drosophila*, a species amenable to genetic analysis, to help direct the search for protein isoforms with subtly different functions in humans. It is crucial that any isoforms are fully investigated, as they will provide the possibility of targeting a specific protein isoform for disease treatment. This may avoid the side effects which are bound to occur when a single gene participates in multiple signalling control pathways under a variety of cellular situations.

**Materials and Methods**

**Midi-preparation of plasmid DNA**

Hybaid (Ashford, Middlesex, UK) midi-prep kits were used to prepare up to 200 µg plasmid DNA, following the manufacturer’s instructions.

**Automated DNA sequencing**

Approximately 400 ng high quality plasmid DNA were used as a template in sequencing reactions. 3·2 picomoles of the appropriate primer and 8 µl reaction mix (ABI PRISM dye terminator cycle sequencing kit, Perkin-Elmer Corporation, Warrington, Cheshire, UK) and double distilled H2O were added to a final volume of 20 µl. The mixture was subjected to thermal cycling (Touchdown thermal cycler, Hybaid), as recommended by the Perkin-Elmer instruction manual. The program used is as follows: 25 cycles of 30 s denaturation at 96 °C; 15 s at the appropriate primer annealing temperature; and a 4 min extension period at 60 °C. The products were analysed using an ABI PRISM 377 DNA sequencer (Perkin-Elmer Corporation).

**Sequence analysis**

DNA sequences were analysed using GenejockeyII (Biosoft, Cambridge, UK) and the Genetics Computer Group (GCG; http://www.hgmp.msc.ac.uk) program package, version 10. BLAST searches (Altschul et al. 1990) were carried out using the NCBI server. This approach identified numerous EST clones and one genomic clone, BACR48E09/AC006169. The structure of the gene was then deduced by comparing the cDNA sequences to the genomic sequence. Protein sequences were analysed for motifs using ProDom (Corpet et al. 1999, 2000), and subcellular localisation was predicted using PSORT (Nakai & Kanehisa 1992).

**Results**

*Dstpk61* is represented 21 times in the EST database

BLAST searches of the EST database were carried out using the sequence of cDNA **B** as a probe. This approach identified 21 ESTs. Sixteen EST cDNAs were identified from an embryonic library, four from the adult head library, and one from the adult ovarian library. Thus, *Dstpk61* is expressed at different stages, and in multiple tissues, during development. We classified the ESTs into three distinct groups based on their 5′ sequence, as shown in Table 1.

The first exon of Group 1 ESTs, which we have named exon 1a, is similar to that of cDNA **B** (see Fig. 1A). This group contains LD16509 (EST09), which we have previously sequenced and published (MacDougall et al. 1999). EST09 is thought to be a more complete version of cDNA **B**. We have sequenced a further two Group 1 cDNAs, LD19385 (EST85) and LD32017 (EST17). These cDNAs appear to represent alternative splice forms of *Dstpk61*.

Group 2 ESTs contain a novel first exon, which we have termed exon 1b (see Fig. 1Aii). Exon 1b is located approximately 3 kb upstream of exon 1a. We have previously reported the sequence of LD19940 (EST40), which we believe to be a female-specific carcass transcript.

---

**Materials and Methods**

**Midi-preparation of plasmid DNA**

Hybaid (Ashford, Middlesex, UK) midi-prep kits were used to prepare up to 200 µg plasmid DNA, following the manufacturer’s instructions.

**Automated DNA sequencing**

Approximately 400 ng high quality plasmid DNA were used as a template in sequencing reactions. 3·2 picomoles of the appropriate primer and 8 µl reaction mix (ABI PRISM dye terminator cycle sequencing kit, Perkin-Elmer Corporation, Warrington, Cheshire, UK) and double distilled H2O were added to a final volume of 20 µl. The mixture was subjected to thermal cycling (Touchdown thermal cycler, Hybaid), as recommended by the Perkin-Elmer instruction manual. The program used is as follows: 25 cycles of 30 s denaturation at 96 °C; 15 s at the appropriate primer annealing temperature; and a 4 min extension period at 60 °C. The products were analysed using an ABI PRISM 377 DNA sequencer (Perkin-Elmer Corporation).

**Sequence analysis**

DNA sequences were analysed using GenejockeyII (Biosoft, Cambridge, UK) and the Genetics Computer Group (GCG; http://www.hgmp.msc.ac.uk) program package, version 10. BLAST searches (Altschul et al. 1990) were carried out using the NCBI server. This approach identified numerous EST clones and one genomic clone, BACR48E09/AC006169. The structure of the gene was then deduced by comparing the cDNA sequences to the genomic sequence. Protein sequences were analysed for motifs using ProDom (Corpet et al. 1999, 2000), and subcellular localisation was predicted using PSORT (Nakai & Kanehisa 1992).

**Results**

*Dstpk61* is represented 21 times in the EST database

BLAST searches of the EST database were carried out using the sequence of cDNA **B** as a probe. This approach identified 21 ESTs. Sixteen EST cDNAs were identified from an embryonic library, four from the adult head library, and one from the adult ovarian library. Thus, *Dstpk61* is expressed at different stages, and in multiple tissues, during development. We classified the ESTs into three distinct groups based on their 5′ sequence, as shown in Table 1.

The first exon of Group 1 ESTs, which we have named exon 1a, is similar to that of cDNA **B** (see Fig. 1A). This group contains LD16509 (EST09), which we have previously sequenced and published (MacDougall et al. 1999). EST09 is thought to be a more complete version of cDNA **B**. We have sequenced a further two Group 1 cDNAs, LD19385 (EST85) and LD32017 (EST17). These cDNAs appear to represent alternative splice forms of *Dstpk61*.

Group 2 ESTs contain a novel first exon, which we have termed exon 1b (see Fig. 1Aii). Exon 1b is located approximately 3 kb upstream of exon 1a. We have previously reported the sequence of LD19940 (EST40), which we believe to be a female-specific carcass transcript.
LD46834 may represent a 5′/p9′ form of di and consists of four earlier publication (MacDougall et al. 1999). Likewise, it is probable that Group 2 truncated and uses P′09, or if it represents a full-length other cDNAs. Similarly, it is not clear if cDNA A is LD13190, which may itself be a 5′ transcript transcribed from a di (see Fig. 1B).

LD46834 may represent a 5′/p9′ form of di and consists of four earlier publication (MacDougall et al. 1999). Likewise, it is probable that Group 2 truncated and uses P′09, or if it represents a full-length other cDNAs. Similarly, it is not clear if cDNA A is LD13190, which may itself be a 5′ transcript transcribed from a di (see Fig. 1B).

**Table 1 CDNA libraries from which EST sequences have been derived**

<table>
<thead>
<tr>
<th>Name</th>
<th>Source</th>
<th>Cloning information</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>Adult heads</td>
<td>EcoR I/Xho I sites of Uni-Zap XR or pOT2</td>
<td>Many clones not full-length or have no inserts</td>
</tr>
<tr>
<td>GH</td>
<td>Adult heads</td>
<td>EcoR I/Xho I sites of Uni-Zap XR or pOT2</td>
<td>Separate from HL library. Made because many HL clones were not full-length or had no inserts</td>
</tr>
<tr>
<td>GM</td>
<td>Ovaries, stage 1–6</td>
<td>EcoR I/Xho I sites of Uni-Zap XR or pOT2</td>
<td>Biased for mRNAs encoding secreted and membrane-bound proteins</td>
</tr>
<tr>
<td>LP</td>
<td>Various larval stages and early pupae</td>
<td>EcoR I/Xho I sites of Uni-Zap XR or pOT2</td>
<td></td>
</tr>
<tr>
<td>LD</td>
<td>Embryos, 0–22 h</td>
<td>EcoR I/Xho I sites of Uni-Zap XR or pOT2</td>
<td></td>
</tr>
<tr>
<td>CK</td>
<td>Rough endoplasmic reticulum from 8–16 h embryos</td>
<td>Hind III/Pst I sites of pBluescript SK(+)</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>Drosophila tissue culture cells</td>
<td>Eco R I/Xho I sites of pOT2</td>
<td></td>
</tr>
</tbody>
</table>

(DMacDougall et al. 1999). A further two Group 2 cDNAs have now been sequenced, LD02269 (EST69) and GM04058 (EST58). EST69 is thought to represent the alternatively spliced non-sex-specific carcass transcript reported in MacDougall et al. 1999.

Group 3 ESTs contain neither exon 1a nor exon 1b (Fig. 1Aii). These cDNAs contain either a novel first exon (GH15751, LD38659, and LD13752), or start transcription midway through a previously characterised exon (LD13190 and LD46834). In the latter case, it is possible that these cDNAs are truncated at their 5′ ends.

**Dstpk61 expression is regulated in many different ways**

Based on the above sequence data, we have determined that the Dstpk61 locus consists of at least 11 different exons and produces at least 10 different transcripts (see Fig. 1B). This variety is achieved by using alternative promoters, alternative polyadenylation sites, and alternative splice sites.

Group 1 transcripts are likely to be produced from one promoter, which we previously termed P′99 (MacDougall et al. 1999). Likewise, it is probable that Group 2 transcripts all use a second distinct promoter, termed P′10 in our earlier publication (MacDougall et al. 1999). Group 3 consists of five ESTs, all of which appear to be transcribed from different promoters. However, it is possible that LD46834 may represent a 5′ truncated version of LD13190, which may itself be a 5′ truncation of any of the other cDNAs. Similarly, it is not clear if cDNA A is truncated and uses P′99, or if it represents a full-length transcript transcribed from a different promoter. Thus, at least five, and possibly eight, different promoters are used to produce mRNAs from the Dstpk61 locus (see Fig. 1B).

Different polyadenylation sites are also used in different transcripts (see Fig. 1B). At least 10 potential polyadenylation signals have been delimited in the 3′ sequence of Dstpk61 (MacDougall et al. 1999). Six different polyadenylation signals are used in the eight cDNAs that have been fully sequenced, and we have recently isolated an additional two cDNAs that use two further poly(A) sites (Sathirana, T Wood, D Clyde & M Bownes, unpublished observations). It is likely that all 10 potential poly(A) signals will be used in vivo to produce transcripts that have variable 3′ untranslated regions (UTRs).

Alternative splicing is also used to further increase the number of transcripts from the Dstpk61 locus. Exon 1a contains three different splice donor sites, giving rise to three alternative forms of this exon. Also, EST40 and GM04058 are spliced differently to LD02269, which does not contain exons 2 and 3. This splice event is likely to be regulated in a sex-specific manner in adult carcass tissue – EST40 (or a similar Group 2 transcript) is detected in female, but not male, carcass RNA by RT-PCR, whereas EST69 is detected in carcass RNA from both sexes (MacDougall et al. 1999).

Considering the variety of promoters, splice sites, and polyadenylation sites available, it is possible that the transcripts we have described represent only a subset of those produced from Dstpk61 in the cell.

**Dstpk61 transcripts encode at least four different protein isoforms**

Despite the large number of potential Dstpk61 transcripts described above, they encode only four predicted protein isoforms (see Fig. 2). EST40 and EST58 encode an 836 amino acid protein, DSTPK61836, the largest of the four isoforms. GH15751 encodes a slightly smaller protein of 819 amino acids (DSTPK61819), which lacks the 22 most N-terminal amino acids of DSTPK61836 and replaces them with the five amino acid sequence MNIIQ. The majority of the predicted transcripts (cDNA A/EST09, cDNA A, EST17, EST85, LD13190, and LD46834) encode a 755 amino acid protein, DSTPK61755. DSTPK61755 is identical to DSTPK61836 except that it lacks 81 amino acids from the N-terminus. The remaining transcripts (EST69, LD38659, and LD13752) encode a 539 amino acid protein, DSTPK61539, which is identical to the most C-terminal 539 amino acids of the other three isoforms. Raising antibodies to bacterially produced
recombinant DSTPK61 (Alessi et al. 1997a) and to peptides from various regions of the proteins will enable us to determine the distribution of these isoforms in Drosophila tissues.

**DSTPK61 contains multiple protein domains**

We have previously shown that DSTPK61 contains three distinct domains: an OPA repeat (or CAX) domain; a Pleckstrin Homology (PH) domain; and a kinase domain (Alessi et al. 1997a, MacDougall et al. 1999). We have also shown that recombinant DSTPK61 is capable of phosphorylating PKB in vitro in a manner similar to its proposed human homologue, PDK1 (Alessi et al. 1997a).

A ProDom (Corpet et al. 1999, 2000) search was carried out to identify further motifs and domains. The amino acid sequence of each isoform was also analysed using the PSORT program (Nakai & Kanehisa 1992) to determine its most likely subcellular location. It was hoped these approaches might help elucidate the functional significance of the alternative protein isoforms.

The ProDom results revealed the presence of a putative DNA binding domain immediately N-terminal of the OPA repeat domain and a bipartite nuclear localisation signal (see Fig. 2). PSORT confirmed that all four isoforms are likely to be nuclear (see Fig. 2). However, DSTPK61 also contains a putative N-terminal signal found in proteins that are located in the mitochondrial intermembrane space (see Fig. 2). This sequence is not present in the other three isoforms. The shortest of the four proteins, DSTPK61, lacks part of the kinase domain and one part of the bipartite nuclear localisation signal.
**Figure 2** Diagram showing the structure of the four predicted DSTPK61 isoforms encoded by Dstpk61 EST cDNAs. The motifs and domains are coded as shown in the key above. The probability, as calculated by PSORT, of the protein being found in the stated subcellular location is shown in brackets.
**PDK1 is likely to produce multiple transcripts**

Three identical 1891 nucleotide human PDK1 mRNA sequences (Accession numbers Y15056, AF017995, and NM_002613), which encode a 555 amino acid protein, have been submitted to the sequence database. These mRNAs show homology to the sequence contained in the P1 clone, AC005591, which maps to human chromosome 16. This contradicts genetic mapping information, which places PDK1 on human chromosome 22.

Comparison of the mRNA sequence in the database to the genomic sequence reveals that the gene is composed of 14 exons, the first of which is not contained within the P1 clone. However, it appears that the clone contains two copies of exons 7–10, and exons 2–10 are in the opposite orientation to exons 7–14 (see Fig. 3A). It is possible that the P1 clone contains a PDK1 pseudogene resulting from a partial gene duplication event during evolution, but the rearrangement could also be an artefact of the cloning and sequencing processes.

The genomic sequence was scanned for additional potential exons using the FGENESH program (Smith et al. 1996), which identified 23 putative exons not contained within the published mRNA sequence (see Fig. 3A). Six of these exons are likely to represent a distinct gene downstream of PDK1, which is transcribed in the opposite direction (see Fig. 3B). It is possible that the other exons could be included in PDK1 mRNAs by alternative splicing to produce alternative PDK1 transcripts. However, it should be noted that gene prediction programs have their limitations. Indeed, FGENESH failed to identify some of the exons used in the published PDK1 mRNA sequence.

It is very likely that a number of PDK1 transcripts are present within the cell. At least 116 EST cDNAs have been classified as potential PDK1 transcripts. These cDNAs range in size from 0.3 to 4.2 kb, and have been isolated from many different tissues, including eye, testes, ovary, brain and blood. Those that have been sequenced appear to share identical 3' ends, although the 5’ ends appear to differ. Many of the smallest clones are likely to be truncated. However, some of the smaller species will not be truncated, as the 5’ sequence of the larger cDNAs is often contained within the smaller cDNAs. This observation suggests that the larger cDNAs may be produced from alternative internal promoters and may contain additional exons. Thus, it is very likely that, as for Dstpk61, multiple PDK1 protein isoforms exist, as the additional exons are likely to be introduced into protein-coding regions.

**Discussion**

*What is the biological significance of alternative Dstpk61 transcripts?*

Although PDK1 homologues exist in a number of organisms, the gene structure and transcriptional activity of these loci have not been studied in detail. To date, only mPDK1 has been shown to produce multiple transcripts and protein isoforms (Dong et al. 1999). No alternative functions have been attributed to the variant mRNAs or proteins, nor has their cellular distribution been reported.

Dstpk61 is the *Drosophila* homologue of PDK1, and we have shown that this locus produces multiple transcripts. Some of these encode different protein isoforms, the significance of which will be discussed in the next section. However, some transcripts encode identical proteins: the mRNAs differ only in the nucleic acid sequence of their 5’ and/or 3’ UTRs. Why do these variant transcripts exist?

Differences in 5’ UTRs occur due to the use of different promoters. These promoters may be regulated by different trans-acting factors. The activity of a given promoter will depend on the availability of the appropriate regulatory proteins, some of which may be present only in certain tissues and/or at certain developmental stages. This may allow their expression to be co-ordinated with other genes in specific response pathways in different combinations in different cells, or at different times.

Similarly, alternative splicing of Dstpk61 transcripts could be regulated by splicing factors with restricted expression patterns. Thus, some Dstpk61 transcripts may be expressed in a temporally and/or spatially restricted manner. Indeed, we have previously reported the presence of a female carcass-specific Dstpk61 transcript that presumably arises from alternative splicing. Many examples exist in *Drosophila* of splicing factors with restricted expression domains. For example, somatic sex is determined in flies by a series of alternative splicing events that are regulated by the female-specific splicing factors, SX1 and TRA (Boggs et al. 1987, Hoshijima et al. 1991). Splicing of the homeotic gene, *Ultrabithorax* (*Ubx*), is also regulated in a tissue and temporal-specific manner (O'Connor et al. 1988).

The 5’ and 3’ UTRs may also contain binding sites for trans-acting factors (proteins or RNAs) that regulate translation or the subcellular localisation of the transcript. Again, this might occur in a spatially or temporally dependent manner according to the distribution of the relevant trans-acting factors. For example, many sperm-specific transcripts in *Drosophila* contain Translational Control Elements (TCEs) in their 5’ UTRs, which ensure that translation is repressed until the appropriate stage of spermatogenesis (for review see Schäfer et al. 1995). *Drosophila* development also requires certain transcripts to be specifically localised. A well understood example is the localisation of *bicoid* and *oskar* mRNAs during embryogenesis and *prospero* during nervous system development (Campos-Ortega 1997). STAUFEN is crucial for this localisation by binding to the 3’ UTRs of the mRNAs (for review on RNA localisation see Bashirullah et al. 1998).
What is the biological significance of alternative DSTPK61 protein isoforms?

Proteins are modular structures that consist of distinct domains with discrete functions. Specific types of domains can be combined to form proteins with very specialised capabilities. Alternative splicing is one method whereby different protein domains can be included or excluded from a family of proteins encoded by a single locus, thereby conferring on them different cellular functions. This is one possible explanation for the production of multiple Dstpk61 transcripts, which encode four protein isoforms.

The four DSTPK61 isoforms contain all or some of the following domains and motifs: a kinase domain; a Pleckstrin Homology (PH) domain; an OPA repeat/DNA binding domain; nuclear localisation signals; and a mitochondrial intermembrane space targeting sequence.

The kinase domain is essential for the role of DSTPK61 as a signalling molecule. It contains two putative ATP binding pockets, which are responsible for the phosphorylation of substrate molecules. DSTPK61^{836}, DSTPK61^{819} and DSTPK61^{755} have identical kinase domains. DSTPK61^{539} lacks 52 N-terminal amino acids of the kinase domain, disrupting one of the two ATP binding pockets. It is not clear if this isoform will be kinase dead, or will have an altered substrate specificity. In the long term, this could be investigated by producing this form of the kinase in vitro and checking its biochemical activity.

The PH domain is involved in membrane and protein:protein interactions. In mammals, the PH domain of PDK1 interacts with phosphoinositol lipids (PI(3,4)P2 and PI(3,4,5)P3) at the cell membrane, increasing its ability to activate certain substrates, such as PKB (Alessi et al. 1997a). Recombinant DSTPK61^{755} has also been shown to activate PKB in a PI(3,4)P2 and PI(3,4,5)P3 dependent manner (Alessi et al. 1997a). However, PDK1 can also interact with other lipids, such as sphingosine, allowing it to phosphorylate different substrates (King et al. 2000). These interactions presumably occur via the PH domain. All four DSTPK61 isoforms contain identical PH domains. It is possible, nevertheless, that different isoforms interact with different lipids or proteins dependent on their tissue distribution and subcellular localisation.

The OPA repeat (or CAX) domain is a stretch of glutamine residues found in an array of proteins in many
organisms, from yeast to humans. In Drosophila, OPA repeats are found mainly in proteins with developmentally restricted expression patterns, for example, the neurogenic protein Notch. Many OPA repeat containing proteins also have DNA binding domains, for example the homeobox proteins in Drosophila, which are transcriptional regulators (Biggin & McGinnis 1997). The three largest DSTPK61 isoforms (DSTPK611836, DSTPK61819, and DSTPK61755) all contain an OPA repeat and upstream putative DNA binding domain. However, DSTPK61539 lacks these domains, suggesting that it might have a ubiquitous expression domain with a more general cellular role, possibly functioning in the cytoplasm. This is in keeping with RT-PCR analysis, which suggests that the transcript represented by EST69 is expressed in both male and female carcasses (MacDougall et al. 1999).

PSORT analysis of the four DSTPK61 isoforms suggests that they all have a high probability of being localised to the nucleus. This theory is substantiated by the presence of two types of nuclear localisation signals. One targeting sequence is located within the kinase domain. The other is a bipartite signal, one part of which is situated at the N-terminus of the kinase domain, with the other part at the C-terminal end of the kinase domain. The three larger DSTPK61 isoforms contain all these targeting sequences, and have a probability of 0.94 of being nuclear. What might be the functional relevance of nuclear DSTPK61 localisation?

It is possible that DSTPK61 has different substrates and/or cellular responses in the nucleus. Indeed, it may be that different DSTPK61 isoforms have slightly different tertiary structures. Thus, some isoforms might be targeted directly to the nucleus. However, the nuclear targeting signals might be hidden in other isoforms and they might translocate to the nucleus only when they interact with specific substrates.

DSTPK61539 lacks the N-terminal part of the bipartite nuclear localisation signal. The predicted probability of this isoform being located in the nucleus is 0.88. Again, this observation suggests that DSTPK61539 might be less restricted in its function and location than the other isoforms.

PSORT analysis also identified an N-terminal motif in DSTPK611836 that targets proteins to the mitochondrial intermembrane space. The other DSTPK61 isoforms lack this sequence. It is interesting that DSTPK611836 is encoded by transcripts that have been detected in female carcass tissue and ovaries, but not in male carcasses. This might reflect the higher energy requirement of females. However, it is interesting to note that the potential DSTPK61 substrate, PKC, is located in the mitochondria, where it has a role in promoting cell survival (for review, see Cross et al. 2000). It could be that it is via this substrate in mitochondria that DSTPK61 can affect cell survival. There are differences in the nerve cell divisions in males and females. Also, interactions are needed between somatic cells and germ cells of the female for germline sex determination to occur and for oogenesis to be initiated (Steinmann-Zwicky et al. 1989, Horabin et al. 1995).

Experiments with Drosophila should reveal the functions of DSTPK61 and PDK1

As yet, little is known about Dstpk61 function in Drosophila, and much work remains to be done. Recombinant DSTPK61755 has already been shown to possess kinase activity (Alessi et al. 1997a). Kinase assays should be carried out on the other three isoforms. It will be particularly interesting to determine if DSTPK61539, which lacks part of the kinase domain, retains kinase activity.

Raising antibodies to different regions of the DSTPK61 isoforms should allow their tissue and developmental expression profiles to be determined, and will confirm which isoforms are found in vivo. Antibodies could also be used for immuno-co-precipitation experiments to identify potential DSTPK61 substrates.

The subcellular localisation of the various protein isoforms should also be investigated—are all isoforms found in the nucleus, and is DSTPK611836 found in the mitochondrial intermembrane space? The in vivo localisation of the various isoforms can be studied by making GFP-tagged recombinant proteins. It will also be interesting to determine if DSTPK611836 can be detected in purified mitochondria.

Currently, we are undertaking screens for mutations in the Dstpk61 locus, to enable us to investigate their phenotypes. Dstpk61 is expressed in early embryos, so it is possible that Dstpk61 null individuals will die early in development. However, it may be possible to create viable fly lines that lack specific transcripts or protein isoforms, which should help determine if different isoforms or transcripts are responsible for different aspects of Dstpk61 function. Transgenic technology will also allow us to over-express, or ‘knock-out’, Dstpk61 function at later developmental stages if viable mutant lines cannot be raised.

Genetic studies should clarify the in vivo roles of DSTPK61 and help identify its substrates and the pathways in which it is involved. This information should aid the design of more informed experiments in mammalian systems that do not have the experimental flexibility of Drosophila.

Acknowledgements

The authors wish to thank Sheila Milne for help in preparing this manuscript, Kevin Leaper for helpful comments and advice, Nicola Preston for assistance with
automated sequencing, and Steven Addicott for assistance with diagrams.

References


Niederberger C & Schweinigruber ME 1999 A Sshizosaccharomyces pombe gene, sgl1, that shows structural homology to the human phosphoinositide-dependent protein kinase PDK1, is essential for growth, mating and sporulation. Molecular and General Genetics 261 177–183.


del Peso L, Gonzalez-Garcia M, Page C, Herrera R & Nunez G 1997 Interleukin-3-induced phosphorylation of BAD through the protein kinase Akt. Science 278 687–689.


Steinmann-Zwicky M, Schmid H & Nother R 1989 Cell-autonomous and inductive signals can determine the sex of the germ line of *Drosophila* by regulating the gene *Sxl*. *Cell* 57 157–166.


Received 30 May 2000

Revised manuscript received 17 July 2000

Accepted 22 August 2000