Structure and nucleotide sequence of a *Drosophila melanogaster* protein kinase C gene

Arnon Rosenthal, Lucy Rhee, Ramin Yadegari, Renato Paro, Axel Ullrich and David V. Goeddel

Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080, and 1Department of Biochemistry, Stanford University Medical Center, Stanford, CA 94305, USA

Communicated by R.A. Flavell

Genomic and cDNA clones encoding a *Drosophila melanogaster* protein kinase C (PKC) homologue were identified using a bovine PKC cDNA probe. The cDNA clones contain a single open reading frame that encodes a 639 amino acid, 75-kd protein having extensive homology with bovine, human and rat PKC and homology with the kinase domains of other serine, threonine and tyrosine kinases. The *Drosophila* PKC gene is localized to region 53E of chromosome 2. The gene spans ~20 kb and contains at least 14 exons. Messenger RNA for PKC could not be detected in 0–3 h *Drosophila* embryos. Adult flies contain three PKC transcripts of 4.3, 4.0 and 2.4 kb.

Key words: *Drosophila melanogaster* / protein kinase C / nucleotide sequence

Introduction

The character of a cell is largely defined by a diverse set of specific cell surface receptors which translate external stimuli into physiological responses. The mechanisms by which cellular signalling cascades generate cell type-dependent and agonist-specific responses are only partially understood but are thought to involve reversible activation of substrate-specific protein kinases (Greengard, 1978; Nishizuka, 1980). These phosphokinases are modulated either directly following receptor stimulation (as in the case of tyrosine kinase receptors; for review see Hunter and Cooper, 1985) or through the generation of secondary messengers, cAMP (activation of cAMP dependent protein kinase; Kuo and Greengard, 1969), cGMP (activation of cGMP dependent protein kinase; Kuo, 1974), and calcium (activation of calcium calmodulin-dependent protein kinases; Cohen et al., 1978; Dabrowska and Hartshorne, 1978; Payne and Soderling, 1980).

Recently another receptor-controlled signalling mechanism that utilizes novel second messengers and activates a different type of protein kinase has been described (Hokin and Hokin, 1954; Durell et al., 1969; Michell, 1975; Takai et al., 1979a,b; for review see Nishizuka, 1984, 1986). In this system, agonist receptor interaction results in a G protein-mediated activation of phospholipase C (Cockcroft and Gomperts, 1985; Paris and Pouyssegur, 1986). Phospholipase C catalyzes the hydrolysis of phosphatidylinositol-4,5-bisphosphate, a minor membrane phospholipid, and transiently generates two secondary messengers: inositol-1,4,5-trisphosphate (Ins-1,4,5-P3) and 1,2-sn-diaclylglycerol (1,2-DG) (Hokin and Hokin, 1954; for review see Hokin, 1985). Ins-1,4,5-P3 promotes calcium release from internal non-mitochondrial sources (Streb et al., 1983; for review see Berridge and Irvine, 1984), while 1,2-DG activates protein kinase C (PKC), a 75–80-kd phospholipid- and calcium-dependent serine and threonine-specific protein kinase (Takai et al., 1979a,b; Kishimoto et al., 1980; for review see Nishizuka, 1984, 1986).

PKC appears to mediate some of the mitogenic effects observed in response to the ‘competence factors’ platelet-derived growth factor (Berridge et al., 1984; Coughlin et al., 1985), fibroblast-derived growth factor (Stiles et al., 1979; Rozengurt, 1981), cartilage-derived growth factor (Macara, 1985), bradykinin (Coughlin et al., 1985; Yano et al., 1984), bombesin (Brown et al., 1984), vasopressin (Rozengurt, 1981), thrombin (Pouyssegur et al., 1982), prostaglandin F2α (MacPhee et al., 1984), phytomegglutinin antigens (Coggeshall and Cambier, 1984), interleukin-2 (Farrar and Anderson, 1985), and interleukin-3 (Farrar et al., 1985), partially through activation of the proto-oncogenes c-fos (Greenberg and Ziff, 1984) and c-myc (Muller et al., 1984). PKC also phosphorylates and possibly mediates diminished response to agonists of receptors for epidermal growth factor (Cochet et al., 1984), insulin (Jacobs et al., 1983), somatomedin C (Jacobs et al., 1983), transferrin (May et al., 1984), interleukin-2 (Shackelford and Trowbridge, 1984), β-adrenergic (Sibley et al., 1984), nicotinic acetylcholine (Huganir et al., 1984), and immunoglobulin E (Teshima et al., 1984), and thus acts as a negative regulator of biological responses. It is involved in differentiation of keratinocytes (Yuspina et al., 1985), murine erythroleukemia cells (Falleti et al., 1985), F9 embryonal carcinoma cells (Kraft and Anderson, 1983), thymocytes (Kaiuchibi et al., 1985) and B lymphocytes (Guy et al., 1985). Activation of PKC through an increase in inositol lipid turnover is thought to mediate part of the transforming phenotype induced by the oncoproteins fes, fins, p68 src (Macara et al., 1984), p68 src (Wakelam et al., 1986), and ras (Wakelam et al., 1986), and its direct activation is involved in the tumor-promoting action of phorbol esters (Castagna et al., 1982; Parker et al., 1984). PKC is also involved in regulation and secretion of enzymes, neurotransmitters and hormones, and plays a role in neuronal excitation, synaptic plasticity, and learning (for review see Nishizuka, 1986).

PKC occurs throughout the animal kingdom, has a broad tissue distribution (Kuo et al., 1980; Blumberg et al., 1983) and acts on a variety of substrates (Nishizuka, 1986). Recent cDNA cloning experiments demonstrated the existence of a family of related PKC molecules (Coussens et al., 1986; Knope et al., 1986) which are likely to function in diverse cellular processes previously attributed to a single PKC.

To investigate the role of PKC in cellular signalling pathways in both cultured cells and during development of intact organisms, we characterized cDNA and genomic clones coding for a *Drosophila* protein with extensive homology to mammalian PKC.
<table>
<thead>
<tr>
<th>nucleotides</th>
<th>amino acids</th>
<th>dPKC gene differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1702</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1312-3243</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Combined nucleotide sequence of two overlapping cDNAs from phase 8 (nucleotides 1–1702) and 101.1 (nucleotides 1312–3243) and predicted amino acid sequence of Drosophila protein kinase C. Introns positions are marked by numbered arrows, translation termination codons are boxed, the putative polyadenylation site is underlined and cysteines are darkly shaded. The putative ATP binding site is marked with asterisks (*), the repeated cysteine structure is lightly shaded, and differences found in the dPKC gene are written below (nucleotides) and above (amino acids) the cDNA sequences.
Documents
identified as part of a *Drosophila* PKC (dPKC) gene and were therefore used as probes for screening cDNA libraries.

Using the genomic EcoRI E-5 and E-8 fragments (Figure 5) as probes in high stringency hybridization experiments, three λ gt10 cDNA libraries representing RNAs from 0-3 h embryo, 5.5-7.5 day early pupal and adult *Drosophila* strain Oregon R (a gift from Dr Thomas Kornberg) (Poole et al., 1985) were screened. Out of 4 × 10^6 phage analyzed (~1.3 × 10^6 phage per library), 12 positives were obtained (seven from male adult, two from 5.5-7.5 day early pupal, and three from early embryo). The two clones having the largest cDNA inserts, phage 101.1 (~1.9 kb, hybridized only with probe E-8), and phage 8 (~1.7 kb, hybridized with both E-5 and E-8 probes) were subcloned and sequenced (Figure 1).

The two cDNAs were found to overlap and together span 3243 nucleotides. A polyadenylation signal which was not used by the cloned cDNA can be found between nucleotides 3159 and 3164. A single open reading frame contained in the combined sequence extends from nucleotide 887 through 2864, encoding a 657 amino acid, 75-kd protein. The putative dPKC translation initiation site ACAATGG is preceded by termination codons in all three reading frames. The triplet ATG appears 10 times in the 5' untranslated region, implying that either internal initiation, termination and reinitiation, or relaxed scanning of ATG sites by the 40S ribosomal subunit is required for generation of the correct protein (Kozak, 1986). dPKC shows 64% identity with bPKCα, 62% identity with bPKCβ, 58% identity with bPKCγ and similar homology with the human and rat PKC family (Figure 2). The amino terminal cysteine-rich repeat domain, which was identified as a conserved feature of mammalian PKC variants (Coussens et al., 1986) is also present in dPKC and shows about 77% identity with mammalian PKC. The carboxy terminal putative kinase domain (aa ~ 340-610) shows ~ 88% identity with mammalian PKC. The cysteine-rich amino terminus region forms a perfect repeat of the structure Cys-X2-Cys-X7-Cys-X2-Cys-X7-Cys-X7-Cys at amino acids 59-95 and at amino acids 124-160 (Figure 1) and could form a metal binding configuration (Figure 3a; Miller et al., 1985; Rosenberg et al., 1985; Vincent et al., 1985; Berg, 1986). Similar structures have been observed in a variety of nucleic acid binding proteins, including steroid receptors, transcription factors and viral DNA binding proteins (Figure 3b; Miller et al., 1985; Weinberger et al., 1985; Greene et al., 1986; Berg, 1986). The existence in PKC of a putative DNA binding domain raises the intriguing possibility of direct and sequence-specific binding of PKC to DNA, enabling phosphorylation of adjacent, already bound factors (i.e. histones, high mobility group proteins, topoisomerases, transcription activators and suppressors) with subsequent modulation of gene expression. Alternatively, the cysteine-rich repeat region might be involved in calcium binding, interaction with other proteins or in formation of the PKC tertiary structure.

The only non-homologous regions span amino acids 1-30 (V1 region), 174-191 (V2 region), 300-340 (V3 region) and 637-657 (V4 region; Coussens et al., 1986) (Figure 2). These regions are also variable in all mammalian PKC variants (Coussens et al., 1986) and might serve as spacer regions between func-
Structure and sequence of Drosophila PKC gene

Fig. 4. Homology between Drosophila protein kinase C (dPKC) and other serine/threonine/tyrosine kinases. Homology between dPKC and bovine cAMP-dependent protein kinase (cAMPc), bovine cGMP-dependent protein kinase (cGMP), rabbit skeletal muscle phosphorylase kinase (phos), human epidermal growth factor receptor (c-erbB), human insulin receptor (HIR). (Amino acid sequences of cAMPc, cGMP, phos, v-src, c-erbB, v-mos and HIR were compiled by the National Biomedical Research Foundation, Georgetown University Medical Center, Washington, DC 20007.)

Fig. 5. Structure of Drosophila protein kinase C gene. Genomic phages are numbered 8, 2, 4, 10. EcoRI fragments are numbered E-1 to E-8 and their respective sizes are written above. E = EcoRI, B = BamHI, C = Cla, H = HindIII.
Fig. 6. Nucleotide sequence and predicted amino acids of the Drosophila protein kinase C gene. Untranslated exons are lightly shaded. Coding exons are darkly shaded. Introns are unshaded. Translation termination codons are boxed, putative polyadenylation site is underlined, putative ATP binding site is marked with asterisks (*).
tional domains. Alternatively, the variable regions might be involved in determining biological specificity and may be sites of interaction with substrate or agonist molecules. In addition, dpKC is 10 amino acids longer at the amino terminus and 35 amino acids shorter at the carboxyl terminus than mammalian PKCs. 

dpKC homology with the kinase domain of other serine, threonine and tyrosine kinases (Figure 4) is highest in a region between amino acids 350 and 585. This region, which contains a consensus sequence for an ATP binding site (amino acids 345–374) (Figure 1) (Hannink and Donoghue, 1985) and is the most conserved region between Drosophila and mammalian PKCs. The fact that removal of the amino terminus results in a constitutively active kinase independent of calcium and phospholipids (Inoue et al., 1977) implicates the amino terminus as a calcium, 1,2-DG and phospholipid binding domain, and indicates that the kinase domain is masked when the protein is inactive. No obvious hydrophobic phospholipid binding sites can be identified in this amino terminal regulatory domain, and although calcium probably binds to the protein directly at multiple sites (Wolf et al., 1985), no homology with putative calcium binding sites from other proteins could be found. The calcium binding site suggested for bPKCa around amino acids 300–310 (Parker et al., 1986) is not conserved in Drosophila.

Structure of the Drosophila protein kinase C gene
In order to identify further putative functional domains (Gilbert, 1978) in dpKC, its gene structure was determined. Genomic phase 2, 4, 8 and 10, initially isolated using the bPKCa probe (see previous section), were mapped using restriction enzymes EcoRI, HindIII, BamHI and Clal. Overlapping regions were determined and eight EcoRI fragments (E-1 to E-8) spanning 32.6 kb of the dpKC gene (Figure 5) were subcloned. Regions that hybridized with the dpKC cDNA were subcloned into M13 vectors and sequenced. The dpKC spans ~20 kb and contains 13 coding exons and at least one untranslated exon at its 5′ end. Exon sizes range from 32 bp to at least 623 bp and intron sizes range from 54 bp to >8000 bp (Figures 5 and 6).

The kinase domain spans at least exons 8, 9, 10, 11, 12 and 13. Homology with other protein kinases starts within the 3′ boundary of exon 8 and ends within exon 13 (Figure 4). Exon 8 entirely contains the region between amino acids 300 and 340, which is variable between the members of the PKC family (V3 region, Coussens et al., 1986) and might define a substrate specificity determining domain. Exon 14 starts exactly where dpKC diverges from mammalian PKCs (Figure 2). The other two regions where dpKC diverges from its mammalian homologs are contained within exon 2 (amino acids 1–30, V1 region) and exon 6 (amino acids 174–189, V2 region). The cysteine-rich repeat region, putative metal and DNA binding regions (amino acids 59–95 and 124–160) are included within exons 2–6. The first repeat is included within exons 2–4 and the second repeat is included within exons 5 and 6. The two repeats are separated by intron 4 and each repeat is split by an intron at the same place, the second amino acid of the second Cys-X2-Cys structure (amino acids 78 and 143), giving the structure Cys-X2-Cys-X13-Cys-X-intron-X-Cys-X7-Cys-X7-Cys (Figures 1 and 6).

The general relation of exon—intron structure to functional domains in dpKC is not always clear (intron 8 splits the putative ATP binding site) and might reflect either disruption of domain structure in evolution by acquisition of introns or our ignorance in defining functional domains. Parker et al. (1986) suggested that the PKC gene, like the cGMP-dependent protein kinase, was derived from a fusion of two genes providing regulatory and catalytic domains. If so, this must have occurred at least 800 million years ago and is not clearly reflected in the exon—intron structure. Looking at the biggest ~8 kb intron (intron 5) as a boundary between domains enables one to include the kinase part
Comparison of nucleotide sequences between dPKC cDNA (Oregon R strain) and genome (Canton S strain) revealed 16 differences in the coding region. Fifteen of them are silent and do not result in amino acid changes, and one (amino acid 428) results in a change from isoleucine (cDNA) to methionine (genome) (Figure 1). All six mammalian protein kinases have methionine at this position (Figure 2). In addition there are three insertion/deletion changes of single nucleotides around poly(A) stretches in the 5'-untranslated region.

To determine cytogenic localization of the dPKC gene, biotinylated full-length cDNA and a genomic fragment (E-8) were used as probes for in situ hybridization to polytene chromosomes of larval salivary glands. In parallel experiments, both probes hybridized to the same single site which was identified as position 53E on chromosome 2 (Figure 7).

Expression of dPKC mRNA

Using dPKC cDNA as a probe, the size and level of mRNA for PKC was measured in 0–3 h embryo and adult flies (Oregon R strain). Northern blot analysis revealed the existence of three equally abundant mRNAs of ~4.3, 4 and 2.4 kb in adult fly tissues. No expression of dPKC RNA could be detected in 0–3 h embryo (Figure 8a). Actin mRNA could be detected in the same blot using an 8.5-kb actin (SC) gene fragment as a probe (Fryberg et al., 1983) in both 0–3 h embryo and adult fly (Figure 8b). The existence of three PKC mRNAs might reflect alternative initiation, termination or splicing events.

The low levels of mRNA in 0–3 h embryo might reflect either the utilization of maternal proteins at this time of development, the use of another PKC gene which was not detected under our hybridization conditions, or indicate that dPKC is required only at later developmental stages. More detailed analysis of dPKC expression patterns during development of both normal and mutated flies will help reveal the protein's function in a whole organism.

Materials and methods

Screening of genomic and cDNA library

E. coli DH5α strain (for Charon 4 phage) and C600 HFL strain (for-lg10 phage) were infected with recombinant phage and replica plated on nitrocellulose filters (30–100 × 105 phage per filter). Filters were baked at 80°C for 2 h, prehybridized for 3–4 h and then transferred to hybridization solution containing 50% formamide, 5 × SSC, 50 mM sodium phosphate pH 6.8, 0.1% sodium pyrophosphate, 5 × Denhardt's, 50 μg/ml salmon sperm at 42°C. DNA probes were labeled with 32P using synthetic random oligomers as primers in DNA polymerase extension reaction and then added to the hybridization solution. Filters were hybridized for 12–16 h before washing in 0.2 × SSC, 0.1% SDS at the same temperature. For low stringency hybridization with non-homologous probes, 30% formamide was used in the hybridization solution and filters were washed as before.

DNA sequencing

DNA sequencing was done according to the standard dideoxy chain termination method following subcloning into M13 derivatives (Smith, 1980). All sequences were determined at least twice.

Northern blot analysis

Twenty μg of polyadenylated RNA from 0–3 h embryo and adult Drosophila Oregon R (a gift from Dr Y. Yarden) were electrophoresed onto a formaldehyde–1.2% agarose gel, and blotted onto nitrocellulose. The nitrocellulose filters were hybridized with 32P-labeled dPKC cDNA probe in 50% formamide, 5 × SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate (pH 6.8), 2 × Denhardt's solution, 10% dextran sulfate at 42°C for 15–20 h. Extensive washings were done in 0.1 × SSC, 0.1% SDS at 60°C.

Received on 20 October 1986; revised on 28 November 1986

Note added in proof

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00042.