The EMBO Journal vol.12 no.5 pp.1853-1870, 1993

Generation of structural and functional diversity in furin-like proteins in *Drosophila melanogaster* by alternative splicing of the *Dfur1* gene

Anton J.M.Roebroek[1], John W.M.Creemers[1], Ilse G.L.Pauli[1], Thierry Bogaert[2] and Wim J.M.Van de Ven[1,3]

[1]Laboratory for Molecular Oncology, Center for Human Genetics, University of Leuven, Herestraat 49, B-3000 Leuven, and [2]Laboratory for Molecular Biology, Celgen, University of Leuven, Herestraat 49, B-3000 Leuven, Belgium

[3]Corresponding author

Communicated by H.Bloemendal

To investigate whether or not alternative splicing might be a mechanism by which in *Drosophila melanogaster* diversity is generated in endoproteases of the novel eukaryotic family of subtilisin-like proprotein processing enzymes, we determined structural and functional characteristics of the *Dfur1* gene. Northern blot analysis revealed *Dfur1* transcripts of 7.6, 6.5, 4.5 and 4.0 kb. By comparative nucleotide sequence analysis of *Dfur1* genomic and cDNA clones, 10 coding exons were identified and, together with Northern blot analysis using exon-specific probes, evidence was obtained that the four transcripts are generated by alternative splicing and polyadenylation. The apparently complete open reading frames of three *Dfur1* cDNAs revealed that these coded for three furin-like proteins, Dfurin1 (892 residues), Dfurin1-CRR (1101 residues) and Dfurin1-X (1269 residues), which possessed common but also unique structural domains. These various isoforms of furin in *Drosophila* were characterized in gene transfer studies using immunoprecipitation analysis. Differential expression of *Dfur1* transcripts was found in Northern blot analysis of RNA from various developmental stages of *Drosophila*. RNA *in situ* hybridization experiments revealed that the Dfurin1-X and Dfurin1-CRR isoforms are expressed in non-overlapping sets of tissues during *Drosophila* embryogenesis. In gene transfer experiments in which the Dfurin1, Dfurin1-CRR and Dfurin1-X proteins were expressed at high levels together with the precursor of the β_A-chain of activin-A, a member of the transforming growth factor β (TGFβ) superfamily, or the precursor of von Willebrand factor, all three proteins appeared capable of processing these substrates. Our studies indicate that the *Dfur1* gene encodes structurally different subtilisin-like proprotein processing enzymes with distinct physiological functions in *Drosophila*.

Key words: Drosophila melanogaster Dfur1/furin isoforms/proprotein processing/subtilisin-like proteins

Introduction

In eukaryotes, post-translational endoproteolytic processing is often a crucial step in the release of bioactive polypeptides from higher molecular weight precursors. Initial endoproteolysis of proproteins generally occurs at cleavage sites consisting of particular sequence motifs of basic amino acids, frequently paired basic residues such as Lys-Arg or Arg-Arg. The first indications for the existence of enzymes with such specificity were obtained some time ago (Chrétien and Li, 1967; Steiner and Oyer, 1967). The physiological importance of these enzymes is evident now since a wide variety of proteins such as neuropeptides, peptide hormones, growth factors, growth factor receptors, particular plasma proteins and envelope glycoproteins of viral pathogens apparently require this type of endoproteolytic cleavage for their production (Docherty and Steiner, 1982; Douglas *et al.*, 1984; Fuller *et al.*, 1988; Sossin *et al.*, 1989). The first known proprotein processing enzyme in eukaryotes with such cleavage specificity is the kexin enzyme (EC 3.4.21.61) of the yeast *Saccharomyces cerevisiae* (Fuller *et al.*, 1988). Encoded by the *KEX2* gene, this enzyme is a membrane-associated, Ca^{2+}-dependent subtilisin-like serine endoprotease with an established cleavage selectivity for paired basic amino acid residues, a neutral pH optimum and a relatively broad substrate specificity. During the past few years, various kexin-like proprotein processing enzymes have been identified (for a review see Van de Ven *et al.*, 1993). Human furin, which is encoded by the widely expressed *fur* gene (Schalken *et al.*, 1987; Barr *et al.*, 1991), is the first known mammalian member of this novel family of processing enzymes. Other mammalian members are the two neuroendocrine-specific enzymes PC1 (Siedah *et al.*, 1990, 1991, 1992a; Creemers *et al.*, 1992) [also called PC3 (Smeekens *et al.*, 1991) and, therefore, referred to hereafter as PC1/PC3] and PC2 (Seidah *et al.*, 1990, 1991; Smeekens and Steiner, 1990) and presumably also the product of the broadly expressed *PACE4* gene (Kiefer *et al.*, 1991), which maps on the long arm of human chromosome 15 at a position cytogenetically close to the *fur* gene, and the testis-specific enzyme encoded by the *PC4* gene (Nakayama *et al.*, 1992; Seidah *et al.*, 1992b). The *fur* gene was initially discovered in the region immediately upstream of the *fes* proto-oncogene (Roebroek *et al.*, 1986a,b) and, later on, structural similarities between its product furin and the kexin enzyme of yeast were noted (Fuller *et al.*, 1989; Van den Ouweland *et al.*, 1990). Experimental evidence that furin is a proprotein processing enzyme with cleavage specificity for paired basic amino acid residues was provided when it was demonstrated that furin was capable of correctly processing proproteins such as the precursors of the von Willebrand factor (pro-vWF) (Van de Ven *et al.*, 1990; Wise *et al.*, 1990) and β-nerve growth factor (pro-β-NGF) (Bresnahan *et al.*, 1990), proalbumin (Misumi *et al.*, 1991), the precursor to complement factor C3 (Misumi *et al.*, 1991) and mutants of prorenin (Hosaka *et al.*, 1991; Nagahama *et al.*, 1991; Watanabe *et al.*, 1992). It is assumed that for proprotein cleavage by furin, the consensus motif Arg-X-Lys/Arg-Arg is required (Hosaka *et al.*, 1991).

In mammalian furin, a number of protein domains have

been predicted (Van de Ven *et al.*, 1990) and these include a 'prepro' domain, a subtilisin-like catalytic domain which is preceded by a potential (auto)proteolytic cleavage site (K-R-R-T-K-R^{107}), a 'middle' domain immediately downstream of the catalytic domain, a cysteine-rich region consisting of two cysteine-rich repeats (Roebroek *et al.*, 1990), a potential transmembrane domain and a cytoplasmic domain. The amino acid sequences of the subtilisin-like catalytic domains of the various enzymes that are presently known are highly similar, especially in the regions flanking the catalytic triad and the oxyanion hole (Van de Ven *et al.*, 1992). Sequences of the 'middle' domain of furin are of importance for catalytic activity (Hatsuzawa *et al.*, 1992) and, therefore, these seem to correspond to the 'P' domain of kexin (Fuller *et al.*, 1991). The role of the cysteine-rich region is still unknown, but this region is not unique for furin since the cysteine-rich repeats are also found in other members of this novel enzyme family; for instance, in human PACE4 (Kiefer *et al.*, 1991), and in two proprotein processing enzymes of *Drosophila melanogaster*, dKLIP-1 (Hayflick *et al.*, 1992) and Dfurin2 (Roebroek *et al.*, 1992). Finally, most of these subtilisin-like proteins are assumed to be associated with membranes, either through hydrophobic transmembrane domains in their carboxy-terminal regions, as is the case for kexin, furin and *Drosophila* Dfurin1, Dfurin2 and dKLILP-1, or by carboxy-terminal amphipathic α-helical structures, as suggested for PC1/PC3 and PC2 (Smeekens *et al.*, 1991). Of special interest in this context are the three *Drosophila* proteins, since they all possess a putative transmembrane domain with an internal signal peptide function in their amino-terminal regions (Roebroek *et al.*, 1991, 1992; Hayflick *et al.*, 1992).

In previous studies, we identified two *fur*-like genes in *D.melanogaster*: *Dfur1* and *Dfur2* (Roebroek *et al.*, 1991, 1992). Our studies on *fur*-like genes in species distantly related to mammals were initiated in an attempt to gather more information about relevant features of the protein domains of furin and other members of this novel family of subtilisin-like processing enzymes. A cDNA of the *Dfur2* gene was recently characterized in detail (Roebroek *et al.*, 1992) and, as a result, a repeated motif of cysteine residues could be recognized in the large cysteine-rich region of its product Dfurin2. This cysteine motif appeared to be conserved in the smaller cysteine-rich regions of furin, PACE4 and *Drosophila* dKLIP-1. As far as the *Dfur1* gene is concerned, we have only described characteristics of two overlapping cDNAs and features in the Dfurin1 protein that could be deduced from its sequences (Roebroek *et al.*, 1991). In those studies, we found indications that expression of the *Dfur1* gene might result in the synthesis of multiple transcripts. The possible interrelationship between these transcripts raised the question as to whether *Dfur1* expression in *Drosophila* could result in the generation of structurally and functionally distinct translational products. To investigate this, we have performed detailed structural and functional studies of this *Drosophila* gene. For the functional studies, we used the SV40/COS-1 expression system and the substrates pro-vWF and a cleavage mutant thereof, as previously applied to assay potential proprotein processing activity of human and mouse furin (Van de Ven *et al.*, 1990; Van Duijnhoven *et al.*, 1992) and *Drosophila* Dfurin2 (Roebroek *et al.*, 1992). Furthermore, we used a vaccinia virus expression system in combination with proβ_A as substrate. Proβ_A (397 amino acids) is the precursor of the β_A-chain (116 amino acids) of the fertility hormone inhibin and of activin-A, which is also known as follicle stimulating hormone (FSH)-releasing protein (Ying, 1988). The two β-chains are known (β_A and β_B) and in combination with another polypeptide chain, the so-called α-chain, they form inhibin-A and -B. Activins are disulfide-linked homodimers of the β-chain; the $\beta_A\beta_A$ dimer (25 kDa) is activin-A. In a variety of cell lines such as CV-1, HeLa S3, HeLa D98/AH2, AtT-20 and Hep3B (Huylebroeck *et al.*, 1990), proβ_A is cleaved after amino acid residue 281 into a 'PRO' fragment (281 amino acids) and the mature β_A-chain (Huylebroeck *et al.*, 1990). At the cleavage site, multiple pairs of basic amino acid residues are present [R-R-R-R-R^{281} (D.Huylebroeck, personal communication)] and these also contain the consensus motif R-X-K/R-R for cleavage by furin (Hosaka *et al.*, 1991). We used PK(15) cells in our

Table I. *Dfur1* probes used in hybridization experiments

Probe name	Origin	Nucleotides[a]	Exons present[b]
pG6	cDNA[c]	609–1495	2 (9 bp), 3, 4
pG7	cDNA[c]	1496–3038	5, 6, 7, 8, 9
		4170–4336	11
		2598–2618[c]	12 (21 bp)[c]
		4337–4378	12 (42 bp)
pG13	cDNA[c]	609–2743	2 (9 bp), 3, 4, 5, 6, 7
pIP44	Genomic	1.2 kbp fragment *Xba*I–*Hin*dIII 5734–6958	Non-coding part of 12
pIP46	Genomic	1.5 kbp fragment *Hin*dIII–*Eco*RI	Part of intron 12, 13, and fragment downstream of 13
pIP95	Genomic	0.9 kbp fragment *Hin*dIII–*Kpn*I	Part of intron 12, major part of exon 13
PCR1	PCR	3125–3967	10 (843 bp)
PCR2	PCR	4615–4952	Non-coding part of 12
PCR3	PCR	5208–5756	Non-coding part of 12
PCR4	PCR	67–636	1 (486 bp), 2, 3 (19 bp)
PCR5	PCR	536–636	1 (17 bp), 2, 3 (19 bp)

[a]Nucleotide numbers correspond to numbers of Figure 2, except for a 21 bp fragment of probe pG7 which is numbered as before (Roebroek *et al.*, 1991).

[b]Exon numbers correspond to numbers in Figure 4.

[c]pG6, pG7 and pG13 are derived from a previously described cDNA (Roebroek *et al.*, 1991).

studies since processing of proβ_A in these cells is very inefficient (96% proβ_A and 4% β_A) as compared for instance to AtT-20 cells (0–2% proβ_A and 98–100% β_A); PK(15) cells transfected with proβ_A DNA almost exclusively secrete proβ_A (Huylebroeck *et al.*, 1990).

In this report, we describe: (i) the isolation and characterization of three *Dfur1* cDNAs apparently containing the complete coding sequences for three proteins, which were designated Dfurin1, Dfurin1-CRR and Dfurin1-X; these proteins appear to have common but also unique structural features; (ii) the genomic distribution of the coding *Dfur1* exons, which was established on the basis of nucleotide sequence analysis of *Dfur1* genomic and cDNA clones; the data indicate that the transcripts from which the cDNAs are derived are generated by alternative splicing and polyadenylation; (iii) differential expression of *Dfur1* transcripts during development of *Drosophila* as established on the basis of Northern blot analysis; (iv) topologically different expression patterns of Dfurin1-CRR and Dfurin1-X mRNAs as observed in *in situ* hybridization analysis of *Drosophila* embryos; (v) characterization of the various *Dfur1*-encoded furin isoforms in *Drosophila* using a gene transfer approach and immunoprecipitation analysis with anti-Dfurin1 antiserum; (vi) proprotein processing activity of the various *Dfur1*-encoded proteins using two distinct heterologous proprotein processing assays.

Results

Isolation of three different Dfur1 cDNAs with complete large open reading frames

In earlier studies (Roebroek *et al.*, 1991), we found indications that expression of the *Dfur1* gene might result in the generation of multiple transcripts. In Northern blot analysis of RNA from *D.melanogaster* with pG6 or pG7 (Table I) as molecular probe, transcripts of 7.6, 6.5, 4.5 and 4.0 kb were detected (data shown in Figures 6 and 8 below, in the context of other experiments). It should be noted that the transcript of 7.6 kb was not detected in our previous studies in which we used RNA from 2–8 h embryos; in this developmental stage, expression levels of this transcript are very low. To define these transcripts, an attempt was made to isolate corresponding cDNAs. To accomplish this, a *Drosophila* cDNA library was screened with pG6, pG7, PCR1, pIP44 and pIP46, five different, non-overlapping *Dfur1* probes; some of these probes were selected because Northern blot analysis revealed that they were discriminative for one or two of the transcripts. From a large number of positive clones that were identified, only those hybridizing to more than one probe were selected for further studies. In this way, four positive clones hybridizing only to pG6 and pG7, two positive clones hybridizing to pG6, pG7 and PCR1, two positive clones hybridizing to pG6, pG7 and pIP46 were selected. The two positive clones hybridizing to pIP44 did not hybridize to any of the other probes. After restriction enzyme analysis of the cDNA inserts and subsequent nucleotide sequence analysis of the 5′- and 3′-ends of these, the positive clone of each category with the largest cDNA insert was selected for complete nucleotide sequence analysis: these were pIP56, pIP58, pIP62 and pIP63. In Figure 1, relevant data of the cDNA inserts of these clones and the degree of overlap are schematically summarized. Clone pIP56 appeared to hybridize to pG6, pG7 and pIP46; clone pIP58 hybridized only to pIP44; clone pIP62 hybridized to pG6, pG7 and PCR1, and clone pIP63 only to pG6 and pG7. The nucleotide

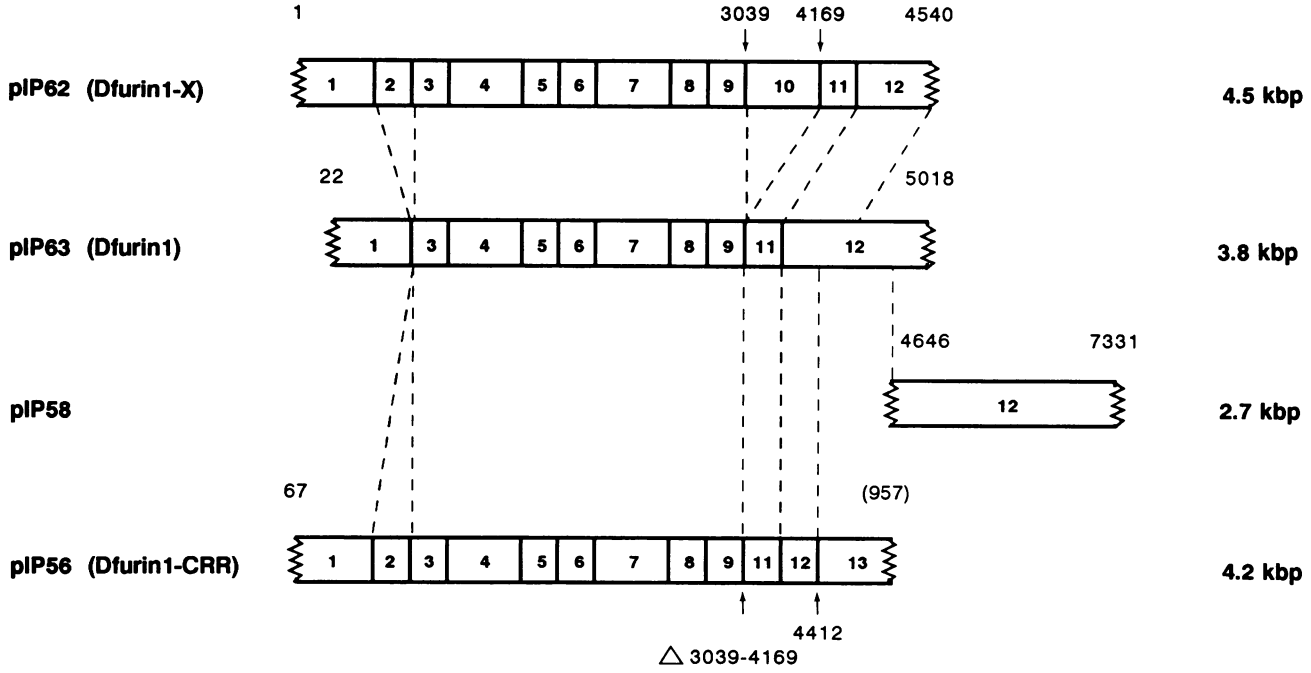

Fig. 1. Schematic representation of *Dfur1* cDNA inserts of clones pIP56, pIP58, pIP62 and pIP63, which were isolated from a cDNA library of *D.melanogaster*. The size of the cDNA inserts is indicated. The *Dfur1* exons present in the four cDNAs are schematically indicated by their number in the boxes. The size of the boxes does not correspond to the size of the exons. The numbers 3039 and 4169 with arrows indicate the first and last nucleotide of exon 10; number 4412 with arrow indicates the last nucleotide of exon 12 in pIP56; the deletion of exon 10 in pIP56 and pIP63 is indicated by Δ3039–4169; 957 nucleotides of exon 13 are present in the cDNA insert of pIP56 and this is indicated by that number in parentheses; the other numbers above the boxes correspond to the first and last nucleotides of the *Dfur1* cDNA inserts as numbered in Figure 2. The names of the deduced *Dfur1*-encoded proteins to which the various inserts correspond are also indicated.

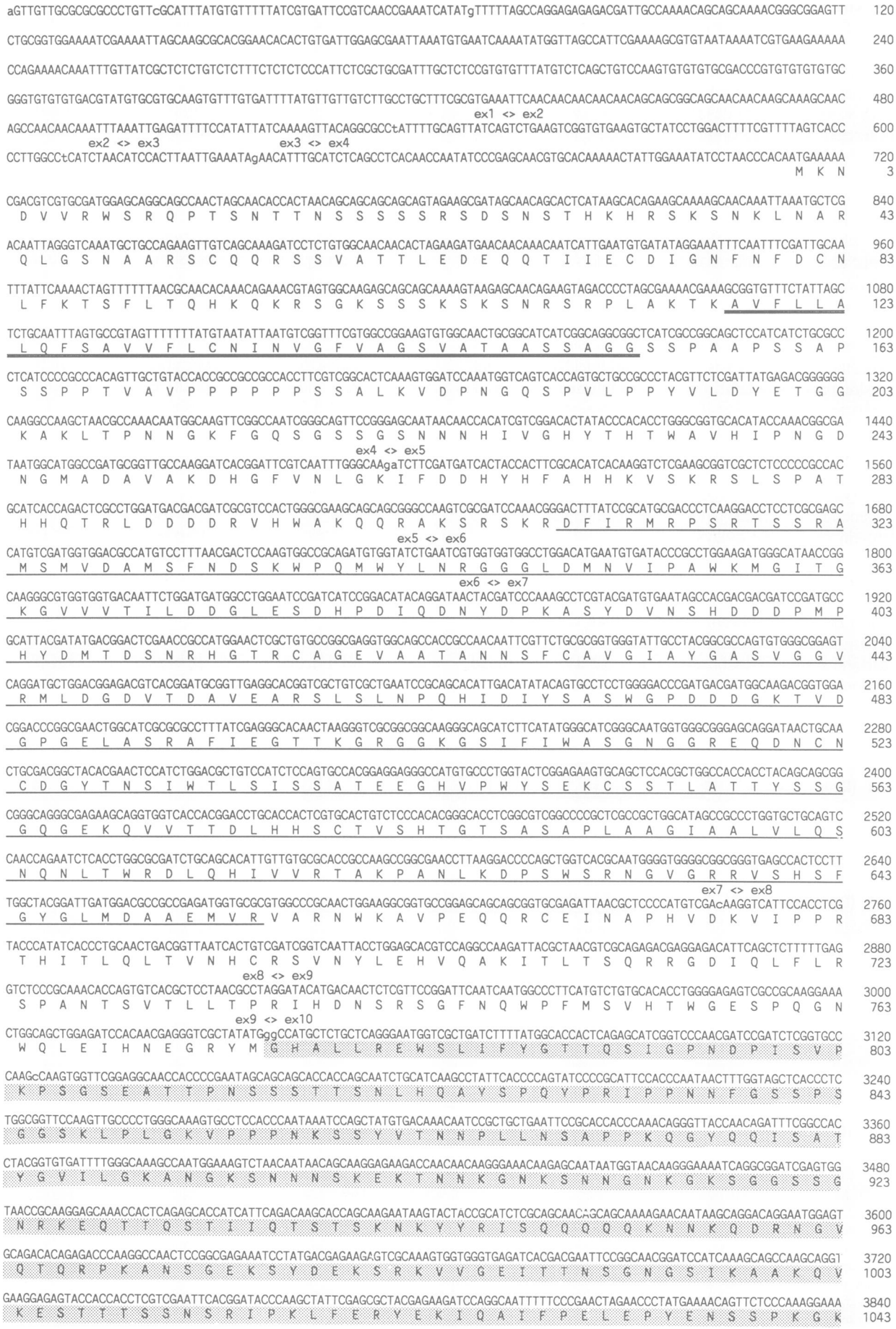
ex1 <> ex2
ex2 <> ex3
ex3 <> ex4
ex4 <> ex5
ex5 <> ex6
ex6 <> ex7
ex7 <> ex8
ex8 <> ex9
ex9 <> ex10

Fig. 2. Nucleotide sequence and predicted amino acid sequence encoded by a composite *Dfur1* cDNA corresponding to the transcript of 7.6 kb. The sequence is a composition of sequence data of pIP56, pIP58, pIP62 and pIP63 (for an explanation see the text). Nucleotides at positions mentioned in Table I or Figure 1 are given in lower case instead of upper case. Numbering of the amino acid residues and the nucleotides is indicated at the end of each line. Boundaries of exons as could be deduced from comparison of cDNA nucleotide sequence data to genomic nucleotide sequence data are indicated above the sequence (< and >); the exons are numbered ex1−ex12. Potential transmembrane domains in the amino- and carboxy-terminal regions of the deduced protein are underlined by a double line. The presumptive subtilisin-like catalytic domain is underlined by a single line; essential amino acid residues of its active site are D (372), H (413) and S (587), and the oxyanion hole N (514). The protein domain encoded by exon 10, domain X (for an explanation see the text), is represented in a shaded background. An internal 5′-splice site in exon 12 is indicated above the sequence (ex12 <), internal polyadenylation sites in exon 12 are indicated underneath the sequence (^^). Two potential polyadenylation signals are underlined by a broken line.

sequence of the newly isolated *Dfur1* cDNAs is given in Figures 2 and 3. It should be noted that Figure 2 shows a composite sequence based on nucleotide sequence data of all four cDNA clones and that it represents the sequence of the large 7.6 kb transcript, as will be discussed below. The nucleotide sequence of the cDNA insert of clone pIP62 (insert size ~4.5 kbp) starts at nucleotide 1 and ends at nucleotide 4540; that of the cDNA insert of clone pIP63 (insert size ~3.8 kbp) starts at nucleotide 22 and ends at nucleotide 5018. It should be noted that nucleotides 553−617 (exon 2) and 3039−4169 (exon 10) are not present in clone pIP63. The nucleotide sequence of the cDNA insert of clone pIP58 (insert size ~2.7 kbp) starts at nucleotide 4646 and ends at nucleotide 7331. The nucleotide sequence of the cDNA insert of clone pIP56 (insert size ~4.2 kbp) includes nucleotides 67−4412 (Figure 2), except nucleotides

Fig. 3. Nucleotide sequence and predicted amino acid sequence encoded by the 3′-end of *Dfur1* cDNA pIP56, which corresponds to the 4.5 kb transcript (for an explanation see the text). Sequence starts within exon 11. Numbering of the amino acid residues and the nucleotides is indicated at the end of each line. Boundaries of exons as could be deduced from comparison of cDNA nucleotide sequence data to genomic nucleotide sequence data are indicated above the sequence (< and >); the exons are numbered ex11−ex13. A potential transmembrane domain in the carboxy-terminal region of the deduced protein is underlined by a double line. The two subdomains of the cysteine-rich region are represented in a shaded background. The sequence data reported have been deposited in the EMBL/GenBank/DDBJ data libraries under the accession numbers L12368 and L12377.

3039−4169 (exon 10). It also contains 957 nucleotides of exon 13, so after nucleotide 4412 of Figure 2 (internal 5′-splice site in exon 12), the pIP56 sequence continues in Figure 3 with nucleotide 3216 (start exon 13). These results indicate that the four cDNAs are most likely derived from at least three different *Dfur1* transcripts. Comparison of the exon compositions of the *Dfur1* cDNAs and the mRNAs should resolve as to which of the four transcripts that were detected in Northern blot analysis each of the cDNAs corresponds (see below).

Analysis of the nucleotide sequences of the cDNAs revealed large open reading frames in the inserts of three of the clones (pIP56, pIP62 and pIP63), but not in clone pIP58. The large open reading frames appeared to encode three different proteins as discussed below. The deduced amino acid sequences for these proteins are given below the corresponding cDNA sequences in Figures 2 and 3.

Identification of Dfur1 exons by comparative analysis of genomic and cDNA clones

To obtain insight into the genomic distribution of the exon sequences present in the various *Dfur1* cDNAs, the available *Dfur1* genomic clones, λYZ3, λYZ4 and λYZ5, were analyzed. Initial hybridization studies revealed the presence of exon sequences downstream of a *Bgl*II restriction site (see Figure 4). From a region of ~11 kbp, starting somewhat downstream of this *Bgl*II site and ending at the 3′-end of λYZ3, the nucleotide sequence was determined (data not shown). Comparison of these with the nucleotide sequence data of the *Dfur1* cDNAs resulted in the identification of the genomic positions and the complexity of nine exons (exons 5−13) (Figure 4). The 5′-end exon sequences corresponding to nucleotides 1−1495 were apparently not present in λYZ3. To determine the genomic position and the complexity of these, a *Drosophila* genomic library was screened with probes pG6 and PCR4. This resulted in the isolation of several positive clones; of these, λDg9, which hybridized to pG6, and λDg21 and λDg22, which hybridized to PCR4, were selected for further analysis. Relatively small DNA fragments of these genomic clones which appeared to hybridize to the respective probes were subcloned and their nucleotide sequences were determined. From these studies, it appeared that λDg21 and λDg22, which are overlapping clones, contained exon 1; exon 1 consists of nucleotides 1−552 and is flanked by a 5′-splice site consensus sequence (Figures 2 and 4A). λDg9, which does not overlap with λDg21 and λDg22, appeared to contain exon 4; exon 4 consists of nucleotides 645−1495 and is flanked by 3′- and 5′-splice site consensus sequences (Figures 2 and 4A). λDg9 also does not overlap with the genomic clones λYZ3, λYZ4 and λYZ5 (Figures 2 and 4A). The exon sequences consisting of nucleotides 553−644 are believed to be non-contiguous in the *Drosophila* genome and to belong to two exons (putative exon 2 and exon 3). We conclude this from the fact that nucleotides 553−617 (putative exon 2) are spliced out from the cDNA insert of pIP63 and, furthermore, from the observation that at the junction of the two putative exons in the cDNAs (Figure 2) no alternative 3′-splice site (C/TAG) is present. The genomic positions of nucleotides 553−617 (exon 2) and of nucleotides 618−644 (exon 3) are still unknown. Exon 2 sequences could not be identified in λDg21 and λDg22 in hybridization studies with the PCR5 probe, which contains only 17 bp of exon 1, complete exon 2 (65 bp) and 19 bp of exon 3. They were also not found in any of the other exon 1-containing genomic clones that were isolated and that hybridized to probe PCR4, which contains 496 bp of exon 1, complete exon 2 and 19 bp of exon 3. Exon 2 is 65 nucleotides long, certainly large enough for detection by hybridization with PCR4 or PCR5. Exon 3, which is only 27 bp in size, might have escaped detection by hybridization analysis under the conditions used. Therefore, DNA sequencing experiments were performed. In these experiments, overlapping DNA subclones of λDg9 containing exon 4 and upstream sequences were used as template; the sequence primer that was used was designed on the basis of the known sequences of exon 3. In these experiments, no primer extension could be demonstrated, indicating that exon 3 was also not present in λDg9.

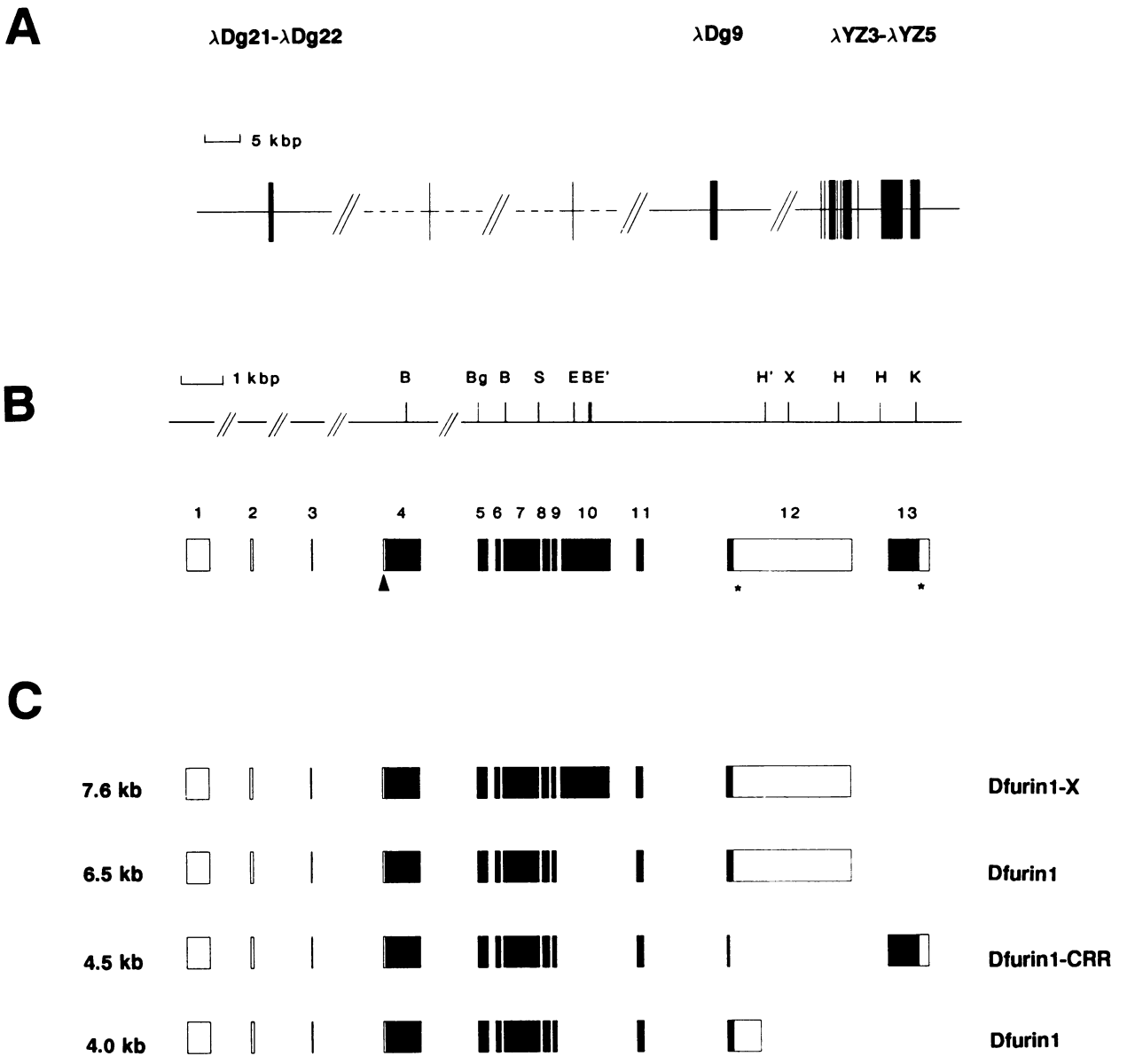

Fig. 4. (**A**) Schematic representation of the genetic organization of the *Dfur1* gene showing the distribution of exons in the gene. The relative positions of the exons in the non-contiguous genomic map are indicated by vertical bars. Gaps in the map are indicated (//), whereas isolated clones are represented by a horizontal solid line and as yet unknown but predicted fragments by a horizontal interrupted line. The relative size of 5 kbp is indicated. (**B**) Schematic restriction map of parts of the *Dfur1* gene showing the exon distribution in more detail. Gaps in the *Dfur1* gene are indicated (//). Restriction endonuclease sites included in the map are *Bam*HI (B), *Bgl*II (Bg), *Eco*RI (E), *Hin*dIII (H), *Kpn*I (K) and *Xba*I (X). E′ and H′ indicate, respectively, two E and two H sites close together. Underneath the restriction map, the relative size and position of the exons is shown. The exons are numbered, protein coding parts are represented as solid boxes, whereas non-coding parts are represented as open boxes. The position of the start codon (closed triangle) in exon 4 and the stop codons (asterisks) in exon 12 and 13 are indicated. The relative size of 1 kbp is shown. (**C**) Schematic representation of the exons present in the four different *Dfur1* transcripts generated by alternative splicing and alternative polyadenylation. The sizes of the transcripts and the *Dfur1* proteins they encode are indicated.

In Figure 4A and B, the deduced genomic organization of the *Dfur1* gene is schematically represented, with the exon/intron distribution of 13 exons. As will be discussed later, alternative splicing in combination with alternative polyadenylation is responsible for the generation of the four different transcripts that are schematically represented in Figure 4C. In Figure 5, the nucleotide sequences of junctions between exons and introns are shown, except for exon 2 and exon 3. These data indicate that all 5′- and 3′-splice sites possess the consensus splice site sequences; all introns start with GT and end with C/TAG.

The nucleotide sequence data of the cDNAs and the genomic DNA clones were obtained from different *D.melanogaster* strains. However, no nucleotide differences were observed within the exons. In contrast, nucleotide differences were found when the data were compared with previously reported *Dfur1* cDNA clones (Roebroek *et al.*, 1991; Hayflick *et al.*, 1992). Differences which do not result in amino acid substitutions are not discussed in detail here; they are considered to represent (sub)strain differences. One observed difference results in an amino acid substitution and will be discussed below.

```
exon1  <-----TTTTGCAGTTATCAGgtgagcttgcactta----->     intron1 (unknown size)   <-----nnnnnnnnnnnnnnnTCTGAAGTCGGTGTG-----> exon2
exon2  <-----TTGGCCTCATCTAACnnnnnnnnnnnnnnn----->     intron2 (unknown size)   <-----nnnnnnnnnnnnnnnATCCACTTAATTGAA-----> exon3
exon3  <-----GAAATAGAACATTTGnnnnnnnnnnnnnnn----->     intron3 (unknown size)   <-----atttttttaattgcagCATCTCAGCCTCACA-----> exon4
exon4  <-----GTCAATTTGGGCAAGgtaagaatcgtttaa----->     intron4 (unknown size)   <-----catcattccttgcagATCTTCGATGATCAC-----> exon5
exon5  <-----CAGATGTGGTATCTGgtaagttatcgaaaa----->     intron5 (170 bp)         <-----taaccttattctcagAATCGTGGTGGTGGC-----> exon6
exon6  <-----ATACAGGATAACTACgtaagtagtgacatt----->     intron6 (63 bp)          <-----tgataacctaaacagGATCCCAAAGCCTCG-----> exon7
exon7  <-----TCCCCATGTCGACAAgtaggagaagcaaat----->     intron7 (63 bp)          <-----ttctctactttccagGGTCATTCCACCTCG-----> exon8
exon8  <-----GCTCCTAACGCCTAGgtaagtggggcagat----->     intron8 (54 bp)          <-----tccctccatcaacagGATACATGACAACTC-----> exon9
exon9  <-----AGGGTCGCTATATGGgtaagtctaagaact----->     intron9 (145 bp)         <-----cccaccggcttgtagGCCATGCTCTGCTCA-----> exon10
exon10 <-----TCAACGGACCAAATGgtagccaaccaattt----->     intron10 (632 bp)        <-----tactctgcgttgcagCACAAATCACACAAT-----> exon11
exon11 <-----AGGAATATGCAGCAGgtaagtttcgcattt----->     intron11

                                          (1956 bp)   <-----cctatcttaatccagcaatatccttttccttccagGTGGGCGAGGTGGGC-----> exon12
                                            or
                                          (1935 bp)   <-----cctatcttaatccagCAATATCCTTTTCCTTTCCAGGTGGGCGAGGTGGGC-----> exon12

exon12 <-----ATCGCAAGTGCTTAGgtttgtccttactct----->     intron12 (3777 bp)       <-----tctaaatttcaacagAGTGCAATGACTCCG-----> exon13
exon13 <-----AATGTAAAAGAAAAAAAATCAAAACTAATTATTTGTAATATTTATGCCTTAATAAAATGATTTTAACTTAgctgctactgcttg  poly-adenylation at end exon13
                                                                                    ^^

                                            or

exon12 <-----ATCGCAAGTGCTTAGGTTTGTCCTTACTCT<----->ACAAACGCATCCTAAATATATTATACACATTAtattta  internal poly-adenylation within exon12
                                                                                  ^^

                                            or

exon12 <-----ATCGCAAGTGCTTAGGTTTGTCCTTACTCT<----->TATAATATCTATATGAGTAATAAATAAATGCATTGACaacgtaaaaa  poly-adenylation at end exon12
                                                                                  ^?
```

Fig. 5. Nucleotide sequence of the 5′- and 3′-splice sites of the *Dfurl* exons as deduced from comparison of *Dfurl* genomic sequences with cDNA sequences. Exon sequences are given in upper case, whereas intron sequences are in lower case. All splice sites are according to the consensus rules: all introns start with GT and end with C/TAG (underlined). Exons and introns are numbered; the sizes of the introns are indicated, when known. Since the intron sequences of the 3′- and 5′-splice sites of exon 2 and exon 3 are not known, these intron sequences are represented by *n*. Potential polyadenylation signals are underlined by a broken line; the two alternative polyadenylation sites at the 3′-end of exon 12, of which only one is exactly known (indicated by ^^) and the other postulated in that region (indicated by ^?), and exon 13 (deduced by comparison with the dKLIP-1 cDNA sequence) are indicated (^^).

Mapping of exons in Dfur1 transcripts by Northern blot analysis

None of the isolated *Dfurl* cDNAs appeared to represent fully any of the four transcripts that were detected by Northern blot analysis (Figures 1 and 4). All cDNAs lacked a poly(dA) tail at their 3′-ends. To establish to which of the transcripts the various cDNAs correspond, Northern blot analysis was performed with informative *Dfurl* probes which were selected on the basis of the deduced exon composition of the cDNAs. RNA samples of the late embryonic stage (22 h) and the adult stage were selected for these studies because together they expressed the four transcripts (7.6, 6.5, 4.5 and 4.0 kb) at relatively high levels (Figure 6, lanes 5A, 5B, 6A and 6B, and Figure 8). It should be noted that after prolonged exposure, very low levels of a transcript >7.6 kb were found in some stages. This transcript could represent a pre-mRNA intermediate or another alternative mRNA of the *Dfurl* gene. Furthermore, low levels of transcripts of ~3.0 and 1.8 kb were sometimes detected. The relationship of these to *Dfurl*, if any, remains to be established. An explanation might be cross-hybridization to transcripts of a highly homologous gene. The hybridization signal at ~1.8 kb is seen in front of the bulk of 18S rRNAs and could therefore be non-specific, due to local overloading of the gel in combination with some RNA degradation. Hybridization of the various *Dfurl* probes to the 7.6 kb transcript is shown in Figure 6 and summarized in Table II. It is of importance to note that pIP62 and pIP58, which has two potential polyadenylation signals near its 3′-end (Figure 2), do not overlap. However, analysis of a polymerase chain reaction (PCR) product, synthesized on first strand cDNA as template using a 5′-end primer derived from exon 10 and a 3′-end primer derived from sequences downstream of the above-mentioned gap, confirmed the linkage of pIP58 and pIP62 sequences in the transcript of 7.6 kb. It should be noted that pIP58 could also be derived

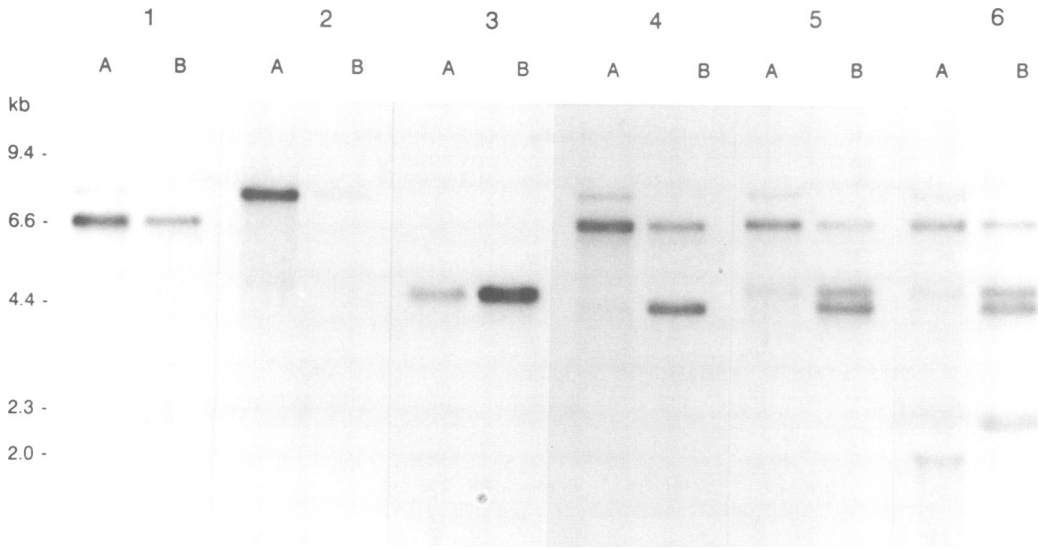

Fig. 6. Northern blot analysis of the four different *Dfur1* transcripts of *D.melanogaster* with different *Dfur1* probes. In each lane, 15 μg of total RNA of embryonic stage (22 h) (lanes 1A−6A) and adult stage (lanes 1B−6B) are analyzed by agarose gel electrophoresis as described in Materials and methods. Lanes 1 were hybridized with probe pIP44, lanes 2 with probe PCR1, lanes 3 with probe pIP46, lanes 4 with probe PCR2, lanes 5 with probe pG7 and lanes 6 with probe PCR4.

Table II. Hybridization of *Dfur1* transcripts with different *Dfur1* probes[a,b]

	PCR4 [1−3]	pG6 [2−4]	pG7 [5−9, 11−12]	PCR1 [10]	PCR2 [12]	PCR3 [12]	pIP44 [12]	pIP46 [13]
7.6 kb transcript	+	+	+	+	+	+	+	−
6.5 kb transcript	+	+	+	−	+	+	+	−
4.5 kb transcript	+	+	+	−	−	−	−	+
4.0 kb transcript	+	+	+	−	+	−	−	−

[a]*Dfur1* probes are described in detail in Table I, exons (partly) present in the *Dfur1* probes are indicated between brackets.
[b]For Northern blot analysis see Figures 6 and 8; hybridization with probe PCR3 is not shown.

from the 6.5 kb transcript. Altogether, the nucleotide sequence data of pIP62 and the Northern blot results suggest that the transcript of 7.6 kb contains exons 1−12 (Figure 4C), accounting for 7.3 kb of the transcript. Assuming that the 7.6 kb transcript has a 3′-end poly(A) tail, only some nucleotides downstream of putative polyadenylation signals are expected to be missing from the data as presented in Figure 2.

As can be seen in Table II, only the 4.5 kb transcript seems to contain exon 13 sequences (see also Figure 6, lanes 3A and B). Therefore, the transcript of 4.5 kb is assumed to correspond to the cDNA insert of pIP56 and to contain exons 1−9, exon 11, a small 5′-end part of exon 12, and exon 13 (Figure 4C); exon 12 appears to possess an internal 5′-splice site and this is apparently used for the generation of the 4.5 kb transcript. The cDNA insert of pIP56 appears to be identical to the cDNA sequence of dKLIP-1 of *D.melanogaster* (Hayflick *et al.*, 1992). The 3′-end of the dKLIP-1 cDNA extends only a few nucleotides further downstream than the *Dfur1* cDNA insert of pIP56. However, a polyadenylation signal and a poly(dA) tail are present in the dKLIP-1 cDNA (Figure 5). It is therefore safe to assume that the pIP56 cDNA insert (4.2 kbp) accounts for almost the complete transcript of 4.5 kb.

Based upon the hybridization pattern of the *Dfur1* probes to the 6.5 kb transcript (Table II and Figure 6), it is assumed that it contains exons 1−9, exon 11 and exon 12 (Figure 4C), accounting for about 6.2 kb of the transcript. As can also be seen in Table II, the transcripts of 4.0 and 6.5 kb displayed highly similar hybridization patterns in the Northern blot studies, except for differences in their 3′-end regions. Screening of *Drosophila* cDNA libraries with probe PCR2 resulted in the isolation of a cDNA of ~400 bp; this clone appeared to contain nucleotides 4632−5032 or 5033 and a short poly(dA) tail (data not shown). Although a classical polyadenylation signal (AATAAA or ATTAAA) is not present immediately before the polyadenylation site, the isolation of this cDNA together with the hybridization data suggest that the transcript of 4.0 kb might be generated by polyadenylation within exon 12 at nucleotide 5032 or 5033 (see also Figure 5). This would account for a size difference between the two transcripts of ~2.3 kb which is similar to the size difference observed in Northern blot analysis. Therefore, the transcript of 4.0 kb is assumed to contain exons 1−9, exon 11 and a portion of exon 12 (Figure 4C). This implies that the coding sequences of the 4.0 kb transcript are identical to those of the 6.5 kb transcript. Sequences of a region of 3.8 kb of the 4.0 kb transcript are accounted for. Since cDNA pIP63 ends at nucleotide 5018, it could be derived from either the transcript of 4.0 or 6.5 kb.

A matter not addressed yet pertains to the differential occurrence of exon 2 in the cDNAs described above. Of all eight *Dfur1* cDNAs that were initially selected on the basis of their hybridization to probe pG6, six were extended at

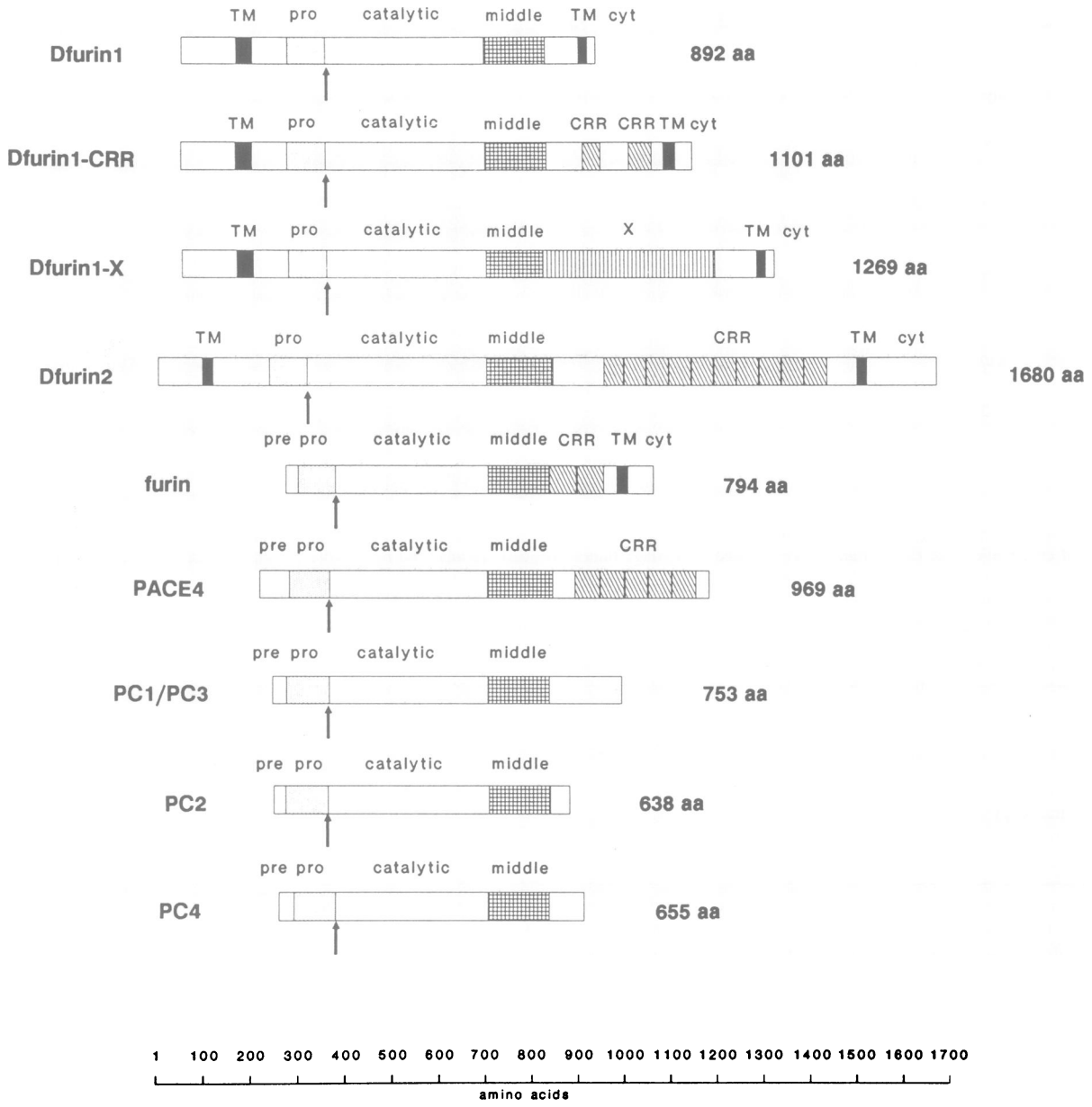

Fig. 7. Schematic representation of known and putative protein domains of Dfurin1, Dfurin1-CRR, Dfurin1-X and Dfurin2 of *D.melanogaster*, human furin, human PACE4, human PC1/PC3, human PC2 and mouse PC4. Protein domains are represented as boxes marked as indicated; signal peptide domain (pre), 'pro' domain (pro), subtilisin-like catalytic domain (catalytic), 'middle' domain (middle), cysteine-rich region (CRR) which is divided into subdomains, transmembrane domain (TM) and cytoplasmic doamin (cyt). The potential (auto)proteolytic cleavage site immediately upstream of the subtilisin-like catalytic domain of each protein is indicated by an arrow.

their 5′-ends up to exon 1 sequences. Of these six clones, only pIP63 did not contain exon 2 sequences. This may indicate that exon 2 is not very often spliced out. However. of these six clones, pIP63 (without exon 2) and pIP64 (with exon 2; data not shown) were deduced to correspond to the 4.0 or 6.5 kb transcript. So it cannot be excluded that exon 2, which presumably is a small (65 bp) non-coding exon, is selectively spliced out of one of these two transcripts. Another site for alternative splicing has already been reported in our previous study (Roebroek *et al.*, 1991). It was noted that for splicing of exon 11 to exon 12 sequences, two alternative 3′-splice sites were used (Figure 5); alternative use resulted in the presence or absence of 21 nucleotides encoding seven amino acids of the Dfurin1 protein. In none of the *Dfur1* cDNAs analyzed in this study were these 21 nucleotides present.

The deduced Dfur1 proteins possess common and unique structural features

On the basis of the nucleotide sequences of the large open reading frames in the various cDNAs, the amino acid sequences of the corresponding *Dfur1* proteins were deduced. One deduced protein, most likely encoded by the 4.0 and 6.5 kb transcripts, appeared to represent previously described (Roebroek *et al.*, 1991) Dfurin1 which consists of 892 amino acid residues (Figure 7). It should be noted that there might be two alternative forms of the Dfurin1 protein because of alternative splicing; in the one previously described, exon 12 is seven amino acid residues larger than the one described here (Roebroek *et al.*, 1991). The transcript of 7.6 kb most likely codes for a protein of 1269 amino acid residues. The difference with Dfurin1 is the insertion of a unique protein domain of 377 amino acid

residues which are encoded by exon 10 (Figures 2 and 4C); this unique domain is inserted after amino acid residue 775. Since analysis of this protein domain did not reveal special structural features, it was designated X and the protein, Dfurin1-X (Figure 7). Finally, the transcript of 4.5 kb is assumed to code for a *Dfur1* protein of 1101 amino acid residues which is identical to Dfurin1 up to amino acid residue 856; the rest of the protein, which is encoded by exon 13, is unique (Figure 4C). In this unique part, a cysteine-rich region is present; dispersed as two subdomains, the cysteine motif is similar to those in the cysteine-rich regions of furin (Roebroek *et al.*, 1986b; Van de Ven *et al.*, 1990), PACE4 (Kiefer *et al.*, 1991) and Dfurin2 (Roebroek *et al.*, 1992) (Figure 7). Because of the presence of this cysteine-rich region (CRR), this third *Dfur1* protein was designated Dfurin1-CRR (Figure 7). Dfurin1-CRR appears to be the same as the previously described dKLIP-1 protein of *D.melanogaster* (Hayflick *et al.*, 1992). The only difference between the two deduced proteins is the presence of a threonine residue at position 1014 in Dfurin1-CRR and an isoleucine residue in dKLIP-1; it is caused by nucleotide substitution at position 3687 from C to T (Figure 3), probably due to strain differences.

As shown in Figure 7, all three *Dfur1* proteins are identical up to amino acid residue 775 (Figure 2) just before the end of the 'middle' domain. Starting at the amino-terminus, this common region of the three proteins includes a domain with a putative transmembrane anchor potentially functioning as an internal signal sequence, a so-called 'pro' domain, a subtilisin-like catalytic domain and a so-called 'middle' domain (Van de Ven *et al.*, 1990; Roebroek *et al.*, 1992). The amino-terminal region of Dfurin1 is similar to that of *Drosophila* Dfurin2 which also starts with a putative transmembrane region with an internal signal sequence (Roebroek *et al.*, 1992). The mammalian subtilisin-like processing enzymes, in contrast, all have a classic signal peptide. Carboxy-terminal of amino acid residue 775, the insertion of the unique X domain is found in Dfurin1-X. Carboxy-terminal of the X domain, Dfurin1 and Dfurin1-X are identical again; this common region contains a potential carboxy-terminal transmembrane domain and a small cytoplasmic tail. Dfurin1-CRR has a unique carboxy-terminal part, starting after amino acid residue 856 (Figure 3); in Dfurin1 and Dfurin1-X, this is the second amino acid residue of the carboxy-terminal transmembrane domain. In this region of divergence between Dfurin1-CRR and the two other *Dfur1* proteins, the first subdomain of the cysteine-rich region of Dfurin1-CRR is found and, more carboxy-terminally, a second cysteine-rich subdomain, a potential transmembrane domain and a small cytoplasmic domain. It should be emphasized that these last two domains have a different amino acid sequence as the transmembrane and cytoplasmic domains in Dfurin1 and Dfurin1-X. The carboxy-terminal transmembrane anchor and small cytoplasmic tail seem to be typical features of furin-like proteins; all other mammalian proprotein processing enzymes described so far, PACE4, PC1/PC3, PC2 and PC4, apparently lack such domains. Comparison of the three *Dfur1* proteins with Dfurin2 of *D.melanogaster*, human furin, human PACE4, human PC1/3, human PC2 and mouse PC4 (Figure 7) revealed that the domain structure of Dfurin1-CRR most closely resembles that of Dfurin2; the cysteine-rich region in Dfurin1-CRR, however, is much smaller. Like in human furin, only two cysteine-rich region subdomains are present in Dfurin1-CRR, whereas Dfurin2 and PACE4 possess, respectively, ten and five such subdomains.

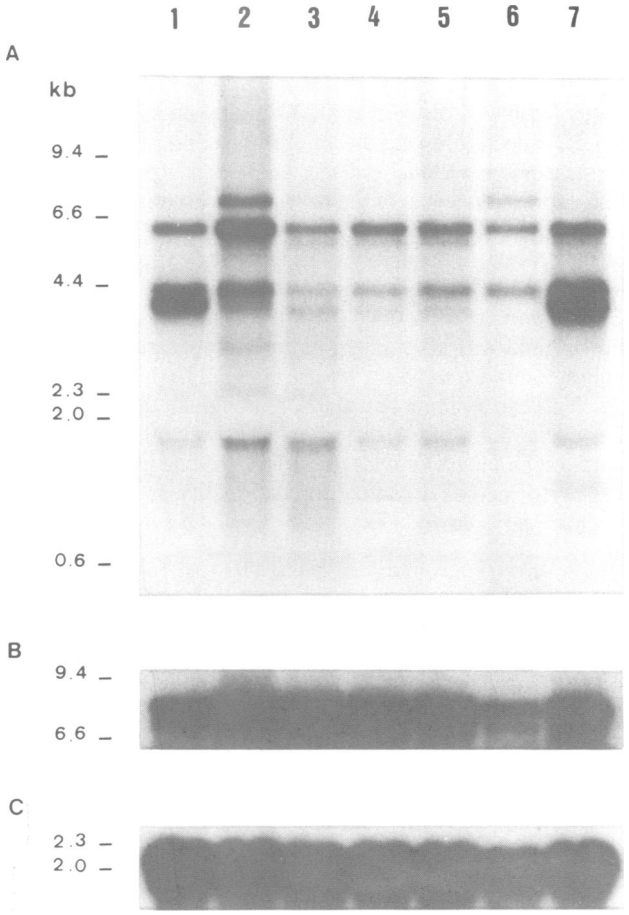

Fig. 8. Northern blot analysis of *Dfur1* gene expression in various developmental stages of *D.melanogaster*. Total RNA (15 μg) was subjected to agarose gel electrophoresis. Lane 1: embryonic stage (2−8 h); lane 2: embryonic stage (22 h); lane 3: larval stage 1 (36 h); lane 4: larval stage 2 (60 h); lane 5: larval stage 3 (96 h); lane 6: pupa stage; lane 7: adult stage. All RNA samples were from the Oregon R strain except for the one in lane 1, which was from the Tübingen strain. (**A**) Hybridization to the *Dfur1* cDNA probe pG6 (72 h exposure). (**B**) Hybridization of a 7.7 kb pre-rRNA and a somewhat smaller processing intermediate to the *Drosophila* rRNA-specific probe dhPR21 (40 h exposure). (**C**) Hybridization of 18S rRNA to the same probe as in B (1 h exposure). Molecular weight markers, λ DNA digested with restriction endonuclease *Hin*dIII, are indicated (kb).

Differential expression of Dfur1 transcripts during development of D.melanogaster

As can be deduced from the results above, the four different *Dfur1* transcripts code for three different *Dfur1* proteins (Figure 4C). In an attempt to get more insight into the expression pattern of the various *Dfur1* transcripts, we performed Northern blot analysis of RNA isolated from various developmental stages of *D.melanogaster*. Stages analyzed are listed in Materials and methods. pG6, which was shown to hybridize to all four transcripts, was used as molecular probe (Figure 8A). All four transcripts appeared to be expressed in all stages tested. However, as can be seen in the figure, expression levels vary; not only the combined expression levels of the four *Dfur1* transcripts, but also the

relative levels of these. The combined expression levels are relatively low in the larval and pupa stages (Figure 8A, lanes 3, 4, 5 and 6); they are relatively high in the embryonic stages (embryos of 2−8 and 22 h old; Figure 8A, lanes 1 and 2, respectively) and the adult stage (Figure 8A, lane 7). As far as the individual transcripts are concerned, relatively high expression levels of the 6.5 and 7.6 kb transcripts were observed in the late embryonic stage (22 h) (Figure 8A, lane 2), whereas relatively high levels of expression of the 4.0 kb transcript were seen in the early embryonic stage (2−8 h) and the adult stage (Figure 8A, lanes 1 and 7). Since the 6.5 and 4.0 kb transcripts have the same large open reading frame (encoding Dfurin1), mRNA levels for this *Dfur1* protein seem more abundant throughout the *Drosophila* life cycle than those encoding the two other proteins. Northern blot analysis with the rRNA-specific probe dhPR21 was performed as a control for the quality and quantity of the RNA samples that were analyzed. A pre-rRNA transcript of ~7.7 kb and a somewhat smaller processing intermediate (Figure 8B) and the bulk of 18S rRNA (Figure 8C) were detected by the dhPR21 probe. The results of these control experiments indicated that the amount and quality of the RNA samples were fairly similar, although the amount of RNA in the pupa sample was apparently somewhat lower.

In situ hybridization analysis of Dfur1 expression during Drosophila embryogenesis

On the basis of the Northern blot results, no specific differentiation could be made as to which of the *Dfur1* transcripts were expressed at what moment in time during *Drosophila* development by which cell types. As an approach to examine when and where during *Drosophila* embryogenesis the different furin isoforms are expressed, *in situ* hybridization of 0.0−24 h wild-type Canton-S *Drosophila*

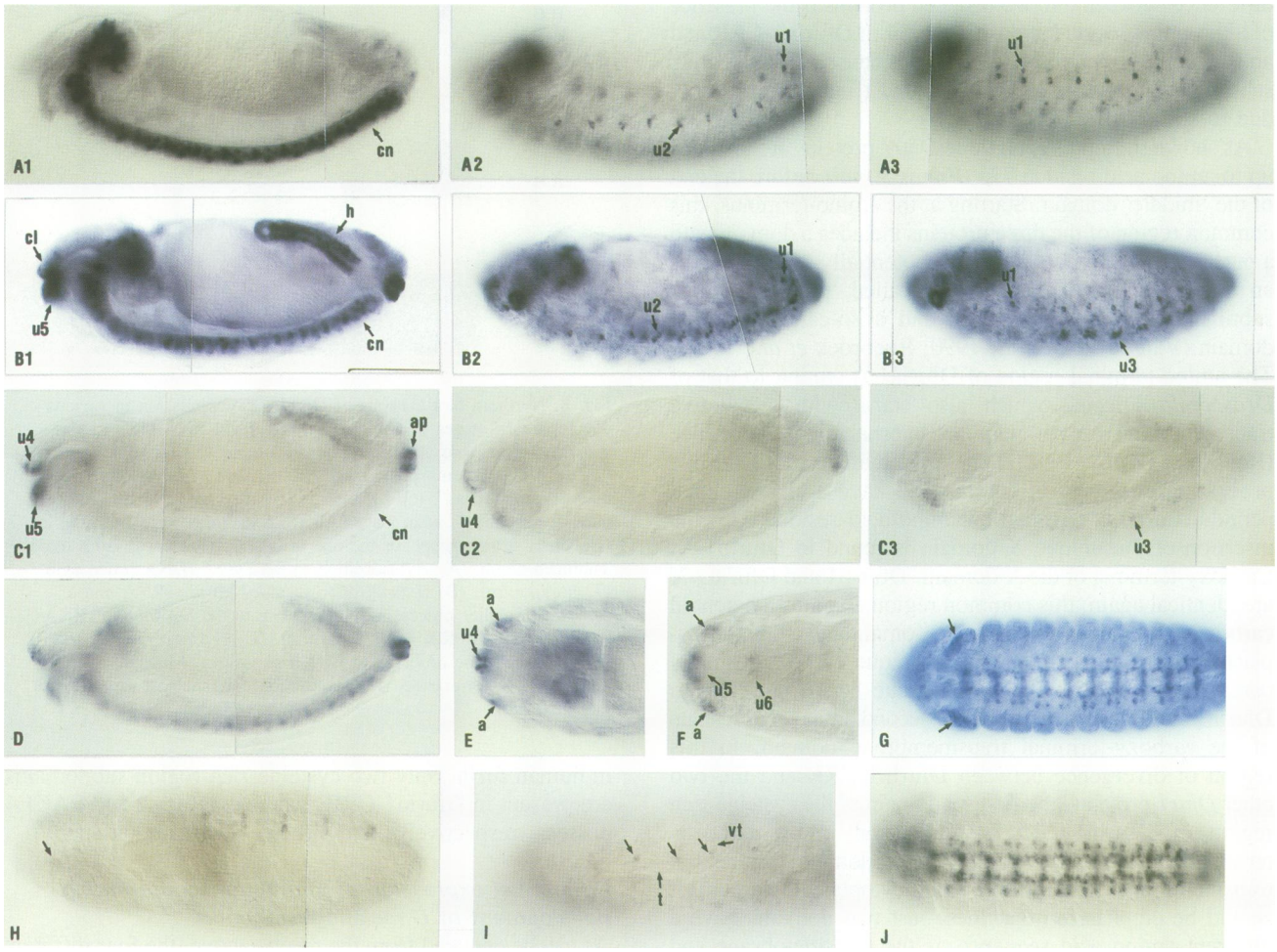

Fig. 9. RNA *in situ* hybridization analysis of *Drosophila* embryos. (**A, B, C, D**) Stage 14 embryos, parasagittal optical sections. A1, B1, C1 and D are mediosagittal. A2−A3, B2−B3 and C2−C3 are progressively more lateral. (A) is hybridized with the PCR1 (Dfurin1-X) probe. (B) and (D) are hybridized with probe pG13 which recognizes all *Dfur1* transcripts. (D) was incubated for a shorter time in the staining mix than (B). (C) is hybridized with the pIP95 (Dfurin1-CRR) probe. (**E, F**) Stage 15 embryos, horizontal optical sections, head region hybridized with the probe that recognizes all *Dfur1* isoforms (E) and the probe specific for Dfurin1-CRR mRNA (F). (**G, J**) Horizontal optical sections of stage 13 embryos at CNS level hybridized, respectively, with the probe that detected all *Dfur1* mRNA isoforms (G) and the probe specific for Dfurin1-X mRNA (J). The patterns of cells of the neurogenic region that show signal with the Dfurin1-X probe or the general *Dfur1* probe are identical. This indicates that the Dfurin1 mRNA isoform is not expressed in additional CNS cells. The arrows in (G) point to the signal in the labial segment. (**H**) Dorsolateral view of a stage 10 embryo hybridized with the probe specific for Dfurin1-X mRNA showing signal in two cells per segment in the midline area of the neurogenic region and one cell in the procephalic neurogenic region (arrow). (**I**) Dorsolateral view of a stage 16 embryo hybridized with the probe specific for Dfurin1-CRR mRNA showing signal (three arrows) in cells at the junction between the trachea (t) and the vertical tracheal branches (vt). Abbreviations: cn, central nervous system; ap, anal plate; a, antennomaxillary complex; h, hindgut; cl, clipeolabrum; t, trachea; vt, vertical tracheal branch; for a description of u1, u2, u3, u4, u5 and u6, see Results.

embryos was performed with two probes with specificity to either Dfurin1-X mRNA (probe PCR1) or Dfurin1-CRR mRNA (probe pIP95) and a third probe common to all transcripts of the *Dfur1* gene. We present here our analysis to the extent of demonstrating that the Dfurin1-X and Dfurin1-CRR isoforms are expressed in non-overlapping sets of tissues during embryogenesis. Results are illustrated in Figure 9. The unambiguous identification of the staining cells u1−u6 (Figure 9) requires double staining with markers for various neuronal and other cell types, and is beyond the scope of this study.

Dfurin1-X mRNA is first observed at stage 10−11 (Campos-Ortega and Hartenstein, 1985) in two cells per segment that are positioned close to the ventral midline of the neurogenic region (Figure 9H). Their small size suggests that they are not neuroblasts. As neurogenesis proceeds, an increasing number of cells in the central nervous system (CNS) express Dfurin1-X mRNA (Figure 9J). At stage 14 embryogenesis, Dfurin1-X mRNA is also expressed in two cells (per segment) that are associated with the tracheal pits (u1) and a more ventral cluster of 2−3 cells (u2) (Figure 9, A1, A2 and A3). Weak staining is occasionally observed in a few cells of the hindgut, posterior spiracle and anal pads. From stage 13−14 of embryogenesis onwards, Dfurin1-CRR mRNA is present in the anal pads, the hindgut, cells of the developing antennomaxillary complex and small clusters of 4−6 ventrolateral cells per abdominal segment in the same position as oenocytes (u3) (Figure 9, C1−C3, F). Expression is also observed in some cells in the tip of the clipeolabrum (u4), an anterior midventral cluster of cells (at the level of the subesophageal ganglion) (u5) and a few bilaterally symmetric positioned cells associated with the floor plate of pharynx cells (u6) (Figure 9, C1−C3, F). Dfurin1-CRR mRNA is expressed at low levels in some cells of the developing posterior spiracles from stage 13 onwards (Figure 9, C1). In stage 15 embryos, two cells per segment positioned at the junctures of the trachea with the main segmental branches per segment express Dfurin1-CRR mRNA (Figure 9 I). Hybridization studies of embryos using probe pG13, a probe that detects all transcripts for the *Drosophila* furin isoforms, Dfurin1-X, Dfurin1 and Dfurin1-CRR, revealed a strong signal in those tissues that expressed Dfurin1-X and Dfurin1-CRR, as described above (Figure 9, B1−B3, D, E, G). At stage 13, approximately the same number of CNS cells stained with the pG13 probe as with the probe specific for Dfurin1-X mRNA Figure 9G and J). Except for the labial bud, at stage 13, there are no additional tissues that stain prominently with pG13 and not with probes specific for exons 10 (probe PCR1) and 13 (probe pIP95). However, when left longer in the alkaline phosphatase mix, ubiquitous staining appears. It is difficult to determine whether this signifies that the Dfurin1 isoform is expressed ubiquitously or whether we observe general background staining. Because probes specific to Dfurin1 mRNA cannot be made, it remains possible that Dfurin1 mRNA is expressed highly in some of the tissues that express Dfurin1-X or Dfurin1-CRR mRNA. In conclusion, the observation that the *Dfur1*-derived mRNAs, which encode the Dfurin1-X and Dfurin1-CRR isoforms, are expressed in non-overlapping sets of tissues during *Drosophila* embryogenesis is of interest in that it points towards differences in the physiological role and function of the corresponding proteins.

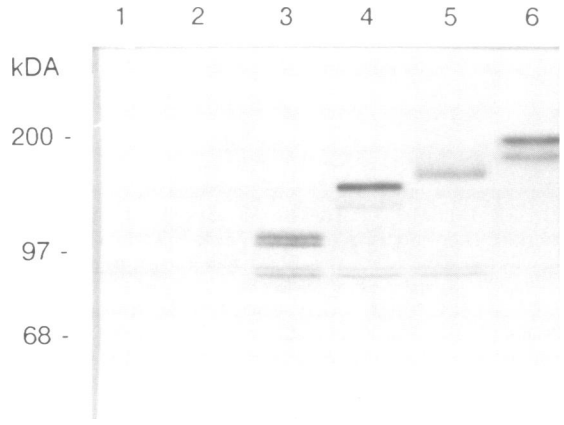

Fig. 10. Biosynthesis of *Dfur1*- and *Dfur2*-encoded proteins. PK(15) cells were infected with V.V.:T7 recombinant vaccinia virus and subsequently transfected with pGEM-Dfur1 (lanes 1 and 3), pGEM-Dfur1-CRR (lane 4), pGEM-Dfur1-X (lane 5) or pGEM-Dfur2 (lanes 2 and 6) DNA, labeled with [^{35}S]methionine as described in Materials and methods. For immunoprecipitation analysis, rabbit anti-Dfurin1 antiserum (lanes 3−5) or anti-Dfurin2 antiserum (lane 6) was used. As controls, corresponding pre-immune sera were used (lanes 1 and 2). Molecular weight markers are indicated.

Biosynthesis of the various Dfur1-encoded furin isoforms and evaluation of their proprotein processing activity

In previous studies, we established the proprotein processing activity of *Drosophila* Dfurin2 by showing that it was capable of correctly cleaving pro-vWF (Roebroek *et al.*, 1992); the same heterologous processing assay was used as in demonstrating the cleavage specificity for paired basic residues of human and mouse furin (Van de Ven *et al.*, 1990; Van Duijnhoven *et al.*, 1992). Since the *Dfur1*-encoded proteins displayed similar structural features as Dfurin2 and mammalian furin, it was of interest to evaluate the potential proprotein processing activity of the *Dfur1*-encoded furin isoforms. This was studied using two different heterologous gene expression systems, as outlined in detail in Materials and methods, and two different cleavage substrates, pro-vWF and the precursor of the β_A-chain of activin-A. To assay the biosynthesis of the various *Dfur1*-encoded furin isoforms in these two expression systems, immunoprecipitation analysis was performed. The results of studies in which expression of the *Dfur1* cDNAs was under the control of a T7 promoter with T7 RNA polymerase produced by a recombinant vaccinia virus are shown in Figure 10; studies in COS-1 cells with expression under the control of the SV40 late promoter gave similar results (data not shown). As can be seen in Figure 10, two proteins with mol. wts of ~110 and 115 kDa, respectively, were immunoprecipitated by the rabbit anti-Dfurin1 antiserum from V.V.:T7-infected PK(15) cells transfected with pGEM-Dfur1 DNA (lane 3). The two proteins were not detected by the corresponding pre-immune serum (lane 1). Similarly, a 155 kDa and low amounts of a 135 kDa protein were immunoprecipitated from lysates of V.V.:T7-infected cells transfected with pGEM-Dfur1-CRR DNA (lane 4). In infected PK(15) cells transfected with pGEM-Dfur1-X DNA, a single protein of ~165 kDa was detected (lane 5). Using a rabbit anti-Dfurin2 antiserum, two proteins with mol. wts of ~200 and 180 kDa were detected in V.V.:T7-infected PK(15) cells transfected

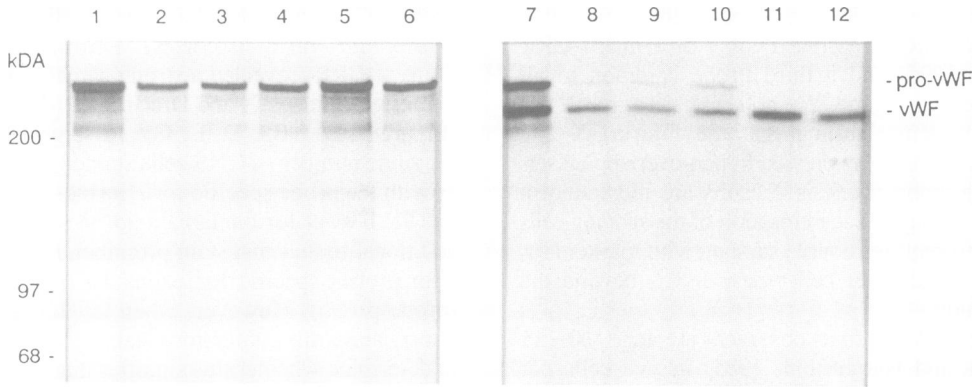

Fig. 11. Analysis of proteolytic processing of pro-vWF by Dfurin1, Dfurin1-CRR, Dfurin1-X and Dfurin2 of *D.melanogaster* and human furin. COS-1 cells were transfected with 10 μg of pSVLvWFgly763 DNA (lane 1), 5 μg of pSVLvWFgly763 DNA and 5 μg of pSVLDfur1 DNA (lane 2), 5 μg of pSVLvWFgly763 DNA and 5 μg of pSVLDfur1-CRR DNA (lane 3), 5 μg of pSVLvWFgly763 DNA and 5 μg of pSVLDfur1-X DNA (lane 4), 5 μg of pSVLvWFgly763 DNA and 5 μg pSVLDfur2 DNA (lane 5), 5 μg of pSVLvWFgly763 DNA and 5 μg of pSVLfur-hu DNA (lane 6), 10 μg of pSVLvWF DNA (lane 7), 5 μg of pSVLvWF DNA and 5 μg pSVLDfur1 DNA (lane 8), 5 μg of pSVLvWF DNA and 5 μg of pSVLDfur1-CRR DNA (lane 9), 5 μg of pSVLvWF DNA and 5 μg of pSVLDfur1-X DNA (lane 10), 5 μg of pSVLvWF DNA and 5 μg of pSVLDfur2 DNA (lane 11), and 5 μg of pSVLvWF DNA and 5 μg of pSVLfur-hu DNA (lane 12). Biosynthesis and processing of vWF-related proteins were studied by immunoprecipitation analysis which was performed as described in Materials and methods. Molecular weight markers and the relative positions of pro-vWF and mature vWF are indicated.

with pGEM-Dfur2 DNA (lane 6); these two proteins were not detected by the corresponding pre-immune serum (lane 2). Finally, the anti-Dfurin1 antiserum did not react with Dfurin2 and the anti-Dfurin2 antiserum did not react with any of the three *Dfur1*-encoded furin isoforms (data not shown). In three of the four cases, two *Drosophila* furin-like proteins were detected in the transfected cells. Although in each case the precise interrelationship between both proteins remains to be resolved, the observed differences in molecular weight are most likely the result of post-translational modification processes such as precursor processing and glycosylation. In any case the transfection experiments clearly indicate that the various *Drosophila* proteins are readily expressed in PK(15) (and also COS-1 cells) under the selected conditions. It should be noted that to obtain the relatively high levels of expression of the *Dfur1*-encoded furin isoforms, sequences from exon 1, that apparently interfered with translation, were deleted (see also Materials and methods).

To assay the potential endoproteolytic cleavage activity of the *Dfur1*-encoded furin isoforms, DNA of pSVLDfur1, pSVLDfur1-X or pSVLDfur1-CRR DNA was co-transfected into COS-1 cells together with pSVLvWF DNA or pSVLvWFgly763 DNA. In control co-transfection experiments, pSVLDfur2 DNA encoding *Drosophila* Dfurin2 and pSVLfur-hu DNA encoding human furin were co-transfected. The results of the experiments are shown in Figure 11. In transfection experiments using DNA of cleavage mutant pSVLvWFgly763 (Figure 11, lane 1), pro-vWFgly763 was synthesized and constitutively secreted into the culture medium as a 360 kDa protein; no processing was observed. Upon transfection of pSVLvWF DNA, which encodes wild-type pro-vWF, the 360 kDa pro-vWF precursor and the 260 kDa mature vWF protein were found in the conditioned medium in almost equal amounts (Figure 11, lane 7). This mature vWF is most likely formed as a result of processing by an endogenous proprotein processing enzyme, possibly endogenous furin (Van de Ven *et al.*, 1990). Co-transfection experiments using pSVLvWFgly763 DNA and pSVLDfur1, pSVLDfur1-CRR or pSVLDfur1-X DNA revealed the production of only the 360 kDa mutant

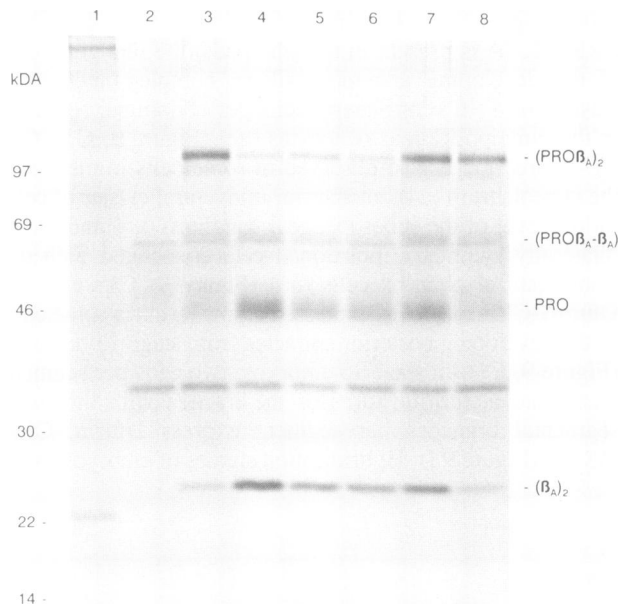

Fig. 12. Analysis of processing of proβ_A by human furin, Dfurin1, Dfurin1-CRR, Dfurin1-X and Dfurin2. PK(15) cells were mock-infected (lane 1) or infected with V.V.:T7 recombinant vaccinia virus (lanes 2–8) and subsequently transfected with 2 μg of pTZR19β13 DNA (Lanes 3–8) and 2 μg of pSelect-wt.fur-hu DNA (lane 4), pGEM-Dfur1 DNA (lane 5), pGEM-Dfur1-CRR DNA (lane 6), pGEM-Dfur1-X DNA (lane 7) or pGEM-Dfur2 DNA (lane 8). Biosynthesis and processing of β_A-related proteins were analyzed by SDS–PAGE as described in Materials and methods. Molecular weight markers and the relative positions of (proβ_A)$_2$, proβ_A–β_A, the 'PRO' polypeptide of proβ_A and (β_A)$_2$ are indicated.

pro-vWFgly763 protein and no processing products (Figure 11, lanes 2, 3 and 4). Co-transfection of pSVLvWF and pSVLDfur1, pSVLDfur1-CRR or pSVLDfur1-X DNA, however, did reveal increased cleavage of pro-vWF in mature vWF (Figure 11, lanes 8, 9 and 10). Pro-vWF was never fully cleaved and the cleavage efficiency of Dfurin-X

appeared to be consistently the lowest of the three. As a control experiment, co-transfection of pSVLvWFgly763 and pSVLDfur2 DNA revealed the production of only the 360 kDa mutant pro-vWFgly763 protein and no processing products (Figure 11, lane 5), whereas co-transfection of pSVLvWF and pSVLDfur2 DNA revealed complete cleavage of pro-vWF into mature vWF (Figure 11, lane 11). Similar results were obtained with human furin (Figure 11, lanes 6 and 12).

In a report by Hayflick *et al.* (1992), proprotein processing activity and specificity for paired basic amino acid residues of the dKLIP-1 protein (which is identical to Dfurin1-CRR in this paper) was demonstrated. In these studies, African green monkey kidney epithelial cells (BSC-40 cells) and a vaccinia virus expression system were used; as substrate, murine pro-β-NGF was used. To evaluate the proprotein processing activity of the *Dfur1*-encoded proteins under similar conditions, but with a different substrate, we performed co-transfection experiments with pig kidney cells [PK(15) cells] which, before transfection, were infected with recombinant vaccinia virus V.V:T7. Recombinant V.V.:T7 encodes T7 RNA polymerase. As substrate, bovine proβ_A was used. DNA constructs expressing cDNA sequences for bovine proβ_A, human furin, Dfurin1, Dfurin2, Dfurin1-X, or Dfurin1-CRR under the control of a T7 promoter were used, as outlined in more detail in Materials and methods. Results of these experiments are presented in Figure 12. In lanes 1 and 2, SDS−PAGE analysis of labeled proteins secreted in the medium by mock-infected PK(15) cells or cells infected with V.V.:T7 virus are shown. Upon lipofection of the infected cells with pTZ19Rβ_A, relatively large amounts of (proβ_A)$_2$ and only very limited amounts of (β_A)$_2$ were found in the medium (lane 3). It should be noted here that mature β_A contains four methionines compared with 10 in the prohormone. As a result of co-transfer of human *fur* cDNA sequences (pSelect-wt.fur-hu DNA), the secreted amounts of (proβ_A)$_2$ found in the medium were reduced whereas those of (β_A)$_2$ and 'PRO' were increased strongly (lane 4). Similar results were obtained in co-transfection experiments with pGEM-Dfur1 (lane 5), pGEM-Dfur1-CRR (lane 6) and pGEM-Dfur1-X (lane 7). When one of these DNAs were co-transferred, the secreted amounts of (proβ_A)$_2$ in the medium were reduced whereas those of (β_A)$_2$ and 'PRO' were strongly increased. It should be noted that also in this assay system, the cleavage efficiency of Dfurin1-X appeared to be the lowest of the three *Dfur1*-encoded furin isoforms. In experiments in which pGEM-Dfur2 DNA was co-transferred with pTZ19Rβ_A DNA, high levels of (proβ_A)$_2$ and only background amounts of secreted (β_A)$_2$ were found in the medium (lane 8). Apparently, the ability of Dfurin2 to process the β_A precursor in this system is highly limited; this is in contrast to the pro-vWF processing in COS-1 cells. Altogether, the results of the processing experiments suggest that all three *Dfur1*-encoded furin isoforms are endoproteolytic processing enzymes and possess proprotein processing activity with specificity for paired basic amino acid residues.

Discussion

Molecular studies of the known genes of the novel family of eukaryotic subtilisin-like proprotein processing enzymes have not yet identified a member that encodes multiple isoforms with distinct physiological functions. In a number of cases, more than one mRNA species was detected in Northern blot analysis. Molecular diversity in the deduced proteins was observed, for instance, in the germ-cell-specific subtilisin-like proprotein convertase PC4 (Seidah *et al.*, 1992b) and a PC3-like protein from a simple metazoan *Hydra vulgaris* (Jin Chan *et al.*, 1992); however, no differences in physiological functions of the isoforms have been reported. In the case of the neuroendocrine-specific PC1/PC3 gene, differences between the transcripts were found only in non-coding sequences (Creemers *et al.*, 1992) and the structural diversity between the human PACE4 and the PACE4.1 transcripts (Kiefer *et al.*, 1991) was only found in one tumor cell line and might merely represent an aberrant form of PACE4. Here, we present evidence that a gene of this novel eukaryotic gene family is transcribed under physiological conditions in multiple mRNAs encoding structurally different protein isoforms. This structural diversity, which seems to be due to alternative splicing, is very likely also to have functional implications, as suggested by our *in situ* hybridization data, which indicate that the two isoforms Dfurin1-X and Dfurin1-CRR are expressed in topologically different and non-overlapping sets of tissues during *Drosophila* embryogenesis. Although the precise degree of functional diversity of these *Dfur1*-encoded *Drosophila* proprotein processing enzymes remains to be established, the observed diversity is intriguing and it would be of interest to know how widespread this phenomenon is in nature. The diversity in the presently known mammalian subtilisin-like processing enzymes is still limited and restricted to the widely expressed enzymes furin and PACE4, the neuroendocrine-specific enzymes PC1/PC3 and PC2, and the testis-specific enzyme PC4. Selective expression of these mammalian genes is likely to be regulated at the promoter level. Regulation of expression of the various *Dfur1*-encoded isoforms is most likely more complex. Our Northern blot studies show that expression of some mRNA isoforms of *Dfur1* are developmentally regulated and our *in situ* hybridization experiments reveal tissue-specific expression. Therefore, expression of the various *Dfur1* transcripts seems to be controlled both at the promoter level and at the level of RNA splicing.

To elucidate the molecular details of the alternative *Dfur1* transcripts, extensive analysis of cDNA and genomic clones was required. Some aspects of the genomic organization of the *Dfur1* gene, such as the positions of exons 2 and 3, the promoter region and also the nature and size of some introns, remain to be resolved. Genomic walking to close existing gaps in the genomic map of λ clones resulted in the isolation of additional *Dfur1* genomic clones (data not shown). However, these still do not contain the missing exon sequences nor do they fully close the intron gaps. On the basis of *Dfur1* genomic sequences presently available, the size of the *Dfur1* gene is estimated to be >100 kbp. Isolation of YAC clones from region 96D of chromosome 3 of *Drosophila*, at which position *Dfur1* was mapped (Hayflick *et al.*, 1992; Roebroek *et al.*, 1992), is likely to facilitate the isolation of the complete *Dfur1* gene.

As far as functional diversity is concerned, it remains to be established whether or not the *Dfur1*-encoded enzymes are *Drosophila* counterparts of some of the known mammalian enzymes or constitute unique enzymes, the mammalian counterparts of which remain to be identified.

This matter may be resolved by comparing physiological roles of the various enzymes, although knowledge about the functional relevance of discriminating structural features of the carboxy-terminal domains of the various enzymes could be informative too. In the generation of the carboxy-terminal structural diversity in the *Dfur1*-encoded enzymes, single exons are involved; exon 10, encoding the so-called X-domain of Dfurin1-X, and exon 13, encoding the carboxy-terminal domain of Dfurin1-CRR, including a unique cysteine-rich region and a potential transmembrane anchor and cytoplasmic tail, both of which are different from those in the other two *Dfur1*-encoded proteins. It is of interest to note that only furin and the furin-like proteins PACE4, Dfurin1, Dfurin1-X, Dfurin1-CRR and Dfurin2 possess a relatively complex multidomain structure at their carboxy-termini (Figure 7). Proprotein processing enzymes like PC1/PC3, PC2 and PC4 have a less complex carboxy-terminal domain structure. The multidomain structure of furin and the furin-like proteins, which are likely to result from exon shuffling (Doolittle, 1985), suggests some functional relevance of these domains. The apparent conservation during evolution of the cysteine motif in members of the family of subtilisin-like proprotein processing enzymes like furin, PACE4, Dfurin1-CRR (dKLIP-1) and Dfurin2 seem to support this suggestion. The X-domain, which possesses no obvious features pointing towards a particular function, has a somewhat similar position in the Dfurin1-X protein as the cysteine-rich subdomains in furin, Dfurin1-CRR and Dfurin2, and the serine/threonine-rich domain in kexin of *S.cerevisiae*. It is possible that the CRR- and X-domains confer a particular substrate selectivity to these presumed *Dfur1*-encoded enzymes. Alternatively, they could be involved in the regulation of the subcellular distribution or modulation of endoproteolytic activity. Since the X-domain is the only structural difference between Dfurin1 and Dfurin1-X, any difference in functional characteristics between the two enzymes could be attributed to the insertion of this X-domain. The carboxy-terminal region of Dfurin1-CRR is different from those of Dfurin1 and Dfurin1-X, which are identical. In addition to a unique cysteine-rich region, Dfurin1-CRR also has therefore another transmembrane domain and cytoplasmic tail, which might provide Dfurin1-CRR with unique functional specifications.

Also of interest is the observed differential expression of the four *Dfur1* transcripts during development of *Drosophila*, not only the combined levels of activity of all four transcripts, but also the relative levels of each of the four transcripts, especially the 4.0 and the 7.6 kb transcript. It should be remembered that the 4.0 kb transcript encodes the same protein as the 6.5 kb transcript. Our data are not in full agreement with those reported for dKLIP-1 (Hayflick *et al.*, 1992). In that study, no expression of dKLIP-1 was detected in the larval and pupa stages. Our data indicate that the four transcripts and, by deduction, the three *Dfur1*-encoded proteins are expressed during the whole lifespan of *D.melanogaster*. *In situ* hybridization analysis of dKLIP-1 expression (Hayflick *et al.*, 1992) revealed the presence of transcripts in adult CNS, fat body and female reproductive tissue. Furthermore, high maternal expression was found in developing oocytes and nurse cells, indicative of a role in early embryogenesis. From the temporal and spatial expression of dKLIP-1 during embryonic development of *Drosophila* it was concluded that the dKLIP-1 gene is probably involved in early embryogenesis and neurogenesis (Hayflick *et al.*, 1992). However, the alternative splicing of the RNA transcripts of this *Drosophila* gene and the resulting structural diversity in the various corresponding proteins, reported here, place the dKLIP-1 *in situ* hybridization results in a new perspective; the dKLIP-1 probe used in those studies was not informative to the extent that it could not discriminate between the various transcripts. Our RNA *in situ* hybridization studies, in which probes were used that specifically detected mRNAs coding for either Dfurin1-X or Dfurin1-CRR, have provided the first specific insights into topological differences of *Dfur1* expression. Furthermore, the observation that, during *Drosophila* embryogenesis, Dfurin1-X and Dfurin1-CRR mRNAs are expressed in cells belonging to non-overlapping tissue types strongly suggests that the enzymes encoded by them are involved in the processing of different spectra of proproteins or prohormones and support the proposed diversity in their physiological roles in *Drosophila*.

Finally, in the light of the differential expression pattern of the various *Dfur1* transcripts and considering the established structural diversity and the presumed functional diversity, the results of our studies suggest that the various *Dfur1*-encoded enzymes might play a role in *Drosophila* during development. Genetic analysis of this locus, which has been mapped to the 96D region of the third chromosome (Roebroek *et al.*, 1992) could resolve this. Until now, we have studied proprotein processing activity of the *Drosophila* furin-like proteins in experiments in which only heterologous substrates were used as targets. By such an approach, the endoproteolytic cleavage activities with specificity for paired basic residues of most of the recently discovered members of this enzyme family have been elucidated and established. Furthermore, heterologous cleavage conditions and the spectrum of potential substrates can also be defined in this way, as for instance for Dfurin2 of *Drosophila*; although the relevance of this is probably of less importance. Dfurin2 appeared to be capable of correctly processing the precursor of von Willebrand factor expressed either in COS-1 cells under the control of the SV40 late promoter (Figure 11) or in V.V.:T7-infected PK(15) cells under the control of the T7 promoter (data not shown), but it was unable to process the precursor of β_A in PK(15) cells (Figure 12). In the light of all this, it would therefore be of more interest to establish the nature of the *Drosophila* proproteins that are the physiological substrates of these enzymes. Because of the diversity in the *Dfur1* enzymes, it is very possible that together they are the physiological processing enzymes for a wide variety of *Drosophila* proproteins.

Materials and methods

Cell lines

COS-1 cells (SV40-transformed African green monkey kidney cells; American Type Culture Collection CRL 1650) were propagated in Iscove's modified minimal medium (Gibco), supplemented with fetal bovine serum (Gibco) (10% v/v) and antibiotics [penicillin (100 U/ml) and streptomycin (100 μg/ml)]. Pig kidney PK(15) cells were grown in Earle's Minimum Essential medium supplemented with 1% non-essential amino acids, 5% fetal bovine serum and antibiotics.

Molecular probes and hybridization

In the present studies, various molecular probes of *Dfur1* were used. In Table I, detailed specifications of the probes are summarized. Probe pG6 is a 5'-end *Eco*RI–*Bgl*II and probe pG7 a 3'-end *Bgl*II–*Eco*RI DNA

fragment of a previously described *Dfur1* cDNA (Roebroek *et al.*, 1991). It should be noted that pG7 does not contain sequences belonging to exon 10; it does contain, however, a unique stretch of 21 nucleotides belonging to an alternatively spliced exon 12. Probe pG13 is a 5′-end *Eco*RI−*Sal*I DNA fragment of the same *Dfur1* cDNA. The genomic probes pIP44, pIP46 and pIP95 are subclones from *Dfur1* genomic clone λYZ3 and represent, respectively, an internal 1.2 kbp *Xba*I−*Hin*dIII fragment, a 1.5 kbp *Hin*dIII−*Eco*RI fragment and a 0.9 kbp *Hin*dIII−*Kpn*I fragment (Figure 4); the *Eco*RI restriction site in pIP46 is from the vector of λYZ3 and, therefore, not indicated in Figure 4. PCR-derived probes were synthesized using appropriate amplimer sets (17- to 20-mers) and genomic clone λYZ3 (for probes PCR1, PCR2 and PCR3) or cDNA clone pIP56 (for probes PCR4 and PCR5) as template. PCR was performed as described by Marynen *et al.* (1990), except that the annealing temperature was adjusted for the various amplimer sets used. After agarose gel electrophoresis, DNA fragments were isolated using Geneclean II (BIO 101) and subsequently [^{32}P]dCTP-labeled probe was synthesized by standard procedures using Klenow polymerase in combination with random hexanucleotide primers or, as in the case of probes PCR2 and PCR5, in combination with the same primers as used for PCR.

Hybridization of replica filters of cDNA and genomic libraries of *D.melanogaster* was performed as previously described (Roebroek *et al.*, 1986a,b). For hybridization of Northern blots, the procedure described by Church and Gilbert (1984) was used.

RNA in situ hybridization analysis

Whole-mount embryos (0−24 h) of wild-type Canton-S *Drosophila* were prepared, fixed, hybridized and washed according to established procedures (Tautz and Pfeiffle, 1989). For the preparation of the probes, a previously described procedure was used; a saturating amount of random hexamers was used during the randomly primed probe labeling reaction (Yang *et al.*, 1991). Using this procedure, the mean probe length was greatly reduced and the signal to noise ratio greatly increased. As molecular probes, pG13 (detects all four *Dfur1* mRNAs), PCR1 (specific for Dfurin1-X mRNA) and pIP95 (specific for Dfurin1-CRR mRNA) were used. Photographs were taken using DIC optics with a Zeiss Axiophot microscope using a 25 × Neofluar n.a. 0.8 lens.

Construction and screening of a cDNA library of D.melanogaster embryos

In order to isolate cDNA clones corresponding to the four *Dfur1* transcripts, a cDNA library was constructed using 15 μg of total RNA of (2−8 h old) embryos of the Tübingen strain of *D.melanogaster*. Oligo-dT-primed cDNA synthesis using total RNA as template was performed with a cDNA synthesis kit of Pharmacia. To the cDNA synthesized in this way, *Eco*RI−*Not*I adaptors were ligated and the products were cloned into the *Eco*RI site of lambda ZAP II. Subsequently, the recombinant phage DNA was packaged (Stratagene). Of ~2 × 10^6 recombinants, 4 × 10^5 were plated and replica filters (*in duplo*) were screened with the probes pG6, pG7, PCR1, pIP44 and pIP46. Positive clones were plaque purified and, subsequently, pBluescriptSK(−) plasmids carrying the *Dfur1* cDNA inserts were produced by the *in vivo* excision procedure. Of these, clones pIP56, pIP58, pIP62 and pIP63 were analyzed in detail.

RNA isolation and Northern blot analysis

Total RNA was isolated from different developmental stages of *D.melanogaster*, including 2−8 h-old embryos, 22 h-old embryos, larvae of stage 1 (36 h), larvae of stage 2 (60 h), larvae of stage 3 (96 h), pupae and adults. The 2−8 h old embryos were of the Tübingen strain of *D.melanogaster*, whereas all other stages were of the Oregon R strain. Total RNA was isolated using the lithium−urea procedure described by Auffray and Rougeon (1980). A total of 15 μg of total RNA were glyoxylated and size fractionated on a 1% agarose gel and transferred to Hybond-N (procedure as recommended by Amersham Corp.). After hybridization with *Dfur1*-specific probes, the amount and quality of the RNA samples on the blot were checked by hybridization with dhPR21, a genomic ribosomal DNA sequence from *Drosophila hydei* (Huijser and Hennig, 1987).

Nucleotide sequence analysis

Nucleotide sequences were determined according to the dideoxy chain termination method using the T7 polymerase sequencing kit of Pharmacia. The DNA fragments to be sequenced were subcloned in pGEM-3Zf(+), pUC18 or pBluescriptSK(−) and sequenced using standard primers or primers synthesized on the basis of newly obtained sequence data. Nucleotide sequence data were analyzed using the sequence analysis computer programs Genepro (Riverside Scientific) and Intelligenetics (IntelliGenetics, Inc.).

Isolation of Dfur1 genomic clones

λYZ3, λYZ4 and λYZ5, the isolation of which was described previously (Roebroek *et al.*, 1991), contain overlapping *Dfur1* genomic DNA inserts. Because hybridization and nucleotide sequence analysis indicated that these clones did not contain the complete *Dfur1* gene, an EMBL-4 phage library of genomic DNA of *D.melanogaster* (Oregon R strain) was screened with probes pG6 and PCR4 under conditions as described before (Roebroek *et al.*, 1991).

DNA transfection, radiolabeling of cells and immunoprecipitation analysis

To test whether or not the *Dfur1* gene encodes proprotein processing enzymes with cleavage selectivity for paired basic amino acid residues, two types of studies were performed. In one type, *Dfur1* cDNAs and pro-vWF cDNAs, both under control of the SV40 late promoter of the pSVL vector, were co-transfected into COS-1 cells and proprotein processing was assayed by immunoprecipitation analysis. These studies were performed as described before (Van de Ven *et al.*, 1990) and both wild-type pro-vWF and mutant pro-vWFgly763 were used as substrates (Verweij *et al.*, 1986; Voorberg *et al.*, 1990). Initially, the *Dfur1* cDNA inserts were cloned as *Not*I fragments into pSVL (Van de Ven *et al.*, 1990). To facilitate this, a *Not*I site was introduced into the *Sma*I site of pSVL using a *Not*I adaptor with blunt ends; *Not*I insert of pIP56 (Dfurin1-CRR), pIP62 (Dfurin1-X), or pIP63 (Dfurin1) subcloned into pSVL-Not resulting in pIP70, pIP69 and pIP71, respectively. In initial co-transfection experiments, no processing of pro-vWF by the *Dfur1*-encoded proteins could be demonstrated. Subsequent immunoprecipitation analysis revealed that only very low levels of the *Dfur1*-related proteins were synthesized in the COS-1 cells and this might be due to alternative and interfering translational start codons (ATGs) upstream of the presumed translation initiation ATG. To eliminate these ATGs, all present in the exon 1 sequences upstream of a *Swa*I restriction endonuclease cleavage site (nucleotides 493−500 in Figure 2), the *Xba*I−*Swa*I DNA fragments (*Xba*I in polylinker of pSVL 5′ of *Not*I insert) in pIP69, pIP70 and pIP71 were removed. Subsequent filling in of the *Xba*I sticky ends and ligation of these to the *Swa*I blunt ends resulted in the generation of clones pSVLDfur1, pSVLDfur1-CRR and pSVLDfur1-X. The following DNA constructs were studied in the COS-1 co-transfection experiments: (i) pSVLDfur1, which encodes Dfurin1; (ii) pSVLDfur1-X, which encodes Dfurin1-X; (iii) pSVLDfur1-CRR, which encodes Dfurin1-CRR; (iv) pSVLDfur2 (Roebroek *et al.*, 1992), which encodes *Drosophila* Dfurin2; (v) pSVLfur-hu, which encodes human furin; (vi) pSVLvWF (Van de Ven *et al.*, 1990), which encodes wild-type pro-vWF; (vii) pSVLvWFgly763 (Van de Ven *et al.*, 1990), which encodes cleavage mutant pro-vWFgly763.

In studies of the second type, pig kidney PK(15) cells (Huylebroeck *et al.*, 1990) were infected with recombinant vaccinia virus V.V.:T7, which encodes the T7 RNA polymerase (Feurst *et al.*, 1986; Andreasson *et al.*, 1989), and subsequently co-transfected with a DNA construct expressing one of the cDNAs of the *Dfur1* gene together with a cDNA construct expressing bovine proβ_A, the precursor of the β_A-chain of activin, both under the control of the T7 promoter. To express each of the three *Dfur1* cDNAs under the control of the T7 promoter, the *Xho*I−*Not*I DNA fragments of pSVLDfur1-X, pSVLDfur1-CRR and pSVLDfur1 were subcloned in pGEM-11Zf(+) digested with *Sal*I and *Not*I. In the resulting clones, pGEM-Dfur1-X, pGEM-Dfur1-CRR and pGEM-Dfur1, the cDNA inserts were present in the proper orientation. As controls, similar experiments were performed with DNA constructs expressing human *fur* cDNA (pSelect-wt.fur-hu) or *Drosophila Dfur2* cDNA [pGEM-Dfur2; *Dfur2* cDNA insert in pGEM-3Zf(+)] with insert under the control of a T7 promoter. Analysis of labeled polypeptides expressed in the vaccinia virus system was performed as described by Huylebroeck *et al.* (1990). Briefly, subconfluent PK(15) cells were washed with phosphate-buffered saline (PBS) and infected with V.V.:T7 virus (multiplicity of infection 5−10) for 1 h at 24°C; subsequently, medium was removed and appropriate combinations of DNAs were transferred into the cells of lipofection during 16 h at 37°C. Thereafter, cells were pulse-labeled for 30 min with [^{35}S]methionine and chased for 4 h as described previously (Huylebroeck *et al.*, 1990). After 4 h at 37°C, medium was collected and centrifuged, and the supernatant was analyzed by SDS−polyacrylamide gel electrophoresis under non-reducing conditions as described. The following DNAs were used in these studies: (i) pTZ19Rβ13 (Huylebroeck *et al.*, 1990), which consists of a cDNA fragment of 1700 bp containing the complete coding sequences for bovine proβ_A in plasmid pTZ19R (Mead *et al.*, 1986) under the control of the T7 promoter; bovine proβ_A possesses a stretch of five arginine residues at its processing site at position 281 (Huylebroeck *et al.*, 1990); (ii) pSelect-wt.fur-hu, which contains wild-type human *fur* cDNA under the control of a T7 promoter; (iii) pGEM-Dfur1, which contains *Dfur1* cDNA encoding Dfurin1; (iv) pGEM-Dfur1-CRR, which contains the *Dfur1* cDNA encoding

Dfurin1-CRR; (v) pGem-Dfur1-X, which contains *Dfur1* cDNA encoding Dfurin1-X; (vi) pGEM-Dfur2, which contains *Dfur2* cDNA encoding Dfurin2.

Biosynthesis of the furin-like proteins corresponding to the various *Dfur1* mRNA forms and *Dfur2* mRNA was demonstrated in gene transfer studies using immunoprecipitation analysis with a rabbit anti-Dfurin1 or anti-Dfurin2 antiserum. The anti-Dfurin1 antiserum was obtained by immunizing rabbits with a partially purified hybrid protein which consisted of glutathione S-transferase (GST) and the 631 carboxy-terminal amino acids of Dfurin1, and was encoded by pIP81. pIP81 was obtained by cloning a 2.4 kbp *BglII*–*Eco*RI DNA fragment from pIP63 in pGEX-3X digested with *Bam*HI and *Eco*RI. The anti-Dfurin2 antiserum was obtained by immunizing rabbits with a partially purified hybrid protein; it consisted of GST and the 1244 carboxy-terminal amino acids of Dfurin2, and was encoded by pIP87. pIP87 was obtained by subcloning a 4.3 kbp *StuI*–*Eco*RI DNA fragment of pIP67 [*Dfur2* cDNA insert in pGEM-3Zf(+)] in pGEX-1N digested with *SmaI*–*Eco*RI. Reactivity of the antisera towards and specificity for *Dfur1*- and *Dfur2*-encoded furin-like proteins was confirmed by Western blot analysis of lysates of bacteria expressing hybrid proteins containing GST and Dfurin1 or Dfurin2 protein sequences. The antisera appeared to display a low reactivity towards GST. Immunoprecipitation analysis was performed as described before (Van Duijnhoven *et al.*, 1992).

Acknowledgements

We thank T.A.Y.Ayoubi for critical reading of the manuscript and helpful suggestions, D.Huylebroeck, E.W.Beek and P.Groot Kormelink for their contributions to the vaccinia virus experiments, D.Gundacker, G.Vandereyken and N.Dewulf for excellent technical assistance, G.Doucet for synthesizing primers, and P.Kools for the art work. This work was supported by the Inter-University Network for Fundamental Research sponsored by the Belgian Government (1991–1995), the 'Geconcerteerde Onderzoekacties 1992–1996', by the Nationaal Fonds voor Wetenschappelijk Onderzoek (NFWO, Levenslijn), by EC contract BIOT-CT91-0302 and by funding to the Laboratory of Molecular Biology from the Faculty of Medicine of the K.U. Leuven and Innogenetics NV.

References

Andreasson,K.I., Tan,W.W.H., Feurst,T.O., Moss,B. and Loh,Y.P. (1989) *FEBS Lett.*, **248**, 43–47.

Auffrey,C.H. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303–314.

Barr,P.J., Mason,O.B., Landsberg,K.E., Wong,P.A., Kiefer,M.C. and Brake,A.J. (1991) *DNA Cell Biol.*, **10**, 319–328.

Bresnahan,P.A., Leduc,R., Thomas,L., Thorner,J., Gibson,H.L., Brake,A.J., Barr,P.J. and Thomas,G. (1990) *J. Cell Biol.*, **111**, 2851–2859.

Campos-Ortega,J.A. and Hartenstein,V.A. (1985) *The Embryonic Development of Drosophila melanogaster*. Springer Verlag, Berlin.

Chrétien,M. and Li,C.H. (1967) *Can. J. Biochem.*, **45**, 1163–1174.

Creemers,J.W.M., Roebroek,A.J.M. and Van de Ven,W.J.M. (1992) *FEBS Lett.*, **300**, 82–88.

Church,G.M. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.

Docherty,K. and Steiner,D.F. (1982) *Annu. Rev. Physiol.*, **44**, 625–638.

Doolittle,R.F. (1985) *Trends Biochem. Sci.*, **10**, 233–237.

Douglas,J., Civelli,O. and Herbert,E. (1984) *Annu. Rev. Biochem.*, **53**, 665–715.

Feurst,T.O., Niles,E.G., Studier,F.W. and Moss,B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8122–8126.

Fuller,R.S., Sterne,R.E. and Thorner,J. (1988) *Annu. Rev. Physiol.*, **50**, 345–362.

Fuller,R.S., Brake,A.J. and Thorner,J. (1989) *Science*, **246**, 482–486.

Fuller,R.S., Brenner,C., Gluschankof,P. and Wilcox,C.A. (1991) In Jörnvall,H. and Höög,J.O. (eds), *Advances in Life Sciences*. Birkhauser Verlag, Berlin.

Hatsuzawa,K., Murakami,K. and Nakayama,K. (1992) *J. Biochem.*, **111**, 296–301.

Hayflick,J.S., Wolfgang,W., Forte,M.A. and Thomas,G. (1992) *J. Neurosci.*, **12**, 705–717.

Hosaka,M., Nagahama,M., Kim,W.S., Watanabe,T., Hatsuzawa,K., Ikemizu,J., Murakami,K. and Nakayama,K. (1991) *J. Biol. Chem.*, **266**, 12127–12130.

Huijser,P. and Henning,W. (1987) *Mol. Gen. Genet.*, **206**, 441–451.

Huylebroeck,D., Van Nimmen,K., Waheed,A., Von Figura,K., Marmenout,A., Fransen,L., De Waele,P., Jaspar,J.M., Franchimont,P., Stunnenberg,H. and Van Heuverswijn,H. (1990) *Mol. Endocrinol.*, **4**, 1153–1165.

Jin Chan,S., Oliva,A.A., LaMendola,J., Grens,A., Bode,H. and Steiner,D. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 6678–6682.

Kiefer,M.C., Tucker,J.E., Joh,R., Landsberg,K.E., Saltman,D. and Barr,P.J. (1991) *DNA Cell Biol.*, **10**, 757–769.

Marynen,P., Devriendt,K., Van den Berghe,H. and Cassiman,J.-J. (1990) *FEBS Lett.*, **262**, 349–352.

Mead,D., Szczesna-Skorupa,E. and Kemper,B. (1986) *Prot. Engng.*, **1**, 67–74.

Misumi,Y., Oda,K., Fujiwara,T., Takami,N., Tashiro,K. and Ikehara,Y. (1991) *J. Biol. Chem.*, **266**, 16954–16959.

Nagahama,M., Ikemitzu,J., Misumi,Y., Ikehara,Y., Murakami,K. and Nakayama,K. (1991) *J. Biochem.*, **110**, 806–811.

Nakayama,K., Kim,W.S., Torii,S., Hosaka,M., Nakagawa,T., Ikemizu,J., Baba,T. and Murakami,K. (1992) *J. Biol. Chem.*, **267**, 5897–5900.

Roebroek,A.J.M., Schalken,J.A., Bussemakers,M.J.G., Van Heerikhuizen,H., Onnekink,C., Debruyne,F.M.J., Bloemers,H.P.J. and Van de Ven,W.J.M. (1986a) *Mol. Biol. Rep.*, **11**, 117–125.

Roebroek,A.J.M., Schalken,J.A., Leunissen,J.A.M., Onnekink,C., Bloemers,H.P.J. and Van de Ven,W.J.M. (1986b) *EMBO J.*, **5**, 2197–2201.

Roebroek,A.J.M., Pauli,I.G.L., Zhang,Y. and Van de Ven,W.J.M. (1991) *FEBS Lett.*, **300**, 133–137.

Roebroek,A.J.M., Creemers,J.W.M., Pauli,I.G.L., Kurzik-Dumke,U., Rentrop,M., Gateff,E.A.F., Leunissen,J.A.M. and Van de Ven,W.J.M. (1992) *J. Biol. Chem.*, **267**, 17208–17215.

Schalken,J.A., Roebroek,A.J.M., Oomen,P.P.C.A., Wagenaar,Sj.Sc., Debruyne,F.M.J., Bloemers,H.P.J. and Van de Ven,W.J.M. (1987) *J. Clin. Invest.*, **80**, 1545–1549.

Seidah,N.G., Gaspar,L., Mion,P., Marcinkiewicz,M., Mbikay,M. and Chrétien,M. (1990) *DNA Cell Biol.*, **9**, 415–424.

Seidah,N.G., Marcinkiewicz,M., Benjannet,S., Gaspar,L., Beaubien,G., Mattei,M.G., Lazure,C., Mbkikay,M. and Chrétien,M. (1991) *Mol. Endocrinol.*, **5**, 111–122.

Seidah,N.G., Hamelin,J., Gaspar,A.M., Day,R. and Chrétien,M. (1992a) *DNA Cell Biol.*, **11**, 283–289.

Seidah,N.G., Day,R., Hamelin,J., Gaspar,A.M., Collard,M.W. and Chrétien,M. (1992b) *Mol. Endocrinol.*, **6**, 1559–1570.

Smeekens,S.P. and Steiner,D.F. (1990) *J. Biol. Chem.*, **265**, 2997–3000.

Smeekens,S.P., Avruch,A.S., LaMendola,J., Chan,S.J. and Steiner,D.F. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 340–344.

Sossin,W.S., Fisher,J.M. and Scheller,R.H. (1989) *Neuron*, **2**, 1407–1417.

Steiner,D.F. and Oyer,P.E. (1967) *Proc. Natl. Acad. Sci. USA*, **57**, 145–148.

Tautz,D. and Pfeiffle,C. (1989) *Chromosoma*, **98**, 81–85.

Van de Ven,W.J.M., Voorberg,J., Fontijn,R., Pannekoek,H., Van den Ouweland,A.M.W., Van Duijnhoven,J.L.P., Roebroek,A.J.M. and Siezen,R.J. (1990) *Mol. Biol. Rep.*, **14**, 265–275.

Van de Ven,W.J.M., Van Duijnhoven,J.L.P. and Roebroek,A.J.M. (1993) *Crit. Rev. Oncogen.*, **4**, 115–136.

Van den Ouweland,A.M.W., Van Duijnhoven,J.L.P., Keizer,G.D., Dorssers,L.C.J. and Van de Ven,W.J.M. (1990) *Nucleic Acids Res.*, **18**, 664.

Van Duijnhoven,J.L.P., Creemers,J.W.M., Kranenborg,M.G.C., Timmer,E.D.J., Groeneveld,A., Van den Ouweland,A.M.W., Roebroek,A.J.M. and Van de Ven,W.J.M. (1992) *Hybridoma*, **11**, 71–86.

Verweij,C.L., Diergaarde,P.J., Hart,M. and Pannekoek,H. (1986) *EMBO J.*, **5**, 1839–1847.

Voorberg,J., Fontijn,R., Van Mourik,J.A. and Pannekoek,H. (1990) *EMBO J.*, **9**, 797–803.

Watanabe,T., Nakagama,T., Ikemizu,J., Nagahama,M., Murakami,K. and Nakayama,K. (1992) *J. Biol. Chem.*, **267**, 8270–8274.

Wise,R.J., Barr,P.J. Wong,P.A., Kiefer,M.C., Brake,A.J. and Kaufman,R.J. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 9378–9382.

Yang,X., Tong Seow,K., Bahri,S.M., Huat Oon,S. and Chia,W. (1991) *Cell*, **67**, 661–673.

Ying,S.Y. (1988) *Endocr. Rev.*, **9**, 267–293.

Received on August 12, 1992; revised on February 16, 1993