The EMBO Journal vol.9 no.1 pp.217-224, 1990

The cecropin locus in *Drosophila*; a compact gene cluster involved in the response to infection

Per Kylsten, Christos Samakovlis and Dan Hultmark

Department of Microbiology, University of Stockholm, S-106 91 Stockholm, Sweden

Communicated by H.G.Boman

Cecropins are antibacterial peptides that are synthesized in insects as a response to infection. As a first step towards a molecular study of the induction of this response, we have isolated genomic clones that cover the cecropin locus in *Drosophila melanogaster*. This locus was found to be unique, and it was mapped cytologically to the chromosomal location 99E. Sequence analysis showed it to be unusually compact, with three expressed genes and two pseudogenes within less than 4 kb of DNA, and with another homologous region less than 4 kb away. Two of the genes, A1 and A2, encode a product that is identical to the major cecropin from *Sarcophaga peregrina*, while the cecropin encoded by the B gene differs in five positions. Cecropin transcripts appear within an hour after bacteria have been injected into the hemocoel, reach a maximum after 2-6 h, and have almost disappeared again after 24 h. The B gene is induced in parallel with the A genes, but on a lower level. The cecropin genes were also induced when the flies were kept on food with the *Drosophila* pathogenic bacterium *Serratia marcescens* Db10 or its non-pathogenic derivative Db1140.

Key words: antibacterial peptides/cecropin/*Drosophila*/gene family/insect immunity

Introduction

The molecular basis for immune recognition in insects has long remained an unsolved problem. Early attempts to demonstrate circulating immunoglobulin-like molecules failed (Bernheimer *et al.*, 1952). Nevertheless, insects are fully capable of reacting when they are invaded by microbes. Hemocytes in the blood have the capacity to identify foreign objects, and to attack them by phagocytosis or encapsulation (for general reviews, see Götz and Boman, 1985; Ratcliffe *et al.*, 1985; Dunn, 1986).

One very striking aspect of the insect's immune response to a bacterial infection is the induction of a powerful humoral bactericidal system. This consists of several broad-spectrum antibacterial peptides which are rapidly synthesized and released in the hemolymph (reviewed in Boman and Hultmark, 1987). The most potent of the inducible antibacterial peptides are the cecropins. They are amphiphilic peptides, 36-40 amino acids long, and they attack the cell membrane of gram positive as well as gram negative bacteria. Cecropins have been characterized from several lepidopterans as well as from a flesh fly, *Sarcophaga peregrina* (see Boman and Hultmark, 1987; Okada and Natori, 1985a), and in some cases cDNA clones were isolated (von Hofsten *et al.*, 1985; Matsumoto *et al.*, 1986; Lidholm *et al.*, 1987; Dickinson *et al.*, 1988). Genomic clones have been described for a cecropin gene from the moth *Hyalophora cecropia* (Xanthopoulos *et al.*, 1988). Recently, a cecropin was also found in pig intestine, showing that this class of peptides may be of general importance throughout the animal kingdom (Lee *et al.*, 1989).

We are interested in the events that lead to the activation of the immune response in the insect, and we believe that the fruit fly, *Drosophila melanogaster*, would be uniquely suited for a molecular approach to this problem. An inducible immune response was first demonstrated in this species by Boman *et al.* (1972), and has been further characterized by Robertson and Postlethwait (1986) and Flyg *et al.* (1987). Here we describe the cloning of a genomic region from *Drosophila* that contains at least three cecropin genes, and demonstrate the induction of these genes by bacterial infection.

Results

Genomic cecropin clones

A cDNA clone corresponding to the major cecropin in *Sarcophaga*, sarcotoxin IA, was recently isolated by Matsumoto

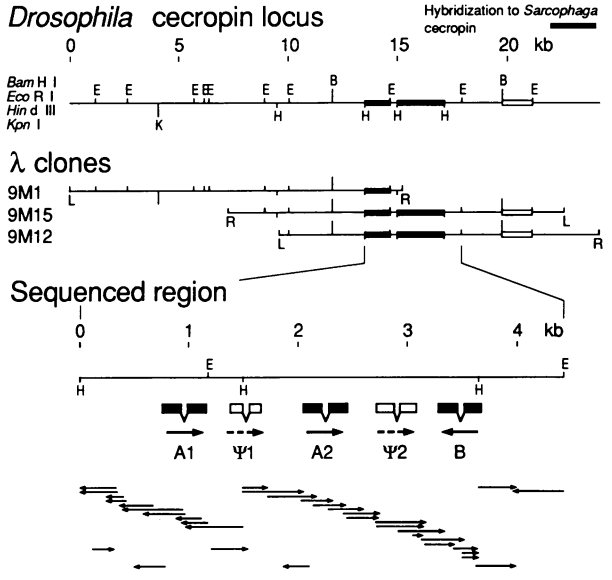

Fig. 1. Organization of the cloned region. The restriction map (**top panel**) is combined from the inserts of the genomic clones shown below (**middle panel**). The positions of the left (L) and right (R) arms of the Charon 4 vector arms are indicated on the inserts. Under the map of the sequenced region (**bottom panel**) the organization of the cecropin genes (filled boxes) and pseudogenes (open boxes) is shown. The sequencing strategy is indicated by thin arrows.

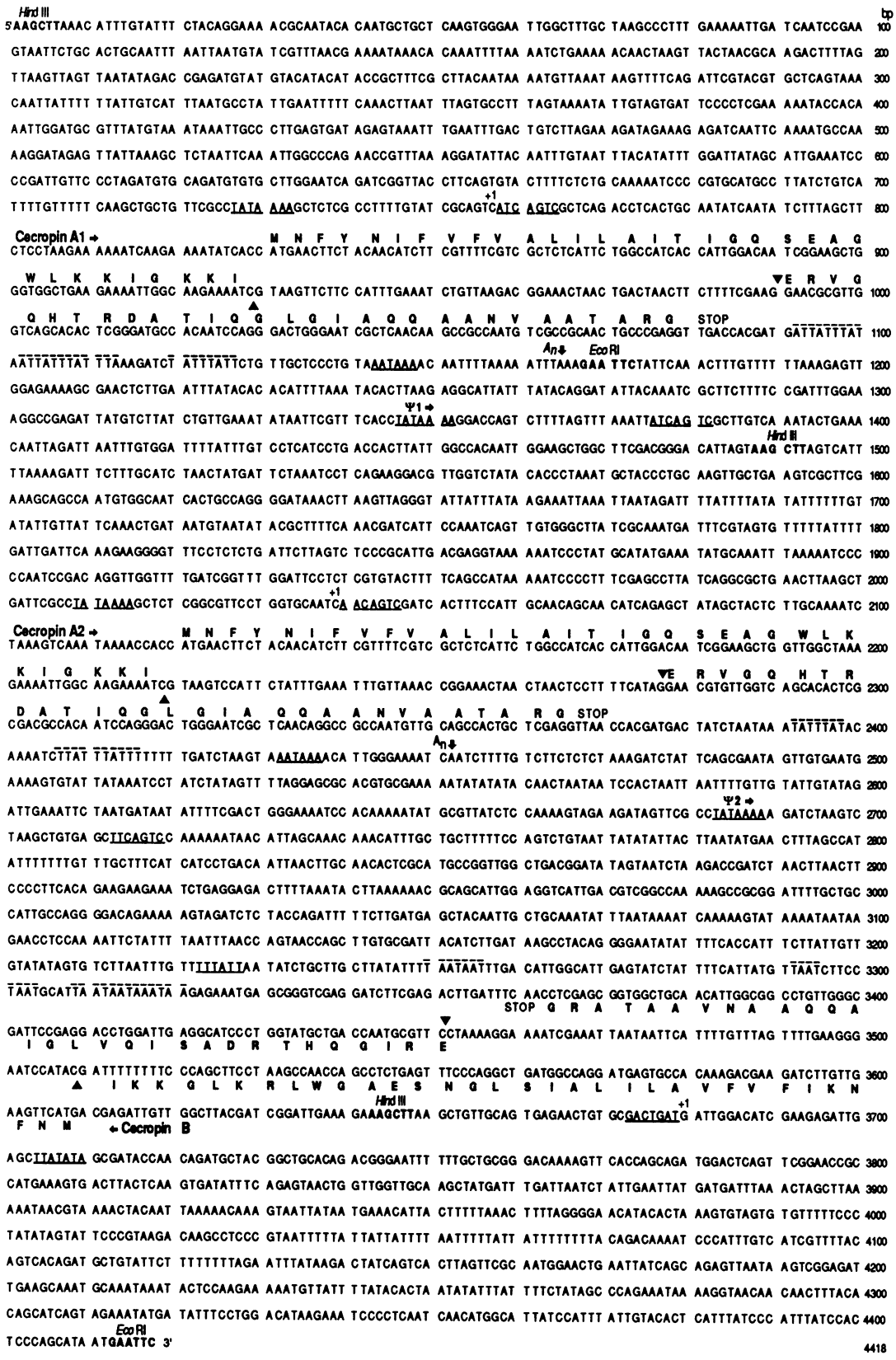

Fig. 2. Sequence of the cecropin locus. Restriction enzyme sites are given above their respective target sequences. Proposed intron splice sites are indicated as ▲ for donor and ▼ for acceptor sites. TATA, CAP, and polyadenylation signal sequences are underlined. The transcription units are indicated by +1 for the starting nucleotides and A_n for the polyadenylation sites, respectively. Cecropins A1 and A2 are translated above their respective exons, whereas cecropin B is translated below and in the opposite direction to indicate its position on the opposite strand. The $AT_{2-3}A$ motif is indicated by a dashed line above the sequence.

et al. (1986). We used the insert from this clone (a kind gift from Professor Natori) as a probe to screen a *Drosophila* genomic library at low stringency. Of 14 positive lambda clones from the primary screen, five apparently contained A-rich stretches, and they no longer hybridized when the poly(A) tail was removed from the probe. Restriction mapping showed that the remaining nine clones fall into three classes, all deriving from a single genomic locus. One clone of each class is represented in Figure 1.

Two restriction fragments in the cloned genomic DNA hybridized strongly to the *Sarcophaga* cecropin probe, a 1.2 kb *Hin*dIII−*Eco*RI, and a 2.2 kb *Hin*dIII fragment. We sequenced a 4.4 kb region (Figure 2) that covers the cross-hybridizing fragments, and found sequence homology to cecropin in five places. Two of these contain stop codons and deletions, and they probably represent pseudogenes, while the remaining regions of homology apparently form three complete genes. In each gene, the coding sequence is interrupted by a single short intron, with consensus splice sites (Shapiro and Senapathy, 1987) at the borders. Two of the genes, A1 and A2, are very similar to each other, and they encode the same peptide, *Drosophila* cecropin A. Furthermore, the predicted amino acid sequence of this cecropin is identical to that of the major cecropin of *S.peregrina*, sarcotoxin IA, except for five amino acid replacements in the signal sequence (Figure 3). However, the similarity to the *Sarcophaga* sequence is less striking on the DNA level. Of the positions where replacements would not affect the amino acid sequence, only 41% are conserved. This figure is not significantly different from the 37% that would be expected from a random assignment of degenerate codons. Outside the coding sequence there is little similarity to the *Sarcophaga* sequence. The B gene is similar to the A genes, with amino acid replacements in ten positions, five of them in the signal sequence (Figure 3). For each of the three genes, a promoter-like sequence with a TATA box (Breathnach and Chambon, 1981) and a cap site (Hultmark *et al.*, 1986) is found upstream of the coding region, and an AATAAA polyadenylation signal (Proudfoot and Brownlee, 1976) can be identified downstream, as indicated in Figure 2.

cDNA clones and transcription units

Preliminary Northern blot analysis, probing RNA from immunized animals with the 7.8 kb *Bam*HI fragment that contains the cecropin homology, demonstrated mRNA with a size of approximately 400 nucleotides. In order to better define what genes are actually transcribed in this region, a cDNA library was constructed from immunized flies, and probed with the 7.8 kb *Bam*HI fragment (Figure 1). Of the 18 positive clones that were analyzed by sequencing, nine corresponded to cecropin A1, seven to A2 and one to a cecropin B transcript. Finally, one clone corresponded to a different gene that was mapped to the region upstream of cecropin A1, and will be described in a later communication. The isolation of the cDNA clones confirms that the A1, A2 and B genes are actively transcribed. For the A1 and A2 genes they also confirm the predicted splicing pattern, and poly(A) tracts define the probable 3′ ends of the mature transcripts, 23 and 21 nucleotides downstream of the AATAAA signal, respectively (Figures 2 and 5). The single B clone isolated contains sequence from the predicted exon 1 only, but extends ~0.8 kb upstream of the coding sequence, long past the promoter-like region. We therefore determined the 5′ ends of the transcripts by extension of

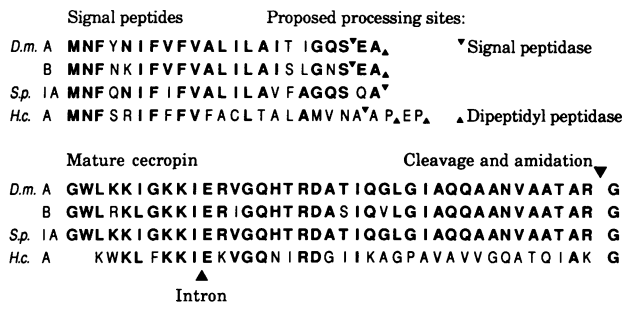

Fig. 3. Deduced amino acid sequences of the *D.melanogaster* (D.m.) cecropins, compared to cecropins from *Sarcophaga peregrina* (S.p.) and *H.cecropia* (H.c.). Identical residues are shown in bold typeface. Formation of the amino terminal by cleavage at the consensus signal peptidase site, and additional processing of amino-terminal dipeptides has been demonstrated in *Hyalophora*. Processing of the carboxy-terminal glycine residue to an amide group is known to occur both in *Sarcophaga* and *Hyalophora*. The position of the intron is known for *Drosophila* and *Hyalophora* only.

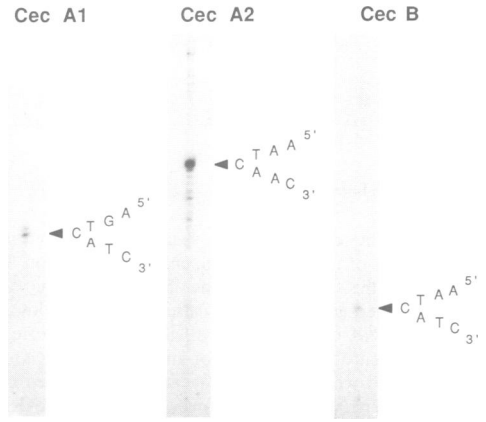

Fig. 4. Determination of transcription start sites by primer extension analysis. The sequence surrounding the first nucleotide is indicated. It was deduced from sequencing ladders primed with the same primers.

synthetic primers with reverse transcriptase (Figure 4). In each case, including the B gene, the major product was extended to a point close to the cap site predicted from the sequence (Figure 2). Thus, each gene is predicted to give rise to a major transcript of about 350 nucleotides, excluding the poly(A) tail, in good agreement with a final size of 400 nucleotides as seen by Northern analysis. The long B gene cDNA clone must originate from a transcript which had been initiated at an alternative upstream promoter. However, this type of transcript is probably rare, and could not be detected in primer extension (Figure 4) or in RNase protection experiments (see below).

Microheterogeneity in the cecropin sequence

The sequences of the individual cecropin cDNA clones showed a considerable microheterogeneity (Figure 5) which we could trace to a heterogeneity in the Canton S stock used to make the cDNA library. We digested genomic DNA from this stock and from a thoroughly inbred Canton S stock (Lüning and Lake, 1985), with two enzymes, *Hha*I and *Mbo*II, whose restriction patterns should be affected by the heterogeneity. Southern blots of the digested DNA were probed with inserts of A1, A2 or B cDNA clones at high stringency. DNA from the inbred Canton S stock gave rise

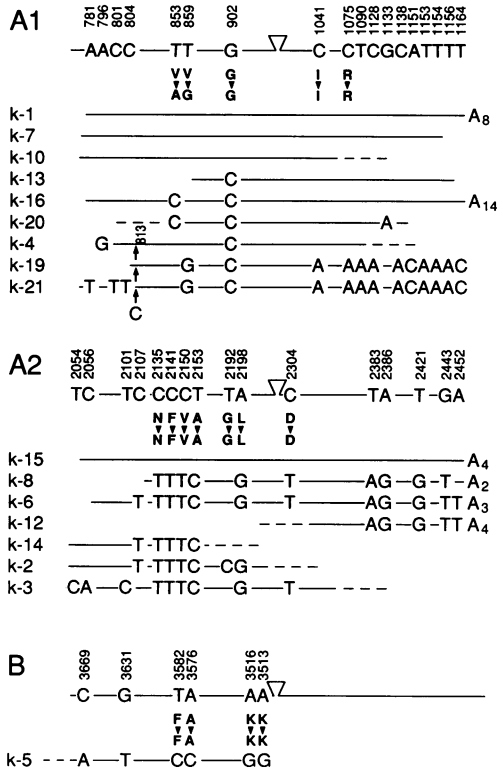

Fig. 5. Variation in the cecropin cDNA sequences. For each gene, the top shows a diagram of the genomic sequence of the Canton S strain, with the residues in the variable positions. The effect of the variability on the predicted amino acid sequence is also indicated. Below, the sequences of the cDNA clones are shown. Solid lines indicate identity to the genomic sequence. The position of an inserted C in A1 is marked. Some inserts were not fully sequenced, as indicated by dashed lines. Additional inserts were seen in clones k-1, 8, 14 and 16. Except in k-8, they were separated from the cecropin insert by *Eco*RI linkers or poly(A) tails, and we consider them to be cloning artifacts. The cecropin B clone k-5 is aberrant, and contains 0.7 kb of genomic sequence upstream of the major transcript start site (see text). At the 3′ end, after exon 1, it contains five additional residues with perfect match to the intron or, with two mismatches, to exon 2.

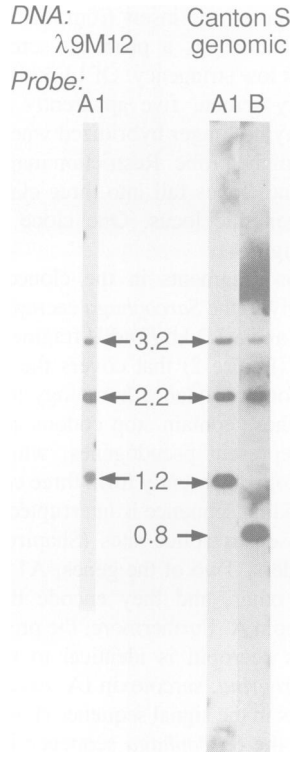

Fig. 6. Genomic fragments with sequence homology to cecropin cDNA. Cloned DNA (0.3 μg), or genomic DNA (4 μg) was digested to completion with *Eco*RI plus *Hin*dIII, separated on 0.4 and 0.6% agarose gels, respectively, blotted, and probed at low stringency with cecropin probes as indicated. The probe for cecropin A1 was the insert of cDNA clone k-7, and for cecropin B that of k-5. Molecular weight markers were run in parallel, and the calculated size of the hybridizing bands is indicated (in kb).

only to the bands predicted from the genomic sequence. In contrast, the stock used to make the library produced a more complex pattern (data not shown). From this we conclude that our fly stock, though it came originally from the same source, contains a mixture of different alleles.

A single cecropin locus at 99E

We further tested for the possible existence of additional cecropin genes in the genome by probing Southern blots of *Eco*RI−*Hin*dIII digested genomic DNA with cecropin A1 or B cDNA probes at reduced stringency (Figure 6). Under these conditions, one new 3.2 kb band appeared in addition to the expected bands at 2.2, 1.2 and 0.8 kb, which are the only bands seen at high stringency. However, the figure shows that a band with the same size also lights up when a digest of the genomic clone 9M12 is probed with A1 at low stringency, indicating that the cross-hybridizing region is within the cloned DNA. By probing further restriction digests of the clones 9M12 and 9M15 (data not shown), we mapped this region to the 1.4 kb *Bam*HI−*Eco*RI fragment indicated with an open box in the top panel of Figure 1.

In conclusion, the cloned cecropin locus appears to be unique, but may contain one or more cecropin genes in addition to the ones we have sequenced. These additional genes may not be closely related to the cecropin A or B genes, since they do not cross-hybridize at high stringency. We mapped the chromosomal position of the cecropin locus by hybridization to salivary gland polytene chromosomes. A single band was seen at position 99E, possibly 99E4-5, in chromosome 3 (Figure 7).

Induction of cecropin mRNA in infected flies

We followed the expression of cecropin mRNA by Northern blotting (Figure 8A). The untreated flies contained undetectable levels of cecropin transcripts. However, soon after the injection of a bacterial suspension, a 0.4 kb cecropin A transcript appears, reaches a maximum after 2−6 h, and then declines. A very faint cecropin B transcript of similar size could also be detected on a Northern blot at the time when cecropin A reached maximum levels (data not shown). As a more sensitive and specific assay, we followed the cecropin B mRNA by RNase protection analysis (Figure 8B). The experiment shows that the cecropin B gene is induced in parallel with the A genes, following similar kinetics. Although the probe used in the cecropin A Northern blot does not distinguish between A1 and A2 (their coding regions have 94% sequence identity), the fact that similar numbers of cDNA clones were isolated for the two genes suggests that they are expressed at comparable levels. Preliminary RNase protection experiments also confirmed that the

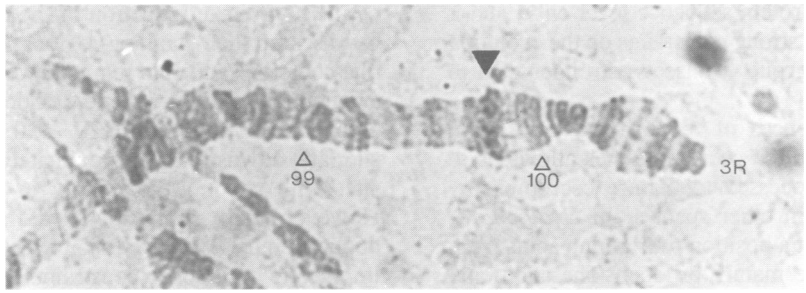

Fig. 7. Mapping of the cecropin locus to cytological position 99E on the right arm of chromosome 3. Squashes of salivary gland chromosomes were probed with a biotinylated cecropin probe as described in Materials and methods. The tip morphology is somewhat atypical (cf. Figure 15 in Lefevre, 1976), but by tracing the chromosome down to the base the identification of 3R is clear (Michael Ashburner, personal communication).

induction of A1 and A2 follow similar kinetics (data not shown).

Injecting bacterial suspensions may represent an artificial way to induce the immune response. As a more natural route of infection, we also investigated the effect of bacteria in the food. For this purpose we used the pathogenic strain *Serratia marcescens* Db10 and its non-pathogenic derivative Db1140 (Flyg and Xanthopoulos, 1983). Figure 8C shows that 24 h after the administration of Db10, when most of the flies were still alive, there was a strong induction of cecropin transcripts. A weaker response was seen with Db1140, which lacks the proteases and envelope determinants that make Db10 resistant to the immune proteins of insects (Flyg and Xanthopoulos, 1983).

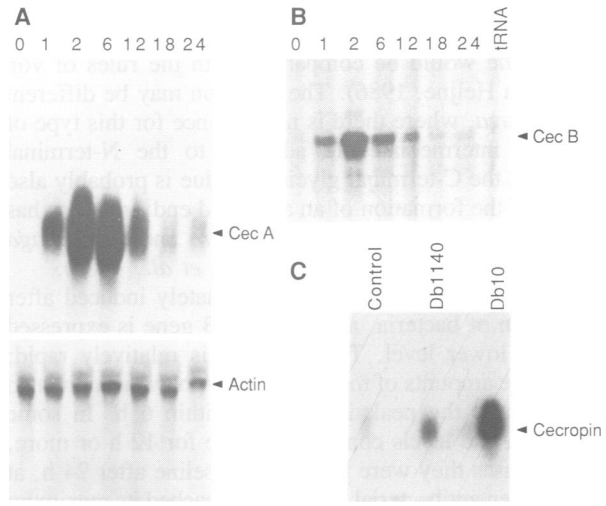

Fig. 8. Expression of the cecropin genes after induction. **Panel A** shows the induction of cecropin A transcripts in untreated flies (0), and 1–24 h after the injection of bacteria, assayed by Northern blotting. The same filter was probed for cecropin A2 with the insert of the cDNA clone k-15 (upper part), and for actin with a 1.6 kb fragment from the 5C actin locus (Fyrberg *et al.*, 1980) to check for equal loading of RNA (lower part). **Panel B**, induction of cecropin B in the same RNA preparations, assayed by RNase protection. **Panel C**, induction of cecropin by bacteria in the food. A Northern blot on total RNA from flies kept on the indicated bacteria was probed with a 1:1 mixture of cecropin A2 and B specific cDNA inserts.

Discussion

We found that the cecropin gene family in *Drosophila* is organized in an unusually tight cluster, with three active genes and two pseudogenes in less than 4 kb of DNA, and with another homologous region within another 4 kb. The genes are also very small, giving rise to transcripts of only 350 nucleotides, each with a single intron of 58–61 nucleotides, near the lower limit for eukaryotic introns (Hawkins, 1988). This compact configuration is in contrast to the situation in *Hyalophora*, where four 12–19 kb lambda inserts were found to contain a single cecropin gene each (Xanthopoulos *et al.*, 1988), and where the intron size was 0.5 kb.

Two pseudogenes, Ψ1 and Ψ2, are interspersed among the other cecropin genes. They are both organized as two blocks of sequence, homologous to exons one and two respectively, separated from each other by a sequence corresponding to the intron. They have diverged considerably from the cecropin genes, only about 50% of the residues are identical. The presence of multiple stop codons and the absence of consensus splice signals make it unlikely that they encode functional products, and we did not find any cDNA clones from these regions. Interestingly, however, both genes have retained a promoter-like region with a TATA box and cap site homology (underlined in Figure 2). Whether these sequences have any functional significance remains to be investigated.

The sequences of the allelic cDNA clones gave us some insight into the variability of the cecropin genes at the nucleotide level. A total of 40 variable positions were found, corresponding to 4.6% of the transcribed DNA. This is a high number, even if a few of the rare exchanges could be due to errors in the reverse transcription reaction. For comparison, in a large study of the *Adh* locus, 1.5% of the transcribed positions were found to be variable (Kreitman, 1983). In the cecropin genes, only two of the variants affect the amino acid sequence (Figure 5). Both of these amino acid replacements represent minor changes in the signal peptide. Thus, in spite of the allelic variability, the protein structure has been conserved.

We were surprised to find that the major cecropins in *Drosophila* and in *Sarcophaga* are identical in sequence. Furthermore, in both species minor cecropins are found that differ very little from the main form. The differences between cecropin A and B in *Drosophila* are largely conservative, and it is too early to speculate on their functional significance at this point. In contrast to the situation in Diptera, the lepidopteran cecropins constitute a relatively heterogeneous group of peptides (see Boman and Hultmark, 1987), rather different from their dipteran counterparts (Figure 3). The homogeneity among dipteran

cecropins is seen at the protein sequence level only, and it appears that there must be strong constraints on the evolution of cecropins within this group of insects, stronger than in Lepidoptera.

The amino acid sequences of the primary translation products from the *Drosophila* cecropin genes allows us to predict how they are processed to mature cecropins, by analogy to the situation in other species. In *Hyalophora*, signal peptidase cleaves at a position four amino acids from the amino terminal of the mature product. The additional amino acids are then removed in two steps by a dipeptidyl peptidase (Boman *et al.*, 1989). A similar pathway may be suggested for the cecropins in *Drosophila*, which are preceded by the dipeptide EA, a known substrate for insect dipeptidyl peptidase-mediated processing (Kreil *et al.*, 1980) (see Figure 3). Although other alternatives were not excluded, a signal peptidase cleavage site upstream of the EA dipeptide would be compatible with the rules of von Heijne (von Heijne, 1986). The situation may be different in *Sarcophaga*, where there is no evidence for this type of processing intermediate. In addition to the N-terminal processing, the C-terminal glycine residue is probably also removed in the formation of an amidated end group, as has been shown for cecropins in *Hyalophora* and *Sarcophaga* (von Hofsten *et al.*, 1985; Matsumoto *et al.*, 1986).

All three cecropin genes are coordinately induced after the injection of bacteria, although the B gene is expressed at a much lower level. The response is relatively rapid; considerable amounts of messenger RNA have accumulated after 1 h, and the peak is reached within 6 h. In some experiments the levels continued to rise for 12 h or more, but in all cases they were back near baseline after 24 h, at the time when antibacterial activity has reached its maximum level (Robertson and Postlethwait, 1986). A conserved AT-rich motif which is believed to mediate selective mRNA degradation (Shaw and Kamen, 1986) has been found in the 3′-untranslated region of several genes involved in the inflammatory response (Caput *et al.*, 1986). We find a similar motif, of the general form $AT_{2-3}A$, repeated in cecropin mRNA (Figure 2), and it may play a role in their rapid turnover.

The induction of an immune response by oral administration of *Serratia* is interesting in several respects. Apparently these successful pathogens have not adopted a strategy in which they avoid triggering the immune response. Instead they rely on their resistance to the antibacterial factors induced. Furthermore, one may ask whether an invasion of the hemocoel is necessary to trigger the immune response. These strains produce a chitinase, but their penetration of the gut is probably slow. Although five cells of strain Db11 (a close relative of Db10) injected into the hemocoel is enough to kill a fly in less than 24 h (Flyg and Xanthopoulos, 1983), it takes nine days until the flies start to die when the same strain is administered by feeding (Flyg and Boman, 1988). Thus, it is uncertain whether any bacteria have penetrated the gut barrier at the time when we observe cecropin induction. Indeed, we occasionally observe a significant level of cecropin expression in our controls (e.g. lane 1 in Figure 8C) that we feel may be related to the presence of bacterial contamination of the food. Finally, the oral induction of the immune response may become a very useful technique, for example in connection with mass-screening of mutations in the immune response or when large numbers of induced flies are needed for biochemical study.

The coordinate induction of the cecropin genes inspired us to search their flanking regions for conserved sequences that could be candidate regulatory elements. In A1 and A2, a continuous region of sequence homology extends from the coding region to a point 120–130 bp upstream of the cap site, and downstream approximately to the polyadenylation site, with about 70% identity in these peripheral parts. However, these similarities could also be due to a recent duplication or gene conversion event, and do not necessarily imply any functional constraints on these sequences. In contrast, we found no obvious sequence homology between the flanking regions of the B gene and those of the A genes, except around the TATA box and the conserved cap sites. This fact raises the question whether there exist independent regulatory elements for each gene, or if they share common elements that coordinately regulate the entire locus. The latter mode of regulation might be facilitated by the very compact configuration of the genes. With access to the cloned genes, it should now be possible to address these questions.

We believe that the most likely function of cecropins and other inducible antibacterial proteins in insects is to act as effector molecules of the immune system, partly analogous to the perforins and the complement system in vertebrates. The major *Sarcophaga* cecropin, identical to *Drosophila* cecropin A, is known to disrupt bacterial cell membranes (Okada and Natori, 1985b). Using chemically synthesized peptides we have confirmed this observation (unpublished data), and are now further investigating the biological activity of the *Drosophila* cecropins. Furthermore, the nature of the primary interactions with bacteria and other foreign bodies that initially trigger the synthesis of these effector molecules should now be accessible to analysis, using the powerful genetic and molecular techniques that are available for *Drosophila*.

Materials and methods

Flies, immunization

Drosophila melanogaster Canton S, obtained from K.G.Lüning, were from a stock that had been inbred by brother–sister mating for at least 350 generations (Lüning and Lake, 1985). We kept them on autoclaved corn meal/yeast food at 25°C with a 10/14 hours light/dark cycle.

To induce cecropins, flies were injected in the abdomen with a suspension of *Enterobacter cloacae* β12 in saline (Flyg *et al.*, 1987). The injected flies were kept in tubes with food. For induction by feeding, a fly tube with 5–10 ml standard fly food was dried for 1–2 h at 37°C. One ml of a suspension of the relevant bacteria, $3-6 \times 10^8$ c.f.u./ml, was allowed to diffuse into the food for 1 h at 37°C. Non-absorbed suspension was poured off. Batches of 25–50 flies were kept on the infected food for 24 h.

Isolation of DNA and RNA from flies

DNA was isolated as described (Bender *et al.*, 1983) and reprecipitated. RNA was extracted from flies frozen in N_2(l) as described (Klemenz *et al.*, 1985) with an additional phenol–chloroform extraction at room temperature. Poly(A)$^+$ RNA was purified on Hybond-mAP paper (Amersham).

Probes

DNA fragments were purified from agarose gels using NA45 DEAE membranes (Schleicher & Schuell Inc.) according to the instructions of the manufacturer. LiCl was used instead of NaCl in the high salt NET buffer. Radioactive probes were made by random priming using the Oligolabeling kit (Pharmacia).

Genomic cloning

10^5 plaque (10 genome equivalents) of a *Drosophila melanogaster* Canton S genomic library (Maniatis *et al.*, 1978) were plated as described by Kaiser and Murray (1985) using Colony/Plaque Screen membranes (New England Nuclear). The filters were probed with the 0.46 kb *Pvu*II–*Pst*I fragment, containing the entire insert of the Sarcotoxin IA cDNA clone pTO19

(Matsumoto *et al.*, 1986). Hybridizations were as described by McGinnis *et al.* (1984), but with 35% formamide, resulting in a very low stringency. All positive clones were rescreened with an internal 276 bp *SpeI–AsuII* fragment, lacking the poly(A) tail.

cDNA library construction and screening

Five μg of poly(A)$^+$ RNA from vaccinated flies was used for cDNA synthesis according to (Gubler, 1988). The blunt-ended cDNA was methylated and ligated to *Eco*RI linkers overnight. Excess of unligated linkers were separated on a Qiagen-tip 5 column (DIAGEN GmbH). The eluted cDNA was ligated to 1.5 μg *Eco*RI-cleaved λ ZAP II arms (Short *et al.*, 1988) and packaged using the Gigapack kit (Stratagene). The yield was 6.8×10^5 recombinant phage. We screened 1.2×10^5 plaques with the 7.8 kb *Bam*HI genomic fragment and 12 clones were isolated with standard procedures. The rest of the library was amplified on *Escherichia coli* BB4 (Stratagene) to a titer of 8×10^{10} p.f.u./ml. After amplifcation 2.5×10^5 plaques were screened with the same probe and 6 positive clones found.

Sequencing

Sequencing was done by the chain termination method using the Sequenase kit (USB).

We sequenced the 1.5 and 2.2 kb *Hin*dIII fragments, subcloned in pTZ18R (Mead *et al.*, 1986). Random size deletions (Barnes, 1987) were made from the *Sac*I site of the vector into the genomic sequence. The following modifications were introduced; KGB buffer (McClelland *et al.*, 1988) was used from the *Sac*I cleavage step and onward and instead of linker tailing the deletion constructs were blunt-end ligated (0.5 Weiss units T4-DNA ligase/μg DNA, 17 h, room temperature). Prior to transformation, unwanted constructs were linearized by *Pst*I cleavage. The 0.8 kb *Hin*dIII–*Eco*RI fragment was sequenced from each end. Most of the genomic sequence was determined on at least two independent templates. Difficult regions were resolved with Taq-polymerase at elevated temperature (Taq-track, Promega), or by sequencing the opposite strand. Sequences 5′ to each gene were also confirmed by sequencing from synthetic primers (see primer extension below).

For the cDNAs, Bluescript plasmids were excised from all positive clones following the protocol of Stratagene. The inserts of k-6, 7, 15 and 21 were sequenced on both strands, other inserts largely on one strand only.

DNA sequence analysis was done on a Macintosh computer using the MacGene Plus software (Applied Genetic Technology, Inc.), and on a microVAX computer with the GCG package from the University of Wisconsin (Devereux *et al.*, 1984).

DNA and RNA analysis

Crude restriction maps of λ9M1 and 12 were obtained by an end-labeling technique. Partially digested DNA was separated on 0.4% agarose gels, blotted, and probed with a 1.49 kb *Bam*HI fragment from the tip of the right arm of the Charon 4 vector. More accurate sizes of the restriction fragments were determined from complete digests, with the enzymes used single or pairwise. The map of λ9M15 was deduced from complete digests only, omitting *Kpn*I.

Low stringency Southern blots were made on Hybond-N membranes (Amersham), and hybridized as described for the screening of the genomic library.

RNA (10 μg) was electrophoresed in a denaturing formaldehyde gel system (Maniatis *et al.*, 1982), blotted overnight on Hybond-N membranes and hybridized overnight according to the manufacturer.

RNase protection. The Bluescript plasmid k-5 with the cDNA insert for the cecropin B gene was linearized with *Xho*I. Transcription from the T3 promoter was done as described by Gilman (1987) resulting in a probe of approximately 900 nucleotides, that protects a 187 nucleotide fragment. Hybridization was at 44°C overnight and RNase digestion at 30°C for 45 min (Gilman, 1987). A DNA sequence ladder was used as marker.

Primer extension. T4 kinase labeled synthetic primers, for A1: GTTCAT-GGTGATATTTTCTTG, for A2: GAAGTTCATGGTGGTTTTATTT and for B: GACGAGATTGTTGGCTTACG (all synthesized by SYMBICOM AB, Umeå, Sweden), were annealed at 25°C to 3 μg poly(A)$^+$ RNA overnight. Extension was done with 50 U M-MuLV reverse transcriptase (Kingston, 1987). Markers were DNA sequencing reactions primed with the same primers.

In situ hybridization to chromosomes

Salivary glands were dissected from late third instar larvae, reared at 18°C. The chromosomes were spread and treated as described by Pardue (1986). The 1.4 kb *Hin*dIII genomic fragment was labeled with biotin-7-dATP using the Nick translation system (both from BRL) and used as probe in a 45% formamide hybridization mix at 37°C. The hybridizing band was visualized with the BluGENE detection kit (BRL), and its position determined by comparison to the maps of Lefevre (1976).

Acknowledgements

We are indebted to Shunji Natori for the gift of the sarcotoxin clone pTO19, Eric Fyrberg for the actin probe, Walter Gehring and collaborators for a sample of the genomic library and for technical advice, Michael Ashburner for help in the interpretation of the banding pattern on salivary gland chromosomes, and K.G.Lüning for advice on the fly work. We also thank Deborah Kimbrell, Hans G.Boman and Ylva Engström for critical comments on the manuscript. The work was supported by grants from the Swedish Natural Science Research Council.

References

Barnes,W.M. (1987) In Berger,S.L. and Kimmel,A.R. (eds), *Methods in Enzymol. Guide to molecular cloning techniques. Vol. 152* Academic Press, Orlando, pp. 538–556.

Bender,W., Spierer,P. and Hogness,D.S. (1983) *J. Mol. Biol.*, **168**, 17–33.

Bernheimer,A.W., Caspari,E. and Kaiser,A.D. (1952) *J.Exp. Zool.*, **119**, 23–35.

Boman,H.G., Boman,I.A., Andreu,D., Li,Z.-q., Merrifield,R.B., Schlenstedt,G. and Zimmermann,R. (1989) *J. Biol. Chem.*, **264**, 5852–5860.

Boman,H.G. and Hultmark,D. (1987) *Annu. Rev. Microbiol.*, **41**, 103–126.

Boman,H.G., Nilsson,I. and Rasmuson,B. (1972) *Nature*, **237**, 232–235.

Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.

Caput,D., Beutler,B., Hartog,K., Thayer,R., Brown-Shimer,S. and Cerami,A. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1670–1674.

Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.

Dickinson,L., Russell,V. and Dunn,P.E. (1988) *J. Biol. Chem.*, **263**, 19424–19429.

Dunn,P.E. (1986) *Annu. Rev. Entomol.*, **31**, 321–339.

Flyg,C. and Boman,H.G. (1988) *Genet. Res. Camb.*, **52**, 51–56.

Flyg,C., Dalhammar,G., Rasmuson,B. and Boman,H.G. (1987) *Insect Biochem.*, **17**, 153–160.

Flyg,C. and Xanthopoulos,K.G. (1983) *J. Gen. Microbiol.*, **129**, 453–464.

Fyrberg,E.A., Kindle,K.L. and Davidson,N. (1980) *Cell*, **19**, 365–378.

Gilman,M. (1987) In Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds), *Current protocols in molecular biology.* John Wiley & Sons, New York, pp. 4.7.1–4.7.8.

Gubler,U. (1988) *Nucleic Acids Res.*, **16**, 2726.

Götz,P. and Boman,H.G. (1985) In Kerkut,G.A. and Gilbert,L.I. (eds), *Comprehensive insect physiology biochemistry and pharmacology.* Pergamon Press, Oxford, pp. 453–485.

Hawkins,J.D. (1988) *Nucleic Acids Res.*, **16**, 9893–9908.

Hultmark,D., Klemenz,R. and Gehring,W.J. (1986) *Cell*, **44**, 429–438.

Kaiser,K. and Murray,N.E. (1985) In Glover,D.M. (ed.), *DNA cloning: a practical approach. Vol. 1*, IRL Press, Oxford, pp. 1–47.

Kingston,R.E. (1987) In Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (eds), *Current protocols in molecular biology.* John Wiley & Sons, New York, pp. 4.8.1–4.8.3.

Klemenz,R., Hultmark,D. and Gehring,W.J. (1985) *EMBO J.*, **4**, 2053–2060.

Kreil,G., Mollay,C., Kaschnitz,R., Haiml,L. and Vilas,U. (1980) *Ann. NY Acad. Sci.*, **343**, 338–346.

Kreitman,M. (1983) *Nature*, **304**, 412–417.

Lee,J.-Y., Boman,A., Sun,C., Andersson,M., Jörnvall,H., Mutt,V. and Boman,H.G. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, in press.

Lefevre,G.,Jr (1976) In Ashburner,M. and Novitski,E. (eds), *The genetics and biology of Drosophila. Vol. 1a*, Academic Press, London, pp. 31–66.

Lidholm,D.A., Gudmundsson,G.H., Xanthopoulos,K.G. and Boman,H.G. (1987) *FEBS Lett.*, **226**, 8–12.

Lüning,K.G. and Lake,S. (1985) *Hereditas*, **102**, 207–217.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular cloning. A Laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connell,C., Quon,D., Sim,G.K. and Efstratiadis,A. (1978) *Cell*, **15**, 687–701.

Matsumoto,N., Okada,M., Takahashi,H., Ming,Q.X., Nakajima,Y., Nakanishi,Y., Komano,H. and Natori,S. (1986) *Biochem. J.*, **239**, 717–722.

McClelland,M., Hanish,J., Nelson,M. and Patel,Y. (1988) *Nucleic Acids*

Res., **16**, 364.
McGinnis,W., Levine,M.S., Hafen,E., Kuroiwa,A. and Gehring,W.J. (1984) *Nature*, **308**, 428–433.
Mead,D.A., Szczesna-Skorupa,E. and Kemper,B. (1986) *Prot. Engineering*, **1**, 67–74.
Okada,M. and Natori,S. (1985a) *J. Biol. Chem.*, **260**, 7174–7177.
Okada,M. and Natori,S. (1985b) *Biochem. J.*, **229**, 453–458.
Pardue,M.L. (1986) In Roberts,D.B. (ed.), *Drosophila: a practical approach*. IRL Press, Oxford, pp. 111–137.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
Ratcliffe,N.A., Rowley,A.F., Fitzgerald,S.W. and Rhodes,C.P. (1985) *Int. Rev. Cytol.*, **97**, 183–350.
Robertson,M. and Postlethwait,J.H. (1986) *Dev. Comp. Immunol.*, **10**, 167–179.
Shapiro,M.B. and Senapathy,P. (1987) *Nucleic Acids Res.*, **15**, 7155–7174.
Shaw,G. and Kamen,R. (1986) *Cell*, **46**, 659–667.
Short,J.M., Fernandez,J.M., Sorge,J.A. and Huse,W.D. (1988) *Nucleic Acids Res.*, **16**, 7583–7600.
von Heijne,G. (1986) *Nucleic Acids Res.*, **14**, 4683–4690.
von Hofsten,P., Faye,I., Kockum,K., Lee,J.Y., Xanthopoulos,K.G., Boman,I.A., Boman,H.G., Engström,Å., Andreu,D. and Merrifield,R.B. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 2240–2243.
Xanthopoulos,K.G., Lee,J.Y., Gan,R., Kockum,K., Faye,I. and Boman,H.G. (1988) *Eur. J. Biochem.*, **172**, 371–376.

Received on September 5, 1989; revised on October 19, 1989

Note added in proof

The sequence data will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence DataBases under the accession number X16972.