Peptidoglycans from bacterial cell walls trigger immune responses in insects and mammals. A peptidoglycan recognition protein, PGRP, has been cloned from moths as well as vertebrates and has been shown to participate in peptidoglycan-mediated activation of prophenoloxidase in the silk moth. Here we report that Drosophila expresses 12 PGRP genes, distributed in 8 chromosomal loci on the 3 major chromosomes. By analyzing cDNA clones and genomic databases, we grouped them into two classes: PGRP-SA, SB1, SB2, SC1A, SC1B, SC2, and SD, with short transcripts and short 5′-untranslated regions; and PGRP-LA, LB, LC, LD, and LE, with long transcripts and long 5′-untranslated regions. The predicted structures indicate that the first group encodes extracellular proteins and the second group, intracellular and membrane-spanning proteins. Most PGRP genes are expressed in all postembryonic stages. Peptidoglycan injections strongly induce five of the genes. Transcripts from the different PGRP genes were found in immune competent organs such as fat body, gut, and hemocytes. We demonstrate that at least PGRP-SA and SC1B can bind peptidoglycan, and a function in immunity is likely for this family.

In recent years, our knowledge of innate immunity in mammals and insects has increased dramatically. Insects and mammals have developed similar mechanisms and molecular pathways to recognize and eliminate invaders (1–4). Insects are therefore good models for innate immunity in vertebrates. In Drosophila, the activation of the immune response is mediated by membrane receptors of the Toll family and by transcription factors related to NF-κB in mammals (5). They signal the presence of invading microorganisms and induce the production of antimicrobial peptides and proteins, more than 500 of which have been described from different animal species.

Although we have a relatively good understanding of the transcriptional activation of the immune genes in insects, this cannot be said of the initial recognition events. As in mammals, lipopolysaccharide, peptidoglycan, and cannot be said of the initial recognition events. As in mammals, transcriptional activation of the immune genes in insects, this from different animal species.

Microorganisms and induce the production of antimicrobial peptides (PGRP, 21, 22). One interesting candidate is the peptidoglycan recognition protein, PGRP (21, 22).

Peptidoglycans are found in the cell walls of almost all bacteria. In many insects, they induce a strong antibacterial response, and they also trigger the activation of prophenoloxidase, leading to the formation of melanin in infected wounds. Humoral PGRPs with a high affinity for peptidoglycan (23) were recently found in the moths Trichoplusia ni (21) and Bombyx mori (22, 24). They are produced in the fat body and in hemocytes (21, 22). The B. mori PGRP is required for the activation of the phenoloxidase cascade (24). A PGRP was also isolated from the cuticle of another lepidopteran, Calpodes ethlius (25), and a homologous expressed sequence tag (EST) found in Anopheles (26). Interestingly, homologs were also found in mouse and humans, where they are strongly expressed in hematopoietic tissue (21).

Recent data from the Drosophila genome project have identified several PGRP homologs. Here we show that they constitute a highly diversified family with at least 12 members. We have characterized these genes in more detail.

Methods

General molecular techniques follow Sambrook et al. (27).

Molecular Cloning of PGRP-SA. A PGRP-SA-specific primer (Dm1: 5′-AGATCTGCTGCGGAGTG) and a vector primer (M13–20) were used to amplify a PGRP-SA cDNA fragment, with a λ DNA library (28) as a template. The PCR product was cloned and sequenced. The insert was used as a probe to screen 5 × 10^5 plaque-forming units of the cDNA library to obtain a full-length cDNA clone according to standard methods. Three identical clones were found, except for length variations. The longest one was named DmVc16.1.9.1.

Sequencing of Plasmid DNA. cDNA clones corresponding to PGRP-LA (GH04960, GH05009, GH110945, GH18280, SD09180), LB (GH11251, GH21008), LC (GH11582, SD04722, LP06740), LD (GH14555, GH13671), LE (GH01554), and SC1B (GH07464) were obtained from the Berkeley Drosophila Genome Project/Howard Hughes Medical Institute EST Project (29). We used the Thermo Sequenase II dye terminator cycle sequencing premix kit from Amersham Pharmacia Biotech and 17- to 18-bases-long primers from Interactiva (Ulm, Germany).

PCR of PGRP Domains. Templates for Northern probes were made with the PCR Supermix kit (GIBCO), 17- to 18-bases-long primers, and genomic fly DNA of the wild-type Canton S strain. Primers were designed to amplify the PGRP domains.

Flies, Cells, and Bacterial Strains. Canton S flies were used for most experiments and kept on a standard mashed-potato diet. Homozygous
gous(3) mbr larvae (30) with blood-cell tumors were used to collect hemocytes. The mbr-2 cell line (31) was grown in Schneider’s medium (44) with 10% FCS. The bacteria Micrococcus luteus M11, Bacillus megaterium Bm11, and Enterobacter cloacae β12 were originally obtained from Hans G. Boman (Karolinska Institute, Stockholm) (see ref. 32 and refs. therein). The Bacillus subtilis wild-type strain was obtained from Sven Bergström (Department of Microbiology, Umeå University).

Collection of Different Developmental Stages; Preparation of Fat Body, Gut, and Hemocytes, and Collection of mbr-2 Cells. All samples were frozen on dry ice and kept at ~80°C until further use. To obtain synchronized developmental stages, newly hatched Canton S flies were kept at 25°C and fed with additional yeast for 6 days before they were allowed to lay eggs on apple juice plates with yeast. Embryos were collected at 5-h intervals. Larvae were kept on fly food. Eggs and larvae were washed in Drosophila Ringer’s solution (182 mM KCl/46 mM NaCl/3 mM CaCl2/10 mM Tris-HCl) and frozen. Prepupae, pupae, and 1-day-old flies were frozen without prior drying.

Fat bodies and intestinal tracts (mid- and hindgut) were dissected from third instar Canton S larvae in ice-chilled Drosophila Ringer’s solution. The separated tissues and remaining carcasses were frozen. For collection of hemocytes, batches of 100 l(3) mbr larvae were washed in Schneider’s medium and rapidly dried on tissue paper. The larvae were opened in 200 μl Schneider’s medium containing a small amount of phenylthiocarbamide on ice. After removal of the dead larvae, the hemocyte suspension was frozen. The mbr-2 cells were centrifuged (500 × g) for 10 min before freezing.

Injection of Bacteria and Peptidoglycan. Newly hatched flies were kept for 6 days at 25°C without additional yeast, washed in 70% ethanol, dried, and injected with approximately 0.1 μl of suspension of M. luteus, B. subtilis, B. megaterium, E. cloacae, or insoluble peptidoglycan isolated from the same bacteria except E. cloacae. Overnight cultures in LB were diluted 1:10 in sterile Ringer’s solution for injection. Insoluble peptidoglycan was prepared according to Morishima (33). A 10-μg/μl stock suspension of peptidoglycan was diluted 1:10 in sterile Ringer’s solution, and approximately 100 ng per fly was injected. Controls were: (i) no injection, (ii) sterile Ringer’s solution, and (iii) sterile Ringer’s solution containing 10% LB.

RNA Preparation, Northern Blot, and Hybridization. Total RNA was prepared with Trizol (GIBCO/BRL) according to the manufacturer. For Northern blots, 15 μg total RNA per lane was run on a 1% agarose gel containing formaldehyde (27). Hybridization was performed under high stringency conditions (50% formamide, 6× sodium saline citrate, 0.15 M NaCl for 30 min at 4°C. After centrifugation at 13,000 × g for 10 min, the supernatant and the pellet were analyzed on a 15% SDs/PAGE gel. The binding of PGRP-SA was assayed by using Coomassie brilliant blue staining of the SDS/PAGE gel. SC1B was detected by Western blot analysis. Proteins were transferred onto a Hybond-C pure membrane (Amersham). The antibody was a rabbit anti-PGRP-SA, and the second antibody and the detection were carried out as before (21).

Results

PGRP Genes Are Widespread Over the Genome. Using tblastn search (35), we routinely checked databases for homologs to the T. ni PGRP. The sequences found were used to search for additional members of the PGRP gene family. We identified 12 homologs in the Drosophila genome, distributed over eight chromosomal loci on the three major chromosomes (Fig. 1). The PGRP-SA gene was first identified in Rph215 flanking sequence (36), the PGRP-SC1B, LA, LB, and LC genes in EST sequences (29), the PGRP-SB1, SB2, SC1A, and SC2 genes in genomic sequence from the Berkeley Drosophila Genome Project, and the PGRP-LD, LC, and LE genes from Celera Genomics (Rockville, MD) (37). We sequenced all relevant cDNA clones from the EST project (29), and for PGRP-SA, we isolated a cDNA clone. By comparison to the genomic sequences, we determined the splicing patterns for all genes except PGRP-SB, SC1A, SC2, and SD. For the PGRP-SB genes, we predicted a splicing pattern, based on conserved ORFs and splice signals. Finally, we assume

Amino Acid Sequence Analysis. After resolution on SDS/PAGE, blotting to Immobilon P membranes (Millipore), and visualization by Coomassie brilliant blue staining, the 20-kDa PGRP-SA band was excised and sequenced by using an Applied Biosystems model 437A Protein Sequencer.

Peptidoglycan-Binding Assay. Insoluble peptidoglycan was prepared from M. luteus as described (34). The peptidoglycan-binding assay was essentially according to Yoshida et al. (24), incubating peptidoglycan with PGRP in 10 mM maleate buffer, pH 6.5/0.15 M NaCl for 30 min at 4°C. After centrifugation at 13,000 × g for 10 min, the supernatant and the pellet were analyzed on a 15% SDS/PAGE gel. The binding of PGRP-SA was assayed by using Coomassie brilliant blue staining of the SDS/PAGE gel. SC1B was detected by Western blot analysis. Proteins were transferred onto a Hybond-C pure membrane (Amersham). The antibody was a rabbit anti-PGRP-SA, and the second antibody and the detection were carried out as before (21).
that PGRP-SC1A, SC2, and SD lack introns, by analogy to PGRP-SC1B. The cDNA sequences identify multiple alternative splicing patterns for the PGRP-LA, LB, and LD genes (Fig. 1).

Near the PGRP-LC gene are four additional blocks of PGRP sequence homology (x, y, z, and w in Fig. 1), which are not represented among the cDNA clones. They could be exons of alternative splice forms of PGRP-LC, independent genes, or pseudogenes.

At Least Two Classes of PGRP Genes. All PGRP genes share a PGRP domain of approximately 160 amino acids, conserved from insects to humans (Fig. 2). Outside the PGRP domain, there is little sequence similarity. Based on their general structure, the PGRP genes can be separated into two classes: the short and the long PGRPs.

The short PGRPs are very similar to the genes previously found in moths and in mammals. The PGRP domain is immediately preceded by a typical signal peptide (processing confirmed for PGRP-SA), and the mature products are predicted to be exported proteins that essentially consist of a single PGRP domain. This class includes PGRP-SA, SB1, SB2, SC1A, SC1B, SC2, and SD. The PGRP-SA gene has two introns in the PGRP domain, and the B. mori and vertebrate PGRP genes also have introns in the corresponding positions (Fig. 2). One of these introns is retained in the PGRP-SB2 gene, whereas all other short genes appear to have lost these introns.

The long PGRP genes, PGRP-LA to LE, are more complex. They all give rise to longer transcripts, in many cases with several splice forms. None of them encodes a typical signal peptide, but PGRP-LA, LC, and LD have a predicted transmembrane region, indicating that they are likely to encode membrane proteins. PGRP-LE has a highly charged N-terminal domain connected to the PGRP domain but no obvious export signal or transmembrane region. Finally, PGRP-LB encodes only the PGRP domain, attached to a relatively long 5′-untranslated region. The splicing pattern in the long genes differs completely from that seen in the short genes. The PGRP domains of the long genes are interrupted by introns at various positions, all different from those seen in the short genes. The exception is PGRP-LB, which shares one intron position with the short genes.

For PGRP-LA, the alternative splice forms encode different proteins. PGRP-LAu has a long N-terminal region with weak sequence similarity to the corresponding region in PGRP-LC. This region is absent in PGRP-LAb. Finally, PGRP-LAc has an alternative splice acceptor site for the second intron, which introduces a stop codon upstream of the PGRP domain. The predicted product would be a short transmembrane protein, similar to PGRP-LAb but without a PGRP domain.

The two PGRP-LD cDNAs have an additional ORF in the 5′-flanking sequence (Fig. 1). The encoded protein is homologous to a human gene of unknown function, PMI (40), and to related ESTs from other species (not shown). Thus, the 3.0- to 3.6-kb cDNA inserts we sequenced appear to correspond to dicistronic transcripts. The situation is further complicated by the fact that the long 3′-untranslated region of PGRP-LD overlaps with a third long ORF on the opposite strand (Fig. 1), homologous to cellular retinaldehyde-binding protein and to α-tocopherol transfer protein.

The previously described human and mouse homologs are typical short PGRPs. However, at least one long gene can also be found in the databases for these species. The human long (L) PGRP gene has
one intron position in common with the short genes. The protein is predicted to have two transmembrane regions (Fig. 2) as well as an N-terminal signal peptide (not shown).

We attempted to deduce the phylogenetic relationship between the different PGRP domains, by using maximum parsimony analysis (41), but most of the relationships found are uncertain and are not supported by bootstrap analysis (Fig. 3), except that the genes within the PGRP-SB and SC clusters are closely related. However, there is a tendency for the long and short genes to form separate monophyletic groups. Again, PGRP-LB is an exception and, depending on which sequences are included in the comparison, it may cluster with the short or with the long genes. Similarly, the position of the vertebrate long PGRP genes is uncertain.

Developmental Expression and Response to Infection of the PGRP Genes. We followed the expression of the PGRP genes during development by using Northern blot analysis (Fig. 4A). PGRP-SC1A and B crosshybridize, because they differ in two bases only, and therefore the panel marked PGRP-SC1 shows the sum of the two transcripts. All short genes gave rise to short transcripts of 0.6–0.8 kb. In addition, we could detect a longer transcript of 1.25 kb from the PGRP-SD gene. The long genes gave rise to several transcripts of varying length. Most of the PGRP genes could be detected in all postembryonic stages. A striking exception is the PGRP-SB2 gene, which is exclusively expressed in prepupa. PGRP-LD and LE transcripts are highly expressed in 0- to 5-h-old embryos and in adult females, as they are probably maternally derived. PGRP-LA, LB, SD, and the longer transcripts of PGRP-LD become active later in the embryonic development. In the early larval stages, additional PGRP genes become constitutively expressed and remain active at variable levels until the adult stage. PGRP-SA, SB1, and SD transcripts were found only at very low levels in uninduced animals, and we were not able to detect PGRP-LC transcripts at all under these conditions (see below).

Because the PGRP genes are thought to play an important role in the innate immune response, we wanted to investigate their inducibility after immune stimulation. We injected 6-day-old flies with different Gram-positive and Gram-negative bacteria and peptidoglycan preparations. Five genes were strongly induced: PGRP-LB, SC1, SC2, and SD. In Fig. 4B, we show the results with whole cells or purified peptidoglycan from B. subtilis (B. megaterium was used for PGRP-LD), but similar effects were seen with all bacteria and peptidoglycans tested. Even sterile Ringer’s solution induced these genes to a variable extent (not shown), and for this reason it is difficult to draw firm conclusions about the specificity of different bacteria or peptidoglycans. Most of the small genes are strongly inducible. PGRP-SC1 and SC2 also show a high constitutive expression. PGRP-SB2 is not expressed at all in adults. Of the long genes, only PGRP-LB is inducible.

PGRP Transcripts Are Expressed in Tissues Known to Play a Role in the Immune Response. We compared the tissue distribution of the different PGRP gene transcripts by using Northern blots of RNA from dissected third instar larvae (Fig. 5). We also included RNA from hemocytes, isolated from the hemocyte-overproducing mutant l(3)mbn, and from the hemocyte-like cell line mbn-2.

For PGRP-SB2, we dissected the corresponding tissues from early prepupae. A few distinct patterns of gene expression can be observed (Fig. 5). First, inducible genes such as PGRP-SB1 and SD are mainly expressed in the fat body. PGRP-SA is also expressed in the fat body, although strong induction is seen only in the remaining carcass, including all epidermal layers. In addition, we detected PGRP-SA and SD in uninduced hemocytes and much more strongly in mbn-2 cells. The high level seen in the mbn-2 cells may be because the cultures of this cell line are often more or less induced. The inducible PGRP-LB gene differs from this pattern. It appears to be ubiquitously expressed with several different transcripts, but we cannot detect it in hemocytes or mbn-2 cells.

Unlike the other short PGRP genes, the PGRP-SC genes are constitutively expressed at high levels in the gut, in addition to their induced expression in the fat body. Yet another pattern is seen for some of the long PGRP genes, including LA, LC, and LD, which appear to be enriched in hemocytes and in mbn-2 cells. Transcripts from the latter genes can also be detected in other tissues, possibly because of associated hemocytes. Particularly PGRP-LD shows a complex pattern with several transcripts of different length. The major 0.65–2.5-kb transcripts are much shorter than the 3.0- to 3.6-kb dicistronic PGRP-LD cDNAs we have sequenced, and they are probably monocistronic. The fact that as many as three independent gene products may come from the PGRP-LD complex makes the interpretation of the Northern pattern difficult for this
gene. Finally, PGRP-LE is very weakly expressed with a 1.4-kb transcript in the gut and a 1.55-kb transcript in hemocytes and in the larval carcass.

**Peptidoglycan Affinity of the PGRP Proteins.** On the basis of previous results with the moth homologs, we assumed that the PGRP domain is a peptidoglycan-binding motif. We tested this directly for two of the Drosophila gene products expressed in the baculovirus system, one of which is expressed in the fat body, PGRP-SA, and one gut specific, PGRP-SC1B (Fig. 6). After incubation with insoluble peptidoglycan, both proteins are quantitatively removed from the free solution. They are retained in the peptidoglycan fraction even after extensive washing, indicating that both PGRP proteins bind peptidoglycan with high affinity. We were not able to isolate any of the long PGRPs in sufficient quantity to test peptidoglycan binding. At least PGRP-LC and PGRP-LA remain insoluble after expression in the baculovirus system, consistent with the proposal that they are membrane proteins.

**Discussion**

The data presented here show that the PGRP family is highly diversified. Previously described PGRPs from insects and vertebrates are small exported proteins, and seven PGRPs in Drosophila belong to this class. Other PGRP genes encode longer proteins, and at least three of them, PGRP-LA, LC, and LD, are likely to be membrane proteins (Fig. 7). A localization in the membrane is likely also for the human long PGRP, although two long internal hydrophobic regions and a likely signal peptide indicate a different membrane topology (Fig. 7). The predicted PGRP-LB and LE have neither obvious signal peptides nor hydrophobic transmembrane regions, and it is possible that they remain in the cytoplasm.

However, these genes give rise to several different transcripts, not all of which correspond to sequenced CDNA. It can therefore not be excluded that signal peptides or transmembrane regions are present in some of these forms.

Several lines of evidence indicate that the divergence of the PGRP family is of ancient origin. First, the Drosophila PGRPs are not obviously more closely related to each other than they are to the vertebrate PGRPs. Only genes within a cluster, such as the PGRP-S family, are similar in sequence and may derive from more recent gene duplications (Fig. 5). Second, the PGRP genes are widely dispersed over the Drosophila genome. The generation of a gene family by duplication typically results in a cluster of closely linked genes, which can remain linked over considerable time before they are broken up by chromosomal rearrangements. The highly dispersed localization is therefore an indication that this gene family is old. Third, the intron positions differ widely between the PGRP genes. All short PGRP genes could in principle be derived from an ancestral pattern with two introns, as in PGRP-SA, assuming that one or both of them may be lost in evolution. PGRP-LB and the human long PGRP also fit this pattern. However, introns are found in five different positions in the other long PGRP genes, none of which corresponds to the introns of the short genes. These facts point to long separate evolutionary histories.

The short PGRPs are a structurally homogenous group and are likely to be exported proteins, secreted in different parts of the body. Three of them, PGRP-SB1, SC2, and SD, are strongly induced in the fat body after infection, which is typical for inducible hemolymph proteins such as the cecropins and many other antibacterial peptides (42). PGRP-SA is also inducible, but not in the fat body or gut. Interestingly, this gene is most closely related to the lepidopteran PGRPs (Fig. 3), at least two of which...
from C. ethlius are known to be expressed in the epidermis (22, 25). The PGRP role in immunity is indicated, at least for PGRP-LB, which is transmembrane proteins. This is consistent with a role as pattern PGRP-LA, LC, and probably also LD are likely to be hemocyte patterns of expression give hints about possible functions. Their general of the short PGRPs are induced after bacterial infection, and of the Drosophila PGRP-lysozymes, and although no enzymatic activity was found in PGRPs dase cascade by peptidoglycan (24). Alternatively, short PGRPs PGRP, which is required for the activation of the phenoloxi- sponse (21, 22). Direct evidence for such a role was shown for pattern recognition molecules, initiating the innate immune re-

Finally, PGRP-LB and LE show a more generalized tissue distribution, and they are likely to be retained in the cytoplasm. A role in immunity is indicated, at least for PGRP-LB, which is induced after infection. These proteins could interact with bacteria that invade the cytoplasmic compartment, or they are exported by alternative routes.

The long PGRPs have complex transcription patterns with several alternative transcripts. For PGRP-LC, there is an interesting possibility that the PGRP homology domains denoted x, y, and w could be spliced to the intracellular and transmembrane parts of PGRP-LC. They all have potential acceptor splice sites in the correct position for an in-frame fusion to PGRP-LC. In this way, the peptidoglycan recognition domain cassette could be exchanged on the PGRP-LC transmembrane anchor. However, we have so far not been able to obtain experimental evidence for this hypothesis.

In conclusion, we have shown that the Drosophila PGRPs constitute a highly diversified gene family, and that much of this diversity probably preceded the protostome–deuterostome split over half a billion years ago. Both short and long PGRPs are present in the human genome, and additional PGRP genes can now be found in the accumulating databases of the human genome project (unpublished observation). It is likely that many, if not all, members of this family play an important role in the interaction with invading bacteria, both in insects and in humans.

Note Added in Proof. The sequences reported in this article have been deposited in GenBank under accession nos. AF207535–AF207542 and AF313389–AF313593. The sequences of PGRP-SB1 (CG9681), PGRP-SB2 (CG9697), PGRP-SC1A (CG14746), PGRP-SC2 (CG14745), and PGRP-SD (CG14746) are predicted by the Berkeley Drosophila Genome Project. The a transcript of PGRP-LA was inferred from partial sequences of cDNA clones GH18280 and GH19045. Full sequencing of GH18280 later showed that the ORF is interrupted in this clone, as in the c transcript.

We thank Michael Williams, Svenja Stoven, Calle Zettervall, Marika Hedengren, and Peter Mellroth for technical advice and helpful discussions. Svenja Stoven for mnb-2 cells, and Anna-Karin Kronhamn for help with the flies. Our initial PGRP-S4 genomic sequence was based on clones isolated by Ulrika Troedsson. This research was supported by grants from the Göran Gustafsson Foundation for Scientific Research, the Swedish Natural Science Research Council, and the Swedish Medical Research Council.