Cloning and Characterization of Four Anopheles gambiae Serpin Isoforms, Differentially Induced in the Midgut by Plasmodium berghei Invasion*

Received for publication, August 9, 2002, and in revised form, November 25, 2002 Published, JBC Papers in Press, November 26, 2002, DOI 10.1074/jbc.M208187200

Alberto Danielli, Fotis K. Kafatos, and Thanasis G. Loukeris

From the European Molecular Biology Laboratory, Meyerhofstrasse 1, Heidelberg 69117, Germany

The genomic locus SRPN10 of the malaria vector Anopheles gambiae codes for four alternatively spliced serine protease inhibitors of the serpin superfamily. The four 40- to 42-kDa isoforms differ only at their C terminus, which bears the reactive site loop, and exhibit protein sequence similarity with other insect serpins and mammalian serpins of the ovalbumin family. Inhibition experiments with recombinant purified SRPN10 serpins reveal distinct and specific inhibitory activity of three isoforms toward different proteases. All isoforms are mainly expressed in the midgut but also in pericardial cells and hemocytes of the mosquito. The cellular localization of SRPN10 serpins is nucleocytoplasmic in pericardial cells, in hemocytes and in a hemocyte-like mosquito cell line, but in the gut the proteins are mostly localized in the nucleus. Although the transcript levels of all SRPN10 isoforms are marginally affected by bacterial challenge, the transcripts of two isoforms (KRAL and RCM) are induced in female mosquitoes in response to midgut invasion by Plasmodium berghei oocenthes. The KRAL and RCM SRPN10 isoforms represent new potential markers to study the oocente midgut invasion process in anopheline mosquitoes.

Successful transmission of malaria parasites to vertebrate hosts requires completion of their sporogonic cycle within the mosquito vector. Therefore, in addition to factors such as feeding behavior or mosquito longevity, vector competence is determined by the ability of the mosquito to support effectively the sporogonic development of a given Plasmodium species. The bird parasite Plasmodium gallinaceum, for example, normally fails to cross the midgut epithelium of the major human malaria vector Anopheles gambiae, because its oocite stages are lysed in the midgut cells of the vector through an unknown mechanism (1). Even in naturally occurring, well adapted mosquito/parasite combinations the parasites experience severe losses. In fact, in the mosquito the parasite faces not only the cellular and humoral immune responses of hemocytes and the fat body but also the well documented local immune responses of the midgut and salivary gland epithelial barriers (2). In the genetically selected A. gambiae strain L3–5 the antiparasitic mechanisms reach an extreme form, in which complete refractoriness is achieved through the melanotic encapsulation of early oocysts (3).

Because the mosquito’s immune system was shown to react to the presence of the parasite, major efforts have been devoted to dissection of the molecular mechanisms underlying these responses (reviewed in Refs. 4 and 5). Continuing efforts are guided by fundamental studies on invertebrate innate immunity performed either in model experimental systems such as Drosophila (6), or in systems that permit consistent biochemical studies such as the Lepidoptera Manduca sexta and Bombyx mori (7) or the horseshoe crab Limulus (8).

Serine proteases play critical roles in the regulation of the invertebrate innate immune responses. In Limulus, for example, cascades of autocatalytically activated proteases triggered either by lipopolysaccharide or glucans converge to activate the hemolymph clotting system, which functions both in coagulation and in defense against pathogens (9). Similarly in M. sexta, the prophenoloxidase cascade (which catalyzes the formation of melanin during the defense reaction and thus plays an important role in the encapsulation of pathogens) is also initiated by proteolytic processing (10).

Not only serine proteases but also their associated regulatory serine protease inhibitors of the serpin superfamily are important modulators of immune responses. The proteases underlying the activation of the clotting cascade in Limulus are known to be inhibited by regulatory serpins (11), as are the Manduca and Hyphantria cunea prophenoloxidase-activating proteases (12, 13). The most dramatic evidence for the regulatory function of serpins comes from Drosophila, where genetic dissection of complex pathways is feasible. The Toll pathway has important roles not only in early development but also in the antifungal response (14). Intracellular signaling is initiated through the binding of the extracellular ligand Spatzle to the Toll receptor, after Spatzle is proteolytically cleaved by Persephone (15). This recently identified protease is thought to be the end point of an antimicrobial recognition cascade that controls activation of Spatzle and, therefore, the Toll pathway-mediated antifungal response. In necrotic mutants (nec) the proteolytic cascade is deregulated, leading to the uncontrollable cleavage of Spatzle and to constitutive expression of the antifungal peptide drosomycin (16). The necrotic (nec) phenotype is associated with two mutations in a serpin gene, the Snp43Ac locus (17). Several immune-responsive serpins have been isolated from various insects, but their specific functions remain unknown (18–21).
Anopheles Serpins Induced by Ookinete Invasion

Anopheles mosquito serpin gene, which gives rise to four alternative spliced serpin isoforms. These isoforms share homology to other insect serpins but also to intracellular cytoprotective mammalian serpins of the ovalbumin family. We characterize biochemically the inhibitory potential of three recombinant serpins and show that the isoforms are expressed in tissues that participate in insect defense reactions, such as hemocytes and the midgut epithelium. Finally, we present evidence that at least two isoforms are transcriptionally up-regulated during parasite passage through the midgut, suggesting that they may be implicated in antiparasitic action or, alternatively, parasite tolerance.

EXPERIMENTAL PROCEDURES

Mosquito and Parasite Techniques—The A. gambiae strain G3 and the hemocyte-like cell line Sua 5.1 were reared and cultured according to previous studies (Refs. 29 and 30, respectively). Mosquito infections were performed as described previously (2), and the mosquitoes were grown until parasitemia reached 0.1% Triton X-100, and gently mixed for 30 min at 4 °C. The procedure resulted in a high solubility of all serpin isoforms. Bacteria were grown on LB (13) and induced with 0.4 mM isopropyl-

β-D-thiogalactopyranoside, and grown for 4 additional hours at 30 °C. The resulting cell lysate was clarified by centrifugation at 5,000 × g. The cell debris was discarded and the supernatant was passed through a French-press, and centrifuged at 20,000 rpm for 40 min. The resulting cell-free supernatant was used to prepare the solubilized cell lysate.

Protease Inhibition Assays—To measure the inhibitory activity of Anopheles serpins, 10 and 100 pmol of purified inhibitor were incubated with 20 μl of a protease solution (10 μg/ml of trypsin or chymotrypsin) in 100 mM Tris, pH 7.8, 50 mM NaCl. The reaction mixtures were incubated at 37 °C for 15 min. After a 15-min incubation period, the reaction was terminated by addition of 100 μl of SDS-PAGE loading buffer (100 mM Tris, pH 6.8, 2% SDS, 5% 2-mercaptoethanol, 10% glycerol). The samples were then boiled for 5 min and loaded onto a 12% SDS-PAGE gel. After electrophoresis, the gels were stained with Coomassie blue or silver stained.

The reaction was allowed to proceed for additional cycles, leading to amplification of both the gene of interest and the internal standard, and permitting normalization of the reaction. After electrophoresis on agarose or polyacrylamide gels, samples were stained with the sensitive SYBR green dye (Molecular Probes) for 45 min and analyzed with a fluorescence imager (Fuji). The primers used for RT-PCRs were as follows.

- Combinant Serpin Expression and Affinity Purification—The ORF encoding the recombinant serpin isoform was amplified from an Anopheles cell-line with primers containing a BamHI (5'-AAAGATCCATG-GCGCAAAATACGACGCT and an SacI overhang (5'-TGGGCTGCA-TGAATTCGGAGGA-ACAT, 5'-CATACATCGTGATGATGATGATG, 5'-GGGTACTACTGTTGGGTGGTTTCC and 5'-TGGGCAAC-CTAATTCCTGCGGTTGGGC) and cloned in a pQE30 receptor plasmid, linearized with the same restriction enzymes. All the expression constructs were sequenced before transformation into Escherichia coli strain TGI. For native protein purification, bacterial cultures were grown until A600 0.7–0.8, induced with 0.4 mM isopropyl-1-thio-

β-D-galactopyranoside, and grown for 4 additional hours at 30 °C. This procedure resulted in a high solubility of all serpin isoforms. Bacteria were lysed in 300 mM NaCl, 50 mM NaH2PO4, 10 mM imidazole, pH 8.0, incubated 40 min on ice with 100 μg/ml lysozyme and 10 μg/ml DNase, passed through a French-press, and centrifuged at 20,000 rpm for 40 min in a refrigerated Beckman ultracentrifuge. The supernatants containing the soluble protein fraction were recovered, incubated with 0.1% Triton X-100, and gently mixed for 30 min at 4 °C. Soluble tagged serpins were purified to over 80% purity with an Ni-NTA column (Qiagen) following the manufacturer's instructions.

RNA Techniques—Total RNA from mosquitoes was isolated using TRIzol (Invitrogen). Northern blot analysis was carried out using hybridization to polytene chromosomes of the A. gambiae Suukok strain (31). The RNA was blotted onto a nylon membrane and probed with a 32P-labeled cDNA probe specific to the ribosomal protein S7 gene. The hybridization was carried out at 65 °C, washed in 2× SSC at 55/60 °C, and 30 s at 72 °C. During a pause at 72 °C, primers complementary to the ribosomal protein S7 gene were added, and the

---

1 The abbreviations used are: RSL, reactive site loop; ORF, open reading frame; GST, glutathione S-transferase; PMSF, phosphate-buffed saline; RT, reverse transcriptase; pNA, p-nitroanilide; Ni-NTA, nickel-nitrilotriacetic acid.
antiserum purification, the RCM-serpin gene was fused in-frame to a GST fusion vector pGEX-5x-1. The construct was sequenced and transformed into an E. coli BL21(DE3) strain for protein expression. The fused GST-serpin was purified through a glutathione-Sepharose 4B matrix (Amersham Biosciences) according to the instructions of the supplier, and eluted fractions were further purified with a Mono Q ion-exchange column (Amersham Biosciences) using a fast protein liquid chromatography device. The peak corresponding to the GST-serpin fusion was collected, and 4 µg of protein was coupled to CNBr-activated Sepharose 4B (Amersham Biosciences). The serpin polyclonal antiserum was incubated with GST-serpin fusion Sepharose overnight at 4 °C, packed in a column, and washed with PBS until A405nm was close to zero. Bound antibodies were eluted with 200 mM glycine-HCl, 200 mM NaCl, pH 2.5, and fractions with A405nm = 0.1 were pooled and tested in immunoblots for activity.

Immunoblotting and Immunocytochemistry—Adult mosquito protein extracts and dissected tissues were prepared essentially as described (29). For immunoblotting, separated polypeptides were transferred to Hybond-P membranes (Amersham Biosciences), blocked with 5% dry milk PBS/0.1% Tween 20, incubated with affinity-purified serpin antiserum (1:1000), and detected by secondary horseradish peroxidase-conjugated goat antibodies (anti-rabbit IgG 1:30000, Promega), using a chemiluminescence kit (ECL detection system, Amersham Biosciences).

For whole mount gut and abdomen immunostainings, dissected tissues were fixed for 45 min in ice-cold 4% paraformaldehyde, washed twice in PBS, blocked for 1 h in PBS (1% bovine serum albumin, 5% normal goat serum, 0.1% Triton X-100 in PBS), and incubated overnight at 4 °C in blocking solution with primary antibodies. Samples were washed three times in PBT for 20 min, incubated with secondary antibodies, and mounted in 80% glycerol or Pro-Long Antifade reagent (Molecular Probes).

RESULTS

Cloning and Characterization of SRPN10—To isolate mosquito serpin genes, degenerate primers were designed based on conserved LVNAVYP and IEVNEGTTEA sequences of both insect and vertebrate serpins (23). A 450-bp fragment of expected size was amplified from an A. gambiae cell line cDNA library and subcloned into a Topo-TA plasmid vector. After confirmation by sequencing, the cloned fragment was used to screen the cell line library, and several positive clones were picked and sequenced. All clones had an identical 5′-end but formed two sequence groups at their 3′-end, indicative of alternative splicing. Therefore, additional clones were analyzed and a fourth instar cDNA larval library was also screened, leading to identification of three distinct full-length serpin clones. They correspond to alternatively spliced 3′-ends, which code for the exposed reactive site loop (RSL) of the serpin.

Because multiple alternatively spliced isoforms had been reported for another insect (32), a mosquito lambda DASH genomic library was screened and used to derive the genomic sequence, thereby determining the full range of possibilities for alternative splicing. Six recombinant bacteriophage clones were isolated by high stringency plaque hybridization with probes corresponding to the sequence of the common serpin backbone and to the RSL of one isoform (CAM), respectively, and analyzed using appropriate restriction enzymes. A single bacteriophage clone spnAG6.1 encompassed the whole gene locus (12551 bp) in three adjacent EcoRI fragments (3917, 1772, and 6862 bp). The genomic structure of the locus is schematically represented in Fig. 1A. It was mapped to subdivision 21F on the left arm of the second chromosome (2L) by in situ hybridization to polytene chromosomes (Fig. 1B). The serpin gene was named SRPN10, and its nucleotide sequence was deposited in the GenBank database with accession number AJ420785 (SP121F).

Several exons encompass the coding region present in the spnAG6.1 genomic clone. Comparison with the three isolated cDNA clones showed that the first three exons (1–3) form the common backbone of all splice variants. Exons R, F, and C code for the distinct isoform specific C-terminal reactive site loops, known to be alternatively spliced from the cDNA studies. Careful search of the genomic sequence revealed the presence of an additional in-frame exon (K), encoding a fourth putative reactive site loop, suggesting an additional splice variant from this locus. Specific primers to exon K were designed and used in combination with a backbone primer to amplify from a larval cDNA library a band, which upon sequencing was shown to correspond to the KRAL transcript, thereby confirming the existence of the fourth serpin isoform. Thus, seven exons of the SRPN10 gene were defined, separated by six introns of highly variable size (85, 114, 125, 108, 2459, and 141 bp, respectively, for introns 1–6). Exon/intron boundaries conform to the GT/AG splice donor/acceptor rule, and all introns were characterized by the presence of a polyadenylyl tract. The fifth intron is large (2459 bp), setting the last two RSL exons F and C far apart from the rest of the gene. This intron sequence was carefully searched for cryptic exons, looking for additional putative RSL sequences, but none were found.

In contrast to the alternatively spliced Mandraea serpin-1 locus (33), each specific RSL exon codes for its own 3′-untranslated region. In-frame translational stop codons are found in each of the RSL exons, followed at a distance ranging from 105 to 257 bp by a consensus polyadenylation site AATAAA, with the exception of exon K in which an obvious polyadenylation sequence is absent. The putative translation initiation site is located in exon 1, with a conserved Kozak consensus (A at −3, G at +4; Ref. 34). Consistent with the computational prediction, a single major transcriptional initiation site was located by primer-extension analysis, 25 nucleotides downstream of a well conserved TATA box (data not shown). Using Genomatix MatInspector software, a 1400-bp region upstream of the transcriptional initiation site was explored for the presence of putative regulatory elements (diagrammed in Fig. 1A). A CCAAT enhancer binding protein β (C/EBPβ) sequence was located 10-bp upstream of the TATA box. Other putative binding sites were found, for morphogenetic factors implicated in embryogenesis (Dorsal (DL), Hunchback (HB), Deformed (DFD), Krupp (KR)), for the ecdysone-inducible DNA binding proteins Broad Complex Z4 (BRZ4) and E74A, for GATA factors, for the Activator Protein 1, and for an activator of the alcohol dehydrogenase gene. Motifs with high similarity to a neuronal cis-element, to nuclear factor AT, to a yeast stress response element, and to binding sites for the ZESTE regulator and c-REL were also observed. This complex organization of the putative SRPN10 promoter region may reflect complex developmental and tissue-specific transcriptional regulation and deserves further investigation. Two REL family factors are also implicated in immune responses in Drosophila (6).

When analyzing the flanking regions of the locus, searching for potential ORFs or additional exons encompassing previously undetected RSLs, sequences with homology to a retroviral element were noted at both ends (depicted with gray bars in Fig. 1A). Upstream of the serpin promoter a coding sequence with significant homology to a retroelement polyprotein was located, whereas downstream of exon C (encoding the last RSL) a sequence was found with similarity to the reverse transcriptase-like protein of the Aedes aegypti LINE transposon Juan-A (35). To check whether the SRPN10 locus might encompass additional, distant RSL exons, we checked the A. gambiae genome sequence that became available on the web after this study was completed (scaffold AAAB01008900.1). None were found in the
24.7-kb genomic sequence spanning from exon C to the next predicted ORF (a putative Zinc-finger DNA-binding protein). Thus, we are confident that the SRPN10 locus encodes only four alternatively spliced isoforms of the serpin superfamily.

Analysis of Sequence Similarities of SRPN10 Serpins—Serpins typically consist of 370–400 amino acid residues and fold into a conserved structure, with an exposed reactive site loop (RSL) located at the C terminus of the molecule. The RSL represents the accessible “bait” region that mimics the cleavage consensus sequence of the target protease. Upon cleavage of the RSL (at the so-called P1–P1’ scissile bond), serpins undergo a drastic conformational reorganization in which part of the RSL and of the adjacent hinge region fold back into the β-sheet of the backbone, conferring a more stable conformation. When conformational change occurs before deacylation of the Michaelis complex, formed between serpin and protease, the latter is trapped in a very stable complex with the inhibitor (24). Thus, serpins act as suicide substrates, and their specificity is largely determined by the scissile bond located within the RSL. The SRPN10 serpins lack a signal peptide and share the first 335 N-terminal residues, differing only in the last 44–60 C-terminal amino acid residues, which encompass the RSL (Fig. 1, C).
and D). The alternative splicing of the RSL coding exons therefore potentially represents a functional multiplication of the inhibitory range of a single serpin gene. The reactive site loops of the four isoforms exhibit sequence conservation both at the nucleotide and the amino acid level. In Fig. 1C we show a matrix comparison of amino acid similarity and identity values between the RSLs of the SRPN10 locus, which are lowest for the KRAL isoform.

BLAST analysis of the whole protein sequences revealed that SRPN10 serpins share high homology to other insect serpins, in particular the Drosophila Sp-4 and Sp-6, with which it forms an orthologous group according to bioinformatic analysis of complete mosquito and fruit fly genomes (36). In addition, SRPN10 shows significant similarity with the mammalian serpins belonging to the neuroserpin and ov-serpin clades (data not shown).

The sequence and conformation of the RSL largely determines the selectivity of inhibition. Thus, sequence alignments (Fig. 1D) and BLAST analysis of the serpin C termini comprising the RSLs were particularly revealing. By these criteria, the four SRPN10 isoforms could be distinguished as follows. The RCM and CAM isoforms resemble not only other insect serpins such as Drosophila Sp-6, B. mori serpin-2 but also neuroserpin and multiple intracellular cytoprotective ov-serpins (e.g. human and bovine PI-6, mouse PTI-6, human PI-8 and PI-9), which are involved in the inflammatory response and in the modulation of pro-apoptotic proteases of epithelial and endothelial tissues as well as of neutrophils and macrophages (37). The Anopheles KRAL isoform and the Drosophila serpin Sp-4 are characterized by multiple basic residues in the scissile bond and further down the C-terminal peptide sequence. Their C terminus is distinguished by a short stretch of residues that closely resemble an endoplasmic reticulum retention signal found also in neuroserpins and in other serpin clades (38). Multiple basic residues are also found in the C-terminal sequences of Hordeum and Arabidopsis serpins, as well as in the chicken MENT protein, which belongs to the intracellular ov-serpins. MENT is known to induce higher order chromatin compaction and is an abundant component in terminally differentiated hematopoietic cells (39). The FCM isoform, characterized by a stretch of hydrophobic residues in the reactive site, resembles most closely mammalian leukocyte elastase inhibitors (intracellular ov-serpin) and two mouse stomach serpins. It presents a Phe residue in the predicted scissile bond sequence, as is the case for MENT and for a viral rabibbivirus virus serpin SPI-1 (the latter is not shown in the figure). Searching for Drosophila orthologs of SRPN10 serpins, we noticed that the Sp-4 gene codes for 10 serpin splice combinations (accession numbers AJ428880 to AJ428889), with the possibility of four alternative RSLs (40). For convenience, we named the Sp-4 RSL variants according to the amino acid residues located at the predicted scissile bond (ASM, TSL, VMA, and KRAL, respectively; Fig. 1D). Protein sequence alignments of the C-terminal regions of SRPN10 and Sp-4 show that Sp-4 KRAL may be considered an ortholog of Anopheles SRPN10 KRAL, unlike the other Sp-4 isoforms. In fact, all the other RSLs diverge significantly from the SRPN10 ones. Finally, additional mammalian and insect serpin RSLs with low similarity to SRPN10 are presented in the lowermost alignment of Fig. 1D.

In conclusion, SRPN10 Anopheles serpins exhibit remarkable sequence homology not only to specific Drosophila serpins but also to a set of vertebrate serpins of the neuroserpin and ov-serpin clades, both at the whole protein level and in the exposed reactive site loop.

**Specific Inhibitory Activity of Anopheles Serpin Isoforms**—Recombinant serpin isoforms with a His tag were produced in a bacterial expression system, and fractions containing a band of the expected size, corresponding to the predicted authentic polypeptides, were purified through their N-terminal His tags on Ni-NTA columns. After purification, bands of expected size were detected on the gel, with the exception of the KRAL isoform that is cleaved under native purification conditions. Lanes marked i are extracts of bacterial cultures induced with 0.4 mM isopropyl-1-thio-β-D-galactoside, whereas p are purified recombinant serpin isoforms. B, SRPN10 protease inhibition assay. The indicated proteases (10 pmol) were incubated for 5 min with an equimolar (10 pmol) or 10-fold higher concentration (100 pmol) of purified serpin. Proteases were trypsin (TRYP), thrombin (THRM), chymotrypsin (CHY), porcine pancreatic elastase (PPE), kallikrein (KAL), human plasmin (PLAS), protease K (ProtK), and subtilisin Carlsberg. Inhibition is reported as percent reduction of the rate cleavage of a protease-specific chromogenic substrate (see “Experimental Procedures”). Inhibition rates are relative to the non-inhibitory serpin ovalbumin and reported values are averages of two independent experiments (different serpin preparations have been used). In each experiment and for each SRPN10 variant/protease combination, diagnostic assays were performed. Therefore, each mean and standard deviation was calculated from four data sets. S.D. < 5%. Inhibition rates of ≤15% were treated as insignificant (- -). Asterisks denote that the KRAL isoform showed no inhibitory activity, apparently due to proteolytic cleavage of the RSL during the purification process.
they might be good inhibitors of elastase, which cleaves preferentially at such target residues. In contrast, only RCM and KRAL have basic residues at the predicted P1 site, making them potential inhibitors of trypsin and thrombin, which prefer Arg or Lys at the N-terminal residue of the target peptide bond. Consistent with these predictions, all the testable isoforms, CAM, FCM, and RCM, are potent inhibitors of elastase, whereas only RCM inhibits in addition trypsin as well as thrombin (Fig. 2B). The pattern of chymotrypsin inhibition is also interesting. This protease tends to cleave at hydrophobic residues, with the catalytic efficiency improving as the side chain increases in size. As predicted, it is efficiently inhibited by FCM, which has a Phe residue in the scissile bond, and least so by the CAM isoform, which is characterized by the presence of small hydrophobic residues (Cys and Ala). Importantly, the CAM isoform proved to be an effective inhibitor of bacterial subtilisin-like proteases, inhibiting both subtilisin Carlsberg and proteinase K. This is intriguing, because microbial subtilisin-like proteases are often associated with pathogenicity.

Expression and Localization of SRPN10 Isoforms—The distribution of the SRPN10-derived transcripts in adult tissues was first monitored by RT-PCR (Fig. 3A), using as template RNAs extracted from dissected adult thoraces, midguts and gut-free abdomens, and a primer pair annealing to the sequences common to all isoforms, or a combination of common and isoform-specific primers. Amplified products were sepa-
rated on agarose gels and visualized with a fluorescence imager after SYBR green staining. Amplification of ribosomal protein S7 transcripts served as an internal standard for sample normalization. Serpin expression is highest in dissected midguts (gt), with weaker expression levels in the thorax (tx) and the gut-free abdomen (ab) (Fig. 3A). This is true both for all isoforms together and for three individual isoforms. The exception is KRAl, which is enriched both in the gut and in the gut-free abdomen. In the adult, RCM is the most abundant of the four isoforms; its transcripts are readily detected with 26 amplification cycles, whereas the other isoforms need four additional cycles to reach comparable amplification levels. Similarly, in the hemocyte-like mosquito cell-line Sua 5.1*, the RCM transcripts are significantly more abundant than the other isoform transcripts (data not shown).

Based on previous experience with the gut-free abdominal fraction, the RT-PCR data suggested that SRPN10 serpin is produced in hemocytes as well as the midgut. To determine the developmental profile as well as the tissue distribution of total SRPN10 serpin, a polyclonal antiserum was raised against a recombinant protein in which amino acid sequences encompassing the common serpin backbone were fused to the His tag.

To determine the developmental profile of serpin levels, total protein was extracted from different stages of the *A. gambiae* G3 strain, and the total protein content of each sample was equalized on the basis of Bradford assays. The samples were then either treated with SDS loading buffer and boiled for 5 min (Fig. 3B) or boiled for 15 min in 8 M urea to promote the complete dissociation and denaturation of serpins and serpin-protease complexes (Fig. 3C) and were then immunoblotted with the serpin antiserum. Serpins are known to bind to their target proteases, forming very stable complexes resistant to SDS denaturation. The predicted molecular masses of the four intact isoforms range from 40 to 42 kDa, and the expected size of each isoform cleaved by its target protease was ~37 kDa. An immunoreactive band at this size range was detected in protein extracts treated with 8 M urea (Fig. 3C). Additional higher molecular size bands were detected in the same extracts in the absence of urea treatment and probably represent serpin-protease complexes as well as uncleaved serpins (Fig. 3B).

SRPN10 serpins are nearly undetectable in early embryos (EE). Their levels increased in late embryos and the first larval stage (L1) and peaked at the fourth larval stage (L4), thereafter declining in the early and late pupal stages (EP and LP). In the adults, 1-day-old sugar-fed females (F) showed higher serpin content than males of the same age (M), and a further increase in serpin levels was recorded in females 24 h after a blood meal (BF).

The antiserum was then used for whole-mount stainings of dissected adult tissues. High serpin levels were detected in selected hemocytes attached to tracheae (Fig. 3D) or to the fat body (Fig. 3E) but not in the fat body itself (Fig. 3E, fb). Consistent with the RT-PCR data, serpins were also localized at high levels in the midgut cells, showing a predominant nuclear localization (Fig. 3F). An additional class of cells that stain strongly with the antibody are the scavenger (detoxifying) pericardial cells (Fig. 3, panels G and H). In all these three cell types serpin was detected in the nucleus. This is better shown in Fig. 3, panels E and H, where an antibody against the serine protease Sp22D (29), which is also expressed in hemocytes and pericardial cells, was used in addition to anti-serpin and anti-histone antibodies. In the pericardial cells serpins were detected in the nucleus (Fig. 3G and arrowheads in Fig. 3H) and distributed throughout the cytoplasm, whereas Sp22D is absent from the nucleus showing a granular cytoplasmic localization (Fig. 3H). In hemocytes Sp22D is restricted to a narrow subpopulation, whereas serpins are present in a broader set of blood cells, only partially overlapping with those which express Sp22D (data not shown). Similarly, in the hemocyte-like cell line Sua 5.1* almost the entire cell population (90%) showed nucleocytoplasmic serpin staining (Fig. 3F), whereas Sp22D was present in secretory vesicles in only 5% of the cells, as previously reported (29).

**Ookinete Midgut Invasion Enhances Transcription from the SRPN10 Locus**—Because SRPN10 serpins are expressed in the midgut and hemocytes of the mosquito, tissues that both have key roles in insect defense against pathogens, we wanted to investigate whether the expression levels of SRPN10 serpins are affected by challenge of mosquitoes with bacteria and *Plasmodium* parasites. Serpins regulating the humoral response pathways in insects (for example the *Drosophila* Srp43Ac) are expected to be secreted in the hemolymph (16). However, aspects of insect defense are cell-mediated, and in *M. sexta* an intracellular hemocyte-specific serpin was shown to be induced upon bacterial challenge (18).

After pricking with a mixture of heat-inactivated Gram+ and Gram- bacteria, mosquito females were dissected and the levels of serpin transcripts were analyzed by RT-PCR using as template total RNA isolated at successive time points after bacterial challenge (Fig. 4A). Although transcriptional up-regulation of the antimicrobial gene defensin was evident 12 h after pricking, the bacterial challenge had no substantial effect on the total level of all SRPN10 serpin transcripts combined and for each serpin variant, as monitored separately by using pairs of common and isoform specific primers (quantified in Fig. 4B). Only the KRAl splice variant showed a modest and transient up-regulation after bacterial challenge (Fig. 4B).
The same type of analysis was applied to female mosquitoes fed on P. berghei-infected mice. Although no apparent differences of serpin transcript levels were detected in mosquitoes 18 and 20 h following an infected blood meal, a remarkable induction was visible at 24 h and persisted at 48 h (Fig. 5A). This response coincides with ookinete invasion and is midgut-specific, because no comparable induction was evident in the gut-free carcasses of the same infected mosquitoes. The results were confirmed by blot analysis of total RNA extracted from naive or infected midguts, dissected 23 h after the blood meal (Fig. 5B).

To check whether the levels of all four isoforms are equally affected during ookinete invasion, pairs of isoform specific primers were used in RT-PCR analysis to amplify isoform specific transcripts (Fig. 5C). Although the levels of FCM and CAM transcripts in mosquitoes fed on infected mice (black bars) did not diverge significantly from the levels present in the RNA of control blood fed mosquitoes (gray bars), the transcript levels for RCM and especially KRAL were markedly enriched 24 h after the infective blood meal and remained higher than the control levels even 48 h later.

To exclude the possibility that this enrichment was due to the parasite presence, by using KRAL- and RCM-specific primers we attempted to amplify any putative contaminating band from genomic DNA or from cDNA derived from in vitro cultured ookinetes (data not shown). No signal was recorded, confirming that RT-PCR results demonstrated an increase in the expression levels of RCM and KRAL isoforms in the infected midguts.

### DISCUSSION

The present work reports the cloning and characterization of an A. gambiae serpin gene (SRPN10) that is transcriptionally regulated during ookinete midgut invasion. Four isoforms were derived from this gene by alternative splicing of exons encoding distinct reactive site loops. This kind of genomic architecture permits multiplication of the functionality of the gene by increasing the number of target-specific bait regions and resembles the organization found in some other insect serpin genes, such as the M. sexta serpin-1 (12 splice variants (12)), the B. mori serpin-1 (2 splice variants (21)), and the Drosophila serpin Sp-4 (4 alternatively spliced RSLs (40)).

Partial clones encoding serpin-like sequences were obtained recently in gene discovery projects, aimed to identify immune-responsive molecules in A. gambiae. Analysis of expressed sequence tags derived from a hemocyte-like cell line library revealed four clone clusters encoding putative serpins (41). One of these clusters, I10, corresponds to SRPN10 serpins. Similarly, a differential display search for immune-responsive genes in the adult females of A. gambiae identified a fragment (AF203339) with a predicted sequence homology to inhibitory serpins (42). Interestingly, in the salivary glands of the mosquito vector A. aegypti, a secreted 48-kDa serpin was identified, which possesses a hemostatic activity and is assumed to inhibit the clotting Factor Xa during mosquito blood feeding (43).

We were able to show experimentally that at least three of the four SRPN10 serpin variants are functional inhibitors of serine proteases. This is consistent with features they share
with inhibitory serpins, which can generally be recognized by a consensus pattern of residues in the hinge region: in inhibitory serpin P15 is usually glycine, P14 is threonine or serine, and residues with short side chains, such as alanine, glycine, or serine usually, occupy positions P12–P9. These residues are essential for inhibitory activity, because they permit a rapid insertion of the RSL into the α/β-sheet, facilitating the conformational change that is necessary for the inhibitory activity of the serpin to be manifested (23). In SRPN10 serpins these essential residues are conserved (shaded gray in Fig. 1D), suggesting the potential for inhibitory activity, which in fact was demonstrated for all three variants that could be tested in vitro. We were unable to test the KRAL serpin isoform, because it is proteolytically cleaved, presumably by endogenous bacterial proteases, during the production and purification procedure. Consistent with the distinct composition of the RSLs, we also showed that the different isoforms exhibit specific protease inhibition spectra in vitro. These biochemical assays did not aim to identify the physiological target(s) of SRPN10 serpins but rather to establish their inhibitory potential. However, we detected high molecular mass complexes in immunoblots using a specific anti-serpin antibody, which are dissociated under harsh denaturing conditions, suggesting that also in vivo SRPN10 serpins are associated with target proteases.

The similarity of SRPN10 serpins to mammalian ov-serpins, both in the whole protein sequence and in some cases in the RSL amino acid composition, leads to intriguing speculations as to their physiological role. Ov-serpins reside in the cytosol and/or in the nucleus of protease-secreting cells, including cytotoxic lymphocytes, monocytes, and epithelial and endothelial cells (44). The physiological role of ov-serpins is still emerging, but for many members of the family a cytoprotective role is envisaged and thought to be exerted through the modulation of pro-inflammatory and pro-apoptotic proteases (24).

Midgut epithelial cells invaded by ookinetes show features indicative of apoptosis, such as loss of cell contacts, genomic DNA fragmentation, and sometimes caspase activation (45, 46). In this context, up-regulation of the inhibitory SRPN10 serpin gene may reflect the activation of anti-apoptotic or cytoprotective mechanisms during ookinete invasion.

Alternatively, the inhibitory activity of the CAM isoform against two distinct bacterial subtilisin-like proteases may support another working hypothesis, according to which SRPN10 serpin isoforms may inhibit ooinete-derived proteases. It is known that ookinetes secrete Sub2, a subtilisin-like protease, during the midgut invasion process (45). Because a pivotal role in red blood cell invasion is predicted for subtilisins secreted by the merozoite stages (47), a similar role of Sub2 during ookinete midgut invasion is an intriguing hypothesis. Provided that functional, purified enzyme becomes available, it would be interesting to test the potential effect of all four SRPN10 serpins toward Sub2. Production and testing of the KRAL isoform would be of special interest, because it is strongly up-regulated during midgut invasion.

In agreement with the lack of an obvious signal peptide, we demonstrated by immunofluorescence that Anopheles serpins have an intracellular nucleocytoplasmic localization, principally in midgut cells, i.e. scavenger pericardial cells and hemocytes, which are well known to mediate, respectively, epithelial and cellular immune responses in the mosquito. Bacterial challenge elicits only marginal up-regulation of serpin transcripts (particular of the KRAL isoform), in contrast to the immune-responsive Spn43Ac Drosophila serpin (16, 17) and to the hemocyte-specific Manduca serpin-2 (18).

A remarkable property of SRPN10 is that two of its isoforms are transcriptionally up-regulated upon ookinete midgut invasion. This differentiates SRPN10 serpins from other described markers such as, nitric-oxide synthase, defensin, and granzyme B, which are transcriptionally regulated by both ookinete invasion and bacterial challenge (2, 5, 48, 49).

Marker genes that are specifically regulated during ookinete invasion are particularly valuable as tools to dissect and compare the physiological responses triggered in the vector when infected with different Plasmodium species or strains. A fragment encoding a gene with sequence similarity to α2-macroglobulin was shown to respond strongly to malaria parasite infection and not to bacteria (42). Several genes that are differentially regulated and may be involved in the defense reaction of the A. gambiae midgut toward P. falciparum have been recently isolated by differential display (50). Curiously, SRPN10 serpins were not among them. This might be due to the different combination of experimental organisms (A. gambiae and P. falciparum) or to the low midgut infection rates in that study (not exceeding 15 oocysts per infected midgut). In our experimental combination (A. gambiae and P. berghei), high infection rates were achieved. Additional studies are necessary to distinguish whether SRPN10 serpin up-regulation takes place only in cells that are invaded by the parasite or is a general response of the midgut after heavy infection. Our highly specific α-SRPN10 antibody is a very valuable tool for such studies.

Only two of the serpin variants (the ones that utilize the most upstream alternative exons) are enriched during ookinete midgut invasion. Combined with the results of primer extension experiments that indicate lack of alternative promoters, these observations point to regulation at a step other than transcriptional initiation. The step in question may affect transcriptional termination, splicing, or relative stability of the mRNAs. The existence of distinct 3′-untranslated regions for each splice variant might be associated with differences in transcriptional termination or mRNA stability. Alternatively, ookinete invasion could result in preferential splicing of KRAL- and RCM-serpin variants: cells penetrated by the parasite show apoptotic phenotypes (45, 46), and it has been reported that cell death can affect profoundly the splicing machinery, favoring maturation of distinct gene products through specific responsive elements present in their introns (51, 52).

Understanding the complex regulation of SRPN10 serpin expression requires systematic studies, starting with the functional dissection of the promoter region, which is characterized by the presence of multiple regulatory elements, and proceeding with analysis of pre-mRNA transcription and in vivo processing and stability. In addition, the effects of purified SRPN10 serpin isoforms on in vitro cultured ookinetes, or ookinete invasion in transgenic midguts overexpressing or inhibiting specific isoforms, will be necessary to clarify whether SRPN10 promotes or inhibits ookinete invasion.

Acknowledgments—We thank Claudia Blass for cytogenetic mapping of SRPN10 cDNA and Belen Minana for assistance with DNA sequencing of the SRPN10 locus. We are grateful to Prof. V. Scarlato for support (to A.D.) during the writing of the article and to Prof. H. Jiang, Dr. E. Levashina, Dr. G. Christophides, and Dr. G. Lycett for fruitful discussions.

REFERENCES

1. Vernick, K. D., Fujioka, H., Seeley, D. C., Taudler, B., Aikawa, M., and Miller, L. H. (1995) Exp. Parasitol. 80, 585–595
2. Dimopoulos, G., Seeley, D., Wolf, A., and Kafatos, F. C. (1998) EMBO J. 17, 6115–6123
3. Collins, F. H., Sukai, R. K., Vernick, K. D., Paskewitz, S., Seeley, D. C., Miller, L. H., Collins, W. E., Campbell, C. C., and Gwadz, R. W. (1986) Science 234, 607–610
4. Dimopoulos, G., Muller, H. M., Levashina, E. A., and Kafatos, F. C. (2001) Curr. Opin. Immunol. 13, 79–86
5. Barillas-Mury, C., Wizel, B., and Han, Y. S. (2000) Insect. Biochem. Mol. Biol. 30, 429–442
Anopheles Serpins Induced by Ookinete Invasion

6. Khush, R. S., Leulier, F., and Lemaitre, B. (2001) Trends Immunol. 22, 260–264
7. Kanost, M. R. (1999) Dev. Comp. Immunol. 23, 291–301
8. Muta, T., and Iwanaga, S. (1996)Curr. Opin. Immunol. 8, 41–47
9. Iwanaga, S., Kawahata, S., and Muta, T. (1996)J. Biochem. (Tokyo) 123, 1–15
10. Kanost, M. R., Jiang, H., Wang, Y., Yu, X. Q., Ma, C., and Zhu, Y. (2001)Adv. Exp. Med. Biol. 484, 319–328
11. Bergner, A., Muta, T., Iwanaga, S., Beisel, H. G., Delotte, R., and Bode, W. (1997)Biol. Chem. 378, 283–287
12. Jiang, H., and Kanost, M. R. (1997)J. Biol. Chem. 272, 1082–1087
13. Park, D. S., Shin, S. W., Hong, S. D., and Park, H. Y. (2000)Mol. Cells 9, 973–983
14. Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M., and Hoffmann, J. A. (1996)Cell 86, 105–114
15. Ligoxygakis, P., Pelte, N., Hoffmann, J. A., and Reichhart, J. M. (2002)Science 297, 114–116
16. Levashina, E. A., Langley, E., Green, C., Gubb, D., Ashburner, M., Hoffmann, J. A., and Reichhart, J. M. (1999)Science 285, 1917–1919
17. Green, C., Levashina, E., McKimmie, C., Dafton, T., Reichhart, J. M., and Gubb, D. (2000)Genetics 156, 1117–1127
18. Gan, H., Wang, Y., Jiang, H., Mita, K., and Kanost, M. R. (2001)Insect Biochem. Mol. Biol. 31, 887–898
19. Han, J., Zhang, H., Min, G., Klemier, D., and Hashimoto, C. (2000)FEBS Lett. 468, 184–198
20. Shin, S. W., Park, S. S., Park, D. S., Kim, M. G., Kim, S. C., Brey, P. T., and Park, H. Y. (1998)Insect Biochem. Mol. Biol. 28, 827–837
21. Sasaki, T. (1991)Eur. J. Biochem. 202, 255–261
22. Potempa, J., Kornus, E., and Travis, J. (1994)J. Biol. Chem. 269, 15957–15960
23. Irving, J. A., Pike, R. N., Lesk, A. M., and Whisstock, J. C. (2000)Genome Res. 10, 1845–1864
24. Silverman, G. A., Bird, P. L., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G., Irving, J. A., Lomas, D. A., Luke, C. J., Moyer, R. W., Pemberton, P. A., Remmler-O'Donnell, E., Salvesen, G. S., Travis, J., and Whisstock, J. C. (2001)J. Biol. Chem. 276, 33293–33296
25. Dahlen, J. R., Foster, D. C., and Kiesler, W. (1997)Biochem. Biophys. Res. Commun. 238, 329–333
26. Dahlen, J. R., Foster, D. C., and Kiesler, W. (1998)Biochem. Biophys. Res. Commun. 244, 172–177
27. Schick, C., Bromme, D., Bartuski, A. J., Uemura, Y., Schechter, N. M., and Silverman, G. A. (1998)Proc. Natl. Acad. Sci. U. S. A. 95, 13465–13470
28. Schick, C., Pemberton, P. A., Shi, G. P., Kamachi, Y., Cataltepe, S., Bartuski, A. J., Gornstein, E. R., Bromme, D., Chapman, H. A., and Silverman, G. A. (1998)Biochemistry 37, 5258–5266
29. Danielli, A., Loukeris, T. G., Lagueux, M., Muller, H. M., Richman, A., and Kafatos, F. C. (2000)Proc. Natl. Acad. Sci. U. S. A. 97, 7136–7141
30. Muller, H. M., Dimopoulos, G., Blass, C., and Kafatos, F. C. (1999)J. Biol. Chem. 274, 11727–11735
31. Kumar, V., and Collins, F. H. (1994)Insect Mol. Biol. 3, 41–47
32. Jiang, H., Wang, Y., Kanzok, S. M., Echle, H., Barillas-Mury, C., and Hoffmann, J. A. (1997)Insect Biochem. Mol. Biol. 27, 693–699
33. Jiang, H., Wang, Y., Huang, Y., Mulnix, A. B., Kandel, J., Cole, K., and Kanost, M. R. (1996)J. Biol. Chem. 271, 28017–28023
34. Kozak, M. (1984)Nucleic Acids Res. 12, 857–872
35. Bensaad-Merchermek, N., Salvador, J. C., and Mouches, C. (1994)Genetica 93, 139–148
36. Christophides, G. K., Zbohonov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P. T., Collins, F. H., Danielli, A., Dimopoulos, G., Hetero, C., Hoa, N. T., Hoffmann, J. A., Kanzok, S. M., Letunic, I., Levashina, E., Loukeris, T. G., Lycett, G., Meister, S., Michel, K., Motta, L. F., Muller, H.-M., Ohta, M. A., Paskewitz, S. M., Reichhart, J. M., Rihetsky, A., Truexler, L., Vercun, K. D., Volau, D., Vola, J., von Mening, C., Xu, J., Zheng, L., Bork, P., and Kafatos, F. C. (2002)Science 298, 159–165
37. Bird, P. I. (1998)Results Probl. Cell Differ. 24, 63–89
38. Atchley, W. R., Lokot, T., Wollenberg, K., Dress, A., and Ragg, H. (2001)Mol. Biol. Evol. 18, 1502–1511
39. Grigoryev, S. A., Bednar, J., and Wodcock, C. L. (1999)J. Biol. Chem. 274, 5626–5636
40. Kruger, O., Ladewig, J., Koster, K., and Ragg, H. (2002)Gene (Amst.) 293, 97–105
41. Dimopoulos, G., Casavant, T. L., Chang, S., Scheetz, T., Roberts, C., Donohue, M., Schultz, J., Benes, V., Bork, P., Ansoor, W., Soares, M. B., and Kafatos, F. C. (2000)Proc. Natl. Acad. Sci. U. S. A. 97, 6819–6824
42. Ohusui, F., Xu, J., Niare, O., Narayanan, R., and Vercun, K. D. (2000)Proc. Natl. Acad. Sci. U. S. A. 97, 11397–11402
43. Stark, K. E., and Aam, A. A. (1998)J. Biol. Chem. 273, 20802–20809
44. Bird, P. I. (1999)Immunol Cell Biol. 77, 47–57
45. Han, Y. S., Thompson, J., Kafatos, F. C., and Barillas-Mury, C. (2000)EMBO J. 19, 6030–6040
46. Zieler, H., and Dvorak, J. A. (2000)Proc. Natl. Acad. Sci. U. S. A. 97, 11516–11521
47. Chitnis, C. E., and Blackman, M. J. (2000)Parasitol. Today 16, 411–415
48. Dimopoulos, G., Richman, A., Muller, H. M., and Kafatos, F. C. (1997)Proc. Natl. Acad. Sci. U. S. A. 94, 11508–11513
49. Luckhart, S., Vlodovitz, Y., Cui, L., and Rosenberg, R. (1998)Proc. Natl. Acad. Sci. U. S. A. 95, 7500–7505
50. Bonnet, S., Prevot, G., Jacques, J. C., Boudin, C., and Bourguinon, C. (2001)Cell Microbiol. 3, 449–458
51. perch, P., Puig, O., Kesler, N., Martinez, C., Granneman, S., Seraphin, B., Anderson, P., and Valcarcel, J. (2000)Mol. Cell 6, 1089–1098
52. perch, P., and Valcarcel, J. (2001)Apoptosis 6, 483–468