The EMBO Journal vol.12 no.7 pp.2891 – 2900, 1993

Members of a trypsin gene family in *Anopheles gambiae* are induced in the gut by blood meal

Hans-Michael Müller, Julian M.Crampton[1], Alessandra della Torre, Robert Sinden[2] and Andrea Crisanti[3]

Università 'La Sapienza', Istituto di Parassitologia, P.le Aldo Moro 5, 00185 Rome, Italy, [1]Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK and [2]Department of Biology, Imperial College of Science, Technology and Medicine, Prince Consort Road, London, UK

[3]Corresponding author

Communicated by C.Louis

Serine proteases are among the enzymes that play a crucial role during the digestion of the blood meal in the gut of mosquitoes. The identification of the corresponding genes would have important implications for the control of mosquitoes and mosquito-borne diseases. Analysis of the genomic organization of these genes may lead to the isolation of a gut-specific, inducible promoter for the expression of anti-parasitic agents in transgenic mosquitoes. Moreover, specific inhibitors could be designed on the basis of the structural properties of the enzymes. We report here on the identification of a trypsin gene family in *Anopheles gambiae*, the mosquito vector of malaria in Africa. Mosquito trypsin-related sequences were amplified by PCR using as template cDNA derived from RNA of blood fed mosquitoes. Cloning of the PCR product revealed two distinct sequences. Corresponding full-length cDNA clones were obtained and sequenced. *Antryp1* and *Antryp2* code for proteins of 274 and 277 amino acids respectively, showing 75% homology at the amino acid level. The deduced amino acid sequences clearly identify them as trypsins. Five additional trypsin sequences were found in overlapping genomic clones. The genes identified are tightly clustered within 11 kb and sequencing indicates that no introns are present. Northern and PCR analysis indicated that the transcription of both *Antryp1* and *Antryp2* is induced by blood feeding. Moreover, the Antryp1 protein was detected among the proteins of a midgut lysate of blood fed mosquitoes using antisera against recombinant Antryp1. In addition, the recombinant polypeptides derived from *Antryp1* and *Antryp2* expressed in *Escherichia coli* showed a strong proteolytic activity against different sets of blood proteins. We conclude that the products of *Antryp1* and *Antryp2* play an important role in the breakdown of the proteins during the digestion of the blood meal in the mosquito gut.

Key words: Anopheles gambiae/blood meal induced/gene cluster/mosquitoes/trypsin

Introduction

Female *Anopheles gambiae* mosquitoes have an absolute requirement to feed at least twice on the blood of a vertebrate organism to allow oogenesis to proceed. The blood meal provides the nutrients necessary to support the biosynthetic burst associated with the process of vitellogenesis (Clements, 1963). Together with the blood, mosquitoes may ingest several types of micro-organisms. *Plasmodium* sp. as well as *Filariae* and some arboviruses have evolved to exploit the feeding behaviour of *A.gambiae* mosquitoes to propagate themselves from one organism to another (White, 1974). The *A.gambiae* complex represents the most efficient vector system for *Plasmodium falciparum* malaria (Coluzzi, 1984) and its control is one of the major health problems in tropical Africa (Sturchler, 1989). The elucidation of the digestive processes leading to the utilization of the blood in the *Anopheles* gut may help both in developing novel vector control strategies and in understanding parasite vector interactions. Disruption of the process of blood utilization in the mosquito gut would have a detrimental effect on the vector population and may therefore represent a tempting target for attack. In fact, it has been shown that immunization of cattle with tick gut antigens results in the death of the insects when they ingest a blood meal containing gut antibodies (Willadsen *et al.*, 1989).

There has been accumulating evidence that trypsins may play an important role in the digestion of the blood meal in the mosquito gut. In a number of mosquito species, including *Anopheles*, enzymes with trypsin-like activity have been purified (Graf *et al.*, 1991). Moreover, trypsin activity is induced in the gut following ingestion of a blood meal (Clements, 1963; Barillas-Mury *et al.*, 1991; Graf *et al.*, 1991). Characterization of purified mosquito trypsins indicates that they have a range of properties in common with other known serine proteases. These similarities include molecular weight range, substrate and inhibitor specificity (Graf *et al.*, 1991) and the level of amino acid homology with any other known trypsin sequence (Barillas-Mury *et al.*, 1991). Trypsin activity does not seem to be uniformly distributed throughout the gut of the mosquito. For example, the trypsin activity is spatially restricted to the posterior region of the midgut lumen in blood fed *Anopheles stephensi* (Billingsley and Hecker, 1991).

The characterization of the genes coding for these trypsins may help in understanding the processes involved in the utilization of blood proteins by *A.gambiae*. Partial peptide sequences of some mosquito trypsins are available but only one gene, from the mosquito *Aedes aegypti*, has so far been cloned (Barillas-Mury *et al.*, 1991). We have focused our attention on the isolation and characterization of trypsin genes in *A.gambiae*. To amplify, by PCR, trypsin-specific sequences from cDNA of blood fed *Anopheles* mosquitoes we have designed degenerate DNA primers on the basis of the available amino acid sequences. We report here the cloning of a trypsin gene family from *A.gambiae*. The expression of two members of this gene family is induced in response to blood feeding. The nucleotide sequence of the corresponding full-length cDNA clones is presented

together with an analysis of their temporal and spatial expression in the mosquito. Recombinant *Anopheles* trypsins expressed in *Escherichia coli* were shown to have selective proteolytic activity against different blood components.

Results

PCR amplification of A.gambiae trypsin sequences

Although trypsin sequences from several organisms are available, the design of PCR primers to amplify *Anopheles* trypsin may be problematic. Few sequences of mosquito trypsins are known, moreover, there is a lack of information concerning the codon usage of *Anopheles* genes. Comparison of trypsin sequences from *A.aegypti* and *Anopheles quadrimaculatus* revealed as much diversity as if the sequences were compared with *Drosophila* or vertebrate trypsins (Graf *et al.*, 1991). We therefore designed the PCR primer pTry1 according to the sequence of the first nine amino acids (VVGGFEIDV) of the amino-terminal peptide sequence of the *A.quadrimaculatus* trypsin (Graf *et al.*, 1991) which was the only sequence information available for anopheline mosquito trypsins. The sequence of the second primer. pTry3, was derived from the amino acid sequence CQGDSGGPLVLV, which encompasses the serine active site. The first eight amino acids encoded by pTry3 were found to be identical in all trypsin genes so far sequenced. Trypsin sequences were amplified from cDNA obtained from blood-fed *A.gambiae* mosquitoes. A PCR product of ~550 bp was obtained and the blunt-end fragment subsequently cloned into the *Hin*dII polylinker site of pDS56/RBSII,6×His/E$^-$ (Hochuli *et al.*, 1988). Restriction analysis of a number of clones indicated that the PCR had amplified two distinct sequences. This finding was also confirmed by analysis of the original PCR product with appropriate restriction enzymes. The sequence analysis of two representative clones derived from the PCR product revealed the presence, in each clone, of an open reading frame coding for a distinct trypsin-related sequence.

cDNA cloning and sequence analysis

The PCR product obtained with the primers pTry1 and pTry3 was used as probe to search for full-length cDNA clones in a λgt11 expression library developed using cDNA generated from mRNA of blood fed *A.gambiae*. The inserts of a number of positive λgt11 plaques were cloned in the *Eco*RI site of the pDS56/RBSII,6×His/E$^-$ polylinker. Two types of clones were identified according to the restriction pattern of the two previously cloned PCR fragments. Sequence analysis of full-length cDNA clones of each type, *Antryp1* and *Antryp2*, showed that they encompassed the respective partial sequences of the cloned PCR products. *Antryp1* and *Antryp2* revealed open reading frames (G/C content of 60 and 59% for *Antryp1* and *Antryp2* respectively) coding for polypeptides of 274 and 277 amino acids respectively. The analysis of the deduced amino acid sequences of *Antryp1* and *Antryp2* indicated that they code for two distinct polypeptides showing 75% homology at the amino acid level (Figure 1). Moreover, both of the predicted polypeptides are highly homologous to known trypsin sequences (Figure 2) and share several features with typical trypsin sequences: (i) highly conserved sequences at the histidine and serine catalytic sites; (ii) glutamine at position 229 (numbering according Figure 2), a trypsin specific amino acid preceding the serine active site; (iii) an aspartic acid at position 136 completing the catalytic triad together with histidine at position 91 and a serine at position 232 (Kraut, 1977); and (iv) the amino acids determining trypsin substrate specificity (aspartate at position 226, two glycine at positions 251 and 259) are also conserved in Antryp1 and Antryp2 (Kraut, 1977). Finally, the cysteine residues in Antryp1 and Antryp2 are located at highly conserved positions that would allow the formation of three intramolecular disulphide bonds. This last characteristic is a common feature of all invertebrate serine protease sequences known so far, including the *A.aegypti* trypsin (Barillas-Mury *et al.*, 1991), the hornet chymotrypsins (Jany *et al.*, 1983; Jany and Haugh, 1983), the *Drosophila* trypsin-like protease (Davis *et al.*, 1985), the *Drosophila* Ser1 protease (Yungdae and Davis, 1989), common cattle grub collagenase (Lecroisey *et al.*, 1987), sandfiddler crab collagenase (Grant *et al.*, 1980) and crayfish trypsin (Titani *et al.*, 1983), and it distinguishes them from the sequences of vertebrate trypsins that usually contain an additional disulfide bridge. The calculated isoelectric points (pI) of Antryp1 (pI = 5.1) and Antryp2 (pI = 4.1) resemble those of trypsins purified from several *Anopheles* species (Graf *et al.*, 1991), which are in the range between pH 4.3 and 4.9. The structural analysis of the deduced amino acid sequences predict the presence of a signal peptide, which is highly conserved between Antryp1 and Antryp2. The signal peptide is followed by a stretch of 29–30 amino acids showing an unusually high content of basic amino acids and prolines that may represent an activation peptide. This sequence is among the most divergent regions in Antryp1 and Antryp2 (50% homology) (Figure 2). Sequence comparison of the predicted active enzyme amino acid sequence of Antryp1 with several serine proteases (Figure 2) revealed significant degrees of identity with *Drosophila* trypsin-like protease (44%), vertebrate trypsins (41–42%), *Aedes* trypsin (40%) and the *Drosophila* Ser1 protease (38%). Similarly, Antryp2 shows 45% identity with the *Drosophila* trypsin-like protease, 39–43% with the vertebrate trypsins, 41% with the *Aedes* trypsin and 39% identity with the *Drosophila* Sera protease. However, highest homology of *Antryp1* and *Antryp2* was found in the EMBL Databank with two unpublished *A.aegypti* trypsin-like sequences (EMBL accession numbers X64362 and X64363), showing 54 and 64% identity to *Antryp1* and 52 and 60% identity to *Antryp2*.

Genomic organization of the A.gambiae trypsin genes

To determine the genomic sequences of *Antryp1* and *Antryp2*, an *A.gambiae* genomic λEMBL3A library was screened with the same probe that was used for the isolation of the λgt11 clones. Several overlapping λEMBL3A clones were obtained, two of them, Ty 3.3 and Ty 4.1, contained both trypsin sequences. Hybridization experiments carried out on the genomic clones (not shown) and on *A.gambiae* DNA digested with a set of restriction enzymes, revealed several bands that could not be explained by the presence of *Antryp1* and *Antryp2* alone (Figure 3). We have sub-cloned and sequenced the internal *Bam*HI fragments of the genomic clone TY3.3 in order to search for additional trypsin-related sequences (Figure 4). Sequence analysis of genomic subclones obtained by *Bam*HI/*Hin*dII shot gun cloning experiments, containing *Antryp1* and *Antryp2*, confirmed the respective cDNA sequences (Figure 1) and indicated that the coding sequences of both genes were not

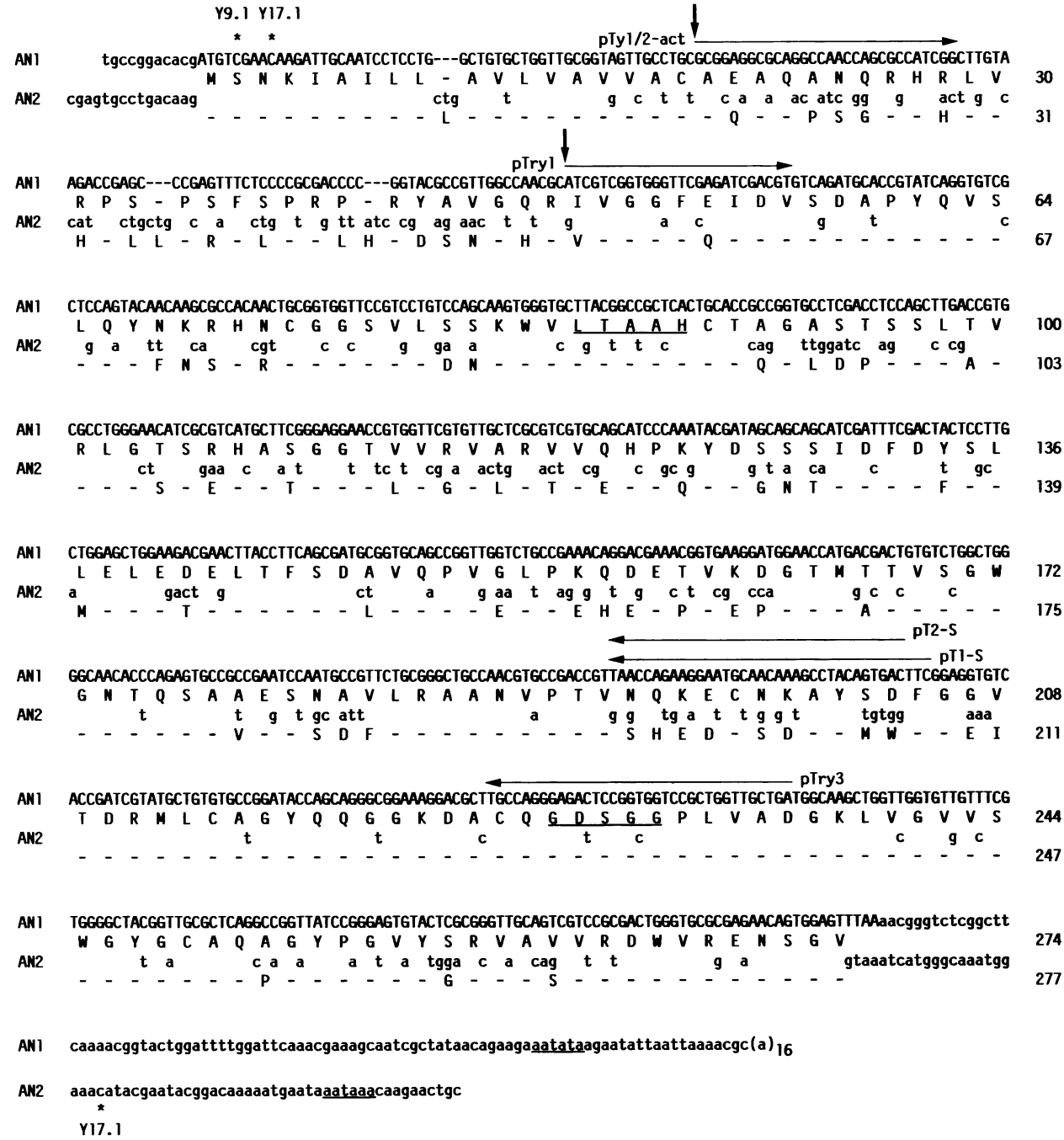

Fig. 1. Comparison of nucleotide and deduced amino acid sequences derived from *Antpyr1* (AN1) cDNA clone Y9.1 and *Antryp2* (AN2) cDNA clone Y17.1. The first nucleotide position of Y9.1 and Y17.1 and the 3'-end of Y17.1 are indicated by asterisks. Additional sequence information completing the amino-termini of Antryp1 and Antryp2 and contributing to the 3'-untranslated region of *Antryp2* is derived from overlapping partial cDNA sequences. Histidine and serine active sites are underlined. Positions and orientation of the various oligonucleotides used in this study are shown above the nucleotide sequence of *Antryp1*. Predicted signal peptide cleavage positions and the putative activation peptide cleavage side are indicated by arrows. Possible polyadenylation signal sequences are underlined.

interrupted by any intron. Moreover, five additional trypsin-related genes were found in close proximity of *Antryp1* and *Antryp2* (Figure 4), and none of them contained introns. Contiguity and relative orientation was confirmed by further restriction fine mapping of the genomic clones TY3.3 and TY4.1, which was in accordance with the results obtained by Southern analysis (Figure 3). The complex *ClaI* pattern in the region coding for Antryp4, 5 and 6 (Figure 4) could be resolved only by sequencing; additional *ClaI* sites occurring in the *SalI*–*Bam*HI fragment containing the amino-terminus of Antryp6 are not yet located. In addition, contiguity was proven by PCR performed on TY3.3 using specific primer pairs, followed by digestion of the PCR products with restriction enzymes specific for each fragment (not shown).

Blood meal-induced expression of the trypsin genes Antryp1 and Antryp2

The expression of the trypsin genes in the mosquito was assessed by Northern analysis using total RNA extracted

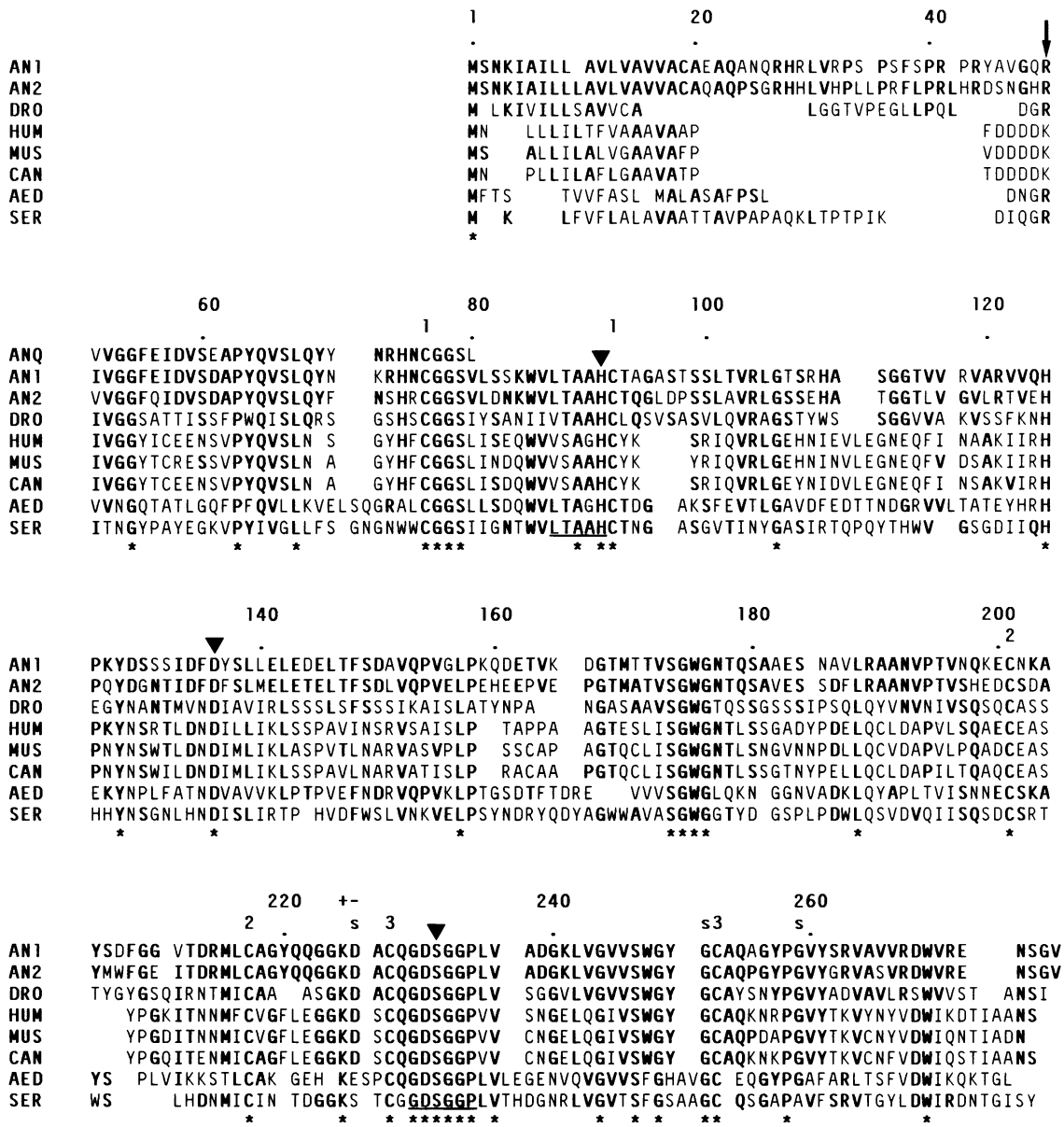

Fig. 2. Multiple alignment of the predicted amino acid sequences of the *A.gambiae* trypsins Antryp1 (AN1) and Antryp2 (AN2) with selected serine proteases. ANQ, partial peptide sequence of an *A.quadrimaculatus* trypsin (Graf *et al.*, 1991); DRO, *D.melanogaster* trypsin-like enzyme (Davis *et al.*, 1985); HUM, human trypsinogen 2 (Emi *et al.*, 1986); MUS, mouse trypsinogen (Stevenson *et al.*, 1986); CAN, canine anionic trypsinogen (Pinksy *et al.*, 1985); *AED*, an *A.aegypti* trypsin (Barillas-Mury *et al.*, 1991); SER, *D.melanogaster* putative serine protease Ser1 (Yungdae and Davis, 1989). Amino acid positions were numbered according to the sequence of Antryp2. The carboxy-end of the activation peptides is indicated by an arrow. Histidine and serine active sites are underlined, the amino acids forming the catalytic triad are marked with triangles. Numbers (1–3) indicate disulfide bonds formed by the conserved cysteine residues, +- indicates the ajdacent positive and negative charge in the trypsin specificity pocket; s, amino acid residues conferring substrate specificity.

from female mosquitoes at different time points after blood feeding. As controls, RNA from male mosquitoes, fourth instar larvae, pupae and non-fed female mosquitoes was also analysed. In this experiment, we used the oligonucleotides pTy1-S and pTy2-S as probes as they have unique sequences specific for *Antryp1* and *Antryp2* respectively. Ty1-S and pTy2-S were derived from one of the most polymorphic regions (Figure 1) of the trypsin genes. In Northern blot, both probes hybridized with an RNA species migrating at ~1 kb (Figure 5), in good accordance with an estimated minimal transcript length of ~950 bases. Adult females, and to some degree pupae, showed a small amount of *Antryp1* transcript. In female mosquitoes following a blood meal, transcription of *Antryp1* steadily increased and reached its peak after 24 h; the amount of *Antryp1* mRNA had decreased 40 h after the blood meal. *Antryp2* transcripts could only be detected in female mosquitoes between 8 and 24 h after blood feeding, the amount of transcript induced is, however, at the detection limit of end-labelled oligonucleotides in Northern analysis. Induction of the *Antryp2* gene could be demonstrated much more clearly in amplification experiments using as template cDNA generated from blood fed mosquitoes (Figure 5). Under this experimental condition, it was possible to demonstrate that

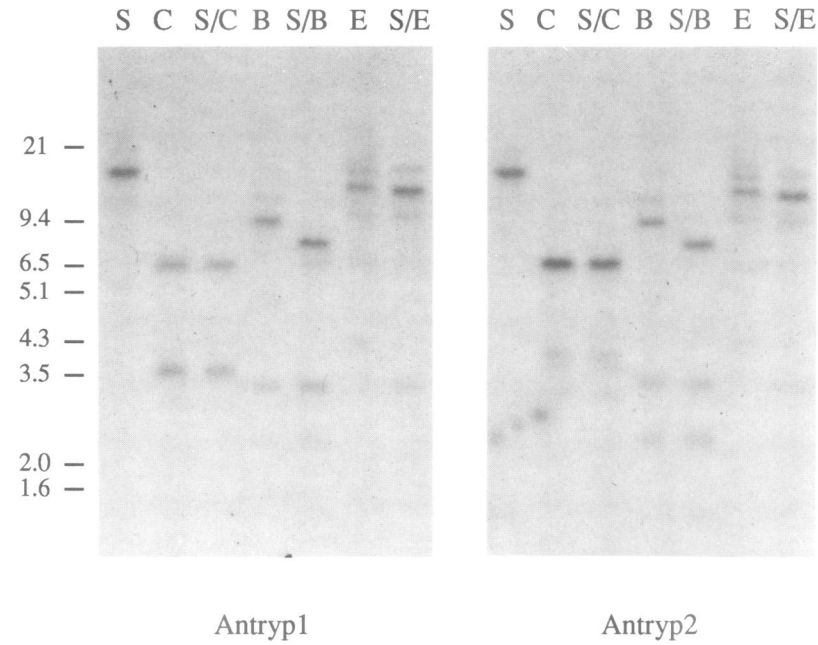

Fig. 3. Genomic Southern blot analysis. *A.gambiae* DNA was digested either with *Sal*I (S), *Cla*I (C), *Bam*HI (B) or *Eco*RI (E) alone or in combinations of *Sal*I with *Cla*I (S/C), *Bam*HI (S/B) and *Eco*RI (S/E). The DNA was separated on a 0.7% agarose gel and hybridized with ^{32}P-labelled *Antryp1* and *Antryp2* full-length cDNA fragments. The selected marker positions indicated are derived from λDNA digested with *Hin*dIII or *Hin*dIII/*Eco*RI, run in parallel.

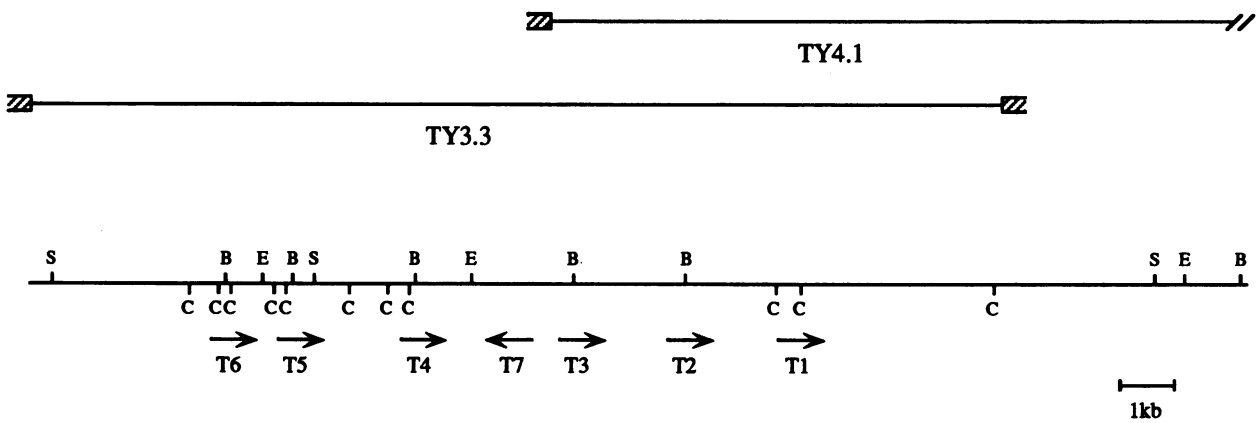

Fig. 4. Genomic restriction map of the *A.gambiae* trypsin gene family. The relative positions of the two overlapping λEMBL3A clones Ty3.3 and Ty4.1 used for restriction mapping and subcloning are shown above the restriction map, λEMBL3A vector borders are shown dashed. The relative position and transcriptional direction of the trypsin genes (T1−T7) are indicated by arrows. B, *Bam*HI; C, *Cla*I; E, *Eco*RI; S, *Sal*I.

Antryp2 is expressed only in blood fed female mosquitoes and that its induction is tightly controlled.

In vitro activity of recombinant Anopheles trypsins

Antryp1 (amino acids 19−274) *Antryp2* (amino acids 20−277) sequences, lacking the signal peptide and containing the putative activation peptide, were expressed in *E.coli* fused to a stretch of six histidines at their amino-termini. The recombinant proteins were purified by nickel affinity chromatography in 8 M urea and the eluate was subsequently dialysed against phosphate-buffered saline, pH 7.4 (PBS). SDS−PAGE analysis showed a difference in migration patterns under reducing and non-reducing conditions, suggesting the presence of intramolecular disulfide bonds (not shown). Since the transcription of both trypsins is induced by blood feeding, we have assayed the proteolytic activity of Antryp1 and Antryp2 on both red blood cell (RBC) lysate and serum proteins (Figure 6). Recombinant Antryp1 and Antryp2, when added to RBC lysate or to serum protein, showed a selective proteolytic activity against different proteins that could be inhibited by the trypsin-specific protease inhibitor 4-amidino-phenylmethane-sulfonyl fluoride (APMSF). Recombinant Antryp1 digested both haemoglobin and serum albumin equally well, whereas Antryp2 seemed to be active mainly against haemoglobin. The two trypsins have different kinetics and the respective breakdown products of haemoglobin have different molecular weights. These *in vitro* digestion

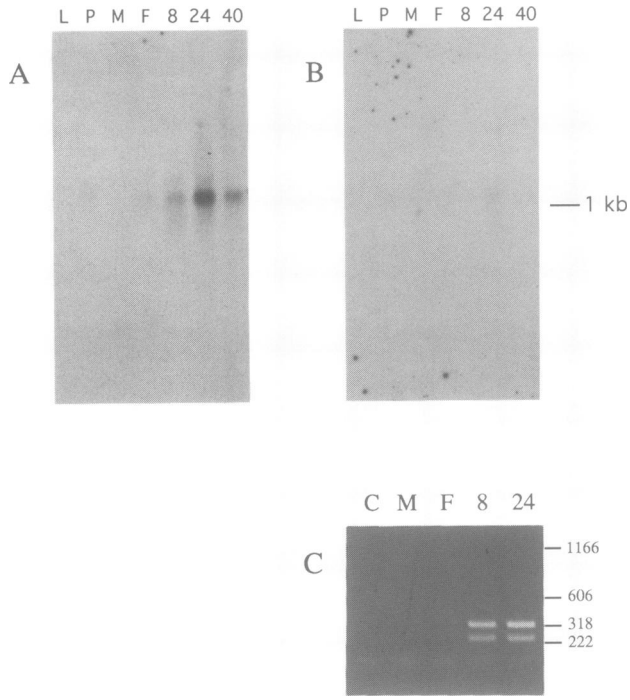

Fig. 5. Induction of *Antryp1* and *Antryp2* mRNA in *A.gambiae* following blood meal. Total RNA of *A.gambiae* fourth instar larvae (L), pupae (P), male (M) unfed female mosquitoes (F) and blood-fed females killed at 8, 24 and 40 h after blood meal was separated on 1.5% denaturing agarose gels and hybridized with ^{32}P-labelled oligonucleotides specific for Antryp1 (pTy1-S) (**A**) and Antryp2 (pTy2-S) (**B**). (**C**) Analysis on 1% agarose gel of RNA-PCR products obtained using the *Antryp2*-specific primer combination pTy2-act/pTy2-S and as template first strand cDNA derived from poly(A)$^+$ RNA of male (M), female (F) mosquitoes and females killed at 8 and 24 h after blood feeding. Control PCR was done on an aliquot of 24 h post-blood meal RNA sample that was treated with RNaseA before reverse transcription. Prior to electrophoresis on 1% agarose, the PCR samples were digested with *Bam*HI which cleaves at a position unique for *Antryp2*, yielding two DNA fragments of 331 and 243 bp. DNA size standards are indicated.

experiments support the notion that the enzymes identified are indeed functional *Anopheles* trypsins. The apparent molecular weight of both trypsins after dialysis against PBS was reduced by 3–4 kDa if compared with molecular weight of the trypsins in the 8 M urea fractions. During the dialysis against PBS, the trypsins may undergo an autoactivation process that may involve the cleavage of the 6×His activation peptide. Interestingly, the *in vitro* activity of both enzymes requires the addition of SDS (0.1%) or β-mercaptoethanol (100 mM) to the digestion mixture, indicating the need for Antryp1 and Antryp2 of additional components which alter the tertiary structure of the protein substrates.

Localization of Antryp1 in the midgut

Any hypothesis on the role of Antryp1 and Antryp2 in the digestion of the blood meal must take into account the localization of the corresponding *Anopheles* proteins. For this purpose, we have used purified, recombinant Antryp1 and Antryp2 proteins dialysed against PBS to raise antisera in mice. Only the mice immunized with Antryp1 produced enough specific antibodies to allow its localization among the proteins present in mosquito tissue lysates in immunoblot. Using recombinant Antryp1 antiserum, the corresponding *Anopheles* trypsin could be identified within a protein lysate generated from guts of blood fed *A.gambiae* mosquitoes and was not detectable in carcasses (blood fed mosquitoes from which the guts were removed). The antiserum recognizes a band of 30 kDa and an additional fine band of 33 kDa, representing probably the trypsin with activation peptide (Figure 7). The recombinant Antryp1 migrates at identical position to its natural counterpart. Antryp1 could not be detected in larvae, pupae, mosquitoes and non-fed females. In female mosquitoes, Antryp1 was detectable already 8 h after blood meal. In addition, the anti-Antryp1 serum strongly recognized, in the protein lysate of fourth instar larvae, a 26 kDa protein. This protein was not detected in the following pupal stage.

Discussion

We have identified two full-length cDNA clones that encode distinct *A.gambiae* trypsins: *Antryp1* and *Antryp2*. Sequence analysis indicated that the predicted amino acid sequences of *Antryp1* and *Antryp2* have a sequence homology of 75% at the amino acid level. *Antryp1* and *Antryp2* are highly homologous to the region corresponding to the oligonucleotide pTry1 derived from an *A.quadrimaculatus* trypsin peptide sequence, so that it is not surprising that the primers pTry1 and pTry3 amplified both *A.gambiae* trypsin sequences. Amino acid sequence comparison analysis with insect and vertebrate trypsins showed that *Antryp1* and *Antryp2* shared the highest degree of homology with the *Drosophila* trypsin-like gene (Davis *et al.*, 1985). The mosquito trypsin from *A.aegypti* (Barillas-Mury *et al.*, 1991) showed a lower degree of similarity to *Antryp1* and *Antryp2* than the vertebrate trypsins. Moreover, *Antryp1* and *Antryp2* showed the highest homology, up to 64%, with two unpublished *A.aegypti* trypsin-like genes (EMBL accession numbers X64362 and X64363).

Analysis of genomic clones containing the *Anopheles* trypsins resulted in the identification of five additional trypsin genes, closely clustered with *Antryp1* and *Antryp2* within 11 kb. None of the genes we have identified contained any intron sequence. After sequencing, the majority of the bands detected on genomic Southern blots could be explained either by specific or cross-hybridization of the *Antryp1* and *Antryp2* probes to the trypsin genes. However, the presence of additional trypsin-like genes near this gene cluster cannot be ruled out. The *Antryp1* and *Antryp2* probes did not cross-hybridize in genomic Southern blots to two *A.gambiae* chymotrypsin genes recently isolated in our laboratory. These chymotrypsins (EMBL accession numbers Z18887 and Z18888) show ~50% homology to the *A.gambiae* trypsin sequences at the nucleic acid level. A genomic organization similar to the *Anopheles* trypsin genes has also been reported for the *Drosophila melanogaster Ser1* gene encoding a serine protease showing the highest homology to crab collagenase (Yungdae and Davis, 1989) and the alpha-gene of *D.melanogaster* encoding a trypsin-like protease (Davis *et al.*, 1985), both proteases are expressed in the gut. The genes of the corresponding gene families are closely clustered and do not contain any intron sequences.

The deduced amino acid sequences of *Antryp1* and *Antryp2* showed several structural features typical of trypsin proteases (Kraut, 1977) including: (i) the amino acid residues considered to determine trypsin substrate specificity; (ii)

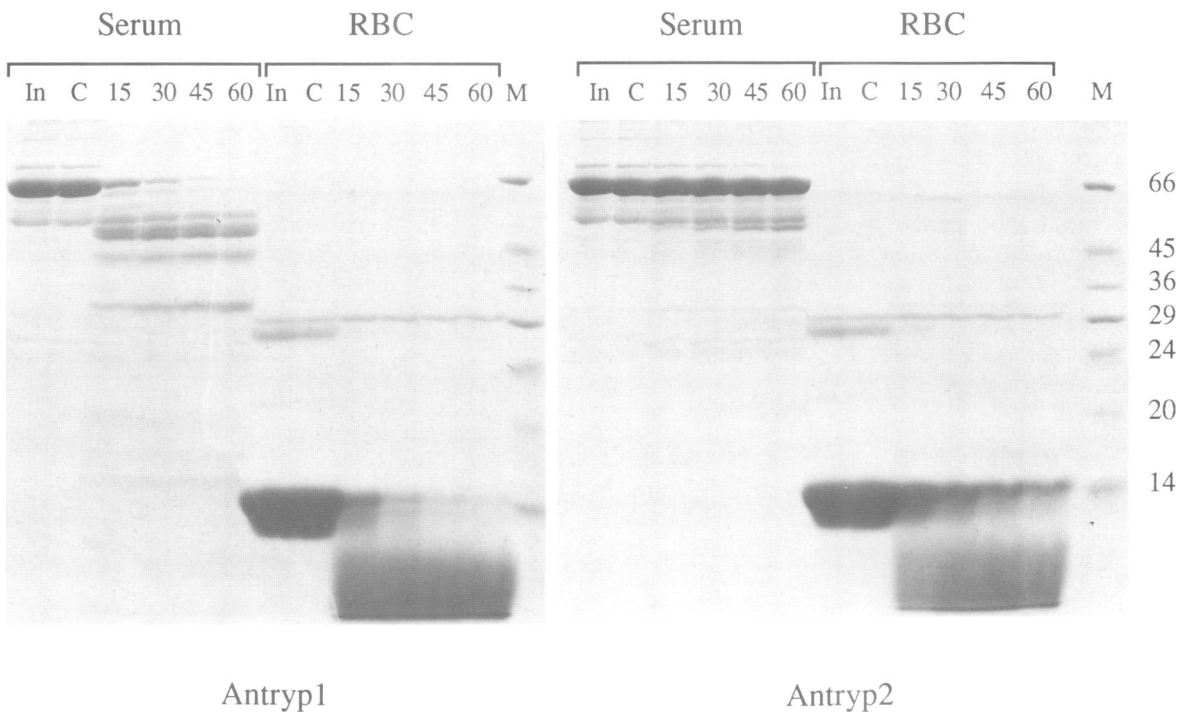

Fig. 6. Digestion of human serum and red blood erythrocyte proteins by recombinant *Anopheles* Antryp1 and Antryp2. Incubation of proteins with recombinant trypsins was allowed for 15, 30, 45 and 60 min. An additional digestion of 60 min was carried out as control in the presence of proteins, *Anopheles* trypsins and the trypsin inhibitor APMSF (In). Serum and RBC components incubated in the absence of proteases are shown (C). The molecular weights of the marker bands are indicated (kDa).

conserved histidine and the serine catalytic sites; (iii) the catalytic triad; and (iv) six cysteine residues, all located at conserved positions. Vertebrate trypsin sequences usually contain more than three cysteine pairs. The lack of a cysteine pair in Antryp1 and Antryp2 is a typical feature of invertebrate trypsin sequences. Both trypsins contained at their amino-termini a highly homologous sequence of 18–19 amino acids that showed the structural characteristics of a signal peptide. Between the predicted signal peptide and the enzyme sequences, Antryp1 and Antryp2 contain a stretch of 29 and 31 amino acids respectively representing putative activation peptides. These sequences are not conserved between Antryp1 and Antryp2 and show no homology with the corresponding vertebrate trypsin sequences. However, Antryp1 and Antryp2 contain an arginine residue at the cleavage site that is shared by several digestive serine proteases including proelastase (MacDonald *et al.*, 1982), chymotrypsinogen (Bell *et al.*, 1984), the Ser1 protease of *Drosophila* (Yungdae and Davis, 1989) as well as by the trypsin-like gene of *Drosophila* (Davis *et al.*, 1985) and *A.aegypti* (Barillas-Mury *et al.*, 1991). The putative activation peptides of *Anopheles* resemble that of the Ser1 protease in its high content of proline and basic amino acids.

If the *Anopheles* trypsins play a role in the blood meal digestion process, the genes should be expressed in the gut of the mosquito upon blood feeding. In addition, the corresponding recombinant proteins expressed in *E.coli* would be expected to exert a proteolytic activity on blood constituents.

Northern blot and RNA-PCR analysis indicated that *Antryp1* and *Antryp2* mRNA was already present 8 h after blood feeding and increased until 24 h thereafter. Although the expression of both *Antryp1* and *Antryp2* is induced by blood feeding, the transcriptional control of the two trypsin

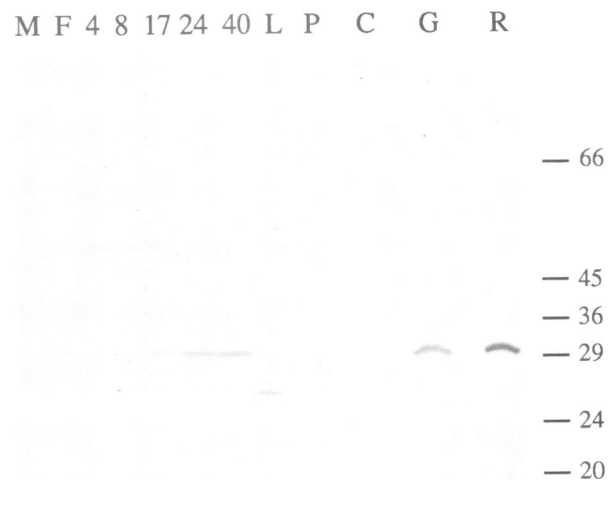

Fig. 7. Immunoblot analysis of the Antryp1 protein in the mosquito using anti-Antryp1 serum. Antryp1 is induced by blood meal in the midgut. M, male; F, unfed female; L, larvae; P, pupae and female mosquitoes killed at several time points after blood meal (4, 8, 17, 24 and 40 h); C, carcass (blood fed female without midgut); G, dissected midgut 24 h after blood feeding; R, 0.2 μg of the purified active form of recombinant Antryp1. Protein lysates were separated on reducing SDS–PAGE (12.5%). Marker positions in kDa are indicated.

genes appears to be quite different. At the time points we examined, the amount of *Antryp1* mRNA was always higher than *Antryp2* mRNA. Moreover, significant levels of *Antryp1* transcript could be detected in non-fed female mosquitoes and to a lesser extent also in males and pupae, whereas the transcription of *Antryp2* appeared to be tightly controlled.

Antryp2 mRNA was found only in female mosquitoes after blood feeding. These observations are consistent with previous reports describing the time-course of the induction of trypsin activity in the mosquito gut upon feeding (Clements, 1963; Graf and Briegel, 1989; Barillas-Mury *et al.*, 1991) as well as the baseline trypsin activity in gut extracts of female mosquitoes (Clements, 1963). The function and the localization of the five *Anopheles* trypsins Antryp3–7 is currently under investigation.

To test the function of the trypsins in the digestion process of the blood meal we have assayed the proteolytic activity of recombinant Antryp1 and Antryp2 against serum proteins and RBC extracts. Our results indicate that these proteases may play an important role in the digestion of the blood meal. Small amounts of recombinant Antryp1 and Antryp2 added to serum protein or RBC extract resulted in the selective breakdown of several proteins. Antryp1 was extremely active against both serum albumin and haemoglobin, whereas Antryp2 was preferentially active against haemoglobin. In order to exert their proteolytic activity, Antryp1 and Antryp2 require the presence of small amounts of either SDS or β-mercaptoethanol in the digestion mixture. The addition of these compounds is likely to alter the tertiary structure of many blood proteins thus making buried amino acid sequences accessible to recombinant *Anopheles* trypsins. Compounds with similar denaturing activity may be produced in the mosquito gut during the digestion process. The mechanism of trypsin activation in the mosquito gut is not yet elucidated. In our experimental system, recombinant trypsins may undergo an auto-activation process, with cleavage of the activation peptide probably taking place during the dialysis from urea to PBS. This possibility is suggested by the results described here as well as by previous observations that showed auto-activation of trypsins under different conditions (Brodrick *et al.*, 1987a,b). After dialysis, the molecular weight by SDS–PAGE of both Antryp1 and Antryp2 shifted from 33 to 30 kDa; the recombinant *Anopheles* trypsin of 30 kDa showed a strong proteolytic activity in PBS. Moreover, the molecular weight shift is compatible with the cleavage, from the trypsin sequence, of a polypeptide that would have the size of the putative activation peptide. Indeed, recombinant, auto-activated Antryp1 shows in SDS–PAGE the same migration behaviour as the *Anopheles* enzyme.

To assess the role of Antryp1 and Antryp2 in the digestion process, we thought it essential to demonstrate the expression of the corresponding *Anopheles* protein in the female mosquito gut after blood feeding. For this purpose we have used purified recombinant Antryp1 and Antryp2 to generate antisera in mice. Using the Antryp1 antisera we have demonstrated that the corresponding *Anopheles* trypsin could be detected in immunoblot among the proteins present in the gut lysate after blood feeding. The serum strongly reacted with a band of 30 kDa, together with a fine band of 33 kDa. The molecular weight of the upper and lower band suggests that they could represent the mosquito Antryp1 with and without the activation peptide. In a protein lysate of mosquito larvae the Antryp1 antiserum recognized a band of the apparent molecular weight of 26 kDa. Although the serum raised against Antryp1 did not react against recombinant Antryp2 and recombinant *Anopheles* chymotrypsins, it is still possible that cross-reacting antibodies against related sequences are present. The apparent molecular weight of this protein is still in the range of that of mosquito trypsins, 25–35 kDa (Graf *et al.*, 1991). As no *Antryp1* message could be detected in larval RNA of the same stage, and a protein of this size was not detected by the anti-Antryp1 serum in female mosquitoes at any time point after blood feeding, the 26 kDa protein is unlikely to represent a processed form of the trypsin Antryp1.

The identification of the trypsin gene family in *A.gambiae* may have important implications for the understanding of the blood meal digestion process as well as for the design of new control strategies. Antibodies raised against trypsins could be tested for their ability to interfere with the digestion process and hence the viability of the mosquitoes. Moreover, it should be possible in the near future to characterize functionally the trypsin promoters specifically transcribed in the gut. These control elements could be used to construct transgenic mosquitoes that express antiparasitic agents in their gut and thus disrupt the disease transmission cycle.

Materials and methods

Mosquito rearing and membrane feeding

The *A.gambiae* s.s. Suakoko strain (karyotype: Xag, 2R+, 2La, 3R+, 3L+), was maintained at adult stage at 26°C and 80% relative humidity with access to 10% sucrose solution. For purification of total RNA used in Northern blotting, PCR and for the preparation of protein extracts for Western blots, 2–4 day-old females were allowed to feed on a guinea pig for 30 min before they were killed at various time points after the blood meal. For the isolation of poly(A)$^+$ RNA and the production of cDNA libaries, female mosquitoes were membrane-fed (Rutledge *et al.*, 1964) on RBC extracts in order to avoid RNA contaminants of blood origin. The RBC extract was prepared as follows. Plasma of heparinized, 2-week-old human blood was removed together with the buffy coat. Without further washing, 2 vol of cold, distilled water were added to the RBC pellet. After haemolysis occurred, the osmolarity of the RBC extract was adjusted to 1 × PBS and insoluble components were removed by centrifugation. After supplementation with 5 mM rATP, the RBC extract was stored in aliquots at −20°C.

cDNA synthesis

Within 3–17 h after feeding on human RBC extracts, female mosquitoes were killed by freezing in liquid nitrogen. Total nucleic acids were isolated from 200 female *A.gambiae* mosquitoes by grinding the pooled mosquitoes in 10 ml of 50 mM Tris–HCl pH 7.6, 100 mM NaCl, 1% SDS, 100 mM EDTA, and 1 mg/ml proteinase K, and incubated at 50°C for 1 h. After phenol extraction and ethanol precipitation, poly(A)$^+$ RNA was isolated using an Invitrogen Micro-Fast Track mRNA Isolation Kit according to the manufacturer's instructions. First strand cDNA synthesis of the isolated poly(A)$^+$ RNA was performed in two 50 μl reactions (Perbal, 1988) using 50 units AMV-reverse trancriptase at 42°C for 45 min. In one reaction, cDNA synthesis was primed with 5 μg oligo dT_{18}, in the second with 0.5 A_{260} units $p(dN)_6$ random primers. Second strand synthesis with RNaseH and PolI, T4 PolI polishing, *Eco*RI methylation, *Eco*RI linker ligation and *Eco*RI digestion were performed as described. The cDNA preparations were separated on a 1% low melting agarose gel. After removing the region of the gel containing the DNA molecules <300 bp, high molecular weight cDNA was bound to a DEAE filter during reverse electrophoresis. Subsequently the DNA was eluded with 1.5 M NaCl–TE at 65°C; after ethanol precipitation, oligo dT_{18} and dN_6-primed cDNAs were resuspended and pooled in 50 μl 1/4 TE buffer. All enzymes, dT_{18} and $p(dN)_6$ random primers were purchased from Boehringer Mannheim. This cDNA preparation was used for the development of a λgt11 library and as template for the PCR reaction with primers pTry1/pTry3.

A.gambiae λgt11 cDNA library

Purified cDNA of blood-fed *A.gambiae* mosquitoes was ligated with dephosphorylated *Eco*RI λgt11 arms according to manufacturer's instructions (Boehringer Mannheim) and incubated with *in vitro* packaging extracts (Boehringer Mannheim). Recombinant phage particles were amplified by preparing plate lysates with *E.coli* strain Y1090, yielding a complexity of 500 000 plaques. Screening was performed on the amplified library at an initial density of 40 000 plaques per 13 cm plate with denatured probes,

^{35}S-labelled by random priming (Feinberg and Vogelstein, 1983) according to standard procedures (Perbal, 1988).

Genomic A.gambiae λEMBL3A library

DNA of 300 male *A.gambiae* mosquitoes was prepared in an initial volume of 10 ml by the proteinase K method as described for cDNA synthesis, 10 μg RNaseA-treated DNA were partially digested in a volume of 100 μl with several dilutions of *Sau*3A (Boehringer Mannheim) at 37°C for 30 min. Aliquots were analysed on a 0.6% agarose gel, digestions with 0.1 and 0.03 units *Sau*3A that resulted in 15–20 kb DNA fragments were phenol extracted and pooled. Aliquots were ligated with dephosphorylated EMBL3A *Bam*HI arms (Promega) and *in vitro* packaged (Boehringer Mannheim). Recombinant phages were amplified in *E.coli* strain LE392 resulting in a complexity of 150 000 plaques, corresponding to ~10 times the estimated genome complexity of *A.gambiae* (2.6×10^8) (Bensansky and Powell, 1992). Library screenings were performed as described for the expression library.

PCR reactions

PCR reactions (Saiki *et al.*, 1988) were performed on a Lab-Line Programmable Thermable Blok machine, in a volume of 100 μl containing 10 mM Tris–HCl pH 8.3 at room temperature, 50 mM KCl, 2 mM $MgCl_2$, 200 μM of each desoxynucleotide, 50 pmol of each primer, 1 unit of Taq polymerase and 2 μl of template (cDNA), overlaid with 80 μl of mineral oil. PCR on cDNA using primers pTry1/pTry3 was done in three steps of 2 min at 48°C, 3 min at 72°C, 2 min at 94°C (35 cycles). PCR using cDNA as template and primers pTy2-act/pTy2S was done in two steps of 2.5 min at 72°C, 2 min at 94°C (25 cycles). Taq DNA polymerase was purchased from Perkin–Elmer.

Southern blot analysis

Genomic DNA of *A.gambiae* mosquitoes was digested with the restriction endonucleases *Sal*I, *Cla*I, *Cla*I/*Sal*I, *Bam*HI, *Bam*HI/*Sal*I, *Eco*RI and *Eco*RI/*Sal*I (all enzymes purchased from Boehringer Mannheim). 5 μg of digested DNA per lane were separated on 0.7% agarose gels and transferred on MSI-Nitropure membranes as described (Perbal, 1988) and hybridized with ^{32}P-labelled random primed *Eco*RI fragments of *Antryp1* and *Antryp2* full-length cDNA inserts. Hybridization and initial washing steps were carried out as described (Perbal, 1988) the final washing was performed in 0.1 × SSC, 0.5% SDS at 70°C for 30 min. Filters were exposed on Kodak X-Omat AR films with Ilford Fast Tungstate screens at −80°C.

Northern blot analysis

Total RNA from fourth instar larvae, pupae, males, non-fed females and females killed 8, 24 and 40 h after blood meal was isolated by the proteinase K method as described for cDNA synthesis, followed by two precipitations in 4 M LiCl. RNA from two mosquitoes was run per lane of denaturing 1.5% formaldehyde gels and transferred on MSi-Nitropure membranes. As probe, ^{32}P end-labelled pTy1-S and pTy2-S primers were used. Hybridization and initial washing steps were done as described (Perbal, 1988), the final washing step was carried out in 0.1 × SSC, 0.5% SDS at 65°C for 20 min. Filters were exposed on Kodak X-Omat AR films with Ilford Fast Tungstate screens at −80°C.

Oligonucleotides

All oligonucleotides were synthesized with an Applied Biosystems 391 PCR-Mate machine. *Bam*HI and *Bgl*II cloning sites in pTy1-act and pTy2-act respectively are underlined, nucleotide mixtures used during synthesis at defined positions in pTry1 and pTry3 are shown in brackets, I = Inosine.

pTry1:5′-GTIGTIGGIGGITTCGA(AG)AT(ACT)GA(TC)GT-3′
pTry3:3′-AC(AG)GT(TC)CC(TA)CT(AG)(AT)(GC)ICCIC-CIGG(TC)GACCACGACCAC-5′
pTy-1S:3′-ATTGGTCTTCCTTACGTTGTTTCGGATGTCACTGAAG-5′
pTy-2S:3′-CTCGGTACTTCTAACATCGCTACGGATGTACACC-5′
pTy1-act:5′-CCCCGGATCCGCGGAGGCGCAGGCCAACCAGCGC CATCGG-3′
pTy2-act:5′-CCCCAGATCTGCGCAAGCACAACCATCCGGGCGG CATCAC-3′

Cloning and sequencing

For the cloning of the PCR fragments of the λgt11 inserts and of the EMBL3A genomic subclones, we used appropriate restriction sites of the pDS56/RBSII,6×His/E$^-$ vector, a derivative of the pDS56/RBSII,His plasmid family (Hochuli *et al.*, 1988) which permits the use of an *Eco*RI site that generates the reading frame of the λgt11 *Eco*RI cloning site. Cloning procedures were as described (Perbal, 1988), all enzymes were purchased from Boehringer Mannheim. Double-strand dideoxy sequencing was performed with the Pharmacia T7 Sequencing Kit using primers flanking the polylinker region of pDS56/RBSII,6×His/E$^-$ and various internal trypsin primers. EMBL Database accession numbers: Antryp1, X18889; Antryp2, Z18890.

Expression of recombinant trypsin in E.coli

The *Antryp1* and *Antryp2* sequences, lacking the signal peptide sequence and containing the putative activation peptide sequence were generated by PCR amplification of the cDNA clones Y9.1 and Y17.1 using the pTy1-act and pTy2-act primers in combination with a primer corresponding to the sequence of the vector region flanking the cloning site. The amplification products contained the original stop codon and the 3′ non-coding regions including the original *Eco*RI cloning site. After digestion, with *Bam*HI/*Eco*RI for *Antryp1* and with *Bgl*II/*Eco*RI for *Antryp2*, the DNA fragments were cloned in the *Bam*HI/*Eco*RI polylinker sites of the expression vector pDS56/RBSII,6×His/E$^-$, resulting in a fusion of six histidines at the amino-termini. Expression of recombinant trypsins was done in *E.coli* strain M15, carrying the plasmid pUHA1 producing *lac* repressor (Hochuli *et al.*, 1988). Induction was performed in LB medium by adding IPTG to a final concentration of 1 mM at an OD_{600} of 0.6 and further incubation for 4 h at 37°C.

Purification of recombinant proteins

The presence of the histidine stretch encoded by the expression vector confers to the recombinant proteins high affinity to nickel ions (Hochuli *et al.*, 1988). Antryp1 and Antryp2 fused to six histidines in the place of the putative signal peptide were purified, in a single step procedure, by nickel-chelate affinity chromatography (Hochuli *et al.*, 1988). In brief: 1 l of an induced culture of M15pUHA1 cells containing the pDS56/RBSII,6×His-constructs were harvested and stirred for 3 h at room temperature in 100 ml 6 M guanidine hydrochloride and 100 mM Na_2HPO_4, pH 8. The suspension was centrifuged at 10 000 *g* and the supernatant was directly loaded on a 5 ml nickel column (Ni-NTA resin, Diagen). After washing with 30 ml 6 M guanidine hydrochloride, 100 mM Na_2HPO_4, pH 8, the column was equilibrated with 20 ml 8 M urea, and 100 mM Na_2HPO_4 and 10 mM Tris, pH 8. The 6×His-proteins were eluted with a pH step gradient (until pH 4) in 8 M urea. Fractions containing the purified protein were pooled and dialysed against PBS, the fractions soluble in PBS were recovered and stored in aliquots at −20°C.

Antibody production

After purification on nickel-chelate chromatography followed by dialysis against PBS, recombinant Antryp1 and Antryp2 proteins were used to immunize Balb/c mice. The animals were immunized intraperitoneally three times with 50 μg of purified protein in complete (first immunization) and incomplete Freund's adjuvant. Three weeks after the last immunization, 1 ml of blood was collected by heart puncture. The antibody titre of the obtained sera was tested on recombinant Antryp1 and Antryp2 protein spotted on nitrocellulose.

Immunoblotting

Mosquitoes or dissected organs and tissues were ground up in sample buffer (30 mosquitoes in 600 μl), 20 μl samples were then boiled for 10 min and run on 12.5% SDS–PAGE. Gels were electroblotted onto nitrocellulose filters (Towbin *et al.*, 1979) in a blotting buffer containing 25 mM Tris, 192 mM glycine, 0.1% SDS and 20% methanol. Filters were incubated with anti-Antryp1 serum diluted 1:100 for 1 h at room temperature. Unspecific binding of antibodies to nitrocellulose was prevented by blocking the filters with 1% BSA in TBST (10 mM Tris–HCl pH 8, 150 mMNaCl and 0.05% Tween 20) for 2 h at room temperature. After extensive washing with 3 × TBST, antibodies bound to the filter were detected by a goat anti-mouse immunoglobulin (H + L) conjugated to alkaline phosphatase (Promega). Phosphatase activity was detected by incubating the filters with 0.3 mg/ml NBT (nitro-blue tetrazolium) and 0.15 mg/ml BCIP (5-bromo-4-chloro-3-indolyl phosphate) in 100 mM Tris–HCl pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$.

In vitro assay for tryptic activity

Proteolytic activity was assayed in 20 μl PBS pH 7.4, 0.1% SDS at 26°C, the temperature at which the mosquitoes were reared. As substrate, 0.8 μl human serum or 0.4 μl human RBC lysate were added to recombinant Antryp1 (0.02 μg) or Antryp2 (0.2 μg) in PBS. RBC extracts were obtained from RDCs hypotonically lysed by adding an equal volume of cold distilled water. Reactions were terminated by adding 2 × reducing sample buffer containing 1 mM of the trypsin-specific inhibitor APMSF (Boehringer Mannheim) and heating at 100°C for 10 min. For the performance of kinetic

experiments, the addition of APMSF after the various incubation times was necessary as the *Anopheles* trypsins tested still showed proteolytic activity in sample buffer (3% SDS and 3% β-mercaptoethanol final concentration).

Acknowledgements

The authors would like to thank Prof. Mario Coluzzi for his continuous support and Jacopo Vizioli and Chiara Bellomo for their collaboration. This work was funded by a grant from the Commission of the European Communities STD programme STS CT92 0044. A contribution of the Rockefeller foundation is kindly acknowledged. J.M.C. is a Wellcome Trust Senior Research Fellow in Basic Biomedical Sciences. H.-M.M. was supported by a fellowship of the Deutsche Forschungsgemeinschaft.

References

Barillas-Mury,C., Graf,R., Hagedorn,H.H. and Wells,M.A. (1991) *Insect Biochem.*, **21**, 825–831.

Bell,G.I., Quinto,C., Quiroga,M., Valenzuela,P., Craik,C.S. and Rutter,W.J. (1984) *J. Biol. Chem.*, **259**, 14265–14270.

Besansky,N.J. and Powell,J.R. (1992) *J. Med. Entomol.*, **29**, 125–128.

Billingsley,P.F. and Hecker,H. (1991) *J. Med. Entomol.*, **28**, 865–871.

Brodrick,J.W., Largman,C., Johnson,J.H. and Geokas,M.C. (1978a) *J. Biol. Chem.*, **253**, 2732–2736.

Brodrick,J.W., Largman,C., Hsiang,M.W., Johnson,J.H. and Geokas,M.C. (1978b) *J. Biol. Chem.*, **253**, 2737–2742.

Clements,A.N. (1963) *The Physiology of Mosquitoes*. Pergamon Press, Oxford.

Coluzzi,M. (1984) *WHO Bulletin*, **62**, 107–113.

Davis,C.A., Riddell,D.C., Higgins,M.J., Holden,J.J.A. and White,B.N. (1985) *Nucleic Acids Res.*, **13**, 6605–6619.

Emi,M., Nakamura,Y., Ogawa,M., Yamamoto,T., Nishide,T. Mori,T. and Matsubara,K. (1986) *Gene*, **41**, 305–310.

Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.

Graf,R. and Briegel,H. (1989) *Insect Biochem.*, **19**, 129–137.

Graf,R., Boehlen,P. and Briegel,H. (1991) *Experientia*, **47**, 603–609.

Grant,G.A., Henderson,K.O., Eisen,A.Z. and Bradshaw,R.A. (1980) *Biochemistry*, **19**, 4653–4659.

Hochuli,E., Bannwarth,W., Döbeli,H., Gentz,R. and Stüber,D. (1988) *Bio/Technology*, **6**, 1321–1325.

Jany,K.D. and Haugh,H. (1983) *FEBS Lett.*, **158**, 98–102.

Jany,K.D., Bekelar,K., Pfleiderer,G. and Ishay,J. (1983) *Biochem. Biophys. Res. Commun.*, **110**, 1–7.

Kraut,J. (1977) *Annu. Rev. Biochem.*, **46**, 331–358.

Lecroisey,A., Gilles,A.M., DeWolf,A. and Keil,B. (1987) *J. Biol. Chem.*, **262**, 7546–7551.

MacDonald,R.J., Swift,G.H., Quinto,C., Swain,W., Pictet,R.L. and Nikorits,W. (1982) *Biochemistry*, **21**, 1453–1463.

Perbal,B. (1988) *A Practical Guide to Molecular Cloning*. Wiley Interscience.

Pinsky,S.D., LaForge,K.S. and Scheele,G. (1985) *Mol. Cell. Biol.*, **5**, 2669–2676.

Rutledge,L.C., Ward,R.A. and Gould,D.J. (1964) *Mosquito News*, **24**, 407–419.

Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R. Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) *Science*, **239**, 487–491.

Stevenson,B.J., Hagenbüchle,O. and Wellauer,P.K. (1986) *Nucleic Acids Res.*, **14**, 8307–8330.

Sturchler,D. (1989) *Parasitology Today*, **5**, 39.

Titani,K., Sasagawa,T, Woodbury,R.G., Ericsson,L.H., Dorsam,H., Kraemer,M., Neurath,H. and Zwilling,R. (1983) *Biochemistry*, **22**, 1459–1465.

Towbin,K., Staehlin,T. and Gordon,J. (1979) *Proc. Natl Acad. Sci. USA*, **76**, 4350–4354.

White,G.B. (1974) *Trans. R. Soc. Trop. Med. Hyg.*, **68**, 278–298.

Willadsen,P., Riding,G.A., McKenna,R.V., Kemp,D.H., Tellam,R.L., Nielsen,J.N., Lahnstein,J., Cobon,G.S. and Gough,J.M. (1989) *J. Immunol.*, **143**, 1346–1351.

Yungdae,Y. and Davis,R.L. (1989) *Mol. Cell. Biol.*, **9**, 692–700.

Received on March 17, 1993; accepted on April 6, 1993