Insect Molecular Biology (2004) **13**(1), 9–18

Cloning and characterization of a trypsin-encoding cDNA of the human body louse *Pediculus humanus*

A. H. Kollien*[1], P. J. Waniek*[1], F. Prölss†, B. Habedank‡ and G. A. Schaub*

**Department of Special Zoology, Ruhr-University, Bochum, Germany; †Institute of Anatomy II, Albert-Ludwigs-University, Freiburg, Germany; and ‡Department of Parasitology and Tropical Veterinary Medicine, Free University, Berlin, Germany*

Abstract

From a cDNA library of the whole insect, a trypsin gene of *Pediculus humanus* has been cloned and sequenced. The 908 bp clone has an open reading frame of 759 bp, which encodes a pre-proenzyme with 253 amino acid residues. A sixteen-residue N-terminal signal peptide is followed by a twelve-residue activation peptide with putative cleavage sites at Gly16 and Tyr28. The deduced amino acid sequence has several features typical of trypsin proteases and an overall identity of 35–43% with the trypsins of several haematophagous Diptera. The 1.0 kb genomic trypsin gene contains three introns of 102, 79 and 80 nucleotides following the codons for Gly16, Gln74 and Ala155, respectively. Only a single gene seems to be present. In Northern blot analysis, unfed first instar larvae have an identical or slightly lower level of trypsin mRNA than fed adult lice, and in adults 2–24 h after the bloodmeal this gene shows a constitutive expression. After *in vitro* transcription and translation, the activation peptide is cleaved by chymotrypsin, a so far unreported phenomenon in trypsin activation.

Keywords: cDNA library, *Pediculus humanus*, lice, serine protease, trypsin.

Introduction

Sucking lice (Anoplura) are a major worldwide problem as debilitating agents and as vectors of human and animal diseases. Of about 500 known species of the suborder Anoplura, only three species are ectoparasites of humans: *Pediculus capitis* (head lice) living in head hair; *P. humanus* (body or clothing lice) occupying the clothes and passing to the body only to feed; and *Phthirus pubis* (crab lice), developing mainly in the hair of the genital region (Schaub, 2001b). All three species cause skin irritations but only *P. humanus* transmits the pathogenic bacteria, *Rickettsia prowazeki*, *Rickettsia quintana* and *Borrelia recurrentis*, which cause louse-borne typhus, trench fever and louse-borne relapsing fever, respectively. Head lice are an ancient but recently increasing problem. Current control methods are largely limited to the use of contact insecticides with the attendant problems of the development of resistance by the ectoparasites to the chemicals and the persistence of toxic chemical residues. All human lice are highly host-specific, obligate parasites, spending their entire life cycle associated with the host and feeding exclusively on human blood. Only one strain of human body lice has been adapted to feed on laboratory rabbits (Culpepper, 1948).

Although all haematophagous insects ingest the same food, the patterns and physiology of digestion are completely different between reduviid bugs, Diptera and lice. Reduviid triatomines have an alimentary tract in the form of a simple tube. All postembryonic stages are exclusively blood feeders, with one large bloodmeal required for development into the next larval instar (Schaub, 2001a). These temporary ectoparasites benefit most from taking very large blood meals, ingesting 6–12 times their original body weight, up to 3.8 ml blood in the largest triatomine (Schaub & Pospischil, 1995). The blood is stored essentially undigested in the wide anterior part of the midgut before concentration and partial enzymatic hydrolysis. Small portions are passed into the digestive and absorptive posterior part of the midgut (Kollien & Schaub, 2000). Compared with other haematophagous insects, an adult female of *Triatoma infestans* has the slowest digestion of a bloodmeal; 336 h on average (Lehane, 1991). As these haematophagous insects have an acidic midgut lumen, they digest blood via cysteine and/or aspartic proteinases, mainly cathepsins (Houseman, 1978; Houseman & Downe, 1980, 1981a,b, 1983; Kollien & Schaub, 2000).

In dipteran haematophagous insects like mosquitoes ingestion of blood and oogenesis are correlated, i.e. because eggs require 48–72 h to develop within the females, they

Received 20 June 2003; accepted after revision 3 September 2003. Correspondence: Astrid H. Kollien, Department of Special Zoology, Ruhr-University, D-44780 Bochum, Germany. Tel.: +49 234 322 4489; fax: +49 234 321 4114; e-mail: astrid.kollien@ruhr-uni-bochum.de

[1]These authors contributed equally to the study.

take blood every 2–4 days (Schaub, 2001c). Sugar meals are stored in the diverticulum while blood is taken directly into the posterior midgut. In mosquitoes digestion proceeds almost simultaneously over the entire surface of the food bolus (Lehane, 1991). Higher Diptera, like tsetse flies, store blood in the midgut and the diverticulae, before it is passed down through the digestive and absorptive middle and posterior regions of the midgut. Digestion of a bloodmeal by adult mosquitoes takes 50–60 h. This is faster than for Hemiptera, reflecting the smaller quantity of meal size (about equal to their unfed weights) (Lehane, 1991). Diptera use serine proteinases, especially trypsins and chymotrypsins, as well as esterases as principal digestive enzymes, consistent with the neutral or alkaline pH of their midgut lumen (Lehane, 1991).

Lice, in contrast to other haematophagous insects, stay permanently associated with their host and frequently require small blood meals. They store the ingested blood in the wider anterior part of the midgut, where it is also digested, before passing it rapidly through the narrow posterior part of the midgut into the hindgut. Lice feed every few hours, taking bloodmeals of only 0.2–0.3 times their unfed weights (Buxton, 1947), and thus have the fastest digestion among the haematophagous insects, *c.* 4 h (Lehane, 1991). Like the majority of haematophagous insects, they use a range of alkaline digestive proteases (Lehane, 1994). To date, a leucine aminopeptidase has been partially characterized. This enzyme possesses a pH optimum at 8, is activated by the bloodmeal, reaches maximum activity at 48 h after feeding and decreases thereafter (Ochanda *et al.*, 1998). Trypsin activity appears 24 h after the bloodmeal, earlier than that of aminopeptidases in the louse midgut (Borovsky & Schlein, 1988).

None of the genes encoding these enzymes has yet been sequenced. In the present investigation we have constructed a cDNA library of *P. humanus* and have isolated and initially characterized a gene from the human body louse with sequence similarity to the digestive enzyme trypsin.

Results

The primary size of the cDNA library of *Pediculus humanus*, constructed in λ-ZAP vector, was about 8×10^7 plaque-forming units, with an average insert size of 1.2 kb as determined from twelve random clones.

The trypsin-gene cDNA (TRY1) had an insert of 908 bp (plus poly A tail) and a single open reading frame of 759 bp with 20 bp of putative 5′ untranslated region (Fig. 1) (GenBank accession no. AY 304498). There are 101 bp of 3′ untranslated region before the start of the poly(A) tail with two AATAAA polyadenylation signal sequences 44 bp and 77 bp downstream of the TAA stop codon. According to the derived amino acid sequence, the pre-proenzyme contains 253 amino acid residues (Fig. 1). The putative signal cleavage site in the nascent trypsinogen precursor protein is at Gly16, and the putative activation cleavage site is at Tyr28 (Figs 1 and 2). The amino-terminal sequence Ile–Val–Gly–Gly indicating the start of the active trypsin is present after Tyr28. His69, Asp114 and Ser206 make up the charge relay system in the active site. The substrate-binding pocket is determined by Asp200–Ser201–Gly225. The six cysteine residues known to form the three disulphide bonds are at positions 54, 70, 177, 193, 202 and 206 (Fig. 1). When compared with matches in the GenBank database (BLASTX using the SWISSPROT database), *P. humanus* trypsin has between 34% and 44% amino acid identity with other insect serine proteases (Fig. 2).

The genomic sequence contains 1020 nucleotides (GenBank accession no. AY 308482). The gene contains three introns of 102, 79 and 80 bp after the codons for Gly, Gln and Ala at positions 16, 74 and 155, respectively (Fig. 1): intron 1: 5′-GTACTCTTTCTTTTGTTTTGATTTTGAAATTTTTTTTCTAAATTTTGAGAATCGCCATTAAATTGAGAGTGGTTGAATTAATATTTTTTATAAAAATTAGGT-3′, intron 2: 5′-GTGAGTTTTTTTTTTTTTTTTGGGAGGGGGGTTTATATTTTTAATTTAAAAAAAAAATTAGTGATTTTAATTTATAAAC-3′, and intron 3: 5′-GTGAGTAGACAAATTTTTTCACACCGAAAATAAACCGGAAATGCTTTTAAACTAACCCTCGATTAATTTAATAAAAATAG-3′.

The fluorescence intensity of DNA fragments obtained under nonspecific PCR conditions was determined across a 60–95 °C temperature gradient. Melting curves plotted on a logarithmic scale showed that all PCR products had the same melting point at about 81 °C (Fig. 3). Only a single band was visible after electrophoresis on the agarose gel (Fig. 4).

In Northern blots, the ^{32}P-labelled cDNA probe T2 hybridized to RNA from first instar larvae and adult lice 2, 6, 12 and 24 h after feeding. However, the hybridization signal was weak although a high amount of total RNA was loaded on to the gel. First instar larvae possessed an identical or slightly lower amount of trypsin mRNA than the adults. Two to 24 h after feeding, adults showed a similar level to the expression levels of actin, indicating a constitutive expression (Fig. 5).

Incubation of the *in vitro* translation product in the presence of chymotrypsin resulted in two bands of 33 and < 26 kDa (Fig. 6, lane 1). In the absence of chymotrypsin (lane 2) or in the presence of chymotrypsin and chymotrypsin inhibitor (lane 4), the uncleaved translation product was present in a single band at 33 kDa (Fig. 6, lanes 2 and 4). The intensity in lane 4 was similar to that in lane 2, with no indication of a cleavage. Both bands were of stronger intensity than that in lane 1. No band was present after incubation without translation product (Fig. 6, lane 3).

Discussion

Serine proteases are a highly conserved group of proteolytic enzymes containing a uniquely reactive serine side

Figure 1. Nucleotide and deduced amino acid sequences of *Pediculus humanus* trypsinogen cDNA (GenBank accession no. AY 304498). The putative signal cleavage site is marked with (+) and the putative activation cleavage site with an arrow (→). The catalytic triad *His–Asp–Ser* is italicized, shaded and underlined. The two polyadenylation signal AATAAA sequences and the positions of the degenerate primers as well as the three introns are marked. Other highly conserved amino acid residues presumed to be important for enzyme structure (cysteine bridges) and function (substrate binding pocket) are denoted in bold type, italicized and underlined. Further details are presented in the legend to Fig. 2 and in the text.

chain (Kraut, 1977). Because this group contains the major proteolytic enzymes, many of its members have been characterized, demonstrating that it is a large and diverse group that can be divided into twenty families (Rawlings & Barrett, 1994). One of the largest and best characterized families includes trypsin and chymotrypsin, both possessing a catalytic triad of His57–Asp102–Ser195 (bovine chymotrypsin numbering convention; Hartley, 1964). Within this three-dimensional structure, the Asp189 at the bottom of the S1 binding pocket is diagnostic of tryptic specificity (Stroud *et al*., 1974). Trypsins differ from other serine proteases by their specificity for Arg and Lys residues and their ability to activate other zymogens (Kraut, 1977). To avoid a self-digestion *in vivo*, within the cell serine proteases are translated as inactive pre-proenzymes containing a signal sequence and proenzyme regions at the amino terminus. At the site of action, e.g. in the lumen of the gut, these regions are cleaved to obtain the active protease. Serine proteases are not only involved in the digestion of 'non-self' food proteins, but also in processing and activation of 'self' proteins and enzymes, respectively (Gooding, 1966; Hagedorn & Judson, 1972; Ikeda *et al*., 1991; Steward & Govind, 1993; Xiong & Jacobs-Lorena, 1995).

Many of the key residues for trypsin structure and function (Kraut, 1977; Huber & Bode, 1978) are conserved in the *P. humanus* gene. From the translation initiation codon

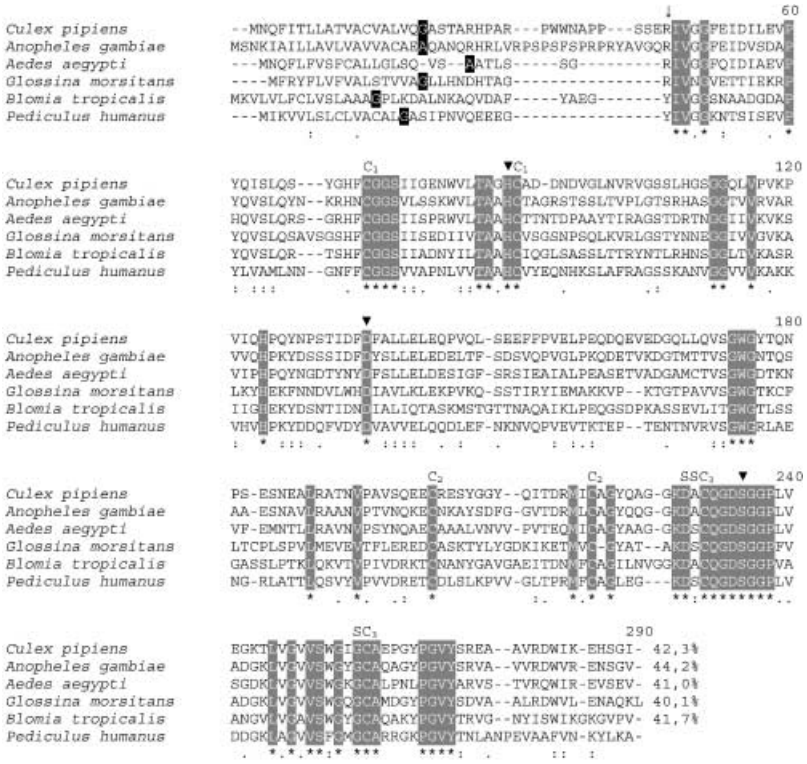

Figure 2. Comparison of the amino acid sequence of *Pediculus humanus* trypsinogen (TRY1) with those of the following other insects: *Culex pipiens* (CPU65412); *Anopheles gambiae* (*Antryp1a*) (Müller *et al.*, 1993); *Aedes aegypti* (AF487426); *Glossina morsitans* (AF252869) (Yan *et al.*, 2001); *Blomia tropicalis* (AY 090091). Identical and similar amino acid residues between *P. humanus* trypsinogen and other trypsinogens are indicated by asterisks and dots, respectively, beneath the alignment sequences. The alignment was optimized by introducing gaps using the CLUSTAL W (1.4) multiple sequence alignment program. The His, Asp and Ser residues essential for enzyme activity are indicated by triangles (▼) above the sequences. Exact matches of all amino acids in an identical position are printed in shaded boxes and marked with stars (*) below the sequence. The black box indicates the end of the predicted signal peptide (according to the respective publication or to calculations using the SIGNALP (CBS) program from the ExPASy Proteomics tools in unpublished sequences), the arrow (→) the activation peptide cleavage site. S, amino acids involved in determining enzyme substrate specificity. C_n, conserved cysteine residues, where n matches pairwise the residues predicting to be involved in S–S bond formation. Further details are given in the text.

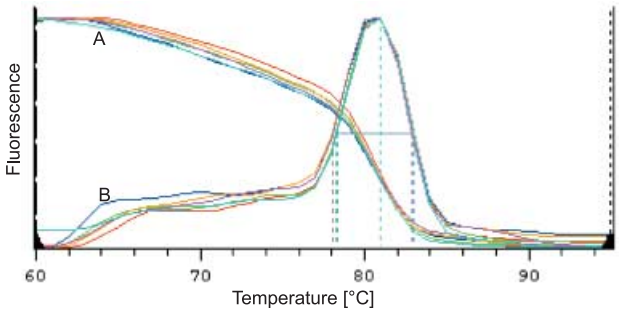

Figure 3. Melting curve analysis of the real-time PCR products, which were obtained at different annealing temperatures. The fluorescence of the DNA fragments was measured in a continuous temperature gradient between 60 and 95 °C. Fluorescence was plotted on a linear (A) or logarithmic scale (B).

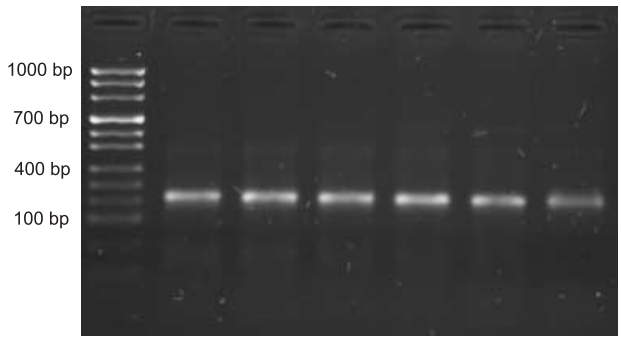

Figure 4. Analysis of PCR amplification products using annealing temperatures of 46.0, 46.9, 48.9, 51.5, 54.4 and 55.8 °C (from left to right) and markers in base pairs (bp) on the left. Amplification products were separated on 2% ethidium-stained agarose gels.

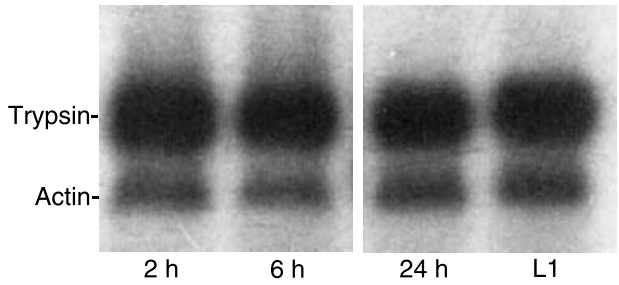

Figure 5. Northern blot analysis of trypsin and actin expression of first instar larvae and adults of *Pediculus humanus* at different times after feeding. In each lane 15 µg total RNA was separated in 1% agarose gel containing 2% formaldehyde, blotted on to a filter and hybridized with ^{32}P-labelled 759 bp probe.

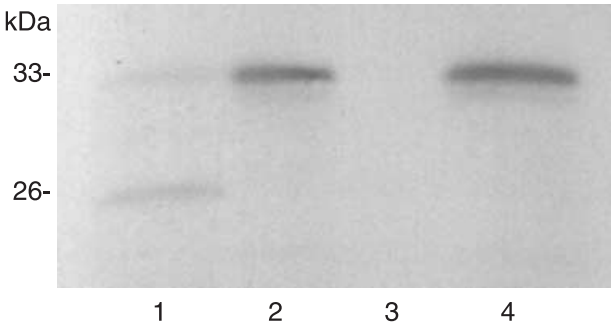

Figure 6. Separation of the *in vitro* expressed radiolabelled trypsinogen by SDS-PAGE. The translated trypsinogen was digested with chymotrypsin (lane 1). Digestion was not observed in the sample lacking chymotrypsin (lane 2) or after incubation with the chymotrypsin inhibitor chymostatin (lane 4). In the lane of the sample without substrate and chymostatin no band was visible (lane 3).

to the start of the active enzyme sequence, the derived amino acid sequences of trypsinogen genes contain a string of hydrophobic amino acids, a characteristic of signal peptides (Gierasch, 1989), which has also been found in *Aedes aegypti* trypsins (Kalhok *et al*., 1993). In the present investigation *c*. 80% of the amino acids in the N-terminal region upstream of the cleavage site of the signal peptide of *P. humanus* were hydrophobic.

Putative trypsinogens have signal and activation peptides. In insects the derived amino acid sequence from the N-terminal end to the first residue of the active trypsin lacks a region of repeated aspartic acid residues preceding the final arginine of the activation peptide, a repeat region found in vertebrates that prevents autocatalysis (Broderick *et al*., 1978). In haematophagous insects, signal and activation peptides are together 25–50 amino acids long (Fig. 2). In rats and *Drosophila melanogaster* the signal peptide has fifteen and twenty-two amino acids, respectively, and the activation peptide only eight amino acids (Craik *et al*., 1984; Davis *et al*., 1985). In *Ae. aegypti* trypsinogens consist of signal peptides of fifteen and twenty or twenty-one amino acids and activation peptides of ten and seven or six amino acids (Barillas-Mury *et al*., 1991; Kalhok *et al*., 1993). The *P. humanus* trypsinogen appears to contain a sixteen amino acid signal peptide (cleaved between Gly16 and Ala17) and a twelve amino acid activation peptide (cleaved after Tyr28) (Fig. 2), both very similar in length to other putative insect trypsinogens.

The N-terminus of the active enzyme, Ile–Val–Gly–Gly, is also conserved in many active trypsins (Fig. 2) (Kalhok *et al*., 1993; Wang *et al*., 1999). However, *Ae. aegypti* amino-terminal trypsin sequences obtained from purified proteins (Graf *et al*., 1991) and a cDNA clone (Barillas-Mury *et al*., 1991) all begin with Val. In *P. humanus* the amino-terminal sequence Ile–Val–Gly–Gly is present after Tyr28. The majority of the known trypsin-like proteinases from insects contain an activation peptide with a trypsin cleavage site after an Arg residue followed by Ile–Val–Gly–Gly–Glu motif (Gaines *et al*., 1999; Zeng *et al*., 2002). In contrast, the trypsin-like proteinase from the human body louse, *P. humanus*, contains a cleavage site for the activation peptide between the aromatic Tyr28 and Ile29, indicating that an enzyme with chymotryptic activity is responsible for cleavage of the activation peptide. This unusual activation was verified by *in vitro* transcription and a chymotrypsin cleavage assay suggesting a novel mechanism of trypsin-like proteinase activation in insects. Most recently an unpublished trypsin-like proteinase of the mite *Blomia tropicalis* (AY090091) was directly submitted to NCBI GenBank; it possesses the same chymotrypsin cleavage site.

The His, Asp and Ser making up the charge relay system in the active site are conserved (Kraut, 1977). In the majority of serine proteases, the two residues flanking the Asp are Asn and Ile. In *Ae. aegypti* Tyr and Phe are in these positions around Asp86 (Kalhok *et al*., 1993), but this differs in other trypsins of this species (e.g. Figure 2). In *P. humanus*, these two residues at Asp114 are Tyr and Val, unlike other serine proteases. The significance of this change remains to be investigated.

Another Asp, located shortly before the Ser of the catalytic triad, lies at the bottom of the substrate binding pocket: it is conserved in all trypsins and plays an important role as part of the specific crevice in stabilizing Lys and Arg residues during catalysis (Stroud *et al*., 1974; Hedstrom *et al*., 1992). All trypsins have an Asp or Glu at this position, whereas in other serine proteases the corresponding amino acid is uncharged (in chymotrypsins it is usually a Ser or Gly) (Stroud *et al*., 1974). In the trypsinogens of the Diptera *Simulium vittatum*, *Culex pipiens*, *Glossina morsitans* and *Ae. aegypti* this Asp176 is present (Kalhok *et al*., 1993), but it is absent in the trypsinogen sequence of the latter species reported by Barillas-Mury *et al*. (1991) (Fig. 2). The sequence of *P. humanus* trypsinogen also has an Asp in position 178.

This negatively charged Asp interacts ionically with the positively charged Lys or Arg residues of the substrate and lines the substrate binding pocket, Asp–Ser–Gly (Figs 1 and 2). In *P. humanus* the substrate specificity is determined by Asp178–Ser179–Gly225 residues that establish ionic and hydrogen bonds with lysyl and arginyl side chains.

Of the twelve cysteine residues known to form the disulphide bonds in vertebrate trypsins (Kauffman, 1965), invertebrate trypsins contain only three cysteine bridges, lacking a bridge from amino acids 136 to 201 (Davis *et al*., 1985). These six cysteines are conserved in *C. pipiens* (U65412), *Anopheles gambiae* (Müller *et al*., 1993), *Ae. aegypti* (Kalhok *et al*., 1993), *S. vittatum* (Ramos *et al*., 1993), *G. morsitans* (Yan *et al*., 2001) and also in *P. humanus* (Fig. 2).

Comparing the derived amino acid sequence of the trypsinogen of *P. humanus* with those of vertebrate and invertebrate serine proteases, considerable similarities are evident, particularly at the conserved catalytic residue sites (Fig. 2). The mature trypsins of haematophagous insects like *C. pipiens*, *An. gambiae*, *Ae. aegypti*, *S. vittatum* and *G. morsitans* contain 264, 275, 254, 247, 255 amino acid residues, respectively (Fig. 2); that of the louse, with 253 amino acids, is quite similar in size.

The amino acid sequence of the *P. humanus* trypsin has 46% similarity with the trypsin of *Drosophila erecta* (P54629) and a slightly lower identity (44%) to that of *D. melanogaster* (P42279). Comparisons with other trypsins of haematophagous insects indicate high similarity levels of 42% and 44% between the louse and the mosquitoes *C. pipiens* and *An. gambiae*, respectively. The identities between the louse and the Diptera *Ae. aegypti* and *G. morsitans* are 41 and 40%, respectively.

Considering the genomic sequences, in the majority of insect trypsins no introns were found (summarized by

Noriega *et al*., 1996). Thus far, the only intron was sequenced in the early trypsin gene of *Ae. aegypti* (Noriega *et al*., 1996). The intron interrupts the codon Val at position 18 of the protein, and this Val is located toward the end of the putative signal sequence of the protein. Mammalian trypsin genes typically contain 5–7 introns, one of which is in a position similar to the *Ae. aegypti* intron described above (Swift *et al*., 1984; Rogers, 1985; Neurath, 1986). *P. humanus* is the first insect in which the trypsin gene has been shown to have three introns. The position of the first intron after Gly16 is located at a similar position toward the end of the signal peptide sequence as described for *Ae. aegypti* and other serine proteinases of mammals. The positions of the others are not related to those of the mammalian trypsinogens.

The Southern blot technique has been used to estimate the number of genes related to a candidate present in the genome. The genomes of many Diptera, e.g. *Drosophila* and mosquitoes, contain trypsin genes as part of multimember gene families (Davis *et al*., 1985; Kalhok *et al*., 1993; Barillas-Mury & Wells, 1993; Müller *et al*., 1995). In the mosquito *An. gambiae*, trypsin genes are arranged as a tightly clustered gene family consisting of seven related coding sequences *Antryp* 1–7 (Müller *et al*., 1995). In *Ae. aegypti,* Southern analysis of genomic DNA indicates that trypsin is encoded by a multigene family (Kalhok *et al*., 1993) and three typsin genes have been reported (Barillas-Mury *et al*., 1991; Kalhok *et al*., 1993), but additional trypsin-related genes may exist (Kalhok *et al*., 1993). A clustered trypsin-like serine protease gene family consisting of four homologous members has been isolated from the sheep blowfly *Lucilia cuprina* (Casu *et al*., 1994). In the blackfly *S. vittatum,* Southern blot analysis indicates that the genome contains at least two trypsin genes (Xiong & Jacobs-Lorena, 1995).

Instead of the usual Southern blot technique in the present investigation, we used a different approach to estimate the number of trypsin genes present in the genome of *P. humanus*. The amplification of genomic DNA using non-specific conditions, degenerate primers and different annealing temperatures should result in an amplification of different PCR products. Analysis of the melting curves of these products should then clarify whether there is more than one PCR product. In *P. humanus* the PCR product(s) showed a single melting curve and a single band on agarose gel, indicating the presence of a single copy of the trypsin encoding gene. The identity was verified by sequence analysis.

Digestive enzymes are expressed by haematophagous insects in a time- and tissue-specific manner. *Ae. aegypti* possesses three midgut trypsin genes, described as early and late trypsins according to their temporal expression. Two late trypsins are induced by blood feeding whereas early trypsin is constitutively expressed only in females (Barillas-Mury *et al*., 1991; Kalhok *et al*., 1993). The trypsin gene cluster in *An. gambiae* has two genes induced by feeding, whereas the other five genes are all constitutively expressed (Müller *et al*., 1995). In the blackfly *S. vittatum* trypsin and carboxypeptidase genes are induced in the midgut after ingestion of the blood meal (Ramos *et al*., 1993). The expression pattern of the trypsin gene of *P. humanus* 2–24 h after feeding resembles that of the late *An. gambiae* trypsin genes or the early *Ae. aegypti* genes, all showing a constitutive expression. These observations are consistent with the biology of lice, which feed several times a day, so that they permanently have blood in their guts to digest; this requires a consistent level of digestive enzymes. In addition, unfed first instar larvae are prepared to suck blood and possess an identical or slightly lower expression than adults.

The biochemical mechanisms that regulate the expression and secretion of trypsin and aminopeptidase in the mosquito midgut have been studied (Lemos *et al*., 1996; Noriega & Wells, 1999). In adults of *Ae. aegypti*, transcription of early trypsin genes begins a few hours after the moult and is controlled by the juvenile hormone level; mRNA is stored in the midgut epithelial cells and is translated after a blood meal (Noriega & Wells, 1999). The synthesis of this early trypsin in *Ae. aegypti* is probably controlled by mechanisms acting at the mRNA translation level, as translation inhibitors inhibited the induction of trypsin activity after feeding (Felix *et al*., 1991). The mechanisms controlling early trypsin synthesis in *Anopheles* seem to differ from those of *Ae. aegypti*, as in *Anopheles* early trypsins are stored as zymogen in the gut epithelium before feeding and their secretion and activation is largely independent of the protein content in the ingested meal (Hörler & Briegel, 1995; Müller *et al*., 1995). Transcription of late trypsin genes in *Ae. aegypti* is regulated by fragments of proteins generated by the proteolysis of blood proteins by early trypsin; the rate of transcription is correlated with the amount of protein in the blood meal (Noriega & Wells, 1999). Because the time pattern of the activity of digestive enzymes of lice is unknown, future studies will need to clarify whether the intermittent feeding of lice, which differs from that of mosquitoes, correlates with a constitutive expression and translation.

In order to understand the mechanisms controlling the activity of trypsin in *P. humanus*, it must be clarified whether chymotrypsins act in a regulatory manner. These genes and the proteases they encode, as well as others, need to be further characterized to identify their biological roles in *P. humanus*, especially because digestive proteases of insects have been considered as immunological targets for vaccine development against different ectoparasites (Müller *et al*., 1993; Elvin *et al*., 1993; Casu *et al*., 1994). Trypsins of *P. humanus* have also been suggested to be a good vaccination target (Mumcuoglu *et al*., 1997).

Experimental procedures

Insects

A strain of the human body louse, *P. humanus*, adapted to feed on rabbits was reared at 31 ± 1 °C and 70–80% relative humidity and on a 23 h/1 h light/dark cycle (Habedank & Schrader, 2001). The lice were fed on rabbits in daily intervals, and during the weekend they were stored for 2 days at 25 °C. First instar larvae were collected in 12 h intervals after hatching, killed by exposure to −80 °C and stored at this temperature. Adult lice were killed at different times after feeding.

RNA isolation

Total RNA and mRNA were isolated from *P. humanus* using the RNeasy® Mini Kit and Oligotex™ mRNA Mini Kit (Qiagen, Hilden) following the manufacturer's protocol. One adult body louse has an average body weight of 1 mg and an average amount of 1 µg of total RNA, about 1% of this RNA being polyA+ mRNA.

Genomic DNA isolation

Genomic DNA was isolated from *P. humanus* by grinding twenty adult lice in liquid N_2 with 2 ml lysis buffer (10 mM Tris-HCl (pH 7.5); 2% SDS, 100 mM NaCl; 1 mM EDTA). After three extractions, first with an equal volume of phenol, then of phenol–chloroform, and finally of chloroform, an isopropanol precipitation followed. The genomic DNA was washed in ethanol and dried at 37 °C, and the pellet was dissolved in water.

Primer selection, PCR, cloning and sequencing

Degenerate primers for PCR were derived from conserved sequence regions in trypsinogens from *Glossina morsitans* (AF252869), *Drosophila melanogaster* (Wang *et al*., 1999) and *Culex pipiens* (CPU65412). The degenerate forward primer was designed from the amino acid sequences –VSGWGT– (positions 149–155 in *D. melanogaster*), and the degenerate reverse primer was based on the amino acid sequences –KDSCQGD– (positions 202–208 in *G. morsitans*). The reverse transcriptase reaction was performed using an oligo-dT_{18}VN primer with mRNA from *P. humanus*. The cDNA product was amplified by conventional polymerase chain reaction (PCR) using primers TRY1-F (5′-GTNWSNGGNTGGGGNACN-3′) and TRY2-R (5′-TCNCCYTGRCANGCRTCYTT-3′) (W = A + T, S = G + C, R = A + G, Y = C + T, N = A + G + T + C). A fragment of the predicted size – approximately 180 bp – was cloned into pGEM®-T Easy vector (Promega, Madison) following the manufacturer's instructions. Sequence analysis of the clone T1 (MWG-BIOTECH, Ebersberg) showed that a trypsin sequence had been cloned.

Specific primers for the trypsinogen gene were designed based on the 5′- and 3′-ends of the complete cDNA sequence of the trypsinogen of *P. humanus* containing *Bam*HI and *Hin*dIII restriction sites at the 5′-end. cDNA and genomic DNA were used as template for PCR with TRY3-F (5′-GGATCCATGATTAAGGTTGTTCTTTCT-3′) and TRY4-R (5′-AAGCTTGGCTTTAAGGTATTTGTTGAC-3′) primers. When cDNA was used, a fragment of about 900 bp was obtained by PCR and cloned, and the clone T2 was sequenced. By PCR with gDNA as template a fragment of about 1 kb was amplified and cloned into pGEM®-T Easy vector (Promega); this clone T3 was then sequenced.

Degenerate actin primers Actin-F (5′-TGGAAYAAYATGGARAARATHTGG-3′) and Actin-R (5′-ATRTCNACRTCRCAYTTGATDAT-3′) (R = A + G, Y = C + T, N = A + G + T + C, H = A + C + T, D = A + G + T) were used to amplify a cDNA fragment of approximately 600 bp, which was then cloned and sequenced.

cDNA library construction and screening

Approximately 500 adult lice from different feeding conditions were used for mRNA isolation to obtain sufficient material (5 µg mRNA) for the construction of a *P. humanus* cDNA library using Uni-ZAP® XR (Stratagene, Amsterdam) and packaged with Gigapack III Gold (Stratagene) following the manufacturer's protocols with slight modifications (Pröls *et al*., 1998, 2001; Kollien *et al*., 2003). A cDNA fragment mixture, obtained by PCR at low annealing temperature with TRY1-F and TRY2-R as the primers and cDNA as PCR template (see above), was labelled with ^{32}P using the Random Prime™ Kit (Ambion, Huntingdon) following the manufacturer's protocol. Prehybridization and hybridization were performed at 68 °C for 2 h and overnight, respectively (Kollien *et al*., 2003). The sizes of the inserts in pBluescript® SK^+ were monitored by PCR with M13 forward and reverse primers. The largest inserts were then fully sequenced (MWG-BIOTECH).

Melting curve analysis of real-time PCR products

An alternative approach to Southern blots was used to analyse the number of trypsin genes present in the genome of *P. humanus*. If genomic DNA is amplified using nonspecific conditions, degenerate primers and different annealing temperatures, different PCR products can be obtained. A continuous increase in temperature up to 95 °C will melt these PCR products at different temperatures depending on their length and nucleotide composition, and this can be determined fluorometrically. Analysis of the melting curves will show how many DNA fragments were amplified and the products can be separated on a gel and then identified by sequencing.

Nonspecific PCR fragments were amplified using the Opticon II Real Time Cycler (MJ Research, Waltham), the QuantiTect™ SYBR® Green RT-PCR Kit (Qiagen), genomic DNA (genomic DNA purification kit, MBI Fermentas, St Leon-Rot), degenerate cysteine proteinase oligonucleotides (TRY1-F, TRY2-R), different annealing temperatures (46.0, 46.9, 48.9, 51.5, 54.4 and 55.8 °C) and forty cycles. After a continuous increase of the temperature to 95 °C the melting curves of the different PCR products were determined (494–521 nm/60–95 °C). The PCR products obtained were separated on a 2.0% agarose gel.

Northern blot analysis

Fifteen micrograms of total RNA of unfed first instars or of adults 2, 6 and 24 h after feeding was separated using 1% agarose gel containing 2% formaldehyde. This was repeated twice, once without first instars but adults 12 h after feeding. The RNA was then transferred to Hybond™-N^+ nylon membranes (Amersham Biosciences, Freiburg) and hybridized to the probe following the manufacturer's instructions. DNA probe T2 was ^{32}P-labelled using the Random Prime™ Kit (Ambion). As actin is a 'housekeeping gene' the actin-PCR-fragment was used in Northern blots as loading control. Filters were washed in 2 × SSC, 0.1% SDS, at 68 °C and exposed to an X-ray film for 24 and 48 h. The quotient of the intensity of the hybridization signals of lysozyme/intensity of the actin signal was calculated after densitometry using Image Master software (Kollien *et al*., 2003).

Sequence and similarity analysis

An analysis of the sequence of *P. humanus* trypsin was performed using the BLASTX v.2.1.1 program (Altschul *et al*., 1997) and comparing it with the SWISSPROT database. In order to calculate its similarity with other trypsins, complete sequences of trypsins of haematophagous arthropods were obtained from the SWISSPROT database available at the NBCI website. The respective insects and accession numbers or references of trypsins are: *C. pipiens* late trypsin (CPU65412), *An. gambiae* trypsin *Antryp1a* (Z18889) (Müller *et al*., 1993), *Ae. aegypti* early trypsin-like (AF487426), *Blomia tropicalis* trypsin (AY090091) and *G. morsitans* trypsin (AF252869) (Yan *et al*., 2001). Multiple sequence alignment was carried out using the CLUSTALW program (v.3.2). In *Ae. aegypti* and *An. gambiae* only those sequences showing the highest similarities were chosen. Because the 5′ and 3′ regions of the *S. vittatum* trypsin gene have not been sequenced, the gene of this species was not used in the final alignment. The theoretical cleavage site of the signal peptide was obtained by using the SIGNALP (CBS) program from the ExPASy Proteomics tools and after von Heijne (1986).

In vitro *transcription and chymotrypsin assay*

The trypsin encoding gene was amplified by PCR using the specific oligonucleotides TRY3-F and TRY4-R and 57 °C annealing temperature. The PCR product, containing a *Bam*HI at the 5′-end and a *Hin*dIII restriction site at the 3′-end, was cloned into the pGEM®-T Easy vector (Promega). After restriction with *Bam*HI and *Hin*dIII, the fragment was analysed on a 1% low-melting-point agarose gel (Biozym, Oldendorf). The band with the expected size of about 800 bp was excised from the gel using a QIAquick™ Gel Extraction Kit (Qiagen) and subcloned into the *Bam*HI and *Hin*dIII restricted pcDNA 3.1(–) vector (Invitrogen, Karlsruhe). The recombinant plasmid was sequenced in order to verify the identity of the subcloned sequence, linearized with *Hin*dIII and purified (QIAquick™ PCR purification kit, Qiagen). For the transcription/translation reaction, the STP3® *in vitro* transcription kit (Novagen, Madison), 0.5 mg of the linearized plasmid and 2 MBq L-[^{35}S]methionine (Amersham Biosciences) were used. The *in vitro* transcription and translation reactions were performed following the manufacturer's protocol.

Chymotrypsin digestion was performed by incubating 40 µl of the translation mix in the presence of 38 mM Tris-HCl (pH 7.8), 53 mM $CaCl_2$ and 51 mU chymotrypsin (Sigma, Taufkirchen) at 25 °C for 1 h. Controls were performed by incubating 40 µl of the translation products in the presence of 51 mU chymotrypsin and the chymotrypsin inhibitor chymostatin (Sigma) at a final concentration of 100 mM (Sigma). Another control was performed without chymotrypsin or the substrate (translation product). The products of these reactions were separated by SDS-PAGE in 15% polyacrylamide gel and exposed overnight to an X-Omat AR film (Sigma).

Acknowledgements

We thank Dr S. Pöggeler for helpful suggestions on molecular biological techniques, Mrs M. Ast for technical assistance and Drs R. Cassada, H.-M. Müller and A.J. Nisbet for correcting the English style and also critical suggestions on the manuscript. This work was supported by the 'Anschubfinanzierung für junge Nachwuchswissenschaftler' by the Ruhr-University Bochum.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-PLAST: a new generation of protein database search programs. *Nucleic Acid Res* **25**: 3389–3402.

Barillas-Mury, C., Graf, R., Hagedorn, H.H. and Wells, M.A. (1991) cDNA and deduced amino acid sequence of a blood meal-induced trypsin from the mosquito, *Aedes aegypti*. *Insect Biochem* **21**: 825–831.

Barillas-Mury, C. and Wells, M.A. (1993) Cloning and sequencing of bloodmeal-induced late trypsin gene from the mosquito *Aedes aegypti* and characterization of the upstream regulatory region. *Insect Mol Biol* **2**: 7–12.

Borovsky, D. and Schlein, Y. (1988) Quantitative determination of trypsinlike and chymotrypsinlike enzymes in insects. *Arch Insect Biochem Physiol* **8**: 249–260.

Broderick, J.W., Largeman, C., Johnson, J.H. and Geokas, M.C. (1978) Structural basis for the specific activation of human cationic trypsin by human enteropeptidase. *J Mol Chem* **253**: 2737–2742.

Buxton, P.A. (1947) *The Louse*. Edward Arnold, London.

Casu, R.E., Jarmey, J.M., Elvin, C.M. and Eisemann, C.H. (1994) Isolation of a trypsin-like serine protease gene family from the sheep blowfly *Lucilia cuprina*. *Insect Mol Biol* **3**: 159–170.

Craik, C.S., Choo, Q.-L., Swift, G.H., Quinto, C., MacDonald, R.J. and Rutter, W.J. (1984) Structure of two related rat pancreatic trypsin genes. *J Biol Chem* **259**: 14255–14264.

Culpepper, G.H. (1948) Rearing and maintaining a laboratory colony of body lice on rabbits. *Am J Trop Med Hyg* **28**: 499–504.

Davis, C.A., Riddell, D.C., Higgins, M.J., Holden, J.J.A. and White, B.N. (1985) A gene family in *Drosophila melanogaster* coding for trypsin-like enzymes. *Nucleic Acid Res* **13**: 6605–6619.

Elvin, C.M., Whan, V. and Riddles, P.W. (1993) A family of serine protease genes expressed in adult buffalo fly (*Haematobia irritans exigua*). *Mol Gen Genet* **240**: 132–139.

Felix, C.R., Betschart, B., Billingsley, P.F. and Freyvogel, T.A. (1991) Post-feeding induction of trypsin in the midguts of *Aedes aegypti* L. (Diptera: Culicidae) is separable into two cellular phases. *Insect Biochem* **21**: 197–203.

Gaines, P.J., Sampson, C.M., Rushlow, K.E. and Stiegler, G.L. (1999) Cloning of a family of serine protease genes from the cat flea *Ctenocephalides felis*. *Insect Mol Biol* **8**: 11–22.

Gierasch, L.M. (1989) Signal sequences. *Biochemistry* **28**: 923–930.

Gooding, R.H. (1966) Physiological aspects of digestion of the blood meal by *Aedes aegypti* (Linnaeus) and *Culex fatigans* (Wiedemann). *J Med Ent* **3**: 53–60.

Graf, R., Boehlen, P. and Briegel, H. (1991) Structural diversity of trypsin from different mosquito species feeding on vertebrate blood. *Experientia* **47**: 603–609.

Habedank, B. and Schrader, G. (2001) Evaluation of nutrition media derived from human blood transfusion units DFP, EC and BC for the *in vitro* feeding of *Pediculus humanus corporis* (Anoplura: Pediculidae). *Altex* **18**: 170.

Hagedorn, H.H. and Judson, C.L. (1972) Purification and site of synthesis of *Aedes aegypti* yolk proteins. *J Exp Zool* **182**: 367–378.

Hartley, B.S. (1964) Amino-acid sequence of bovine chymotrypsinogen-A. *Nature* **201**: 1284–1287.

Hedstrom, L., Szilagyi, L. and Rutter, W.J. (1992) Converting trypsin to chymotrypsin: the role of surface loops. *Science* **255**: 1249–1253.

von Heijne, G., (1986) A new method for predicting signal sequence cleavage sites. *Nucleic Acid Res* **14**: 4683–4690.

Hörler, E. and Briegel, H. (1995) Proteolytic enzyms of female *Anopheles* – biphasic synthesis, regulation, and multiple feeding. *Arch Insect Biochem Physiol* **28**: 189–205.

Houseman, J.G. (1978) A thiol-activated digestive proteinase from adults of *Rhodnius prolixus* Stål (Hemiptera: Reduviidae). *Can J Zool* **56**: 1140–1143.

Houseman, J.G. and Downe, A.E.R. (1980) Endopeptidase activity in the posterior midgut of *Rhodnius prolixus* Stål (Hemiptera: Reduviidae). *Insect Biochem* **10**: 363–366.

Houseman, J.G. and Downe, A.E.R. (1981a) Exopeptidase activity in the posterior midgut of *Rhodnius prolixus* Stål (Hemiptera: Reduviidae). *Insect Biochem* **5**: 579–582.

Houseman, J.G. and Downe, A.E.R. (1981b) Identification and partial characterization of digestive proteinases from *Triatoma phyllosoma pallidipennis* Stål (Hemiptera: Reduviidae). *Comp Biochem Phys* **70B**: 713–717.

Houseman, J.G. and Downe, A.E.R. (1983) Activity cycles and the control of four digestive proteinases in the posterior midgut of *Rhodnius prolixus* Stål (Hemiptera: Reduviidae). *J Insect Phys* **29**: 141–148.

Huber, R. and Bode, W. (1978) Structural basis of the activation and action of trypsin. *Acc Chem Res* **11**: 114–122.

Ikeda, M., Yaginuma, T., Kobayashi, M. and Yamashita, O. (1991) cDNA cloning, sequencing and temporal expression of the protease responsible for vitellin degradation in the silkworm *Bombyx mori*. *Comp Biochem Phys* **99B**: 405–411.

Kalhok, S.E., Tabak, L.M., Prosser, D.E., Brook, W., Downe, A.E.R. and White, B.N. (1993) Isolation, sequencing and characterization of two cDNA clones coding for trypsin-like enzymes from the midgut of *Aedes aegypti*. *Insect Mol Biol* **2**: 71–79.

Kauffman, D.L. (1965) The disulphide bridges of trypsin. *J Mol Biol* **12**: 929–932.

Kollien, A.H., Fechner, S., Waniek, P.J. and Schaub, G.A. (2003) Isolation and characterization of a cDNA encoding for a lysozyme from the gut of the reduviid bug *Triatoma infestans*. *Arch Insect Biochem Physiol* **53**: 134–145.

Kollien, A.H. and Schaub, G.A. (2000) The development of *Trypanosoma cruzi*. Triatominae. *Parasitol Today* **16**: 381–387.

Kraut, J. (1977) Serine proteases: structure and mechanisms of catalysis. *Ann Rev Biochem* **8**: 331–358.

Lehane, M.J. (1991) *Blood-Sucking Insects*. Harper Collins Academic, London.

Lehane, M.J. (1994) Digestive enzymes, haemolysins and symbionts in the search for vaccines against blood-sucking insects. *Int J Parasitol* **24**: 27–32.

Lemos, F.J.A., Cornel, A.J. and Jacobs-Lorena, M. (1996) Trypsin and aminopeptidase gene expression is affected by age and food composition in *Anopheles gambiae*. *Insect Biochem Mol Biol* **26**: 651–657.

Müller, H.-M., Catteruccia, F., Vizioli, J., della Torre, A. and Crisanti, A. (1995) Constitutive and blood meal-induced trypsin genes in *Anopheles gambiae*. *Exp Parasitol* **81**: 371–385.

Müller, H.-M., Crampton, J.M., della Torre, A., Sinden, R. and Crisanti, A. (1993) Members of a trypsin family in *Anopheles gambiae* are induced in the gut by blood meal. *EMBO J* **12**: 2891–2900.

Mumcuoglu, K.Y., Ben-Yakir, D., Ochanda, J.O., Miller, J. and Galun, R. (1997) Immunization of rabbits with faecal extract of *Pediculus humanus*, the human body louse: effects on louse development and reproduction. *Med Vet Entomol* **11**: 315–318.

Neurath, H. (1986) Limited proteolysis, domains and the evolution protein structure. *Chemica Scripta* **26B**: 221–227.

Noriega, F.G., Wang, X.-Y., Pennington, J.E., Barillas-Mury, C.V. and Wells, M.A. (1996) Early trypsin, a female-specific midgut protease in *Aedes aegypti*: isolation, aminoterminal sequence determination, and cloning and sequencing of the gene. *Insect Biochem Mol Biol* **26**: 119–126.

Noriega, F.G. and Wells, M.A. (1999) A molecular view of trypsin synthesis in the midgut of *Aedes aegypti*. *J Insect Phys* **45**: 613–620.

Ochanda, J.O., Oduor, E.A.C., Galun, R., Imbuga, M.O. and Mumcuoglu, K.Y. (1998) Partial characterization and post-feeding activity of midgut aminopeptidase in the human body louse, *Pediculus humanus humanus*. *Physiol Entomol* **23**: 382–387.

Pröls, F., Loser, B. and Marx, M. (1998) Differential expression of osteopontin, PC4, and CEC5, a novel mRNA species, during *in vitro* angiogenesis. *Exp Cell Res* **239**: 1–10.

Pröls, F., Mayer, M.P., Renner, O., Czarnecki, P.G., Ast, M., Gässler, C., Wilting, J., Kurz, H. and Christ, B. (2001) Upregulation of the cochaperone Mdg1 in endothelial cells is induced by stress and during *in vitro* angiogenesis. *Exp Cell Res* **269**: 42–53.

Ramos, A., Mahowald, A. and Jacobs-Lorena, M. (1993) Gut-specific genes from the blackfly *Simulium vittatum* encoding trypsin-like and carboxypeptidase-like proteins. *Insect Mol Biol* **1**: 149–163.

Rawlings, N.D. and Barrett, A.J. (1994) Families of serine peptidases. *Methods Enzymol* **244**: 19–61.

Rogers, J. (1985) Exon shuffling and intron insertion in serine protease genes. *Nature* **315**: 458–459.

Schaub, G.A. (2001a) Kissing bugs. In *Encyclopedic Reference of Parasitology*, Vol. 1. *Biology, Structure, Function*, 2nd edn (Mehlhorn, H., ed.), pp. 326–329. Springer-Verlag, Heidelberg.

Schaub, G.A. (2001b) Lice. In *Encyclopedic Reference of Parasitology*, Vol. 1. *Biology, Structure, Function*, 2nd edn (Mehlhorn, H., ed.), pp. 339–343. Springer-Verlag, Heidelberg.

Schaub, G.A. (2001c) Mosquitoes. In *Encyclopedic Reference of Parasitology*, Vol. 1. *Biology, Structure, Function*, 2nd edn (Mehlhorn, H., ed.), pp. 378–384. Springer-Verlag, Heidelberg.

Schaub, G.A. and Pospischil, R. (1995) Chagas (Teil 2) – Epidemiologie, Vektorbiologie und Bekämpfung. *Reisemedizin No.* **2**: 19–20.

Steward, R. and Govind, S. (1993) Dorsal–ventral polarity in the *Drosophila* embryo. *Curr Op Genet Dev* **3**: 556–561.

Stroud, R.M., Kay, L.M. and Dickerson, R.E. (1974) The structure of bovine trypsin: electron density maps of the inhibited enzyme at 5 Å and 2.7 Å resolution. *J Mol Biol* **83**: 185–208.

Swift, G.H., Craik, C.S., Stary, S.J., Quinto, C., Lahaie, R.G., Rutter, W.J. and MacDonald, R.J. (1984) Structure of the two related elastase genes expressed in the rat pancreas. *J Biol Chem* **259**: 14271–14278.

Wang, S., Magoulas, C. and Hickey, D. (1999) Concerted evolution within a trypsin gene cluster in *Drosophila*. *Mol Biol Evol* **16**: 1117–1124.

Xiong, B. and Jacobs-Lorena, M. (1995) The blackfly *Simulium vittatum* trypsin gene: characterization of the 5′-upstream region and induction by the blood meal. *Exp Parasitol* **81**: 363–370.

Yan, J., Cheng, Q., Li, C.-B. and Aksoy, S. (2001) Molecular characterization of two serine proteases expressed in gut tissue of the African trypanosome vector, *Glossina morsitans morsitans*. *Insect Mol Biol* **10**: 47–56.

Zeng, F., Zhu, Y.-C. and Cohen, A.C. (2002) Molecular cloning and partial characterization of a trypsin-like protein in salivary glands of *Lygus hesperus* (Hemiptera: Miridae). *Insect Biochem Mol Biol* **32**: 455–464.