

Characterization of genes expressed in the salivary glands of the tsetse fly, *Glossina morsitans morsitans*

S. Li, J. Kwon and S. Aksoy

Yale University School of Medicine, Department of Epidemiology and Public Health, Section of Vector Biology, New Haven, CT, USA

Abstract

Salivary gland products of haematophagous insects including tsetse flies (Diptera: Glossinidae) are involved in antihaemostasis to allow for efficient blood feeding. In addition, salivary products of tsetse are thought to indirectly support the metacyclogenesis and eventual transmission of the African trypanosome protozoan parasites to their mammalian hosts. We have previously characterized the major anticoagulant, Tsetse Thrombin Inhibitor (TTI), from salivary extracts, and described molecular aspects of its cDNA from a *Glossina morsitans morsitans* salivary gland cDNA library. In addition, a family of two related genes with growth factor and adenosine-deaminase motifs (*TSGF-1* and *TSGF-2*) have also been described. Here, we report on the molecular aspects of three different cDNAs and their putative products expressed in salivary glands: cDNAs *TAg5*, *Tsal1* and *Tsal2*. The full-length transcript encoded by *Tsetse Antigen 5 (TAg5)* cDNA is 926 bp excluding the poly(A) stretch, and has an open reading frame of 259 amino acids that can encode for a protein of 28 925 Da. The putative product of *TAg5* shows extensive similarities to cDNAs characterized from *Drosophila* (*Agr* and *Agr2*) and sandfly *Lutzomyia* (*LuLoAG5*). The cDNAs *Tsal1* and *Tsal2* are predicted to encode for mature proteins of 45 612 Da (399 amino acids) and 43 930 Da (389 amino acids), respectively, and their putative products exhibit over 42% identity to one another. The N terminus of each putative protein contains a hydrophobic region with signal peptide characteristics indicating that they may be secretory in

nature. Transcripts specific for *TAg5* and *Tsal2* genes can be detected in all developmental stages of tsetse while *Tsal1* expression is limited to adult and larval stages. A reverse transcription polymerase chain reaction based amplification approach indicates that *TAg5* transcripts can be detected from proventriculus and midgut tissues of the fly in addition to salivary glands, while *Tsal1* and *Tsal2* expression is restricted to salivary gland and proventriculus. The salivary glands of adult males are found to express higher levels of *TAg5* and *Tsal2* in comparison to females while no significant sex-based difference is observed for *Tsal1* expression. The expression of these cDNAs in different tsetse species (*G. m. morsitans*, *Glossina austeni* and *Glossina fuscipes*) shows wide variations.

Keywords: salivary gland, genus *Glossina*, Antigen-5, African trypanosome, vector biology.

Introduction

Tsetse flies (Diptera: Glossinidae) are medically and agriculturally important vectors as they are involved in the transmission of the African trypanosome parasite. Recently, tsetse-transmitted diseases have been on the rise, causing severe economic hardships and public health crisis in already stressed communities in sub-Saharan Africa (Ekwanzala *et al.*, 1996). The successful transmission of trypanosomes involves two developmental stages in their tsetse host: first, differentiation of the ingested mammalian form parasites to insect-stage procyclic cells in midgut and proventriculus tissues, and second, the invasion and maturation of procyclic cells to metacyclic forms in salivary glands. The metacyclic forms are subsequently introduced to the bite site with saliva during feeding. It has been shown that attachment of the parasites via their flagella to the microvilli of the epithelial cells lining the salivary gland lumen is necessary for the generation of infective metacyclic forms (Vickerman *et al.*, 1988). It has also been shown that the gut-form procyclic cells can undergo metacyclogenesis *in vitro* when cultivated in medium containing explants of tsetse head salivary glands (Cunningham & Honigberg, 1977). Recently, proventriculus products have also been implicated for the successful maturation of trypanosomes (Van den Abbeele *et al.*, 1995;

Received 26 April 2000; accepted after revision 24 August 2000. Correspondence: Serap Aksoy, PhD., Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College St., 606 LEPH, New Haven, CT 06510, USA. Tel.: (203) 737-2180; fax: (203) 785-4782; e-mail: serap.aksoy@yale.edu

Van den Abbeele *et al.*, 1999). The specific tsetse proteins trypanosomes interact with and the nature of these interactions, however, are unknown. While different tsetse species exhibit wide differences in their ability to establish mature salivary gland infections, the genetic basis of these variations are unknown.

In order to ensure efficient feeding, haematophagous insects have evolved a diverse array of strategies to interfere with host haemostasis among which are substances including anticoagulants, platelet antiaggregating agents and vasodilators (Ribeiro, 1987). An antithrombin anticoagulant activity (Lester & Lloyd, 1926) and two platelet antiaggregating activities have been described in salivary gland extracts of *Glossina morsitans morsitans* (Mant & Parker, 1981). A thirty-two amino acid peptide, Tsetse Thrombin Inhibitor (TTI), has been purified as the major anticoagulant from *G. m. morsitans* saliva and its cDNA has been cloned and molecularly characterized (Cappello *et al.*, 1998). Purified TTI has also been shown to abolish thrombin-induced platelet aggregation (Cappello *et al.*, 1996). Recently, salivary extracts have been shown to have an adenosine deaminase activity and two cDNAs (*TSGF-1* and *TSGF-2*) whose products display growth factor and adenosine deaminase motifs have been molecularly described (Li & Aksoy, 2000). Because adenosine has been shown to have a platelet antiaggregating activity, its deamination to inosine might potentially interfere with host haemostasis at the bite site (Luthje, 1989).

In addition to aiding in successful feeding, saliva of blood-sucking insects have also been shown to contain pharmacologically active substances that indirectly enhance the transmission of pathogens (Titus & Ribeiro, 1990; Edwards *et al.*, 1998). Because replicating trypanosomes initially reside in the dermal microenvironment and undergo differentiation from non-dividing metacyclic forms to trypomastigotes, vector-induced inflammatory reactions may be significant for the development and dissemination of the parasites (Akol & Murray, 1982). While no immune modulatory molecules have yet been characterized from tsetse saliva, it has been shown that rabbits exposed to feeding tsetse flies develop cutaneous hypersensitivity responses with immediate and delayed-type characteristics (Ellis *et al.*, 1986). Tsetse saliva has been found to contain more than twenty proteins some of which could be recognized by sera from chronically exposed rabbits and from hypersensitive rabbits with only two exposures to fly saliva (Patel *et al.*, 1981; Ellis *et al.*, 1986).

Here, we have molecularly characterized three abundant cDNAs from a *G. m. morsitans* salivary gland cDNA library. The putative product of the *Tsetse Antigen 5* (*TAg5*) cDNA has the hallmarks of the CAP family (Crisp-Antigen 5, Plant pathogenesis protein) proteins. The function(s) of the closely related putative protein products encoded by the other two cDNAs (*Tsal1* and *Tsal2*) are at present unknown.

We describe the molecular characteristics of the cDNAs and their tissue and developmental stage specific expression in *G. m. morsitans*. We also describe the expression profiles of these cDNAs as well as the previously characterized genes encoding TTI, Tsetse Salivary Gland Factor (*TSGF-1* and *TSGF-2*) from three species of *Glossina* displaying different vector competence abilities (*G. m. morsitans*, *G. austeni* and *G. fuscipes*).

Results

Characterization of the full-length cDNA sequences

The sequences and the corresponding ORFs encoded by *TAg5*, *Tsal1* and *Tsal2* cDNAs are shown in Fig. 1(A–C), respectively. The 5' end of each transcript has been determined by 5' rapid amplification of cDNA ends (RACE). The potential signal peptide sequences associated with the N terminus of each putative protein have been determined based on hydrophilicity (Kyte–Doolittle) plots and SignalP program analysis.

The full-length *TAg5* cDNA is 926 bp excluding the poly(A) tail, and encodes for a 259 amino acid putative polypeptide (Fig. 1A). The sequence has a 48 bp 5' untranslated region (*UTR*) followed by a segment encoding for a twenty amino acid hydrophobic peptide that has the hallmarks of a cleaved leader peptide. The open reading frame (ORF) is followed by a 103 bp 3'-*UTR*, which has a polyadenylation signal, AATAAA, 28 bp upstream from the poly(A) stretch. The molecular mass of the predicted mature protein is 28 924 Da, and has an isoelectric point of 8.27.

The full-length *Tsal1* cDNA is 1329 bp excluding the poly(A) tail (Fig. 1B). It has a 20 bp 5'-*UTR*, and at the 5' end of its 399 amino acid ORF has a 22 amino acid hydrophobic leader sequence indicating that it may be secretory in nature. It has a 79 bp 3'-*UTR* with a AATAAA polyadenylation signal 50 bp upstream of the (A)₁₇ stretch. The *Tsal1* cDNA can encode for a mature protein of 45 612 Da with an isoelectric point of 4.96 and has one potential N-glycosylation and three N-myristolation sites.

The full-length *Tsal2* cDNA is 1255 bp excluding its poly(A) tail with a putative ORF of 389 amino acid. It has a 24 bp 5'-*UTR* followed by a nineteen amino acid hydrophobic signal peptide sequence. Following the ORF, it has a 65 bp 3'-*UTR* with a AATAAA polyadenylation signal present 21 bp upstream of the (A)₁₈ stretch. The ORF in *Tsal2* can encode for a mature protein of 43 900 Da with an isoelectric point of 7.18.

Putative protein products of *TAg5*, *Tsal1* and *Tsal2*

The putative mature protein products of *TAg5*, *Tsal1* and *Tsal2* cDNA sequences were used to search GenBank and EMBO databases using BLAST, and the sequences from other species showing high scores were retrieved and analysed in multiple alignment using CLUSTAL W analysis

Tag5

ATCAGTGAGAGTATTTAATCATCACACGATAAATAAAGGAAATGAATTCGTGCTTGCACCTTAAGCCTTTAAT 80
 M N F V L A T L S L L I
 TTTGGGTAGTGTGACGAGGGTGGAGATTTATCGCGCTTGCGCAGCATGATGATGCTGCGTGAACAAATGATTTC 160
 L G S A A V G G G C G C L C D N H D A C V M Q N V F
 AGAGTGGATGCCGTCGGCGCAAAATGATAGACTTGAACAAATATCAGAGCGCTTTGTTGACGCGCAATAAGAAA 240
 Q S G C P S G A K M I D L N K Y C S A L L D A H N K K
 CGTAACCATGTAGTGGCGGGGGAATCGAAATGAAGCGAGCTTGTGCGAGTGGCAGTATGAAATGGATTCAGAGTT 320
 R N H V A G G G S K L K R A C O M A T M K W D S E L
 GGCAAGCTGGCGAATCAATGTCAAAATGTCAATGAATCAGCAGGTCGGTAAATCGGTAAATTTAAATATG 400
 A K L A E Y N V K Q C O M N H D R C R N T V K F K Y
 CTGGTCAAGACTTACGCCAGCTTGGCGATCAGGAGGACCCCGCATACCGCAATATGAGAAAAGCTGTGAC 480
 A G Q N L L A E L G R S G G P P P D Y R K L I E K A V D
 AAATGGTATGAAGAGTTAAGACTGCATCAAGGCTATATGACTCTTCAATGAATTAAGAGACCGACCTGG 560
 K W Y E E V K D C N Q G Y I D S Y P H N Y R G P A I G
 CCATTTACAGTCTGGTGGCGGAAAGAACCGCATGTGGTGTGGCGCTTGTGAGTACACAAATCAATGTTTC 640
 H F T V M V A E R N T H V G C A A S E Y T K S N G F
 AATATTTCTGATGGCTTGCATATACGCCACCAATATGATGAGTTCCTCATCTATAATCGTGGTTCATCGGCT 720
 Q Y F L M A C N Y A T T N N M E F P I Y K S C G S S A
 CAAGATTGTGAAGCGGTGAAAATTCGAATATCCAAATCTATGCTCGCCCAACGAGAAATCAGAGGTCAATAAGAGT 800
 Q D C K S G K N S K Y P N L C S P N E K Y E V N K W I
 CAAAATGGTGTGCAATCATTAACCTTTGATCTGTGCAACAGCGGACTTCTGCTCAAGTGTATTATGTTGAAA 880
 K N G V E Y H
 TAAAAAATAAAAGATGAATAAAAAATTTAACTAGAGCAAAAAAATAAAAAA 943

Tsal1

CCATCAAAAGTGTTTTGAAGAACTTCCAATATGCAATGGATTGGTTACGGAGTTTTCACCTTGGCTTTAATAGGCA 80
 M A L K L V Y G V F T L A L I G
 TCAGTTCAGTTAAAGCTGATTTGCTTTAAATATCCAGAGAGTGTGAAATGAAAACCTCAGTCATCATGGTAAG 160
 I S V N A D C S L K I P E S V E N E K I P V I M V R
 AAGTCTCGAAGACTTTCCAAATGATTTATTCACCACTGGAGAGTAACAAATTTCTGAGCTACTGAGCTTCT 240
 K S L K T F E Y T T G E V T L N F P E R T E L L
 TTTGGCTGTGCTGGCAGCAAAATCTTTAAGAATGGAGAGAACTGTCTACTCTAATGTGTCATAATAAGATTCG 320
 L A C T G A O N Y F K N G E S V V T L M C H N E F
 ACAGTGGCAAGGTAAAGGCTGATCTTACTTGTGCTGAAGACCTTACAGCGAGTGTGAGGAAACGAAGAGGT 400
 D Q G N G L D L F T C T G C T P T A E L R K T K R
 TGCTCGTATGAGACTTATGCTATGAGTGTGCTTCTGCGATCGGAGAGGAACTGTGGCCGCTTTATGAGAT 480
 C S L G D L Y V Y D G V T F R I S E E E L V G P V Y D I
 TTGCTATAATGAGTATCGCAAGGCTTATGAGCTTAATAATAATGGTCAGCGCTTAATATTCGGCTTCCCG 560
 C Y N E Y S Q K P A Y S R N I N A A V N Y R V P
 AAAGTGAACAGATCTCAGTGTGCTATGCTAAGATTTTACTACAGAGAGATGTAAATATTTTAAATAGAAAC 640
 E S E T D L T V S D L F T C T T R D V N N Y F K L E N
 CAAAGCAAGCTTCAAGGCTATATAGCGGTAAGCAGCTTGTGGATGATAAAATCTTTCACCTGGCCAAAT 720
 Q K Q R F Q G Y T I D G K P L V D D K N F F A P G Q L
 AGCAGCTGATACCTCGATGATCCTCCGCTGACAGATTCACCTACGATATGCGAATATGTCACCAATATAAA 800
 A P D T S M I T P A D K L S T Y D Y A N I V P Q Y K
 CAGTTATGAGGCTAATGTATGGCGCTAGAGAGATTCCTGCTGATCTTCTGCTGTTAATGCTCAAGCTAAGTTGAAGT 880
 T V Y D G N V W R V E I T R D L A V N R Q A K F V
 TATACCTGGCGCTATGAACGCTAATCTTCAAGCGCAAGAGAGAGATTTCTGAGTCAAGAAATATGT 960
 Y T G G Y E R L T A L H N G K E E E I F L S A K K Y E
 ACATGAATACCGAAATATATCTACAATTTGTAGTGACAGGAGATGCTGGAATGTATTTTACTCTAACCA 1040
 H E I P K Y I Y K F V D K E N D A G I V F T L N
 ATCTCTATGCAAAATGTAAAGGCACTGAATTTTGCAGAGATCAATGCGAGGACCAATGTTCTGACTCAATTT 1120
 N P H Y K N V K G T E F C K N O C E E A N V L D S N F
 AAGATGACCAAGGATATATCTTCTGCTGCTGCAAGTGTAAAGGATGTAAAGGCTTACCTCAGGATAT 1200
 K D W T Q G Y T F C C T W N N V K D Y V R A L P D I
 TGAATTAATAATTTGTTAAATTTAATTAAGAGCAAAAGTAAAGAAAGAAATGAATAACAGATATATATATATA 1280
 E I N N L L K F N
 TATATGTATATACATGTAGTGAAGCATAAAAAATAAAAAA 1326

Tsal2

TTAGTGTGTTAAATCAACAAATGCTGCTTTTACGGCTTTTAACTTGTGCTTTTACCAGGCTGCTGTGTGT 80
 M S L L Y G L L I L A F T R S C L Y V
 CGGCTAATGTCATTAACTACCCGATGCTTAAATGAGGAGAGCTGCTGCTTCTAGTAAACCGGTAACATG 160
 G C C S I N I P D D L N G E E A P V V L V K T G N N
 V TAAATATTCGGCTGAGGAAAAAGGACCACTTCCCAAGAGGAGCTGAATCTCTTGGCTTGCACAGGAGAGGT 240
 V K L F R P E E K T T F P K G T E L L L A C T G E G
 AAGCGTTGAAATCTAATGGCAAGAGCAACATTTGCTGCAATGAATCAATTTGAGAGTGTGCAAGAGAGAA 320
 N G L K S N G O E T T T L S C N G N O F E S A A K E K
 ACTCAGGAGATGCTATGAAAGTATGGCGAAGGCTGGTGGAAGCAACCAAAAGCTTGTATGGTAAAGATTCA 400
 L K D M S C K S M A K A V V E Q T T K R C M G N D F
 ATCTTACGAAGCTGGCTATAAGATGAATGGAATTTCTATGGTCTGTTTATGATATTTGTAACATGCAAGGTCAA 480
 N L Y E A G Y K V N G K F V G E Q T T K R C M G N D F
 AGTAGCGATACAGCAACATTTATCTATGCTGCGACATGGAATTAAGCTTCCGAGAACTTATGAGCATTTTC 560
 S S G Y T H N F I Y G R T W K Y K L P E K P Y E H Y
 TTCAGAGATGCTCCAGCTGGAAGATCTTACCAATTTGTAAGAGCAAAAGGAAAGGTTCAAAATACTAAGTAAG 640
 S R D P O A G K D L D K L Y K E O K E R F K N T K V
 ATGTTAAACCTTTGCTAGACGATGAACATTTACTTACGAGGTCACTAACACGAGACCTTCCATATAACCGGCT 720
 N G K P L L D D E H Y F T E G O L T P D T S I I T G A
 GATAAGCTGTGACGATACGATACGATTAATCGCCCAATGTTTAAAGACATATGATGGCAATATCTGAGGATGA 800
 D K L S T Y D Y A N I A P L F K D I Y D G N I W R Y E
 AAACATGACCAAGAACTGGCTGACCAAGCTCAATTTGAGGAGATACAGCGGAGTTTACTCATACGAAGTGG 880
 N M T Q E L A D Q R O A T F E E Y T G G F Y S Y E
 AGAAGTGAACCTATTGGTTTGAATGATGCTGAGTTCTACTCATCGAGTACCGAAATACATATAAAGCTGGATG 960
 E K W K P I G L D N A E F T R V P K Y I Y K L V
 GATACGAAAGTAAAGTGGATGATTTGTTACTCTCAATGATCTATCATAGAGCCCGCTTCCGAAATTTATG 1040
 D T E S K D G I V F T V T L N D P Y H K S P A S E N L C
 TAAAGATATATGCAAGTGAAGCAATTAACGAGAGATTTCAAGATGTGCAAGAGGTACACAAATGCTGGCAGTT 1120
 K D I C S E A N N E K I S R M S K R V I T S A R S
 ATGGGGATTTCCGAAATCGTATACCACTTACCAAGATATATATGCTCAAGGCTTTTAAATATGAAGAGCCACAA 1200
 Y G D F G N Y K G T L L K Y
 AGATTATGAAGTTTAAATTCCTTAAGCATTAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT 1273

Figure 1. Complete nucleotide sequence of *Tag5*, *Tsal1* and *Tsal2* and their putative ORFs. The putative signal peptide sequences in the ORFs are italicized. The putative polyadenylation signal (AATAAA) prior to the poly(A) tail is underlined. The underlined glycine residues in bold in *Tsal1* sequence show the potential N-myristoylation sites. The underlined asparagine residue at nucleotide position 831 shows the potential N-glycosylation site.

(Altschul *et al.*, 1990). Extensive similarities were found for *Tag5* with members of the CAP family, in particular with *Agr* and *Agr2* described from *Drosophila* and *LuloAG5* from *Lutzomyia*, respectively. An alignment depicting similarities between the putative *Tag5*, and the three arthropod *Ag5*-related products is shown in Fig. 2(A). Sequence pair distance analysis indicates that *Tag5* shows 40% identity in its amino acid sequence to *Agr*, 42% identity to *Agr2*, while it exhibits only a 28% overall identity to *LuLoAG5* (Fig. 2B). The putative coding segments of *Tsal1* and *Tsal2* cDNAs were found to display over 42% identity to one another indicating that they represent a family of related sequences but neither sequence exhibited extensive similarities to other known gene products.

Expression and developmental regulation of salivary gland cDNAs

Northern analysis from *G. m. morsitans* salivary glands has shown that *Tag5* hybridizes to an abundant transcript of approximately 900 bp, while *Tsal1* and *Tsal2* hybridize to transcripts of approximately 1.3 kb in size (summarized in Table 1). The developmental stage-specific expression of these transcripts as well as the previously characterized *TTI*, *TSGF-1* and *TSGF-2* cDNAs was evaluated by a semiquantitative reverse transcription polymerase chain reaction (RT-PCR) analysis in *G. m. morsitans* (Table 1). Based on this analysis, the expression levels for all six genes were found to be comparable between male and female glands. PCR-amplification products corresponding to *Tag5* and *Tsal2* could be detected from adult, larval and pupal developmental stages. *Tsal1* expression, however, could not be detected from the pupal stage but is present in the adult and larva. The previously characterized cDNAs (*TTI*, *TSGF-1*, *TSGF-2*) were found to be expressed only in adult glands. With the exception of *TTI*, none of the transcripts was significantly induced upon blood feeding. Although this analysis is semiquantitative in nature, the amplification products specific for *Tag5* and *TTI* were significantly more abundant than those detected for the other four genes indicating that *Tag5* and *TTI*-specific transcripts represent abundant products in salivary glands.

Expression of salivary gland cDNAs in different tsetse species

The expression of the three cDNAs as well as the previously characterized *TTI*, *TSGF-1* and *TSGF-2* were studied by Northern analysis from female and male glands of *G. m. morsitans*, *G. austeni* and *G. fuscipes* (Fig. 3A). The ribosomal RNA (rRNA) was used to adjust for variations in input RNA levels in different salivary gland preparations. The autoradiograms were scanned, corrected for input RNA levels and the results are schematically presented in Fig. 3(B). Male glands in all species had slightly higher levels of messenger RNA (mRNA) for *Tag5* and *Tsal2* while the other

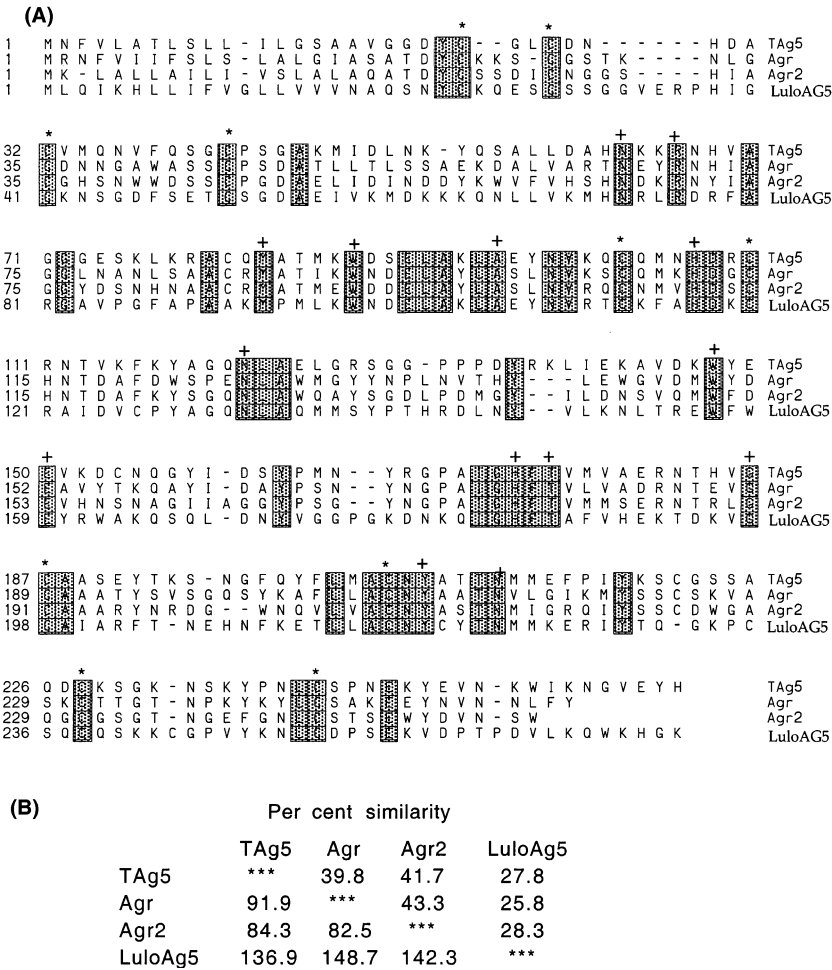


Figure 2. (A) Alignment of putative TAg5 protein with other known CAP family proteins. The dashes in the sequence indicate gaps inserted to allow for optimal alignment. The boxes denote residues found identical in the four sequences. The ten cysteine residues found conserved in the arthropod Ag5 products are denoted by *, and the + indicates the residues found conserved in CAP family sequences including plant pathogenesis related proteins and mammalian CRISPs. (B) Pair distance analysis of the above alignment was performed using the distance analysis function of DNASTar software. Below * absolute distances; above *, mean distances.

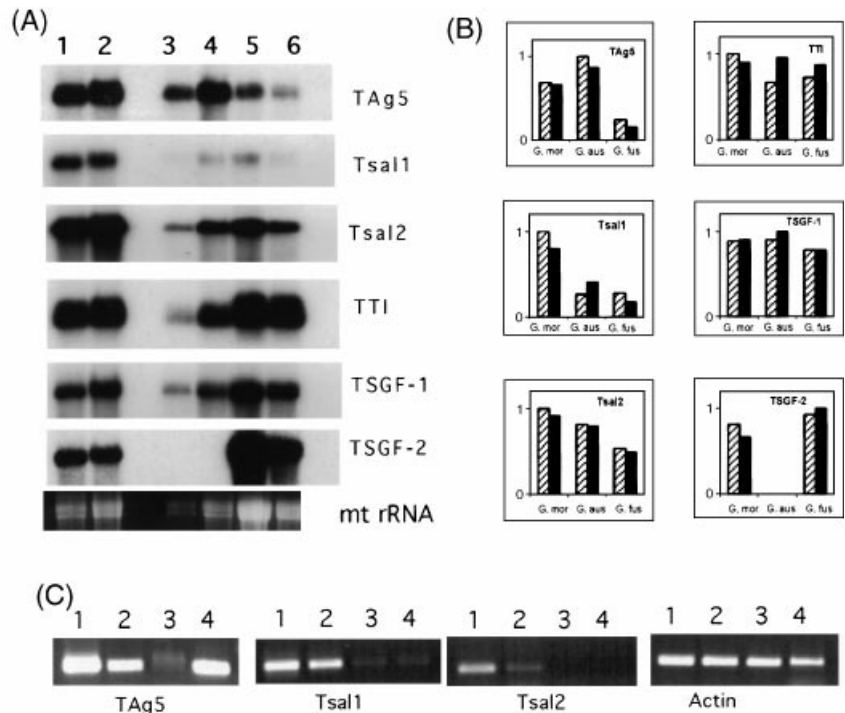
Table 1. Full-length cDNA clones expressed in salivary glands of *Glossina morsitans* and characteristics of their putative products. The developmental regulation of different genes was determined by a semiquantitative RT-PCR assay where the quantitative presence of the generated amplification product is denoted as + to +++, depending on the intensity of the scanned amplification product visualized by ethidium bromide staining after gel electrophoresis

Clone	cDNA size (bp)	Mature protein (Da)	Predict pI	Developmental regulation						Sequence similarity/Identity
				Female	Male	Larva	Pupae	Fed	Unfed	
TTI	303	3530	3.71	++++	++++	-	-	++++	++	Tsetse Thrombin Inhibitor (TTI)
TAg5	943	28 924	8.27	+++	++++	++	++	+++	+++	<i>Drosophila</i> (Ag5, Ag5-2) <i>Lutzomyia</i> (LuLoAG5)
TSGF-1	1673	56 637	5.5	+	+	-	-	+	+	<i>Sarcophaga</i> (IDGF) <i>Apylasi</i> (AGSA, MDSF)
TSGF-2	1594	58 300	8.3	++	++	-	-	+	++	<i>Sarcophaga</i> (IDGF) <i>Apylasi</i> (AGSA, MDSF)
Tsal1	1326	45 612	4.9	++	++	++	-	++	++	No significant matches
Tsal2	1273	43 930	7.2	++	++	++	+	++	++	No significant matches

transcripts did not exhibit a sex-based regulation of expression as was determined by RT-PCR analysis described earlier. The level of *TAg5* expression was found to be highest in *G. austeni* while only about 60% of this level could

be detected in *G. morsitans* and 10–20% in *G. fuscipes*, respectively (Fig. 3B). The expression of *TAg5* was also analysed from *G. palpalis palpalis* which is a closely related species to *G. fuscipes* in the same subgenera of *Glossina*

Figure 3. Expression of salivary gland cDNAs in different tsetse species and tissues. (A) Northern blot analysis of *G. m. morsitans* (Lanes 1 and 2), *G. austeni* (Lanes 3 and 4) and *G. fuscipes* (Lanes 5 and 6) salivary gland RNA analysed from male (Lanes 1, 3 and 5) and female flies (Lanes 2, 4 and 6). The same blot was stripped and hybridized to different labelled gene probes as indicated. As an internal marker, rRNA was visualized on the agarose gel by staining with ethidium bromide. (B) The autoradiograms were scanned, corrected for input RNA levels and data is schematically shown to indicate the expression of each gene product in male and female glands of different tsetse species. (C) Tissue-specific expression of *TAg5*, *Tsal1* and *Tsal2* gene products by RT-PCR analysis. Amplification products were generated using cDNAs from salivary gland (Lane 1), proventriculus (Lane 2), fat body (Lane 3) and midgut (Lane 4) tissues. Actin specific primers were used as control. The gene-specific primers used in each analysis are indicated below each figure (actin, *TAg5*, *Tsal1* and *Tsal2*).



(*Nemorhina = palpalis* group), and again significantly less *TAg5* expression was detected in this species (data not shown). The prevalence of *Tsal1* and *Tsal2* mRNAs is found to be the highest in *G. m. morsitans*. The expression of *Tsal1* in *G. austeni* and *G. fuscipes* is lower in comparison to *G. m. morsitans*, being only about 25% and 15%, respectively. In *G. austeni* *Tsal2* expression is at about 80% of the level detected in *G. m. morsitans* while in *G. fuscipes*, it is only about 50%. In contrast, the mRNA levels corresponding to *TTI* and *TSGF-1* do not vary extensively in the different species analysed. Both mRNA and the genomic locus for *TSGF-2* are apparently missing in *G. austeni* while it is expressed at comparable levels in *G. fuscipes* and in *G. morsitans*.

Tissue-specific expression of *TAg5*, *Tsal1* and *Tsal2*

The expression of *TAg5*, *Tsal1* and *Tsal2* mRNAs was analysed from salivary gland, proventriculus, midgut and fat-body tissues by Northern analysis. Transcripts corresponding to all three genes could only be detected in salivary glands (data not shown). A more sensitive RT-PCR amplification approach was used with cDNAs generated from mRNAs prepared from these same tissues (Fig. 3c). Amplification products specific for *TAg5* could be detected from salivary gland, proventriculus and midgut tissue, but not reproducibly from fat body. Because in *Drosophila* two related gene products (*Agr* and *Agr2*) were found to be expressed in different tissues, the PCR-amplified *TAg5*-specific fragments from proventriculus and midgut tissues

were cloned and sequenced, but were found to be identical to the sequence characterized from the salivary glands (data not shown). Products corresponding to *Tsal1* and *Tsal2* could be reproducibly amplified only from salivary gland and proventriculus tissue. It remains to be seen whether the transcripts detected in different somatic tissues represent biologically functional gene products. Alternatively, these could represent basal level transcripts detected due to the high sensitivity of PCR amplification-based techniques.

Discussion

The molecular features of three cDNAs expressed in salivary glands of *G. m. morsitans* are described. Based on its extensive amino acid similarity, one of the cDNAs (*TAg5*) encodes for a member of the CAP family proteins. The functional significance of the other two putative polypeptides encoded by the cDNAs *Tsal1* and *Tsal2*, is unclear.

The species *G. morsitans*, *G. austeni* and *G. fuscipes* belong to separate subgenera of *Glossina* and exhibit very different vector competence abilities for the transmission of the pathogenic trypanosomes. While the genetic basis of vector competence in tsetse is largely unknown, *G. morsitans* subgroup flies have been shown to be highly susceptible, whereas *G. austeni* and *G. fuscipes* flies are increasingly refractory for the transmission of the *T. brucei* complex parasites (Moloo & Kutuza, 1988a, 1988b). *TTI* and the two related putative proteins *TSGF-1* and *TSGF-2* which encode

for the thrombin-inhibiting anticoagulant and putative adenosine deaminating activities, respectively, were found to be expressed in adult glands only. This might indicate that they might be functioning primarily in the blood-feeding process and play a significant role in antihemostasis. The high levels of transcripts corresponding to the anticoagulant TTI detected from all tsetse species may also reflect its indispensable role in salivary gland functions. The protein (and gene) for TSGF-2 is apparently missing in *G. austeni*, suggesting that at least TSGF-2 function(s) in adult tsetse glands might be dispensable (Li & Aksoy, 2000). High levels of transcripts are however, detected for the related cDNA TSGF-1 in all species. Again, both TSGF-1 and TSGF-2 protein products are only detected in adult glands implying a possible role for these products in the blood-feeding process. The *TAg5*, *Tsal1* and *Tsal2* transcripts, however, can be detected from different developmental stages, suggesting more extensive physiological roles than blood feeding for their putative proteins. While based on Northern analysis, salivary glands are found to be the major tissue where all genes are expressed, the more sensitive RT-PCR based amplification analysis has detected the presence of these mRNAs in the proventriculus tissue and midgut. It is interesting that the expression levels of *Tsal1* and *Tsal2* display extensive differences between the susceptible *G. morsitans* and the refractory species, *G. austeni* and *G. palpalis*. The expression of *Tsal1* is significantly less in the refractory species in comparison to the susceptible *G. m. morsitans*.

Antigen 5 is a member of the larger CAP family proteins which have been identified from as diverse organisms as mammals cysteine-rich secretory proteins (CRISP) (Brooks *et al.*, 1986), nematodes Ag5–Ag3 proteins (Hawdon *et al.*, 1996), vespids Antigen 5 (A) and plants pathogenesis-related proteins (P) (Niderman *et al.*, 1995). These secreted proteins share a core sequence of about 200 amino acids that has been adapted to play in diverse processes in various animals although their precise function remains largely unknown. In vespids (hornets, yellow jackets and wasps), Antigen 5 has been characterized as one of the three major venom allergens causing allergies in humans (Fang *et al.*, 1988), and again in fire ant venom it has been shown to be the major allergen (Hoffman, 1995). In nematodes, they are expressed early in the transition of the free-living larva to an infectious parasitic form (Hawdon *et al.*, 1996), and in plants they have been shown to be induced in response to infection by viral, bacterial and fungal pathogens and are supposed to participate in local acute-phase defence responses. Two related proteins have been characterized from *Drosophila*, the first gene (Antigen 5 related, *Agr*) is primarily expressed in proventriculus of larva and adult and as such may be a constituent of the peritrophic matrix (Kovalick *et al.*, 1998). The second gene (Antigen 5 related 2, *Agr2*) is expressed in posterior midgut cells in adults, and may be involved in mediating immune reactions in the gut (Megraw *et al.*, 1998).

Recently, an Antigen 5-related sequence (*LuloAG5*) has also been isolated from a salivary gland cDNA library from the sandfly vector *Lutzomyia longipalpis* (Charlab *et al.*, 1999). Using low stringency hybridization conditions, similar genes were also detected from *Sarcophaga* indicating their overall conservation in Diptera (Kovalick *et al.*, 1998). In tsetse, *TAg5* is primarily expressed in salivary glands but could also be detected in midgut and proventriculus tissues by RT-PCR analysis. It remains to be seen if *TAg5* protein can be detected in tsetse guts or whether the wide tissue tropism observed by RT-PCR analysis reflects a baseline transcription level detected due to the high sensitivity of this assay.

Tsetse saliva (*G. m. centralis*) has been studied for its immune functions and rabbits have been found to exhibit type I hypersensitivity reactions upon secondary or tertiary exposure to tsetse bite (Ellis *et al.*, 1986). However, in one study, three out of seven rabbits chronically exposed to tsetse saliva failed to elicit an immune response (Ellis *et al.*, 1986). When sera and IgG from hypersensitive rabbits exposed to two tsetse bites were used to analyse tsetse salivary proteins, three major proteins with 160, 42 and 28 kDa were detected by immunoblotting. The chronically exposed animals also reacted well with the 42 and 28 kDa antigens. As whole sera was found to detect more polypeptides with greater intensity than purified IgG, it is possible that IgE antibody might be recognizing saliva antigens and contributing to hypersensitivity reactions in this study, similar to what has been found in venom antigens (Hoffman, 1981). It remains to be seen if the 28 kDa antigen in tsetse salivary extracts detected by hypersensitive serum corresponds to the putative *TAg5* protein encoded by the *TAg5* cDNA. In addition, a number of the secreted proteins in saliva were found to be glycosylated including a prominent 42 kDa protein (Ellis *et al.*, 1986). The presence of predicted glycosylation sites and the expected similar size of *Tsal2* and *TSGF-2* cDNA products make these proteins candidates for this activity.

The extensive antigenic variation phenomenon displayed by the African trypanosomes in their mammalian host has hampered efforts to develop effective vaccines. Disease-management strategies currently rely on the treatment of infected hosts by chemotherapy and on reduction of tsetse challenge by eradication or suppression approaches. Recent developments in molecular technologies and their application to vector genetics have now opened up new avenues for the development of alternative vector-based strategies. The success of these approaches, however, largely depends on the characterization of vital insect targets that can be manipulated to block transmission of pathogenic agents. As salivary gland gene products and saliva may play a major role in pathogen transmission, their characterization and comparative analysis from species with susceptible and refractory phenotypes can shed light on trypanosome maturation and transmission and may lead to the development of novel control strategies.

Experimental procedures

Insects

The *G. m. morsitans* puparia obtained from the Tsetse Research Laboratory (Bristol University, Bristol, UK) were originally established from puparia collected in Zimbabwe. The colony is maintained in the insectary at Yale University's Laboratory of Epidemiology and Public Health at 24–26 °C with 65% humidity and the flies are fed daily on defibrinated bovine blood (Crane Laboratories, Syracuse, NY) using an artificial membrane system (Moloo, 1971). Puparia of colonized *G. austeni*, Newstead, *G. fuscipes* and *G. palpalis palpalis* (Robinean-Desvoidy) have been kindly supplied by the Seibersdorf Agricultural Research laboratory, Vienna, Austria.

Construction and screening of salivary gland cDNA library from salivary glands

The construction of the cDNA library in Uni-ZAP XR Vector (Stratagene Cat. number 200400, La Jolla, CA) has been described (Li & Aksoy, 2000). Briefly, mRNA was purified from 500 pairs of dissected salivary glands from 1- to 2-week-old *G. m. morsitans* flies, and was used to construct the library according to the manufacturer's instructions. About 15 000 plaques were differentially screened using two labelled cDNAs populations prepared from salivary gland and muscle tissue specific poly(A)⁺ RNAs, respectively. The plaques that hybridized preferentially to the salivary gland probe were rescreened and twenty-six were selected and the cDNA containing phagemids were rescued from the λZAP phages using the *in vivo* excision protocol from the UNI-ZAP vector manual. The cDNAs were characterized by DNA sequencing on both strands using an automated sequencer (Keck Center, Yale University). Three clones, pTT1, pTSGF-1 and pTSGF-2, have been previously described; here, we describe the clones, pTag5, pTsal1, pTsal2.

The 5' ends of all three genes were confirmed using 5'-RACE system (v. 2.0) from Gibco-BRL with gene-specific primers. Briefly, a specific minus strand primer was annealed to poly(A)⁺ salivary gland RNA (50–100 ng), followed by reverse transcription of cDNA using SuperScript II transcriptase (Gibco-BRL) at 42 °C. The RNA was degraded with ribonuclease (Rnase) I, and the cDNA was purified using a MAX Spin cartridge (Gibco/BRL). The cDNA was modified by the addition of a dCTP tail using terminal transferase, and this product was PCR amplified using a second gene-specific internal primer (Tag5: 5'-GCTTGGCCAACTCTGAATCC, Tsal1: 5'-ACGGCTATAGGCAGGCTTTTGC; Tsal2: 5'-GCCAGCTTCGTAAAGATTG) and an Abridged Anchor Primer, respectively. The generated PCR product(s) were reamplified using a third specific internal oligonucleotide as the 3'-end primer (Tag5: 5'-GAATCCATTTTCATAGTGGC; Tsal1: 5'-TTCCTCTCCGATATGCG; Tsal2: 5'-CCACCACAGCCTTCGCCATAC) and the Abridged Universal Amplification Primer (AUAP-Gibco/BRL), respectively. The resulting amplification product(s) from two independent PCR-reactions were cloned into the pGEM-T vector (Promega) and the DNA sequence of two clones from each ligation product were obtained by sequencing. The full length cDNA sequences have been submitted to GenBank: Accession numbers are *Tag5* (AF259957), *Tsal1* (AF259958) and *Tsal2* (AF259959).

Analysis of DNA sequence data

Computer analysis of the sequences was performed using gapped BLAST analysis in National Center for Biotechnology Information

(Bethesda, MD), v. 1.4.6 MP. The nucleotide sequences were aligned using the MegAlign program of the DNASTar software. The putative protein sequences were analysed using the Protean option of the DNASTar software, SignalP v. 1.1 (Center for Biological Sequence Analysis, Technical University of Denmark, <http://genome.cbs.dtu.dk/htbin/nph>), and PSORT II analysis (Prediction of Protein Sorting Signals and Localization Sites in Amino Acid Sequences, <http://sprot.nibb.ac.jp:8800>).

Developmental and tissue-specific expression

For Northern analysis, total RNA was prepared from thirty male and female adult salivary glands of *G. m. morsitans*, *G. austeni* and *G. fuscipes* flies, and was analysed on 1.5% agarose gel containing 2.2 M formaldehyde. rRNA level from a scanned image of the ethidium bromide stained gel was used to quantify the input RNA levels. The RNAs were transferred to Nytran and hybridized in a buffer containing 7% sodium dodecyl sulphate (SDS)/0.5 M sodium phosphate, pH 7.2/1% bovine serum albumin (BSA)/1 mM EDTA at 60 °C. For hybridization, radiolabelled antisense RNA probes were used as previously described (Li & Aksoy, 2000). The filter was washed at 60 °C for 20 min in low stringency buffer (5% SDS + 40 mM phosphate buffer, pH 7.2) and subsequently in high stringency buffer (1% SDS + 40 mM phosphate buffer, pH 7.2) (Morrison & Walsh, 1988). The autoradiograms were scanned and the intensity of hybridization level was determined using the software NIH Image 1.61 (<http://rsb.info.nih.gov/ni-image>).

The developmental regulation of transcripts was determined by a semiquantitative RT-PCR analysis. mRNA was extracted from male and female salivary glands, second- to third-instar larvae and 2-week-old puparia and from teneral glands and glands dissected 24 h after blood feeding, respectively. First strand cDNAs were synthesized using Superscript II (GibcoBRL, Cat. number 18064–014) according to instructions. The cDNA products were diluted (1 : 50, 1 : 200, 1 : 500, 1 : 1000) and subjected to PCR amplification with tsetse actin gene-specific primer set: (actin F: 5'-GAGAAGATGACCCAGATCATG; actin R: 5'-TCGTGCGCATGGTGATAACCTG) in order to evaluate the quality and quantity of the cDNAs generated. The products were separated by electrophoresis on 1.0% agarose gels and visualized by EtBr staining. If quantitative differences were not observed in the amplification products, the cDNA templates were further diluted and subjected to amplification. The cDNA dilutions from different tissues that generated comparable amounts of the actin amplification product were then used in subsequent PCR amplifications with gene-specific primers, respectively. In each experiment, the actin-specific primers were also included as control. The PCR amplification conditions were 94 °C for 2 min; thirty cycles of 30 s at 94 °C, 1 min at 55 °C and 1 min at 72 °C in an MJ Research PTC-200 thermocycler. The amplification products corresponding to different genes were analysed by electrophoresis analysis on 1% agarose gels, and visualized by EtBr staining. A densitometer was used to evaluate the relative abundance of the products. The reactions were scored as +, ++, +++ and ++++ to indicate increasing abundance of the gene-specific product amplified. In case of negative amplification, the PCR experiments were repeated with higher DNA concentrations to ensure that the dilutions used were not below the sensitivity of the PCR technique.

For tissue-specific expression analysis, cDNAs were prepared from mRNA isolated from salivary gland, proventriculus, midgut and fat-body tissues. RT-PCR analysis was conducted with gene-specific primer sets (Tag5: 5'-Tag5–5'-ATGGCCACTATGAAATGG and

3'-TAG5-5'-CAGAAGCGGCAACCC; Tsal1: 5'-Tsal1-5'-CACTC-CCGCTGACAACTTA and 3'-Tsal1-5'-TATCCTGAGGTAAGGCTC; Tsal2: 5'-Tsal2-5'-CTTTGCTAGACGATGAACA and 3'-Tsal2-5'-TAAC TGCCAGCAGAT TG). As control tsetse actin gene specific primers were used (5'-Actin: 5'-GAGAAGATGACCCAGATCATG and 3'-actin: 5'-TCGT TGCCGATGGTGAT TACCTG). The PCR amplification conditions were 94 °C for 2 min; thirty cycles of 30 s at 94 °C, 1 min at 55 °C and 1 min at 72 °C in an MJ Research PTC-200 thermocycler. Amplification with actin-specific primers was conducted for twenty-five cycles only.

Acknowledgements

NIH/NIAID-AI-34033 and NSF-9511806 to S.A. supported this work, and S.L. has been funded through the Li Foundation. The authors thank the Entomology Unit of the Joint FAO/IAEA Division in Seibersdorf, Vienna, Austria for providing tsetse puparia.

References

- Akol, G. and Murray, M. (1982) Early events following challenge of cattle with tsetse infected with *Trypanosoma congolense*: development of the local skin reaction. *Vet Rec* **110**: 295.
- Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990) Basic local alignment search tool. *J Mol Biology* **215**: 403–410.
- Brooks, D., Means, A., Wright, E., Singh, S. and Tiver, K. (1986) Molecular cloning of a cDNA for androgen-dependent sperm-coating glycoproteins secreted by the rat epidymus. *Eur J Biochem* **161**: 13–18.
- Cappello, M., Bergum, P.W., Vlasuk, G.P., Furmidge, B.A., Pritchard, D.I. and Aksoy, S. (1996) Isolation and characterization of the tsetse thrombin inhibitor: a potent antithrombotic peptide from the saliva of *Glossina morsitans morsitans*. *Am J Trop Med Hyg* **54** (5): 475–480. [Published erratum appears in *Am J Trop Med Hyg* **55** (1): 118.]
- Cappello, M., Li, S., Chen, X., Li, C., Harrison, L., Narashimhan, S., Beard, C. and Aksoy, S. (1998) Tsetse thrombin inhibitor: bloodmeal induced expression of a novel anticoagulant in salivary glands and gut tissue of *Glossina morsitans morsitans*. *Proc Natl Acad Sci USA* **95**: 14290–14295.
- Charlab, R., Valenzuela, J., Rowton, E. and Ribeiro, J. (1999) Toward and understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis*. *Proc Natl Acad Sci USA* **96** (26): 15155–15160.
- Cunningham, I. and Honigberg, B. (1977) Infectivity reacquisition by *Trypanosoma brucei brucei* cultivated with tsetse salivary glands. *Science* **197**: 1279–1282.
- Edwards, J., Higgs, S. and Beaty, B. (1998) Mosquito feeding-induced enhancement of Cache Valley Virus (Bunyaviridae) infection in mice. *J Med Entomol* **35** (3): 261–265.
- Ekwanzala, M., Pepin, J., Khonde, N., Moliho, S., Bruneel, H. and De Wals, P. (1996) In the heart of darkness: sleeping sickness in Zaire. *Lancet* **348**: 1427–1430.
- Ellis, J., Shapiro, S., Ole Moi-Yoi, O. and Moloo, S. (1986) Lesions and saliva-specific antibody responses in rabbits with immune and delayed hypersensitivity reactions to the bites of *Glossina morsitans centralis*. *Vet Pathol* **23**: 661–667.
- Fang, K., Vitale, M., Fehlner, P. and King, T. (1988) cDNA cloning and primary structure of a white-face hornet venom allergen, antigen-5. *Proc Natl Acad Sci USA* **85** (3): 895–899.
- Hawdon, J., Jones, B., Hoffman, D. and Hotez, P. (1996) Cloning and characterization of *Ancylostoma* secreted protein: a novel protein associated with the transition to parasitism by infective hookworm larvae. *J Biol Chem* **271**: 6672–6678.
- Hoffman, D. (1981) Allergens in Hymenoptera venom VI: cross reactivity of human IgE antibodies to the three vespid venoms and between vespid and paper wasp venoms. *Annu Allergy* **46**: 304–309.
- Hoffman, D. (1995) Fire ant venom allergy. *Allergy* **50** (7): 535–544.
- Kovalick, G., Schreiber, M., Dickason, A. and Cunningham, R. (1998) Structure and expression of the *Antigen 5-related* gene of *Drosophila melanogaster*. *Insect Biochem Mol Biol* **28**: 491–500.
- Lester, H. and Lloyd, L. (1926) Notes on the process of digestion in tsetse flies. *Bull Entomol Res* **19**: 39–60.
- Li, S. and Aksoy, S. (2000) A family of genes with growth factor and adenosine-deaminase similarity are preferentially expressed in the salivary glands of *Glossina m. morsitans*. *Gene* **252**: 83–93.
- Luthje, J. (1989) Extracellular adenine compounds, red blood cells and haemostasis: facts and hypotheses. *Blut* **59** (4): 367–374.
- Mant, M.J. and Parker, K.R. (1981) Two platelet aggregation inhibitors in tsetse (*Glossina*) saliva with studies of roles of thrombin and citrate in *in vitro* platelet aggregation. *Br J Haematol* **48** (4): 601–608.
- Megraw, T., Kaufman, T. and Kovalick, G. (1998) Sequence and expression of *Drosophila Antigen 5-related 2*, a new member of the CAP gene family. *Gene* **222**: 297–304.
- Moloo, S.K. (1971) An artificial feeding technique for *Glossina*. *Parasitology* **63** (3): 507–512.
- Moloo, S.K. and Kutzua, S.B. (1988a) Comparative study on the infection rates of different laboratory strains of *Glossina* species by *Trypanosoma congolense*. *Med Vet Entomol* **2** (3): 253–257.
- Moloo, S.K. and Kutzua, S.B. (1988b) Comparative study on the susceptibility of different *Glossina* species to *Trypanosoma brucei brucei* infection. *Trop Med Parasitol* **39** (3): 211–213.
- Morrison, J. and Walsh, C. (1988) Advances in enzymology and related areas. *Mol Biol* **61**: 201–300.
- Niderman, T., Genetet, I., Bruyere, T., Gees, R., Stintzi, A., Legrand, M., Fritig, B. and Mosinger, E. (1995) Pathogenesis-related PR-1 proteins are antifungal. *Plant Physiol* **108**: 17–27.
- Patel, N., Youdeowli, A. and Odhiambo, T. (1981) The composition of the salivary gland secretion of the tsetse, *Glossina morsitans morsitans*. Westwood 1850. *Insect Sci Applications* **1**: 383.
- Ribeiro, J.M.C. (1987) Role of saliva in blood-feeding by arthropods. *Ann Rev Entomol* **32**: 463–478.
- Titus, R. and Ribeiro, J. (1990) The role of the vector saliva in transmission of arthropod-borne disease. *Parasitol Today* **6**: 157–160.
- Van den Abbeele, J., Rolin, S., Claes, Y., Le Ray, D., Pays, E. and Coosemans, M. (1995) *Trypanosoma brucei*: stimulation of adenylate cyclase by proventriculus and esophagus tissue of the tsetse fly, *Glossina morsitans morsitans*. *Exp Parasitol* **81** (4): 618–620.
- Van den Abbeele, J., Claes, Y., Bockstaele, D.V., le Ray, D. and Coosemans, M. (1999) *Trypanosoma brucei* spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology* **118**: 469–478.
- Vickerman, K., Tetley, L., Hendry, A. and Turner, M. (1988) Biology of African trypanosomes in tsetse fly. *Biol Cell* **64**: 109–119.