

Identification of the most abundant secreted proteins from the salivary glands of the sand fly *Lutzomyia longipalpis*, vector of *Leishmania chagasi*

Jesus G. Valenzuela^{1,*}, Mark Garfield², Edgar D. Rowton³ and Van M. Pham¹

¹Vector Molecular Biology Unit, Laboratory of Malaria and Vector Research, NIAID, National Institutes of Health, 12735 Twinbrook Parkway, Room 2E-22C, Rockville, MD 20852, USA, ²Structural Biology Unit, NIAID, Rockville, MD 20852, USA and ³Department of Entomology, Walter Reed Army Institute of Research, Washington, DC 20307, USA

*Author for correspondence (e-mail: jvalenzuela@niaid.nih.gov)

Accepted 12 July 2004

Summary

Using massive cDNA sequencing, proteomics and customized computational biology approaches, we have isolated and identified the most abundant secreted proteins from the salivary glands of the sand fly *Lutzomyia longipalpis*. Out of 550 randomly isolated clones from a full-length salivary gland cDNA library, we found 143 clusters or families of related proteins. Out of these 143 families, 35 were predicted to be secreted proteins. We confirmed, by Edman degradation of *Lu. longipalpis* salivary proteins, the presence of 17 proteins from this group. Full-length sequence for 35 cDNA messages for secretory proteins is reported, including an RGD-containing peptide, three members of the yellow-related family of proteins, maxadilan, a PpSP15-related protein, six members of a family of putative anticoagulants, an antigen 5-related protein, a D7-related protein, a cDNA belonging to the Cimex apyrase family of proteins, a

protein homologous to a silk protein with amino acid repeats resembling extracellular matrix proteins, a 5'-nucleotidase, a peptidase, a palmitoyl-hydrolase, an endonuclease, nine novel peptides and four different groups of proteins with no homologies to any protein deposited in accessible databases. Sixteen of these proteins appear to be unique to sand flies. With this approach, we have tripled the number of isolated secretory proteins from this sand fly. Because of the relationship between the vertebrate host immune response to salivary proteins and protection to parasite infection, these proteins are promising markers for vector exposure and attractive targets for vaccine development to control *Leishmania chagasi* infection.

Key words: salivary gland transcript, salivary protein, New World sand fly, *Lutzomyia longipalpis*, saliva.

Introduction

The New World sand fly, *Lutzomyia longipalpis*, is an important vector of *Leishmania chagasi*, causal agent of the visceral form of leishmaniasis in humans. These sand flies, as well as many other blood feeders, contain potent pharmacological components in their saliva that assist them in obtaining their blood meals and evading host inflammatory and immune responses. Components such as anticoagulants, inhibitors of platelet aggregation and vasodilators have been described in blood feeders (Ribeiro, 1987, 1995). In the saliva of *Lu. longipalpis*, apyrase, anticoagulant, vasodilatory and immunomodulatory activities have been reported; some of these molecules were characterized at the molecular level (Lerner et al., 1991; Charlab et al., 1999; Gillespie et al., 2000).

Arthropod saliva modifies the physiology of the host at the site of the bite, making it more permissive for pathogen invasion (Titus and Ribeiro, 1988; Ribeiro, 1989; Kamhawi, 2000; Nuttall et al., 2000). In some vector/parasite systems, immune response to arthropod feeding or arthropod bites (probably to salivary proteins) precludes the establishment of

the pathogen in the vertebrate host (Bell et al., 1979; Jones and Nuttall, 1990; Wikel et al., 1997; Nazario et al., 1998; Kamhawi et al., 2000). There are two hypotheses (not mutually exclusive) to explain this protection. First, vertebrate immune response to insect salivary proteins, particularly antibodies, may neutralize the effect of the salivary component(s) responsible for pathogen establishment. There is evidence supporting this hypothesis; Morris et al. (2001) demonstrated that animals vaccinated with maxadilan, a potent vasodilator and immunomodulator from *Lu. longipalpis*, raised antibodies against this salivary protein, which resulted in animals protected against *Le. major* infection. Second, vertebrate host immune response to salivary proteins creates an inhospitable environment to the pathogen, which is either killed or its future development negatively affected at the site of the bite. In this case, the anti-salivary immune response may be mainly a cellular rather than humoral response. Evidence to support this hypothesis comes from the work of Kamhawi et al. (2000), where animals pre-exposed to *Phlebotomus papatasi* sand fly

bites generated a strong delayed-type hypersensitivity response at the site of the bite; this response was related to protection against *Le. major* infection. Mice vaccinated with a 15 kDa salivary protein (PpSP15) produced a strong delayed-type hypersensitive response (DTH) and antibodies in black/6 mice, resulting in protection against *Le. major* infection when parasites were co-inoculated with salivary gland homogenate (SGH). Further evidence suggested that DTH was responsible for the observed anti-*Leishmania* protection and that antibodies were not necessary (Valenzuela et al., 2001a).

In humans, we have reported a correlation between immune response to *Lu. longipalpis* salivary proteins and cellular response to *Leishmania chagasi*, a protective mechanism against leishmaniasis (Barral et al., 2000). Additionally, we have reported a number of salivary antigens recognized by sera of humans, which may be related to protection against leishmaniasis (Gomes et al., 2002); however, the identity of these proteins remains to be solved.

We have hypothesized that the protective effect of sand fly salivary proteins is related to a cellular response to these molecules, probably from CD4 T cells (Valenzuela et al., 2001b). Accordingly, these responses depend on the immunogenetic background of the vertebrate host; thus, targeting a single protein may not be an adequate approach to identify the right vaccine for a population. A broader approach or selection of multiple candidates may be required. High-throughput approaches based on massive cDNA sequencing, proteomics and customized computational biology approaches are helping us to reveal the proteins present in the salivary glands of different blood-feeding arthropods, including the mosquitoes *Anopheles gambiae* (Francischetti et al., 2002), *Anopheles stephensi* (Valenzuela et al., 2003), *Anopheles darlingi* (Calvo et al., 2004) and *Aedes aegypti* (Valenzuela et al., 2002a), the tick *Ixodes scapularis* (Valenzuela et al., 2002b) and the bug *Rhodnius prolixus* (Ribeiro et al., 2004). In the present work, we explored the proteins and transcripts encoded in the salivary gland of *Lu. longipalpis*, targeting the isolation and full sequencing of the secreted and putative secreted proteins, which are potential markers for vector exposure and vaccine candidates to control *Le. chagasi* infection.

Materials and methods

Sand fly rearing

Lutzomyia longipalpis (Lutz and Neiva 1912), Jacobina strain, were reared at the Walter Reed Army Medical Research Institute and at the Laboratory of Parasitic Diseases at the NIH, using as larval food a mixture of fermented rabbit feces and rabbit food. Adult sand flies were offered a cotton swab containing 20% sucrose and were used for dissection of salivary glands at 0–2 days following emergence (for cDNA library construction) and 4–7 days following emergence for proteomic analysis. Salivary glands were stored in groups of 20 pairs in 20 µl NaCl (150 mmol l⁻¹):Hepes buffer (10 mmol l⁻¹, pH 7.4) at -70°C until needed.

Salivary gland cDNA library

Lu. longipalpis salivary gland mRNA was isolated from 80 salivary gland pairs using the Micro-FastTrack mRNA isolation kit (Invitrogen, San Diego, CA, USA). The PCR-based cDNA library was made following the instructions for the SMART cDNA library construction kit (BD-Clontech, Palo Alto, CA, USA) with some modifications (Valenzuela et al., 2001a, 2002c). The obtained cDNA libraries (large, medium and small sizes) were plated by infecting log phase XL1-blue cells (BD-Clontech), and the amount of recombinants was determined by PCR using vector primers flanking the inserted cDNA and visualized on a 1.1% agarose gel with ethidium bromide (1.5 µg ml⁻¹).

Massive sequencing of cDNA library

Lu. longipalpis salivary gland cDNA libraries were plated to a number of ~200 plaques per plate (150 mm Petri dish). The plaques were randomly picked and transferred to a 96-well polypropylene plate (Novagen, Madison, WI, USA) containing 75 µl of water per well. Four microliters of the phage sample was used as a template for a PCR reaction to amplify random cDNAs. The primers used for this reaction were sequences from the triPLEX2 vector. PT2F1 (5'-AAG TAC TCT AGC AAT TGT GAG C-3') is positioned upstream of the cDNA of interest (5' end), and PT2R1 (5'-CTC TTC GCT ATT ACG CCA GCT G-3') is positioned downstream of the cDNA of interest (3' end). Platinum Taq polymerase (Invitrogen) was used for these reactions. Amplification conditions were: 1 hold of 75°C for 3 min, 1 hold of 94°C for 2 min and 30 cycles of 94°C for 1 min, 49°C for 1 min and 72°C for 1 min 20 s. Amplified products were visualized on a 1.1% agarose gel with ethidium bromide. PCR products were cleaned using the PCR multiscreen filtration system (Millipore, Bedford, MA, USA). Three microliters of the cleaned PCR product was used as a template for a cycle-sequencing reaction using the DTCS labeling kit from Beckman Coulter (Fullerton, CA, USA). The primer used for sequencing, PT2F3 (5'-TCT CGG GAA GCG CGC CAT TGT-3'), is upstream of the inserted cDNA and downstream of the primer PT2F1. Sequencing reaction was performed on a Perkin Elmer 9700 thermacycler (Foster City, CA, USA). Conditions were 75°C for 2 min, 94°C for 2 min and 30 cycles of 96°C for 20 s, 50°C for 10 s and 60°C for 4 min. After cycle-sequencing the samples, a cleaning step was done using the multiscreen 96-well plate cleaning system from Millipore. Samples were sequenced immediately on a CEQ 2000XL DNA sequencing instrument (Beckman Coulter) or stored at -30°C.

Bioinformatics

Detailed description of the bioinformatic treatment of the data can be found elsewhere (Valenzuela et al., 2002c). Briefly, primer and vector sequences were removed from raw sequences, compared against the GenBank non-redundant (NR) protein database using the stand-alone BlastX program found in the executable package at <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/> (Altschul et al., 1997) and searched against the Conserved Domains

Database (CDD) (<ftp://ftp.ncbi.nlm.nih.gov/pub/mmd/cdd/>), which includes all Pfam (Bateman et al., 2000) and SMART (Schultz et al., 1998, 2000) protein domains. The predicted translated proteins were searched for a secretory signal through the SignalP server (Nielsen et al., 1997). Sequences were clustered using the BlastN program (Altschul and Lipman, 1990) as detailed before (Valenzuela et al., 2002c), and the data presented in the format of Table 1. The electronic version of the table (available on request from jvalenzuela@niaid.nih.gov) has additional hyperlinks to ClustalX alignments (Jeanmougin et al., 1998) as well as FASTA-formatted sequences for all clusters.

Full-length sequencing of selected cDNA clones

An aliquot (4 µl) of the λ-phage containing the cDNA of interest was amplified using the PT2F1 and PT2R1 primers

(same conditions as described above). The PCR samples were cleaned using the multiscreen PCR 96-well filtration system (Millipore). Cleaned samples were sequenced first with PT2F3 primer and subsequently with custom primers. Full-length sequences were again compared with databases as indicated for the nucleotide sequences above, and the data displayed as in Table 2.

SDS-PAGE

Tris-glycine gels (4–20%), 1 mm thick (Invitrogen), were used. Gels were run with Tris-glycine SDS buffer according to manufacturer's instructions. To estimate the molecular mass of the samples, SeeBlue™ markers from Invitrogen (myosin, bovine serum albumin, glutamic dehydrogenase, alcohol dehydrogenase, carbonic anhydrase, myoglobin, lysozyme, aprotinin, and insulin, chain B) were used. SGH were treated

Table 1. First 40 clusters from the salivary gland of the sand fly *Lutzomyia longipalpis*

Cluster no.	No. of sequences in cluster	NCBI best match	E values of NCBI match	Pfam best match	E value of Pfam	SignalP result	Comments/ annotations
1	69	gi 4887112 putative RGD-containing peptide	7.00E-38	pfam01028 Topoisomerase_I	1.00E-04	SIG	RGD-like peptide
2	46	gi 4887116 putative yellow-related protein	1.00E-120	No matches found		SIG	Yellow protein
3	30	gi 266511 gi 266511 sp P30659 MAX_LO	5.00E-29	pfam01127 Sdh_cyt	8.00E-07	SIG	Maxadilan
4	30	No matches found		pfam01028 Topoisomerase_I	9.00E-06	SIG	Unknown protein
5	28	gi 4887114 SL1 protein (<i>Lutzomyia longipalpis</i>)	1.00E-65	pfam02386 TrkH	3.00E-05	SIG	SL1-protein
6	25	No matches found		pfam02414 Borrelia_orfA	5.00E-06	SIG	Unknown protein
7	22	No matches found		pfam01604 7tm_5	4.00E-04	SIG	Unknown protein
8	20	gi 4928272 anticoagulant (<i>L. longipalpis</i>)	2.00E-55	Smart smart00034 CLECT	7.00E-20	SIG	Anticlotting 2
9	20	gi 4887116 putative yellow-related protein	2.00E-47	No matches found		SIG	Yellow protein 2
10	19	gi 4887102 antigen 5-related protein	1.00E-115	Smart smart00198 SCP	9.00E-20	SIG	Antigen 5
11	18	gi 4928272 anticoagulant (<i>L. longipalpis</i>)	5.00E-22	Smart smart00034 CLECT	2.00E-20	SIG	Anticlotting
12	17	gi 7301811 CG7592 gene product	8.00E-04	No matches found		SIG	D7 protein – PBP
13	16	gi 4928274 putative apyrase...	1.00E-116	No matches found		SIG	Apyrase
14	15	gi 4928272 anticoagulant (<i>L. longipalpis</i>)	3.00E-90	Smart smart00034 CLECT	2.00E-20	SIG	Anticlotting
15	9	No matches found		No matches found		No ORF	Unknown protein
16	8	No matches found		No matches found		SIG	Unknown protein
17	8	gi 4928272 anticoagulant (<i>L. longipalpis</i>)	5.00E-33	Smart smart00034 CLECT	9.00E-22	SIG	Anticlotting
18	7	No matches found		pfam01490 Aa_trans	1.00E-05	SIG	Unknown protein
19	7	gi 4928272 anticoagulant (<i>L. longipalpis</i>)	1.00E-18	Smart smart00034 CLECT	2.00E-16	SIG	Anticlotting
20	6	No matches found		pfam02414 Borrelia_orfA	1.00E-04	SIG	Unknown protein
21	5	No matches found		No matches found		SIG	Unknown protein
22	5	gi 4887116 putative yellow-related protein	1.00E-43	No matches found		SIG	Yellow protein
23	4	No matches found		No matches found		SIG	Unknown protein
24	4	No matches found		pfam01943 Polysacc_synt	1.00E-04	No SIG	Unknown protein
25	3	No matches found		pfam02414 Borrelia_orfA	6.00E-06	SIG	Unknown protein
26	3	No matches found		pfam01838 DUF40	3.00E-05	SIG	Unknown protein
27	2	gi 7302779 CG4837 gene product	8.00E-36	pfam01009-5_nucleotidase	4.00E-26	SIG	5'-nucleotidase
28	2	gi 2500418 gi 2500418 sp Q24186 RS5_M	2.00E-78	pfam00177 Ribosomal_S7	5.00E-37	No SIG	Ribosomal protein
29	2	gi 4887104 putative alpha-amylase	1.00E-116	pfam00128 alpha-amylase	1.00E-41	SIG	Amylase
30	2	No matches found		No matches found		ANCH	Unknown protein
31	2	gi 4588920 ribosomal protein S14	4.00E-55	pfam00411 Ribosomal_S11	2.00E-37	No ORF	Ribosomal protein
32	2	No matches found		No matches found		No SIG	Unknown
33	2	gi 8745398 putative adenosine deaminase	8.00E-57	No matches found		No ORF	Adenosine deaminase
34	2	No matches found		pfam02695 DUF216	6.00E-05	SIG	Unknown protein
35	1	gi 1709616 gi 1709616 sp P54399 PDI_DROM	4.00E-70	pfam01216 Calsequestrin	1.00E-05	No ORF	PDI Prolyl hydroxylase
36	1	gi 1168228 gi 1168228 sp P41570 6PGD_CER	3.00E-92	pfam00393 6PGD	3.00E-98	No SIG	6-phosphogluconate dehydrogenase
37	1	gi 10726855 CG7866 gene product	2.00E-10	No matches found		No SIG	Unknown protein
38	1	gi 2499905 gi 2499905 sp Q10715 ACE_HAEI	7.00E-43	pfam01401 Peptidase_M2	2.00E-15	SIG	Dipeptidyl-peptidase
39	1	gi 12848978 putative (<i>Mus musculus</i>)	1.00E-100	pfam00189 Ribosomal_S3_C	4.00E-22	No SIG	Ribosomal protein
40	1	gi 7300077 glob1 gene product (alt 1)	3.00E-21	pfam00042 globin	6.00E-16	No ORF	Globin

SIG, signal secretory peptide present; No ORF, no open reading frame; No SIG, no signal peptide present; ANCH, probably anchored protein.

with equal parts of 2×SDS sample buffer (8% SDS in Tris-HCl buffer, 0.5 mol l⁻¹, pH 6.8, 10% glycerol and 1% bromophenol blue dye). Thirty pairs of homogenized salivary glands per lane (approximately 30 µg protein) were applied when visualization of the protein bands stained with Coomassie blue was desired. For amino-terminal sequencing of the salivary proteins, 40 homogenized pairs of salivary glands were electrophoresed and transferred to polyvinylidene difluoride (PVDF) membrane using 10 mmol l⁻¹ CAPS, pH 11.0, 10% methanol as the transfer buffer on a Blot-Module for the XCell II Mini-Cell (Invitrogen). The membrane was stained with Coomassie blue in the absence of acetic acid. Stained bands were cut from the PVDF membrane and subjected to Edman degradation using a Procise sequencer (Perkin-Elmer Corp.). To find the cDNA sequences corresponding to the amino acid sequence obtained by Edman degradation, we used a search program (written in Visual Basic by J. M. C. Ribeiro) that checked these amino acid sequences against the three possible protein translations of each cDNA sequence obtained in the DNA sequencing project. A more detailed account of this program is found elsewhere (Valenzuela et al., 2002c).

Results

Massive cDNA sequencing

An unamplified cDNA library from the salivary glands of the female sand fly *Lu. longipalpis* was plated, and 550 plaques were randomly picked and sequenced. The resulting sequences were clustered using BlastN with a cutoff of 10E-20, obtaining 143 unique clusters of related sequences. All sequences within each cluster were compared with the non-redundant (NR) protein database using the BlastX program (Altschul et al., 1997) and with the CDD database, containing all Pfam and

SMART motifs (Bateman et al., 2000; Schultz et al., 2000), using the RPS-BLAST program (Altschul et al., 1997). The three possible reading frames of each sequence were inspected for long-reading frames with an initial methionine residue followed by at least 40 residues; these were submitted to the SignalP server (<http://www.cbs.dtu.dk/services/SignalP-2.0/>) for verification of secretory signal peptide. This resulted in the identification of 35 clusters coding for proteins containing a secretory signal peptide. These analyses are summarized in Table 1. Because of space limitations, we show only the first 40 clusters; the remaining clusters can be obtained from the electronic version of this table (available on request from jvalenzuela@niaid.nih.gov).

Table 1 shows the most abundant clusters in descending order, from the most abundant sequences to the least. From these 40 clusters, 28 contain sequences with predicted secretory proteins, and those remaining represent clusters for housekeeping genes or unknown sequences without a clear secretory signal peptide. We found 65 sequences (out of 550) of probable housekeeping genes, arranged in 61 clusters, an average of 1.07 sequences per cluster. Examples of these housekeeping genes are: protein disulphide isomerase (cluster 35); 6-phosphogluconate dehydrogenase (cluster 36); ribosomal protein (cluster 39); actin (cluster 54); adenylate kinase (cluster 92); ATP-synthase (cluster 65) and cytochrome oxidase (cluster 57). These clusters represent only the clusters from the middle of Table 1 downwards, with only a few sequences per cluster (most are only one sequence per cluster) representing the low abundant messages.

Another set of cDNA found in this library contains clones that do not have similarities to other genes in the NCBI databank, do not have an assigned function (by CDD analysis) and do not have a secretory signal peptide. We found 34 sequences (out of 550) of these types of genes arranged in 23 clusters (1.48 sequences per cluster).

The most abundant clones found in this cDNA library are the ones coding for secretory proteins. We found 451 cDNA (out of 550 sequences) of potentially secreted proteins arranged in 40 clusters (an average of 11.27 sequences per cluster). Although four of these clusters (cluster 33 with two sequences, cluster 84 with one sequence, cluster 88 with one sequence and cluster 110 with one sequence) resulted in no signal peptide in our analysis, we included these sequences

Fig. 1. Proteome analysis of *Lutzomyia longipalpis* salivary proteins. Salivary proteins were separated by SDS-PAGE and transferred to PVDF membrane as described in the Materials and methods section. N-terminal sequences are annotated on the left, and their corresponding cDNAs are annotated on the right.

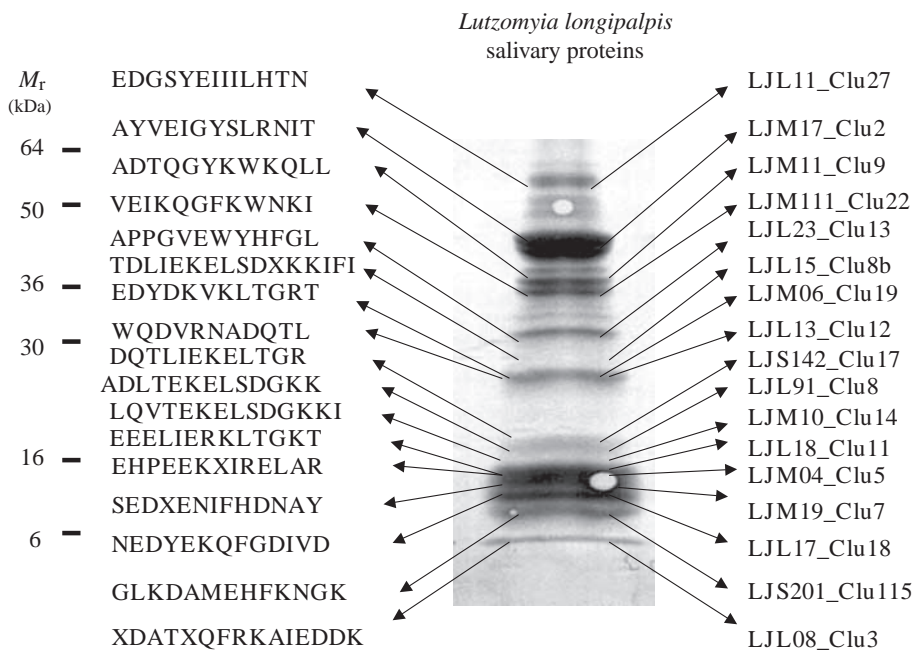


Table 2. Salivary gland cDNA coding for secreted proteins

Sequence	Cluster no.	SignalP result	Cleavage position	Mature M_r (kDa)	pI	NR best match	E value	Match extent	% Match	Present in proteome
LJL35	1	SIG	23-24	6.04	3.97	Putative RGD-containing peptide	2E-038	76/76	100	
LJM17	2	SIG	18-19	45.15	5.71	Yellow-related protein	0	399/412	96	Yes
LJL08	3	SIG	23-24	6.87	8.91	Maxadilan <i>Lutzomyia longipalpis</i>	4E-028	60/86	69	Yes
LJL38	4	SIG	20-21	2.46	3.34					Yes
LJM04	5	SIG	20-21	13.76	9.09	SL1 protein	9E-072	129/139	92	Yes
LJS192	6	SIG	23-24	9.64	4.21					
LJM19	7	SIG	22-23	10.74	4.15					Yes
LJL91	8	SIG	19-20	16.32	5.83	Anticoagulant (<i>L. longipalpis</i>)	8E-056	97/160	60	Yes
LJL15	8b	SIG	19-20	16.45	6.11	Anticoagulant (<i>L. longipalpis</i>)	4E-055	97/160	60	Yes
LJM11	9	SIG	18-19	43.22	9.32	Yellow-related protein	1E-110	197/405	48	Yes
LJL34	10	SIG	19-20	28.79	9.09	Antigen-5 related protein	1E-165	271/271	100	
LJL18	11	SIG	19-20	16.29	6.49	Anticoagulant (<i>L. longipalpis</i>)	1E-022	53/160	33	Yes
LJL13	12	SIG	19-20	26.39	4.93	D7 salivary protein (<i>L. longipalpis</i>)	1E-150	247/247	100	Yes
LJL23	13	SIG	21-22	35.03	9.14	Putative apyrase (<i>L. longipalpis</i>)	0	325/325	100	Yes
LJM10	14	SIG	19-20	16.64	8.61	Anticoagulant (<i>L. longipalpis</i>)	3E-093	160/160	100	Yes
LJL143	16	SIG	23-24	32.41	8.45	Hypothetical protein (<i>P. falciparum</i>)	0.030	33/115	28	
LJS142	17	SIG	20-21	16.62	7.07	Anticoagulant (<i>L. longipalpis</i>)	3E-033	63/139	45	Yes
LJL17	18	SIG	20-21	10.11	4.31	CCAAT-box-binding transc. factor	0.091	23/67	34	Yes
LJM06	19	SIG	19-20	16.32	8.9	Anticoagulant (<i>L. longipalpis</i>)	8E-024	59/155	38	Yes
LJL04	20	SIG	17-18	29.21	10.16	32 kDa salivary protein (<i>P. papatasi</i>)	3E-018	53/150	35	
LJM114	21	SIG	24-25	14.17	5.55					
LJM111	22	SIG	18-19	42.95	4.85	44 kDa salivary protein (<i>P. papatasi</i>)	1E-095	182/399	45	Yes
LJM78	23	SIG	20-21	37.18	7.7					
LJS238	25	SIG	20-21	4.63	7.32					
LJS169	26	SIG	22-23	11.55	4.54	PDZ domain protein (<i>P. yoelii</i>)	0.022	27/109	24	
LJL11	27	SIG	25-26	60.55	6.94	5'-nucleotidase_ <i>L. longipalpis</i>	0	540/560	96	Yes
LJS105	34	SIG	19-20	7.29	4.53					
LJL09	38	SIG	18-19	71.06	5.58	Angiotensin converting enzyme	0	404/595	67	
LJM26	56	SIG	17-18	48.77	5.78	Serine protease inhibitor SERPIN	3E-080	169/423	39	
LJS03	58	SIG	19-20	15.06	4.02	Erythrocyte membrane-associated	4E-008	30/97	30	
LJL124	67	SIG	20-21	6.01	4.31					
LJL138	71	SIG	20-21	43.72	9.45	agCP9602 (<i>Anopheles gambiae</i>)	3E-041	117/376	31	
LJS138	97	SIG	20-21	16.11	5.84	agCP10141 (<i>Anopheles gambiae</i>)	4E-052	103/173	59	
LJS193	113	SIG	20-21	32.17	6.34	Palmitoyl-(protein) hydrolase	2E-089	154/275	56	
LJS201	115	SIG	23-24	8.56	4.86	orf 48 (ateline herpesvirus 3)	0.092	25/77	32	Yes

as secretory proteins because they are truncated versions of the salivary adenosine deaminase cDNA and the salivary hyaluronidase cDNA, previously analyzed and reported to be secretory proteins (Charlab et al., 1999). Therefore, the number of cDNAs coding for secretory proteins is 10.53 times greater than the cDNA coding for housekeeping genes and 7.61 times greater than the cDNA coding for proteins with unknown function; overall, cDNA coding for secretory proteins is 4.4 times greater than cDNA coding for non-secreted proteins. The cDNAs coding for secretory proteins represent 82% of the cDNAs sequenced in this sand fly library. In fact, out of the first 40 clusters, 28 belong to cDNA coding for secretory proteins (Table 1).

Proteome analysis of *Lu. longipalpis* salivary proteins

The supernatant of *Lu. longipalpis* SGH was separated by one-dimension SDS-PAGE, the proteins transferred to PVDF membrane, and Edman degradation performed on the Coomassie blue-stained bands. Fig. 1 (left side) shows the resulting N-terminal sequence of 17 proteins. By searching our database using an in-house program (written by J. M. C.

Ribeiro) similar to the BLOCK program, we were able to identify all the coding cDNA, which correspond to the 14 N-terminal sequences (right side of Fig. 1). We identify by Edman degradation the N-terminus of the cDNA corresponding to clusters 2, 3, 5, 7, 8, 8b, 9, 11, 12, 13, 14, 17, 18, 19, 22, 27 and 115. With the exception of cluster 115, all clusters were within the first 40 clusters. Not all the attempted Edman degradation experiments resulted in a sequence, probably because of the small amount of protein or because these proteins were blocked at the N-terminal end.

Full-length sequence of clones coding for secreted proteins

Because we are primarily interested in the identification and characterization of secreted proteins from this cDNA library, we selected the 35 clusters containing sequences with a clear signal peptide and obtained full sequence on all of them. Table 2 shows the analysis of these sequences, including the signal peptide cleavage site, molecular mass, isoelectric point and best match to NCBI database. The last column of Table 2 indicates whether the clone has been found in the proteome analysis. Extensive analysis is presented in the electronic

A

```

gi_15963519_Phlebotomus -----
gi_15963517_Phlebotomus -----
Clu2_Lutzomyia_ -----
Clu9_Lutzomyia_ -----
Clu22_Lutzomyia_ -----
gi_21297027_Anopheles -----RSQQSFSTVSVFFSKMAKWAIVFGIIFLIG--LFS--E
gi_18487033_Drosophila MLSPQFLILSLISGLQLLSLADADPMIEVFRWKQMDFFNRGDHLSGGGR
gi_19335662_Aedes -----ML

gi_15963519_Phlebotomus -----MKFFLSVIALASFOYVFCDDVERFYAWRNITFDVKEGTYKPGD
gi_15963517_Phlebotomus -----MMKLLILCVLSPLSLQVALSDDVGRAYEISEIKLVGRVFNAYDSGN
Clu2_Lutzomyia_ -----MRFFVFLVLAIVLFCGIGHGAYVEIGYSLRNITFDGLDITDYNPKF
Clu9_Lutzomyia_ -----MKVFFSIFTLVLFQGTGADT--QGYKWKQLLYNNVTPGSGYNPDN
Clu22_Lutzomyia_ -----MKLFFFLYTFGLVQTFIFGVEIKQGFKNKILYEGDTSNFNPN
gi_21297027_Anopheles ELRITSAEKQLLVYQWSQLKSSNSSEASQNVVFAANSVPAEDT--FNAYG
gi_18487033_Drosophila KDRPSFSAVVPVPGKYSRRKRKIMTSRDTPVVSSRADSDDDNASYIPYN
gi_19335662_Aedes RLVLFCVVAFAALTVVSGQRQVEEVLKQVQVEFDPVASVLSAPDGYIPIN

gi_15963519_Phlebotomus VIPTGVTHDAKTKKLYFGVPRRYSNIPVLAIEIDTRNYSRSEIRSP---
gi_15963517_Phlebotomus IIVPTGVAYDAASKMLPFGIPRIYSRVPITFAQLSTRYSNASEIPNPE---
Clu2_Lutzomyia_ NIPITGLAVDPEGYRIFLIAIPRRKPKVPIYVAELNLMVNNPFPVVERASFE
Clu9_Lutzomyia_ MISTAFAYDAEGEKFLAVLPRKLPVPIYVAEVDTKNSLVGKGHSE---
Clu22_Lutzomyia_ NILLTAFAYDPEQKLFPLTVPRKYPETMYLAIEVDTEKNSPESGDTSE---
gi_21297027_Anopheles NLEPMGVTHHKG--FLYVTIPRRRPGIPALNVIDLARNPANNPTLD---
gi_18487033_Drosophila NVPMGATHFRG--RLFVTIPRRRVGIPSLNVIDLAEDGSGNRSPKLR---
gi_19335662_Aedes NIPMSGVHYKN--RVFVTIPRRRWGIPSLNVIIELEPPYVPTNPVLK---

gi_15963519_Phlebotomus -FSKFNQSG-----KEFTSIVQPVIDDORRLWLDVGOVDYKK-HG
gi_15963517_Phlebotomus -LDKFSGSKS-----QPLTSVYQPVIDDORRLWLDVGHVENE-ER
Clu2_Lutzomyia_ KFKKENGEGK-----KDLNVVYQPVIDDORRLWLDVCKVEYTGGA
Clu9_Lutzomyia_ LLNKFESGHTG-----KELTSIVQPVIDDORRLWLDVCSVEYRSGA
Clu22_Lutzomyia_ LLGKFSGHGTG-----KELTSVYQPVIDDORRLWLDVCSVERNSDGT
gi_21297027_Anopheles PYPNLLVNSLRDFTADPKRLISVYRTQVDRCDRLWFDVPCVLEYPG--G
gi_18487033_Drosophila AYPNFALN-----QFNASAENLVSVYRTSDVACDRLWFDVPCVLEYPN--N
gi_19335662_Aedes PYPSEFELNELRADLQPDANRLVTVYRPRVDRCORLWFDVPCVMEIPG---

gi_15963519_Phlebotomus NEYPTKPEIIAFLDNLQEGNPEVHRVYKLEGDVARSPLGFGFPAVDVINPN
gi_15963517_Phlebotomus KTYPIKPESLIAFOLDTKSNYPEIHRVELTGEAGKNPLGYGGFPAVDVWPK
Clu2_Lutzomyia_ DQYPKGPEFLIAYDLKDKDHTPEIHRPEIPDDLSSQVEFGFPAVDVWNTK
Clu9_Lutzomyia_ KDYPSHPEALIVAYDLQKNYPEVIVYFPFLRVEKPTVYFGFPAVDVWPK
Clu22_Lutzomyia_ EGQPSHPEALIVAYDLKKEANYPEVIVYFPFDNSLEKPTVYFGFPAVDVWPK
gi_21297027_Anopheles AGRQVQPEALWLDLITIN--RSVRPEIPEMVEFGYIPNIEVVDVDRD
gi_18487033_Drosophila R--QQRPEPSIHWVVDLADT--QVLKPEFVPSIEAETGRGLASITVYKAGQ
gi_19335662_Aedes NFTVQVQPEPSIWSIDLKTN--QPLSPVEIPQKDVETGYGLTSLITLWDPDD

gi_15963519_Phlebotomus GNAKSDETYIITNFDNALIVYDMKNKNAMKFNDDSEKPEPKSVFNH
gi_15963517_Phlebotomus RCDKNEKTYIIANFDENSLIVYDKKGEANSLKDDSEKPE--GVTTFTL
Clu2_Lutzomyia_ GDCT---ESFVLTNFKDNSLIVYDQTKKAMKFTDKTEAD--KESTFSY
Clu9_Lutzomyia_ GDCS---ETFTVITNFKLGFALIVYDHHKQDSNIVPTPEKAE--RFTKFDY
Clu22_Lutzomyia_ -ECS---ETFTVITNFKLGFALIVYDHHKNDSTVQDSTEGPD--KKSFDH
gi_21297027_Anopheles --CG---RAYAIPDYEWRRLLYVYGLTEGRMRFPHNVEFPEPRYCDFINV
gi_18487033_Drosophila --CG---DAYAIPDLYRRLYVYHLRNDRIYDFEHNVEFDFPLSGDLSI
gi_19335662_Aedes --CS---KVFVYISDLQYRMVYVYDHEHQSRFLHNYEFLNPLEGDFNI

gi_15963519_Phlebotomus KEEQ--YSYIACHGEGITLGDNRKD--GHRPAYVIAGSSTKLYSVNTEASEKE
gi_15963517_Phlebotomus NCKE--HKFKACHGEGIALGDRNKE--GNRPAYVLACSSTKLYRLDKLKK
Clu2_Lutzomyia_ SBEQMKYKACHGEGIALGDRDEM--GHRPAYVIAGSSTKLYSVNTEASEKE
Clu9_Lutzomyia_ GCKE--YEFKACHGEGITLGDNRDSE--GNRPAYVLACSSTKLYSVNTEASEKE
Clu22_Lutzomyia_ DEQQ--YEYEAACHGEGITLGERDNE--GNRQAYVLVASTKLYHSINTKEDKQ
gi_21297027_Anopheles AQQQ--FTWYDCHGFSVAVGGRWENSAABRTVYLHAMASVSEIAVSNRLQD
gi_18487033_Drosophila GEQT--FRWDDCHGFSITLGAQKLD--GSDRAYFHPMASTNEFVVSNRVLQD
gi_19335662_Aedes QETP--FAWDDCHGFSIALSNPDPM--KFRPAYFHALSSNSFEFTVSTAVLRN

gi_15963519_Phlebotomus KGASLKPR-----LLEPRGFKTEALALAYDEPKTKVIFVFEVSDSRQVS
gi_15963517_Phlebotomus KSKKLEPK-----LLEDRGFKTEALALAYDEPKTKVLFVFAEDSRQVS
Clu2_Lutzomyia_ ENGQLNPQ-----LLEDRGKYTDAALALAYDEPHKVLVFAEDSRQVS
Clu9_Lutzomyia_ KGGKLNPE-----LLENRGKYNDALALAYDEPKTKVIFVFAEANTQVS
Clu22_Lutzomyia_ KGSKNVAN-----YLDRGSESTDAICLVLYDEPKTKVIFVFEVSNKRQVS
gi_21297027_Anopheles ETMAQLGSGAYGRAFHHLGARGPNTQSSSHAYDPRTGVLFYAEINRNAIG
gi_18487033_Drosophila ESNAARSD--HGNDFRVLGSRGPSTQSTHAYDPRTGVIFFDEIQRNQVG
gi_19335662_Aedes ETASKRGY--HGDDFKLLEVRGAQSQSSIHGFPEETGVIFFALVQLNAVS

gi_15963519_Phlebotomus CWNIKQE--LIPKNVGVLYTN--AYVFGCDIMVADSTLIPMSNAHPETEL
gi_15963517_Phlebotomus CWNIKHE--LKPENVMGIYAN--PNFNFQDLMVSKGFLIPMSNGQPEID-
Clu2_Lutzomyia_ CWNVME--LKPNDTDFVFS--ARFTFGDILVSKGMLIMMANGHPEVVD
Clu9_Lutzomyia_ CWNQKMPLRMKNNTDVVYTS--SRFVFGDLSVSKGGLIPMSNGFPBIRK
Clu22_Lutzomyia_ CWNQQT--LNDKIDVLYHN--ADFSFGDLSISQDNLWPLANGLPELEN
gi_21297027_Anopheles CWNTHKQ--PTPENHGVHLDNEEMIYPADLISDGLVDESDGLVSNRLPTWY
gi_18487033_Drosophila CWKTSKP--ISAENYGSVDSNAEDMIYPSDLSIDEDGTIWMNSMPLIFY
gi_19335662_Aedes CWDTRPK--FAPQNMALIVYKNDRIIYPNLSDIDQEGNVPMNSIISKLLY

gi_15963519_Phlebotomus ---PKLDFDKRQIRLMYVP--THRAIRNLGCEVRKPK-----
gi_15963517_Phlebotomus ---EKMEYDVPQIRLMKVK--TKRAIKGERKQG-----
Clu2_Lutzomyia_ QEKIKMRPFVNRKIRIMKVD--TERVFKYSRQNPYKPKKEIEV---
Clu9_Lutzomyia_ S---EKFKYDPPRYLRIMMDTQEAIACTACDMNA-----
Clu22_Lutzomyia_ S---DKFVFTKPRYQIFKVN--IQEAIACTKCEKLN-----
gi_21297027_Anopheles S---QLNRT--EVNRYLRQVS--ASRAVAGTVCA-----
gi_18487033_Drosophila S---TLDTS--IYNFRWQK--ASLAKRGTVC-----
gi_19335662_Aedes ---TQLSLEEDFPHIWRAN--IKEIKGTVDPTVPPNVHDGQRFGGEGN
    
```

B

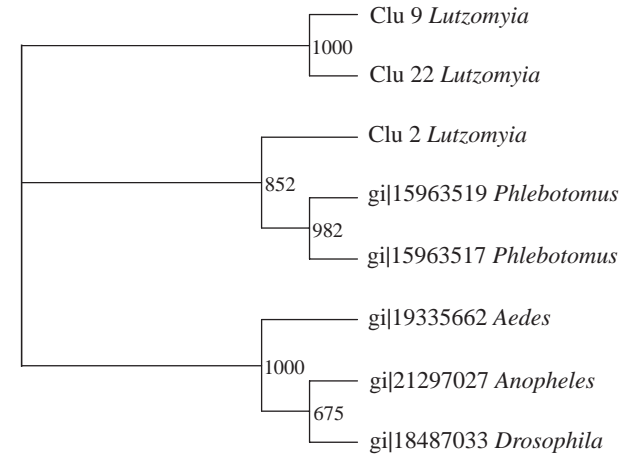


Fig. 2. (A) ClustalW analysis of yellow-related proteins from different Diptera, including *Phlebotomus papatasi* salivary proteins, *Aedes aegypti* dopachrome converting enzyme, *Anopheles gambiae* and the three yellow-related proteins found in the salivary gland cDNA library of *Lutzomyia longipalpis*. Reverse-shaded amino acids represent identical amino acids among all these proteins; gray-shaded amino acids represent identical amino acids among phlebotomine yellow-related proteins. (B) Dendrogram of the yellow-related proteins indicating the distribution of three distinct clades. Bootstrap values are indicated in their respective nodes.

version of Table 2 (available upon request from jvalenzuela@niaid.nih.gov).

The selected cDNAs in Table 2 are arranged in descending order from the cluster containing the largest number of cDNA to the clusters containing the least number of cDNA to the most abundant transcripts to the least abundant transcripts in this cDNA library).

Description of cDNAs coding for secreted proteins Cluster 1 (LJL35; GenBank acc. no. AF132516) – RGD peptide

The cDNA from this cluster codes for a secreted 6 kDa peptide with no similarities to other proteins in databanks. We previously isolated this transcript by PCR subtraction technique (Charlab et al., 1999). The function of this protein remains unknown, although it may inhibit platelet aggregation by interfering with the binding of platelet receptor to fibrinogen by the RGD motif present on it. We did not find this protein in our proteomic analysis, probably because of its small size or because the N-terminal amino acid is blocked.

Clusters 2, 9 and 22 – family of yellow-related proteins

Proteins from this family were found in cluster 2 (GenBank acc. no. AF132518), cluster 9 (GenBank acc. no. AY445935) and cluster 22 on this cDNA library.

Cluster 2 codes for a secreted protein of 45 kDa

LJL38_Clu4
MKTFALIFLALAVFVLCIDGAPTFVNLDDVQEEVEVNTYEP

LJS192_Clu6
 MVKYSLVLAIFLLAGPYGVVGCENDLTEAAKYLQDECNAGEIADFPLPFS
 EEEVGEALSDKPEVQEVNTIVRGCFEAEQAKEHGKRCERFSALSQCYLEKNL
 CQFF

LJM19_Clu7
 MKFFYLIFSALFLLADPALVKCEDCENIFHDNAYLLKLDCEAGRVPVEYDDI
 SDEEYIETVDVGVSSSEDQEKVAKIIRECIAQVSTQDCTKFSEIYDCYMKKKIK
 NYYPENM

LJL17_Clu18
 MQNFLLVSLALAAALMLCAEAKPYDFPLYQDLIQGVQRESQAEREKRSPNED
YEKQFGDIVDQIKIEISFNVMMKMPHFSSDDNRDDGEYVDHGHYDDEDRDYK
 HY

LJS238_Clu25
 MLKIVLFLSVLAVLVICVAAMPGNSVNPWHISREELEKLRARKSNHKALEKAID
 ELIDKYL

LJS169_Clu26
 MKFSCPVEVAIFLLCGFYRVEGSSQCEEDLKEEAEAFFKDCNEAKANPGEYEN
 LTKEEMFEELKEYGVADTDMETVYKLVVEECWNELTDTDCRFLFEEAECKF
 KKNICKYFPDEVKLLKKK

LJS105_Clu34
 MNVLFVSTLITILLCVKARPEDFVALQDQANFQKCLEQYPEPNQSGEVLACL
 KKREGAKDFREKRSLDDIEGTQESGNLWGA

LJL124_Clu67
 MVSILLISLILNLLVFYAKARPLEDISSDLSPDYITTEGYDGVKEKREIELVPVT
 FGIFNIHTTPAPRITFEW

LJS201_Clu115
 MRNFAVVSALAVAVLLFCAWPINAEDNEEVGKAREKRGLKDAMEHFKNQFKK
LTKDQKFLPSLPSLPGFGKPKPESGSSSEDSGDKTETDSGSKDDQSKDNTVEES

Fig. 3. Small proteins and peptides of novel sequences present in the salivary glands of the sand fly *Lutzomyia longipalpis*. Protein sequences represent the translation of their corresponding cDNAs. The signal secretory peptide is underlined. Amino acids of mature peptides are in bold. N-terminus peptides observed in proteomics analysis that differ from predicted SignalP programs are highlighted in gray.

previously characterized by PCR subtraction (Charlab et al., 1999). The gene coding for this protein was first described in *Drosophila melanogaster* (Geyer et al., 1986). When comparing with other Diptera, this protein was not found in the salivary glands of *Ae. aegypti* (Valenzuela et al., 2002a), *An. gambiae* (Francischetti et al., 2002) or *An. stephensi* (Valenzuela et al., 2003). It was, however, described in *Ae. aegypti* whole-larvae extract and purified as a dopachrome-converting enzyme (Johnson et al., 2001). The homologue of *Lu. longipalpis* yellow protein was also described in the salivary glands of the sand fly *P. papatasi* (Valenzuela et al., 2001b). The function of this protein in the saliva remains to be elucidated. In the proteomic analysis (Fig. 1), this cluster codes for one of the most abundant proteins (AYVEIGYSLRNIT) in the saliva of this sand fly. Interestingly, a protein with similar molecular mass was the most recognized antigen by the sera of individuals who were exposed to sand flies and developed a cellular immunity to *Le. chagasi* (Gomes et al., 2002). Therefore, this protein is a candidate marker for epidemiological studies on vector exposure and may be a potential vaccine candidate to control *Le. chagasi* infection.

This yellow-related protein is highly similar to proteins

found in cluster 9 (GenBank acc. no. AY445935) and cluster 22. Fig. 2 shows the ClustalW alignment of this protein with related proteins, including yellow-related salivary proteins from *P. papatasi* (Valenzuela et al., 2001b), *Ae. aegypti* midgut dopachrome conversion enzyme (Johnson et al., 2001) and *D. melanogaster* yellow protein. Similarities among the phlebotomine's yellow proteins are greater than non-salivary yellow from other Diptera and are marked as gray-shaded amino acids (Fig. 2A). Bootstrap neighbor-joining analysis (Fig. 2B) resulted in grouping insect yellow proteins in three distinct clades: the first group contains the *Aedes*, *Anopheles* and *Drosophila* yellow-related proteins (non-salivary); the second group contains the two *P. papatasi* yellow-related salivary proteins and the *Lu. longipalpis* yellow salivary protein from cluster 2; and the third group contains the *Lu. longipalpis* yellow salivary proteins from clusters 9 and 22.

Cluster 3 (*LJL08*; GenBank acc. no. M77090) – maxadilan

This cluster codes for the most studied salivary protein from a sand fly. It was discovered to be the most potent vasodilator from any organism (Ribeiro et al., 1989) and later as an immunomodulatory molecule (Soares et al., 1998). Recently, Morris et al. reported that mice vaccinated with this 6 kDa peptide produced antibodies against this molecule, and these animals were protected against *Le. major* infection (Morris et al., 2001). The N-terminal of this protein, XDATXQFRKAIEDDK, was found in the proteomic analysis. Surprisingly, the cDNA we found in this library is only 69% identical to the previously reported maxadilan. Differences may be due to strain differences or the polymorphism reported for this molecule (Lanzaro et al., 1999).

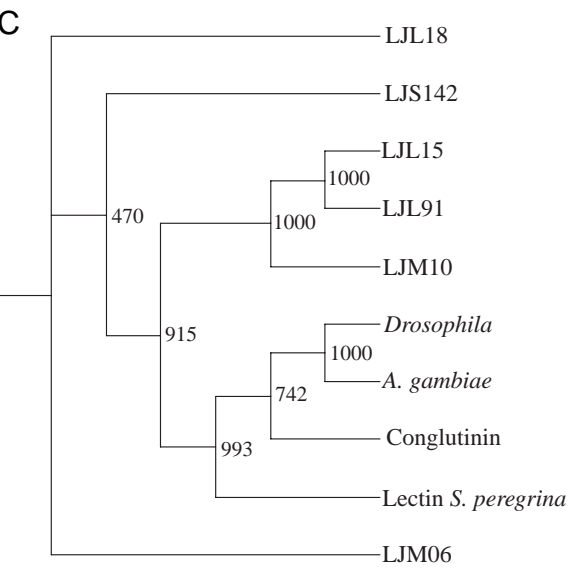
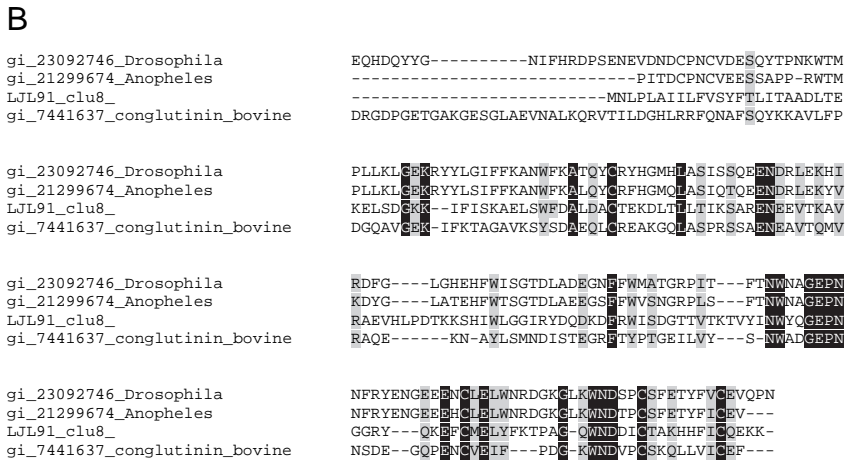
Clusters 4, 6, 7, 18, 25, 26, 34, 67 and 115 – novel peptides from the saliva of *Lu. longipalpis*

With our approach, we identified a number of small peptides in this cDNA library. Only two peptides were previously reported from the salivary glands of *Lu. longipalpis*: maxadilan (Ribeiro et al., 1989) and an RGD-containing peptide (Charlab et al., 1999). Following is the description of the nine clusters containing these novel peptides. Sequences of these peptides are shown in Fig. 3.

(1) Cluster 4 (*LJL38*; GenBank acc. no. AY438269). This cDNA codes for a small secreted peptide of only 2.5 kDa, with no similarities to other proteins in the searched databases. This peptide was previously isolated and fully sequenced from the saliva of *Lu. longipalpis* by molecular sieving and reverse-phase HPLC (J. M. C. Ribeiro, unpublished results). The function of this protein remains unknown.

(2) Cluster 6 (*LJS192*; GenBank acc. no. AY438270). This cDNA codes for a secreted protein of 10 kDa, with no similarities to other proteins in the searched databases. Searching a motif database, this protein matched pfam02414 from *Borrelia* orfA. This cDNA appears to be unique to sand flies and has not been found in any mosquito, tick or triatomine salivary gland cDNA libraries.

(3) Cluster 7 (*LJM19*; GenBank acc. no. AY438271). This



cDNA codes for a secreted protein of 11 kDa, with no similarities to other proteins in the databank. The N-terminus of the predicted salivary protein, SEDXENIFHDNAY, was found in the proteomic analysis.

(4) Cluster 18 (LJL17; GenBank acc. no. AY452695). This cluster codes for a secreted protein of 10 kDa, with no

Fig. 4. (A) ClustalW analysis of the six C-type lectin proteins found expressed in the salivary glands of *Lutzomyia longipalpis*. (B) ClustalW analysis of C-type lectin-related proteins from various organisms, including insects and a mammal. Reverse-shaded amino acids indicate identical amino acids among all these proteins; gray-shaded amino acids represent highly conserved amino acids. (C) Dendrogram of the C-type lectin-related proteins from *Lu. longipalpis* and other organisms. Bootstrap values are indicated in their respective nodes.

similarities to other proteins in the databank. The N-terminus sequence, NEDYEKQFGDIVD, found in the proteomic analysis, is not exactly in the position predicted by the SignalP program (Fig. 3). The observed N-terminus sequence suggests that a cleavage occurred at position P₄₉-N₅₀, resulting in a peptide of 7 kDa. The cleavage was probably due to either a protease in the saliva or the fact that this peptide bond is unstable at the conditions where the protein was isolated.

(5) Cluster 25 (LJS238; GenBank acc. no. AY455909). This cluster codes for a small secreted protein of 5 kDa, with no similarities to other proteins in databases searched.

(6) Cluster 26 (LJS169; GenBank acc. no. AY455912). This cluster codes for a secreted

protein of 11.5 kDa, with no significant similarities to other proteins in databases.

(7) Cluster 34 (LJS105; GenBank acc. no. AY455910). This cDNA codes for a protein of 7 kDa, with no significant similarities to other proteins in the databases searched.

(8) Cluster 67 (LJL124; GenBank acc. no. AY455915). This cDNA codes for a protein of 6 kDa, with no significant similarities to other protein in the searched databases.

(9) Cluster 115 (LJS201; GenBank acc. no. AY455919). This cDNA codes for a protein of 9 kDa, with no significant similarities to other proteins in the databases searched. The N-terminus sequence, GLKDAMEHFKNKGKELTTKDF, found in the proteomic analysis, is not exactly in the position predicted by the SignalP program (Fig. 3). The observed N-terminus sequence suggests that a cleavage occurred at position R₃₆-G₃₇, resulting in a peptide of 7.1 kDa. The cleavage was probably due to a protease in the saliva or the fact that this peptide bond is unstable at the conditions where the protein was isolated.

Cluster 5 (LJM04; GenBank acc. no. AF132517) – PpSP15-like proteins

This cDNA was previously isolated by PCR subtraction and named SL1. This cDNA codes for a protein of 14 kDa and is similar to the protein of 15 kDa from *P. papatasi* (PpSP15) that

1 MIKEVFLALLVALAQCANEIPINRQGDYVPVPIIDPNKSSDDY
 46 FDDRFYPDIDDEGIAEAPKDNRGKSRGGGAAGAREGRLGTNGAKP
 91 **GQG**GTRP**GQG**GTRP**GQG**GTRP**GQG**GTRP**GQG**GTRP**GQG**RTKPAQG
 136 TTRPAQGTRNPGSVGTKEAQDASKQ**GQG**KRRPGQVGGKRPQANA
 181 PNAGTRKQKQKSGRQVGRPDLRSRYKDAPAKFVFKSPDFSGEGKTPT
 226 VNYFRTKKKEHIVTRGSPNDEFVLEILDGDPTGLGLKSETIGKDT
 271 RLVLENPNGNSIVARVKIYKNGYSG

Fig. 5. Predicted protein sequence from the cDNA of cluster 20 (LJL04). Reverse-shaded amino acids and gray-shaded amino acids represent the repeats found in this protein. Bolded amino acids represent a putative signal secretory peptide.

conferred protection against *Le. major* infection. The N-terminus predicted from this cDNA, EHPEEKXIRELAR, was present in our proteomic analysis (Fig. 1). Previously, we reported three similar proteins (three different clusters) on the salivary glands of *P. papatasi* (PpSP12, PpSP14 and PpSP15). We found only one cluster of this family of proteins in the *Lu. longipalpis* cDNA library.

Clusters 8, 8b, 11, 14, 17 and 19 – family of putative anticoagulants (C-type lectin)

One of the most abundant families of protein found in this cDNA library is the family of putative anticoagulants with homology to C-type lectins. Six clusters (8, 8b, 11, 14, 17 and 19) belong to this family.

Cluster 8 (LJL91; GenBank acc. no. AY445934). This cDNA codes for a secreted protein of 16 kDa and is similar to the previously described anticoagulant from *Lu. longipalpis* (Charlab et al., 1999). These proteins contain a C-type lectin or C-type lectin-like domain. This domain functions as a calcium-dependent carbohydrate-binding pocket involved in extracellular matrix organization, pathogen recognition and cell-to-cell interactions. Factor IX/X and Von Willebrand factor binding proteins contain these domains, suggesting that the anticoagulant(s) of *Lu. longipalpis* may be binding some of the targets of these coagulation factors and inhibiting the blood coagulation cascade. An alignment showing all the C-type lectin proteins found in this cDNA library is shown in Fig. 4A. A ClustalW alignment of this protein and other proteins containing lectin-like domains is shown in Fig. 4B. Phylogenetic tree analysis of C-type lectin-like proteins from *Lu. longipalpis* and other organisms is shown in Fig. 4C. C-type lectin-like salivary proteins have only been described in *Lu. longipalpis* sand flies.

Cluster 10 (LJL34; GenBank acc. no. AF132511) – antigen 5-related proteins

This cluster codes for a secreted protein of 29 kDa similar to antigen 5-related protein from vespid venom (Lu et al., 1993). Similar proteins have been isolated from the salivary glands of *Ae. aegypti* (Valenzuela et al., 2002a) and *An.*

gambiae (Francischetti et al., 2002). We previously isolated this cDNA from the salivary glands of *Lu. longipalpis* by PCR subtraction (Charlab et al., 1999).

Cluster 12 (LJL13; GenBank acc. no. AF420274) – D7-related protein

This cluster codes for a secreted protein of 26 kDa and belongs to the D7 family of proteins (Valenzuela et al., 2002c). One salivary D7 protein from *An. stephensi* was shown to be an inhibitor of the blood coagulation Factor XII (Isawa et al., 2002). The N-terminus of the predicted salivary protein, WQDVRNADQTL, was found in the proteomic analysis (Fig. 1).

Cluster 13 (LJL23; GenBank acc. no. AF131933) – apyrase

This cluster codes for a secreted protein of 35 kDa previously characterized by PCR subtraction (Charlab et al., 1999). This protein belongs to a family first described in *Cimex lectularius* and shown to be a salivary apyrase (Valenzuela et al., 1998). Later, the clone coding for an homologous cDNA from *P. papatasi* was expressed and demonstrated to be a calcium-dependent apyrase (Valenzuela et al., 2001b). Recently, a Cimex homologue from humans was demonstrated to have ATPase and low ADPase activity (Murphy et al., 2003), while the insect apyrases have high ADPase activities. The N-terminus of the predicted salivary protein, APPGVEWYHFGGL, was found in the proteomic analysis (Fig. 1). A protein with similar molecular mass to the cDNA prediction for this cluster (35 kDa) was recognized with high frequency among serum of individuals exposed naturally to sand flies who developed a cellular immunity to *Le. chagasi* (Gomes et al., 2002). Therefore, this protein may be a good marker for vector exposure and potential vaccine candidate to control *Le. chagasi* infection.

Clusters 16, 21, 23, 58 – novel sequences

Cluster 16 (LJL143; GenBank acc. no. AY445936). This cluster codes for a secreted protein of 32 kDa. It has no significant homology to known proteins from GenBank. This protein may only be present in phlebotomines, because we did not find similar sequences in either the *Drosophila* or *Anopheles* genome or any other databases.

Cluster 21 (LJM114; GenBank acc. no. AY455907). This cluster codes for a secreted protein of 14 kDa, with no similarities to other proteins in the databank. It is a relatively abundant cDNA in this library and is therefore a novel protein in sand fly saliva.

Cluster 23 (LJM78; GenBank acc. no. AY455908). This cluster codes for a secreted protein of 37 kDa, with no similarities to other proteins on the databases searched. This cDNA represents a novel protein from the saliva of a sand fly.

Cluster 58 (LJS03; GenBank acc. no. AY455914). This cDNA codes for a secreted protein of 15 kDa, with no significant similarities to other proteins in the searched databases.

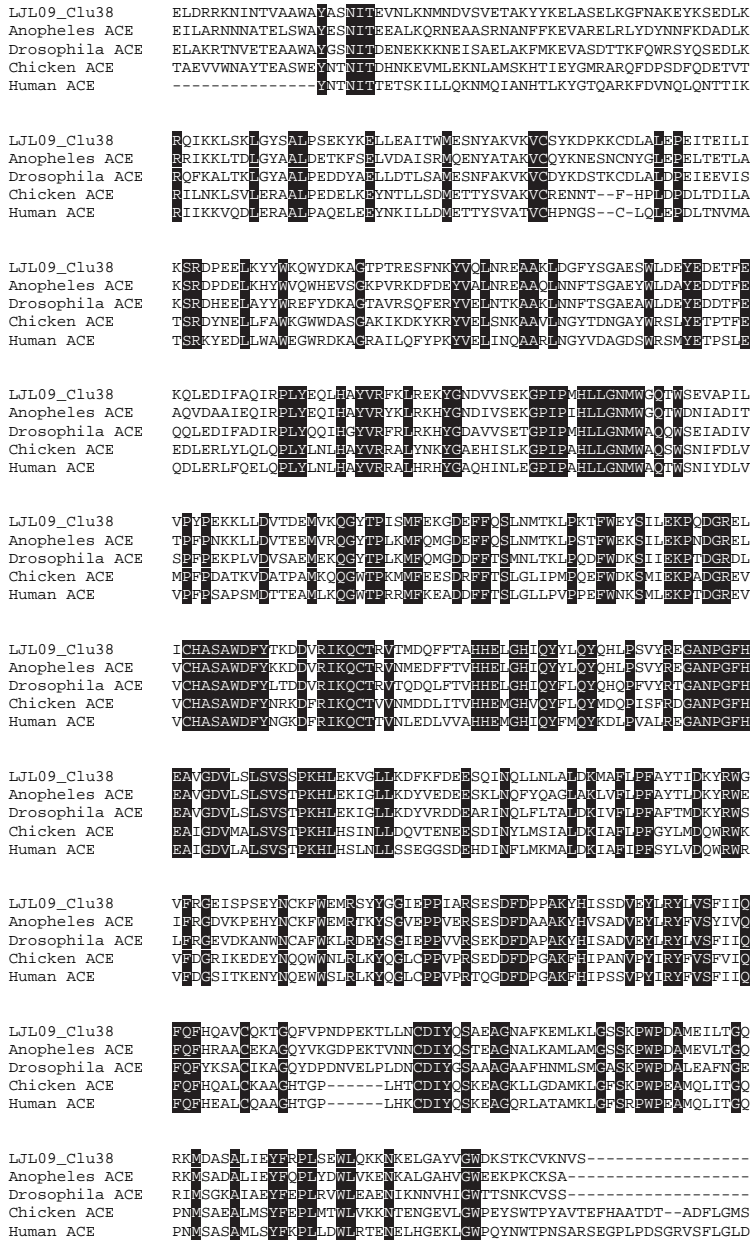


Fig. 6. ClustalW alignment of angiotensin converting enzyme-related proteins from different organisms, including *Lutzomyia longipalpis*, *Anopheles gambiae*, *Drosophila melanogaster*, chicken and human. Reverse-shaded amino acids represent identical amino acids among all these proteins.

bind to the extracellular matrix proteins of the vertebrate host.

Cluster 27 (LJL11; GenBank acc. no. AF132510) – 5'-nucleotidase

This cluster codes for a secreted protein of 61 kDa. It is the secreted 5'-nucleotidase from *Lu. longipalpis*, previously isolated by PCR subtraction (Charlab et al., 1999). The N-terminus of the predicted salivary protein, EDGSYEIHLHTN, was found in the proteomic analysis (Fig. 1), suggesting that this protein is very abundant in the saliva of this insect.

Cluster 38 (LJL09; GenBank acc. no. AY455911) – angiotensin converting enzyme

This cDNA codes for a secreted protein of 71 kDa, with high similarities to angiotensin converting enzyme (ACE) from *An. gambiae*, *D. melanogaster*, chicken and human (Fig. 6). The function as a peptidase remains to be elucidated. The activity may be similar to the kininase activity observed in the tick *I. scapularis* (Ribeiro and Mather, 1998).

Cluster 56 (LJM26; GenBank acc. no. AY455913) – serine protease inhibitor

This cDNA codes for a secreted protein of 49 kDa, with high similarities to the serpin family of protease inhibitors (Fig. 7). This protein may function as an anticoagulant in the saliva or may be responsible for regulation of the insect immune response. Anticomplement activity in *Lu. longipalpis* has been recently described (Cavalcante et al., 2003); because of its putative anti-protease activity, this cDNA may code for the salivary anti-complement activity.

Cluster 20 (LJL04; GenBank acc. no. AY455906) – collagen binding-like proteins

This cluster codes for a secreted protein of 29 kDa, with similarities to PpSP32 salivary protein from *P. papatasi* and to the collagen adhesion protein from *Bacillus cereus* (acc. no. NP_830673). This 29 kDa protein is rich in glycine (51 amino acids), arginine (25 amino acids), proline (25 amino acids) and lysine (23 amino acids). These four amino acids represent 44% of the amino acids in this protein. Additionally, the pI of this protein is 10.2, making it a very basic protein. Different short repeats were identified in this protein, particularly the repeats GQG and GTRP (Fig. 5). Because of its similarities to the collagen-binding protein from *B. cereus*, the high pI, the richness in glycine and the repeated amino acid motifs, this salivary protein may

Cluster 71 (LJL138; GenBank acc. no. AY455916) – endonuclease

This cDNA codes for a protein of 44 kDa, with high similarities to non-specific RNA/DNA endonucleases from different organisms. A similar cDNA was isolated from the salivary glands of the tsetse fly (*Glossina morsitans*) and *Culex quinquefasciatus* (Ribeiro et al., 2004) and the protein was named TsAI (Li et al., 2001). The function as an endonuclease remains to be elucidated. The presence of endonuclease in the saliva of an insect is puzzling and it is interesting to note that enzymes directed to nucleotide metabolism, such as 5'-nucleotidase and adenosine deaminase, are present in this sand fly (Charlab et al., 2000).

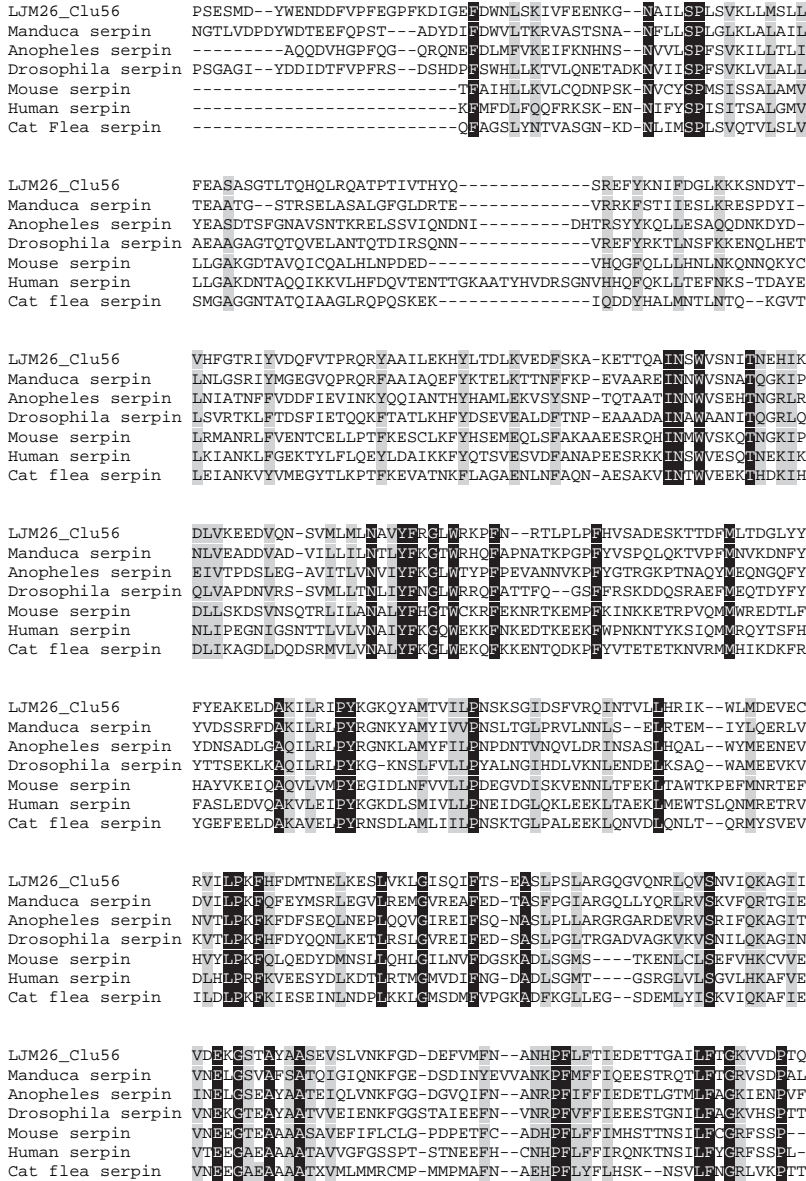


Fig. 7. ClustalW alignment of serpin-related proteins from different organisms, including *Lutzomyia longipalpis*, *Manduca sexta*, *Anopheles gambiae*, *Drosophila melanogaster*, cat flea, mouse and human. Reverse-shaded amino acids represent identical amino acids among all these proteins. Gray-shaded amino acids represent highly conserved amino acids.

glands of the sand fly *Lutzomyia longipalpis*. When comparing this PCR-based cDNA library construction/massive sequencing approach to the PCR subtraction technology we have used before (Charlab et al., 1999), we have tripled the finding of cDNA coding for secreted proteins. Additionally, the isolated cDNA are, in the majority, full-length cDNA, making the process of isolation and sequencing less laborious than in the PCR subtraction method, which cuts the full-length cDNA into smaller pieces.

It is interesting to note that we needed to sequence only 550 cDNA to identify the majority of secretory salivary proteins for this sand fly. The most abundant transcripts in this *Lu. longipalpis* salivary gland cDNA library correspond to putative secreted proteins, which indicates that most transcripts in this organ are directed to secretion. From these 550 cDNA, we obtained 143 different clusters of related sequences, suggesting that the cDNA library was diverse enough to have a good representation of the transcripts present in the salivary gland of this sand fly.

From the 143 different clusters or families of proteins, we identified 35 clusters of proteins containing a secretory signal peptide. Interestingly, the majority of the cDNAs sequenced (451 out of 550) were in these 35 clusters. We confirmed by Edman degradation the presence of 17 secreted proteins in the salivary gland of this sand fly. These 17 proteins are included in these 35 clusters. In fact, all the proteins that resulted with N-terminus data had a corresponding cDNA in these 35 clusters, and the predicted signal cleavage site (SignalP program for these 35 clusters) matched perfectly with the N-terminus obtained, with the exception of two proteins. These two proteins probably underwent further processing by a protease.

When comparing these 35 clusters with sequences deposited in existing databases, we found that nine of these cDNAs correspond with already described *Lu. longipalpis* proteins. We found eight cDNAs with low homology to already described *Lu. longipalpis* proteins, suggesting that they belong to related families of proteins but are different enough to be clustered in different groups. We found 10 cDNAs with homology to other proteins in databases, including an angiotensin converting enzyme, a protease and a palmitoyl-hydrolase; the rest of the cDNAs (seven) did not match to any

Cluster 97 (LJS138; GenBank acc. no. AY455917) – translocon-associated protein

This cDNA codes for a secreted protein of 16 kDa, with high similarities to translocon-associated protein from different organisms and may be associated with an ER protein export function.

Cluster 113 (LJS193; GenBank acc. no. AY455918) – palmitoyl thioesterase

This cDNA codes for a protein of 32 kDa, with high similarities to palmitoyl-(protein) thioesterase from different organisms.

Discussion

In this work, we have isolated and characterized the most abundant secretory proteins and transcripts from the salivary

protein in the existing databases. These are probably novel proteins only present thus far in this sand fly.

The sand fly *Lu. longipalpis* is the main vector of *Le. chagasi*, the causal agent of visceral leishmaniasis. The relationship between *Lu. longipalpis* saliva and human visceral leishmaniasis was recently undertaken. Barral et al. (2000) studied serum from children living in an endemic area of visceral leishmaniasis and found a positive correlation between children producing antibodies against *Lu. longipalpis* and delayed-type hypersensitive response (DTH) to *Le. chagasi*. DTH to *Leishmania* is a marker for protection against this parasite, while positive serology is a sign of poor prognosis. On the other hand, no correlation was found between children producing antibodies to *Lu. longipalpis* and positive serology to *Le. chagasi*. In an extension of this work, we recently reported that children from an endemic area of leishmaniasis, who in a 6-month period developed antibodies to *Le. chagasi*, did not produce detectable antibodies to *Lu. longipalpis* salivary proteins. On the other hand, individuals that developed a cellular response to *Le. chagasi* (DTH) produced IgG, IgG1 and IgE antibodies to *Lu. longipalpis* salivary proteins (Gomes et al., 2002). These data support the hypothesis that induction of an immune response to salivary proteins from *Lu. longipalpis* may facilitate a protective immune response to *Le. chagasi*.

The salivary proteins, as well as the cDNA coding for secreted proteins identified in this work, are good candidates as markers for vector exposure to continue performing epidemiological studies with single recombinant proteins instead of whole SGH. In previous work, we identified five salivary antigens by western blot that resulted in molecular masses of 45, 44, 43, 35 and 17 kDa (Gomes et al., 2002). In the present study, we have identified a limited number of cDNAs that code for a molecular mass similar to these antigens; these are the yellow-related proteins (clusters 2, 9 and 22), an apyrase (cluster 13) and an anticoagulant protein (clusters 8, 14, 19). Recombinant expression of these proteins will determine whether these are the antigens recognized by individuals exposed to sand flies that showed protection to *Le. chagasi* infection.

Ultimately, these salivary proteins represent good vaccine candidates to control *Leishmania* infection. We reported that a delayed-type hypersensitivity response to a sand fly salivary protein was responsible for the protective effect against *Le. major* infection in mice (Valenzuela et al., 2001a). Because this type of protection is T cell dependent and the response may depend on the immunogenetic background of the host, focusing on a single protein may not be a proper vaccine strategy in an outbred population. This high-throughput approach is providing a larger candidate repertoire to select the best vaccine candidate or the best cocktail of protective salivary components.

The authors thank Dr José M. C. Ribeiro for help in bioinformatic analysis and critical review and comments on this manuscript, and Nancy Shulman for editorial assistance.

References

- Altschul, S. F. and Lipman, D. J. (1990). Protein database searches for multiple alignments. *Proc. Natl. Acad. Sci. USA* **87**, 5509-5513.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.
- Barral, A., Honda, E., Caldas, A., Costa, J., Vinhas, V., Rowton, E. D., Valenzuela, J. G., Charlab, R., Barral-Netto, M. and Ribeiro, J. M. (2000). Human immune response to sand fly salivary gland antigens: a useful epidemiological marker? *Am. J. Trop. Med. Hyg.* **62**, 740-745.
- Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L. and Sonnhammer, E. L. (2000). The Pfam protein families database. *Nucleic Acids Res.* **28**, 263-266.
- Bell, J. F., Stewart, S. J. and Wikel, S. K. (1979). Resistance to tick-borne *Francisella tularensis* by tick-sensitized rabbits: allergic klendusity. *Am. J. Trop. Med. Hyg.* **28**, 876-880.
- Calvo, E., Andersen, J., Francischetti, I. M., del Capurro, M., deBianchi, A. G., James, A. A., Ribeiro, J. M. and Marinotti, O. (2004). The transcriptome of adult female *Anopheles darlingi* salivary glands. *Insect Mol. Biol.* **13**, 73-88.
- Cavalcante, R. R., Pereira, M. H. and Gontijo, N. F. (2003). Anti-complement activity in the saliva of phlebotomine sand flies and other haematophagous insects. *Parasitology* **127**, 87-93.
- Charlab, R., Valenzuela, J. G., Rowton, E. D. and Ribeiro, J. M. (1999). Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly *Lutzomyia longipalpis*. *Proc. Natl. Acad. Sci. USA* **96**, 15155-15160.
- Charlab, R., Rowton, E. D. and Ribeiro, J. M. (2000). The salivary adenosine deaminase from the sand fly *Lutzomyia longipalpis*. *Exp. Parasitol.* **95**, 45-53.
- Francischetti, I. M., Valenzuela, J. G., Pham, V. M., Garfield, M. K. and Ribeiro, J. M. (2002). Toward a catalog of the transcripts and proteins (sialome) from the salivary gland of the malaria vector *Anopheles gambiae*. *J. Exp. Biol.* **205**, 2429-2451.
- Geyer, P. K., Spana, C. and Corces, V. G. (1986). On the molecular mechanism of gypsy-induced mutations at the yellow locus of *Drosophila melanogaster*. *EMBO J.* **5**, 2657-2662.
- Gillespie, R. D., Mbow, M. L. and Titus, R. G. (2000). The immunomodulatory factors of blood-feeding arthropod saliva. *Parasite Immunol.* **22**, 319-331.
- Gomes, R. B., Brodskyn, C., de Oliveira, C. I., Costa, J., Miranda, J. C., Caldas, A., Valenzuela, J. G., Barral-Netto, M. and Barral, A. (2002). Seroconversion against *Lutzomyia longipalpis* saliva concurrent with the development of anti-*Leishmania chagasi* delayed-type hypersensitivity. *J. Infect. Dis.* **186**, 1530-1534.
- Isawa, H., Yuda, M., Orito, Y. and Chinzei, Y. (2002). A mosquito salivary protein inhibits activation of the plasma contact system by binding to factor XII and high molecular weight kininogen. *J. Biol. Chem.* **277**, 27651-27658.
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. and Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* **23**, 403-405.
- Johnson, J. K., Li, J. and Christensen, B. M. (2001). Cloning and characterization of a dopachrome conversion enzyme from the yellow fever mosquito, *Aedes aegypti*. *Insect Biochem. Mol. Biol.* **31**, 1125-1135.
- Jones, L. D. and Nuttall, P. A. (1990). The effect of host resistance to tick infestation on the transmission of Thogoto virus by ticks. *J. Gen. Virol.* **71**, 1039-1043.
- Kamhawi, S. (2000). The biological and immunomodulatory properties of sand fly saliva and its role in the establishment of Leishmania infections. *Microbes Infect.* **2**, 1765-1773.
- Kamhawi, S., Belkaid, Y., Modi, G., Rowton, E. and Sacks, D. (2000). Protection against cutaneous leishmaniasis resulting from bites of uninfected sand flies. *Science* **290**, 1351-1354.
- Lanzaro, G. C., Lopes, A. H., Ribeiro, J. M., Shoemaker, C. B., Warburg, A., Soares, M. and Titus, R. G. (1999). Variation in the salivary peptide, maxadilan, from species in the *Lutzomyia longipalpis* complex. *Insect Mol. Biol.* **8**, 267-275.
- Lerner, E. A., Ribeiro, J. M., Nelson, R. J. and Lerner, M. R. (1991). Isolation of maxadilan, a potent vasodilatory peptide from the salivary glands of the sand fly *Lutzomyia longipalpis*. *J. Biol. Chem.* **266**, 11234-11236.
- Li, S., Kwon, J. and Aksoy, S. (2001). Characterization of genes expressed

- in the salivary glands of the tsetse fly, *Glossina morsitans morsitans*. *Insect Mol. Biol.* **10**, 69-76.
- Lu, G., Villalba, M., Coscia, M. R., Hoffman, D. R. and King, T. P.** (1993). Sequence analysis and antigenic cross-reactivity of a venom allergen, antigen 5, from hornets, wasps, and yellow jackets. *J. Immunol.* **150**, 2823-2830.
- Morris, R. V., Shoemaker, C. B., David, J. R., Lanzaro, G. C. and Titus, R. G.** (2001). Sandfly maxadilan exacerbates infection with *Leishmania major* and vaccinating against it protects against *L. major* infection. *J. Immunol.* **167**, 5226-5230.
- Murphy, D. M., Ivanenkov, V. V. and Kirley, T. L.** (2003). Bacterial expression and characterization of a novel, soluble, calcium-binding, and calcium-activated human nucleotidase. *Biochemistry* **42**, 2412-2421.
- Nazario, S., Das, S., de Silva, A. M., Deponte, K., Marcantonio, N., Anderson, J. F., Fish, D., Fikrig, E. and Kantor, F. S.** (1998). Prevention of *Borrelia burgdorferi* transmission in guinea pigs by tick immunity. *Am. J. Trop. Med. Hyg.* **58**, 780-785.
- Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G.** (1997). A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int. J. Neural Syst.* **8**, 581-599.
- Nuttall, P. A., Paesen, G. C., Lawrie, C. H. and Wang, H.** (2000). Vector-host interactions in disease transmission. *J. Mol. Microbiol. Biotechnol.* **2**, 381-386.
- Ribeiro, J. M.** (1987). Vector salivation and parasite transmission. *Mem. Inst. Oswaldo Cruz* **82**(Suppl. 3), 1-3.
- Ribeiro, J. M.** (1989). Vector saliva and its role in parasite transmission. *Exp. Parasitol.* **69**, 104-106.
- Ribeiro, J. M.** (1995). Blood-feeding arthropods: live syringes or invertebrate pharmacologists? *Infect. Agents Dis.* **4**, 143-152.
- Ribeiro, J. M. and Mather, T. N.** (1998). *Ixodes scapularis*: salivary kininase activity is a metallo dipeptidyl carboxypeptidase. *Exp. Parasitol.* **89**, 213-221.
- Ribeiro, J. M., Vachereau, A., Modi, G. B. and Tesh, R. B.** (1989). A novel vasodilatory peptide from the salivary glands of the sand fly *Lutzomyia longipalpis*. *Science* **243**, 212-214.
- Ribeiro, J. M., Andersen, J., Silva-Neto, M. A., Pham, V. M., Garfield, M. K. and Valenzuela, J. G.** (2004). Exploring the sialome of the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol.* **34**, 61-79.
- Schultz, J., Milpetz, F., Bork, P. and Ponting, C. P.** (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. USA* **95**, 5857-5864.
- Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. and Bork, P.** (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**, 231-234.
- Soares, M. B., Titus, R. G., Shoemaker, C. B., David, J. R. and Bozza, M.** (1998). The vasoactive peptide maxadilan from sand fly saliva inhibits TNF-alpha and induces IL-6 by mouse macrophages through interaction with the pituitary adenylate cyclase-activating polypeptide (PACAP) receptor. *J. Immunol.* **160**, 1811-1816.
- Titus, R. G. and Ribeiro, J. M.** (1988). Salivary gland lysates from the sand fly *Lutzomyia longipalpis* enhance *Leishmania* infectivity. *Science* **239**, 1306-1308.
- Valenzuela, J. G., Charlab, R., Galperin, M. Y. and Ribeiro, J. M.** (1998). Purification, cloning, and expression of an apyrase from the bed bug *Cimex lectularius*. A new type of nucleotide-binding enzyme. *J. Biol. Chem.* **273**, 30583-30590.
- Valenzuela, J. G., Belkaid, Y., Garfield, M. K., Mendez, S., Kamhawi, S., Rowton, E. D., Sacks, D. L. and Ribeiro, J. M.** (2001a). Toward a defined anti-*Leishmania* vaccine targeting vector antigens: characterization of a protective salivary protein. *J. Exp. Med.* **194**, 331-342.
- Valenzuela, J. G., Belkaid, Y., Rowton, E. and Ribeiro, J. M.** (2001b). The salivary apyrase of the blood-sucking sand fly *Phlebotomus papatasi* belongs to the novel Cimex family of apyrases. *J. Exp. Biol.* **204**, 229-237.
- Valenzuela, J. G., Pham, V. M., Garfield, M. K., Francischetti, I. M. and Ribeiro, J. M.** (2002a). Toward a description of the sialome of the adult female mosquito *Aedes aegypti*. *Insect Biochem. Mol. Biol.* **32**, 1101-1122.
- Valenzuela, J. G., Francischetti, I. M., Pham, V. M., Garfield, M. K., Mather, T. N. and Ribeiro, J. M.** (2002b). Exploring the sialome of the tick *Ixodes scapularis*. *J. Exp. Biol.* **205**, 2843-2864.
- Valenzuela, J. G., Charlab, R., Gonzalez, E. C., de Miranda-Santos, I. K., Marinotti, O., Francischetti, I. M. and Ribeiro, J. M.** (2002c). The D7 family of salivary proteins in blood-sucking Diptera. *Insect Mol. Biol.* **11**, 149-155.
- Valenzuela, J. G., Francischetti, I. M., Pham, V. M., Garfield, M. K. and Ribeiro, J. M.** (2003). Exploring the salivary gland transcriptome and proteome of the *Anopheles stephensi* mosquito. *Insect Biochem. Mol. Biol.* **33**, 717-732.
- Wikel, S. K., Ramachandra, R. N., Bergman, D. K., Burkot, T. R. and Piesman, J.** (1997). Infestation with pathogen-free nymphs of the tick *Ixodes scapularis* induces host resistance to transmission of *Borrelia burgdorferi* by ticks. *Infect. Immun.* **65**, 335-338.