Published in final edited form as:

Gene. 2005 August 1; 355: 11–27. doi:10.1016/j.gene.2005.05.002.

A catalog for the transcripts from the venomous structures of the caterpillar *Lonomia obliqua*: identification of the proteins potentially involved in the coagulation disorder and hemorrhagic syndrome

Ana B. G. Veiga^{a,b}, José M. C. Ribeiro^a, Jorge A. Guimarães^b, and Ivo M.B. Francischetti^{a,*}

^aVector Biology Section, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Disease, National Institutes of Health, 12735 Twinbrook Parkway, Twinbrook III, Room 2E-28, Rockville, MD 20852 USA

^bCenter of Biotechnology, UFRGS, Av. Bento Gonçalves, 9500, 43421, CP 15005, Porto Alegre-RS, Brazil

Abstract

Accidents with the caterpillar *Lonomia obliqua* are often associated with a coagulation disorder and hemorrhagic syndrome in humans. In the present study, we have constructed cDNA libraries from two venomous structures of the caterpillar, namely the tegument and the bristle. High-throughput sequencing and bioinformatics analyses were performed in parallel. Over one thousand cDNAs were obtained and clustered to produce a database of 538 contigs and singletons (clusters) for the tegument library and 368 for the bristle library. We have thus identified dozens of full-length cDNAs coding for proteins with sequence homology to snake venom prothrombin activator, trypsin-like enzymes, blood coagulation factors and prophenoloxidase cascade activators. We also report cDNA coding for cysteine proteases, Group III phospholipase A2, C-type lectins, lipocalins, in addition to protease inhibitors including serpins, Kazal-type inhibitors, cystatins and trypsin inhibitor-like molecules. Antibacterial proteins and housekeeping genes are also described. A significant number of sequences were devoid of database matches, suggesting that their biologic function remains to be defined. We also report the N-terminus of the most abundant proteins present in the bristle, tegument, hemolymph, and "cryosecretion". Thus, we have created a catalog that contains the predicted molecular weight, isoelectric point, accession number, and putative function for each selected molecule from the venomous structures of *L. obliqua*. The role of these molecules in the coagulation disorder and hemorrhagic syndrome caused by envenomation with this caterpillar is discussed. All sequence information and the Supplemental Data, including Figures and Tables with hyperlinks to FASTA-formatted files for each contig and the best match to the Databases, are available at <http://www.ncbi.nih.gov/projects/omes>.

Keywords

caterpillar venom; coagulation; fibrinolysis; hemorrhage; toxins

*Corresponding author. Tel.: +1-301-402-2748; fax: +1-301-480-2571 ifrancischetti@niaid.nih.gov.

1. Introduction

Many toxins that affect hemostasis are produced by venomous animals such as snakes and spiders, which use these compounds to facilitate capture and digestion of prey (Markland, 1997; Aird, 2002). Some lepidopteran larvae, notably those of the family Saturniidae, also produce venoms that adversely impact hemostasis. Unlike snake and spider venoms, caterpillar venoms are used solely for defense against predators, which are envenomed by touching the bristles of the caterpillars (Kelen et al., 1995).

Lonomia sp. caterpillars are known for causing a hemorrhagic syndrome characterized by ecchymoses, hematuria, bleeding from scars and mucous membranes, intracerebral bleeding and acute renal failure (Arocha-Piñango and Guerrero, 2001). In southern Brazil, accidents caused by *L. obliqua* caterpillars are reportedly increasing (Kelen et al., 1995). Accidents usually occur when the victim, leaning against tree trunks containing dozens or hundreds of caterpillars, comes into contact with the bristles of the caterpillars. Often, the whole animal is smashed in the accident. In the latter case, the insect's cuticle is broken and many secretions, including hemolymph, penetrate the human skin and enter the circulation (Veiga et al., 2001). While some toxic principles are found in bristle extract, others are present in the hemolymph of *L. obliqua* (Donato et al., 1998; Veiga et al., 2003). Some of the active principles produced by *Lonomia* sp. that interfere with the hemostatic system have been characterized: fibrinolytic proteases in the hemolymph of *L. achelous* (Amarant et al., 1991), prothrombin or factor X activators in *L. obliqua* bristle (Donato et al., 1998) and α -fibrinogenases found in both bristles and in a secretion obtained after freezing *L. obliqua* caterpillars (Veiga et al., 2003; Pinto et al., 2004). Other active compounds such as phospholipases were also described in *Lonomia* sp. (for a review, see Arocha-Piñango and Guerrero, 2001).

Remarkably, structural information on *Lonomia* venom is almost nonexistent. In fact, only partial amino acid sequences of two fibrinolytic proteases from *L. achelous* have been reported before (Amarant et al., 1991). Furthermore, a GenBank search for "*Lonomia*", in November, 2004, revealed only the amino acid sequences of these two fibrinogenases and the sequence of polyhedrin, a multiple nucleopolyhedrovirus from *L. obliqua*. The need for information on the molecular constituents of *Lonomia* venom led us to create cDNA libraries from the tegument and bristles of *L. obliqua* followed by high-throughput sequencing and bioinformatics analysis. In addition, Edman degradation of the most abundant protein has been performed in parallel. This approach allowed us to generate a comprehensive catalog of *L. obliqua* transcripts (cDNAs) and proteins. The roles of *L. obliqua* molecular components probably involved in the coagulation disorder and in the hemorrhagic syndrome are also discussed.

2. Materials and methods

2.1. Reagents

All water used was of 18-M Ω quality and was produced using a MilliQ apparatus (Millipore, Bedford, MA, USA). Organic compounds were obtained from Sigma Chemical Corporation (St. Louis, MO, USA) or as stated otherwise.

2.2. Caterpillars and venomous samples

L. obliqua caterpillars were provided by the Health Department of the city of Videira (Santa Catarina, Brazil) after being collected by local inhabitants directly from trees. Bristles were collected from caterpillars by excision at the base of the scoli. Tegument, which included bristles and surrounding tissues, was collected after total dissection of the animals. Both tissues were homogenized in 500 μ l of water for preparation of the respective extracts.

Hemolymph was collected with a syringe through an incision at the pseudopodia of the caterpillar. Cryosecretion was obtained as described in Pinto et al. (2004). Briefly, caterpillars were placed over Petri dishes and kept at -20°C for 24 h. Frozen caterpillars were then washed with 20 mM HEPES (pH 7.4), and the released protein-rich fluid was collected from the dish. All venomous preparations were then centrifuged to remove debris ($12,000 \times g$ for 20 min) and stored at -80°C until use.

2.3. SDS-PAGE

Analysis of the protein components of each *L. obliqua* venomous preparations (bristle, tegument, hemolymph, and cryosecretion) was performed as in Francischetti et al. (2004) and is described in detail in the Supplemental Data (Section 2.7).

2.4. Construction and sequencing of *L. obliqua* cDNA libraries

Two cDNA libraries were constructed: one using mRNA isolated from *L. obliqua* bristles and the other from *L. obliqua* tegument. mRNA was obtained using a Micro-Fast Track mRNA isolation kit (Invitrogen, San Diego, CA, USA) according to manufacturer's instructions. The PCR-based cDNA libraries were constructed using the SMART cDNA library construction kit (Clontech, Palo Alto, CA, USA). Cycle-sequencing reactions used the DTCS labeling kit (Beckman Coulter Inc., Fullerton, CA, USA); nucleotide sequencing was performed in a CEQ 2000 DNA instrument (Beckman Coulter Inc.) as described in Francischetti et al. (2004) and in the Supplemental Data (Sections 2.8 and 2.9).

2.5. Bioinformatics analyses and full-length sequencing of selected cDNA clones

cDNA sequences analyses and obtaining of full-length sequences were performed as described in detail in Francischetti et al. (2004) and in the Supplemental Data (Sections 2.10 – 2.13). Electronic version of the complete tables (Microsoft Excel format) with hyperlinks to web-based databases and to BLAST results is available at <http://www.ncbi.nih.gov/projects/omes>.

2.6. Statistical tests

Statistical tests were performed with SigmaStat version 2.0 (Jandel Software, San Rafael, CA, USA). Kruskal-Wallis ANOVA on ranks was performed, and multiple comparisons were done by the Dunn method. Dual comparisons were made with the Mann-Whitney rank sum test.

3. Results and Discussion

Studies on the venom of *Lonomia* sp. caterpillars have so far characterized two fibrinolytic proteins from *L. achelous*, named Achelase I and Achelase II (Amarant et al., 1991). The present study attempts to understand the molecular components likely involved in the envenomation by *L. obliqua*. Accordingly, we have performed SDS-PAGE and constructed cDNA libraries using tissues potentially involved in the envenomation process.

3.1. SDS-PAGE of *L. obliqua* bristle and tegument extracts, hemolymph, and cryosecretion

Because accidents with *L. obliqua* are usually accompanied by smashing of the whole animal, we investigated four types of venomous preparations in the present study: bristle and tegument extracts, hemolymph, and cryosecretion. Figure 1 (left) shows the pattern of protein separation from these preparations applied side-by-side in the same gel followed by staining with Coomassie blue. To identify these proteins, they were transferred to PVDF membrane, and the bands were cut from the membrane and submitted to Edman degradation. Amino-terminal information was successfully obtained for 25 bands.

Sequences were compared with the non-redundant (NR) GenBank database and with the cDNAs obtained in the mass sequencing project of *L. obliqua* (see Section 2). Results are presented in Figure 1 (right). We found many sequences displaying matches to cDNAs from both libraries, coding for diverse proteins such as trypsin, serpin, lipocalin, transferrin, lectin, and others.

Band 19-LOB is the most prominent band in the bristle extract and other samples, and was identified as a lipocalin. A trypsin-like protein (11-LOB) and a lectin (1-LOB) were also identified in the bristle preparation. Other bands that were successfully sequenced in the tegument extract have similarities primarily to proteins involved in cell metabolism or cell structure (housekeeping function). On the other hand, the hemolymph contains a serpin (8-LOH) and a protease inhibitor (23-LOH), among other proteins. The cryosecretion also contains a serpin (8-LOC) and other housekeeping proteins. Finally, in some cases, the N-terminal sequence from a given band matches more than one cDNA; for example, 9-LOB can be assigned to cDNAs coding for hemolin and laminin in the bristle library, while band 16-LOB can be assigned to cDNAs coding for lectin and for a ribosomal protein in the tegument library.

3.2. cDNA libraries of *L. obliqua*

Two different cDNA libraries were constructed: one from a section of the tegument of *L. obliqua*, which includes the bristle, surrounding tissues, and chitinous processes, and a second from homogenized bristle only. A total of 1,152 independent clones from the tegument library and 960 independent clones from the bristle library were 5' sequenced, yielding 938 and 730 sequences, respectively, of good quality for analysis. Next, these sequences were followed by bioinformatics analysis as described in Francischetti et al. (2004) and in the Supplemental Data, and included: *i*) clustering at high stringency levels (98% identity over 100 nucleotides), *ii*) BLAST search against the non-redundant and protein motifs databases, *iii*) comparison with the PFAM, SMART and Kog databases using RPSBlast and *iv*) submission of the translated sequences to the Signal P server (Nielsen et al., 1997) (see Section 2). This initial approach allowed us to obtain a fingerprint of the protein families or "clusters" present in these tissues. Several sequences were then selected based on novelty or the protein family it assigns for and extension of their corresponding cDNAs were performed until the poly A was reached.

3.2.1. cDNA library of the tegument—In the cDNA library of *L. obliqua* tegument, 592 sequences (63.1%) showed database hits and were organized into 278 clusters, while the other 346 sequences showed no similarity to any known sequences and formed 260 clusters (complete supplemental table available for download as Excel spreadsheet). As far as the absolute number of sequences is concerned, we found that 16.1% have a potential signal peptide, 62.6% code for housekeeping proteins, and 21.3% are indeterminate. After clustering the sequences, 12% of the clusters were found to have a potential signal peptide, 58% code for housekeeping proteins, and 30% are indeterminate. Moreover, 477 sequences matched database sequences obtained from a variety of insects, including the lepidopterans *Bombyx mori*, *Manduca sexta*, *Heliothis virescens*, *Antheraea pernyi*, *Galleria mellonella* and others, as well as to the mosquito *Anopheles gambiae* and to the fruit fly *Drosophila* spp..

3.2.2. cDNA library of bristles—In the cDNA library of *L. obliqua* bristles, 408 sequences (55.9%) showed database hits and were organized into 176 clusters, while the other 322 unknown sequences formed 192 clusters (complete supplemental table available for download as Excel spreadsheet). Forty percent of all sequences code for proteins with a potential signal peptide, 43.3% for housekeeping proteins, and 16.7% are indeterminate.

Considering the number of clusters, 17% had a potential signal peptide, 53% code for housekeeping proteins, and 30% were indeterminate. A total of 311 sequences had hits to sequences from such insects as *M. sexta*, *B. mori*, *Samia cynthia*, *Spodoptera frugiperda*, and *Apis mellifera*.

3.2.3. Secretory and housekeeping genes—Comparing these results with those from other studies describing transcriptomes of venomous or blood-sucking animals (Valenzuela et al., 2002; Francischetti et al., 2002; Francischetti et al., 2004), the *L. obliqua* display fewer secreted peptides, especially in the tegument. This can be attributed to the fact that those studies are based on cDNA libraries constructed with mRNAs extracted from venom/salivary glands, which are rich in transcripts and proteins with specific roles in envenomation/blood-feeding. An estimate of the percentage of sequences and clusters of putative toxins for both bristle and tegument is shown in Figure 2, a–d. Concerning the tegument library, the most abundant sequences are those coding for serpins (25.8%), followed by serine proteases (16.1%) and for lipocalin (16.1%) (Fig. 2a). A similar distribution was found when analyzing the percentage of clusters (Fig. 2b).

In the bristle library, over 50% of cDNAs that were sequenced code for a lipocalin, followed by kininogen (16.5%) and serine proteases (14.7%) (Fig. 2c). However, the number of clusters found for serine protease is the most abundant (25%), followed by lectins (15%) and serpins (10%). Such a distribution of transcripts in the bristle library reveals that sequences from the same family (e.g. serine proteases) and found in low copy number may display significant differences in their primary sequence. Therefore, these sequences were independently clustered (Fig. 2d) suggesting functional diversity in the same protein family. On the other hand, lipocalin is the most abundant cDNA in the bristle, but forming only one cluster, indicative of high similarity in their primary sequences. Finally, in contrast to other venoms, no metalloproteases or disintegrins (Fox et al., 2002) were found in our libraries.

Both libraries also contained sequences coding for housekeeping proteins apparently devoid of toxic properties. Accordingly, we report cDNAs coding for proteins involved in transcription and translation (elongation factors, ribosomal proteins), metabolism (juvenile hormone metabolism and transport, NADH-dehydrogenase, cytochrome *c* oxidase, heme oxygenase, ATP synthase, glutathione S-transferase, aldehyde dehydrogenase, glyceraldehyde 3-phosphate dehydrogenase), protein processing (chaperones, proteasome), and other functions. Sequences related to cell structure, such as cuticle proteins, myosin, actin, tubulin and tropomyosin, were found primarily in the cDNA library of tegument.

3.3. Table 1: An estimate of cDNAs coding for putative secretory proteins from bristle and tegument libraries

In an attempt to maximize the identification of putative toxins in *L. obliqua* tissues, we constructed separate cDNA libraries from both bristle and tegument mRNAs. Table 1 presents an overview of sequences coding for secreted proteins that are potentially associated with toxic functions, obtained after the first round of sequence and bioinformatics analysis. The number of sequences found in a given cluster is also available as an approximate estimate of cluster representation and complexity. Matches to the NR and protein motifs databases with their corresponding *e* value are also reported. Table 1 also shows, when available, the corresponding N-terminus obtained by Edman degradation of specific proteins.

In the tegument cDNA library (Table 1), we found 34 sequences forming 21 clusters coding for putative toxins, including protease inhibitors, trypsin-like serine proteases, lectins, esterases, and others. Among the putative secreted proteins, cluster 1, with 36 sequences,

codes for a cuticle protein. Other clusters are less represented and code for proteins with sequence homology to ferritin and attacin, as well as to unknown proteins.

In the bristle cDNA library, 116 sequences may be involved in the envenomation process, and they were grouped in 23 clusters. Among these, the most abundant is cluster 1, coding for a lipocalin with 54 sequences, followed by cluster 8, which codes for a kininogen precursor. Other clusters code for proteins of diverse families, including serine proteases, lectins, serpins, cystatins, trypsin-like inhibitors, and cysteine-rich molecules. Among the putative secreted proteins, we found sequences coding for cecropin, chemosensory protein, and others.

Table 1 also describes clusters that contain sequences where an ORF with or without signal peptide could be identified, but for which there were no database hits. Accordingly, clusters 330, 349, 387, and others in the tegument library assign for proteins of unknown function. Some of these sequences were identified by SDS-PAGE from both the bristle and tegument libraries, demonstrating that these cDNAs code for actual proteins (see Fig. 1 and Section 3.1). The spreadsheet containing the complete list of the sequences coding for proteins with secretory, housekeeping, or indeterminate functions, with or without database hits, can be downloaded at <http://www.ncbi.nlm.nih.gov/projects/omes/> and extracted to a personal computer for viewing. The spreadsheets also include *i*) hyperlinks to the best match of the NR database, *ii*) links to NR matches found for the cluster, *iii*) matches to the conserved domain database, *iv*) FASTA-formatted files for each cluster, and *v*) CAP assembled alignments of each cluster having two or more sequences.

3.4. Table 2: A catalog for full-length *L. obliqua* cDNAs coding for proteins probably involved in envenomation

To maximize the information for the proteins presented in Table 1, selected molecules were re-sequenced and extended to obtain, when applicable and possible, their full-length cDNAs. The full coding sequences were then searched against the NR protein database and SignalP server to confirm, respectively, sequence similarity and the presence of a signal peptide. When a signal peptide was predicted to exist, the molecular weight and the isoelectric point of the mature protein were also calculated and the putative function annotated. The summary of our findings is presented in Table 2, which also includes the accession numbers in the NCBI databases, the length of the sequences (partial or full-length), the tissue source of the transcript (bristle or tegument), and the putative function. A brief discussion of selected sequences from tegument and bristle and their possible participation on the envenomation by *L. obliqua* is presented below.

3.4.1. Serine proteases—Serine proteases can be found in microorganisms, plants and animals, being involved in many physiological processes, and can be classified as trypsin-like, subtilisin-like, and carboxypeptidase-like serine proteinases. Human serine proteinases are involved in hemostasis, digestion, cellular differentiation, and many other processes; those involved in hemostasis are typically trypsin-like proteases. In insects, various serine proteases participate in hemolymph coagulation and prophenoloxidase-activating cascades in response to wounding and infection (Kanost et al., 2004). Both processes involve serine protease activation of zymogens and culminate in formation of the coagulum and phenoloxidase activation, respectively. These enzymes display high sequence and structural similarity and have evolutionarily diverged into two main groups. One group, whose members include prothrombin, FIX, X, FVII, and protein C, characteristically contains a Gla domain, an EGF motif, and eventually a Kringle domain. The other group is represented by prekallikrein, FXI, plasminogen, urokinase, T-PA, FXII, and trypsin. These proteins have no Gla domain but have selected other modules such as Apple, Kringle, EGF, and fibronectin-

like domains (Larson and Katherine, 1998). Notably, our bristle library contains full-length sequences matching both groups of serine proteases, which are briefly discussed below.

LOqua-SP1 (cluster 220, tegument library). This protein is homologous to a serine protease involved in the prophenoloxidase cascade in *M. sexta* (gi|39655053) as well as to human coagulation factors IXa (gi|67596), VII (gi|10518503), II (gi|4503635), and Xa (gi|538554) (Fig. 3). It is also similar to trocarin (gi|54402099), a prothrombin activator from *Tropidechis carinatus* (Joseph et al., 1999). Of interest, unlike the coagulation factors, *LOqua-SP1* lacks a Gla domain and the EGF-1 and EGF-2 motifs but contains a conserved serine proteinase domain. In addition, the molecular mass of mature *LOqua-SP1* (55.2 kDa) is very similar to those of coagulation factors IX (56 kDa), VII (50 kDa), and X (58.8 kDa). A cladogram for *LOqua-SP1* and these serine proteases is shown in Figure 3b, available for download and in the Supplemental Data. Consistent with a pro-hemostatic function for these sequences, previous studies have described a prothrombin activator from bristle extract of *L. obliqua* (Donato et al., 1998). Cloning and expression of *LOqua-SP1* may help to elucidate its substrate specificity and role in envenomation.

LOqua-SP6 (cluster 12, bristle library). This serine protease of ~30 kDa is homologous to serine proteases from the mosquito *A. gambiae* (gi|31238311) and from the flea *Ctenocephalides felis* (gi|4530052). It is also similar to a plasminogen activator from the centipede *Scolopendra subspinipes* (gi|4098568) (Fig. 8, available for download and in the Supplemental Data). Consistent with these sequences, α -fibrinogenase activity was recently reported in *L. obliqua* crude bristle extract (Veiga et al., 2003). This protein, with an apparent molecular mass of 35 kDa, was purified from the cryosecretion (Pinto et al., 2004). *LOqua-SP6* is also similar to a venom proteinase from the bumblebee *Bombus pennsylvanicus* (gi|1079195) and to human plasma kallikrein (gi|4504877). At present, the specificity of this enzyme is unknown, but an α -fibrinogenase (crotalase) displaying kallikrein-like activity has been described in *C. adamanteus* venom (Markland et al., 1982).

3.4.2. Cysteine proteinases—Cysteine proteinases are endopeptidases that rely on the presence of a cysteinyl thiol group for their catalytic activity. Most of the well characterized cysteine proteinases of the papain family, which includes the cathepsins, have the active site motif QGCGSCWAF (Chapman et al., 1997). Sequences in the cDNA libraries of bristles and tegument of *L. obliqua* showed homology to cathepsins.

LOqua-CysPep2 (cluster 55, bristle library). This is a 34.7-kDa cathepsin-like cysteine proteinase containing the motif QGSCGSCWAF and homologous to cathepsins from *Helicoverpa armigera* (gi|7537454), *Aedes aegypti* (gi|5031250), *Triatoma infestans* (gi|38147393), and *Araneus ventricosus* (gi|31872149) (Fig. 9, available for download and in the Supplemental Data). While in some hematophagous arthropods cathepsin-like enzymes have an important role in yolk degradation during embryogenesis (Sappington and Raikhel, 1998), in others, such as *Rhodnius prolixus*, they have a role in digestion of the blood meal (Lopez-Ordóñez et al., 2001). Considering that the present study is on a larval phase of *L. obliqua*, it is not possible to assign an oogenic function for a cathepsin-like protease. On the contrary, these enzymes may be somehow involved in envenomation or in cellular lysosomal function.

3.4.3. Protease inhibitors—Insects and other arthropods are a rich source of protease inhibitors such as serine protease inhibitors (serpins), Kunitz-type, and Kazal-type inhibitors (Kanost et al., 2004). These protease inhibitors found in arthropod hemolymph are likely to function in protecting their hosts from infection by pathogens or parasites and in regulating physiological processes. Some may inhibit fungal or bacterial proteinases, while others probably have roles in regulating endogenous proteinases involved in coagulation,

prophenoloxidase activation, or cytokine activation (Kanost, 1999). Further, it has been shown that some Kazal-type inhibitors found in arthropods play a role in blood-feeding processes by inhibiting thrombin and other coagulation factors (Friedrich et al., 1993; Campos et al., 2004). Here we report the full-length sequences for eight protease inhibitors, including Kazal-type, trypsin inhibitor-like, cysteine protease inhibitors (cystatin), and serpins (Table 2).

3.4.3.1. Serpins: *LOqua-Serp2* (cluster 131, tegument library). This is a 41.6-kDa secreted serpin similar to other members of this family. Alignments are shown in Figure 4 and the phylogenetic tree is shown as Figure 4b, available for download and in the Supplemental Data. *LOqua-Serp2* is homologous to serpin-2 from *B. mori* (gi|7341330) and to serpins from two hematophagous insects, the cat flea *C. felis* (gi|25527494) and the mosquito *A. gambiae* (gi|6759386). Of interest, this serpin also showed homology to bovine plasminogen activator inhibitor (PAI-1, gi|27806497), to antithrombin from zebrafish (AT, gi|33504509) and human (ATIII, gi|113936), and to serpins from the ticks *R. appendiculatus* (gi|17223662) and *Ixodes ricinus* (gi|14140097).

LOqua-Serp3 (cluster 42, tegument library). This serpin showed homology to a coagulation inhibitor from horseshoe crab (gi|31872149), to ATIII (gi|113936), and to PAI-1 (gi|27806497), as well as to a serpin from the tick *Haemaphysalis longicornis* (gi|42662201) (alignments not shown). Consistent with the sequence in cluster 42 of the tegument library, we found by Edman degradation the sequence TETDLQRILRdgNDqFtaKM for bands 8-LOB, 8-LOH, and 8-LOC. The sequence FHYFGHEYDRTALGDAIdKA was also obtained for band 8-LOH and matched the sequence of cluster 247 of the same library, which also codes for a serpin (*LOqua-Serp6*) (Fig. 1).

Considering the participation of serpins in many physiological processes in arthropods, such as hemolymph coagulation and immunity, it is possible that one or more *L. obliqua* serpins may affect human blood coagulation by inhibiting coagulation factors. However, the function and specificity of *L. obliqua* serpins and their participation in envenomation remain to be determined.

3.4.3.2. Other protease inhibitors: *LOqua-Cyst1* (cluster 125, tegument library). This sequence displayed homology to cysteine proteinase inhibitors, mostly of the cystatin family, including a cystatin found in hemocytes of the horseshoe crab *Tachypleus tridentatus* (gi|47115697) and to cystatins from the puff adders *Bitis arietans* (gi|118194) and *B. gabonica* (gi|38570038), as well as to murine kininogen (gi|12963497) (alignments not shown).

LOqua-PI3 (cluster 374, tegument library). This is a secreted trypsin inhibitor-like protein with homology to an immune-induced metalloproteinase inhibitor from the moth *G. mellonella* (gi|34485736) and also similar to an anticoagulant from the venom of the scorpion *Buthus martensii* (gi|14388599) (Fig. 10, available for download and in the Supplemental Data); however, while the *Lonomia* full-length sequence has 398 amino acids, the metalloproteinase inhibitor and the scorpion toxins consist of 170 and 89 residues, respectively. The function of this trypsin inhibitor-like molecule deserves further study.

3.4.4. Phospholipase A2 (PLA2)—Phospholipases are almost universally present in animal venoms. PLA2 promotes hydrolysis of phospholipids with generation of free fatty acids and lysophospholipids, thus producing indirect hemolysis. PLA2 may inhibit blood coagulation by direct interaction with coagulation factors or through degradation of phospholipids involved in coagulation complexes, in addition to antiplatelet effects, neurotoxicity, cardiotoxicity, and myotoxicity (Kini, 2003).

Of note, *L. obliqua* bristle extract induces in vitro hemolysis on rat and human erythrocytes as well as intravascular hemolysis in rats, probably due to a PLA2 (Seibert et al., 2004). Both cDNA libraries contain sequences homologous to PLA2 and other proteins that may have hemolytic activity. Of interest, the full-length sequence of cluster 154 of the bristle library (LOqua-PLA1) did not display homology to any snake Group I and Group II PLA2 known to cause coagulation disturbances and myotoxicity (Lambeau and Lazdunski, 1999). In fact, no Asp-49 or Lys-49 PLA2 were found in our libraries. On the other hand, as depicted in Figure 5, this sequence matched Group III PLA2s from the insects *D. melanogaster* (gi|21627771) and *A. mellifera* (gi|16904372), from the venom of the gila monster *Heloderma suspectum* (gi|104338), and from the scorpions *Anuroctonus phaeodactylus* (gi|46484897) and *Pandinus imperator* (gi|37079101), as well as a Group IX PLA2 toxin from the marine medusa *Rhopilema nomadica* (gi|999133). Furthermore, LOqua-PLA1 also displays homology to the carboxy domain of human Group III PLA2 (gi|7657126) whose function is related to arachidonate release from cells (Murakami et al., 2003). At present, the substrate specificity and function for this putative secreted PLA2 is a matter of speculation, but it may be somehow involved in the hemolysis observed after envenomation (Seibert et al., 2004).

3.4.5. Lipocalins—The lipocalin family is a large group of proteins that exhibit great structural and functional variation. They are typically small extracellular proteins (~20 kDa) sharing common molecular recognition properties: the binding of small, mainly hydrophobic molecules; binding to specific cell-surface receptors; and formation of covalent and non-covalent complexes with other soluble macromolecules (Åkerstrom et al., 2000). In arthropods, the most studied proteins of this family are the lobster colorant crustacyanin and the colorant bilin-binding protein and insecticyanin from butterfly larvae. On the other hand, lipocalins with antihemostatic functions have been isolated from the saliva of blood-sucking arthropods. These proteins are known to inhibit platelet aggregation and reduce blood coagulation in addition to reducing inflammation and promoting vasodilatation (Ribeiro and Francischetti, 2003; Ribeiro et al., 2004).

Both cDNA libraries of *L. obliqua* contain sequences with homology to lipocalins. The tegument library has one cluster formed by five sequences, while the bristle library has three clusters, two of them containing only one sequence and the third, LOqua-Lipcl1, formed by 54 sequences. The function of this lipocalin may be related to its ligand-binding property.

3.4.6. Lectins—Lectins are molecules that usually bind sugars. We found a total of ten sequences with homology to lectins, including c-type lectins, which have a typical carbohydrate recognition domain (Harrison et al., 2003). These sequences formed clusters 120 and 129 in the tegument library, and clusters 33, 140, and 184 in the bristle library (Table 1). We obtained the full-length sequences of all except cluster 129 of the tegument library; however, matching this truncated cluster we found, by Edman degradation, the sequence TVLKDLAK for band 16-LOB. Below is a brief discussion on the major *L. obliqua* lectins.

LOqua-Lect5 (cluster 33, bristle library). Figure 6 shows that this protein has homology to lectins found in snake venoms. Typically, they are 30-kDa proteins consisting of the covalent association of two identical or homologous polypeptides, and they play significant roles in envenomation by binding to coagulation factors and to platelet receptors, as well as through fibrinogen clotting inhibition (Braud et al., 2000). Considering the predicted molecular weight of mature LOqua-Lect5 (16.3 kDa), we suggest that it is a subunit of a dimeric lectin.

LOqua-Lect1 and *LOqua-Lect3* (clusters 120, tegument library, and 140, bristle library, respectively). These two lectins, with molecular masses of 33.6 and 33.9 kDa, respectively, displayed homology to a lipopolysaccharide-binding protein from *B. mori* (gi|4469126), *H. sapiens* chondroitin sulfate proteoglycan 2, also known as versican (gi|21361116), and to a lectin that appears to be involved in the envenomation by the fish *Thalassophryne nattereri* (gi|51863395) (Lopes-Ferreira et al., 1998) (Fig. 11, available for download and in the Supplemental Data).

3.4.7. Antibacterial proteins—Bacterial infection in lepidopteran insects induces production of antibacterial factors in the hemolymph such as attacin and cecropin (Kanost et al., 2004). In the present study, we found three sequences with homology to attacin, one in the tegument library (*LOqua-Def2*, truncated) and the other two forming one cluster in the bristle library (*LOqua-Def3*). Furthermore, we found five sequences in the bristle library forming three clusters with homology to cecropin, one of which had its full-length sequence determined (*LOqua-Def4*). Even though these peptides may have an immunity function in *Lonomia*, the possibility exists that they may also affect vertebrate physiology.

3.5. A catalog toward the full-length *L. obliqua* cDNAs coding for housekeeping protein with various putative functions

Full-length sequences of other proteins, secreted or not, were obtained, such as cuticle proteins, sensory proteins, proteins related to cell metabolism, and ribosomal proteins, among others. Results for these proteins are in Table 3, available for download and in the Supplemental Data.

3.6. *L. obliqua* venomous components: insights for their role in the coagulation disorder and in the hemorrhagic syndrome

The primary symptoms characterizing accidental contact with *Lonomia obliqua* caterpillars are severe hemorrhage and acute renal failure (Kelen et al., 1995). The coagulation disorder presented by these patients appears to be related to the action of many caterpillar toxins acting in a redundant manner on the hemostatic system. Accordingly, molecules that activate factor X and prothrombin (Donato et al., 1998) and likely coded by sequences such as *LOqua-SP1* and *LOqua-SP6*, among others, may play an important role in envenomation. The action of these enzymes leads ultimately to thrombin formation that, among other effects, induces platelet aggregation and fibrin formation. Thrombin also induces compensatory fibrinolysis by a mechanism involving release of tissue plasminogen activator (t-PA) by endothelial cells. Plasmin, thus formed, acts on FXIIIa-crosslinked fibrin and generates D-dimers (Williams, 1998).

D-dimers are found at remarkably high levels in the blood of caterpillar-envenomed patients (Zannin et al., 2003). However, these levels appear to be disproportionate to the activation of the blood coagulation cascade. Actually, the laboratorial profile of these patients resembles "primary fibrinolysis", a term that has been applied to cases of atypical disseminated intravascular coagulation in which the laboratorial and clinical manifestations are dominated by the effects of fibrinolysis (Williams, 1998). These cases, as reported for *L. obliqua* envenomation (Zannin et al., 2003) or after infusion of snake venom α -fibrinogenases such as Ancrod (Bell, 1997; Markland, 1997), are characterized by a marked hypofibrinogenemia and very high titers D-dimers, but normal FII and FX levels, and platelet counts; therefore, a consumption coagulopathy did not take place. Of note, Ancrod induces a thrombin-independent, FXIII proenzyme-dependent formation of crosslinked fibrin oligomers that act as cofactors in the activation of plasminogen by t-PA, with plasmin generation and appearance of D-dimers (Dempfle et al., 2000; Dempfle et al., 2001). In this regard, *L. obliqua* cryosecretion, hemolymph (Pinto et al., 2004) and, to a lesser extent the

bristle extract (Veiga et al., 2003), present marked fibrinogenolytic activity, attributed to an enzyme named lonofibrase, which degrades the α -chain of fibrinogen (Pinto et al., 2004). We speculate that, in addition to prothrombin activator and other enzymes, lonofibrase may also contribute to the consumption of fibrinogen by directly degrading the molecule and indirectly, and perhaps most critically, through activation of the fibrinolytic system, as described for Ancrod (Dempfle et al., 2000; Dempfle et al., 2001). Molecular identification, cloning, and expression of lonofibrase, in addition to other molecules, will be an important step in our understanding of the envenomation process. A diagram containing the potential targets for some *L. obliqua* toxins is presented in Figure 7.

Finally, manipulation of the fibrinolytic system appears to be the rule rather than the exception regarding accidents with *Lonomia* sp. caterpillars. In fact, *L. achelous* venom contains enzymes that affect fibrinolysis in a distinct manner when compared with *L. obliqua*: a urokinase-like protease that degrades FXIII, and two plasmin-like activities (Arocha-Piñango and Guerrero, 2001). Other components such as PLA2, serpins, and lectins, in addition to the host humoral response to envenomation, may also participate in the symptoms manifested by most patients, including hemolysis, bleeding, and acute renal failure (Kellen et al., 1995). It appears that our attempt to generate a catalog containing the putative toxic proteins of the caterpillar is an effective approach to generate testable hypotheses on the molecular basis of envenomation. It may help to find candidate proteins for development of a reliable diagnostic kit for this envenomation and for the improvement of serum production (Rocha-Campos et al., 2001; Theakston et al., 2003).

Abbreviations

NR	non-redundant
PLA2	phospholipase A2

Acknowledgments

We thank Drs. Thomas E. Wellems, Robert W. Gwadz, and Thomas J. Kindt (NIAID/NIH) for encouragement and support and the Brazilian agency CAPES (Ministry of Science and Technology Department) for the fellowship to A.B.G. Veiga. We thank Van My Phan (LMVR/NIAID) for technical assistance and Dr. Mark K. Garfield (Research Technologies Branch/NIAID) for protein sequencing. We acknowledge Brenda Rae Marshall (NIAID) for editorial assistance. We express special thanks to the Health Department of the city of Videira (Santa Catarina, Brazil) for providing *L. obliqua* caterpillars. The authors are grateful to both reviewers for their time, comments and valuable suggestions in the manuscript.

References

- Aird SD. Ophidian envenomation and the role of the purines. *Toxicon* 2002;40:335–393. [PubMed: 11738231]
- Åkerstrom B, Flower DR, Salier J. Lipocalins: unity in diversity. *Biochim. Biophys. Acta* 2000;1482:1–8. [PubMed: 11058742]
- Amarant T, Burkhart W, LeVine H, Arocha-Piñango CL, Parikh I. Isolation and complete amino acid sequence of two fibrinolytic proteinases from the toxic *Saturnid* caterpillar *Lonomia achelous*. *Biochim. Biophys. Acta* 1991;1079:214–221. [PubMed: 1911844]
- Arocha-Piñango CL, Guerrero B. *Lonomia* genus caterpillar envenomation: clinical and biological aspects. *Haemostasis* 2001;31:288–293. [PubMed: 11910197]
- Bell WR Jr. Defibrinogenating enzymes. *Drugs* 1997;54:18–31. [PubMed: 9360849]
- Braud S, Bon C, Wisner A. Snake venom proteins acting on hemostasis. *Biochimie* 2000;82:851–859. [PubMed: 11086215]

- Campos IT, Tanaka-Azevedo AM, Tanaka AS. Identification and characterization of a novel factor XIIIa inhibitor in the hematophagous insect, *Triatoma infestans* (Hemiptera: Reduviidae). *FEBS Lett* 2004;19:512–516. [PubMed: 15556638]
- Chapman HA, Riese RJ, Shi G. Emerging roles for cysteine proteases in human biology. *Ann. Rev. Physiol* 1997;59:63–88. [PubMed: 9074757]
- Dempfle CE, Argiriou S, Kucher K, Muller-Peltzer H, Rubsamen K, Heene DL. Analysis of fibrin formation and proteolysis during intravenous administration of ancrod. *Blood* 2000;96:2793–2802. [PubMed: 11023513]
- Dempfle CE, Argiriou S, Alesci S, Kucher K, Muller-Peltzer H, Rubsamen K, Heene DL. Fibrin formation and proteolysis during ancrod treatment. Evidence for des-A-profibrin formation and thrombin independent factor XIII activity. *Ann. N.Y. Acad. Sci* 2001;936:210–214. [PubMed: 11460476]
- Donato JL, Moreno RA, Hyslop S, Duarte A, Antunes E, Le Bonniec BF, Rendu F, De Nucci G. *Lonomia obliqua* caterpillar spicules trigger human blood coagulation via activation of Factor X and prothrombin. *Thromb. Haemost* 1998;79:539–542. [PubMed: 9531036]
- Fox, JW.; Shannon, JD.; Stefansson, B.; Kamiguti, AS.; Theakston, RDG.; Serrano, SMT.; Camargo, ACM.; Sherman, N. Role of discovery sciences in toxinology: examples in venom proteomics. In: Ménez, A., editor. *Perspectives in Molecular Toxinology*. West Sussex, UK: Wiley; 2002. p. 97-106.
- Francischetti IM, Valenzuela JG, Pham VM, Garfield MK, Ribeiro JM. Toward a catalog for the transcripts and proteins (sialome) from the salivary gland of the malaria vector *Anopheles gambiae*. *J. Exp. Biol* 2002;205:2429–2451. [PubMed: 12124367]
- Francischetti IM, My-Pham V, Harrison J, Garfield MK, Ribeiro JM. *Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. *Gene* 2004;337:55–69. [PubMed: 15276202]
- Friedrich T, Kroger B, Bialojan S, Lemaire HG, Hoffken HW, Reuschenbach P, Otte M, Dodt J. A Kazal-type inhibitor with thrombin specificity from *Rhodnius prolixus*. *J. Biol. Chem* 1993;268:16216–16222. [PubMed: 8344906]
- Harrison RA, Oliver J, Hasson SS, Bharati K, Theakston RD. Novel sequences encoding venom C-type lectins are conserved in phylogenetically and geographically distinct *Echis* and *Bitis* viper species. *Gene* 2003;315:95–102. [PubMed: 14557069]
- Joseph JS, Chung MC, Jeyaseelan K, Kini RM. Amino acid sequence of trocarnin, a prothrombin activator from *Tropidechis carinatus* venom: its structural similarity to coagulation factor Xa. *Blood* 1999;15:621–631. [PubMed: 10397729]
- Kanost MR. Serine proteinase inhibitors in arthropod immunity. *Dev. Comp. Immunol* 1999;23:291–301. [PubMed: 10426423]
- Kanost MR, Jiang H, Yu XQ. Innate immune responses of a lepidopteran insect, *Manduca sexta*. *Immunol. Rev* 2004;198:97–105. [PubMed: 15199957]
- Kelen EMA, Picarelli ZP, Duarte AC. Hemorrhagic syndrome induced by contact with caterpillars of the genus *Lonomia* (Saturniidae, Hemileucinae). *J. Toxic. - Toxin Rev* 1995;14:283–308.
- Kini RM. Excitement ahead: structure, function and mechanism of snake venom phospholipase A2 enzymes. *Toxicon* 2003;42:827–840. [PubMed: 15019485]
- Lambeau G, Lazdunski M. Receptors for a growing family of secreted phospholipases A2. *Trends Pharmacol. Sci* 1999;20:162–170. [PubMed: 10322502]
- Larson, PJ.; Katherine, AH. Molecular genetics of blood coagulation. In: Loscalzo, J.; Schafer, AI., editors. *Thrombosis and Hemorrhage*. Baltimore: Williams & Wilkins; 1998. p. 29-43.
- Lopes-Ferreira M, Barbaro KC, Cardoso DF, Moura-Da-Silva AM, Mota I. *Thalassophryne nattereri* fish venom: biological and biochemical characterization and serum neutralization of its toxic activities. *Toxicon* 1998;36:405–410. [PubMed: 9620588]
- Lopez-Ordoñez T, Rodríguez MH, Hernández-Hernández FC. Characterization of a cDNA encoding a cathepsin L-like protein of *Rhodnius prolixus*. *Insect Mol. Biol* 2001;10:505–511. [PubMed: 11881815]

- Markland FS, Kettner C, Schiffman S, Shaw E, Bajwa SS, Reddy KN, Kirakossian H, Patkos GB, Theodor I, Pirkle H. Kallikrein-like activity of crotalase, a snake venom enzyme that clots fibrinogen. *Proc. Natl. Acad. Sci. U.S.A* 1982;79:1688–1692. [PubMed: 7043462]
- Markland FS Jr. Snake venoms. *Drugs* 1997;54:1–10. [PubMed: 9360847]
- Murakami M, Masuda S, Shimbara S, Bezzine S, Lazdunski M, Lambeau G, Gelb MH, Matsukura S, Kokubu F, Adachi M, Kudo I. Cellular arachidonate-releasing function of novel classes of secretory phospholipase A2s (groups III and XII). *J. Biol. Chem* 2003;278:10657–10667. [PubMed: 12522102]
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 1997;10:1–6. [PubMed: 9051728]
- Pinto AFM, Dobrovolski R, Veiga ABG, Guimarães JA. Lonofibrase, a novel α -fibrinogenase from *Lonomia obliqua* caterpillars. *Thromb. Res* 2004;113:147–154. [PubMed: 15115670]
- Ribeiro JM, Francischetti IM. Role of arthropod saliva in blood feeding: sialome and post-sialome perspectives. *Annu. Rev. Entomol* 2003;48:73–88. [PubMed: 12194906]
- Ribeiro JM, Andersen J, Silva-Neto MA, Pham VM, Garfield MK, Valenzuela JG. Exploring the sialome of the blood-sucking bug *Rhodnius prolixus*. *Insect Biochem. Mol. Biol* 2004;34:61–79. [PubMed: 14976983]
- Rocha-Campos AC, Goncalves LR, Higashi HG, Yamagushi IK, Fernandes I, Oliveira JE, Ribela MTCP, Sousa-e-Silva MCC, Dias-da-Silva W. Specific heterologous F(abV)₂ antibodies revert blood incoagulability resulting from envenoming by *Lonomia obliqua* caterpillars. *Am. J. Trop. Med. Hyg* 2001;64:283–289. [PubMed: 11463118]
- Sappington TW, Raikhel AS. Molecular characteristics of insect vitellogenins and vitellogenin receptors. *Insect Biochem. Mol. Biol* 1998;28:277–300. [PubMed: 9692232]
- Seibert CS, Oliveira MRL, Gonçalves LRC, Santoro ML, Sano-Martins IS. Intravascular hemolysis induced by *Lonomia obliqua* caterpillar bristle extract: an experimental model of envenomation in rats. *Toxicon* 2004;44:793–799. [PubMed: 15500855]
- Theakston RD, Warrell DA, Griffiths E. Report of a WHO workshop on the standardization and control of antivenoms. *Toxicon* 2003;4:541–557. [PubMed: 12676433]
- Valenzuela JG, Pham VM, Garfield MK, Francischetti IM, Ribeiro JMC. Toward a description of the sialome of the adult female mosquito *Aedes aegypti*. *Insect Biochem. Mol. Biol* 2002;32:1101–1122. [PubMed: 12213246]
- Veiga ABG, Blochtein B, Guimarães JA. Structures involved in production, secretion and injection of the venom produced by the caterpillar *Lonomia obliqua* (Lepidoptera, Saturniidae). *Toxicon* 2001;39:1343–1351. [PubMed: 11384722]
- Veiga ABG, Pinto AFM, Guimarães JA. Fibrinogenolytic and procoagulant activities in the hemorrhagic syndrome caused by *Lonomia obliqua* caterpillars. *Thromb. Res* 2003;111:95–101. [PubMed: 14644086]
- Williams, E. Disseminated intravascular coagulation. In: Loscalzo, J.; Schafer, AL., editors. *Thrombosis and Hemorrhage*. Baltimore: Williams & Wilkins; 1998. p. 963-985.
- Zannin M, Lourenço DM, Motta G, Costa LRD, Grando M, Gamborgi GP, Noguti MA, Chudzinski-Tavassi AM. Blood coagulation and fibrinolytic factors in 105 patients with hemorrhagic syndrome caused by accidental contact with *Lonomia obliqua* caterpillar in Santa Catarina, Southern Brazil. *Thromb. Haemost* 2003;89:355–364. [PubMed: 12574817]

Fig. 1. SDS-PAGE analysis of *L. obliqua* bristle and tegument extracts, hemolymph, and cryosecretion. Amino acid sequences obtained by Edman degradation of the protein bands are as depicted in the gel figure (left) and detailed in the table (right). Letters in lower case indicate PTH-amino acids that did not match the sequence predicted by the corresponding cDNA. LOB, bristle; LOT, tegument; LOH, hemolymph; LOC, cryosecretion.

Fig. 2. Distribution of putative toxins in *L. obliqua* tegument and bristle cDNA libraries. (a) and (b) Percentage of sequences and clusters of putative toxins in the tegument library, respectively. (c) and (d) Percentage of sequences and clusters of putative toxins in the bristle library, respectively.

Fig. 3. Alignment of *L. obliqua* serine protease 1 (SP1) with *M. sexta* serine protease of the prophenoloxidase (PPO) cascade (gi|39655053), *H. sapiens* factors IXa (gi|67596), VII (gi|10518503), II (gi|4503635), and Xa (gi|538554), and trocarin from *T. carinatus* (gi|54402099). Black and grey shading indicate identity and high conservation of amino acids, respectively.

Fig. 4.
Alignment of *L. obliqua* serine proteinase inhibitor 2 (Serp2) with *B. mori* serpin-2 (gi|7341330), *C. felis* serpin 5 (gi|25527494), *A. gambiae* serine protease inhibitor (gi|6759386), *I. ricinus* protein (gi|14140097), *D. rerio* antithrombin (AT, gi|33504509), *R. appendiculatus* serpin-1 (gi|17223662), *H. sapiens* antithrombin III (ATIII, gi|999514), and *B. taurus* plasminogen activator inhibitor-1 (PAI-1, gi|27806497). Black and grey shading indicate identity and high conservation of amino acids, respectively.

Fig. 5.

Alignment of *L. obliqua* phospholipase A2 (PLA2) with PLA2s from *D. melanogaster* (gi|21627771), *H. suspectum* (gi|104338), *A. mellifera* (gi|16904372), *R. nomadica* (gi|999133), *A. phaiodactylus* (gi|46484897), *P. imperator* imperatoxin (gi|37079101), and *H. sapiens* Group III-PLA2 (gi|7657126). Black and grey shading indicate identity and high conservation of amino acids, respectively.

Fig. 6. Alignment of *L. obliqua* lectin 5 (Lect5) with *G. halys* fibrinogen clotting inhibitor (gi|4337050), *T. flavoviridis* FIX/FX binding protein (gi|2119527), *C. durissus terrificus* crotoctetin (gi|33332307), *E. carinatus* echicetin (gi|7993934), *M. lebetina* FX activator (gi|33391736), *B. jararaca* botrocetin (gi|264835), and *B. arietans* bitiscetin (gi|37928175). Black and grey shading indicate identity and high conservation of amino acids, respectively.

Fig. 7.

A diagram for the potential targets of *L. obliqua* toxins (section 3.6). “*L. obliqua* FX-activator” and “*L. obliqua* prothrombin-activator” acts on Factor X (FX) and prothrombin, respectively, leading ultimately to thrombin generation and fibrin formation (Donato et al., 1998). Fibrin is cross-linked by FXIIIa in a thrombin-dependent manner, and is subsequently degraded by plasmin with D-dimers formation (Williams, 1998). In addition, “*L. obliqua* α-fibrinogenase” acts on fibrinogen and fibrin (Veiga et al., 2003; Pinto et al., 2004), may lead to formation of FgDP (Fibrinogen-degradation products) and FbDP (fibrin-degradation products) that may act as cofactors for activation of tissue plasminogen activator (T-PA) (see Discussion) (Bell et al., 1997; Dempfle et al., 2000; Dempfle et al., 2001). The target for caterpillar putative anticoagulants (e.g.serpins) are not shown. The ultimate result is a blood coagulation disorder often associated with a hemorrhagic syndrome.

Table 1
cDNAs coding for toxins and putative secretory proteins from *L. obliqua* bristle and tegument

Cl no.	Seq ²	Match to NR protein database ³	E val ⁴	Species ⁵	Best match to CDD ⁶	Sig ⁷	Comments	Proteome match
<i>Tegument cDNA library</i>								
<i>putative toxins</i>								
37	5	biliverdin binding protein-II [Samia	6e-050	<i>Samia cynthia ricini</i>	lipocalin	Cyt	lipocalin - biliverdin-binding prot	19-LOB/LOT/LOH/LOC - DVVIDGAcPDMKAVSKFDMN
42	3	serpin 1 327 2e-088	2e-088		serpin	Cyt	serpin	
43	2	The Structure Of Active Serpin K From Manduca sexta	5e-076	<i>Manduca sexta</i>	serpin	Cyt	serpin	8-LOB/LOH/LOC - TETDLQRLRLRdgNDqFtaKM
86	2	predicted CDS, trypsin inhibitor-like	0.011	<i>Caenorhabditis elegans</i>	TIL	SIG	trypsin inhibitor	
112	2	CG5966 PA [Drosophila melanogaster]	9e-051	<i>Drosophila melanogaster</i>	lipase	SIG	lipase	
118	2	trypsin-like serine protease [Ctenocep	1e-023	<i>Ctenocephalides felis</i>	trypsin	Cyt	trypsin-like serine protease	
120	2	lipopolysaccharide binding protein [B	3e-065	<i>Bombyx mori</i>	lectin_c	Cyt	complement activation, lectin	
125	2	cystatin precursor - horseshoe crab (Tach	0.004	<i>Tachyplesus tridentatus</i>	cystatin	SIG	cystatin	
129	2	lipopolysaccharide binding protein [B	1e-057	<i>Bombyx mori</i>	lectin_c	SIG	lectin	16-LOB - TVIKDIAK
130	1	serpin-1 [Bombyx mori] 303 e-100	1e-100	<i>Bombyx mori</i>	serpin	SIG	serpin	
131	1	serpin-2 [Bombyx mori] 357 3e-097	3e-097	<i>Bombyx mori</i>	serpin	SIG	serpin	
220	1	ENSANGP00000018359 [Anopheles gambi	3e-008	<i>Anopheles gambiae str. PEST</i>	ldl_recept_a	SIG	serine protease	
247	1	chymotrypsin inhibitor CI-8	1e-105	<i>Bombyx mori</i>	serpin	SIG	serpin	8-LOH - FHYFGHEYDRTALGDAIdKA
317	1	CG33055-PG [Drosophila melanogaster	3e-009	<i>Drosophila melanogaster</i>	kazal	SIG	Kazal-type protease inhibitor	
333	1	ENSANGP00000010929 [Anopheles gambi	4e-042	<i>Anopheles gambiae str. PEST</i>		Cyt	serine protease	
351	1	ENSANGP00000014050 [Anopheles gambi	9e-048	<i>Anopheles gambiae str. PEST</i>	abhydrolase	Ind	probable phospholipase	
374	1	ENSANGP00000010965 [Anopheles gambi	2e-015	<i>Anopheles gambiae str. PEST</i>	TIL	Cyt	trypsin inhibitor-like	
396	1	30kP protease A (43k peptide) precurs	3e-068	<i>Bombyx mori</i>	trypsin	Cyt	trypsin-like serine protease	
423	1	cysteine endopeptidase (EC 3.4.22.-) prec	3e-024	<i>Bombyx mori</i>	Peptidase_C1	Ind	cathepsin-like cysteine peptidase	
426	1	kazal-type proteinase inhibi	1e-015	<i>Manduca sexta</i>		Ind	Kazal-type proteinase inhibitor	
454	1	Esterase-like protein (ESR-LP) [Bomb	7e-016	<i>Bombyx mori</i>		Cyt	esterase	
<i>putative secreted proteins</i>								
1	36	LCP18 [Bombyx mori] >gn BL_ORD_ID 66	1.00E-35	<i>Bombyx mori</i>	Chitin_bind_4	SIG	cuticle protein LCP18	
18	8	ENSANGP00000014835 [Anopheles gambi	8e-018	<i>Anopheles gambiae str. PEST</i>	Chitin_bind_4	SIG	cuticle protein LCP18	

bioRxiv preprint doi: <https://doi.org/10.1101/201001>; this version posted July 1, 2010. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

CI no.	Seq ²	Match to NR protein database ³	E val ⁴	Species ⁵	Best match to CDD ⁶	Sig ⁷	Comments	Proteome match
20	11						Unknown	
57	4	hypothetical protein [Haemophilus	4.00E-19	<i>Haemophilus somnus</i> 2336	dehydrin	SIG	prior protein, DNA binding prot	
79	3	ferritin heavy chain-like protein pre	6.00E-70	<i>Manduca sexta</i>	ferritin	SIG	ferritin	
106	2	CG32405-PA [Drosophila melanogaster	2.00E-20	<i>Drosophila melanogaster</i>	Chitin_bind_4	SIG	cuticle protein LCP18	
169	2	immune-induced protein 1 [M	8.00E-39	<i>Manduca sexta</i>	Reeler	SIG	phosphotyrosine phosphatase	
181	2	Larval cuticle protein 16/17 pre	7.00E-39		Chitin_bind_4	SIG	cuticle protein LCP18	
330	1	hypothetical protein [Streptomyces	2.00E-17	<i>Streptomyces avermitilis</i> MA-4680	DUF612	SIG	unknown	
349	1	ENSANGP00000019272 [Anopheles gambi	4.00E-05	<i>Anopheles gambiae</i> str. PEST		SIG	unknown	
352	1	p27K [Bombyx mori] 250 3e-065	3e-065	<i>Bombyx mori</i>		SIG	hemolymph glycoprotein	18-LOB - DIEIPNqKREQLKqILTeQc
381	1	tetraspanin D107 [Manduca se	1e-115	<i>Manduca sexta</i>	transmembrane4	SIG	tetraspanin - defense protein	
387	1	predicted protein [Neurospora crass	2.00E-07	<i>Neurospora crassa</i>	DUF726	SIG	unknown	
420	1	CG33806-PB [Drosophila melanogaster	0.025	<i>Drosophila melanogaster</i>	DUF40	SIG	unknown	
430	1	Attacin A precursor >gn BL_ORD_	3e-016	<i>Hyalophora cecropia</i>	Attacin_C	SIG	attacin	
473	1	Bombyxin F-1 precursor (BBX-F1)	0.012	<i>Bombyx mori</i>	Insulin	SIG	bombyxin	
Bristle cDNA library								
putative toxins								
1	54	bliverin binding protein-II	2E-052	<i>Samia cynthia ricini</i>	lipocalin	SIG	lipocalin	19-LOB/LOT/LOH/LOC - DVVIDGAcPDMKAVSKFDMN
8	18	Unknown (protein for MGC:66347)	3E-019	<i>Danio rerio</i>	dehydrin	SIG	kininogen precursor	
12	13	ENSANGP000000006359	2E-047	<i>Anopheles gambiae</i>	trypsin	Cyt	trypsin-like serine protease	11-LOB - VEgEAEDPskNdeS / gGTEAaFGdwPwMVyI
13	1	CG1184-PA	3E-034	<i>Drosophila melanogaster</i>	trypsin	Cyt	trypsin-like, venom serine protease	
33	4	lipopolysaccharide binding protein	8E-010	<i>Bombyx mori</i>	lectin_c	SIG	lectin, c type	
39	3	CG33045-PG	4E-009	<i>Drosophila melanogaster</i>	kazal	SIG	thrombin inhibitor-like protein	
49	3	cysteine-rich protein	0.004	<i>Vigna radiata</i>	Gamma-thionin	SIG	knot domain of toxins	
55	2	cathepsin B-like cysteine proteinase	3E-071	<i>Helicoverpa armigera</i>	Peptidase_C1	SIG	cathepsin B-like cysteine peptidase	
56	2	serpin 1	1E-020		serpin	Cyt	serpin	
109	1	ENSANGP00000020324	5E-042	<i>Anopheles gambiae</i>	trypsin	Cyt	trypsin-like serine protease	
130	1	trypsin-like serine protease	3E-048	<i>Ctenocephalides felis</i>	trypsin	Cyt	trypsin-like serine protease	
140	1	lipopolysaccharide binding protein	9E-044	<i>Bombyx mori</i>	lectin_c	SIG	c-type lectin	1-LOB - TKstQKcVL
154	1	Phospholipase A2 precursor	9E-017	<i>Apis mellifera</i>		Ind	phospholipase A2	

Cl no. ¹	Seq ²	Match to NR protein database ³	E val ⁴	Species ⁵	Best match to CDD ⁶	Sig ⁷	Comments	Proteome match
184	1	ENSANGP00000012234	1E-080	<i>Anopheles gambiae</i>	lectin_c	Cyt	c-type lectin	
212	1	V-CATH	5E-008	<i>E. postvittana nucleopolyhedrovirus</i>	Spore_permease	SIG	cathepsin-like cysteine peptidase	
245	1	putative odorant-degrading enzyme	7E-029	<i>Antheraea polyphemus</i>	COesterase	Cyt	esterase	
252	1	predicted CDS, trypsin inhibitor-li	0.007	<i>Caenorhabditis elegans</i>	TIL	Cyt	trypsin inhibitor-like protein	
264	1	masquerade-like serine proteinase hom	1E-134	<i>Bombyx mori</i>	trypsin	Ind	trypsin-like, venom serine protease	
272	1	similar to RIKEN cDNA 2310046M08	3E-004	<i>Rattus norvegicus</i>	serpin	Ind	serpin	
287	1	biliverdin binding protein-I	1E-022	<i>Samia cynthia ricini</i>		SIG	lipocalin	
297	1	ENSANGP00000014256	5E-020	<i>Anopheles gambiae</i>	COesterase	Ind	esterase	
333	1	kazal-type proteinase inhibi	5E-027	<i>Manduca sexta</i>	kazal	SIG	protease inhibitor	23-LOH - DvNLTNLkAAQAAQrAcL
putative secreted proteins								
10	12	RNA polymerase ECF-type sigma facto	0.059	<i>Bacteroides thetaiotaomicron VPI-5482</i>		SIG	unknown	
21	2	antimicrobial peptide cecropin 6	1E-006	<i>Manduca sexta</i>	cecropin	SIG	cecropin	
51	2	AttacinE and F precursor (Immune	2E-098	<i>Hyalophora cecropia</i>	Attacin_C	SIG	Attacin	
93	2	sensory appendage protein 1	1E-007	<i>Manduca sexta</i>	OS-D	SIG	chemosensory protein	
104	1	ENSANGP00000020072	3E-007	<i>Anopheles gambiae</i>	PBP_GOBP	SIG	pheromone binding protein	7-LOB - SHqKLQhFLK
176	1	Cecropin A precursor (Cecropin C)	6E-024	<i>Hyalophora cecropia</i>	cecropin	SIG	cecropin	

Cluster number

Number of sequences in the cluster

Best protein match by blastX to the non-redundant protein database of NCBI

E-value, significance of match to NR sequence of previous column

Organism with best match to NR protein database

Best match to the Conserved Domain Database

Potential presence of signal P: SIG, signal-P; CYT, cytoplasmic; Ind, indeterminate

Table 2

A catalog of the full-length *L. obliqua* cDNAs coding for proteins probably involved in envenomation

Sequence name ¹	Cl no. ²	GenBank ³	Seq size ⁴	SigP Result	Cleavg Pos. ⁵	MW ⁶	Mature MW ⁷	pI ⁸	Match to NR protein database ⁹	E val ¹⁰	Full-length ¹¹	Comments	Tissue ¹²
<i>Cysteine proteinases</i>													
LOqua-CysPep1	423	AY829805	65	IND		7.08		6.9	cathepsin L-like protease [Helicoverp	5E-026	N	cysteine proteinase, cathepsin-like	T
LOqua-CysPep2	55	AY829838	338	SIG	20-21	36.9	34.773	5.7	cathepsin B-like cysteine proteinase [	1E-172	Y	cysteine proteinase, cathepsin-like	B
LOqua-Cyst1	125	AY829806	120	SIG	22-23	13.3	10.59	5.1	L-cystatin precursor >gn BL_OR	6E-004	Y	cystatin	T
<i>Defense proteins</i>													
LOqua-Def2	430	AY829737	121	CYT		13.3		9.4	Attacin A precursor >gn BL_ORD_	2E-020	N	attacin	T
LOqua-Def3	51	AY829840	233	SIG	17-18	25.1	22.968	5.4	Attacin E precursor (Immune prote	1E-107	Y	attacin precursor	B
LOqua-Def4	21	AY829848	65	SIG	25-26	7.23	4.255	11	antimicrobial peptide cecropin 6 [Man	5E-007	Y	cecropin	B
LOqua-Def6	36	AY829857	75	SIG	20-21	8.23	5.786	9.2	gallerimycin [Spodoptera frugiperda]	4E-004	Y	antimicrobial toxin	B
LOqua-Ease1	454	AY829807	63	CYT		7.61		6	Esterase-like protein (ESR-LP) [Bomb	4E-019	N	esterase	T
<i>Lectins</i>													
LOqua-Lect1	120	AY829822	321	SIG	19-20	35.8	33.653	6.1	lipopolysaccharide binding protein [B	9E-072	Y	Immunelectin	T
LOqua-Lect3	140	AY829836	320	SIG	20-21	36.3	33.922	6.3	lipopolysaccharide binding protein [B	2E-073	Y	lectin	B
LOqua-Lect4	184	AY829849	148	CYT		17.2		4.9	ENSANGP00000012234 [Anopheles gambi	6E-077	N	c-type lectin	B
LOqua-Lect5	33	AY829846	162	SIG	16-17	18.3	16.299	5.6	lipopolysaccharide binding protein [B	5E-010	Y	lectin	B
<i>Lipocalins</i>													
LOqua-Lipcl1	1	AY829833	201	SIG	16-17	22.4	20.558	6.1	biliverdin binding protein-II [Samia	2E-052	Y	lipocalin	B
LOqua-Lipcl3	104	AY829856	137	SIG	16-17	15.4	13.493	4.9	odorant-binding protein AgamOBP26 [An	2E-007	Y	lipocalin	B
LOqua-Lipcl4	37	AY829809	198	CYT		21.6		6.1	biliverdin binding protein-II [Samia	5E-050	Y	lipocalin	T
<i>Serine proteases</i>													
LOqua-PPOAF1	264	AY829844	398	IND		42.9		5.2	masquerade-like serine proteinase hom	1E-168	N	serine protease, PPO-activating factor	B
LOqua-SP1	220	AY829818	519	SIG	19-20	57.4	55.244	6.3	pattern recognition serine proteinase	1E-128	Y	serine protease	T
LOqua-SP2	333	AY829819	322	CYT		36.1		6.5	ENSANGP00000010929 [Anopheles gambi	4E-042	Y	serine protease	T
LOqua-SP3	396	AY829820	272	CYT		29.7		8.2	30kP protease A (43k peptide) precurs	2E-068	Y	serine protease, chymotrypsin-like	T
LOqua-SP4	118	AY829821	229	CYT		25.9		5.9	trypsin-like serine protease [Ctenocep	1E-030	N	serine protease, PPO-activating factor	T
LOqua-SP5	109	AY829842	242	IND		26.8		5.9	ENSANGP00000020324 [Anopheles gambi	7E-043	N	serine protease, PPO-activating factor	B
LOqua-SP6	12	AY829843	302	SIG	34-35	34.1	30.136	4.9	ENSANGP00000006359 [Anopheles gambi	2E-047	Y	serine-protease, plasminogen activator	B
LOqua-SP7	130	AY829841	280	CYT		30.5		6.1	trypsin-like serine protease [Ctenocep	3E-076	N	serine-protease, plasminogen activator	B

Sequence name ¹	Cl no. ²	GenBank ³	Seq size ⁴	SigP Result	Cleav ^g Pos. ⁵	MW ⁶	Mature MW ⁷	pI ⁸	Match to NR protein database ⁹	E val ¹⁰	Full-length ¹¹	Comments	Tissue ¹²
<i>Serpins</i>													
LOqua-Serp1	130	AY829814	476	SIG	18–19	52.4	50.199	5.1	serpin-2 [Bombyx mori]	1E-152	Y	serpin	T
LOqua-Serp2	131	AY829815	395	SIG	18–19	43.8	41.601	5.1	serpin-2 [Bombyx mori]	1E-153	Y	serpin	T
LOqua-Serp3	42	AY829816	383	CYT		43.2		6.1	serpin 461 e-128	1E-128	N	serpin	T
LOqua-Serp4	43	AY829817	322	IND		36.4		6.7	serpin 1 399 e-110	1E-110	N	serpin	T
LOqua-Serp6	247	AY829847									Y	serpin	T
<i>Other protease inhibitors</i>													
LOqua-P11	317	AY829810	154	SIG	19–20	17	14.644	6	similar to CG8806-PA [Apis mellifera]	3E-011	Y	protease inhibitor	T
LOqua-P13	374	AY829811	398	SIG	21–22	44.4	41.753	8.7	ENSANGP00000010965 [Anopheles gambiae]	2E-015	Y	trypsin inhibitor-like	T
LOqua-P14	86	AY829812	102	SIG	20–21	11	8.583	5.6	predicted CDS, trypsin inhibitor-Ii	0.005	Y	trypsin inhibitor-like	T
LOqua-P15	420	AY829813	86	SIG	19–20	9.7	7.268	5	CG33196-PB [Drosophila melanogaster]	0.035	N	trypsin inhibitor-like	T
LOqua-P16	252	AY829839	86	SIG	19–20	9.31	7.157	7.7	Hypothetical protein CBG19175 [Caeno]	0.002	Y	trypsin-inhibitor like	B
LOqua-P17	333	AY829835	94	SIG	18–19	10.2	7.962	8.3	kazal-type proteinase inhibitor [Mandu]	4E-028	Y	kazal-type protease inhibitor	B
LOqua-P19	49	AY829837	56	SIG	17–18	6.1	4.057	9	plant defensin precursor [Vigna radiata]	0.002	Y	proteinase inhibitor	B
<i>Phospholipase A2</i>													
LOqua-PLA1	154	AY829845	107	SIG	22–23	12.3	9.593	6.9	RE08605p [Drosophila melanogaster]	2E-010	Y	PLA2	B

¹ Putative toxins

² Cluster numbers as in Table 1

³ GenBank accession number

⁴ Length of aminoacid sequence

⁵ Signal peptide cleavage position

⁶ Molecular mass before removal of signal peptide

⁷ Molecular mass of the mature protein

⁸ pI of the mature protein

⁹ Best protein match by blastX to the non-redundant protein database of NCBI

¹⁰ E-value, significance of match to NR sequence of previous column

¹¹ Complete sequence (Y) or truncated (N)

¹² Tissue used for cDNA library construction: B, bristles; T, tegument

Table 3
A catalog toward the full-length *L. obliqua* cDNAs coding for housekeeping protein with various putative functions

Sequence name ¹	Cl no. ²	GenBank ³	Seq size ⁴	SigP Result	Cleavg Pos. ⁵	MW ⁶	Mature MW ⁷	pI ⁸	Match to NR protein database ⁹	E val ¹⁰	Full-length ¹¹	Comments	Tissue ¹²
<i>Sensory proteins</i>													
LOqua-Sens1	22	AY829852	127	SIG	17-18	14.502	12.465	5.67	chemosensory protein CSP1 [Bombyx mori] 162 1e-039	1E-039	Y	sensory protein	B
LOqua-Sens2	93	AY829853	112	SIG	19-20	12.719	10.438	8.38	sensory appendage protein 1 [Manduca s	2E-013	Y	sensory protein	B
<i>Cuticle proteins</i>													
LOqua-Cp1	1	AY829732	107	SIG	16-17	11.502	9.613	4.22	cuticle protein LCP18 [Bombyx mori] >	4E-036	Y	cuticle protein	T
LOqua-Cp2	106	AY829733	120	SIG	15-16	12.886	11.084	4.23	cuticle protein 14 [Antheraea pernyi] 3e-035 148	3E-035	Y	cuticle protein	T
LOqua-Cp3	18	AY829734	196	SIG	19-20	22.743	20.57	5.26	ENSANGP00000014835 [Anopheles gambia	1E-021	Y	cuticle protein	T
LOqua-Cp4	181	AY829735	126	SIG	17-18	14.02	11.689	4.65	larval cuticular protein 1 [Helicoverp	2E-039	Y	cuticle protein	T
<i>Heat shock proteins</i>													
LOqua-Hsp1	7	AY829746	192	CYT			21.922		ENSANGP00000018891 [Anopheles gambia	4E-059	Y	heat shock protein	T
LOqua-Hsp2	92	AY829747	126	CYT			13.389		ribosomal protein S14 [Plutella xylo	3E-054	Y	heat shock protein	T
LOqua-Hsp3	93	AY829748	188	CYT			21.473		heat shock protein hsp20.4 [Bombyx mori] 303 2e-081	2E-081	Y	heat shock protein	T
LOqua-Hsp4	14	AY829851	654	CYT			71.639	5.33	Heat shock 70 kDa protein cogna	0.0	Y	hsp	B
<i>Structural proteins</i>													
LOqua-Act1	10	AY829794	294	CYT			32.499		ENSANGP00000009996 [Anopheles gambia	1E-165	Y	actin	T
LOqua-LIM1	19	AY829796	93	CYT			10.018	8.59	Lim protein [Bombyx mori] 202 1e-051	1E-051	Y	LIM protein	T
LOqua-Myo1	13	AY829797	141	CYT			13.729	5.43	similar to putative MLC1 protein [A	1E-043	Y	myosin	T
LOqua-Myo2	14	AY829798	150	CYT			16.693	4.57	similar to putative MLC1 protein [A	9E-062	Y	myosin	T
LOqua-Myo3	25	AY829799	400	CYT			42.9	4.94	similar to ENSANGP00000018155 [Apis	1E-069	Y	myosin	T
LOqua-Trmyo1	68	AY829800	325	CYT			36.011	4.98	tropomyosin 1 [Plutella xylostella] 490 e-137	1E-137	Y	tropomyosin	T
LOqua-TrpnC2	104	AY829802	172	CYT			19.26	3.97	TPA: troponin C type IIa [Apis melli	2E-070	Y	troponin C	T
LOqua-TrpnI1	24	AY829803	382	CYT			45.721	5	TPA: troponin T isoform 2 [Drosophil	1E-164	Y	troponin	T
LOqua-TrpnI2	35	AY829804	141	CYT			16.703	10.32	ENSANGP00000020935 [Anopheles gambi	2E-042	Y	troponin	T
<i>General metabolism</i>													
LOqua-ATP1	55	AY829739	161	CYT			17.024		RE24457p [Drosophila melanogaster] >g	8E-049	Y	ATP synthase	T

Sequence name ¹	Cl no. ²	GenBank ³	Seq size ⁴	SigP Result	Cleavag Pos. ⁵	MW ⁶	Mature MW ⁷	pI ⁸	Match to NR protein database ⁹	E val ¹⁰	Full-length ¹¹	Comments	Tissue ¹²
LOqua-ATP2	101	AY829740	63	IND		6.906			ATP synthase beta [Drosophila melanog	5E-025	N	ATP synthase	T
LOqua-Cyt1	72	AY829741	272	BL	37-38	30.908	26.563	6.51	cytochrome c oxidase subunit I [Ant	1E-148	N	cytochrome c oxidase	T
LOqua-Ease1	454	AY829807	63	CYT		7.608			Esterase-like protein (ESR-LP) [Bomb	4E-019	N	esterase	T
LOqua-EF1	32	AY829742	428	CYT		46.724			translation elongation factor eEF-1 alpha	0.0	Y	elongation factor	T
LOqua-Fer1	79	AY829743	226	SIG	20-21	25.708	23.555	5.92	ferritin heavy chain-like protein pre	7E-079	Y	ferritin	T
LOqua-Hmn1	473	AY829744	96	SIG	18-19	10.957	8.679	5.26	Bombyxin F-1 precursor (BBX-F1)	0.003	Y	Bombyxin	T
LOqua-HO1	307	AY829745	243	IND		27.763			ENSANGP00000014279 [Anopheles gambi	5E-029	Y	heme oxygenase	T
LOqua-Hyp1	381	AY829738	255	SIG	35-36	27.715	23.369	5.41	tetraspanin D107 [Manduca sexta] e-115	1E-115	Y	tetraspanin	T
LOqua-Hyp10	47	AY829749	211	IND		22.502			CG31075-PA [Drosophila melanogaster	6E-074	N	aldehyde dehydrogenase 2	T
LOqua-Hyp11	110	AY829750	285	CYT		30.299			methylenetetrahydrofolate dehydrogenase	1E-139	N	C-1-tetrahydrofolate synthase	T
LOqua-Hyp12	119	AY829751	195	BL		20.9			CG1236-PA [Drosophila melanogaster]	1E-057	N	glyoxylate reductase/hydroxypyruvate	T
LOqua-Hyp13	78	AY829752	152	CYT		16.938			myo-inositol 2-dehydrogenase [Salmo	5E-018	N	inositol 2-dehydrogenase	T
LOqua-Hyp14	64	AY829753	264	CYT		28.275			ribosome-associated protein P40 [Bom	1E-103	N	laminin binding protein	T
LOqua-Hyp15	102	AY829754	53	CYT		6.21			phosphoenolpyruvate carboxykinase [Li	3E-009	N	phosphoenolpyruvate carboxykinase	T
LOqua-Hyp16	76	AY829755	162	CYT		17.251			PREDICTED: similar to PABPC4 protei	6E-043	N	poli-A binding protein	T
LOqua-Hyp17	56	AY829764	130	CYT		14.359			ubiquitin/ribosomal protein S30e fus	2E-057	Y	ribosomal protein	T
LOqua-Hyp18	121	AY829765	277	CYT		30.355			Y-box protein [Bombyx mori] 353	2E-096	Y	Y box protein	T
LOqua-Hyp2	12	AY829756	220	CYT		24.953			ENSANGP000000014344 [Anopheles gambi	3E-082	Y	beta-ureidopropionase	T
LOqua-Hyp20	87	AY829795	100	CYT		11.281			similar to ENSANGP000000013163 [Apis	7E-035	N	paramyosin	T
LOqua-Hyp21	112	AY829808	274	IND		30.927			LD47264p [Drosophila melanogaster] >g	2E-052	N	lipase	T
LOqua-Hyp22	443	AY829823	162	CYT		19.133			arylphorin precursor; ArH [Hyalophora	4E-071	N	arylphorin	T
LOqua-Hyp23	50	AY829824	157	CYT		17.42			putative multifunctional protein ADE2	4E-065	Y	multifunctional protein	T
LOqua-Hyp3	91	AY829757	217	CYT		24.387			glutathione-S-transferase-like protei	8E-074	Y	glutathione-S-transferase	T
LOqua-Hyp35	10	AY829850	88	SIG	18-19	10.16	7.737	6.49	RNA polymerase ECF-type sigma facto	0.10	Y	secreted peptide	B
LOqua-Hyp36	8	AY829834	102	SIG	17-18	10.886	8.966	6.18	membrane-associated histidine-rich pr	2E-022	Y	secreted peptide	B
LOqua-Hyp4	171	AY829758	147	CYT		16.176			ENSANGP000000009316 [Anopheles gambi	3E-039	N	lipin	T
LOqua-Hyp5	142	AY829759	112	ANC		13.381			NADH dehydrogenase subunit 6 [Anthe	7E-037	N	NADH dehydrogenase	T
LOqua-Hyp6	53	AY829760	172	CYT		19.945			translationally controlled tumor pro	5E-087	Y	translationally controlled tumor protein	T
LOqua-Hyp7	122	AY829761	113	CYT		12.782			receptor for activated protein kinas	9E-048	N	activated protein kinase C receptor	T

Sequence name ¹	Cl no. ²	GenBank ³	Seq size ⁴	SigP Result	Cleavag Pos. ⁵	MW ⁶	Mature MW ⁷	pI ⁸	Match to NR protein database ⁹	E val ¹⁰	Full-length ¹¹	Comments	Tissue ¹²
LOqua-Hyp8	111	AY829762	138	CYT		15.057	10.43	10.43	ADP/ATP translocase [Manduca sexta] 229 7e-060	7E-060	N	ADP/ATP translocase	T
LOqua-Hyp9	137	AY829763	186	CYT		20.514	5.51	5.51	ENSANGP00000013314 [Anopheles gambi	7E-062	N	aldehyde dehydrogenase 2	T
Ribosomal proteins													
LOqua-RbP1	107	AY829766	158	CYT		17.972	9.99	9.99	ribosomal protein S10 [Spodoptera fru	5E-084	Y	ribosomal protein	T
LOqua-RbP10	58	AY829767	131	CYT		14.804	10.01	10.01	ribosomal protein S15A [Spodoptera fr	2E-067	Y	ribosomal protein	T
LOqua-RbP11	61	AY829768	116	CYT		13.434	10.47	10.47	ribosomal protein L44 [Choristoneura	3E-056	Y	ribosomal protein	T
LOqua-RbP12	63	AY829769	109	CYT		11.924	11.46	11.46	ribosomal protein L37 [Spodoptera fru	2E-043	Y	ribosomal protein	T
LOqua-RbP13	66	AY829770	190	CYT		21.502	9.88	9.88	ribosomal protein L9 [Spodoptera frug	3E-097	Y	ribosomal protein	T
LOqua-RbP14	74	AY829771	148	CYT		16.765	11.2	11.2	ribosomal protein L24 [Spodoptera fru	2E-075	Y	ribosomal protein	T
LOqua-RbP15	75	AY829772	139	CYT		15.734	9.74	9.74	ribosomal protein L10A [Spodoptera fr	2E-073	N	ribosomal protein	T
LOqua-RbP16	88	AY829773	70	CYT		8.253	10	10	Ribosomal protein L38 [Plutella xylo	3E-032	Y	ribosomal protein	T
LOqua-RbP17	97	AY829774	115	CYT		11.852	11.33	11.33	ribosomal protein L8 [Spodoptera frug	7E-058	Y	ribosomal protein	T
LOqua-RbP18	92	AY829775	129	IND		15.51	10.79	10.79	ribosomal protein L27 [Spodoptera fru	6E-064	N	ribosomal protein	T
LOqua-RbP19	113	AY829776	81	BL	22-23	6.839	11.06	11.06	Ribosomal protein S26 [Plutella xylo	1E-030	N	ribosomal protein	T
LOqua-RbP2	117	AY829777	147	CYT		17.025	10.44	10.44	ribosomal protein S15 [Spodoptera fru	2E-079	Y	ribosomal protein	T
LOqua-RbP20	116	AY829778	131	CYT		15.245	10.55	10.55	QM protein [Heliothis virescens] 252 1e-066	1E-066	N	ribosomal protein	T
LOqua-RbP21	134	AY829779	119	CYT		13.522	10.15	10.15	ribosomal protein L26 [Spodoptera fru	3E-059	N	ribosomal protein	T
LOqua-RbP22	160	AY829780	110	CYT		13.375	10.62	10.62	similar to CG2746-PA [Apis mellifera] 174 4e-043	4E-043	N	ribosomal protein	T
LOqua-RbP23	29	AY829781	300	CYT		33.568	11	11	similar to CG5502-PA [Apis mellifera] 363 2e-099	2E-099	N	ribosomal protein	T
LOqua-RbP24	38	AY829782	280	CYT		31.935	10.2	10.2	ribosomal protein L3 [Spodoptera frug	1E-155	N	ribosomal protein	T
LOqua-RbP25	69	AY829783	97	CYT		11.539	10.5	10.5	Ribosomal protein S18 [Plutella xylo	2E-047	N	ribosomal protein	T
LOqua-RbP26	84	AY829784	86	CYT		9.604	10.16	10.16	Ribosomal protein L27A2 [Plutella xy	4E-039	N	ribosomal protein	T
LOqua-RbP27	90	AY829785	200	CYT		23.378	10.1	10.1	ribosomal protein L7 [Spodoptera frug	5E-091	N	ribosomal protein	T
LOqua-RbP28	98	AY829786	137	CYT		15.244	10.75	10.75	ribosomal protein S19 [Spodoptera fru	4E-066	N	ribosomal protein	T
LOqua-RbP3	123	AY829787	42	CYT		4.937	5.99	5.99	ubiquitin / ribosomal protein S27a - tobac	2E-007	Y	ribosomal protein	T
LOqua-RbP30	53	AY829854	139	CYT		15.96	11.68	11.68	Ribosomal protein L28 [Plutella xylo	2E-067	Y	ribosomal protein	B
LOqua-RbP31	87	AY829855	187	CYT		21.539	10.38	10.38	ENSANGP00000011784 [Anopheles gambi	8E-084	Y	ribosomal protein	B
LOqua-RbP4	124	AY829788	159	CYT		18.393	10.51	10.51	ribosomal protein S11 [Bombyx mori] 305 3e-082	3E-082	Y	ribosomal protein	T

Sequence name ¹	Cl no. ²	GenBank ³	Seq size ⁴	SigP Result	Cleav ⁵ Pos.	MW ⁶	Mature MW ⁷	pI ⁸	Match to NR protein database ⁹	E val ¹⁰	Full-length ¹¹	Comments	Tissue ¹²
LOqua-RBP5	126	AY829789	143	CYT		16.04	10.66	10.66	ribosomal protein S23 [Bombyx mori] 292 8e-079	8E-079	Y	ribosomal protein	T
LOqua-RBP6	188	AY829790	159	CYT		17.964	10.97	10.97	Ribosomal protein L21 [Plutella xylo	1E-081	Y	ribosomal protein	T
LOqua-RBP7	31	AY829791	111	BL		11.322	4.1	4.1	60S acidic ribosomal protein P1 [Spod	7E-050	Y	ribosomal protein	T
LOqua-RBP8	52	AY829792	56	CYT		6.692	9.79	9.79	ribosomal protein S29 [Spodoptera fru	2E-028	Y	ribosomal protein	T
LOqua-RBP9	54	AY829793	25	IND		3.419	12.96	12.96	REG3504p [Drosophila melanogaster] >g	1E-006	Y	ribosomal protein	T
Unknown proteins													
LOqua-Hyp24	351	AY829825	277	IND		31.282	6	6	ENSANGP00000014050 [Anopheles gambi	5E-041	N	unknown	T
LOqua-Hyp25	57	AY829826	102	SIG	17-18	10.883	8.952	6.18	membrane-associated histidine-rich pr	2E-022	Y	unknown	T
LOqua-Hyp28	20	AY829827	88	SIG	17-18	9.668	7.511	4.79	hypothetical protein AN8479.2 [Asperg	0.17	Y	unknown	T
LOqua-Hyp29	330	AY829828	233	SIG	16-17	25.173	23.035	4.61	surface layer protein [Bacillus cer	7E-013	Y	unknown	T
LOqua-Hyp30	349	AY829829	62	SIG	17-18	7.073	5.037	7.99			Y	unknown	T
LOqua-Hyp31	352	AY829830	225	SIG	17-18	24.999	22.853	5.28	27k hemolymph protein [Galleria mell	7E-063	Y	unknown	T
LOqua-Hyp32	387	AY829831	74	SIG	15-16	6.975	5.363	5.52	predicted protein [Neurospora crass	3E-007	Y	unknown	T
LOqua-Hyp33	65	AY829832	66	CYT		7.836			CG6770 [Drosophila yakuba] >gnl BL_OR	5E-026	Y	unknown	T
LOqua-Hyp37	202	AY829858	109	CYT		12.539			ENSANGP00000010299 [Anopheles gambi	6.7	Y	unknown	B
LOqua-Hyp38	212	AY829859	130	SIG	19-20	15.243	12.614	4.5	V-CATH [Epiphyas postvittana nucleopo	3E-008	Y	unknown secreted peptide	B

¹Proteins probably involved in envenomation

²Cluster numbers as in Table 1

³GenBank, NR database accession number

⁴Length of aminoacid sequence

⁵Signal peptide cleavage position

⁶Molecular mass before removal of signal peptide

⁷Molecular mass of the mature protein

⁸pI of the mature protein

⁹Best protein match by blastX to the non-redundant protein database of NCBI

¹⁰E-value, significance of match to NR sequence of previous column

¹¹ Complete sequence (Y) or truncated (N)

¹² Tissue used for cDNA library construction: B, bristles; T, tegument