Insect Biochemistry and Molecular Biology 39 (2009) 125–134

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

# The *Anopheles gambiae* adult midgut peritrophic matrix proteome

R.R. Dinglasan [a,1], M. Devenport [a,1], L. Florens [b,1,2], J.R. Johnson [b,3], C.A. McHugh [a], M. Donnelly-Doman [d], D.J. Carucci [c], J.R. Yates, III [b], M. Jacobs-Lorena [a,*]

[a] Department of Molecular Microbiology & Immunology, Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA
[b] Cell Biology Department, The Scripps Research Institute, La Jolla, CA, USA
[c] Grand Challenges in Global Health, Foundation for the National Institutes of Health, 45 Center Drive, Bethesda, MD, USA
[d] Department of Genetics, Case Western Reserve University School of Medicine, Cleveland, OH, USA

ARTICLE INFO

*Article history:*
Received 21 July 2008
Received in revised form
5 October 2008
Accepted 23 October 2008

*Keywords:*
*Anopheles gambiae*
Peritrophic matrix
Mosquito
Proteome

ABSTRACT

Malaria is a devastating disease. For transmission to occur, *Plasmodium*, the causative agent of malaria, must complete a complex developmental cycle in its mosquito vector. Thus, the mosquito is a potential target for disease control. *Plasmodium* ookinetes, which develop within the mosquito midgut, must first cross the midgut's peritrophic matrix (PM), a thick extracellular sheath that completely surrounds the blood meal. The PM poses a partial, natural barrier against parasite invasion of the midgut and it is speculated that modifications to the PM may lead to a complete barrier to infection. However, such strategies require thorough characterization of the structure of the PM. Here, we describe for the first time, the complete PM proteome of the main malaria vector, *Anopheles gambiae*. Altogether, 209 proteins were identified by mass spectrometry. Among them were nine new chitin-binding peritrophic matrix proteins, expanding the list from three to twelve peritrophins. Lastly, we provide a model for the putative interactions among the proteins identified in this study.



## 1. Introduction

The peritrophic matrix (PM) is an extracellular envelope composed predominantly of chitin fibrils and glycoproteins which line the digestive tract of most invertebrates (reviewed by Devenport and Jacobs-Lorena, 2005; Shao et al., 2001; Tellam et al., 1999). The PM completely surrounds the food bolus, separating it from the midgut epithelium and acts as a molecular sieve that mediates the traffic of molecules (e.g., enzymes, metabolites and hydrolyzed digestion products) to and from the midgut lumen. In order to effectively control particle passage, the PM is predicted to be highly structured and dynamic, capable of manipulating its thickness and porosity to varying degrees (Peters, 1992; Shen and Jacobs-Lorena, 1997). The structure of the PM of hematophagous insects, such as mosquitoes, are of special interest since it has been shown to have an additional role, that of limiting *Plasmodium* establishment in the mosquito midgut.

There are two types of mosquito PM (1 and 2) that have very different properties (reviewed by Devenport and Jacobs-Lorena, 2005). Adult mosquitoes do not have a PM prior to a blood meal. The dramatic distention of the midgut epithelium caused by blood ingestion induces the secretion by most posterior midgut epithelial cells of a relatively thick (~20 μm) type 1 PM. In contrast, larval mosquitoes have a type 2 PM. This thin (<1 μm) tubular sleeve lines the entire midgut and hindgut of the mosquito larvae and is constitutively secreted by a small group of cells at the foregut/midgut junction called the cardia. Adult mosquito type 1 PMs are secreted immediately into the lumen of the gut in response to feeding. In contrast, the larval mosquito type 2 PM is constitutively synthesized, independent of food ingestion. The best evidence for a role of the type 1 PM as a barrier to pathogen invasion comes from the study of the *Plasmodium* life cycle in the mosquito. Transmission of *Plasmodium* is initiated when mosquitoes ingest an infected blood meal. Within the blood mass, the *Plasmodium* gametes mate and develop into motile ookinetes, a process that takes 16–24 h depending on the *Plasmodium* species. To traverse the gut epithelium, the ookinetes must first cross the PM which has fully matured by this time (Fig. 1). Thus, the PM constitutes a potential barrier for *Plasmodium* invasion.

To traverse the PM, the *Plasmodium* parasite secretes its own chitinase, which allows it to literally "punch a hole" through this chitin-containing structure (Huber et al., 1991). The release of the *Plasmodium* chitinase may be triggered by an interaction of the parasite with PM proteins. Interestingly, the chitinase of certain

* Corresponding author.
*E-mail address:* mlorena@jhsph.edu (M. Jacobs-Lorena).
[1] The authors contributed equally to this work.
[2] Present address: The Stowers Institute for Medical Research, 1000 E. 50th Street, Kansas City, MO 64110, USA.
[3] Present address: Department of Protein Chemistry, Genentech, Inc., South San Francisco, CA 94080, USA.

0965-1748/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ibmb.2008.10.010

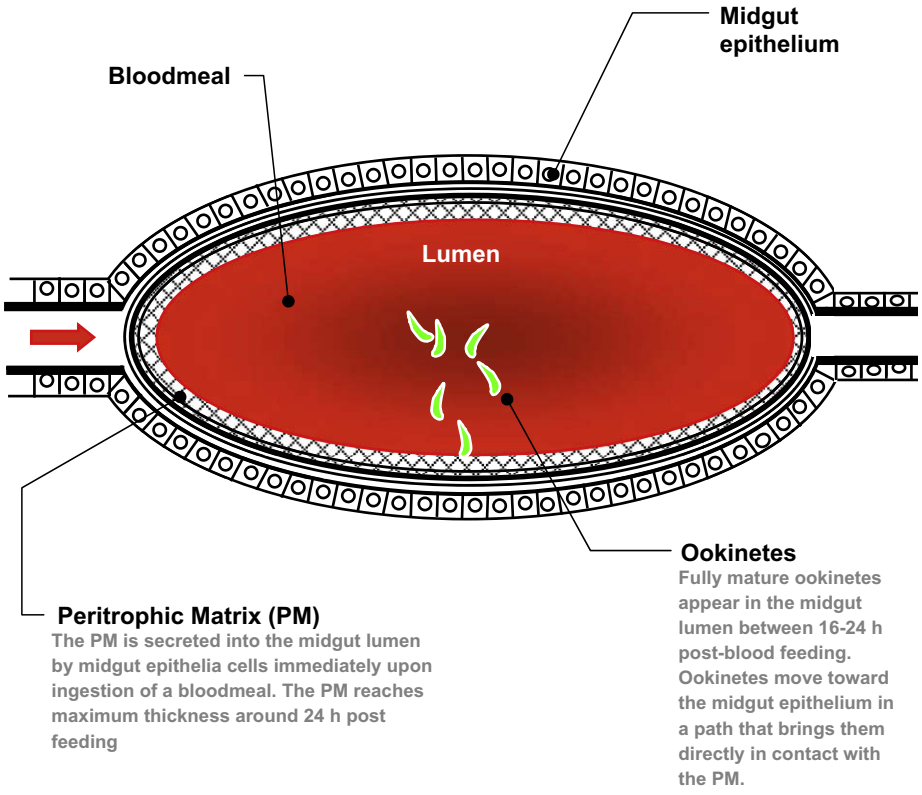


**Fig. 1.** The adult *Anopheles gambiae* peritrophic matrix (PM). Blood feeding (arrow shows blood entering from the foregut and anterior midgut) induces the secretion of the PM into the lumen by midgut epithelial cells lining the gut. The PM (hatched area) encloses the blood meal bolus and protects the midgut epithelial cells from microbes, heme and other toxins. The *Plasmodium* parasite ookinete stage develops within the blood meal and must traverse the PM in order to access and cross the midgut epithelial cells.

*Plasmodium* species is secreted as a pro-enzyme with little enzymatic activity. Activation of the pro-enzyme depends upon the action of trypsins which are secreted by the mosquito host for blood digestion. Blocking mosquito trypsin activity with inhibitors, or with an anti-trypsin antibody, prevents *Plasmodium gallinaceum* from penetrating the PM (Shahabuddin et al., 1993,1996). Further evidence that the mosquito PM constitutes a barrier for *Plasmodium* invasion was obtained by Billingsley and Rudin (1992) who showed that thickening of the PM by experimental manipulation resulted in decreased *Plasmodium* infectivity. Complementary work showed that by inhibiting the parasite chitinase by antibody or enzymatic inhibitors results in a massive reduction in parasite invasion of the midgut (Li et al., 2004; Langer et al., 2002a,b). If PM proteins play a role in chitinase secretion by the parasite, then antibodies to these PM proteins might also interfere with the parasite crossing of the PM.

Two-dimensional gel electrophoresis analysis revealed that the black fly type 1 PM has only 2 major proteins (Ramos et al., 1994), that the *Aedes aegypti* PM has about 20 major proteins and that the *Anopheles gambiae* PM has about 40 major proteins (Moskalyk et al., 1996). The first mosquito type 1 PM protein gene to be cloned is *AgAper1* (Shen and Jacobs-Lorena, 1998). *AgAper1* has two chitin-binding domains and is believed to function as a molecular cross-linker of chitin fibrils. We predicted initially that chitin fibrils and proteins with chitin-binding domains are the most abundant molecules since they contribute directly to the chitin scaffold, i.e., PM structure. While a consensus existed about the role of the PM in limiting ookinete invasion, we had also recently observed that the PM has an unsuspected role in protecting the gut from the toxic effects of heme resulting from blood digestion. The mosquito ingests more than its own weight of blood and all of the hemoglobin is digested during the course of the next 24–48 h releasing large amounts of heme. Heme generates free radicals and is highly toxic to cells in part by damaging membrane lipids. We found that the PM (that completely surrounds the blood meal) plays a protective role against heme damage by first demonstrating that it can bind approximately the same amount of heme as that released during digestion of a typical blood meal (Pascoa et al., 2002). Subsequently, we identified a specific *Ae. aegypti* PM protein that efficiently sequesters heme. Surprisingly, the heme-binding domain overlaps with the chitin-binding domain (Devenport et al., 2006). Therefore, we suspected that other PM proteins containing chitin-binding domains possess similar properties and that not all predicted chitin-binding domains can actually bind chitin (Shi and Paskewitz, 2004) or has limited binding affinity to chitin alone. Additional evidence suggests that the PM contains a few glycoproteins with mucin-like domains (Devenport, unpublished data). These mucin domains are rich in Pro–Ser/Thr motifs that are predicted sequons for O-linked glycosylation. Mucin O-glycans are short glycan modifications that help retain water and therefore act in lubricating surfaces and protecting tissues. In order to get a comprehensive understanding of PM structure, we sought to characterize the protein components of the PM and subsequently, propose a model of how the PM proteins are assembled and interact with malaria parasites.

## 2. Methods

### 2.1. Preparation of PM samples for analysis

PMs were dissected from adult female *An. gambiae* that were fed a protein-free latex meal. Protein-free meals consisted of 16% v/v latex beads (LB-5, Sigma; Billingsley and Rudin, 1992) suspended in 150 mM NaCl, 10 mM $NaHCO_3$, pH 7.0 (Moskalyk et al., 1996) with 1 mM ATP added as a phagostimulant (Galun et al., 1963; Moskalyk et al., 1996; Billingsley and Rudin, 1992). This would induce the secretion of the thickest PM using a protein-free meal (Moskalyk

et al., 1996). Intact PMs (including the enclosed latex beads) were dissected at 8 h post-feeding in 50% ethanol in PBS and transferred to a 1.5 ml eppendorf tube in a dry ice methanol bath. PMs were stored at −80 °C until a total of approximately 750 PMs were accumulated. Extraction buffer was added to the sample and homogenized with a plastic pestle and vortexed. Approximately 750 PMs were dissected and extracted sequentially in 250 μl buffer containing either 50 mM Tris pH 8.5 (Buffer A), Buffer A + 0.5% Triton X-100 (Buffer B), or Buffer A + 2% SDS (Buffer C). The initial Tris fraction was quantified by Bradford assay (0.44 μg/μl). The samples were incubated on ice for 1.5 h and centrifuged in a microcentrifuge at maximum speed. The supernatant (containing the extracted proteins) was removed for subsequent SDS-PAGE analysis. For the sequential extractions, either Buffer B or Buffer C was added to the pellet fraction and the extraction procedure above was repeated. Fractions were loaded onto the gel (~2 μg/lane).

### 2.2. MudPIT analysis

Proteins that were sequentially extracted from *An. gambiae* peritrophix matrix with (i) Buffer A ("Tris" fraction); (ii) Buffer B ("Tris–Triton" fraction); or (iii) Buffer C ("Tris–SDS" fraction) were processed as follows. The leftover pellets were washed twice with an excess of Buffer A to wash away detergent and analyzed as well ("Pellet" fraction). No unique proteins were identified in the pellet and this fraction is not considered further. Two independent batches of peritrophic matrix extractions were digested differently, one with trypsin and one with a non-specific enzyme, proteinase K.

The first batch of "Tris", "Tris–Triton", and "Tris–SDS" fractions were digested following the high pH proteinase K procedure (Wu et al., 2003). Briefly, solid urea was added to 8 M to the protein samples (vol. *ca.* 200 μl), followed by 1 M TCEP (Tris (2-Carboxyethyl)-Phosphine Hydrochloride, Pierce) to 5 mM at room temperature for 30 min, and 0.5 M IAM (Iodoacetamide, Sigma) to 20 mM at room temperature for 30 min, in the dark. pH was adjusted to 11.5 with 30% ammonium hydroxide, then 0.25 μg of proteinase K (Roche) was added for 4 h at 37 °C. The reaction was quenched by adding formic acid to 5%.

The second batch of "Tris", "Tris–Triton", and "Tris–SDS" fractions were TCA-precipitated, and protein pellets were solubilized in 30 μl 100 mM Tris–HCl, pH 8.5, 8 M urea, reduced with 5 mM TCEP and alkylated with 10 mM IAM. As described in (Wolters et al., 2001), endoproteinase Lys-C (Roche) was added to 0.5 μg at 37 °C overnight, then the samples were diluted to 2 M urea with 100 mM Tris–HCl, pH 8.5. Calcium chloride was added to 2 mM and the digestion with trypsin (0.5 μg) was let to proceed overnight at 37 °C while shaking. The digestion was quenched by adding formic acid to 5%.

"Pellets" from both peritrophic matrix batches were dried under vacuum, solubilized in 100 μl cyanogen bromide at 500 mg/ml in 88% formic acid, and left in the fume hood overnight in the dark (Wolters et al., 2001). The samples were neutralized by adding 30% ammonium hydroxide drop by drop. The pH was adjusted to 8.5 by adding 1 M Tris–HCl to 100 mM. The samples were denatured with 8 M urea, reduced with 5 mM TCEP, and alkylated with 20 mM IAM, then digested with endoproteinase Lys-C and trypsin as described above.

Peptide mixtures were loaded onto 250 μm fused silica microcapillaries, which had a filtered union on one end (Upchurch) and were packed with 4 cm of 5-μm strong cation exchange material (Partisphere SCX, Whatman), followed by 2 cm of 5-μm $C_{18}$ reverse phase (Aqua, Phenomenex). The loaded 250 μm microcapillaries were connected using a filtered union to a 100 μm fused silica column, with a tip pulled to a 5 μm opening and packed with 9–10 cm of reverse phase material (Florens and Washburn, 2006).

The loaded columns were placed in-line with a Quaternary Agilent 1100 series HPLC pump. Overflow tubing was used to decrease the flow rate from 0.1 ml/min to about 200–300 nl/min. Fully automated 10–12 step chromatography runs were carried out (Florens and Washburn, 2006). Three different elution buffers were used: 5% acetonitrile, 0.1% formic acid (Buffer A); 80% acetonitrile, 0.1% formic acid (Buffer B); and 0.5 M ammonium acetate, 5% acetonitrile, 0.1% formic acid (Buffer C). Peptides were sequentially eluted from the SCX resin to the reverse phase resin by increasing salt steps, followed by an organic gradient. The application of a 2.5 kV distal voltage electrosprayed the eluting peptides directly into a Deca ion trap mass spectrometer equipped with a nano-LC electrospray ionization source (ThermoFinnigan). Full MS spectra were recorded on the peptides over a 400 to 1600 $m/z$ range, followed by three tandem mass (MS/MS) events sequentially generated in a data-dependent manner on the first, second, and third most intense ions selected from the full MS spectrum (at 35% collision energy). Mass spectrometer scan functions and HPLC solvent gradients were controlled by the Xcalibur data system (ThermoFinnigan).

SEQUEST (Eng et al., 1994) was used to match MS/MS spectra to peptides in a database consisting of 38318 *An. gambiae* proteins downloaded from ENSEMBL in January 2007 (http://www.ensembl.org/info/data/download.html: Anopheles_gambiae.AgamP3.42.pep.abinitio.fa [genscan and snap predictions from Ag genome] and Anopheles_gambiae.AgamP3.42.pep.all.fa.gz [known or novel gene predictions]). To estimate false discovery rates, each protein entry was "reversed" and added to the forward sequences, leading to a final search space of 76636 amino acid sequences. The validity of peptide/spectrum matches was assessed using the SEQUEST-defined parameters, cross-correlation score (XCorr) and normalized difference in cross-correlation scores (DeltCn). Spectra/peptide matches were only retained if they had a DeltCn of at least 0.08 for the CNBr and/or trypsin digests or at least 0.15 for the proteinase K digests (Zybailov et al., 2005), a maximum Sp rank of 10, and a minimum XCorr of 1.8 for singly, 2.0 for doubly, and 3.0 for triply charged spectra. In addition, the peptides had to be at least 7 amino acids long and their ends had to conform to the digestion(s)/cleavage used when applicable. Combining all runs, proteins had to be detected by at least 2 such peptides, or 1 peptide with 2 independent spectra, and proteins that were subset of others were removed. DTASelect and CONTRAST (Tabb et al., 2002) was used to select, sort, and compare peptide/spectrum matches passing this criteria set. To estimate relative protein levels, spectral counts were normalized (Zybailov et al., 2006) using an in-house developed script (*contrast-report-add-nsaf*): for each protein $k$ detected in a particular MudPIT analysis, Normalized Spectral Abundance Factors (NSAFs) were calculated as follows:

$$(\mathrm{NSAF})_k = \frac{(\mathrm{SpC}/\mathrm{Length})_k}{\sum_{i=1}^{N}(\mathrm{SpC}/\mathrm{Length})_i}$$

Protein families: all 36532 arthropoda proteins from SwissProt and SPTREMBL (inc. whole predicted *An. gambiae* proteome) clustered into families using TribeMCL algorithm (Enright et al., 2002, NAR 30). Accounts for multi-domain structure of proteins 10461 families in total 7511 contain anopheles ENSEMBL predicted peptides 5661/7511 contain only one anopheles peptide. Conserved domains were identified by BLAST and parallel analysis using prediction algorithms for N-glycan and O-glycan modification sites (Expasy). A putative mucin domain is characterized by multiple Pro–Ser/Thr repeats.

### 2.3. RT-PCR analysis

The following RNA samples were examined for the presence or absence of transcripts from genes encoding the hypothetical proteins identified by mass spectrometry. Total RNA was extracted from larval guts, larval carcasses (all tissues minus gut), and midguts from sugar-fed adult female *An. gambiae*, blood-fed *An.*

*gambiae* and carcasses from blood-fed females. The cDNA template from each tissue sample was synthesized according to the manufacturer's instruction (Invitrogen). Genomic DNA was used as a positive control in the RT-PCR experiments. Primers used in this analysis are found in Supplementary Table S4.

## 3. Results and discussion

### 3.1. Proteomic analysis of the PM

*An. gambiae* PMs were dissected and the proteins fractionated according to solubility in buffers and detergents (Fig. 2). These fractions were then subjected to mass spectrometric analysis. Analysis of the three fractions identified a total of 209 non-redundant proteins across the different samples (Fig. 3, Supplementary Tables S1–3) that produced at least two identifying peptides and had predicted annotations on the ENSEMBL database (version 46, August 2007). The results can be further summarized as follows: 30 were hypothetical proteins of unknown function, and 28 proteins have glycan-binding motifs, of which, 12 proteins had predicted chitin-binding domains (CBDs) (Fig. 3a, b). Not surprisingly, extracellular matrix associated, glycan-binding proteins, peptidases and proteins of unknown function are relatively more abundant than proteins involved in lipid or energy metabolism and membrane transport.

Of the 209 proteins, 123 had predicted signal peptides of which 17 also had predicted transmembrane domains and are therefore candidate midgut surface proteins rather than PM proteins that are intimately intercalated with the PM chitin fibrils but were co-extracted from the gut during sample processing (Fig. 3c, d). Of those with signal peptide sequences, 64 proteins are predicted to be non-membrane bound (i.e., soluble), which would include the putative PM proteins as well as secreted peptidases. As expected, the number of proteins with signal peptides is relatively more abundant than those with transmembrane domains only. Based on their annotations, the proteins fell into different classes, with diverse putative functions ranging from immunity to blood digestion (Fig. 3a). Only 5% of the detected proteins were either known PM proteins or contained clear CBDs. A considerable number of proteins were hypothetical and of unknown function (15%), but corresponded to known ESTs. A significant proportion of the proteins with unknown function may correspond to true PM proteins, given that this structure is poorly characterized molecularly in any organism. Considering only those proteins with a predicted signal peptide sequence ($N = 123$), approximately 45% were either putative proteases or glycoside hydrolases. Given the relatively simple make up of the PM (estimated ~40 proteins; Moskalyk et al., 1996) this large number of identified proteins was at first surprising. We attribute it to two factors: (1) the sensitivity of the mass spectrometry detection system and (2) the fact that the microvilli of the gut epithelial cells are probably embedded into the PM at the time of dissection. This may have resulted in the inclusion of epithelial cell surface and even cytoplasmic proteins in the PM samples.

Mass spectrometry detected 178 proteins following extraction of the PM using Tris buffer alone (Fig. 2b, lane 1, Supplementary Table S1–3). This was surprising since extraction with low ionic strength buffers would presumably reveal those proteins which are only loosely associated with the PM or categorized as "Class I PM proteins" (Elvin et al., 1996). Following this categorization scheme, those proteins which can only be extracted using buffer supplemented with detergents would be considered "integral PM proteins". These definitions cannot be precise since it is difficult to determine if proteins that are "loosely" associated (i.e., Class I PM proteins) with the PM are true PM components or contaminants emanating from gut proteins that have been secreted into the gut lumen during feeding (Elvin et al., 1996; Sarauer et al., 2003). In our experiments, it is difficult to differentiate according to these classifications since MS analysis was successful in detecting minute amounts of protein. However, based on the staining patterns (Fig. 2b) observed for each of the individual fractions, several characteristics become evident. Similar to other insect PMs that have been studied to date, the *An.*

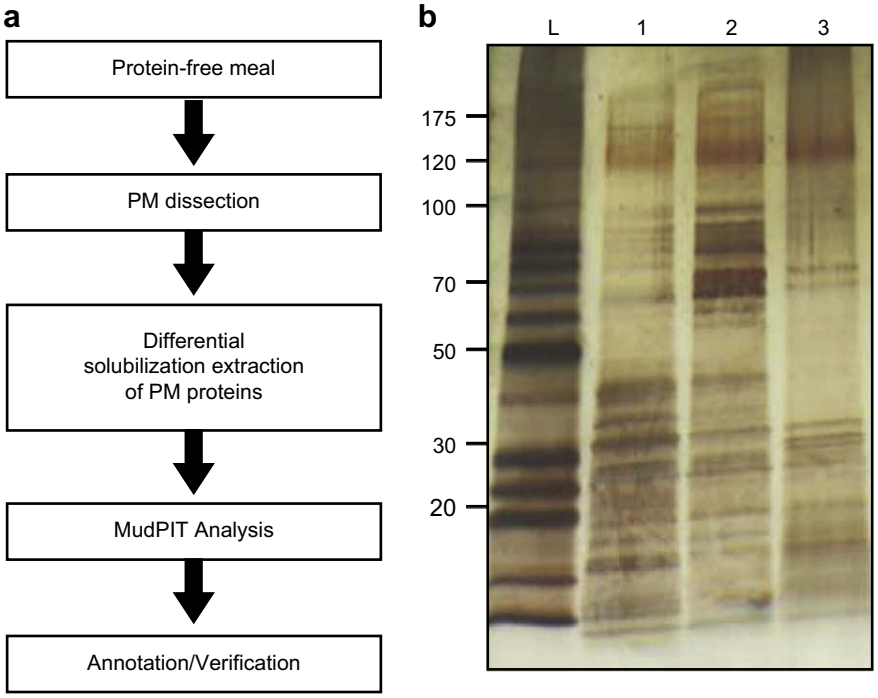


**Fig. 2.** Analysis of *An. gambiae* PM proteins. (a) Experimental flow chart. A protein-free meal was fed to starved female *An. gambiae* mosquitoes to produce a PM without non-mosquito protein contamination. MudPIT, Multidimensional Protein Identification Technology (i.e., liquid chromatography tandem mass spectrometry, LC-MS/MS). (b) Gel electrophoretic analysis of PM proteins (silver stained). L: size markers (kDA, shown on the left). The other lanes show proteins eluted sequentially with Tris buffer (lane 1), with Tris + 0.5% Triton X-100 (lane 2) and with Tris + 2% SDS (lane 3).

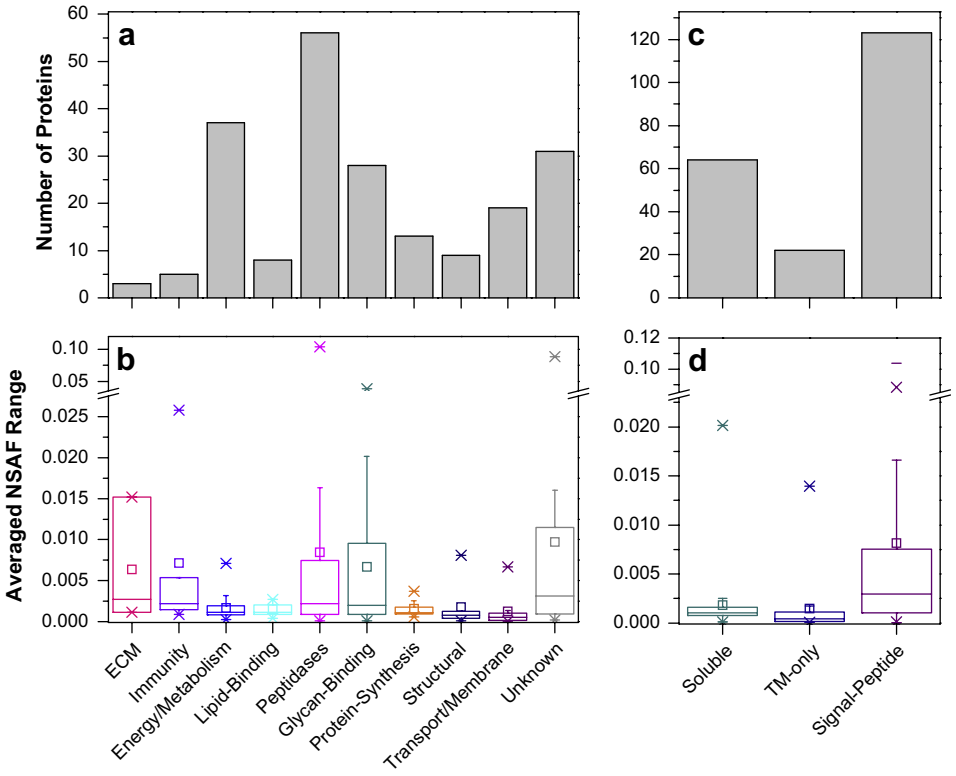


**Fig. 3.** The *An. gambiae* type 1 PM proteome. (a–b) Categorization of 209 proteins identified by mass spectrometry according to their putative function and relative abundance (NSAF values). ECM, extracellular matrix. (c–d) Categorization of identified proteins according to predicted structural properties such as the presence of a classical signal peptide (SP), transmembrane domains without signal peptide (TM only), and soluble proteins (without predicted SP or TM). The summary statistics are shown for averaged NSAF values for each functional (b) or structural (d) category. In the box plots, the 25th and 75th percentile are represented by the upper and lower boundaries of the box, with the median being the line dissecting the box, and the mean being the small square symbol. The 5th and 95th percentiles are shown as errors bars, the 'X' represents the 1st and 99th percentiles, and the stand alone dashes '–' represent the complete range. The number of total protein identifications in each category is shown in (a) and (c).

*gambiae* adult PM is composed of proteins ranging in size between <10 kDa and >200 kDa (Fig. 2b). There are several prominent protein bands that were observed from the Tris fraction, between 15 and 40 kDa, whereas in the detergent fractions, the more distinct bands correspond to proteins >70 kDa. This may suggest that higher molecular weight proteins, which required detergent extraction, may form the structural and as such, integral components of the PM. However, these larger MW proteins are not necessarily the most abundant species in the PM (Table 1). In fact, excluding the two proteases, chymotrypsin and trypsin, as well as mosquito chitinase, the other abundant proteins are relatively small, <35 kDa. AgAper1 (AGAP006795, Shen and Jacobs-Lorena, 1998) was found to be the most abundant chitin-binding domain (CBD) containing protein or peritrophin. No other peritrophins were among the top 10 abundant proteins across all our samples. The most abundant proteins in our sample included the peptidases and conserved proteins that did not contain CBDs and of unknown function.

### 3.2. Characterization of the PM genes

Several PM genes (the peritrophins) containing CBDs have been characterized and there is partial knowledge about their presumed

**Table 1**
Most abundant proteins identified by mass spectrometry from the *Anopheles gambiae* peritrophic matrix[a].

| Rank | ENSEMBL accession number[b] | Predicted kDA, $M_r$ | Annotation/comments/references[c] |
|---|---|---|---|
| 1 | AGAP001199 | 29.2 | O-glycosylated; SP; chymotrypsin (Dimopoulos et al., 1996) |
| 2 | AGAP000570 | 18.1 | Conserved hypothetical protein; aglycosylated; SP, UF |
| 3 | AGAP008296 | 29.1 | Trypsin-1; O-glycosylated[d]; SP |
| 4 | AGAP007745 | 34.1 | Conserved hypothetical protein; proline-rich extension domain, heavily glycosylated, SP, UF |
| 5 | AGAP006414 | 57.0 | 1 CBD; N-glycosylated and heavily O-glycosylated; mucin domain; SP; chitinase (Shen and Jacobs-Lorena, 1998) |
| 6 | AGAP009313 | 26.3 | Conserved hypothetical protein; similar to a chemotactic protein; SP, UF |
| 7 | AGAP006795 | 16.6 | 2 CBD, N- and O-glycosylated; SP; AgAper1 (Devenport et al., 2004) |
| 8 | AGAP009593 | 49.1 | Zinc carboxypeptidase A 1; N-glycosylated; SP |
| 9 | ENSANGESTP00000003082 SNAP_ANOPHELES00000018835 | 54.8 | Conserved hypothetical protein; heavily glycosylated; 1 of 2 putative protein products of ENSANGESTG00000002412; no SP, UF |
| 10 | AGAP006194 | 21.5 | Conserved hypothetical protein; aglycosylated; putative paralogue of AGAP006195; SP, UF |

[a] Mass spectrometry data resulting from all PM extraction approaches were analyzed and compared to generate the table based on detection across all samples and ranked according to Normalized Spectral Abundance Factors (NSAF) values. See Supplementary Tables S1–3 for complete information.
[b] ENSEMBL mosquito database, version 48, January 2008 (current accession numbers).
[c] SP, predicted signal peptide; CBD, chitin-binding domain; UF, unknown function. *N*- and *O*-glycosylation status is based on prediction algorithms (CBS Prediction Servers).
[d] Trypsin has not been shown to be glycosylated but has predicted *O*-glycosylation sites (see text for details about other serine proteases).

**Table 2**
Putative peritrophic matrix proteins identified by mass spectrometry with predicted chitin-binding domains.

| ENSEMBL accession number[a] | Protein ID | Predicted kDa, $M_r$ | Annotation/comments/references[b] |
|---|---|---|---|
| AGAP006795 | AgAper1 | 16.6 | 2 CBD, N- and O-glycosylated; SP; AgAper1 (Devenport et al., 2004) |
| AGAP006796 | AgAper9 | 9.0 | 1 CBD; aglycosylated; short sequence; SP |
| AGAP009830 | AgAper14 | 10.1 | 1 CBD; aglycosylated; SP; apparent $M_r$ of 14 kDa (Devenport et al., 2005) |
| AGAP010364 | AgAper25a | 25.0 | 2 CBD; O-glycosylated; No SP |
| AGAP001819 | AgAper25b | 25.8 | 1 CBD (domain 3 of baculovirus spindolin protein); N-glycosylated; SP |
| AGAP011615 | AgAper26 | 26.3 | 3 CBD; N-glycosylated; SP |
| AGAP010363 | AgAper29 | 28.9 | 2 CBD; N- and O-glycosylated; SP |
| AGAP011616 | AgAper30 | 30.0 | 3 CBD; aglycosylated; SP |
| AGAP006433 | AgAper34 | 33.6 | 4-CBD; N- and O-glycosylated; No SP |
| AGAP006434 | AgAper57 | 57.2 | 4-CBD; O-glycosylated and heavily N-glycosylated; SP |
| AGAP006432 | AgICHIT | 40.6 | 2 CBD; Pro-rich; N-glycosylated, heavily O-glycosylated; mucin domain; SP; immune-responsive gene (Dimopoulos et al., 1998) |
| AGAP006414 | AgChitinase | 57.0 | 1 CBD; N-glycosylated and heavily O-glycosylated; mucin domain; SP; chitinase (Shen and Jacobs-Lorena, 1998) |

[a] ENSEMBL, *An. gambiae* mosquito database, version 48, January 2008 (current accession numbers). The protein ID designations are based on predicted molecular mass except where indicated.
[b] CBD, chitin-binding domain(s); SP, signal peptide.

function: binding to chitin fibers and binding heme (Shen and Jacobs-Lorena, 1998; Devenport and Jacobs-Lorena, 2005; Tellam et al., 1999). Although it should be noted that proteins with predicted CBDs do not necessarily indicate functional chitin-binding activity (Shi and Paskewitz, 2004) or a direct role in peritrophic matrix structure (He et al., 2007), for the purposes of this discussion on PM structure, we will maintain that genes that are expressed specifically in the gut and contain putative CBD have primary functions as peritrophins. Proteomic analysis of the PM has expanded the list of CBD containing PM proteins from *An. gambiae* from three *(AgAper1*, *AgAper14*, *AgICHIT*) to 12 (Table 2).

### 3.3. PM proteins with chitin-binding domains

We analyzed the chitin-binding domains of several invertebrate proteins, including chitinases and PM proteins, and found that they share significant amino acid sequence similarity (data not shown). There are 87 predicted proteins with CBD domains (corresponding to 77 genes) in the *An. gambiae* genome. The data suggest that these domains evolved from a common ancestor (Shen and Jacobs-Lorena, 1999) and that duplication and transposition of these domains contributed to the functional diversification of chitin-binding domains. Moreover, invertebrate and plant chitin-binding domains do not share significant amino acid sequence similarity, suggesting that they are not co-ancestral. However, both invertebrate and plant domains are cysteine-rich and have several highly conserved aromatic residues. These observations led to the proposal that invertebrate and plant chitin-binding domains evolved by convergent evolution and that the disulfide bonds and aromatic residues are hallmarks for glycan-binding proteins (Shen and Jacobs-Lorena, 1999). Chitin-binding domains are also present in the mosquito cuticle (He et al., 2007) and in proteins that are members of the 6- and 8-Cys domain family of proteins, which includes macromolecules with lectin-like and non-lectin activities (Dinglasan and Jacobs-Lorena, 2005).

A key criterion for the assignment of *bona fide* PM proteins is midgut tissue-specific expression of the candidate genes. We determined the presence in larval and adult mosquito stages of several transcripts encoding CBD containing proteins identified by MS (Table 2). As expected, several of these genes were found to be expressed in adult mosquito midguts (either sugar-fed or blood-fed) with varying levels of expression (Fig. 4). Of these genes, we observed that AgAper14 (Devenport et al., 2005) was, as predicted, strongly expressed in the larval gut, with lesser expression in the larval carcass (data not shown). AgAper9, AgAper26 and AgAper30 also exhibited low level but tissue-specific expression in the larval gut. The *An. gambiae* database (January 2008, v.48) has AgAper26 (AGAP011615) and AgAper30 (AGAP011616) annotated as two distinct ORFs. However, we found that one set of primers was successful in amplifying from cDNA a product that spans the two ORFS (*AgAper26* and *AgAper30*), suggesting that these two genes give rise to a single transcript (Fig. 4). We have also previously isolated a partial cDNA from larval cardia which we had originally ascribed to be involved in larval PM structure. This cardia protein corresponds to AgAper14 in our study. AgAper57, AgICHIT and AgChitinase have restricted expression to the adult midgut. AgAper57 exhibits significant homology to *Ae. aegypti* AEIMUC1 (Devenport et al., 2005) which is also similar to the *Ae. aegypti* peritrophin AeAper50 in domain/protein architecture (Shao et al., 2005). AeAper50 protein has five CBDs and intervening mucin domains that are found in other insect PMs. Like AeAper50, AgAper57 also contains putative Pro-rich extensin-like motifs (Wu et al., 2001) which provide tensile strength under mechanical stress to tissues, e.g., during extension of the gut while blood feeding, and is predicted to contain two N-glycan modification sites and multiple *O*-glycosylation sites (data not shown).

It also appears that several peritrophins are expressed in non-gut tissues (i.e., carcass). Most prominent of these genes were *AgAper25a*, *AgAper30*, *AgICHIT*, and *AgChitinase*. AgAper14 also showed expression in the carcass (data not shown). The varying expression of these genes in sugar- or blood-fed midguts may suggest tissue-specific differences in transcriptional and/or global post-transcriptional control. However, since the data do not

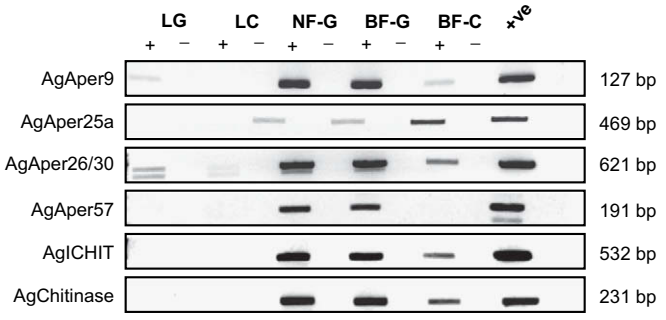


**Fig. 4.** Tissue-specific expression of peritrophins (chitin-binding domain containing proteins). Reverse-transcriptase polymerase chain reaction (RT-PCR) analysis of genes that encode putative chitin-binding proteins (described in Table 2). The tissues that were examined are as follows: LG, larval midgut; LC, larval carcass (all tissues minus midgut); NF-G, midguts of sugar-fed adults; BF-G, midguts of blood-fed adults; and BF-C, carcasses of blood-fed minus midgut. Amplicons were generated from a plasmid clone of the corresponding gene as a positive control (+ve). The amplicon size (base pairs, bp) is shown on the right. The (+) indicates the addition of reverse transcriptase and (−) indicates the absence of enzyme.

**Table 3**
Proteins identified by mass spectrometry from the peritrophic matrix with no predicted chitin-binding domains.

| ENSEMBL accession number[a] | Predicted kDa, $M_r$ | Annotation/comments[b] |
|---|---|---|
| AGAP006442 | 12.9 | Conserved hypothetical protein; heavily N-glycosylated; SP, UF |
| AGAP004883 | 19.7 | Conserved hypothetical protein; snake toxin-like protein folds/disulfide rich; Thr-rich domains; both heavily N- and O-glycosylated; SP, UF |
| AGAP007860 | 33.8 | Conserved hypothetical protein; putative protein binding motifs; N-glycosylated; SP, UF |
| AGAP007612 | 92.2 | Conserved hypothetical protein; snake toxin-like protein folds/disulfide rich; aglycosylated; SP, UF |
| AGAP002851 | 16.3 | Conserved hypothetical protein; MD2-lipid recognition domain; aglycosylated; SP |
| AGAP001352 | 28.1 | Conserved hypothetical protein; odorant/hormone binding domain; O-glycosylated; SP |
| AGAP010132 | 52.1 | Conserved hypothetical protein; Croquemort (CD36) scavenger receptor Class B domain; N-glycosylated; SP |
| AGAP006398 | 31.2 | Conserved hypothetical protein; galactose-like binding protein/lectin-like domain; N- and O-glycosylated; No SP |
| AGAP004916 | 35.0 | Conserved hypothetical protein; Fibrillin/fibrinogen-like; globular domain (α/β/γ chains); N-glycosylated; SP, UF |

[a] ENSEMBL, *An. gambiae* mosquito database, version 48, January 2008 (current accession numbers). The remaining eight highly abundant non-CBD proteins are included in Table 1. The proteins were ranked based on their abundance rank according to Normalized Spectral Abundance Factors (NSAF) values. See Supplementary Tables S1–3 for complete information.
[b] SP, signal peptide; UF, unknown function.

represent normalized amplifications, we cannot directly compare relative abundance of transcript between tissues. Moreover, it should be noted that the apparent intensity of the AgAper25a amplicon in the gut does not correlate with the level of detectable protein in the PM sample. The midgut expression of *AgICHIT* has been shown previously to be an immune-responsive gene following the introduction of bacteria into adult mosquitoes (Dimopoulos et al., 1998). A proline-rich domain on AgICHIT may facilitate protein–protein aggregation (Williamson, 1994) as well. These data suggest a possible structural role for AgICHIT as glue between CBD and non-CBD containing proteins in the PM.

### 3.4. Non-chitin-binding proteins

We were also particularly interested in exploring the properties and function of novel classes of PM proteins (i.e., do not contain CBDs). We hypothesized that a subset of these genes may be involved in protein–protein interactions and in the formation of the three-dimensional structure of the PM. Our criteria for identification of candidate adult PM proteins includes the presence of a signal peptide, the absence of a transmembrane domain or GPI anchor, and adult midgut specific gene expression. Using these criteria, we identified 31 additional candidate PM proteins (containing a SP but no TMD, CBD or peptidase signature) which brings the total number of putative PM proteins identified in our analysis to 45, which is close to the ~40 proteins in our initial estimate (Moskalyk et al., 1996).

A total of 14/31 genes (Tables 1 and 3) were analyzed successfully by RT-PCR (Fig. 5). Transcripts of one gene (AGAP006398) appears to be of low abundance in the larval gut and carcass as well as sugar-fed adult guts but is not expressed in blood-fed guts or adult carcass. Six of the hypothetical proteins (AGAP000570, AGAP007745, AGAP009313, AGAP006194, and ENSANGESTP00000003082) were found to be highly abundant across all PM fractions (see Section 2) analyzed (Table 1). The most abundant PM protein, AGAP00570, is conserved in *Ae. aegypti* mosquitoes and is predicted to contain parallel beta-alpha-beta sheets but the function of this protein remains unknown. AGAP007745 is also a conserved hypothetical protein among mosquitoes and, like some of the peritrophins and CBD-binding proteins that have been discussed, contains a Pro-rich domain, an extensin-like motif and is predicted to be heavily glycosylated. The absence of a CBD in these 14 hypothetical proteins and their relative abundance suggests a role in protein–protein interactions in the structural framework of the PM.

Antibodies that were raised against PM proteins were also used to screen an *An. gambiae* midgut cDNA expression library (Lemos et al., 1996). A partial cDNA clone that was identified during this screen and was provisionally named PM9 corresponds to AGAP009313 in our current analysis. AGAP009313 is a conserved hypothetical protein that is also found in *Ae. aegypti*. The protein has predicted coiled-coil domain and interestingly, a chemotaxis methyl-accepting protein/histidine kinase domain more commonly present in chemical transducer molecules in bacteria (Kondoh et al., 1979; Hazelbauer et al., 2008). The function of this highly abundant protein is unclear but since BLAST searches did not match any bacterial proteins, it is likely that this protein has evolved similar domain architecture but not necessarily the same function as what has been described for bacteria. For example, the absence of a transmembrane domain on AGAP009313 excludes the possibility of a role in signal transduction, characteristic of the bacterial molecule.

What should be noted is that these transcripts are produced in blood-induced PMs whereas the proteomic data stem from latex bead-induced PMs. Although it has been argued that there are structural differences between latex-induced and blood-induced PMs (Billingsley and Rudin, 1992; Moskalyk et al., 1996), our RT-PCR

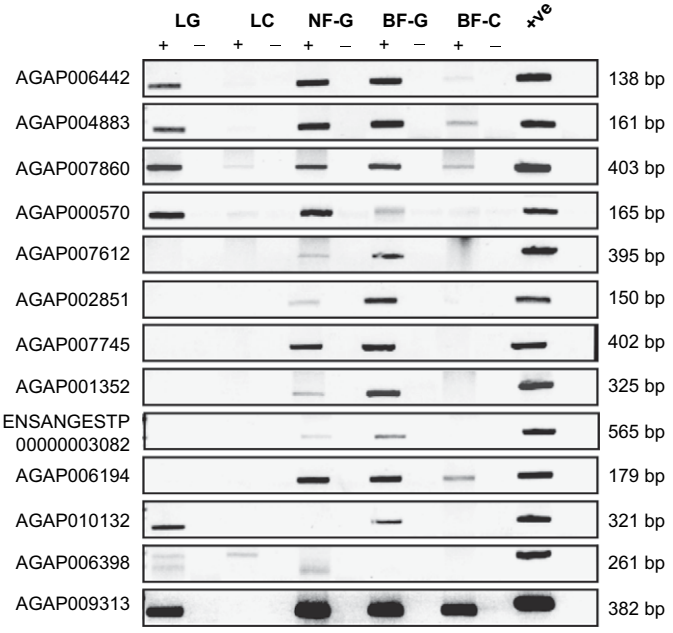


**Fig. 5.** Tissue-specific expression of putative secreted proteins that do not contain a chitin-binding domain. Reverse-transcriptase polymerase chain reaction (RT-PCR) analysis of genes that encode for conserved hypothetical proteins that do not contain CBDs (described in Table 3). The tissues that were examined are as follows: LG, larval midgut; LC, larval carcass (all tissues minus midgut); NF-G, midguts of sugar-fed adults; BF-G, midguts of blood-fed adults; and BF-C, carcasses of blood-fed minus midgut. Amplicons were generated from a plasmid clone of the corresponding gene as a positive control (+ve). The amplicon size (base pairs, bp) is shown on the right. The (+) indicates the addition of reverse transcriptase and (−) indicates the absence of enzyme.

data suggest that the genes described in Figs. 4 and 5 are conserved between the two experimental manipulations.

A genomic survey revealed that the peritrophins clustered to distinct regions of the chromosome, so-called "PM production" clusters. Genome localization suggests at least seven PM production clusters. In contrast, the hypothetical proteins did not commonly fall into clusters by themselves nor did they, in general, fall into PM production clusters. One cluster containing AgICHIT, AgAper57, chitinase, also included AGAP006442. However, these seven clusters represent less than half of the hypothetical PM proteins that we have identified in this analysis, where a majority of the genes identified do not group together. It is not clear at present, how chromosomal localization regulates or modulates stage-specific PM gene transcription and translation.

### 3.5. Proteolytic enzymes

Several mosquito midgut enzymes were identified across the fractions analyzed in this study, including chymotrypsin and trypsin-1 and AgAPN1, an alanyl aminopeptidase (AGAP004809-PA) that was found to be an effective malaria transmission-blocking antigen (Dinglasan et al., 2007) (Table S1). While the protein-free latex meal used for these experiments is known to be a poor inducer of digestive enzyme expression (Billingsley and Rudin, 1992; Lemos et al., 1996), identification of these enzymes among the PM proteins is not too surprising given that base-level enzymatic activity is detectable in sugar-fed mosquitoes (Muller et al., 1993a, b; Edwards et al., 1997). It is likely that a blood-induced PM contains a higher proportion and possibly a different subset, of protease family members, which in turn contributes to the higher enzymatic activity observed in earlier studies (Billingsley and Rudin, 1992; Lemos et al., 1996).

AgAPN1 is a membrane bound, GPI-anchored glycoprotein on the surface microvilli. It is likely that during dissection of the PM from the gut epithelium, AgAPN1 comes off of the microvilli since it may be highly intertwined between the chitin fibrils of the PM. Alternatively, it is possible that the GPI anchor of AgAPN1 is cleaved during PM maturation and separation immediately following blood

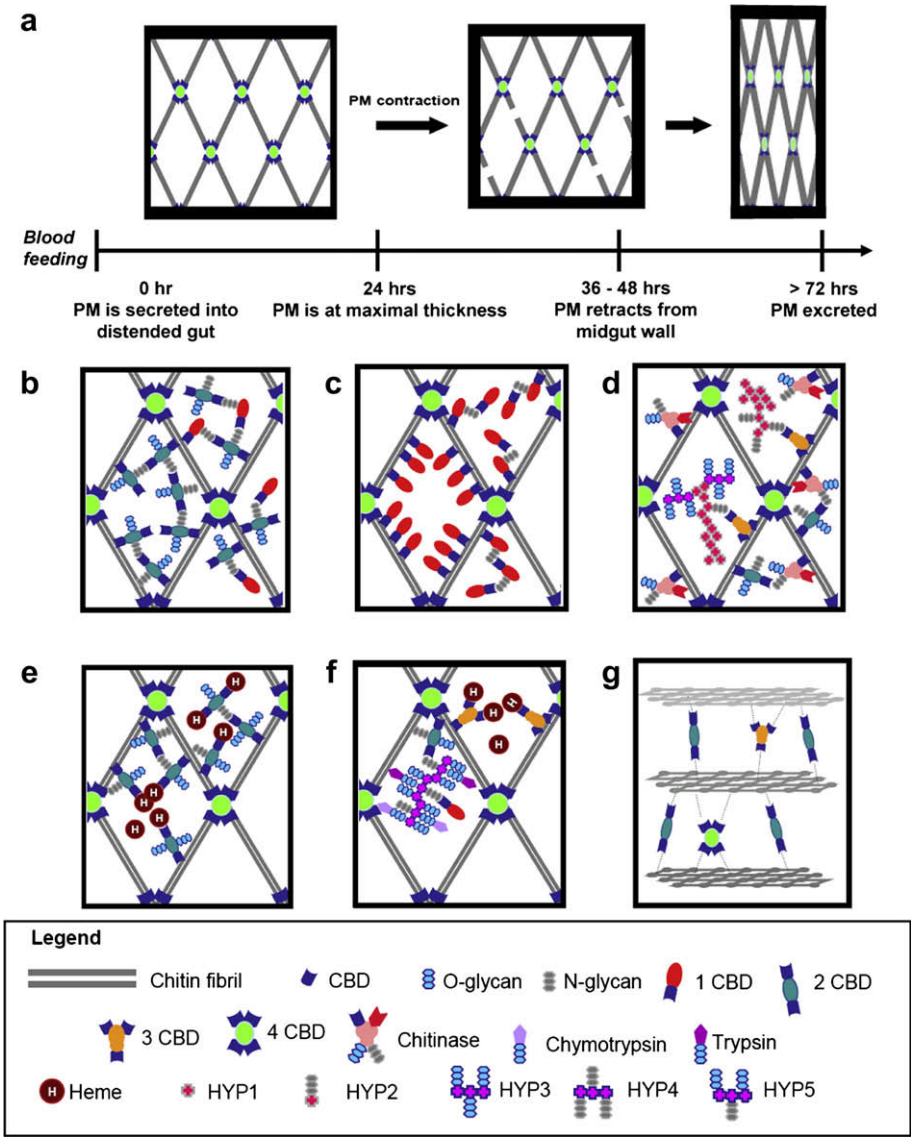


**Fig. 6.** Proposed structural model of the adult *An. gambiae* PM. (Panel a) It is speculated that intersecting chitin fibrils form cross-hatch apertures and provide the structural scaffold of the PM. Peritrophins with chitin-binding domains (CBD) interact with other peritrophins and digestive enzymes such as trypsin and chymotrypsin get caught in the PM during transit to the lumenal blood meal. Peritrophins also interact with chitinase and non-CBD hypothetical proteins (HYP1-5) as well as trapping toxic heme (Devenport et al., 2006) all within the apertures of the PM. Peritrophins and these other molecules strengthen the PM and provide structural support and control of porosity, contractility, and thickness. See text for more details.

feeding. What is unclear is if the interaction of AgAPN1 or other enzymes with the PM is a result of a non-covalent interaction. The involvement of strong hydrophobic interactions between these enzymes and domains on other PM peritrophins and glycoproteins may result in a tight aggregate of molecules that come off easily with the PM. Note that some of the gut enzymes were found in the detergent extraction fractions suggesting that they may in fact be termed "integral PM proteins". However, their contribution to PM structural integrity is unknown and we speculate that their intercalation into the PM scaffold may affect PM porosity through steric hindrance (Peters, 1992; Moskalyk et al., 1996).

### 3.6. *The* An. gambiae *PM: a structural model*

Based on the data generated by our proteomic analysis, we have elaborated a model for several types of possible molecular interactions that, in concert, could give rise to the adult mosquito PM (Fig. 6). During the course of blood ingestion the PM is rapidly secreted by the distended midgut epithelium. The chitin fibrils are assembled into a wide cross-hatched pattern connected by peritrophins containing multiple CBDs. Between 0 and 24 h, the cross-hatch is fully stretched and then slowly contracts like an accordion between 36 and 48 h post-blood feeding (Fig. 6a) as blood digestion proceeds to completion. One hypothesis is that mosquito chitinases break down chitin fibrils to facilitate this contraction (Shen and Jacobs-Lorena, 1997). There are several scenarios by which each of the different classes of peritrophins that we have identified by proteomic analysis assemble together to produce a PM as well as perform several of its functions. AgAper1, the most abundant peritrophin identified in our study, contains 2 CBDs and is predicted to be modified by O-glycans. AgAper9 and AgAper14 are smaller peritrophins with 1 CBD, are putatively aglycosylated, and are also relatively abundant (Table S1). We speculate that AgAper1 is arrayed within these large cross-hatches with their O-glycans extending into the center aperture (Fig. 6b). In this fashion, AgAper1 anchors to the PM via its CBD and controls the properties and selectiveness of the pores by virtue of its O-glycan moieties, providing for water retention within the PM. Thus, the PM would constitute a secondary mucous layer, which separates from the midgut microvillar mucins during PM retraction (Shen et al., 1999). Glycans may also serve another important function in protecting molecules from degradation by proteases, which are replete in the gut during digestion, thereby maintaining PM integrity. AgAper1 also has a predicted *N*-glycosylation site. N-linked glycans all contain a chitobiose (β1–4GlcNAc dimer) anchor that is covalently linked to the Asn on the polypeptide core (Kornfeld and Kornfeld, 1985). Chitobiose is structurally related to chitin and as such, the N-glycan can be used to array several AgAper1 proteins within the aperture, thereby further imparting distinct porosities to different apertures along the PM. An analogous array can be envisioned with the involvement of the aglycosylated peritrophins with 1 CBD, such as the relatively abundant AgAper14 (Fig. 6c) or with 3 CBDs, such as AgAper30 (Fig. 6d). The presence of aglycosylated peritrophins may suggest that these proteins are left open to degradation by midgut enzymes that are secreted sequentially into the gut lumen. The possible degradation of these aglycosylated peritrophins may play a role in the dynamic change in the PM's thickness and size. Mosquito-derived chitinase (Shen and Jacobs-Lorena, 1997) and the non-CBD containing hypothetical proteins identified in our study can also play a role in modulating the structure and porosity of the PM (Fig. 6d, f). In addition, there is evidence that PM proteins can bind heme residues via their CBDs (Pascoa et al., 2002; Devenport et al., 2006), suggesting that free CBDs on multi-CBD peritrophins function as such in the sequestration of heme and consequent protection of the midgut epithelial cells from its toxic effects (Fig. 6e, f). The hypothetical proteins can either aggregate by virtue of their proline-rich domains. Trypsin and chymotrypsin are commonly known to be aglycosylated although several trypsin-like and chymotrypsin-like proteases have been shown to be glycosylated (Gorman et al., 2000; Peng et al., 2003; Broehan et al., 2007; Paes Leme et al., 2008). If the predicted glycosylation sites on the trypsin and chymotrypsin molecules that we have identified from our proteomic study are indeed glycosylated, one can speculate that these proteases can be transiently associated with peritrophins via their N- or O-linked glycans (Fig. 6f) as they transit through the PM into the midgut lumen. Lastly, we propose a three-dimensional structure for the PM wherein 2-, 3- or 4-CBD containing peritrophins bridge multiple layers of chitin fibril sheets which are also arrayed in a cross-hatch pattern (Fig. 6g). Further examination of the relationship and intimate interaction between each of these molecules both *in vitro* and *in vivo* will lead to a better understanding of the dynamic nature of PM assembly and structure.

## Acknowledgements

This work was supported by a grant from the NIAID, NIH (R01AI040971) and by the Johns Hopkins Malaria Research Institute and the Bloomberg Family Foundation. R.R.D. is a NIH Ruth L. Kirschstein National Research Service Award Scholar (5F32AI068212).

## AppendixA. Supplementary information

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ibmb.2008.10.010

## References

Billingsley, P.F., Rudin, W., 1992. The role of the mosquito peritrophic membrane in bloodmeal digestion and infectivity of *Plasmodium* species. J. Parasitol. 78, 430–440.

Broehan, G., Zimoch, L., Wessels, A., Ertas, B., Merzendorfer, H., 2007. A chymotrypsin-like serine protease interacts with the chitin synthase from the midgut of the tobacco hornworm. J. Exp. Biol. 210, 3636–3643.

Devenport, M., Jacobs-Lorena, M., 2005. The peritrophic matrix of hematophagous insects. In: Marquardt, W. (Ed.), Biology of Disease Vectors, second ed. Elsevier Academic Press, New York, pp. 297–310.

Devenport, M., Fujioka, H., Jacobs-Lorena, M., 2004. Storage and secretion of the peritrophic matrix protein Ag-Aper1 and trypsin in the midgut of *Anopheles gambiae*. Insect Mol. Biol. 13, 349–358.

Devenport, M., Fujioka, H., Donnelly-Doman, M., Shen, Z., Jacobs-Lorena, M., 2005. Storage and secretion of Ag-Aper14, a novel peritrophic matrix protein, and Ag-Muc1 from the mosquito *Anopheles gambiae*. Cell Tissue Res. 320, 175–185.

Devenport, M., Alvarenga, P.H., Shao, L., Fujioka, H., Bianconi, M.L., Oliveira, P.L., Jacobs-Lorena, M., 2006. Identification of the *Aedes aegypti* peritrophic matrix protein AeIMUCI as a heme-binding protein. Biochemistry 45, 9540–9549.

Dimopoulos, G., Richman, A., della Torre, A., Kafatos, F.C., Louis, C., 1996. Identification and characterization of differentially expressed cDNAs of the vector mosquito, Anopheles gambiae. Proc. Natl. Acad. Sci. U.S.A. 23, 13066–13071.

Dimopoulos, G., Seeley, D., Wolf, A., Kafatos, F.C., 1998. Malaria infection of the mosquito Anopheles gambiae activates immune-responsive genes during critical transition stages of the parasite life cycle. EMBO J 21, 6115–6123.

Dinglasan, R.R., Jacobs-Lorena, M., 2005. Insight into a conserved lifestyle: protein-carbohydrate adhesion strategies of vector-borne pathogens. Infect. Immun. 73, 7797–7807.

Dinglasan, R.R., Kalume, D.E., Kanzok, S.M., Ghosh, A.K., Muratova, O., Pandey, A., Jacobs-Lorena, M., 2007. Disruption of Plasmodium falciparum development by antibodies against a conserved mosquito midgut antigen. Proc. Natl. Acad. Sci. U.S.A. 33, 13461–13466.

Edwards, M.J., Lemos, F.J., Donnelly-Doman, M., Jacobs-Lorena, M., 1997. Rapid induction by a blood meal of a carboxypeptidase gene in the gut of the mosquito *Anopheles gambiae*. Insect Biochem. Mol. Biol. 27, 1063–1072.

Elvin, C.M., Vuocolo, T., Pearson, R.D., East, I.J., Riding, G.A., Eisemann, C.H., Tellam, R.L., 1996. Characterization of a major peritrophic membrane protein, peritrophin-44, from the larvae of *Lucilia cuprina*. cDNA and deduced amino acid sequences. J. Biol. Chem. 271, 8925–8935.

Eng, J.K., McCormack, A.L., Yates, J.R.I., 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 979.

Enright, A.J., Van Dongen, S., Ouzounis, C.A., 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 7, 1575–1584.

Florens, L., Washburn, M.P., 2006. Proteomic analysis by multidimensional protein identification technology. Methods Mol. Biol, 159–175.
Galun, R., Avi-Dor, Y., Bar-Zeev, M., 1963. Feeding response in Aedes aegypti: stimulation by adenosine triphosphate. Science 3600, 1674–1675.
Gorman, M.J., Andreeva, O.V., Paskewitz, S.M., 2000. Sp22D: a multidomain serine protease with a putative role in insect immunity. Gene 251, 9–17.
Hazelbauer, G.L., Falke, J.J., Parkinson, J.S., 2008. Bacterial chemoreceptors: high-performance signaling in networked arrays. Trends Biochem. Sci. 33, 9–19.
He, N., Botelho, J.M., McNall, R.J., Belozerov, V., Dunn, W.A., Mize, T., Orlando, R., Willis, J.H., 2007. Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass spectrometry. Insect Biochem. Mol. Biol. 37, 135–146.
Huber, M., Cabib, E., Miller, L.H., 1991. Malaria parasite chitinase and penetration of the mosquito peritrophic membrane. Proc. Natl. Acad. Sci. U.S.A. 88, 2807–2810.
Kondoh, H., Ball, C.B., Adler, J., 1979. Identification of a methyl-accepting chemotaxis protein for the ribose and galactose chemoreceptors of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 76, 260–264.
Kornfeld, R., Kornfeld, S., 1985. Assembly of asparagine-linked oligosaccharides. Annu. Rev. Biochem. 54, 631–664.
Langer, R.C., Li, F., Vinetz, J.M., 2002a. Identification of novel *Plasmodium gallinaceum* zygote- and ookinete-expressed proteins as targets for blocking malaria transmission. Infect. Immun. 70, 102–106.
Langer, R.C., Li, F., Popov, V., Kurosky, A., Vinetz, J.M., 2002b. Monoclonal antibody against the *Plasmodium falciparum* chitinase, PfCHT1, recognizes a malaria transmission-blocking epitope in *Plasmodium gallinaceum* ookinetes unrelated to the chitinase PgCHT1. Infect. Immun. 70, 1581–1590.
Lemos, F.J., Cornel, A.J., Jacobs-Lorena, M., 1996. Trypsin and aminopeptidase gene expression is affected by age and food composition in *Anopheles gambiae*. Insect Biochem. Mol. Biol. 26, 651–658.
Li, F., Templeton, T.J., Popov, V., Comer, J.E., Tsuboi, T., Torii, M., Vinetz, J.M., 2004. *Plasmodium* ookinete-secreted proteins secreted through a common micronemal pathway are targets of blocking malaria transmission. J. Biol. Chem. 279, 26635–26644.
Moskalyk, L.A., Oo, M.M., Jacobs-Lorena, M., 1996. Peritrophic matrix proteins of *Anopheles gambiae* and *Aedes aegypti*. Insect Mol. Biol. 5, 261–268.
Muller, H.M., Crampton, J.M., della Torre, A., Sinden, R., Crisanti, A., 1993a. Members of a trypsin gene family in Anopheles gambiae are induced in the gut by blood meal. EMBO J 7, 2891–2900.
Muller, H.M., Vizioli, I., della Torre, A., Crisanti, A., 1993b. Temporal and spatial expression of serine protease genes in Anopheles gambiae. Parassitologia, 73–76.
Paes Leme, A.F., Prezoto, B.C., Yamashiro, E.T., Bertholim, L., Tashima, A.K., Klitzke, C.F., Camargo, A.C., Serrano, S.M., 2008. Bothrops protease A, a unique highly glycosylated serine proteinase, is a potent, specific fibrinogenolytic agent. J. Thromb. Haemost..
Pascoa, V., Oliveira, P.L., Dansa-Petretski, M., Silva, J.R., Alvarenga, P.H., Jacobs-Lorena, M., Lemos, F.J., 2002. *Aedes aegypti* peritrophic matrix and its interaction with heme during blood digestion. Insect Biochem. Mol. Biol. 32, 517–523.
Peng, Q., McEuen, A.R., Benyon, R.C., Walls, A.F., 2003. The heterogeneity of mast cell tryptase from human lung and skin. Eur. J. Biochem. 270, 270–283.
Peters, W., 1992. Peritrophic Membranes. Springer-Verlag, Berlin, 238 pp.
Ramos, A., Mahowald, A., Jacobs-Lorena, M., 1994. Peritrophic matrix of the black fly *Simulium vittatum*: formation, structure, and analysis of its protein components. J. Exp. Zool. 268, 269–281.
Sarauer, B.L., Gillott, C., Hegedus, D., 2003. Characterization of an intestinal mucin from the peritrophic matrix of the diamondback moth, *Plutella xylostella*. Insect Mol. Biol. 12, 333–343.
Shahabuddin, M., Toyoshima, T., Aikawa, M., Kaslow, D.C., 1993. Transmission-blocking activity of a chitinase inhibitor and activation of malarial parasite chitinase by mosquito protease. Proc. Natl. Acad. Sci. U.S.A. 90, 4266–4270.
Shahabuddin, M., Lemos, F.J., Kaslow, D.C., Jacobs-Lorena, M., 1996. Antibody-mediated inhibition of *Aedes Aegypti* midgut trypsins blocks sporogonic development of *Plasmodium gallinaceum*. Infect. Immun. 64, 739–743.
Shao, L., Devenport, M., Jacobs-Lorena, M., 2001. The peritrophic matrix of hematophagous insects. Arch. Insect Biochem. Physiol. 47, 119–125.
Shao, L., Devenport, M., Fujioka, H., Ghosh, A., Jacobs-Lorena, M., 2005. Identification and characterization of a novel peritrophic matrix protein, Ae-Aper50, and the microvillar membrane protein, AEG12, from the mosquito, *Aedes aegypti*. Insect Biochem. Mol. Biol. 35, 947–959.
Shen, Z., Jacobs-Lorena, M., 1997. Characterization of a novel gut-specific chitinase gene from the human malaria vector *Anopheles gambiae*. J. Biol. Chem. 272, 28895–28900.
Shen, Z., Jacobs-Lorena, M., 1998. A type I peritrophic matrix protein from the malaria vector *Anopheles gambiae* binds to chitin. Cloning, expression, and characterization. J. Biol. Chem. 273, 17665–17670.
Shen, Z., Jacobs-Lorena, M., 1999. Evolution of chitin-binding proteins in invertebrates. J. Mol. Evol. 48, 341–347.
Shen, Z., Dimopoulos, G., Kafatos, F.C., Jacobs-Lorena, M., 1999. A cell surface mucin specifically expressed in the midgut of the malaria mosquito *Anopheles gambiae*. Proc. Natl. Acad. Sci. U.S.A. 96, 5610–5615.
Shi, L., Paskewitz, S.M., 2004. Identification and molecular characterization of two immune-responsive chitinase-like proteins from *Anopheles gambiae*. Insect Mol. Biol. 13, 387–398.
Tabb, D.L., McDonald, W.H., Yates 3rd, J.R., 2002. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J. Proteome Res. 1, 21–26.
Tellam, R.L., Wijffels, G., Willadsen, P., 1999. Peritrophic matrix proteins. Insect Biochem. Mol. Biol. 29, 87–101.
Williamson, M.P., 1994. The structure and function of proline-rich regions in proteins. Biochem. J. 297 (Pt 2), 249–260.
Wolters, D.A., Washburn, M.P., Yates 3rd, J.R., 2001. An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. 23, 5683–5690.
Wu, H., de Graaf, B., Mariani, C., Cheung, A.Y., 2001. Hydroxyproline-rich glycoproteins in plant reproductive tissues: structure, functions and regulation. Cell Mol. Life Sci. 58, 1418–1429.
Wu, C.C., MacCoss, M.J., Howell, K.E., Yates 3rd, J.R., 2003. A method for the comprehensive proteomic analysis of membrane proteins. Nat. Biotechnol. 5, 532–538.
Zybailov, B., Coleman, M.K., Florens, L., Washburn, M.P., 2005. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. Anal. Chem. 19, 6218–6224.
Zybailov, B., Mosley, A.L., Sardiu, M.E., Coleman, M.K., Florens, L., Washburn, M.P., 2006. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. J. Proteome Res. 9, 2339–2347.