A cell surface mucin specifically expressed in the midgut of the malaria mosquito *Anopheles gambiae*

ZHICHENG SHEN*, GEORGE DIMOPOULOS†, FOTIS C. KAFATOS†, AND MARCELO JACOBS-LORENA*‡

*Department of Genetics, School of Medicine, Case Western Reserve University, Cleveland, OH 44106-4955; and †European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

ABSTRACT An invertebrate intestinal mucin gene, *AgMuc1*, was isolated from the malaria vector mosquito *Anopheles gambiae*. The predicted 122-residue protein consists of a central core of seven repeating TTTTVAP motifs flanked by hydrophobic N- and C-terminal domains. This structure is similar to that of mucins that coat the protozoan parasite *Trypanosoma cruzi*. Northern blot analysis indicated that the gene is expressed exclusively in the midgut of adult mosquitoes. A length polymorphism and *in situ* hybridization were used to genetically and cytogenetically map *AgMuc1* to division 7A of the right arm of the second chromosome. The subcellular localization of the encoded protein in tissue culture cells was examined by using a baculovirus vector to express AgMuc1 protein tagged with the green fluorescent protein (GFP). The results indicated that this protein is found at the cell surface and that both hydrophobic domains are required for cell surface targeting. We propose that AgMuc1 is an abundant mucin-like protein that lines the surface of the midgut microvilli, potentially protecting the intestinal epithelium from the proteinase-rich environment of the gut lumen. An intriguing possibility is that, as an abundant surface protein, AgMuc1 may also interact with the malaria parasite during its invasion of the mosquito midgut.

Anopheles gambiae is the principal vector for transmission of human malaria, a disease that kills in excess of 2 million people worldwide (mostly children) every year. To be transmitted from one host to another, malaria parasites have to complete a complex life cycle in vector mosquitoes, starting in the midgut lumen, crossing through the midgut epithelial barrier, and finally invading the salivary glands, from where they can be inoculated into the next host during blood feeding.

The insect midgut is composed of a single layer of epithelial cells, which are lined at their basal side by a continuous extracellular layer, the basal lamina. On the apical side, the epithelial cell membranes are folded into numerous actin-filled microvilli. Microvilli greatly increase the surface area and play an important role in absorption of nutrients (1). The microvilli are exposed to the harsh environment of the gut lumen, and they are subjected to damage caused by food particle abrasion, digestive hydrolases, and attack by pathogens and parasites. Two extracellular structures have been proposed to provide protection to the microvilli: the peritrophic matrix and the glycocalyx (2–4). The peritrophic matrix is an extracellular sac, composed of chitin, proteins, and proteoglycans (2, 3), which completely surrounds the ingested food and is secreted by the gut epithelial cells. All the recently cloned peritrophic matrix proteins from *Lucilia cuprina* (5, 6), *Trichoplusia ni* (7), and the mosquito *Anopheles gambiae* (8) have at least two chitin-binding domains that are presumed to function in the cross-linking of the chitin fibrils. However, some of these proteins

also have mucin-like domains (6, 7), suggesting that the insect peritrophic matrix resembles the vertebrate intestinal mucus, a structure largely composed of mucins. In addition to providing protection, the peritrophic matrix may also facilitate digestion by compartmentalization of digestive enzymes (9). Another protective structure is the glycocalyx (glyco = sweet/sugar, calyx = shell), which is an integral part of the microvillar membrane and appears as an electron-dense fuzzy coat on the outside of the microvillar surface (4, 10). The glycocalyx, including that of mosquitoes, is rich in carbohydrates, as it is recognized by a variety of lectins (11, 12). However, no experimental data are currently available on the molecular composition of the insect midgut glycocalyx. This is a subject of potential importance, because components of the glycocalyx may serve as receptors or attachment sites for invasion of parasites such as malaria.

Plasmodium development begins in the mosquito gut by formation of gametes, fertilization, meiosis, and differentiation into an ookinete. About 24 h later, the ookinete crosses the midgut epithelium from the luminal to the hemocoel side. Although the recognition of the gut epithelial cell surface by ookinetes is a crucial step in the life cycle of *Plasmodium*, little molecular information is available about this process. There is evidence to suggest that carbohydrate components of the gut epithelial cell surface are involved in interactions with the parasite (13, 14). However, no candidate glycosylated proteins from the adult *Anopheles* midgut have been identified.

Even in *Anopheles* species that can serve as vectors, not all strains support the development of *Plasmodium* with the same efficiency. Selected refractory mosquito strains exist in which the invading ookinetes are killed in the midgut epithelial cells either by lysis or, later, by melanotic encapsulation (15, 16). Other refractory mechanisms may act prior to or during midgut invasion by the ookinetes. The mechanisms controlling the demonstrated refractory traits remain unknown. Genetic mapping has shown that the melanotic encapsulation phenotype is controlled by three quantitative trait loci (17), whereas the lytic refractoriness is believed to be controlled by a different locus (15). The genetic elements controlling infection intensities at pre-encapsulation stages appear to be genetically unlinked to the encapsulation quantitative trait loci (QTL) and remain to be defined (17).

In this report, we describe the cDNA cloning of a mucin-like protein, named AgMuc1 for *Anopheles gambiae* mucin 1. Our results suggest that this putative mucin is an abundant surface-associated protein component of the midgut. Interestingly, a length polymorphism within the mucin domain was detected between strains that are susceptible and refractory to the malaria parasite. By use of this polymorphism, the *AgMuc1*

---

gene was mapped to a chromosomal region near but not at *Pen1*, the main locus controlling the melanotic encapsulation phenotype.

MATERIALS AND METHODS

cDNA Cloning and Sequencing. One *AgMuc1* cDNA clone, pm2, was obtained by screening an *A. gambiae* adult midgut Lambda ZAP cDNA library (18) with an antiserum against peritrophic matrix proteins; it was sequenced in both directions by using a dye-termination procedure. An identical *AgMuc1* cDNA clone was obtained independently by screening a Lambda ZAP cDNA library from adult female abdomens (19) with a ^{32}P-labeled dCTP trinucleotide repeat probe encoding a polythreonine stretch.

Northern Analysis. Total RNA samples were isolated from *A. gambiae* adult midguts, adult carcasses (whole body minus midgut), pupae, larval guts, and larval carcasses. About 5 μg of total RNA from each sample was fractionated by electrophoresis on denaturing 1% agarose gels and transferred onto a nylon membrane. The full-length cDNA insert was labeled with [α-^{32}P]dCTP by random labeling (20) and hybridized to the RNA blots at high stringency.

Isolation of the *Muc1b* Allele from the 4A r/r (Pink Eye) Strain. A 318-bp PCR fragment was amplified by PCR from 4A r/r strain cDNA by using the *AgMuc1*-specific primers: MUCPA, 5′-AAAGTTGTTGTGGGACTAGTG-3′ and MUCPB, 5′-CGACATTGCCACGTATGCTCC-3′ with 26 cycles of 1 min at 94°C, 1 min at 56°C, and 1 min at 72°C. The fragment was cloned in a TOPO TA cloning vector (Invitrogen) prior to sequencing.

Genetic Mapping. Genetic mapping of *AgMuc1* was kindly performed by Rui Wang, using the same primers as for the cloning of the polymorphic allele, on the DNA of 55 individuals of families previously used in mapping the QTL for refractoriness of *A. gambiae* to *Plasmodium cynomolgi* B (17).

Cytogenetic Mapping. The *AgMuc1* cDNA clone was mapped by *in situ* hybridization to Suakoko strain *Anopheles gambiae* polytene chromosomes as previously described (21).

Constructs for Baculovirus-Based Protein Expression. Four baculovirus transfer plasmid constructs were made (see Fig. 4A). A control green fluorescent protein (GFP) expression vector (pBacPak8-GFP) was constructed by excising the GFP-encoding insert from plasmid EGFP-C3 (CLONTECH) with *Nhe*I and *Pst*I and cloning into the *Xba*I (*Xba*I and *Nhe*I fragments have compatible ends) and *Pst*I sites of the pBacPak8 plasmid (CLONTECH). To obtain the pBacPak8–72GFP fusion construct, the cDNA fragment encoding the N-terminal 72 amino acid residues of *AgMuc1* was amplified by PCR using cDNA clone pm2 as the template. The two PCR primers had additional sequences at their 5′ ends to provide restriction sites as follows. Forward primer: 5′-AGC-CGCTAGCAACATGTTGAAAGTTGTT; reverse primer: 5′-AGCTACCGGTCCAGGTGCCACTGTG; the restriction sites for *Nhe*I and *Age*I are underlined. This PCR product was digested with *Nhe*I and *Age*I and was inserted into the EGFP-C3 plasmid digested with the same restriction enzymes. The 72GFP cDNA fragment was then excised with *Nhe*I and *Pst*I and inserted into the *Xba*I and *Pst*I sites of pBacPak8 plasmid. To construct pBacPak8-GFP27 and pBacPak8–72EGFP27, the cDNA fragment encoding the C-terminal 27 amino acids was amplified by PCR using a *AgMuc1*-specific primer (forward primer: 5′ GCTCAAGCTTTCAAGTGC-CCCACAGG, *Hin*dIII site underlined) and a T3 primer. After digestion with *Hin*dIII and *Pst*I (from the multiple cloning site of pBluescript), the PCR product was inserted into the *Hin*dIII and *Pst*I sites of pBacPak8-GFP and pBacPak8–72GFP, respectively.

Cell Culture and Generation of Recombinant Viruses. Sf21 (*Spodoptera frugiperda*) cells were grown at 27°C in Grace's

insect medium plus 10% fetal bovine serum. Recombinant baculoviruses were generated by homologous recombination by cotransfection of plasmid constructs and viral DNA digested with *Bsu*36 I (from the CLONTECH pBacPAK Baculovirus Expression System kit). The procedures recommended by the supplier were followed. Cells infected with recombinant viruses were easily identified with a UV dissection microscope to visualize GFP. Recombinant viruses were collected from the cell culture medium and used to infect Sf21 cells.

Preparation of Antibodies Against Microvilli. Microvilli were prepared from midguts dissected from 5- to 10-day-old *A. gambiae* adult females according to English *et al.* (22). About 500 midguts were suspended in 2 ml of ice-cold 50 mM sucrose/0.1 mM PMSF/2 mM Tris·HCl, pH 7.4, and homogenized on ice with 15 strokes of a Dounce homogenizer. Then CaCl$_2$ was added to a final concentration of 10 mM. After 15 min of shaking, the sample was centrifuged at 4,300 × g for 10 min at 4°C. The supernatant was collected and centrifuged at 26,000 × g for 10 min. The microvillar pellet was resuspended with PBS and used to immunize rabbits.

Analysis of Baculovirus-Expressed GFP Fusion Proteins. Sf21 cells infected with the various recombinant viruses were collected 60 h after infection. Total proteins from the infected cells were first fractionated by electrophoresis on SDS/12% polyacrylamide gels and then blotted onto a nitrocellulose membrane. The fusion proteins were detected by Western blot analysis using polyclonal antibody against GFP (CLONTECH) or rabbit antiserum against microvilli.

Confocal Microscopy. The subcellular localization of GFP or of GFP-fusion proteins was detected by the fluorescence of the GFP moiety by using a Bio-Rad MRC-600 laser scanning confocal imaging system with a 488-nm excitation wavelength.

Detection of 72GFP27 Protein on the Cell Surface. Sf21 cells infected with 72GFP27 or 72GFP recombinant baculoviruses were collected by centrifugation at 2,000 × g for 2 min. The cells were then resuspended and incubated with basic insect medium containing rabbit anti-GFP antibody (CLONTECH, 1:500 dilution) and 1% BSA for 1 h at room temperature. The cells were then washed three times with basic insect cell medium and incubated for 30 min with basic insect medium containing 1% BSA and rhodamine-conjugated goat anti-rabbit antibody (Pierce; 0.02 mg/ml). The cells were washed three times and examined with a confocal fluorescence microscope. Exactly the same parameters were used to collect the fluorescence images of cells expressing 72GFP27 and 72GFP. Images from the green (inherent GFP fluorescence anywhere in the cell) and red (surface labeling by antibody) channels were merged.

RESULTS

Cloning of a Mucin cDNA. cDNA clones were obtained independently in our two laboratories by unrelated screens of *A. gambiae* midgut or abdominal Lambda ZAP cDNA expression libraries. One of the clones contains a 572-bp insert that includes an ORF encoding a protein of 122 amino acids (Fig. 1A). The cDNA has a putative polyadenylation signal 18 bp from the beginning of the poly(A) tail (AATAAA, boldface in Fig. 1A). The deduced protein has two hydrophobic fragments, one at the N terminus and the other at the C terminus flanking a central core of seven repeating TTTTVAP motifs, six of which are near perfect repeats (Fig. 1 B–D). The amino acid composition of the predicted protein is heavily biased, as it has 39 threonines (accounting for 44% of the residues in the central core), 11 prolines (13%), 9 alanines (11%), 9 valines (11%), 5 glycines (6%), 4 serines (5%), 4 glutamines (5%), 3 aspartic acids (4%), 1 isoleucine (1%), and 1 lysine (1%). The other 10 amino acids are absent. Repeating threonine- and proline-rich motifs are the hallmark of the extensively char-
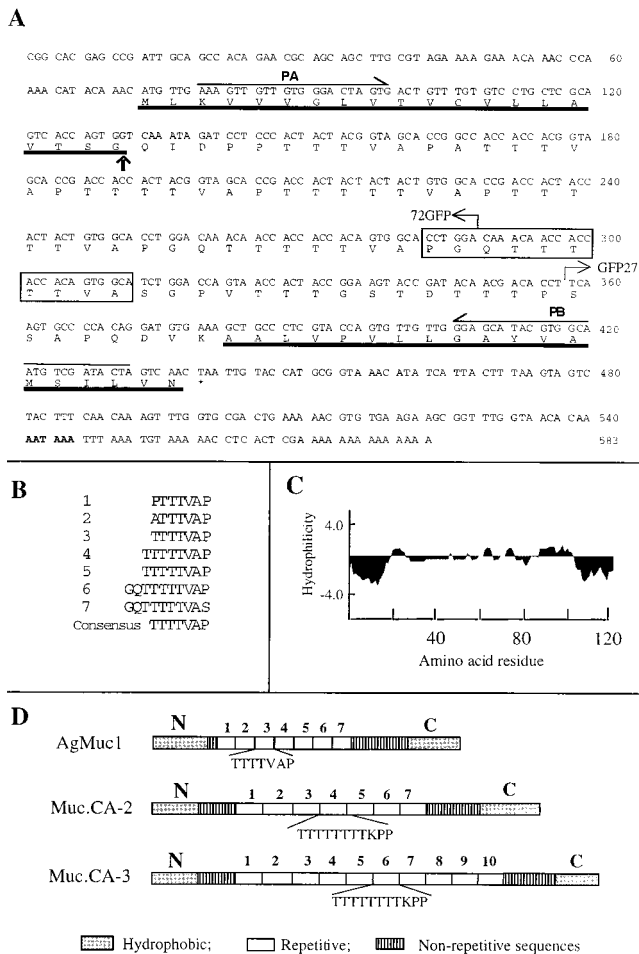
A

B

```
1    PTTTVAP
2    ATTTVAP
3    TTTTVAP
4    TTTTTVAP
5    TTTTTVAP
6    GQTTTTVAP
7    GQTTTTVAS
Consensus  TTTTVAP
```

C

D

☐ Hydrophobic; ☐ Repetitive; ☐ Non-repetitive sequences

FIG. 1. Nucleotide and predicted amino acid sequences of *AgMuc1.* (*A*) The N- and C-terminal hydrophobic protein domains are underlined. The putative cleavage site of the signal peptide is indicated by a vertical arrow and the borders of GFP fusion constructs by bent arrows. The putative polyadenylation signal sequence is in boldface. The polymorphic region missing in the *Muc1b* allele is boxed and the PCR primers used for genetic mapping (PA, PB) are indicated above the sequence. (*B*) The seven repeated motifs of the central protein core are aligned. The consensus repeat sequence is given at the bottom. (*C*) A Kyte–Doolittle hydropathy plot was generated with an average hydrophilicity window of 7 residues. (*D*) Diagrammatic comparison of the predicted amino acid sequence of the *A. gambiae* AgMuc1 mucin with that of *Trypanosoma cruzi* mucins, MUC.CA-2 and MUC.CA-3 (31, 32). AgMuc1 mucin contains TTTTVAP repeats, whereas the MUC.CA-2 and MUC.CA-3 mucins contain TTTTTTT-TKPP motifs. All three mucins contain nonrepeated sequences between the repeat array and the N- and C-terminal hydrophobic sequences.

acterized vertebrate mucins (23–27). Hence, this protein was named AgMuc1, for *Anopheles gambiae* mucin.

According to the rules of von Heijne (28), the N-terminal hydrophobic domain is likely to be a signal peptide with the cleavage site at the carboxyl side of Gly-20 (Fig. 1*A*). The C-terminal hydrophobic domain may serve to anchor the protein to the cell membrane. This sequence has the features of glycosyl-phosphatidylinositol (GPI) anchor signals: a region of 15–20 hydrophobic residues at the extreme C terminus preceded by a few hydrophilic residues (29), as in the case of GPI-anchored *Trypanosoma cruzi* mucins (30–32).

Although sequence similarity to known proteins was not detected by BLAST searches, the amino acid composition and sequence pattern of AgMuc1 show significant similarity to a mucin family isolated from the protozoan parasite *T. cruzi* (Fig. 1*C*). In contrast, vertebrate mucins are often much longer,

have more numerous repetitive units, and lack an extreme C-terminal hydrophobic sequence. For instance, the human MUC2 has about 100 PTTTPITTTTTVTPTPTPTGTQT repeats, the number varying among different alleles (33). Both vertebrate and *T. cruzi* mucins are highly glycosylated by O-linkage to threonine or serine residues of the protein backbone (30, 34), suggesting that this may also be true for AgMuc1.

AgMuc1 Is Exclusively Expressed in Adult Midguts. Developmental- and tissue-specific expression of the *AgMuc1* gene was investigated by Northern blot analysis (Fig. 2). A single band of ≈0.8 kb was detected in the RNA isolated from adult midguts, but not in the RNA from adult carcasses, whole pupae, larval carcasses, or larval guts, suggesting that *AgMuc1* is expressed only in adult midguts. The size of the mRNA detected on Northern blots agrees with the length of the cloned cDNA, assuming that the poly(A) tail contains ≈100 adenosine residues. The signal detected on Northern blots is very strong, suggesting that the mRNA is abundant in the adult gut.

Polymorphic *AgMuc1* Alleles and Mapping. When reverse transcription–PCR primers flanking the repeat domain (Fig. 1*A*) were used a shorter *AgMuc1* allele was detected in the malaria-susceptible *A. gambiae* 4A r/r strain (Fig. 3*A*). This allele, named *AgMuc1b*, lacks 30 bp that encode one repeat (PGQTTTVA). All other *A. gambiae* laboratory strains tested (G3, Suakoko, and L3–5) were found to have the *AgMuc1* allele. The L3–5 strain is refractory to malaria parasites. The invading ookinetes and early oocysts are encapsulated soon after crossing the midgut epithelium (16). The length polymorphism was used to map *AgMuc1* genetically. The female F_1 progeny of a cross between the susceptible 4A r/r and the refractory L3–5 strains were backcrossed with 4A r/r males, and segregation of *AgMuc1* was followed in the backcross progeny. Recombination distances of *AgMuc1* to previously mapped microsatellite markers (17) placed this gene in the general vicinity of the major refractory QTL, *Pen1,* on the right arm of the second chromosome (Fig. 3*C*). Cytogenetic mapping of the *AgMuc1* cDNA clone to *A. gambiae* polytene chromosomes localized this gene to division 7A on the right arm of the second chromosome (Fig. 3*B*). This location correlates with the genetic and cytogenetic location of the adjacent microsatellite markers (data not shown) but is clearly separated from that of *Pen1,* which is most likely at 8C (F. Collins, personal communication).
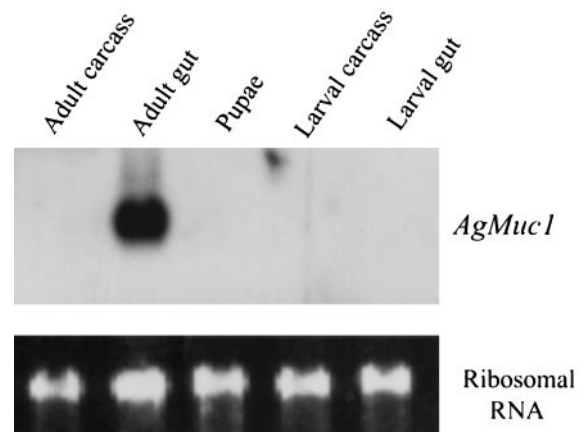

FIG. 2. Developmental and tissue specificity of *AgMuc1* expression. (*Upper*) Autoradiogram of a Northern blot of RNAs (≈5 μg per lane) isolated from the indicated tissues. The blot was hybridized overnight with a ^{32}P-labeled *AgMuc1* cDNA probe, washed, and exposed to film for 4 h. (*Lower*) Staining of ribosomal RNA with ethidium bromide to indicate the amount of RNA analyzed in each lane.
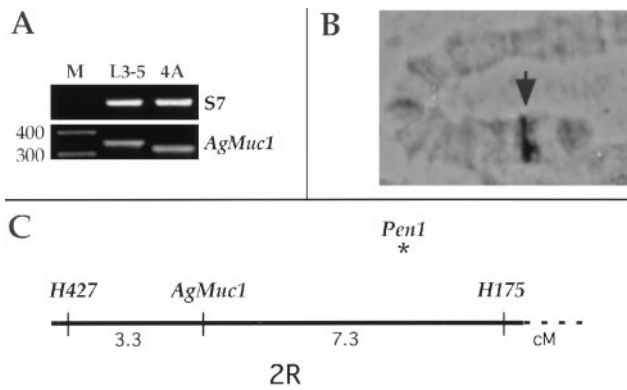
Genetics: Shen *et al.*

Proc. Natl. Acad. Sci. USA 96 (1999)   5613

FIG. 3. Polymorphism and genetic mapping of *AgMuc1*. (*A*) Ethidium bromide-stained agarose gel with the PCR-amplified *Ag-Muc1* segment containing the length polymorphism from the refractory L3-5 strain and susceptible 4A r/r strain. Lane M, length markers. (*B*) The cytogenetic location of *AgMuc1* in division 7A on the second chromosome is indicated by an arrowhead. (*C*) Genetic distances of *AgMuc1* from adjacent microsatellite markers are shown in centimorgans (cM). The gene is within the same region as the *Plasmodium* refractory QTL *Pen1*, but the latter locus is distinct from *AgMuc1* and is located closer to H175.

Expression of GFP-Mucin Fusion Proteins in Insect Cells. GFP-AgMuc1 fusion proteins (Fig. 4*A*) were expressed in insect cells from a baculovirus vector. The recombinant proteins were detected by Western blot analysis using an anti-GFP antibody (Fig. 4*B*). The relative mobilities of the different recombinant proteins were those expected from their calculated molecular mass, indicating that the fusion proteins are faithfully expressed. However, the apparent molecular masses of the fusion proteins containing repeat sequences (72GFP and 72GFP27; Fig. 4*A*) were higher than the calculated molecular masses (34 and 36 kDa). This discrepancy may be due to glycosylation of the fusion proteins at the core mucin domain. However, the proteins may not be fully glycosylated (in humans carbohydrates make up 50% of the mass), either because of strong and rapid protein expression in the baculovirus system or because the Sf21 cells, which are ovarian in origin, do not have all the necessary enzymes.

An antiserum against midgut microvilli was used to probe control Sf21 cell lysates or lysates expressing recombinant proteins (Fig. 4*C*). A single band (of the same mobility as that detected by GFP antibody) was detected in cells expressing 72GFP27, but not in uninfected Sf21 cells or in cells expressing GFP. No bands were detected on similar blots with preimmune serum (data not shown). These results indicate that the antiserum is specific for the mucin sequences in the fusion protein and suggest that the mucin is a component of midgut microvilli.

Both Hydrophobic Domains Are Required to Target AgMuc1 to the Cell Surface. The products of all four fusion constructs shown in Fig. 4*A* could be detected and localized easily by fluorescence microscopy when expressed in Sf21 cells (Fig. 5*A*). This observation indicates that GFP fluorescence is not lost, even when this reporter protein is placed in the middle of the mucin protein (72GFP27). For all constructs, the fluorescence was strongly detectable at both 30 h and 60 h after infection, and it revealed final spatial localization by the latter time (*A2, A4, A6,* and*A8* in Fig. 5*A*). In the control experiment, GFP was distributed throughout the cytoplasm and the nucleus (*A1* and *A2*), except for some dark areas in the cytoplasm, which may correspond to secretory vesicles and/or lysosomes. In general, less GFP was detected in the nucleolus. Nuclear localization is consistent with the idea that GFP is a nonsecretory soluble protein that is able to diffuse into the nucleus because of its small size ($\approx$28 kDa). The distribution of
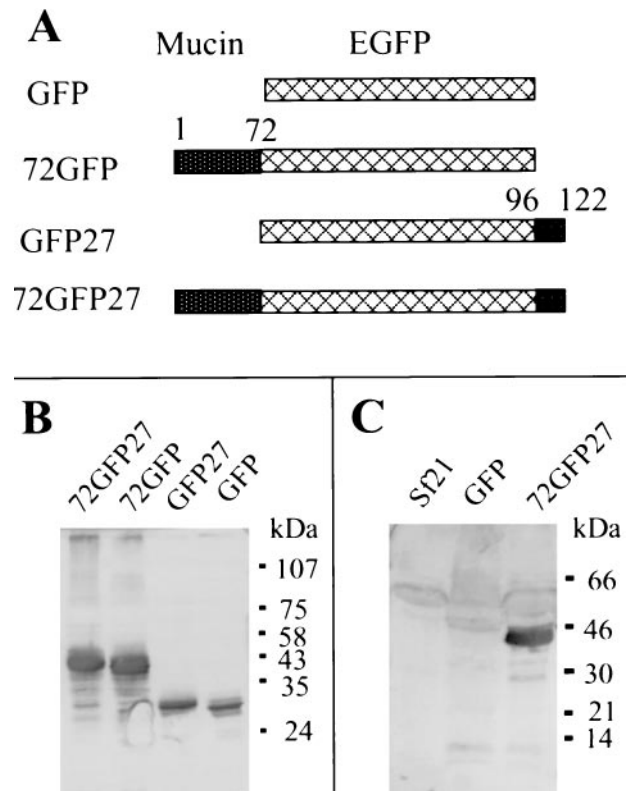

FIG. 4. Expression and characterization of mucin-GFP fusion proteins encoded by recombinant baculoviruses. (*A*) Diagrams showing recombinant protein structures. Lattice-patterned rectangles, GFP reporter protein; stippled rectangles, AgMuc1 sequences (see Fig. 1*A*). (*B*) Western blotting analysis of baculovirus-expressed GFP and GFP-mucin fusion proteins with an anti-GFP antibody. Total proteins from cells infected with different recombinant baculoviruses were loaded on each lane. A polyclonal antibody against GFP was used to probe the Western blots. The structure of the recombinant proteins indicated at the top of each lane is given in *A*. (*C*) Western blotting analysis of the baculovirus-expressed mucin-GFP protein with antiserum against total midgut microvilli. The Sf21 lane contains control cell lysate. Migration of marker proteins is indicated on the right side of *B* and *C*.

recombinant 72GFP27 protein, consisting of GFP fused on either side to the AgMuc1 N- and C-termini, differed dramatically from that of GFP alone. During the early stages of infection, the fusion protein was observed both at the cell surface and in the cytoplasm (*A3*). At later stages of infection, the protein was detected predominantly at the cell surface (*A4*). Therefore, the combination of AgMuc1 N- and C-terminal sequences is sufficient to direct the fusion protein to the cell surface. Fusion protein 72GFP accumulated predominantly in cytoplasmic vesicle-like structures and on the nuclear periphery (*A5* and *A6*). The vesicle-like structures may be secretory vesicles, while fluorescence at the nuclear periphery may represent the condensed endocytoplasmic membrane as previously observed in baculovirus-infected cells (35). This pattern of distribution was expected because the 72 residues fused to GFP include the putative signal peptide. The GFP27 fusion protein showed a different distribution. At early times the protein appeared more or less uniformly distributed in the cytoplasm and in the nucleus. In many (but not all) cells the protein was concentrated at the periphery of the nuclear envelope (*A7*), suggesting that the fusion protein may be targeted to the endocytoplasmic membrane and/or the nuclear envelope. At later times, the protein tended to be more concentrated in the nucleus than in the cytoplasm (*A8*; compare with *A2*), indicating that the C-terminal sequences may contain a cryptic nuclear localization sequence. Hong *et al.*
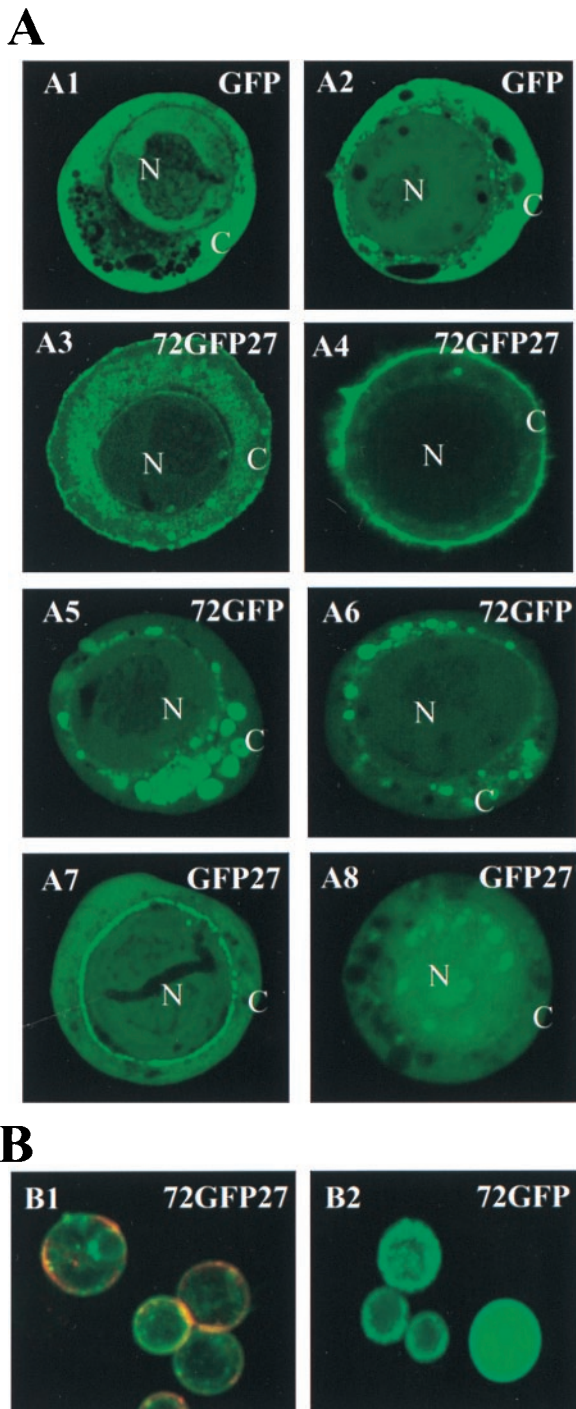
A

B

FIG. 5. Localization of baculovirus-expressed GFP fusion proteins in Sf21 cells. (*A*) Sf21 cells were infected with recombinant baculoviruses that express the GFP, 72GFP27, GFP27, and 72GFP recombinant constructs (Fig. 4*A*) under the control of the viral polyhedrin promoter. Representative images recorded at 30 h after infection (*A1, A3, A5,* and *A7*) and at 60 h after infection (*A2, A4, A6,* and *A8*) are shown. The green is due to fluorescence of GFP or of GFP fusion proteins. N, Nucleus; C, cytoplasm. (×1000.) (*B*) Immunological localization of mucin-GFP fusion proteins. Living Sf21 cells expressing fusion proteins 72GFP27 or 72GFP (as shown in *A4* and *A6*, respectively) were incubated with a rabbit anti-GFP antibody followed by incubation with a rhodamine-conjugated anti-rabbit IgG secondary antibody, thus detecting GFP-containing proteins only if exposed to the surface (red channel). The distribution of GFP, independent of its location in the cell, was detected by its inherent fluorescence (green channel). The images shown in *B1* and *B2* were collected while using identical settings for the green and red channels, and images from the red and green channels were then merged. (×250.)

(35) found that a short N-terminal sequence of the occlusion-derived virus (ODV) envelope protein, composed predominantly of hydrophobic amino acids, is sufficient to direct proteins to cytoplasmic membranes, nuclear envelope, and nucleus when expressed in Sf9 cells from baculoviruses. It is possible that the C-terminal hydrophobic sequence of AgMuc1 is playing a similar role.

To further corroborate the association of the baculovirus-expressed mucin with the cell surface, rabbit anti-GFP antibody and rhodamine-conjugated goat anti-rabbit IgG were used to detect the GFP fusion protein on the surface of living cells infected with recombinant AgMuc1-GFP baculoviruses (red channel). The inherent fluorescence of GFP, anywhere in the cell, was also recorded (green channel). Fig. 5*B* shows the fluorescent images. A red or orange ring indicating the presence of externalized GFP-containing protein was observed around almost every cell expressing 72GFP27 (Fig. 5, *B1*). In contrast, with the same staining and detection conditions, no rhodamine signal was detected in any cells expressing 72GFP (Fig. 5, *B2*). These results indicate that the 72GFP27 protein is accessible to antibodies on the surface of living cells, whereas the 72GFP protein is not. In conclusion, both the N-terminal and the C-terminal AgMuc1 amino acid sequences are required to direct the protein to the plasma membrane. It appears that the N-terminal sequence targets the protein to the secretory vesicle pathway (Fig. 5, *A5* and *A6*) and that sequences contained in the 27-amino acid C-terminal hydrophobic stretch subsequently target the protein to the cell surface (Fig. 5, *A4* and *B1*).

DISCUSSION

We have isolated and molecularly characterized a cDNA encoding an invertebrate intestinal mucin whose properties suggest potential roles in midgut physiology. The high abundance of *AgMuc1* mRNA and the probable membrane association of the protein raises the possibility of its involvement in mosquito–parasite interactions.

AgMuc1 is likely to be displayed on the surface of the microvilli of midgut epithelial cells. This possibility is strongly suggested by two of our experiments. First, the GFP-mucin fusion protein (72GFP27) is localized on the surface of Sf21 cells (Fig. 5); second, an antiserum raised against microvilli recognizes the recombinant mucin fusion protein (Fig. 4). Furthermore, certain sequence features distinguish AgMuc1 from vertebrate intestinal mucins and from insect peritrophic matrix mucins. Two vertebrate intestinal mucins, MUC1 and MUC2, lack putative GPI-anchor and hydrophobic C-terminal sequences, and have a significantly larger number of threonine-rich repeat units (25, 26). The *A. gambiae* Ag-Aper1 peritrophic matrix protein also lacks GPI-anchor sequences and contains cysteine-rich carbohydrate-binding domains which may bind to the chitinous fibers of the peritrophic matrix (8, 36). In contrast, AgMuc1 contains a putative GPI-anchor sequence, has hydrophobic domains at both termini which are separated from the central repeat region by nonrepetitive sequences, and lacks carbohydrate-binding domains. It shares these features with the mucins of the parasite *T. cruzi*, potentially defining a new family of membrane-associated mucins.

AgMuc1 is presumed to be highly glycosylated and may constitute an important component of the carbohydrate-rich layer previously observed by electron microscopy at the microvillar surface (10, 11). Similar membrane-associated intestinal mucins may also exist in vertebrates, but the membrane-associated mucins that have been characterized to date in vertebrates are nonintestinal and have different functions. For instance, human membrane-associated mucin MUC1 may play a role in tumor progression, in metastasis (37–39), and in cell–cell or cell–matrix interactions (40, 41).

Mucins play important roles in cell–cell interactions in multicellular organisms (reviewed in ref. 42) and have been implicated in the interactions between unicellular parasites with their hosts. For instance, the galactose-specific adherence protein from *Entamoeba histolytica* binds to human colonic mucins and epithelial cells (43). The mucin-type glycoproteins of *T. cruzi* have also been suggested to be involved in the interaction with and/or invasion of mammalian host cells (44, 45). Accumulating evidence suggests that the carbohydrate on the mosquito microvillar surface might serve as a receptor for the initial association between the malaria parasites and the gut epithelial cells. Ramasamy *et al.* (13, 14) found that chitotriose and antibodies against midgut glycoproteins inhibited malaria parasite development. Moreover, Shahabuddin and co-workers (46) found in an *in vitro* assay that *P. gallinaceum* ookinetes could not bind to *Aedes aegypti* midguts after chemical modification of epithelial cell surface carbohydrates. As a major component of the microvillar glycocalyx, AgMuc1 might prove to be involved in parasite–midgut interactions prior to invasion.

The existence of a length polymorphism within the AgMuc1 mucin domain enabled the genetic localization of this gene to a chromosomal region neighboring the determinants for refractoriness to the malaria parasite *P. cynomolgi* B. This finding is useful, as it provides a genomic landmark within this interesting chromosomal region. However, the cytogenetic localization of *AgMuc1* does not coincide with the estimated cytological localization of *Pen1* (F. Collins, personal communication) and is thus unlikely to directly affect the *Pen1*-determined encapsulation refractory trait. This is not surprising because the *Pen1* QTL was shown to affect encapsulation capacity, but not other traits relating to *Plasmodium* infection (17). However, the lack of identity with *Pen1* does not exclude the possibility that AgMuc1 might be involved in early stages of parasite–mosquito interactions. Polymorphic AgMuc1 proteins might exhibit altered binding affinity or specificity to the ookinete, affecting its capacity to interact with the midgut epithelial cell surface prior to invasion, and thus might lead to differences in the vectorial capacity in strains expressing different alleles. Interaction assays of ookinetes with polymorphic AgMuc1 proteins and genetic experiments monitoring potential cosegregation of limited infection intensity phenotypes and *AgMuc1* alleles could test the potential role of this intestinal mucin in mosquito–parasite interactions. More broadly, it is relevant that the insect gut is an important target for pest control (47–49). Characterization of insect intestinal mucins could lead to the development of new strategies for disease and pest control.

1. Billingsley, P. F. & Lehane, M. J. (1996) in *Biology of the Insect Midgut*, eds. Lehane, M. J. & Billingsley, P. F. (Chapman & Hall, London), pp. 3–25.
2. Jacobs-Lorena, M. & Oo, M. M. (1996) in *The Biology of Disease Vectors*, eds. Beaty, B. J. & Marquardt, W. C. (Univ. Press of Colorado, Niwot), pp. 318–332.
3. Tellam, R. L. (1996) in *Biology of the Insect Midgut,* eds. Lehane, M. J. & Billingsley, P. F. (Chapman & Hall, London), pp. 86–114.
4. Del Bene, G, Dallai, R. & Marchini, D. (1991) *Int. J. Morphol. Embryol.* **20,** 15–24.
5. Elvin, C. M., Vuocolo, T., Pearson, R. D., East, I. J., Riding, G. A., Eisemann, C. H. & Tellam, R. L. (1996) *J. Biol. Chem.* **271,** 8925–8935.
6. Casu, R., Eisemann, C., Pearson, R., Riding, J., East, I., Donaldson, A., Cadogan, L. & Tellam, R. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 8939–8944.
7. Wang, P. & Granados, R. R. (1997) *J. Biol. Chem.* **272,** 16663–16669.
8. Shen, Z. & Jacobs-Lorena, M. (1998) *J. Biol. Chem.* **273,** 17665–17670.
9. Terra, W. R., Ferreira, B. P. & Dillon, R. J. (1996) in *Biology of the Insect Midgut,* eds. Lehane, M. J. & Billingsley, P. F. (Chapman & Hall, London), pp. 153–194.
10. Kitajima, E. W. (1995) *Cytobiologie* **11,** 299–303.
11. Rudin, W. & Hecker, H. (1989) *Parasitol. Res.* **75,** 268–279.
12. Lane, N. J., Dallai, R. & Ashhurst, D. E. (1996) in *Biology of the Insect Midgut,* eds. Lehane, M. J. & Billingsley, P. F. (Chapman & Hall, London), pp. 115–150.
13. Ramasamy, R., Wanniarachchi, I. C., Srikrishnaraj, K. A. & Ramasamy, M. S. (1997) *Biochim. Biophys. Acta* **1361,** 114–122.
14. Ramasamy, M. S., Kulasekera, R., Wanniarachchi, I. C., Srikrishnaraj, K. A. & Ramasamy, R. (1997) *Med. Vet. Entomol.* **11,** 290–296.
15. Vernick, K. D., Fujioka, H., Seeley, D. C., Tandler, B., Aikawa, M. & Miller, L. H. (1995) *Exp. Parasitol.* **80,** 583–595.
16. Collins, F. H., Sakai, R. K., Vernick, K. D., Paskewitz, S., Seeley, D. C., Miller, L. H., Collins, W. E., Campbell, C. C. & Gwadz, R. W. (1986) *Science* **234,** 607–610.
17. Zheng, L., Cornel, A. J., Wang, R., Erfle, H., Voss, H., Ansorge, W. & Kafatos, F. C. (1997) *Science* **276,** 425–428.
18. Lemos, F. J., Cornel, A. J. & Jacobs-Lorena, M. (1996) *Insect Biochem. Mol. Biol.* **26,** 651–658.
19. Dimopoulos, G., Seeley, D., Wolf, A. & Kafatos, F. C. (1998) *EMBO J.* **17,** 6115–6123.
20. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
21. Kumar, V. & Collins, G. H. (1996) *Insect Mol. Biol.* **3,** 41–47.
22. English, L. H., Readdy, T. L. & Bastin, A. E. (1991) *Insect Biochem.* **21,** 177–184.
23. Gum, J. R. (1995) *Biochem. Soc. Trans.* **23,** 795–799.
24. Gendler, S. R. & Spicer, A. P. (1995) *Annu. Rev. Physiol.* **57,** 607–634.
25. Gum, J. R., Jr., Hicks, J. W., Toribara, N. W., Siddiki, B. & Kim, Y. S. (1994) *J. Biol. Chem.* **269,** 2440–2446.
26. Gum, J. R. Jr., Ho, J. J. L., Pratt, W. S., Hicks, J. W., Hill, A. S., Vinall, L. E., Roberton, A. M., Swallow, D. M. & Kim, Y. S. (1997) *J. Biol. Chem.* **272,** 26678–26686.
27. Spicer, A. P., Parry, G., Patton, S. & Gendler, S. J. (1991) *J. Biol. Chem.* **266,** 15099–15109.
28. von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690.
29. Englund, P. T. (1993) *Annu. Rev. Biochem.* **62,** 121–138.
30. Schenkman, S., Ferguson, M. A., Heise, N., de Almeida, M. L. Mortara, R. A. & Yoshida, N. (1993) *Mol. Biochem. Parasitol.* **59,** 293–303.
31. Reyes, M. B., Pollevick, G. D. & Frasch, A. C. (1994) *Gene* **140,** 139–140.
32. Di Noia, J. M., Sanchez, D. O. & Frasch, A. C. (1995) *J. Biol. Chem.* **270,** 24146–24149.
33. Gum, J. R., Jr., Hicks, J. W., Lagace, R. E., Byrd, J. C., Toribara, N. W., Siddiki, B., Fearney, F. J., Lamport, D. T. A. & Kim, Y. S. (1991) *J. Biol. Chem.* **266,** 22733–22738.
34. Di Noia, J. M., Pollevick, G. D., Xavier, M. T., Previato, J. O., Mendoca-Previato, L., Sanchez, D. O. & Frasch, A. C. (1996) *J. Biol. Chem.* **271,** 32078–32083.
35. Hong, T., Summers, M. D. & Braunagel, S. C. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 4050–4055.
36. Shen, Z. & Jacobs-Lorena, M. (1999) *J. Mol. Evol.* **48,** 341–347.
37. Burdick, M. D., Harris, A. & Reid, C. J. (1997) *J. Biol. Chem.* **272,** 24198–24202.
38. Cao, Y., Blohm, D., Ghadimi, B. M., Stosiek, P., Xing, P. X. & Karsten, U. (1997) *J. Histochem. Cytochem.* **45,** 1547–1557.
39. Makiguchi, Y., Hinoda, Y. & Imai, K. (1996) *J. Cancer Res.* **87,** 505–511.
40. Hilkens, J., Vos, H. L., Wesseling, J., Boer, M., Storm, J., van der Valk, S., Calafat, J. & Patriarca, C. (1995) *Cancer Lett.* **90,** 27–33.
41. Wesseling, J., van der Valk, S. W., Vos, H. L., Sonnenberg, A. & Hilkens, J. (1995) *J. Cell Biol.* **129,** 255–265.
42. van Klinken, B. J., Dekker, J., Buller, H. A. & Einerhand, A. W. (1995) *Am. J. Physiol.* **269,** 913–627.
43. Ravdin, J. I., Shain, D. C. & Kelsall, B. L. (1993) *Vaccine* **11,** 214–246.
44. Serrano, A. A., Schenkman, S., Yoshida, N., Mehlert, A., Richardson, J. M. & Ferguson, M. A. (1995) *J. Biol. Chem.* **270,** 27244–27253.
45. Burleigh, B. A. & Andrews, N. W. (1995) *Annu. Rev. Microbiol.* **49,** 175–200.
46. Zieler, J., Nawrocki, J. & Shahabuddin, M. (1999) *J. Exp. Biol.* **202,** 485–495.
47. Vadlamudi, R. K., Weber, E., Ji, I., Ji, T. H. & Bulla, L. A., Jr. (1995) *J. Biol. Chem.* **270,** 5490–5494.
48. Francis, B. R. & Bulla, L. A., Jr. (1997) *Insect Biochem. Mol. Biol.* **27,** 541–550.
49. Knight, P. J., Knowles, B. H. & Ellar, D. J. (1995) *J. Biol. Chem.* **270,** 17765–17770.