JOURNAL OF VIROLOGY, July 2005, p. 7979–7989 Vol. 79, No. 13
0022-538X/05/$08.00+0 doi:10.1128/JVI.79.13.7979–7989.2005

A Cellular *Drosophila melanogaster* Protein with Similarity to Baculovirus F Envelope Fusion Proteins

Oliver Lung† and Gary W. Blissard*

Boyce Thompson Institute, Cornell University, Ithaca, New York

Received 5 December 2004/Accepted 8 March 2005

Baculovirus F (fusion) proteins are found in the envelopes of budded virions. Recently a *Drosophila melanogaster* gene (CG4715) that encodes a protein with sequence similarity to baculovirus F proteins was discovered. To examine similarities and differences with baculovirus F proteins, we cloned the *D. melanogaster* cellular F (*Dm-cF*) protein gene and analyzed *Dm-cF* expression and localization. The predicted Dm-cF protein sequence lacks a furin cleavage site, and transiently expressed Dm-cF showed no protein cleavage and no detectable membrane fusion activity. In cell localization studies, transiently expressed Dm-cF was localized to intracellular organelles in *D. melanogaster* S2 cells, unlike baculovirus F proteins, which localize to cellular plasma membranes. Using reverse transcriptase PCR and Western blot analysis to examine *Dm-cF* expression in animals, we detected *Dm-cF* expression in both larval and adult *D. melanogaster* cells. However, *Dm-cF* expression was detected only in third instar larvae and adults, suggesting that *Dm-cF* expression may be developmentally regulated. We also identified genes related to *Dm-cF* in the genomes of two other *Drosophila* species, *Drosophila yakuba* and *Drosophila pseudoobscura*, and the mosquito *Anopheles gambiae*. These observations suggest that *f* genes may be present in the genomes of many insects. Conservation within and between 22 baculovirus and 4 insect F proteins was examined in detail. These studies demonstrate that *Dm-cF* is expressed in *D. melanogaster* and suggest that if baculovirus *f* genes are derived from a host cellular *f* gene, the function appears to have changed substantially upon adaptation to the baculovirus infection cycle.

Baculoviruses are a large and diverse group of double-stranded DNA viruses that are pathogens of arthropods in the class *Insecta* (7, 16). To date, the genomes of approximately 26 baculoviruses have been sequenced, and these include viruses that infect insects in the orders *Lepidoptera*, *Diptera*, and *Hymenoptera*. Baculovirus virions are enveloped, and the envelope fusion proteins are critical structural proteins that mediate the fusion of the viral envelope and the host membrane during virus entry. Two different types of envelope fusion proteins have been described from budded virions of lepidopteran baculoviruses: GP64 and F (fusion). In the best-studied baculovirus, *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV), the GP64 protein is essential for both viral entry and efficient viral budding (34, 37). Although GP64 is essential in AcMNPV and closely related lepidopteran baculoviruses, many other baculoviruses do not encode a *gp64* gene. All baculoviruses infecting lepidopteran hosts appear to have a gene encoding an envelope protein called F. Thus, some lepidopteran baculoviruses (the group I nucleopolyhedroviruses [NPVs]) carry a *gp64* gene and an *f* gene while others (group II NPVs and granuloviruses [GVs]) appear to carry only an *f* gene (reviewed in reference 40). In the *gp64*-minus baculoviruses that have been studied in detail, the F protein appears to serve the same essential role as GP64. For example, F proteins from *Lymantria dispar* MNPV (LdMNPV) and *Spodoptera exigua* MNPV (SeMNPV) have a low pH-triggered membrane fusion activity (22, 39) and can rescue infectivity by a *gp64*-null mutant AcMNPV virus (28, 34, 37). In contrast, in AcMNPV, which encodes both GP64 and F, the F protein is not essential for virus propagation (29).

F proteins from the group II NPVs, LdMNPV and SeMNPV, are processed by cleavage at a consensus furin cleavage site (22, 42, 49). Mutational analyses of these two F proteins suggest that both the furin cleavage site and a fusion peptide are necessary for membrane fusion activity (40, 42, 48, 49). However, F proteins from group I NPVs, AcMNPV and OpMNPV (viruses with both GP64 and F), do not contain a predicted furin cleavage site, and low pH-triggered membrane fusion activity was not observed when these F proteins were examined in standard syncytium formation assays (reference 41 and unpublished observations). Thus, in baculoviruses that contain *gp64* genes, the F proteins are thought to be fusion incompetent. In nonlepidopteran baculoviruses, some viruses contain *f* genes, while others do not. An *f* gene is found in the genome of the dipteran baculovirus, *Culex nigripalpus* NPV (CuniNPV) (2), although its expression and function have not been examined. Neither an *f* gene nor a *gp64* gene is present in the genomes of the two sequenced sawfly (hymenopteran) baculoviruses (18, 27). Thus, different envelope protein genes appear to have been acquired and/or lost during the evolution of the different groups of baculoviruses.

In addition to the baculovirus F proteins, endogenous insect retroviruses (errantiviruses) also encode related F proteins (30, 43, 46). Furthermore, an F protein gene that appears to be of host cell origin was identified in the genome of *Drosophila melanogaster* (1, 43). We have called the later gene *Dm-cF* (for *D. melanogaster* cellular *f*). In the current study, we describe and examine similar *f*-like genes in the genomes of *Drosophila yakuba*, *Drosophila pseudoobscura*, and *Anopheles gambiae*. Us-

* Corresponding author. Mailing address: Boyce Thompson Institute, Cornell University, Tower Road, Ithaca, NY 14853. Phone: (607) 254-1366. Fax: (413) 480-4762. E-mail: gwb1@cornell.edu.

† Present address: Department of Biological Sciences, University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K 3M4, Canada.

ing *f* genes identified in the genomes of 22 baculoviruses, we performed a detailed comparative analysis of F protein sequence conservation within the *Baculoviridae*. We also examined F protein conservation between baculovirus F proteins and the recently identified cellular F proteins. We then examined *Drosophila Dm-cF* gene expression and asked whether the Dm-cF protein exhibits properties similar to viral F proteins. To characterize the *Dm-cF* gene and its product, we cloned *Dm-cF*, examined it for possible membrane fusion activity, characterized its expression in *D. melanogaster* larvae and adults, and examined its subcellular localization in *D. melanogaster* S2 cells.

MATERIALS AND METHODS

Dm-cF cloning and expression. To clone the *Dm-cF* gene, *D. melanogaster* genomic DNA (100 ng; OrR strain) was used as a template for PCR amplification of the Dm-cF coding region with Vent DNA polymerase (New England Biolabs). The primers P5′DmCG4715GW (5′-CACCATGAAAGCGATTAGT TTGGTATTC-3′) and P3′DmCG4715Δstop (5′-GATTTCATGTTTTTCCAGC AGC-3′) were used to amplify the Dm-cF open reading frame (ORF). PCR conditions were set for 94°C for 2 min; five cycles of 94°C for 10 s, 52°C for 30 s, and 72°C for 1 min 40 s; 20 cycles of 94°C for 10 s, 55°C for 30 s, and 72°C for 1 min 40 s; and a final step at 72°C for 10 min. The PCR product was cloned into pENTR-D (Invitrogen) and confirmed by sequencing the entire insert. The Dm-cF ORF was introduced into vectors for expression in *Escherichia coli* and *D. melanogaster* S2 cells using the Gateway system (Invitrogen). LR (*attL* and *attR* recombination) reactions were performed with pDEST49 (Invitrogen) and pGW-MAL (a gift from Mike Goldberg, Cornell University) to generate expression plasmids for expressing histidine-tagged (Dm-cF-His6) and maltose binding protein (MBP)-tagged (MBP-Dm-cF) Dm-cF fusion proteins in *E. coli*, under the control of a T7 promoter. The resulting Dm-cF-His6 and MBP-Dm-cF fusion proteins were used for antibody production and antibody purification as described below. The Dm-cF ORF was also introduced into a Gateway destination vector pMET-GFP (where GFP is green fluorescent protein; a gift from John Lis, Cornell University) by LR reaction to generate an expression plasmid for expressing Dm-cF-GFP under the control of a metalothionine promoter in *D. melanogaster* S2 cells. To generate an epitope-tagged Dm-cF protein in insect cells, plasmid pAc5-1/V5-His6 (Invitrogen) was first converted into a Gateway destination vector by digesting with EcoRV, dephosphorylating the ends with calf intestinal alkaline phosphatase (Promega), and blunt-end ligating it to reading frame cassette C (RFC; Invitrogen) to generate pDEST Ac5-1/V5His6-RFC. The expression plasmid was then generated by LR reaction with pDEST Ac5-1/V5His6-RFC and pENTR-Dm-cF, and the resulting construct (pAc5-1P-Dm-cF-/V5) was used to express a Dm-cF-V5 fusion protein with a V5 epitope tag under the control of a *Drosophila* actin promoter.

Identification of *Dm-cF*-like genes in other insect genomes. An *A. gambiae* putative cellular *f* gene (*Ag-cF*) was identified as a protein fragment (AgCP15405; accession no. Q7Q8H1) by a FASTA search using Dm-cF as a query sequence. AgCP15405 did not contain an initiating methionine codon or a stop codon. Therefore, a TBLASTN search was performed to find the genomic sequence that encoded the AgCP15405 translated product. *A. gambiae* sequence accession number AAAB01008944 was found to contain an entire ORF with start and stop codons that could encode an F-like protein. The region containing the C9 residue (see Results) was confirmed by PCR amplifying *Ag-cF* from genomic DNA using primers P5′Ag-cF1 (5′-GGCGAATTCGATAGTCATGT ATCCACTGTTC-3′) and P3′Ag-cF1 (5′-GGCCGAATTCTTACCTTGGCAG GGATGTATCG-3′) and sequencing with primer P5′Ag-cF4 (5′-TCCACCAC CGACGACTG). A *D. Yakuba* putative cellular *f* gene (*Dy-cF*) was identified with a TBLASTN search of the whole genome shotgun sequence database with the Dm-cF protein sequence. *D. yakuba* contigs 14.19 (accession no. AAEU0103795) and 14.18 (accession no. AAEU01003794) encoded the amino and carboxy termini of the F protein, respectively. Therefore, nucleotides 120 to 790 from the contig 14.19 sequence were combined with nucleotides 130569 to 129555 from contig 14.18 to form the entire ORF for Dy-cF. The entire Dy-cF ORF was amplified by PCR from *D. yakuba* genomic DNA using the primers P5′Dy-cF1 (5′-GGCCGAATTCAGACATATATGCTCTACC-3′) and P3′Dy-cF1 (5′-GGCCGAATTCTCTTATCTATTCGGTATTTCCAC-3′) and confirmed by sequencing the entire ORF. TBLASTN searches of the whole genome shotgun sequence database using Dm-cF as the query showed that a putative cellular *f* gene from *Drosophila pseudoobscura* (*Dp-cF*) was encoded by contig 3880 (accession no. AADE01000622). Contig 3880 encoded an ORF that showed similarity to the first 365 amino acids (aa) of Dm-cF and a separate ORF from an overlapping frame that had similarity with aa 370 to 490 of Dm-cF, suggesting a frame-shift mutation or a sequencing error. To examine this, the *Dp-cF* gene was PCR amplified from *D. pseudoobscura* genomic DNA using P5′Dp-cF1 (5′-GGCCGAATTCAAACAAAATTAAACACTCAGC-3′) and P3′Dp-cF1 (5′-GGCCGAATTCACTCACTCCACATTGTTCCTCTC-3′) and sequenced with primer P5′Dp-cF4 (5′-AGCCCCATCAAATCACCTT-3′). Resequencing led to the identification of a corrected sequence (nucleotides 29853 to 29865 from AADE01000622 were deleted and replaced with ACAGTCACCAA in the corrected sequence) that connected the two separate ORFs into a single large ORF that encoded a product with similarity across the entire Dm-cF sequence. Multiple sequence alignment of insect cellular F sequences was performed on the four full-length putative insect F protein sequences with MultAlign (9; http://prodes.toulouse.inra.fr/multalin/multalin.html) using default parameters and a gap initiation penalty of 3. The alignment was formatted using Boxshade 3.2.1 (http://www.ch.embnet.org/software/BOX_form.html).

Motif predictions and multiple sequence alignments. Sequence similarity searches were performed with FASTA3 using default parameters (http://www.ebi.ac.uk/fasta33/), and BLAST searches were performed using default parameters (3; http://www.ncbi.nlm.nih.gov/BLAST/). Signal sequence prediction was performed with SignalP 3.0 (6; http://www.cbs.dtu.dk/services/SignalP/), and transmembrane domain prediction was performed with TMHMM server V2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0). Coiled-coil domain prediction was performed by LearnCoil-VMF (44; http://web.wi.mit.edu/kim/pub/Computing/computing.html). Multiple sequence alignment of F protein sequences was performed by the ClustalW method with the Megalign program of Lasergene (DNASTAR Inc). Default parameters were used with the exception of the Gonnet series of protein weight matrix and a gap penalty of 3. F protein sequences were obtained from the GenBank database from the following sources and used in the sequence alignment: AdhoNPV ORF 118 (accession no. GI29567203), AcMNPV ORF 23 (accession no. GI9627765), BmNPV ORF 14 (accession no. GI9630831), CfMNPV ORF 21 (accession no. GI47174745), EppoNPV ORF 19 (accession no. GI5320676), HearNPV ORF 133 (accession no. GI2957618), HzSNPV ORF 137 (accession no. GI8028707), LdMNPV ORF 130 (accession no. GI9631097), MacoNPV-A ORF 9 (accession no. GI20069888), MacoNPV-B ORF 9 (accession no. GI2254910), OpMNPV ORF 21 (accession no. GI11038432), RaouNPV ORF 21 (accession no. GI23577830), SeMNPV ORF 8 (accession no. GI9634229), SpliNPV ORF 136 (accession no. GI5617604), AdorGV ORF 23 (accession no. GI32698562), AgseGV ORF 25 (accession no. GI46309429), CrleGV ORF 30 (accession no. GI33622230), CpGV ORF 31 (accession no. GI460068), PhopGV ORF 27 (accession no. GI21637008), PlxyGV ORF 26 (accession no. GI11067782), XecnGV ORF 23 (accession no. GI9635277), CuniNPV ORF 104 (accession no. GI532098), and *D. melanogaster* CG4715 (accession no. GI19920492).

RT-PCR analysis of *Dm-cF* expression in *D. melanogaster*. Using an RNeasy kit (QIAGEN), *D. melanogaster* RNAs were isolated from 4- to 16-h embryo; first, second, and third instar larva; white prepupa; late pupa; adult head; adult thorax plus abdomen; S2 cells; and Kc cells. DNase was inactivated in each 0.2-μg total RNA sample by the addition of 5 μl DNA-free (Ambion) and reverse transcriptase PCRs (RT-PCRs) were performed in the presence and absence of Superscript II (Invitrogen). One microliter of the RT product was used in each PCR with *Taq* DNA polymerase (Promega) and Dm-cF-specific primers P5′ CG4715RT (5′-GATGCGGGGATTCAACAACAGTC-3′) and P3′4715RT (5′-GTGGGCAGAGATTATGGCTTCAGT-3′). Primers (rp49+1f, 5′-AGCTTCA AGATGACCATCCGCCC-3′, and rp49+333r, 5′-TGAGCAGGACCTCCAGC TCGCGC-3′, specific for the ribosomal protein 49 (*rp49*) gene were used as a positive control. Conditions for PCRs were as follows: 94°C for 5 min; 30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s; and a final step at 72°C for 10 min. Products of the PCR were examined in a 0.8% agarose gel.

Antibody production and Western blot analysis. MBP-Dm-cF and Dm-cF-His6 fusion proteins were prepared by growing DH5α cells containing either plasmid pExpMBP-Dm-cF or pExpDm-cF-His6 in LB plus glucose at 37°C to an optical density at 600 nm of 0.5 and then inducing expression with isopropyl-β-D-thiogalactopyranoside (0.4 mM final concentration) for 2 h at 37°C. Cell pellets were collected and proteins were denatured in 2× Laemmli disruption buffer at 100°C for 10 min, and then loaded onto a 6% sodium dodecyl sulfate-polyacrylamide gel. Gels containing Dm-cF fusion proteins were stained with Coomassie blue dissolved in distilled H_2O and destained in distilled H_2O until the band that corresponds to the fusion protein was visible, and the Dm-cF-His6 protein band was excised from the gel. The gel slice containing the Dm-cF-His6 protein was used for antibody production in rabbits (Cocalico Biologicals, Reamstown, PA).

Anti-Dm-cF antibodies were affinity purified from rabbit serum by incubating the rabbit anti-Dm-cF crude serum with Immobilon-P membrane (Millipore) containing MBP-Dm-cF protein overnight at room temperature. Membranes were then washed three times with 100 ml of Tris-buffered saline, pH 7.5, for 10 min each, and the antibody was eluted twice with 1.8 ml each of 100 mM glycine, pH 2.5. The pooled purified antibody solution (3.6 ml) was neutralized with 0.4 ml 1 M Tris, pH 8.0, and used for Western blotting at a 1:50 dilution. *D. melanogaster* extracts for Western blotting were collected from 30 first instar, 18 second instar, and 9 third instar larvae, as well as from nine unmated adult males and nine unmated adult females. Samples were homogenized in 30 μl of phosphate buffered saline (PBS), pH 7.4, plus protease inhibitor cocktail (Boehringer Mannheim), and then 30 μl of 2× Laemmli buffer was added and proteins were denatured by heating to 100°C for 5 min. One-third (20 μl) of each sample was loaded into a lane of a 6% sodium dodecyl sulfate-polyacrylamide gel. Western blotting was performed as described previously (28).

Protein localization by immunofluorescence microscopy. For transient expression of Dm-cF in insect cells, plasmid DNA (5 μg) was used to transfect approximately 1×10^6 *D. melanogaster* S2 cells using a calcium phosphate transfection kit (Invitrogen). After 48 h, cells were fixed with 4% paraformaldehyde for 10 min, washed three times with PBS (pH 7.4), and then permeabilized by incubation in 0.2% Triton X-100 in PBS for 10 min. Cells were washed three times, blocked with 5% bovine serum albumin in PBS for 1 h, incubated with a 1:200 dilution of primary antibody for 1 h, washed three times (10 min each) with PBS, incubated with secondary antibody for 1 h, washed three times with PBS, and mounted for observation with a Leica TCS SP2 confocal microscope. Primary antibodies for the above studies included mouse anti-V5 (Invitrogen), mouse anti-rat-KDEL (Calbiochem), mouse anti-*Drosophila* Golgi (Calbiochem), anti-M6PR (8), and anti-cathepsin D (38). Secondary antibodies included Alexafluor 488-conjugated anti-mouse antibody (Molecular Probes) and Cy3-conjugated anti-rabbit and anti-mouse antibodies (Jackson Immunoresearch). For mitochondrial localization, cells were transfected as above and incubated for 30 min with MitoTracker Red CMXRos (Molecular Probes) and then fixed with 3.7% formaldehyde and processed for immunocytochemistry.

Analysis of membrane fusion activity. For studies of possible membrane fusion activity, *D. melanogaster* S2 cells were transfected with plasmids encoding Dm-cF-V5 and Dm-cF-GFP as described above. At 48 h posttransfection, cells were washed three times with PBS (pH 7.4), incubated in PBS at pH 5.0 for 30 min, and then washed three times with PBS (pH 7.4) and placed in fresh Schneider's complete medium plus 10% fetal bovine serum. Cells expressing Dm-cF-GFP were examined for GFP fluorescence to confirm protein expression, and both Dm-cF-GFP and Dm-cF-V5 expressing cells were examined for signs of syncytium formation at various times from 30 min to 24 h after treatment with low pH. For ionophore treatment, similarly transfected *D. melanogaster* S2 cells were incubated for 30 min in either 4Br-23187 or ionomycin (5 μM in Schneider's complete medium without serum). Cells were then prepared for immunofluorescence microscopy as described above using an anti-V5 antibody (Invitrogen) as the primary antibody and a goat anti-mouse fluorescein-conjugated antibody (Sigma) as the secondary antibody. Cells were examined with an Olympus IX70 fluorescence microscope.

RESULTS

Cloning and comparative analysis of F proteins. Recently, a gene encoding a protein with sequence similarity to baculovirus F proteins was identified in the genome of the fruit fly, *D. melanogaster* (43). The *D. melanogaster f* ortholog (*CG4715*), which we refer to as *Dm-cF* to denote its relationship to baculovirus *f* genes, encodes a predicted protein of 531 aa with a predicted mass of 58.9 kDa. The gene maps to cytological position 21F1 on chromosome 2L of the *D. melanogaster* genome. Through similarity searches with the Dm-cF protein sequence (see Materials and Methods), we also identified *f*-like genes from each of two other *Drosophila* species and from the mosquito *A. gambiae* (19) (Fig. 1). A *D. yakuba* gene (*Dy-cF*, for *D. yakuba* cellular F) that also maps to chromosome 2L encodes an F protein with ~61% amino acid sequence identity to Dm-cF. A similar *D. pseudoobscura f* ortholog (*Dp-cF*, for *D. pseudoobscura* cellular *f*) encodes a protein with ~37% amino acid identity to Dm-cF. A gene mapped to chromosome 2R of the *A. gambiae* genome, *Ag-cF* (*A. gambiae* cellular *f*) encodes a predicted protein with ~21.6% identity to Dm-cF. The observation that there are no identifiable retrovirus *gag* and/or polymerase (reverse transcriptase) genes upstream of these *A. gambiae*, *D. melanogaster*, *D. pseudoobscura*, and *D. yakuba f*-like genes suggests that these genes are not part of an integrated retrovirus genome. We also identified approximately 16 *f*-like genes in the currently available genome sequence of the silk moth *Bombyx mori* (33). However, all *f*-like genes identified so far from *B. mori* appear to be associated with integrated retrovirus genomes (data not shown). The presence of cellular *f* genes in the genomes of four dipteran insect species raises the possibility that F proteins may be present in many insects, and we cannot rule out the possibility that a lepidopteran cellular *f* gene may yet be identified.

Alignment of the three putative *Drosophila* cellular F proteins shows that both Dy-cF and Dp-cF have extended N-terminal sequences when compared with the Dm-cF sequence (Fig. 1). Several observations suggest that perhaps the second ATG in the Dy-cF ORF and the third ATG in the Dp-cF ORF may code for the initiator methionine. First, like baculovirus F proteins, Dm-cF and the proteins encoded by the *f* genes from *D. yakuba* and *D. pseudoobscura* contain a predicted signal sequence and a transmembrane domain (Fig. 2A). However, the length of the signal peptide would be 49 aa for Dy-cF and 62 aa for Dp-cF if the first ATG in each ORF codes for the initiator methionine. If the second ATG in the Dy-cF ORF and the third ATG in Dp-cF ORF coded for the initiator methionine, the signal peptide lengths would be 19 and 20 aa for Dy-cF and Dp-cF, respectively. This is more typical of eukaryotic signal peptides and is comparable to the signal peptide length of 20 aa predicted for Dm-cF. Supporting this possibility is the observation that the contexts of the first ATGs of the Dy-cF and Dp-cF ORFs do not match either the −3 or +4 positions of the consensus Kozak translation initiation sequence (G/A)NNATGG (25), while the second ATG for both Dy-cF and Dp-cF and the third ATG for Dp-cF would match the most critical −3 position of the Kozak consensus sequence (data not shown). Thus, these downstream ATG codons may serve as initiator codons by leaky scanning (24). Unlike the baculovirus F proteins from group I NPVs and GVs (those considered "fusion competent") (22, 39, 42, 48, 49), the F proteins from all three *Drosophila* species lack (i) a predicted furin cleavage site, (ii) a predicted fusion peptide, and (iii) a predicted coiled-coil domain. In contrast, the predicted *Anopheles* protein, Ag-cF, does have a potential furin cleavage site RTKR (Fig. 1), although it is not known whether Ag-cF is cleaved at this site.

F proteins from different lepidopteran baculoviruses all have a predicted signal peptide and a transmembrane domain, but F proteins are relatively divergent and may show <20% amino acid sequence identity. Therefore, identifying conserved residues shared among different F proteins will be important for identifying distantly related F orthologs. Rohrmann and Karplus (43) reported the presence of 10 conserved cysteine residues in the ectodomain and one conserved cysteine residue in the cytoplasmic domain of six lepidopteran baculovirus F proteins. In this study we performed a multiple sequence alignment of F proteins from 21 lepidopteran baculoviruses (14 NPVs and 7 GVs) to determine whether the conserved cys-

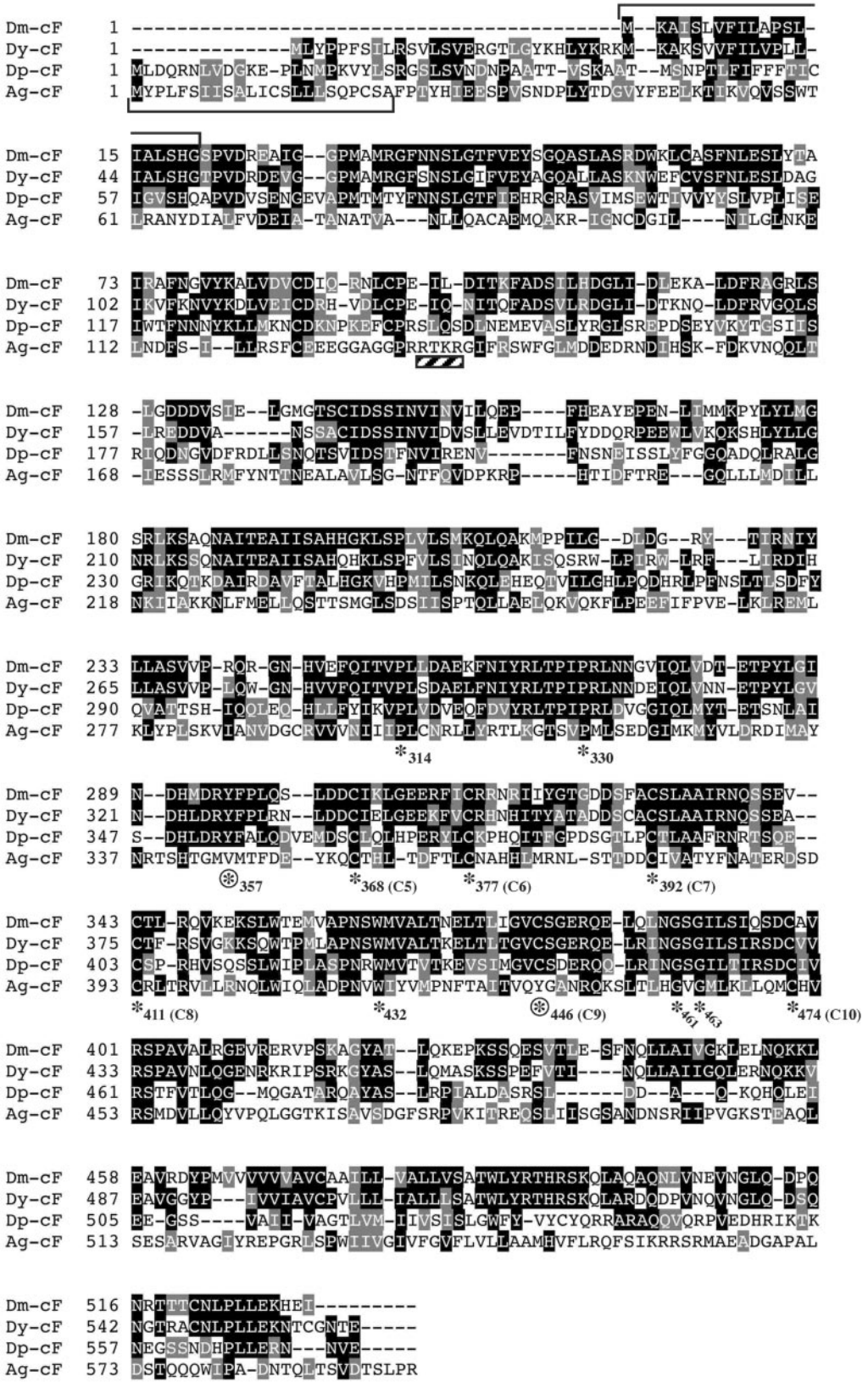
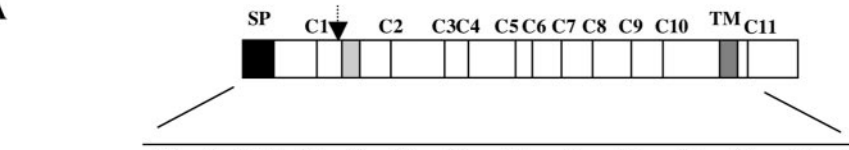


FIG. 1. Multiple sequence alignment of F proteins encoded by the cellular *f* genes from three *Drosophila* species and *Anopheles gambiae*. The predicted signal peptide of the three *Drosophila* proteins is marked by a bracket above the sequence, and the predicted signal peptide of the *A. gambiae* protein is marked by a bracket below the sequence. The consensus furin cleavage site ($RX^R/_KR$) in the Ag-cF gene is indicated by a hatched box below the sequence. The positions of highly conserved residues that are conserved in both baculovirus and insect cellular F proteins are marked by an asterisk below the residue, and conserved cysteine residues are indicated as C5 to C10 (see Fig. 2). For reference, numbers beside the asterisks indicate the corresponding amino acid position of the LdMNPV F protein, and circled asterisks indicate those positions conserved between baculovirus and *Drosophila* cellular proteins but not in the *Anopheles* or CuniNPV F protein.

A



	SP	C1	Furin	FP	C2	C3	C4	C5	C6	C7	C8	C9	C10	TM	C11
Lep Bac F	Y	Y	Y/N	Y/N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
CuniNPV F	Y	N	Y	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	N
Dm-cF	Y	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	N
Dy-cF	Y	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	N
Dp-cF	Y	N	N	N	N	N	N	Y	Y	Y	Y	Y	Y	Y	N
Ag-cF	Y	N	Y	N	N	N	N	Y	Y	Y	Y	N	Y	Y	N

B

Conserved Cys residue spacing

		C1-2	C2-3	C3-4	C4-5	C5-6	C6-7	C7-8	C8-9	C9-10	C10-11
Lep Bac F	Avg spacing	150	117	6	28	9	14	15	35	24	126
	max	150	121	19	35	10	14	18	37	31	133
	min	101	113	4	25	8	13	12	34	20	114
	SD	[10]	[3]	[3]	[2]	[1]	[0]	[2]	[1]	[3]	[6]
CuniNPV F		NA	NA	NA	NA	7	15	12	36	20	NA
Dm-cF		NA	NA	NA	NA	9	15	12	33	20	NA
Dy-cF		NA	NA	NA	NA	9	15	12	33	20	NA
Dp-cF		NA	NA	NA	NA	9	15	14	33	20	NA
Ag-cF		NA	NA	NA	NA	8	14	14	33*	21*	NA

C

Conserved Cys and Non-Cys residues and spacing

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
conserved residue	*G*	*F*	*C1*	*D*	*L*	*V*	*C2*	*L*	*W*	*P*	*P*	*C3*	*C4*	*Y*	*C5*	*C6*	*C7*	*C8*	*W*	*C9*	*G*	*G*	*C10*	*L*	*C11*
LdMNPV F residue	*33*	*51*	*88*	*168*	*172*	*189*	*214*	*260*	*277*	*314*	*330*	*332*	*338*	*357*	*368*	*377*	*392*	*411*	*432*	*446*	*461*	*463*	*474*	*532*	*603*
Avg spacing	17	33	78	3	15	25	46	16	35	16	1	6	18	9	9	14	15	21	13	13	1	8	57	69	
max	18	40	106	3	16	28	47	17	37	17	1	19	26	10	10	14	18	22	14	19	1	10	62	74	
min	17	26	44	3	11	21	44	14	34	15	0	4	16	8	8	13	12	20	13	8	1	7	51	63	
SD	0	4	9	0	2	2	1	1	1	1	0	3	2	1	1	0	2	1	0	3	0	1	4	3	
CuniNPV F										15				10	7	15	12	(-)		10	1	7			
Dm-cF										15				8	9	15	12	19	13	10	1	7			
Dy-cF										15				8	9	15	12	19	13	10	1	7			
Dp-cF										15				10	9	15	14	19	13	10	1	7			
Ag-cF										15				(-)		8	15	8	20	(-)		1	7		

D

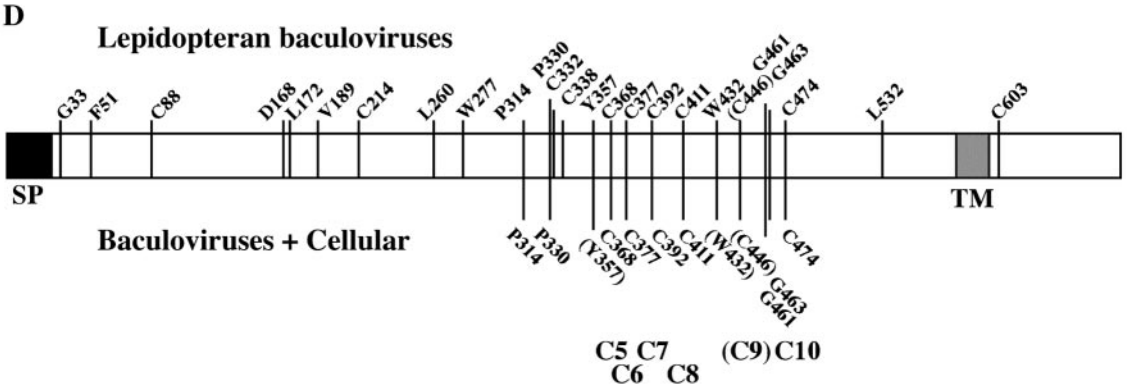


FIG. 2. Conserved positions in baculovirus and cellular F proteins. (A) The top panel is a schematic representation of an F protein showing the predicted signal peptide (SP), transmembrane domain (TM), and 10 conserved cysteine residues (C1 to C10) found in the ectodomains of F proteins from the 21 lepidopteran baculoviruses (Lep Bac F; see Materials and Methods for the baculoviruses included in the analysis). The bottom panel lists the presence (Y) or absence (N) of conserved domains and cysteine residues. A furin cleavage site (furin) and a fusion peptide (FP) are found in F proteins from group II NPVs and GVs but not in F proteins from group I NPVs and are therefore listed as Y/N. (B) The average spacings (Avg spacing), maximum spacings (max), minimum spacings (min), and standard deviations (SD) between neighboring conserved cysteines in the lepidopteran baculovirus F proteins (Lep Bac F) used in the analysis are listed (top grouping) and compared with similar spacings from CuniNPV F, Dm-cF, Dy-cF, Dp-cF, and Ag-cF (below). Note that because C9 is absent in Ag-cF, as indicated by an asterisk (*), the Tyr substitution was used for calculation of spacing. (C) Highly conserved residues in lepidopteran baculovirus F proteins are compared with similar positions in CuniNPV F, Dm-cF, Dy-cF, Dp-cF, and Ag-cF proteins. A list of 25 residues conserved in F proteins from all 21 lepidopteran baculoviruses are shown at the top. Conserved cysteine residues are indicated by a number (C1 to C11) corresponding to the conserved cysteine number in panel A. For reference, the residue number in the LdMNPV F protein is indicated. The spacing between neighboring conserved residues is shown below. Four underlined residues indicate residues that are identical in 20 of 21 lepidopteran baculovirus F proteins, whereas all others were identical in all 21 proteins. (D) Comparison of F protein residues conserved in lepidopteran baculovirus F proteins (top panel) with residues conserved in F proteins from both baculovirus and cellular genomes (bottom panel). Residues in parenthesis are those conserved between baculovirus and *Drosophila* cellular F proteins but not conserved in the *Anopheles* or CuniNPV F protein.

teine residues identified by Rohrmann and Karplus (43) were present in F proteins of other baculoviruses and in F proteins from the insect genomes. Sequence alignment and examination of the spacing between conserved cysteines show that all 21 lepidopteran baculovirus F proteins included in our analysis have the 10 conserved cysteine residues in the ectodomain and the conserved cysteine in the cytoplasmic domain (Fig. 2A). However, the F proteins from the mosquito baculovirus CuniNPV (CuniNPV F) and the fruit flies *D. melanogaster*, *D. yakuba*, and *D. pseudoobscura* contain only 6 (C5 to C10) of the 11 conserved cysteine residues found in the lepidopteran baculovirus F proteins (Fig. 2A). The putative cellular F protein from the mosquito *A. gambiae* (Ag-cF) similarly lacks conserved cysteines C1 to C4 and C11, but surprisingly C9 is also absent (Fig. 2A). Given the highly conserved nature of the C9 cysteine residue among all baculovirus F proteins and the *Drosophila* cellular F proteins, we resequenced this region of the *Ag-cF* gene (from DNA derived from field-collected *A. gambiae* insects) and confirmed the substitution of a Tyr codon for the C9 (Cys) codon in the Ag-cF coding region (Fig. 1) (note that Tyr 428 corresponds to C9 at position 446 in the LdMNPV F protein). An analysis of the spacing between the conserved cysteines in CuniNPV F, Dm-cF, Dy-cF, Dp-cF, and Ag-cF shows that they are in excellent general agreement and within the same range of spacings observed in the lepidopteran baculovirus F proteins (Fig. 2B).

Multiple sequence alignments of the lepidopteran baculovirus F proteins also identified a number of noncysteine residues that were conserved. In addition to the 11 conserved cysteine residues, we identified 14 additional highly conserved amino acid residues in the ectodomains of all 21 lepidopteran baculovirus F proteins (Fig. 2C and D). Of these 14 noncysteine residues, 10 are identical in all 21 proteins, and 4 are conserved in 20 of 21 proteins (Fig. 2C). The spacings between these conserved residues are also highly conserved (Fig. 2C). Of the conserved noncysteine residues, two proline residues (Fig. 2C, 10 and 11) and residues 14 to 23 also appear to be mostly conserved in the CuniNPV F and cellular F proteins Dm-cF, Dy-cF, Dp-cF, and Ag-cF (Fig. 1 and 2C). We noted that in the four cases where a conserved residue (noncysteine) is missing in just one of the 21 lepidopteran baculovirus F proteins analyzed, the nonconserved residue and the consensus residue belong to the same functional group. Additionally, in these four cases, the codon for the nonconserved residue and the conserved residue differ by a single nucleotide. Therefore, the differences may have resulted from either a sequencing error or a point mutation resulting in a conservative substitution.

Comparison of the presence and absence of conserved residues (Fig. 2B and C) also suggests that CuniNPV F protein is more similar to dipteran cellular F proteins than to lepidopteran baculovirus F proteins. This observation is also supported by similarity searches using FASTA and BLAST (3). The highest score for FASTA (version v3.4t23) searches with the predicted Dm-cF protein sequence was observed from the CuniNPV F protein, with ~24% amino acid identity (E = 3.2e-09) over a region of 255 aa across the central half of the protein. In BLASTP searches of the nr database with the CuniNPV F sequence, Dm-cF was identified as the highest-scoring non-CuniNPV F sequence identified (E = 2e-11).

Thus, there are 12 residues that are highly conserved (most are invariant) among all lepidopteran baculovirus F proteins and the cellular F proteins identified in this study (Fig. 1, asterisks, and 2D). These residues include a subset (six) of the previously identified conserved cysteine residues and six additional residues. The overall trend suggests that there is greater divergence in the N-terminal portion of F proteins and greater conservation near the central and C-terminal portion of the protein (Fig. 2D). Because a better understanding of host F proteins may provide valuable insights into mechanisms of viral evolution, we selected the *Dm-cF* gene and protein for further analysis. Next, we asked whether Dm-cF was expressed in the animal or in cell lines of *D. melanogaster* and whether Dm-cF was processed, localized, or served similar functions as baculovirus F proteins.

Dm-cF expression is developmentally regulated. To begin to characterize Dm-cF expression in insects, we examined the expression of the *Dm-cF* gene during the developmental stages of *D. melanogaster*. RNA was isolated from *D. melanogaster* embryos, first instar, second instar, third instar, white prepupae, pupae, adult males, and adult females. *D. melanogaster* RNA was used for RT-PCR analysis with primers specific to the Dm-cF ORF. Primers specific to the ribosomal protein 49 (*rp49*) gene were used as a positive control in parallel experiments. Dm-cF-specific RNA was detected only in RNAs isolated from the third instar and later developmental stages (Fig. 3A, top panels), while control *rp49* mRNAs were detected in mRNAs from all stages examined (Fig. 3A, lower panels). Dm-cF RNA was detected in adults of both sexes including both head and body (thorax and abdomen) fractions (Fig. 3A, top panel, lanes 10 to 13). No Dm-cF-specific RNA was detected in RNAs isolated from *D. melanogaster* cell lines S2 and Kc (Fig. 3A, top panel, lanes 5 and 6). Also, no Dm-cF-specific RNA was detected in parallel control experiments in which the reverse transcriptase was omitted from the RT reaction (Fig. 3A, middle panels, −RT), while the positive control RNA (*rp49*) was detected in all RT reactions (Fig. 3A, bottom panel, RT control). Thus, these results suggest that Dm-cF RNA expression is developmentally regulated in *D. melanogaster*.

Using Western blot analysis of *D. melanogaster* extracts with an affinity-purified anti-Dm-cF antibody, we identified a band of approximately 60 kDa that corresponds to the predicted molecular mass of Dm-cF in extracts of third instar larvae and adult males and females (Fig. 3B, lanes 3 to 5). The same ~60-kDa protein was not detected in extracts of first and second instar larvae (Fig. 3B, lanes 1 and 2). Consistent with studies of Dm-cF RNA expression (above), Western blots also indicate that Dm-cF protein is expressed in later stages of *D. melanogaster* development.

Dm-cF does not appear to be proteolytically processed in S2 cells. We used Western blots of epitope- and GFP-tagged Dm-cF proteins to examine possible proteolytic processing of Dm-cF. *D. melanogaster* S2 cells transiently expressing Dm-cF-V5 or Dm-cF-GFP fusion proteins (tagged at the C terminus of Dm-cF) were used for Western blot analysis with an affinity-purified anti-Dm-cF polyclonal antibody. Bands corresponding to the sizes predicted for the full-length Dm-cF-V5 (~65.3 kDa) and Dm-cF-GFP (~84 kDa) fusion proteins (Fig. 4, lanes 2 and 3) were detected, and truncated products were not identified. This result is consistent with the absence of a consensus furin cleavage site in the Dm-cF sequence. This is in

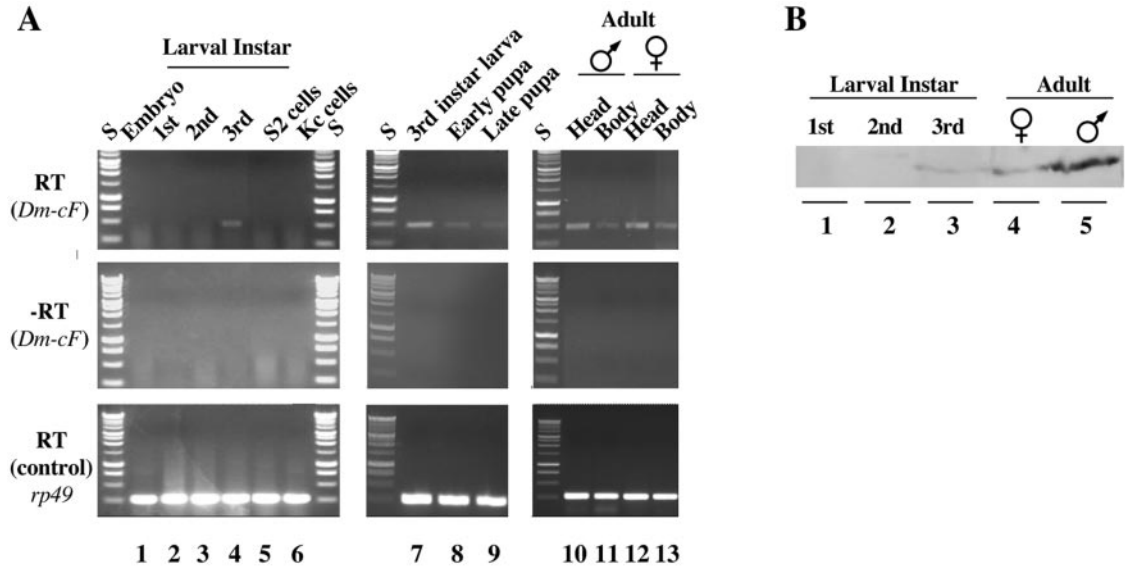


FIG. 3. Dm-cF expression is developmentally regulated. (A) Analysis of Dm-cF mRNA in *D. melanogaster*. RT-PCR analysis was performed in the presence of reverse transcriptase (top and bottom panels) or in the absence of reverse transcriptase (middle panel) with RNA isolated from *D. melanogaster* 4- to 16-h embryos (lane 1), first instar larvae (lane 2), second instar larvae (lane 3), third instar larvae (lanes 4 and 7), early pupae (lane 8), late pupae (lane 9), adult male head (lane 10), adult male body (lane 11), adult female head (lane 12), adult female body (lane 13), S2 cells (lane 5), and Kc cells (lane 6). PCR analysis was performed with primer pairs specific to either the *Dm-cF* gene (top two panels) or the ribosomal protein *rp49* gene (bottom panel). (B) Western blot analysis of Dm-cF in various developmental stages of *D. melanogaster* larvae and adults. An affinity purified rabbit anti-Dm-cF antibody was used to identify Dm-cF in proteins from first instar larvae (lane 1), second instar larvae (lane 2), third instar larvae (lane 3), adult female flies (lane 4), and adult male flies (lane 5).

contrast to F proteins from LdMNPV and SeMNPV, which contain a furin cleavage site and are proteolytically cleaved into F_1 and F_2 subfragments (22, 42, 49). Consistent with the lack of detectable *Dm-cF* mRNA in S2 cells by RT-PCR (Fig. 3A, lane 5), a band corresponding to the Dm-cF protein was not detected in mock-transfected S2 cells, suggesting that Dm-cF is not expressed endogenously at detectable levels in S2 cells (Fig. 4, lane 1).

Dm-cF transiently expressed in S2 cells localizes to intracellular organelles. To examine the subcellular localization of Dm-cF, we transiently expressed Dm-cF in S2 cells, either as a Dm-cF-GFP fusion protein or with a C-terminal V5 epitope tag (Dm-cF-V5). Both the Dm-cF-GFP and Dm-cF-V5 fusion proteins localized to intracellular organelles that are about 1 μm in diameter (Fig. 5). Frequently, Dm-cF fusion proteins were localized in a ring-like structure with an interior showing no apparent signal (Fig. 5C, arrow). This localization pattern suggests that Dm-cF may localize to the membrane of these intracellular organelles. To further examine the subcellular localization of Dm-cF in S2 cells, we performed colocalization experiments using a variety of markers for intracellular organelles. Dm-cF-GFP (Fig. 6) and Dm-cF-V5 (data not shown) did not colocalize with organelles labeled with antibodies directed against endoplasmic reticulum (ER) marker proteins (Fig. 6A to C, anti-KDEL and anti-bovine mannose-6-phosphate receptor) (data not shown), a Golgi marker (Fig. 6D to F, anti-*Drosophila* Golgi), or a lysosomal marker (Fig. 6G, I, and J, anti-bovine cathepsin D). However, Dm-cF-GFP partially colocalized with MitoTracker (Fig. 7), a marker for mitochondria.

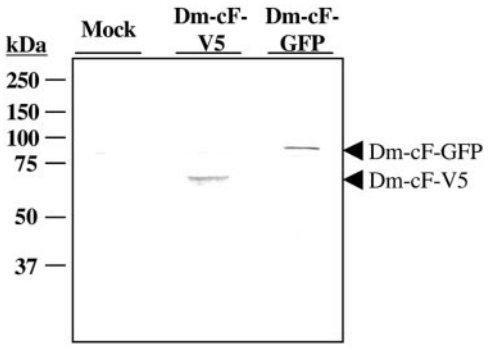


FIG. 4. Western blot analysis of Dm-cF fusion proteins. An affinity purified rabbit anti-Dm-cF antibody was used for Western blot analysis of lysates from *D. melanogaster* S2 cells that were either mock transfected (lane 1) or transfected with plasmids expressing Dm-cF-V5 (lane 2) or Dm-cF-GFP (lane 3). Marker protein sizes are indicated on the left.

Membrane fusion activity assays. Because many baculovirus F proteins are membrane fusion proteins, we asked whether Dm-cF might have a similar function or activity in *D. melanogaster* cells. Transient expression of Dm-cF-V5 or Dm-cF-GFP in S2 cells did not result in any obvious signs of membrane fusion activity between cells or between intracellular compartments. Some baculovirus F proteins fuse membranes in response to low pH. Also, since some intracellular membrane fusion events (such as neurotransmitter release) are initiated by calcium flux, we examined cells expressing the above Dm-cF fusions in the presence of either low pH or ionophores to increase intracellular Ca^{++} levels. Cells were incubated in

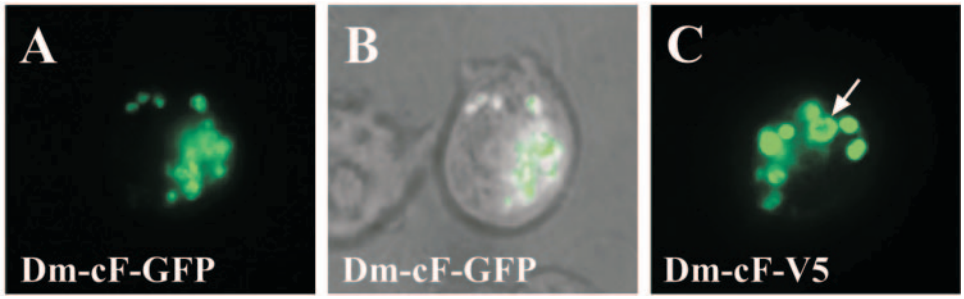


FIG. 5. Transiently expressed Dm-cF localizes to intracellular compartments of *D. melanogaster* S2 cells. (A) GFP fluorescence of Dm-cF-GFP in transiently transfected S2 cells. (B) The image from panel A overlaid with transmitted light. (C) Immunofluorescence detection of Dm-cF-V5 transiently expressed in S2 cells and detected with an antibody directed against the V5 epitope.

either PBS at pH 5.0 for 30 min or in 4Br-23187 or ionomycin (5 μM) for 30 min (see Materials and Methods). Neither treatment resulted in a detectable increase in the sizes of Dm-cF-containing intracellular vesicles or organelles, as judged by microscopic observations (data not shown). In control experiments, Sf9 cells expressing the AcMNPV GP64 protein showed robust syncytium formation after similar low pH treatment (data not shown). Thus, under these experimental conditions, we detected no membrane fusion activity by Dm-cF. However, we cannot rule out the possibility that Dm-cF-mediated fusion activity may be subtle or was not induced or detectable by these methods.

DISCUSSION

Phylogenetic analyses of large DNA viruses such as pox viruses, herpesviruses, and baculoviruses support the hypothesis that certain viral genes were acquired from the host. Examples include cytokines and cytokine receptors in the pox and herpesviruses (20, 26) and the inhibitor of apoptosis (*iap*) and ecdysteroid UDP-glucosyltransferase (*egt*) genes in the baculoviruses (20). The results of the current studies suggest that baculovirus virion envelope proteins may also be acquired from the host. The majority of baculoviruses that have been characterized produce two different virion phenotypes: occlu-

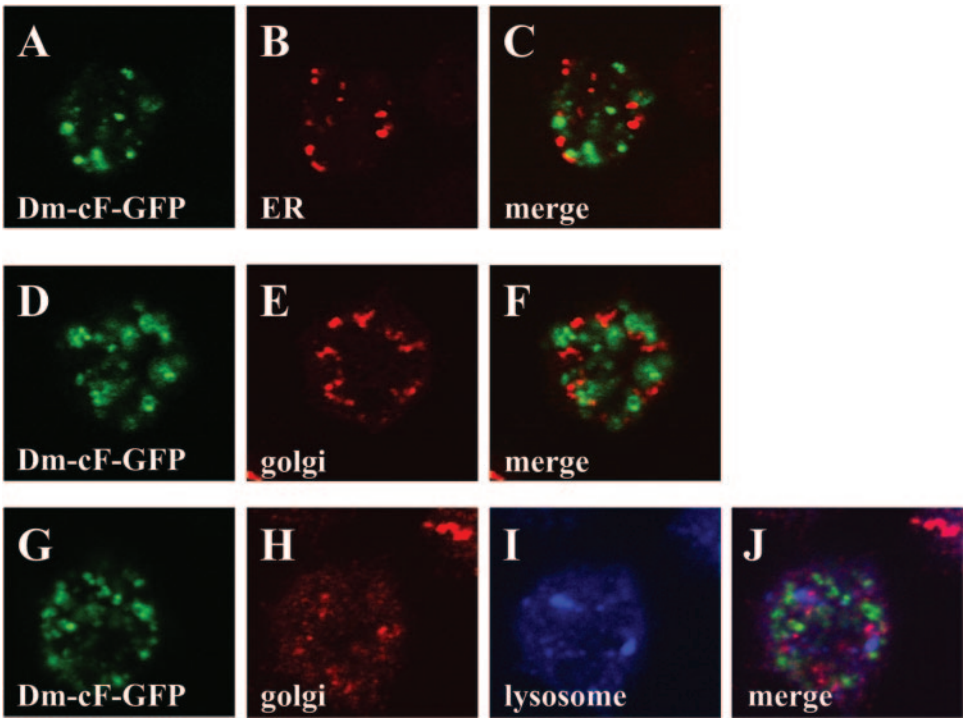


FIG. 6. Colocalization studies of Dm-cF-GFP and marker proteins for ER, Golgi, and lysosomes. Images represent Dm-cF-GFP fluorescence in *D. melanogaster* S2 cells transiently transfected with plasmids expressing Dm-cF-GFP (A, D and G). Panels B, E, and H show immunofluorescence with antibodies specific for an ER protein (B; KDEL receptor), a Golgi protein (E and H), and a lysosomal protein (I; cathepsin D). Merged images of Dm-cF-GFP and immunofluorescence of organelle-specific antibodies are shown on the right of each horizontal series C, F, and J).

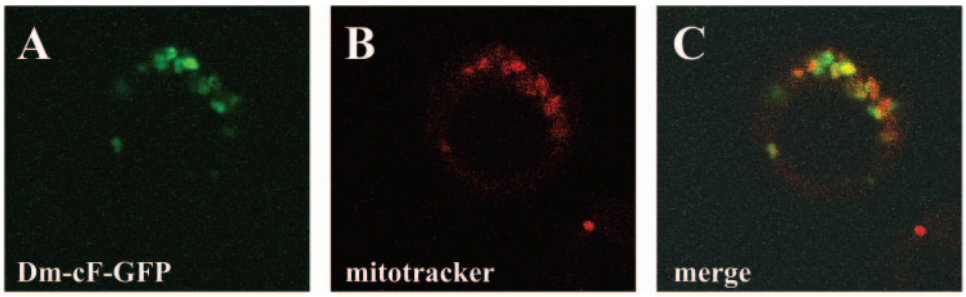


FIG. 7. Colocalization studies of Dm-cF-GFP and mitochondria. Dm-cF-GFP fluorescence in *D. melanogaster* S2 cells transiently transfected with a plasmid expressing Dm-cF-GFP is shown in panel A, and fluorescence of cells treated with MitoTracker Red is shown in panel B. Panel C shows a merged image with yellow representing the colocalization of Dm-cF-GFP and MitoTracker.

sion derived virions and budded virions (BV). Of the two different types of envelope fusion proteins (GP64 and F) that have been identified from BV of various baculoviruses, GP64 has been reported only in the closely related baculoviruses belonging to the group I NPVs, and GP64 is very highly conserved in those viruses. Thus, it has been suggested that the *gp64* gene may represent a more recent acquisition in baculoviruses (39). Although the *gp64* gene has a homolog in arthropod (tick)-borne orthomyxoviruses (Thogoto viruses) (35), a *gp64* homolog has not been reported in any host genome so far. In contrast to *gp64*, *f* genes are much more widely distributed in the baculoviruses. *f* genes have been reported in the genomes of all sequenced lepidopteran baculoviruses and in the genome of the recently sequenced dipteran baculovirus, CuniNPV. In the current study, we have identified and analyzed *f* genes in the genomes of three *Drosophila* species and the mosquito *A. gambiae*. The *Drosophila* and *Anopheles f* genes examined here do not appear to be derived from endogenous retroviruses (*f* genes associated with endogenous retroviruses have been examined elsewhere [43, 46]). Although these four dipteran species are not thought to be permissive hosts for baculoviruses, baculoviruses are known to infect dipterans including several species of mosquitoes (4, 5, 10–12, 36, 45). Therefore, it is likely that *f* genes are present in insects that serve as permissive hosts for baculovirus infection. Cellular *f* genes have not yet been reported in lepidopteran insects. However, we have identified several *f* gene homologs in the recently sequenced *B. mori* genome (33) although all *B. mori f* genes identified to date appear to be associated with endogenous retroviruses (data not shown). Because the *B. mori* genome sequence was recently released and cellular *f* genes may be poorly conserved, the question of whether a cellular *f* gene is found in the *B. mori* genome remains open. However, the presence of *f* genes in insect genomes in general suggests the intriguing possibility that baculoviruses may have obtained the *f* gene directly or indirectly from an insect host. The above observations also raise interesting questions regarding the possible function(s) of cellular F proteins in insects. To examine some of these questions, we performed comparative sequence analysis with baculovirus and insect F proteins, and analyzed Dm-cF expression, intracellular localization and potential for membrane fusion activity in *D. melanogaster*.

The current studies suggest that Dm-cF expression in *D. melanogaster* is developmentally regulated. *Dm-cF* mRNA and protein were not detected in embryos, first or second instar larvae, but they were detected in third instar larvae and in adults. Because some viral F proteins go through a cleavage-maturation step that produces two distinct subfragments and activates the fusion function (21, 42, 49), we also examined Dm-cF for similar proteolytic cleavage. When Dm-cF was transiently expressed in S2 cells, a protein of the size predicted for full-length Dm-cF was identified, and no discernible cleavage was detected. The results are consistent with the absence of a predicted furin cleavage site in the Dm-cF sequence. We also examined intracellular localization of the Dm-cF protein by both immunocytochemistry of an epitope-tagged Dm-cF protein and immunofluorescence of a Dm-cF-GFP fusion protein. Unlike baculovirus F proteins, Dm-cF localized to intracellular organelles. Dm-cF did not colocalize with ER, Golgi, or lysosomal marker proteins but partially colocalized with MitoTracker, a mitochondrial marker. The ring-like localization that was frequently observed suggests that Dm-cF may localize to outer membranes of these intracellular organelles. The sequence similarity between Dm-cF and baculovirus F proteins suggested that Dm-cF may be a cellular membrane fusion protein. However, transient expression of Dm-cF in S2 cells and treatment with either low pH or ionophores did not result in an obvious or detectable membrane fusion activity. Membrane fusion activity also has not been detected from transiently expressed F proteins from group I NPVs such as AcMNPV (Ac23) and OpMNPV (Op21) (reference 41 and unpublished results). Dm-cF, Ac23 and Op21 all lack the highly conserved region that contains a furin cleavage site and the fusion peptide. Thus, the absence of detectable membrane fusion activity in Dm-cF, Ac23, and Op21 is consistent with previous observations that an intact furin cleavage site and fusion peptide are both important for pH-induced membrane fusion activity in the SeMNPV and LdMNPV F proteins (42, 48, 49). We cannot however rule out the possibility that Dm-cF is an intracellular fusion protein with fusion activity regulated by mechanisms other than those tested in our assays.

The evolution of baculovirus envelope fusion proteins within the *Baculoviridae* is interesting in that two different envelope fusion proteins (GP64 and F) are found in BV of the *Baculoviridae*, and both *gp64* and *f* genes appear to have been ac-

quired and/or lost during the evolution of the different subgroups of baculoviruses. An intriguing possible explanation for the presence or absence of *f* and/or *gp64* in various subgroups within the family is the following: ancestral baculoviruses may have replicated only in the epithelial cells of the insect midgut, producing only occlusion derived virions and occlusion bodies as progeny. The ability to invade the insect hemocoel may have evolved through the acquisition of a gene such as *f* that promoted or mediated budding from the cell surface to generate a BV phenotype. One subgroup of baculoviruses may have later acquired a second envelope protein gene (*gp64*), and GP64 may have eventually supplanted some or many of the functions of the F protein in that lineage of baculoviruses. Thus, according to such a theory, viruses descended more directly from ancestral midgut-limited baculoviruses might be represented by the sawfly baculoviruses (such as *Neodiprion lecontei* NPV and *Neodiprion sertifer* NPV) which are midgut-limited and contain neither *gp64* nor *f* in their genomes. A relatively large group of the baculoviruses (group II NPVs, GVs, and the mosquito baculovirus CuniNPV) all contain an *f* gene but no *gp64* gene. Many or most of these viruses are believed to invade the hemocoel during infection. Exceptions are viruses such as CuniNPV and *Harrisina brillians* GV, which are midgut-limited (13, 36). A third major group, the group I NPVs, encodes both *gp64* and *f* genes. Group I NPVs appear to be much more closely related to each other than are members of the group II NPV or GV subgroups, suggesting that the divergence of the group I NPV subgroup was a more recent event in the evolution of the *Baculoviridae*. Thus, as suggested earlier (40), the group I NPVs may represent descendants of a virus that acquired the *gp64* gene more recently. Consistent with these speculations are prior studies showing that the *f* gene homolog (*Ac23*) from the group I NPV, AcMNPV, is not essential (29), whereas the *f* gene from the group II NPV, SeMNPV, is essential (47) and serves a role similar to that of *gp64* (28).

The identification of host insect genes encoding proteins with similarity to baculovirus F proteins in four insect genomes suggests that *f* genes may be present in many insects and further supports the possibility that the baculovirus *f* gene may have been acquired directly or indirectly from a cellular gene in an insect host. The abundance of insect endogenous retroviruses encoding *f*-like genes (30, 32, 43, 46) combined with the observation that endogenous retroviruses and transposable elements can insert into the baculovirus genome (14, 15, 17, 23, 31) also suggests a possible indirect route for baculovirus acquisition of host insect *f* genes. Although it was proposed previously that errantiviruses may have acquired the *env* gene from baculoviruses (30), the direction of horizontal gene transfer remains uncertain. Also intriguing is a possible link to the paramyxoviruses, since baculovirus F proteins are very similar in structure to F proteins from paramyxoviruses (49). To date baculovirus *f* orthologs have been found in endogenous retroviruses from dipteran and lepidopteran insects and in the genomes of *Drosophila* and *Anopheles* species. Further studies of the distributions and function of cellular F proteins will be important for understanding how host cellular proteins may be acquired and adapted to viral infection cycles.

ACKNOWLEDGMENTS

The authors thank John Ewer for providing *D. melanogaster* and for valuable suggestions. We also thank Guey-Jen Lai for assistance with RT-PCR, Mike Goldberg for providing plasmid pGW-MAL, and John Lis for providing vector pMET-GFP. We are grateful to Chip Aquadro for providing *D. pseudoobscura* and *D. yakuba* genomic DNA and Brian Lazarro for providing *A. gambiae* genomic DNA. We also thank Bill Brown for providing anti-M6PR and anti-cathepsin-D antibodies.

This work was supported by NIH grant AI33657 and BTI project 1255.

REFERENCES

1. **Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, et al.** 2000. The genome sequence of *Drosophila melanogaster*. Science **287:**2185–2195.
2. **Afonso, C. L., E. R. Tulman, Z. Lu, C. A. Balinsky, B. A. Moser, J. J. Becnel, D. L. Rock, and G. F. Kutish.** 2001. Genome sequence of a baculovirus pathogenic for *Culex nigripalpus*. J. Virol. **75:**11157–11165.
3. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
4. **Becnel, J., S. White, B. Moser, T. Fukuda, M. Rotstein, A. Undeen, and A. Cockburn.** 2001. Epizootiology and transmission of a newly discovered baculovirus from the mosquitoes *Culex nigripalpus* and *C. quinquefasciatus*. J. Gen. Virol. **82:**275–282.
5. **Becnel, J. J., S. E. White, and A. M. Shapiro.** 2003. *Culex nigripalpus* nucleopolyhedrovirus (CuniNPV) infections in adult mosquitoes and possible mechanisms for dispersal. J. Invertebr. Pathol. **83:**181–183.
6. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340:**783–795.
7. **Blissard, G., B. Black, N. Crook, B. A. Keddie, R. Possee, G. Rohrmann, D. Theilmann, and L. E. Volkman.** 2000. *Baculoviridae*: taxonomic structure and properties of the family, p. 195–202. *In* M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle, and E. B. Wickner (ed.), Virus taxonomy: classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego, Calif.
8. **Brown, W. J.** 1990. Cation-independent mannose 6-phosphate receptors are concentrated in trans Golgi elements in normal human and I-cell disease fibroblasts. Eur. J. Cell Biol. **51:**201–210.
9. **Corpet, F.** 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16:**10881–10890.
10. **Federici, B. A.** 1980. Mosquito baculovirus: sequence of morphogenesis and ultrastructure of the virion. Virology **100:**1–9.
11. **Federici, B. A., and D. W. Anthony.** 1972. Formation of virion-occluding proteinic spindles in a baculovirus disease of *Aedes triseriatus*. J. Invertebr. Pathol. **20:**129–138.
12. **Federici, B. A., and R. E. Lowe.** 1972. Studies on the pathology of a baculovirus in *Aedes triseriatus*. J. Invertebr Pathol. **20:**14–21.
13. **Federici, B. A., and V. M. Stern.** 1990. Replication and occlusion of a granulosis virus in larval and adult midgut epithelium of the Western grapeleaf skeletonizer *Harrisina brillians*. J. Invertebr Pathol. **56:**401–414.
14. **Fraser, M. J., J. S. Brusca, G. E. Smith, and M. D. Summers.** 1985. Transposon-mediated mutagenesis of a baculovirus. Virology **145:**356–361.
15. **Fraser, M. J., G. E. Smith, and M. D. Summers.** 1983. Acquisition of host cell DNA sequences by baculoviruses: relationship between host DNA insertions and FP mutants of *Autographa californica* and *Galleria mellonella* nuclear polyhedrosis virus. J. Virol. **47:**287–300.
16. **Friesen, P. D., and L. K. Miller.** 2001. Insect Viruses, p. 599–628. In D. M. Knipe and P. Howley (ed.), Fields Virology, Forth edition ed.
17. **Friesen, P. D., and M. S. Nissen.** 1990. Gene organization and transcription of TED a lepidopteran retrotransposon integrated within the baculovirus genome. Mol. Cell Biol. **10:**3067–3077.
18. **Garcia-Maruniak, A., J. E. Maruniak, P. M. Zanotto, A. E. Doumbouya, J. C. Liu, T. M. Merritt, and J. S. Lanoie.** 2004. Sequence analysis of the genome of the *Neodiprion sertifer* nucleopolyhedrovirus. J. Virol. **78:**7036–7051.
19. **Holt, R. A., G. M. Subramanian, A. Halpern, G. G. Sutton, et al.** 2002. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science **298:**129–149.
20. **Hughes, A. L., and R. Friedman.** 2003. Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses. Mol. Biol. Evol. **20:**979–987.
21. **IJkel, W. F., E. C. Roode, R. W. Goldbach, J. M. Vlak, and D. Zuidema.** 2002. Characterization of *Spodoptera exigua* multicapsid nucleopolyhedrovirus ORF17/18, a homologue of *Xestia c-nigrum* granulovirus ORF129. J. Gen. Virol. **83:**2857–2867.
22. **IJkel, W. F., M. Westenberg, R. W. Goldbach, G. W. Blissard, J. M. Vlak, and D. Zuidema.** 2000. A novel baculovirus envelope fusion protein with a proprotein convertase cleavage site. Virology **275:**30–41.

23. **Jehle, J. A., A. Nickel, J. M. Vlak, and H. Backhaus.** 1998. Horizontal escape of the novel Tc1-like lepidopteran transposon TCp3.2 into *Cydia pomonella* granulovirus. J. Mol Evol. **46:**215–224.
24. **Kozak, M.** 1995. Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. Proc. Natl. Acad. Sci. USA **92:**2662–2666.
25. **Kozak, M.** 1987. At least six nucleotides preceding the *AUG* initiator codon enhance translation in mammalian cells. J. Mol. Biol. **196:**947–950.
26. **Lalani, A. S., and G. McFadden.** 1999. Evasion and exploitation of chemokines by viruses. Cytokine Growth Factor Rev. **10:**219–233.
27. **Lauzon, H. A. M., C. J. Lucarotti, P. J. Krell, Q. Feng, A. Retnakaran, and B. M. Arif.** 2004. Sequence and organization of the *Neodiprion lecontei* nucleopolyhedrovirus genome. J. Virol. **78:**7023–7035.
28. **Lung, O., M. Westenberg, J. M. Vlak, D. Zuidema, and G. W. Blissard.** 2002. Pseudotyping *Autographa californica* multicapsid nucleopolyhedrovirus (Ac*M*NPV): F proteins from group II NPVs are functionally analogous to Ac*M*NPV GP64. J. Virol. **76:**5729–5736.
29. **Lung, O. Y., M. Cruz-Alvarez, and G. W. Blissard.** 2003. Ac23, an envelope fusion protein homolog in the baculovirus *Autographa californica* multicapsid nucleopolyhedrovirus, is a viral pathogenicity factor. J. Virol. **77:**328–339.
30. **Malik, H. S., S. Henikoff, and T. H. Eickbush.** 2000. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res. **10:**1307–1318.
31. **Miller, D. W., and L. K. Miller.** 1982. A virus mutant with an insertion of a copia-like transposable element. Nature (London) **299:**562–564.
32. **Misseri, Y., M. Cerutti, G. Devauchelle, A. Bucheton, and C. Terzian.** 2004. Analysis of the *Drosophila* gypsy endogenous retrovirus envelope glycoprotein. J. Gen. Virol. **85:**3325–3331.
33. **Mita, K., M. Kasahara, S. Sasaki, Y. Nagayasu, et al.** 2004. The genome sequence of silkworm, *Bombyx mori*. DNA Res. **11:**27–35.
34. **Monsma, S. A., A. G. P. Oomens, and G. W. Blissard.** 1996. The GP64 envelope fusion protein is an essential baculovirus protein required for cell-to-cell transmission of infection. J. Virol. **70:**4607–4616.
35. **Morse, M. A., A. C. Marriott, and P. A. Nuttall.** 1992. The glycoprotein of Thogoto virus (a tick-borne orthomyxo-like virus) is related to the baculovirus glycoprotein gp64. Virology **186:**640–646.
36. **Moser, B., J. Becnel, S. White, C. Afonso, G. Kutish, S. Shanker, and E. Almira.** 2001. Morphological and molecular evidence that *Culex nigripalpus* baculovirus is an unusual member of the family *Baculoviridae*. J. Gen. Virol. **82:**283–297.
37. **Oomens, A. G. P., and G. W. Blissard.** 1999. Requirement for GP64 to drive efficient budding of *Autographa californica* multicapsid nucleopolyhedrovirus. Virology **254:**297–314.
38. **Park, J. E., R. K. Draper, and W. J. Brown.** 1991. Biosynthesis of lysosomal enzymes in cells of the End3 complementation group conditionally defective in endosomal acidification. Somat. Cell Mol. Genet. **17:**137–150.
39. **Pearson, M. N., C. Groten, and G. F. Rohrmann.** 2000. Identification of the *Lymantria dispar* nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the *Baculoviridae*. J. Virol. **74:**6126–6131.
40. **Pearson, M. N., and G. F. Rohrmann.** 2002. Transfer, incorporation, and substitution of envelope fusion proteins among members of the *Baculoviridae*, *Orthomyxoviridae*, and *Metaviridae* (insect retrovirus) families. J. Virol. **76:**5301–5304.
41. **Pearson, M. N., R. L. Russell, and G. F. Rohrmann.** 2001. Characterization of a baculovirus-encoded protein that is associated with infected-cell membranes and budded virions. Virology **291:**22–31.
42. **Pearson, M. N., R. L. Russell, and G. F. Rohrmann.** 2002. Functional analysis of a conserved region of the baculovirus envelope fusion protein, LD130. Virology **304:**81–88.
43. **Rohrmann, G. F., and P. A. Karplus.** 2001. Relatedness of baculovirus and gypsy retrotransposon envelope proteins. BMC Evol. Biol. **1:**1.
44. **Singh, M., B. Berger, and P. S. Kim.** 1999. LearnCoil-VMF: computational evidence for coiled-coil-like motifs in many viral membrane-fusion proteins. J. Mol. Biol. **290:**1031–1041.
45. **Stiles, B., P. E. Dunn, and J. D. Paschke.** 1983. Histopathology of a nuclear polyhedrosis infection in *Aedes epactius* with observations in four additional mosquito species. J. Invertebr. Pathol. **41:**191–202.
46. **Terzian, C., A. Pelisson, and A. Bucheton.** 2001. Evolution and phylogeny of insect endogenous retroviruses. BMC Evol. Biol. **1:**3.
47. **Westenberg, M.** 2004. Functional analysis of a novel baculovirus envelope fusion protein. Ph.D. thesis. Wageningen University, Wageningen, The Netherlands.
48. **Westenberg, M., F. Veenman, E. C. Roode, R. W. Goldbach, J. M. Vlak, and D. Zuidema.** 2004. Functional analysis of the putative fusion domain of the baculovirus envelope fusion protein F. J. Virol. **78:**6946–6954.
49. **Westenberg, M., H. Wang, W. F. IJkel, R. W. Goldbach, J. M. Vlak, and D. Zuidema.** 2002. Furin is involved in baculovirus envelope fusion protein activation. J. Virol. **76:**178–184.