Baculovirus envelope fusion protein F structure and function Qiushi Wang 2015

Structural and functional analysis of the baculovirus envelope fusion protein F

Qiushi Wang

Structural and functional analysis of the baculovirus envelope fusion protein F

Qiushi Wang

Thesis committee

Promotors

Prof. Dr J.M. Vlak
Personal Chair at the Laboratory of Virology
Wageningen University

Prof. Dr P.J.M. Rottier
Professor of Veterinary Virology
Virology Division, Department of Infectious Diseases and Immunology
Faculty of Veterinary Medicine, Utrecht University

Prof. Dr M.M. van Oers
Professor of Virology
Wageningen University

Co-promotor

Dr.ir. B.J. Bosch
Assistant Professor
Virology Division, Department of Infectious Diseases and Immunology
Faculty of Veterinary Medicine, Utrecht University

Other members

Prof. Dr G.F. Wiegertjes, Wageningen University
Dr G.W. Blissard, Boyce Thompson Institute at Cornell University
Prof. Dr J.M. Smit, Groningen University
Dr J.A. Kortekaas, Central Veterinary Institute, Lelystad

This research was conducted under the auspices of the Graduate School of Production Ecology and Resource Conservation.

Structural and functional analysis of the baculovirus envelope fusion protein F

Qiushi Wang

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 18th November 18 2015
at 11.00 a.m. in the Aula.

Qiushi Wang
Structural and functional analysis of the baculovirus envelope fusion protein F
242 pages

PhD thesis, Wageningen University, Wageningen, NL (2015)
With references, summaries in English, Dutch and Chinese

ISBN 978-946257-563-9

To Paul and my parents

Cover art is from the painting album "Exploded view" (2014). Bruce Riley, a contemporary Chicago artist (www.bruce-riley.com) has permitted its usage in this thesis.

Contents

Chapter 1

General introduction

Qiushi Wang

Viruses are small infectious particles that only replicate within host cells. The viruses must enter the cells and deliver their genomes to the replication sites in the target cells. Enveloped viruses release their genomes into the host cells through fusion between the viral envelope and a target cellular membrane. A membrane fusion event is an energy-dependent reaction and is essential for living organisms for many physiological and intracellular activities. In case of viral infection, fusion is mediated by specialized viral glycoproteins called envelope fusion proteins (EFPs). The EFPs anchored in the viral envelope facilitate membrane fusion through interaction with the target membrane and undergo conformational rearrangement to bring the viral envelope into contact with the target membrane. EFPs can be divided into three classes on the basis of their structural characteristics. Viruses within the family *Baculoviridae* possess F or GP64 fusion proteins that belonging to distinct classes of EFPs. Much is known about the structural and functional properties of GP64, but much less about baculovirus F proteins. This thesis focuses on the baculovirus F protein. In this chapter an overview of membrane fusion mechanisms between different types of biological membranes is given as well as an introduction to the baculovirus infection process. At the end of this chapter, the research described in this thesis will be introduced.

MEMBRANE FUSION

Biological membrane fusion

Membrane fusion is an energy-dependent process where two opposing lipid bilayers merge to become one interconnected structure (153) (Fig. 1). It plays an essential role in many biological processes including intracellular vesicle trafficking, neurotransmission, fertilization, organ and tissue development, carcinogenesis, and cell entry of enveloped viruses (118, 326, 358). Most membrane fusion events are mediated by specialized proteins, which are commonly called fusion proteins. Over the past 30 years, efforts have been made to model the process of lipid bilayer fusion on the basis of structures of fusion proteins and physical and chemical properties of lipid leaflets. At the current stage we are starting to understand the common features shared by all the protein-mediated biological membrane fusion events with instrumental viral and intracellular vesicle fusion models.

It is generally believed that the fusion proteins interact with lipid leaflets or proteins in the opposing membrane (Fig. 1, ii) and disturb the lipid bilayer to form an extreme curvature through protein conformational changes (Fig. 1, iii). Membrane fusion is believed to proceed through a hemifusion intermediate after the membranes are brought into close proximity. At this intermediate stage of membrane fusion, only outer monolayers of the opposing membranes are merged (Fig. 1, iv). Before the full fusion stage the inner monolayers are possibly brought into contact to form the hemifusion diaphragm (180, 378) (Fig. 1, v). Further fusion protein folding results into two new leaflets, each of which contains a mixed lipid composition derived from two initial membranes (Fig. 1, vi).

In the membrane fusion process several steps demand great energy consumption. The first one is to bring the two distant membranes separated by a small water layer into close apposition, which requires to overcome repulsive membrane charges (281). This hydration repulsion is subsequently minimized by protrusion of one or both of the opposing membranes. The redistribution of lipids in the outer leaflets of both membranes to establish the hemifusion diaphragm forms another major energy barrier. Before fusion activation (pre-fusion) the fusion proteins are metastable. Upon fusion activation the fusion proteins undergo conformational rearrangements leading to a stable post-fusion form. The amount of energy released during this process is equivalent to the difference in free energy between these states, and helps to overcome the above mentioned energy barriers (240).

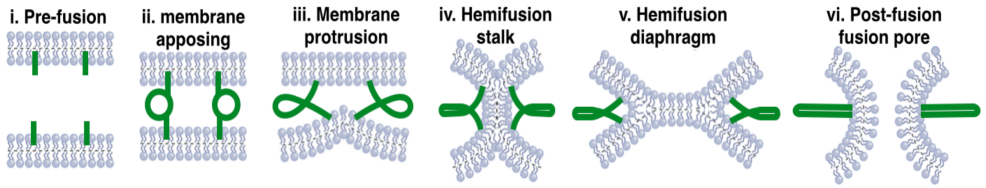


Fig. 1. Hemifusion pathway of lipid bilayer fusion [adapted from (58)]. (i) Pre-fusion. The fusion proteins (green) located at one or both membranes are in a metastable state. (ii) Opposing membranes brought in close proximity. Upon activation the fusion proteins interact with either the lipid layer directly or with proteins on the opposing membrane and undergo conformational rearrangements. (iii) Membrane protrusion to reduce the hydration repulsion. (iv) Hemifusion stalk formation. The high energy level resulting from membrane curvature and hydration repulsion is further reduced when the proximal leaflets are fused and distal leaflets remain separated (v) Stalk expansion into hemifusion diaphragm. The energy is released from the fusion proteins to provide more extreme curvature of membranes. (vi) Post-fusion pore formed either in the hemifusion diaphragm bilayer or directly after hemifusion stalk formation. The fusion proteins are in a stable post-fusion state and the energy of the membranes has dropped to the lowest level as the two newly formed membranes separate in space alleviating repulsion forces.

Cellular membrane fusion through SNAREs

Many forms of intracellular cargo transport and vesicle biogenesis require membrane fusion. In eukaryotic cells early endosomes fuse homotypically with each other, as well as with late endosomes or lysosomes to recycle or degrade their contents. In yeast the functional equivalent of the mammalian lysosome is the vacuole, which undergoes homotypic fusion with other vacuoles to maintain low copy numbers of this organelle and enrich vacuolar proteins (204, 417). Another example of vesicle fusion is the fusion of synaptic vesicles with the axon plasma membrane, which allows the release of neurotransmitters enabling intercellular communication (359). The common mechanism of these fusion events can be described as follows: Both of the membranes to be fused possess a number of membrane-anchored SNARE proteins (Soluble N-ethylmaleimide-sensitive factor attachment protein receptor). The opposing membranes are brought into proximity and connect through *trans* interaction of the SNAREs resident in different membranes (Fig. 2A). This *trans* interaction progresses like zipping up a zipper until the SNAREs are arranged into a four-helix bundle, resulting in complete fusion and new membrane formation. At this state the completely zipped complex consisting of the SNARE proteins originated from different membranes translocate to the same newly formed membrane and are therefore named as '*cis* SNARE complex' (Fig. 2B) [reviewed in (154)].

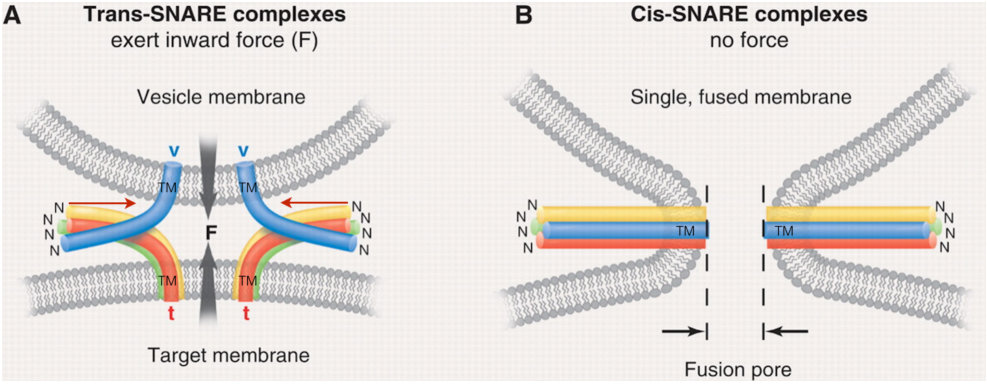


Fig. 2. Transition from *trans* to *cis* SNARE complexes leading to membrane fusion [adapted from (360)]. (A) Upon membrane apposing, three helices anchored in the target membrane (target SNAREs in green, red and yellow) assemble with the fourth helix anchored in the vesicle membrane (vesicle SNARE in blue) to form a *trans* complex. The assembly starts at the N-termini and progresses towards the C-termini of the four SNAREs (red arrows), generating a pulling force (F) to merge the two membranes. (B) At the final state the SNAREs form four-helix bundle in the *cis* configuration, resulting in fusion pore formation.

Table 1. Classification of viral envelope fusion proteins.

	Class I	Class II	Class III
Examples	*Influenza A virus* HA [45, 124, 418] HIV-1 Env [159, 186, 208, 228, 233] PIV5[405]/RSV[247] F *Ebola virus* GP [195] SeMNPV F [Chapter 3]	*Dengue virus* E [264, 265, 433] SFV E1 [109, 199, 202] RVFV G_c [77] *Rubella virus* E1 [83]	VSV G [311, 313] HSV-1 gB [132] AcMNPV GP64 [161] Thogoto virus [105]
Proteolytic cleavage required for fusogenicity priming	Yes	Yes	No
Fusogenicity activation	Low pH, receptor binding or receptor binding followed by low pH	Low pH	VSV G & AcMNPV GP64: low pH; HSV-1 gB: receptor binding or low pH
Major secondary structure of fusogenic form	α-helix	β-sheet	α-helix and β-sheet
Oligomeric state of native form	Trimer	Dimer	VSV G: trimer; HSV-1 gB/AcMNPV GP64: not known
Oligomeric state of fusogenic form	Trimer	Trimer	Trimer
Oligomeric state of post-fusion form	Trimer-of-hairpins (central α- helical coiled-coil, 6HB)	Trimer-of-hairpins (mainly β-structure)	Trimer-of-hairpins (central α-helical coiled- coil surrounded β-structure)
Location of fusion peptide of fusogenic form	Propelled by α-helical coiled coils	Propelled by β-strands	VSV G/AcMNPV GP64: propelled β-strands; HSV-1 gB: not known
Orientation of native form to viral membrane	Perpendicular (project as a spike)	Parallel (close to viral membrane)	VSV G: perpendicular; HSV-1 gB/AcMNPV GP64: not known

Abbreviations of species and families belonged mentioned in the table: *Influenza A virus* (*Orthomyxoviridae*) HIV-1, *human immunodeficiency virus type 1* (*Retroviridae*); PIV5, *parainfluenza virus 5* (*Paramyxoviridae*); RSV: *respiratory syncytial virus* (*Paramyxoviridae*); *Ebola virus* (*Filoviridae*); SeMNPV, *Spodoptera exigua* multiple nucleopolyhedrovirus (*Baculoviridae*); *Dengue virus* (*Flavivirus*); SFV: *Semliki Forest virus* (*Togaviridae*); RVFV: *Rift Valley fever virus* (*Bunyaviridae*); *Rubella virus* (*Togaviridae*); VSV: *vesicular stomatitis virus* (*Rhabdovirus*); HSV-1: *herpes simplex virus type 1* (herpes virus); AcMNPV: *Autographa californica* multiple nucleopolyhedrovirus (*Baculoviridae*)

Viral fusion proteins

Enveloped viruses deliver their genomes into the target cells by fusion of viral and cellular membranes. Viruses may fuse with the endosomal membrane after the virions enter host cells via endocytosis or may directly fuse at the plasma membrane [reviewed in (119)]. In general, one viral envelope protein is directly responsible for membrane fusion, which is called envelope fusion protein (EFP), while for some viruses such as parainfluenza virus 5 and herpes simplex virus type 1, a second envelope protein is necessary for triggering the fusion process in the latter case through (co-) receptor binding (299, 388). The mechanism used by EFPs to achieve virus-cell fusion is similar to what is seen in other biological membrane fusion events. The EFPs have membrane-anchoring domains for both virus envelope and target membrane, which are called a transmembrane domain (TMD) and a fusion peptide (FP), respectively. Based on their structural organization and fusion mechanism the EFPs can be divided into three classes (415) (Table 1).

All class I EFPs are synthesized as fusion-inactive precursors, which require proteolytic cleavage to liberate their FP proximal to the cleavage site and generate a metastable 'fusogenic form' (415). This step to gain fusion competence is also known as 'fusogenicity priming'. Most class I EFPs such as influenza virus hemagglutinin (HA) (355), paramyxovirus F (114), human immunodeficiency virus (HIV) GP160 (122), Ebola virus GP (389) and some coronavirus S proteins (22) are processed by the endoprotease furin after synthesis and routing through the secretory system of host cells prior to assembly into virions. The metastable EFP anchors on the surface of the viral envelope via the TMD as a lollipop-like structure consisting of a TMD-proximal stalk domain and a large globular head domain [Fig. 3A (a)]. The conformational rearrangement of class I proteins leading to membrane fusion is initiated by acidic pH, receptor binding or a combination of both. This step is also known as 'fusion activation'. There are at least two major heptad repeat regions (HRs), of which one is adjacent to the FP and the other is adjacent to the TMD. Upon fusion activation the FP-proximal HR region extends from the globular domain as a trimeric coiled-coil structure and inserts the FP into the target membrane. The protein forms pre-hairpin structure [Fig. 3A (b)]. The TMD-proximal HR region, as a second trimeric coiled-coil structure, folds into the grooves of the FP-proximal coiled-coil in an antiparallel orientation, thus forming a six-helix bundle (6HB) [Fig. 3A (c)]. This conformational change brings the viral membrane into contact with the target cellular membrane and allows fusion (14, 86, 338).

Among all the identified class I EFPs influenza virus HA and paramyxovirus F proteins have been well characterized and both of their pre- and post-fusion structures are known. Typically, the structures of both influenza virus HA and paramyxovirus F proteins at pre- and post-fusion states are trimeric and can be divided into two structural parts: globular head and stem. Although these two types of proteins form the same structural hallmark of extended 6HB at post-fusion state, the position and interactions of their globular head to the stem during membrane fusion exhibit large distinction.

Influenza virus HA—Influenza virus HA is the fusion-competent form of its precursor HA0, which proteolytically cleaved into disulphide-linked HA1 (receptor binding) and HA2 (fusion) subunits. HA1 subunit contains a receptor-binding domain with a globular head conformation connecting to a stem domain that is mainly composed by fusion subunit HA2. The crystal structures of H3N2 (X:31) HA2 subunit of ectodomain at pre- and post-fusion states are shown in Fig. 4A. The HA1 subunit is not visible in the post-fusion structure (45) therefore, is not shown in both pre- and post-fusion structures for comparison. The pre-fusion form (Fig. 6A i) is a trimer including a globular head (above the central blue helices, not shown), a central coiled coil composed with three long helices and three shorter outer helices. The central helix and outer helix are connected by a flexible loop, also referred as 'B loop' (orange). The N-terminal FP (Fig. 6A i and ii, red) linked to the outer shorter helix by two β-strands (blue) is sequestered in the trimer interface. The C-terminal TMD (purple triangles) anchoring in the viral membrane is extended from the central helix by two β-strands (purple). Upon fusion triggering by activation, the HA1 (not shown) dissociates with HA2 (79) without changing its overall structure (338), followed by two major structural rearrangements (Fig. 6A iii and iv), FP extension into target membrane and TMD inversion (FP: red triangles, TMD: purple triangles). FP is extended through conversion of B loop to a α-helix (orange), which loads the short helix (blue) and FP on the tip of the central helix. TMD is inverted through conversion of a α-helix into a short flexible loop (green), which flips the two helices (purple), a long leash-like loop (purple) and TMD into antiparallel orientation to the central helix. The central helices (Fig. 4A iv, blue) are decorated by the small outer helices (Fig. 4A, purple) at membrane distal end. These structural rearrangements result interchange of elements in central and outer helices and formation of hairpin structure to allow fusion between viral envelope and target membrane.

Paramyxovirus F—Paramyxovirus F proteins are the only type of class I EFPs activated by receptor binding through an individual envelope protein, HN (189). Like influenza virus HA, paramyxovirus F is fusogenic after proteolytic cleavage of precursor protein, F0. This cleavage separates F0 into F2 and F1, which are linked by a disulfide bond (188). The crystal structures of PIV5 F (pre-fusion) and hPIV3 F (post-fusion) are shown in Fig. 4B. The trimeric pre-fusion structure (Fig. 4B i) is composed of a large globular head and a small stem of coiled coil. FP (Fig. 4B i and ii, red) connecting to N-helix (blue) via a helical linker (also known as HRA, orange) is embedded between domain III (cyan) and domain II (yellow) from different protomers. TMD (Fig. 4B i and ii, purple triangles) is extended from C-helix (also known as HRB, purple), which is linked to domains II and I (yellow) via a helical linker (green). Upon fusion activation by receptor binding, paramyxovirus F protein undergoes conformational changes, in which the major ones, FP extension and TMD inversion result a hairpin structure similar to influenza virus HA protein. The structural changes of peptides that facilitate the major conformational changes in paramyxovirus F are different from influenza virus HA. First, evidence suggests C-helices in the trimer (purple) dissociates from each other (324). Next, the linker (Fig. 4B, orange) between N-helix (blue) and FP (red) converts from small helices and a β-sheet into a large helix, extending FP into the target membrane. This intermediate structure forms a pre-hairpin appearance with two membrane-anchoring domains FP and TMD pointing at opposite directions [Fig. 3A (b)]. Finally, a small helix (green) between domain II (yellow) and C-helix (purple) at pre-fusion state (Fig. 4B i and ii) inverts, unwinds and becomes an extended loop to bring the C-helix and TMD next to N-helix and FP in an antiparallel orientation at post-fusion state (Fig. 4B iii and iv). Unlike influenza virus HA (Fig. 4A iv), the 6HB of paramyxovirus F at post-fusion state is located proximally to the fused membrane (Fig. 4B iv).

Class II EFPs have mainly been identified in arboviruses, including members of the genera *Flavivirus* (*Flaviviridae*), *Alphavirus* (*Togaviridae*) and *Phlebovirus* (*Bunyaviridae*) but recently also identified in the non-arbovirus genus *Rubivirus* (*Togaviridae*). The pre- and postfusion structures have been best characterized in *Flavivirus* E and *Alphavirus* E1 proteins (202). All class II EFPs form heterodimers with a cognate viral glycoprotein. The EFPs require fusogenicity priming by a proteolytic cleavage to remove the second glycoprotein (77, 83, 203, 272, 283). In case of the *Flavivirus* E protein, the cleavage requires the low pH in the trans-Golgi network to first trigger a conformational rearrangement to allow

the cleavage site accessible to proteolytic cleavage (185). The cleaved prefusion form of class II EFPs appears to be a β-sheet-enriched, head-to-tail homodimer, parallel to the viral envelope [Fig. 3B (a)]. Each protomer of the EFP homodimer is organized into three structural domains: a β-sandwich domain proximal to the TMD (DI), an elongated domain with a fusion loop located at the tip of the domain (DII) and an immunoglobulin-like domain (DIII). The fusion loop is buried within the interface of the subunits of the homodimer. Upon fusion activation through acidification in the endosome, the homodimers transiently dissociate and reorganize into homotrimers. Domain DII rotates around the hinge between domain DI and DII to a position perpendicular to the viral envelope to propel the fusion loop into the target membrane [Fig. 3B (b)]. Domain DII rotates back to DI around the hinge to form a hairpin structure and to merge the viral envelope with the target membrane [Fig. 3B (c)] (202, 354). The β-sheet dominated postfusion form is a characteristic feature of class II EFPs, and is clearly distinguished from the class I EFP postfusion form, which contains a high content of α-helices.

Class III EFPs include rhabdovirus G, herpes virus gB and baculovirus GP64. So far the only protein from this class for which pre- and postfusion structures have been determined is the G protein of vesicular stomatitis virus (VSV). Low pH induces conformational changes that render the G protein to mediate membrane fusion. Distinct from EFPs in the other two classes the conformational change is reversible and there is a pH-dependent equilibrium of different conformations between pre- and postfusion state (108). During trafficking through the low-pH environment of the Golgi compartment the G protein is in transition towards the postfusion form, and thus in a fusion-inactive state (107). Once the G protein is transported to the plasma membrane the prefusion form is restored (106). Like class I EFPs, the G protein is trimeric in both pre- and postfusion forms. Each protomer contains four structural domains: a TMD-proximal domain (DI), a trimerization domain (DII), a pleckstrin homology domain (pfam accession number: PF15405) (DIII) and a fusion domain (DIV). The fusion loops are located at the tip of the β-strands of the fusion domain. The native G trimer on the viral surface appears to have a tripod-like organization with the fusion loops proximal to the viral envelope and centered by the TMD-proximal domain [Fig. 3C (a)]. Upon low pH activation the tripod legs flip upwards driving the fusion loops to the target membrane. The changes mainly exist of secondary structure rearrangements in the hinge region between DIV and DIII as well as in refolding

of the DII domain, resulting in the reorientation of the fusion domain DIV and eventually forming a hairpin-like postfusion conformation (6).

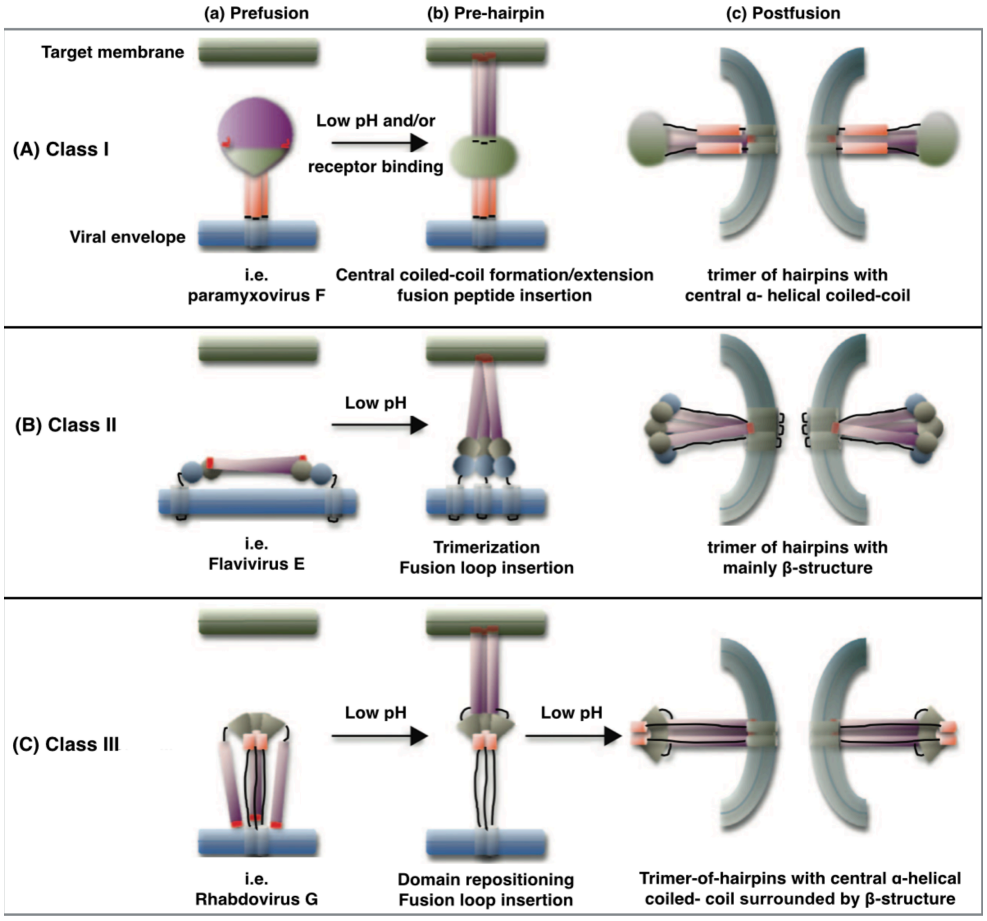


Fig. 3. Schematic representation of the general structural reorganization of the of EFPs in the three classes [adapted from (302)]. Fusion peptides/loops are in red. (A) Class I (i.e. paramyxovirus F): TM-proximal HR domains (orange rod) and FP-proximal HR domains (purple rod) form the trimeric coiled-coils. The former ones harbor the FP to the target membrane and jackknife to the latter ones to form a 6HB postfusion structure. (B) Class II (i.e. flavivirus E): DI (blue ball), DII (purple rod) and DIII (green ball) are indicated. DII rotates and propels the fusion loops into the target membrane. (C) Class III (i.e. rhabdovirus G): DI (black string), DII (orange blocks), DIII (green wedge) and DIV (purple rod) are indicated. DIV (fusion domain) and DII (trimerization domain) refold as hairpin-like formation.

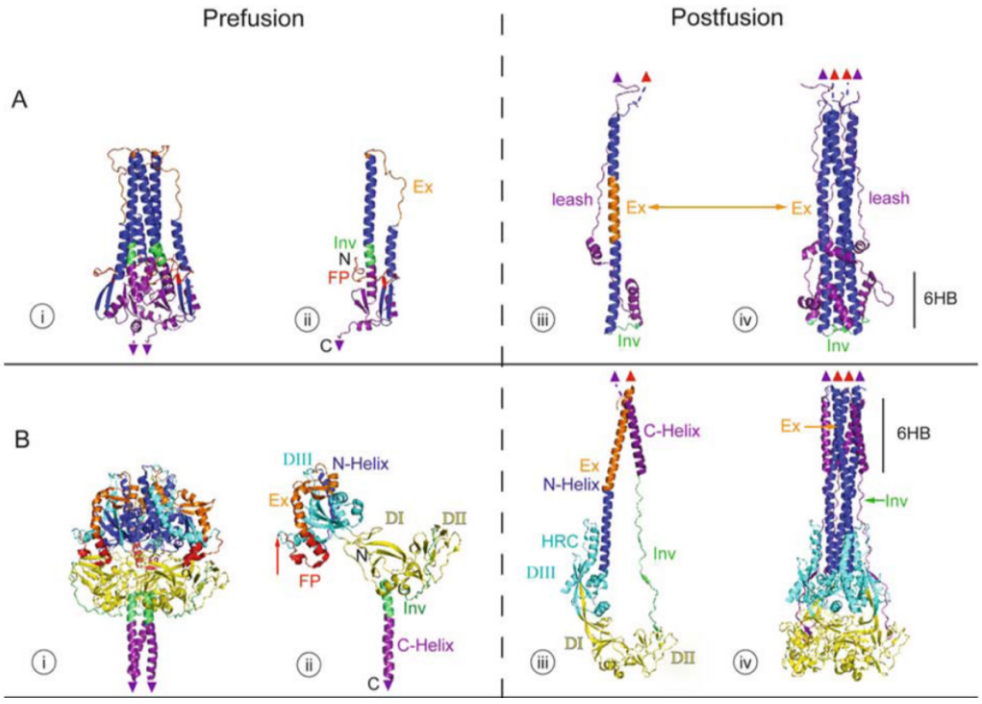


Fig. 4. Crystal structures of class I viral envelope fusion proteins in their pre- and post-fusion forms, represented by influenza virus HA2 (A) and paramyxovirus F (B) [adapted from (415)]. The pre-fusion forms (i and ii) are on the left and post-fusion forms (iii ad iv) are on the right. Both monomeric (ii and iii) and trimeric (i and iv) structures are shown. Fusion peptides (FP) with visible structure are in red and the ones without known structure are indicated as red triangles. C-terminal domains are in purple. Transmembrane domains (TMD) are indicated as purple triangle. N-terminal domains are in blue. The regions that facilitate TMD inversion are in green and labeled as Inv. The regions that facilitate FP extension to target membrane are in orange and labeled as Ex. In paramyxovirus F structures, domains I and II (DI and DII) are in yellow and domains III (DIII) are in cyan. PDB accession numbers of these protein structures are: 2HMG (A i and ii), IQU1 (A iii and iv), 2B9B (B i and ii) and 1ZTM (B iii and iv).

BACULOVIRUS

Genome organization and taxonomy

Baculoviruses are a family of large enveloped, circular double-stranded DNA viruses infecting exclusively arthropods and predominantly insects. The non-segmented genome of baculoviruses ranges from 80 to 180 kbp and encodes between 90 and 180 genes (134). To date, the complete genomes of 63 baculovirus species have been sequenced, revealing 35 core genes, conserved in all baculovirus species. The core genes are involved in essential biological functions, including gene transcription, DNA replication, genome packaging, virion

assembly and release, cell cycle arrest, interaction with host proteins and oral infectivity (102, 255). On the basis of alignment and phylogeny of the core genes the baculoviruses have been taxonomically divided into four genera: *Alphabaculovirus* (lepidopteran-specific), *Betabaculovirus* (lepidopteran-specific), *Gammabaculovirus* (hymenopteran-specific) and *Deltabaculovirus* (dipteran-specific) (Table 2). The alphabaculoviruses are subdivided into two groups - group I and group II according to the phylogenetic analysis of polyhedrin genes (74) and other conserved genes (156). The separation is correlated with presence of two distinct BV envelope glycoproteins, i.e. GP64 (group I) and F (group II), for virus-cell fusion and receptor binding (224).

Table 2. Baculovirus taxonomy with some representative members with complete sequenced genomes in each genus.

Genus	Species	Abbreviation	Accession
Alphabaculovirus **Group I**	*Autographa californica* MNPV	AcMNPV	NC_001623
	Bombyx mori NPV	BmNPV	NC_001962
	Epiphyas postvittana NPV	EppoNPV	NC_003083
	Orgyia pseudotsugata MNPV	OpMNPV	NC_001875
	Plutella xylostella MNPV	PlxyMNPV	NC_008349
Alphabaculovirus **Group II**	*Chrysodeixis chalcites* NPV	ChchNPV	NC_007151
	Helicoverpa armigera NPV	HearNPV	NC_003094
	Helicoverpa zea SNPV	HzSNPV	NC_003349
	Lymantria dispar MNPV	LdMNPV	NC_001973
	Spodoptera exigua MNPV	SeMNPV	NC_002169
	Spodoptera frugiperda MNPV	SfMNPV	NC_009011
	Spodoptera litura NPV	SpliNPV	NC_011616
	Trichoplusia ni SNPV	TnSNPV	NC_007383
Betabaculovirus	*Cydia pomonella* GV	CpGV	NC_002816
	Helicoverpa armigera GV	HearGV	NC_010240
	Spodoptera litura GV	SpliGV	NC_009503
Gammabaculovirus	*Neodiprion abietis* NPV	NeabNPV	NC_008252
	Neodiprion lecontei NPV	NeleNPV	NC_005906
	Neodiprion sertifer NPV	NeseNPV	NC_005905
Deltabaculovirus	*Culex nigripalpus* NPV	CuniNPV	NC_003084

The genome accession numbers are from NCBI (http://www.ncbi.nlm.nih.gov/). Abbreviations: NPV, nucleopolyhedrovirus; MNPV, multiple nucleopolyhedrovirus; SNPV, single nucleopolyhedrovirus; GV, granulovirus.

Phenotypes and infection cycle

Baculovirus virions appear in two phenotypes: the budded virus (BV) and the occlusion-derived virus (ODV) form (Fig. 5). BVs generally contain a single NC (282), whereas ODVs, on the other hand may contain one or more nucleocapsids (NC) within a surrounding envelope and is protected by a proteinaceous

occlusion body (OB), which contains one or more ODVs (see below). Although the two phenotypes contain identical viral genomes and NCs, they differ in the surrounding envelope structure and lipoproteic composition. The ODV envelope is more rigid than the BV envelope due to the presence of more saturated fatty acids and a larger percentage of nonglycosylated proteins (41, 339). The functional difference between these two forms is, that BVs are responsible for systemic infection within the insect larvae through cell-to-cell transmission and the ODV is responsible for the initial infection of the larval midgut and is therefore important for horizontal transmission of the infection in the insect population.

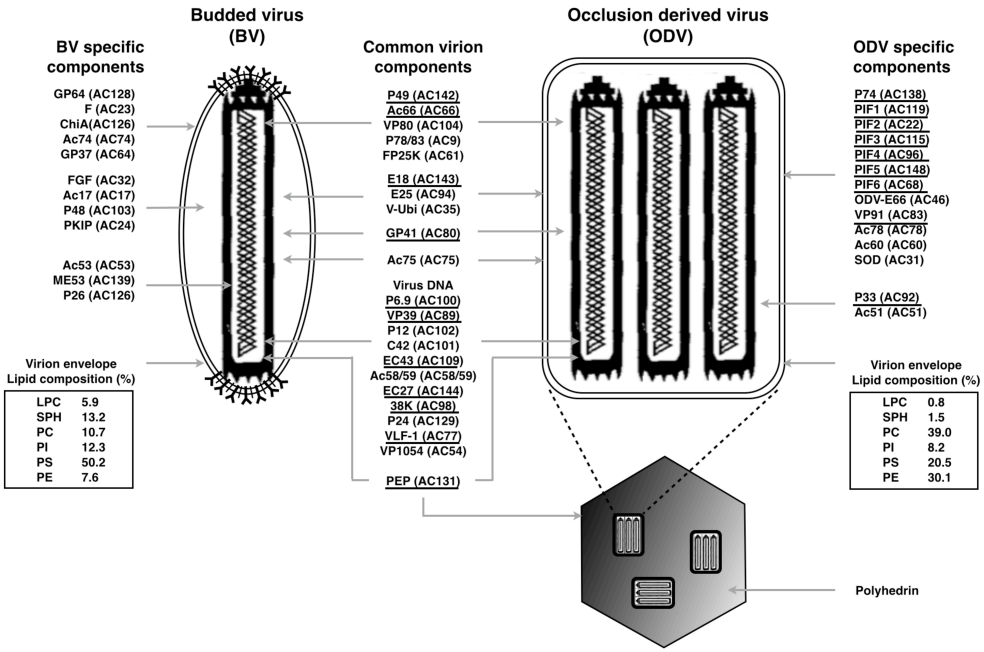


Fig. 5. Overview of virion structure and lipid/protein composition of the budded virus (BV) and occlusion-derived virus (ODV) in the type species Autographa californica multiple nucleopolyhedrovirus (AcMNPV) [adapted from (28, 144, 402)]. BV and ODV specific proteins are listed at the left and right, respectively. ODV is further embedded in a proteinaceous occlusion body (OB). Common components for both phenotypes are indicated in the center. The locations of all the proteins are indicated with arrows. The ORF numbers of AcMNPV are indicated in brackets. The underlined proteins are conserved in all baculoviruses (135, 402). The lipid composition of the BVs and the ODVs has been determined for isolated virions of AcMNPV-infected Sf9 cells (41). LPC (lysophoshatidylcholine), SPH (sphingomyelin), PC (phosphatidylcholine), PI (phosphatidylinositol), PS (phosphatidylserine), PE (phosphatidylethanolamine).

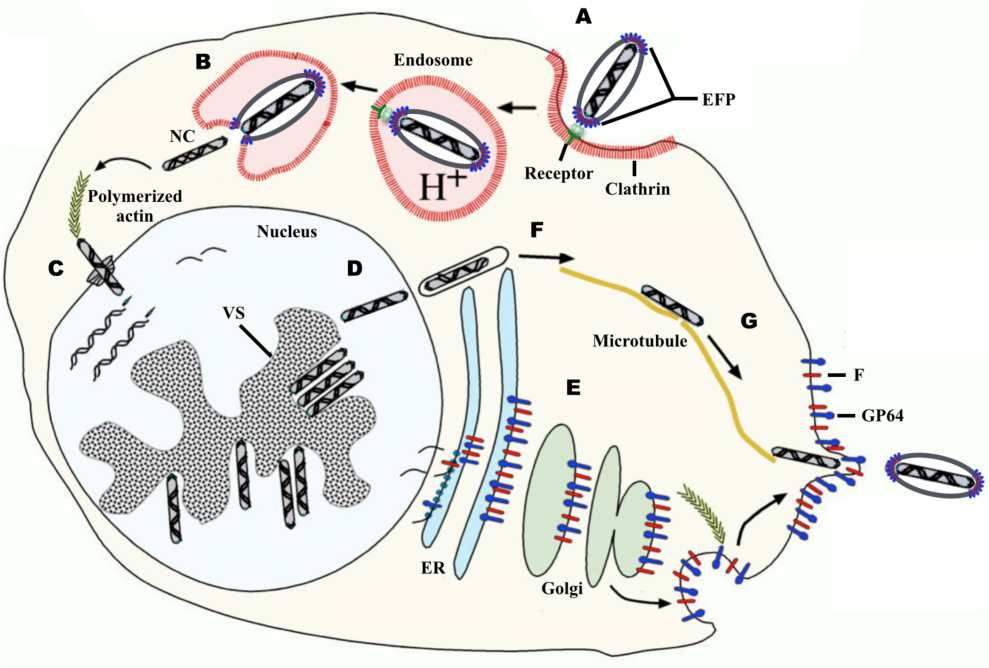


Fig. 6. Infection model of group I alphabaculovirus BVs [adapted from (315)]. BVs attach to the cell surface by receptor binding through a EFP and enter the cells via clathrin-dependent endocytosis (215, 392) (A). The acidified pH in endosome triggers the conformational changes of the EFP, which causes fusion between the viral envelope and the endosomal membrane. Membrane fusion results in fusion pore formation, releasing the nucleocapsid (NC) into the cytoplasm (B). The NC is transported via filamentous actin (F-actin) to the nucleus and enters the nucleus through a nuclear pore complex (112, 280) (C). The viral genes are transcribed (219, 293), DNA is replicated (49, 178) and new NCs are assembled (390) in the virogenic stroma (VS) (D). From there they are transported along nuclear F-actin to the nuclear periphery (127, 128). The envelope proteins are synthesized and glycosylated in the endoplasmic reticulum (ER) and Golgi apparatus. They are subsequently transported to the plasma membrane with F-actin (121) prior to NC transport. (E). Group I alphabaculoviruses have two BV envelope proteins, GP64 and a truncated F (See Baculovirus structural protein section). Group II alphabaculovirus, betabaculovirus and deltabaculovirus have only a single major envelope protein F, which is cleaved by furin in the Golgi (411). The NCs are thought to acquire a temporary envelope after budding through the nucleus and lose the envelope during anterograde trafficking (10) (F). The de-enveloped NCs are transported to EFP-enriched areas at the cell periphery by the motor protein kinesin-1 that moves along microtubules (71) and finally bud through the plasma membrane obtaining envelopes (G).

Based upon the morphology of their OB baculoviruses can be recognized as: nucleopolyhedrovirus (NPV in the genera *Alphabaculovirus*, *Gammabaculovirus* and *Deltabaculovirus*) or granulovirus (GV in the genus *Betabaculovirus*). NPVs produce polyhedron-shaped OBs with multiple ODVs containing single (S) or

multiple (M) NCs, which are denoted as SNPV and MNPV (no taxonomic meaning but are morphological types with practical convenience), respectively. The GV is featured with its granule-shaped OBs with single virions embedded (126).

The infection of insects occurs by oral intake of OBs, when arthropod larvae consume OB-contaminated plants. The OBs travel through the digestive tract and fall apart due to the alkaline pH in the midgut of the insect host, where ODVs are released. Upon membrane fusion of the ODV envelope and the membrane of the microvilli of midgut epithelial cells (116, 169, 361), the NCs are transported through the nuclear pore and enter the nucleus for gene transcription, DNA replication and progeny NC assembly. In the early phase of infection the progeny NCs are transported to the cell periphery and assembled into BVs by budding through the plasma membrane. The BV infection is best characterized in group I alphabaculoviruses, which is illustrated in Fig. 6. The BVs transmit the virus from cell to cell within the host, causing systemic infection. In the late phase of infection, BV production decreases and ODVs are generated within the nucleus of infected cells and occluded within a proteinaceous capsule constituted by the polyhedrin (NPVs) or granulin (GVs) protein to form OBs. At the end of the infection cycle larval cadavers liquefy and OBs are released to spread the infection to new hosts [reviewed in (315)].

Genome expression

Similar to other large DNA viruses, baculoviruses hijack the host transcriptional machinery using their transcriptional enhancers and activators. Transcription of baculovirus genes is divided into four sequential phases but overlapping during the infection cycle: immediate-early, delayed-early, late and very late. All the genes are distributed randomly with short intergenic sequences rather than according to the transcriptional phase in the circular genome (134). The early genes are transcribed with host RNA polymerase II (95) and are expressed into components for DNA replication and other early functions. It has been suggested that the major early transcriptional activator IE-1 is involved in regulating DNA replication by binding to homologous region (*hr*) sequences, which are found interspersed throughout the baculovirus genomes (64, 207, 314). Two major sequences in the genome are necessary for early gene transcription including a TATA promoter motif and a transcriptional initiation consensus motif CAGT (29). The late and very late genes are transcribed by a virus-encoded RNA polymerase on a conserved promoter motif (A/G)TAAG (30). The delayed-early

gene products include LEFs (late expression factors), which are required for DNA replication and late transcription (129, 330). In the late phase, nucleocapsid and envelope-associated proteins such as GP64 and F are synthesized (285, 297). Some genes like *gp64* and *F* genes the expression is regulated by transcription from both early and late promoters and is expected to occur throughout the infection (31, 297). During the very late phase polyhedrin and p10 are expressed for ODV occlusion and OB release, respectively (382, 412).

Baculovirus applications

Because of the inherent insecticidal activities to a broad range of insect pests and the complete safety to vertebrates the baculoviruses have been used as insecticides since the 1940s (259). Up to date about 60 baculovirus-based pesticides have been used to control lepidopteran, hymenopteran and coleopteran pests in agriculture, forestry and horticulture all over the world (27, 68). For example, SeMNPV as a member of group II in the genus *Alphabaculovirus* (134) is specifically infects *Spodoptera exigua* (beet army worm). The larvae of this moth form a widely distributed polyphagous pest to vegetables and flowers including cotton, tomato, cabbage, soybean and tobacco (332). SeMNPV has been used as commercialized biopesticide in Europe, USA and Thailand (26, 177, 191, 343).

Although the natural baculovirus pesticides have achieved a lot of success several obstacles remain to be overcome to make them more economical and effective. Compared to chemical pesticides the baculoviruses have slow killing speed, low field stability and high production costs (151). Therefore, various recombinant baculoviruses by inserting insect enzymes, hormones, toxins or deleting endogenous viral hormones have been successfully developed to enhance the killing speed and efficacy [reviewed in (18)]. Moreover, the first case of resistance to baculoviruses has been reported concerning the apple codling moth and the virus *Cydia pomonella* GV (CpGV) (3, 9, 434). Currently, efforts are taken to use CpGV strains that are effective against resistant moth larvae (25).

Extensive studies on baculoviruses, initiated by bio-control purposes, have allowed the further exploitation of their application. High gene expression levels under polyhedrin and p10 promoters (342, 381) as well as its large insertion capacity of at least 38 kb (61) and post-translational modifications similar to those in mammalian cells (167) made the baculovirus expression vector system into a competent expression system to produce recombinant proteins, including antigens for subunit vaccines for veterinary and human use (73, 260) as well as proteins for structural studies (171). Although baculoviruses exclusively infect arthropods, group I alphabaculoviruses containing GP64 can transduce

mammalian cells without any viral replication or integration of viral DNA into the host genome (54, 252, 376, 391). Hence the baculovirus have been shown an extraordinary safety advantage compared with other virus-based vectors. The detailed examples of trials as expression and delivery vectors were reviewed in (53, 69). An interesting observation here is that transduction of group II alphabaculoviruses is much less efficient.

Virion-associated structural proteins

The BVs and ODVs carry different major envelope proteins. The ODV entry into midgut epithelial cells is facilitated by a large protein complex (named as PIF complex) including the envelope proteins P74, PIF1, PIF2, PIF3, PIF4 and P95 (298). The essential envelope proteins related to BV entry are identified as GP64 (class III EFP; in group I alphabaculoviruses) and F (class I EFP; in other baculoviruses except the gammabaculoviruses), respectively. The GP64 protein is the major envelope glycoprotein on the surface of BVs of group I alphabaculoviruses. In AcMNPV, GP64 is one of the three most abundant viral encoded proteins in BVs, along with the capsid structural protein VP39 and the DNA binding protein P6.9 (402). GP64 is responsible for receptor binding (130, 437), low-pH dependent membrane fusion (32) and efficient budding (286). GP64 is characterized as a class III EFP with an internal fusion loop, an extended α-helical domain, a stem domain proximal to transmembrane domain (105) and shares structural homology with VSV G protein and HSV-1 gB protein (161) (see also Table 1). As the receptor binding protein GP64 possesses several critical hydrophobic residues in the fusion loop, which have been demonstrated in AcMNPV to be involved in attachment of the fusion loop to the host cellular membrane prior to low-pH triggered conformational changes of GP64 (81). This suggests that GP64 does not require specific receptors. Although baculoviruses are infectious to arthropods, the group I alphabaculoviruses have been found to enter mammalian cells without replication (391). Further studies also reported that the receptors to GP64 could be common molecules in mammalian cells such as phospholipids (371) and heparan sulfate (420), whereas they could be more specific receptors in insect cells (420).

The F proteins are present in BVs in all the baculovirus genera except *Gammabaculovirus*, for which the BV phenotype is absent (156). The F protein in group II alphabaculoviruses is the most abundant BV envelope protein (144) and a functional analog to GP64 in receptor binding (401), low-pH triggered membrane fusion (150) and virus budding (214). Group I alphabaculoviruses

possess a truncated F homolog (also known as F-like protein or Ac23) which somehow appears to enhance viral infectivity (225, 398). In betabaculovirus, the F protein from *Agrotis segetum* GV (428) and *Plutella xylostella* GV (427) were found to functionally replace GP64 in a GP64-null AcMNPV. The putative F protein in *Deltabaculovirus* (dipteran-specific) has not yet been studied in detail.

The following content in this paragraph will focus on the characteristics of F proteins of group II alphabaculoviruses as they have been studied in relative detail. The F protein shares common functional and structural domains with class I EFPs (Fig. 7) including a proteolytic cleavage site, hydrophobic fusion peptide, heptad repeat regions (HRs) in addition to a signal peptide and a transmembrane domain for all classes of EFPs. The cleavage of baculovirus F proteins occurs during post-translational modification by furin and prior to low pH activation of the F protein. Mutation of the furin cleavage site resulted in the loss of F fusogenicity. The furin-cleaved F protein is separated into a small N-terminal fragment (F_2) and a large C-terminal fragment (F_1), which are connected by a disulfide bridge (411). The fusion peptide, located directly downstream of the furin cleavage site and at the N-terminus of F_1, is liberated upon cleavage.

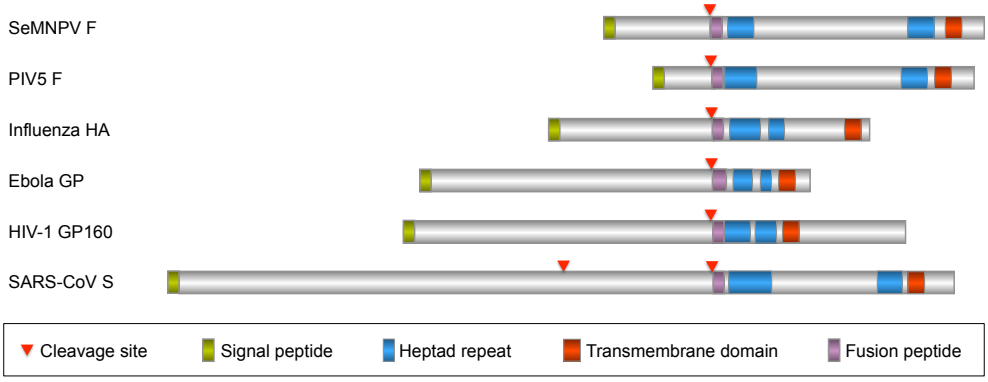


Fig. 7. Schematic representation of class I EFPs aligned at the proteolytic cleavage site: SeMNPV F (*Spodoptera exigua* multiple nucleopolyhedrovirus, *Baculoviridae*), PIV5 F (Parainfluenza virus 5, *Paramyxoviridae*), Influenza A virus HA (*Orthomyxoviridae*), Ebola virus GP (*Filoviridae*), HIV-1 GP160 (Human immunodeficiency virus, *Retroviridae*) and SARS-CoV S (Severe acute respiratory syndrome coronavirus, *Coronaviridae*). Conserved regions are indicated.

From an evolutionary perspective the baculovirus is unique in that it is the only found DNA virus encoding a functional class I EFPs until now. Such F proteins are primarily found in enveloped RNA viruses of vertebrates, such as ortho- and

evolution within this group mimics the evolution of baculoviruses as a whole. Since insects, and hence baculoviruses, are much older than vertebrates and their viruses, the baculovirus F protein can be considered ancestral to vertebrate virus EFP proteins. GP64 is a class III envelope protein and very distinct from F, both in protein sequence and structure. The gp64 gene appears to be acquired from a different virus family and entered the *Baculoviridae* later than the *F* gene (157). The significant amino acid sequence homology of GP64 with the glycoprotein encoded by members of the genus *Thogotovirus* (tick-borne orthomyxovirus) particularly with their conservation of cysteine residues provides evidence of their evolutionary relationship (268). Since its functional replacement by GP64, the F protein of the group I alphabaculoviruses may have evolved under a relaxed selection pressure eventually yielding an F-like protein that has lost its function as an EFP (157). In addition, the GP64 proteins in group I alphabaculoviruses have very high amino acid sequence identity (>74%), whereas the F proteins in both group II alphabaculoviruses and betabaculoviruses are relatively diverse (20-40%), but structurally related (295). On the other hand, the baculovirus F protein has significant sequence homology with the envelope protein of the gypsy retrovirus [also known as *Errantivirus* (33)]. Evidence suggests that these two proteins might have a common ancestor (232, 319). The errantiviruses are thought to have acquired a baculovirus ancestral F protein as their envelope (Env) protein by recombination after integration of a retrotransposon into a baculovirus genome. The transposition of retrotransposon was observed in a mutated AcMNPV genome in which a gypsy retrovirus TED encoding Gag, Pol and Env proteins integrated from an AcMNPV host, *Trichoplusia ni* (258). The sequence relatedness of the *env* gene and baculovirus *F* gene suggests that the acquisition of baculovirus *F* gene could have occurred (295).

A number of essential structural proteins associated to the nucleocapsid (NC) orchestrate NC formation of BVs as well as ODVs (Fig. 5). The major capsid protein conserved in all the baculoviruses is VP39, which distributes on the surface of the NC sheath (40, 373). Another capsid protein VP80, which is only conserved in alphabaculoviruses (234) and typically is located on one end of the NC sheath (235), plays a key role in viral DNA encapsidation and NC release from virogenic stroma during virion production (218, 235, 369). Other essential, conserved NC proteins 38K (422), AC109 (206), 49K (244, 385), VP1054 (284), BV/ODV-EC27 (385), BV/ODV-C42 (385) and P78/83 (112) are directly or indirectly responsible for NC assembly (40, 144, 402). BV/ODV-EC27, BV/ODV-C42 and P78/83 form complexes and are located on the NC

EC27, BV/ODV-C42 and P78/83 form complexes and are located on the NC sheath (39). P78/83 is a phosphoprotein associated with the basal structure of the capsid (325, 387). It activates actin polymerization, which plays an important role in nuclear transport of the NC after virus entry (112) and translocation to the cell surface for BV production (280). BV/ODV-C42 appears to recruit P78/83 to nuclei by means of its nuclear location signal (403). P6.9 (VP12) is a basic protein rich in arginine and histidine residues that binds with high affinity to the negatively charged phosphates of the viral DNA chain to condensate the genome and to organize the supercoiled structure of viral DNAs in the NCs (40, 379, 399, 419). VLF-1 is the only known integrase encoded by baculoviruses and is involved in DNA supercoiling and packaging through site-specific cleavage, recombination and ligation of two DNA strands; therefore, it is essential for capsid morphogenesis (245, 256, 383, 384).

SCOPE OF THE THESIS

Baculovirus F proteins are unique in that they occur in this case in a DNA virus, since they are more common in RNA viruses. Another interesting feature is that baculovirus F proteins, in contrast to GP64, do not readily infect mammalian cells, despite the large degree of structural homology between class I EFPs. Although much information has been acquired over the years on the functional aspects of baculovirus F proteins (163), the fusion mechanism is still not resolved and a 3D structure is still lacking. Such information may be important to understand the functional differences between baculovirus F proteins and F proteins of vertebrate origin with reference to entering insect versus mammalian cells.

In this dissertation my goal is to understand the fusogenic properties of baculovirus F proteins taking the SeMNPV F protein as a model. An important aim in this process is to solve the crystal structure of the SeMNPV F protein. In **Chapter 2** the generation of a soluble form of F in *Drosophila* cells is described and biochemical properties in comparison with the membrane-anchored F protein are characterized. In this way, it was confirmed that the acidified soluble form can be used as a model for the post-fusion form of the F protein as present in the viral envelope. In **Chapter 3** the 3D structure of the acidified soluble F protein is determined using X-ray crystallography in combination with cryo-electron microscopy (cryo-EM). In **Chapter 4** an F homolog is described in the mosquito *Anopheles darlingi* (Ad-F), which shares structural features and amino acid homology with baculovirus F proteins. The ability of Ad-F to mediate cell-cell fusion is analyzed, as well as it is evolutionary relationships to baculovirus F

triggered conformational changes of SeMNPV F. Since the fusion activation pH turned out to be lower than the pKa value of histidine, and 6 out of a total of 20 histidine residues are highly conserved among baculovirus F proteins, I hypothesize that one or more conserved histidines mediate the pH triggering. This hypothesis is tested transiently expressing various F mutants in cultured cells and performing fusion assays. In **Chapter 6** the baculovirus BV morphology is analyzed in detail by cryo-EM in order to better maintain the native structure of BVs than in previously adopted negative staining strategies. Finally, in **Chapter 7** by reviewing the results of the thesis, including the 3D structure of the baculovirus F protein and the morphology of the BV particle, the comprehensive functions and evolutionary aspects of the baculovirus F protein, as well as the scientific challenges and possible strategies for future research are discussed.

Chapter 2

Acid-induced rearrangement of the baculovirus fusion protein F

Qiushi Wang[1,2], Ying Tan[4], Ieva Vasiliauskaite[3], Scott Jeffers[3], Thomas Krey[3], Felix Rey[3], Peter Rottier[1] and Berend Jan Bosch[1]

Virology Division, Department of Infectious Disease and I mmunology, Faculty of Veterinary Science, Utrecht University, Utrecht, the Netherlands[1]

Virology Laboratory, Plant Science Group, Wageningen University, Wageningen, the Netherlands[2]

Structural Virology Unit, Department of Virology, Pasteur Institute[3]

Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, China P. R. [4]

Manuscript in preparation

ABSTRACT

The enveloped baculovirus budded virion (BV) carries envelope fusion proteins on its surface, which mediate membrane fusion upon its exposure to acidic pH encountered during endocytosis. Here we biochemically characterize the acid-induced changes of the baculovirus F protein - the BV fusion protein of most baculoviruses - which is thought to gain its fusion function by furin cleavage of a precursor during biosynthesis. We expressed and purified the F ectodomain (sF) of the F-containing baculovirus *Spodoptera exigua* multicapsid nucleopolyhedrovirus, and a furin cleavage site (FCS) deletion variant (sF^{FCSmut}), which were secreted as monomers. Trypsin proteolysis of sF as well as sF^{FCSmut} yielded a protein core fragment lacking the fusion peptide and part of a heptad repeat region, both located downstream of the FCS. The core fragment of sF undergoes an irreversible conformational rearrangement and trimerizes when exposed to acidic pH, which we anticipate to resemble the post-fusion F conformation. We were able to obtain X-ray diffracting, protein crystals of the acid-induced, trimeric trypsin-resistant core of the baculovirus BV F protein. Ultimately, we anticipate that the structure of the post-fusion F protein will not only give molecular insight into the fusion process but also may reveal the evolutionary ties of the BV F with F-homologous found in RNA viruses and eukaryotes.

INTRODUCTION

Baculoviruses, a family of enveloped double stranded DNA viruses, are highly pathogenic to arthropods, including a broad range of agricultural insect pests. Besides the direct application as biopesticides, baculoviruses are utilized as a platform for production of recombinant proteins and vaccines (90, 137, 220). Recently their potential as a safe gene delivery vector has been uncovered due to their capacity of mammalian cell transduction (8, 53). During infection two distinct virion phenotypes, the budded virus (BV) and occlusion-derived virus (ODV) are produced. BVs are generated at early phase of infection and are responsible for systemic infection of insect larvae. ODVs are produced at very late phase of infection and mediate transmission among hosts (259). Baculoviruses are divided into four genera: alpha-, beta-, gamma- and deltabaculoviruses. Alphabaculoviruses can be further divided into two subgroups, group I and group II (156), marked by the presence of two distinctive envelope fusion proteins GP64 and F proteins function, respectively. Within group I alphabaculoviruses, both GP64 and a F-homologue are present. However, membrane fusion is mediated solely by GP64 protein. The non-fusogenic F homologue in this group

was reported to enhance the infectivity but lost its fusion function possibly due to acquisition of GP64 during evolution (224, 398).

The baculovirus F fusion protein found in group II alphabaculoviruses, betabaculovirus and deltabaculoviruses shares structural features with several envelope fusion proteins of mammalian viruses, including the influenza A virus HA, paramyxovirus F and coronavirus S, which have been classified as class I viral fusion proteins (35, 216, 411, 415). The trimeric precursors of class I viral fusion proteins require proteolytic cleavage to generate a metastable pre-fusion form or sometimes referred to as 'fusogenic form' (56). After fusion activation by acidic pH and/or receptor binding, an internal hydrophobic region – the fusion peptide – is exposed to interact with the target cell membrane. This intermediate form undergoes subsequent conformational changes towards a trimeric, stable hairpin-like post-fusion form, during which a six-helix bundle is formed through the antiparallel interaction of two heptad repeat (HR) regions located down- and upstream of the fusion peptide and transmembrane domain, respectively (161, 304, 415). Like most class I fusion proteins, previous biochemical studies on F protein in *Spodoptera exigua* multicapsid nucleopolyhedrovirus (SeMNPV) demonstrated cleavage of a precursor protein F (F_0) by Golgi-resident furin (-like) proteases, which is required for fusion activity. This cleavage generates a C-terminal membrane anchored subunit (F_1) and an N-terminal subunit (F_2), which remain covalently linked via an intersubunit disulphide bond (411) (Fig.1A). In addition, the fusion peptide was located directly downstream of the furin cleavage site (410). Membrane fusion mediated by baculovirus F proteins is activated by acidic pH, which the virus encounters during endocytosis (150).

The long-term aim of our research was to obtain structural and functional insight into the fusion process of the baculovirus budded virion, by solving the X-ray structure of the soluble F ectodomain (sF) of *Se*MNPV. In order to achieve this we here characterize the sF protein using biochemical analyses and try to understand the biological relevance of sF compared to the full-length F protein. We show that sF does not undergo pH-triggered conformational change unless it is cleaved at the F_1 and F_2 junction. In addition, we demonstrate low-pH of cleaved sF induces irreversible trimerization of the monomeric ectodomain. In this study we demonstrated the conformational transitions of purified F ectodomain of SeMNPV, which resembles the features of viral envelope fusion protein during virus entry.

MATERIALS AND METHODS

Cells

Human embryonic kidney (HEK) 293T cells were cultured at 37°C with 5% CO_2 in DMEM (Lonza) supplemented with 10% fetal calf serum (PAA).

Plasmid construction

In order to construct a vector expressing the F ectodomain of F (sF), the region encoding the F ectodomain (Acc. no. AAQ11029) lacking its signal peptide (residues 18 to 553) was PCR amplified with Pwo high-fidelity DNA polymerase (Roche) using plasmid p166-AcV5-Se8 as a template (411). The PCR product was cloned into the pCD5 expression vector in frame with DNA sequence coding for CD5 signal peptide for efficient expression and secretion in mammalian cells (435). At the 3′ site, the gene fragment encoding the F-ectodomain was cloned in frame with a C-terminal enterokinase cleavage site for possibility of removal of downstream tags followed by three strep-tags for detection and purification with streptactin. The soluble F expression vector was designated as pCD5-sF-ST3. To construct a vector expressing full-length F protein, the SeMNPV F gene was PCR amplified with Pwo DNA polymerase using the p166-AcV5-Se8 plasmid as a template. The PCR product was cloned into the pCAGGS expression vector (138) in frame with a 3′-terminal strep-tag. The membrane-bound F expression vector was designated pCAGGS-F-ST2. Similarly, the expression vector pCAGGS-GP64-ST2 expressing the AcMNPV GP64 protein was constructed. Using the pCD5-sF-ST3 vector as a template, the pCD5-sF^{FCSmut}-ST3 vector encoding the F ecotodomain with the furin cleavage site mutated (RRXKR→SSXKK) was generated by QuikChange II Site-directed mutagenesis (Stratagene). Plasmids containing PCR-amplified regions were confirmed DNA sequencing.

Protein expression and purification

Expression of soluble F proteins was performed by transient transfection of 293T cells with the expression vector using polyethylenimine (PEI). The supernatant was harvested 5 days after transfection and centrifuged for 3000 x g for 15 min to remove cells and cellular debris. The protein was subsequently affinity-purified from the culture medium by Streptactin sepharose beads (iBA) and eluted from the beads using a biotin elution buffer (2.5 mM biotin, 100 mM Tris (pH8), 150 mM NaCl and 1 mM EDTA). Proteins were additionally purified by gel filtration (AKTA system, Amersham Biosciences) using the Superdex 200

10/300 column (Amersham Biosciences) with HNE buffer (0.15 M NaCl, 0.1 mM EDTA, 5 mM HEPES, pH 7.4) as a running buffer.

Limited proteolysis

Purified sF^{FCSmut} protein in HNE buffer (1 mg/ml) was incubated with indicated concentrations of TPCK (L-1-tosylamide-2-phenylmethyl chloromethyl ketone)-treated trypsin (Sigma) for 30 min at room temperature. Digestion was terminated by addition of 5 μg/ml of soybean trypsin inhibitor (Sigma). The cleaved protein was subsequently analyzed by gel filtration (Superdex 200 10/300 column, HNE buffer), by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) stained with Coomassie Blue, or by Western blotting (see below). Target protein bands were analyzed by N-terminal sequencing (Cambridge Peptides Ltd., Birmingham, UK).

Acid-treatment of sF^{FCSmut}

Purified sF^{FCSmut} protein in HNE buffer was incubated with TPCK-treated trypsin at a sF^{FCSmut}-trypsin ratio of 20:1 (w/w) and incubated for 30 min at room temperature. Digestion was stopped by adding soybean trypsin inhibitor (SBTI) at a SBTI-trypsin ratio of 5:1 (w/w). The cleaved protein was exposed to low pH by adding 0.5 M NaAc pH5 at a HNE-NaAc ratio of 11:1 (v/v) for 1 hour at room temperature and returned to neutral pH by adding 1M NaOH at a HNE-NaOH of 27:1 (v/v).

Western blot analysis

For electrophoresis under reducing and non-reducing conditions, samples were taken up in Laemmli sample buffer (125 mM Tris-HCl, 2% sodium dodecyl sulfate (SDS), 10% glycerol and 0.001% bromophenol blue, pH 6.8) with or without 5% β-mercaptoethanol, respectively. Samples were denatured for 5 min at 95°C. Proteins were subjected to SDS-PAGE electrophoresis and subsequently transferred to PVDF membranes (Bio-Rad) by wet electrophoresis transfer (Bio-Rad). The membrane was blocked for 1h at room temperature in PBS-T (PBS with addition of 0.1% Tween 20) containing 5% FBS. For detection of strep-tagged proteins, the membrane was incubated for 1 h at room temperature with horseradish peroxidase conjugated Streptacin (iBA) at a dilution of 1:5000 in blocking buffer. For detection of SeMNV-F protein or F_2 subunit, the membrane was incubated for 1 h at room temperature with polyclonal antiserum against SeMNPV-F_2 subunit raised from chicken (411) at a dilution of 1:2000 in blocking buffer and subsequently with horse radish peroxidase conjugated with goat anti-chicken immunoglobulin (SouthernBiotech) in blocking buffer with

three times washing between the incubations of two antibodies. After washing the membrane three times for 15 min each time in PBS-T and once for 15 min in PBS, proteins were detected using the ECL kit (GE Healthcare).

Deglycosylation assay

Purified sF^{FCSmut} protein, 2 μg in HNE buffer was mixed with 50 units of PNGaseF glycosidase and each reaction mix was incubated at 37°C for 0, 15, 30, 60 or 120 min individually. The reactions were stopped by adding non-reducing Laemmli sample buffer. Protein samples were analyzed by SDS-PAGE and gel filtration chromatography.

Multiangel light-scattering analysis

The molecular weights of different sF^{FCSmut} samples were determined by gel filtration combined with detection using Multiangel light scattering (MALS) and refractrometry (424). Purified protein (90 μg) was loaded onto a Superdex 200 10/300 column connected to an MALS instrument and an interferometric refractometer (DAWN HELEOS II, Wyatt Technologies, Santa Barbara, CA). The column was equilibrated either at pH5 (10 mM NaAc pH5, 150 mM NaCl, 0.1 mM EDTA) or pH7 (5 mM HEPES pH7, 150 mM NaCl, 0.1 mM EDTA), respectively. The absolute molecular masses were calculated using the ASTRA software (Wyatt Technology Corp., Santa Barbara, CA).

RESULTS

Expression of soluble F protein

In order to study the biochemical properties of soluble baculovirus F protein we expressed the F ectodomain of SeMNPV (sF) by transient expression in HEK-293T cells (Fig. 1A). To prevent rearrangement to the post-fusion form, we also expressed an sF variant containing a mutated furin cleavage site (sF^{FCSmut}: RRSKR→SSSKK). Proteins secreted into the cell culture supernatant were affinity-purified by Streptactin-sepharose chromatography and subsequent gel filtration chromatography. Under reducing SDS-PAGE conditions, purified sF migrates as three bands representing an uncleaved form of sF (sF_0) and the two furin-cleavage products, sF_1 and sF_2. Under non-reducing conditions the sF migrates as a single band around the 80kDa position, due to the presence of a disulphide bond connecting the sF_2 and sF_1 fragments. Purified sF^{FCSmut} migrates as a single band under both reducing and non-reducing conditions at the same position as sF_0, confirming the absence of furin cleavage during biosynthesis (Fig. 1B). The molecular weight of sF calculated from the amino acid sequence is 70 kDa. Assuming that each of the 6 predicted N-glycosylation sites contains a

glycan of ~2.5 kDa in size (276), the size of the expressed F ectodomain would measure ~82.5 kDa, which would be in accordance with the SDS-PAGE analysis.

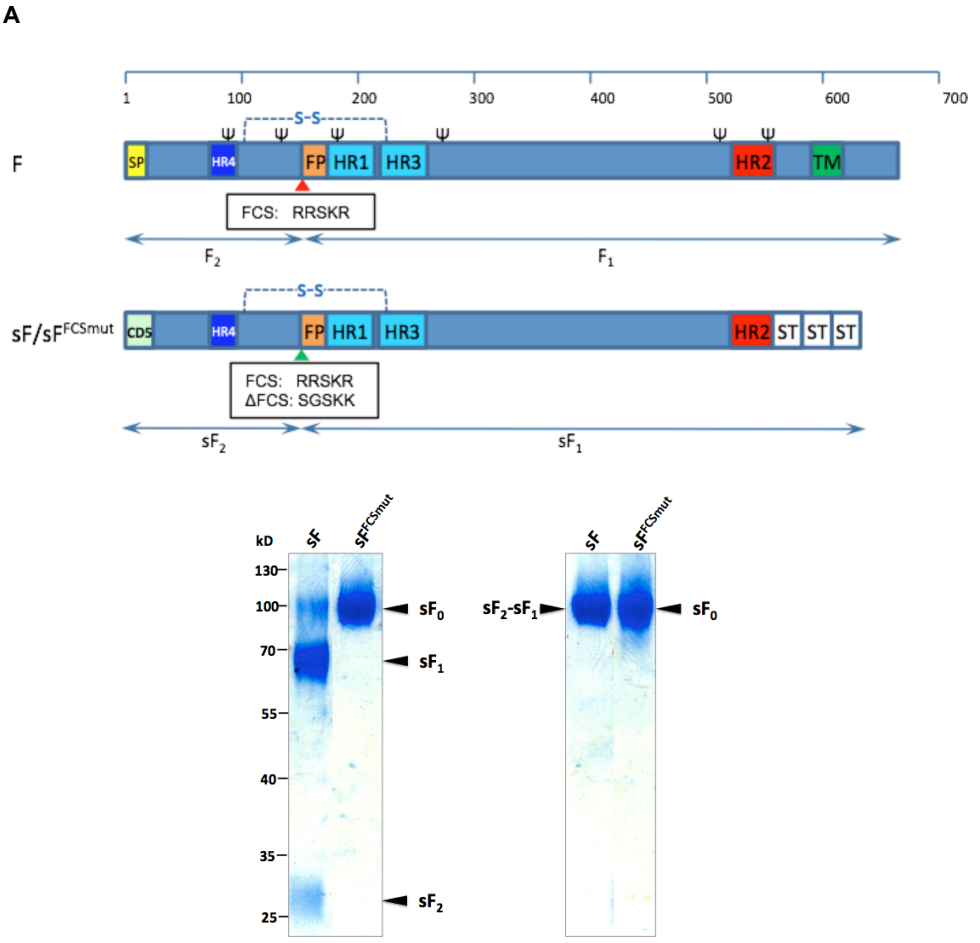
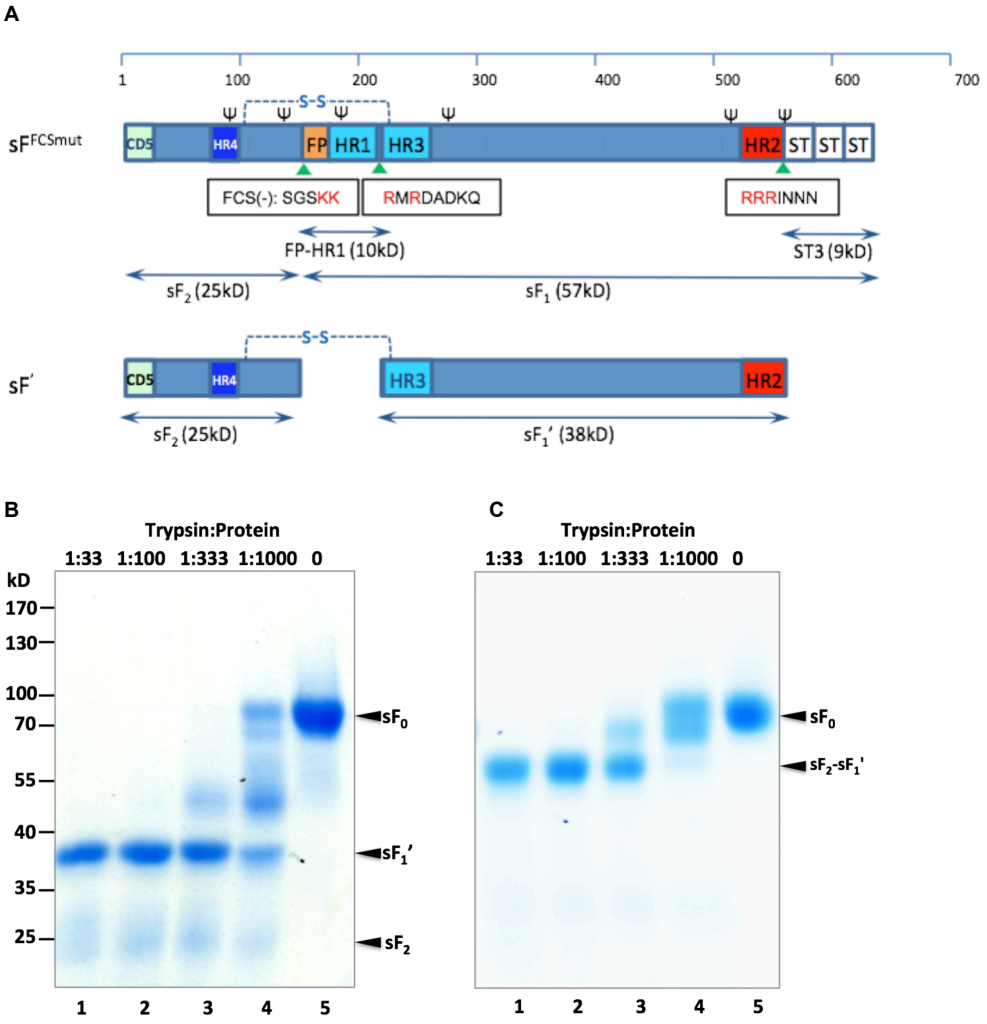


Fig. 1. Construction and expression of soluble F proteins. (A) Schematic representation of primary structure of SeMNPV-F protein and designed F ectodomain variants sF or sF^{FCSmut}. The domains indicated are signal peptide (SP) of F protein, replaced signal peptide (CD5), fusion peptide (FP), four heptad repeat regions (HR1, HR2, (211, 411), predicted HR3 and HR4 (Coiled coil prediction server: npsa-pbil.ibcp.fr), transmembrane domain (TM). Also indicated are the furin cleavage site (FCS, red arrowhead, (150), mutagenized FCS (green arrowhead) F1-F2 connecting disulfide bond (S-S, dotted line), strep-tag (ST) and computer predicted N-glycosylation sites (ψ) (NetNGlyc server: www.cbs.dtu.dk/services/NetNGlyc). The purified sF and sF^{FCSmut} were analyzed by SDS-PAGE under reducing (B) and non-reducing (C) conditions and stained with Coomassie Blue. Positions of marker protein and the uncleaved sF (sF_0), the sF_1 and sF_2 subunits and the disulphide-connected sF_1 and sF_2 subunits (sF_1-sF_2) are indicated.

Trypsin proteolysis demonstrates a stable core fragment of F ectodomain

We subjected the sF^{FCSmut} to limited trypsin proteolysis to biochemically characterize the protein. Proteolysis with increasing concentrations of trypsin lead to the formation of a predominant product appearing as a double band at 38 kD and 25 kD under reducing condition (Fig. 2B) and as a single band at 63 kD under non-reducing condition (Fig. 2C). This indicates that the 38 and 25 kDa are connected by an intersubunit disulphide bridge. Thus, the trypsin cleavage occurs between the two cysteines, which form disulfide bond. The 25 kD fragment under reducing condition is consistent with the molecular weight with



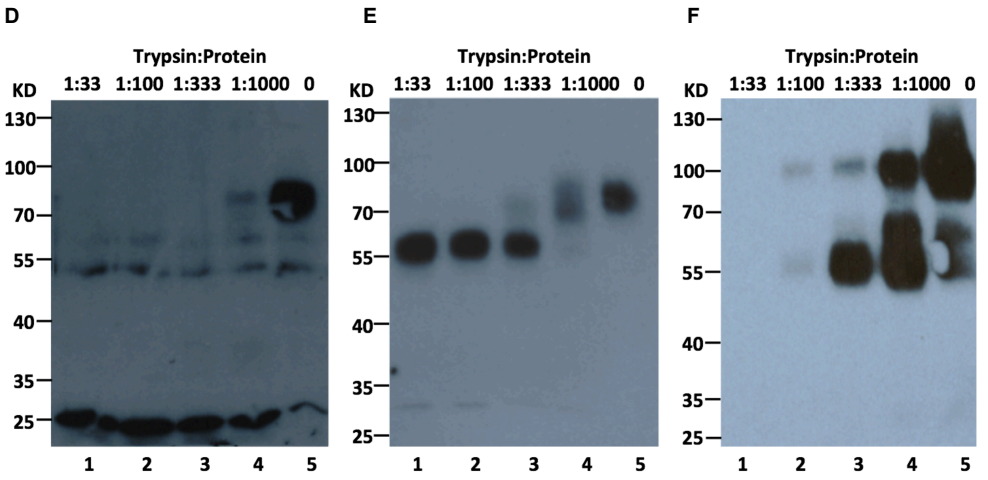


Fig. 2. Limited proteolysis of sF^{FCSmut}. (A), Schematic presentation of sF^{FCSmut} and its trypsin-resistant form, sF'. The trypsin-accessible loci are indicated with green arrowheads and red-coloured arginines (R) in the inidicated sequence. sF^{FCSmut} was mixed with trypsin in different ratio's and incubated at 22°C for 30 min and analyzed by SDS-PAGE under reducing (B) and non-reducing (C) conditions. Gels were stained with coomassie blue. The sF^{FCSmut} trypsin cleaved products were subjected to western blotting using an SeMNPV-F_2 antibody, performed under reducing (D) and non-reducing (E) conditions, and by western blotting using streptactin-HRP under reducing conditions (F).

sF_2 (Fig. 1B) which was confirmed by western blot using antibody against SeMNPV-F_2 subunit (α-SeF_2) (Fig. 2D), suggesting trypsin cleavage to occur at the F1/F2 boundary at the two lysine trypsin target sites present in mutated furin cleavage site (SGSKK). Under non-reducing conditions, the sF_2 is detected as part of the disulfide bond linked sF_1-sF_2 (Fig. 2E). The molecular weight of glycosylated sF_1 is approximately 57 kD from SDS-PAGE under reducing condition (Fig. 1B). Hence, the 38 kD fragment represents partially proteolysed sF_1 (sF'_1) lacking ~19kDa of the sF_1 subunit. N-terminal sequencing of the 38 kD sF'_1 fragment over 5 amino acids yielded two sequences (MRDAD and DADKQ) indicating that trypsin cleavage occurs after two closely spaced arginines (R↓MR↓DADKQ) in between the HR1 and HR3 regions. Trypsin cleavage at the sF_1-sF_2 junction and in between the HR1-HR3 domains, would remove a sF_1 fragment (FP-HR1; Fig. 2A) including the fusion peptide and HR1 domain, which contributes to a size reduction of approximately 10 kD including a predicted *N*-glycan. In addition, we observed that the detection of the C-terminal strep-tags in sF^{FCSmut} protein by western blot was gradually lost upon treatment with increasing amounts of trypsin indicating that the triple strep-tag is absent in the 38 kDa sF_1 fragment (Fig. 2F). The most adjacent trypsin target

sites (Arg or Lys) are located within the C-terminal end of the HR2 domain, which is five amino acids upstream of the Strep-tags (RRRINNNT-Strep-tags). Trypsin cleaves at one of these arginines, would account for the removal of a fragment (ST3; Fig. 2A) located at C-terminal of sF_1 of approximately 9 kD in size. Together, trypsin removal of the 10 kDa FP-HR1 and the 9 kDa 3x-streptag is in accordance with the observed size reduction of sF_1 (57 kDa → 38 kDa). The trypsin resistance pattern suggests that only a limited number of arginines or lysines in sF are accessible to trypsin due to the folding of the soluble protein. To summarize, trypsin cleaves sF at three positions; i) at the F_2-F_1 boundary, ii) between HR1-HR3 domains and iii) directly upstream of strep-tag cluster (Fig. 2A, green arrows), leaving the sF_2 and sF_1´disulphide connected.

Acid-treated soluble F monomer undergoes a conformational change and forms trimers in a trypsin-cleavage dependent manner

Furin-cleaved baculovirus F proteins utilize low pH to activate membrane fusion as indicated by their ability to mediate cell-cell fusion upon acid treatment (411). To analyse the impact of proteolytic cleavage and low pH on the sF protein, we investigated the effects of acid treatment of sF^{FCSmut} and the trypsin-cleaved sF^{FCSmut} using a combination of biochemical assays. The sF^{FCSmut} protein was mock-treated or treated with trypsin, subsequently exposed to low pH (pH5) or neutral pH, and analysed by SDS-PAGE, native-PAGE, gel filtration and multiangle light scattering (MALS). SDS-PAGE analysis confirmed the sF^{FCSmut} trypsin cleavage pattern seen before (Fig.2B). In contrast to neutral pH, the pH5-exposed, trypsin-treated sF^{FCSmut} protein shifts to a higher position in the native-PAGE (Fig. 3C, Lane 2) and elutes earlier during gel filtration (Fig. 3D, red solid line), suggesting a change in its oligomerization. Shift to a higher molecular weight form was only seen for a proportion of the neutral pH exposed, trypsin-treated sF^{FCSmut}, as demonstrated by the appearance of a double band by native PAGE and a double peak in the size exclusion chromatography profile. Gel filtration followed by MALS analysis was subsequently used to accurately determine the absolute molecular mass of the differentially treated sF proteins independent. MALS analysis indicated the non-cleaved, neutral pH treated sF^{FCSmut} to be in a monomeric form (Table 1). The non-cleaved, pH5-treated sF^{FCSmut} appeared to be a mixture of monomers and trimers. A full conversion to trimers was only seen for the trypsin- cleaved, pH5-treated sF^{FCSmut} protein. The acid-induced trimerization of sF' is irreversible since back titration of the sample to neutral pH prior to native PAGE analysis did not lead to trimer dissociation (Fig. 3C, Lane 5). We conclude that acidic pH triggers trimerization of sF^{FCSmut}

and the trimerization is more efficient upon trypsin cleavage at the F_2-F_1 boundary.

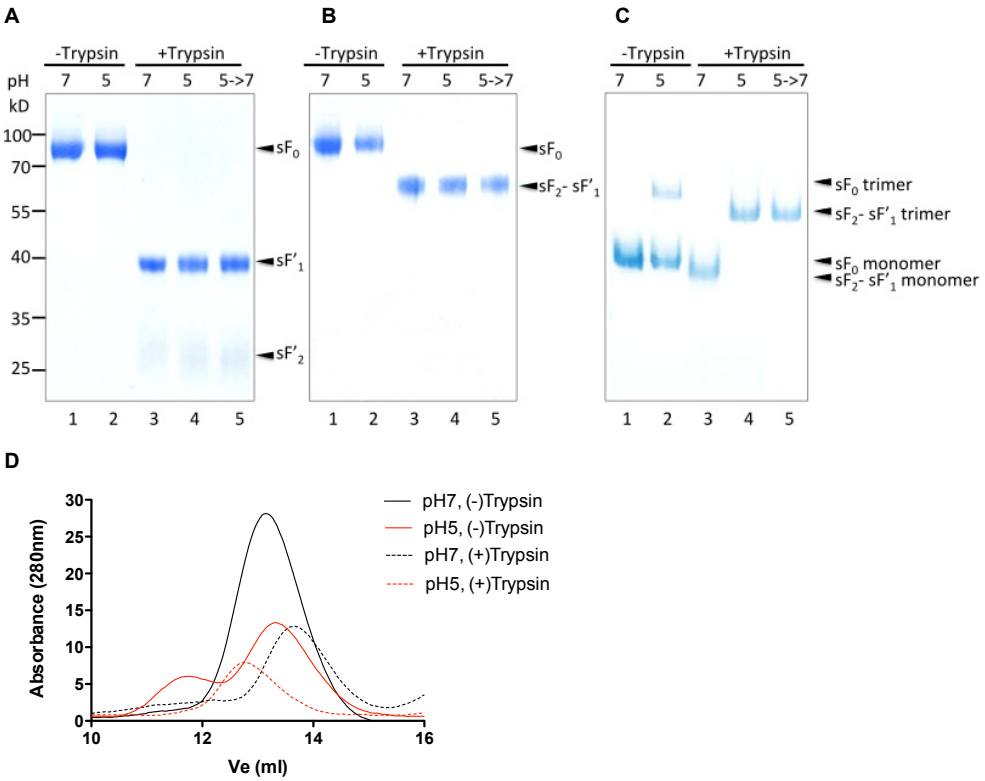


Fig. 3. Effects of acid treatment on sF^{FCSmut} and trypsin-cleaved sF^{FCSmut}. Mock-digested (-) or trypsin digested (+) sF^{FCSmut} protein was exposed neutral (pH7) or acidic pH (pH5) and analyzed by SDS-PAGE under non-reducing (A) and reducing (B) conditions, native-PAGE analysis (C), and gel filtration chromatography (D).

Table 1. MALS analysis of the oligomeric state of differentially treated sF^{FCSmut}

	Molecular weight [kD]	Mass fraction (%)	Oligomeric
- Trypsin pH7	95.4 (±5.5%)	100	Monomer
- Trypsin pH5	237.4 (±0.1%) and 91.3 (±0.2%)	36.4 and 63.6	Trimer and
+ Trypsin pH7	77.5(±0.2%)	100	Monomer
+ Trypsin pH5	188.0 (±1.9%)	100	Trimer

Membrane anchored F and F^{FCSmut} are folded into oligomers

Chemical cross-linking experiments have indicated a trimeric oligomeric assembly for the F protein of HearMNPV, a group II alphabaculovirus

homologue of *Se*MNPV (216). Since the sF^{FCSmut} protein was secreted as a monomer, we investigated the oligomeric state of the membrane anchored *Se*MNPV F and F^{FCSmut} proteins using coimmunprecipitation (coIP) analysis of differentially tagged full-length F proteins. We expressed C-terminal Strep- and HA-tagged versions of the SeMNPV F together or alone in HEK-293T cells. Similarly tagged versions of SeMNPV F^{FCSmut} proteins and the AcMNPV GP64 protein, the latter known to form disulfide-linked homotrimers, (216, 287) were also taken along. Lysates were prepared of cells co-expressing the HA- and ST-tagged protein. As a control, cell lysates of individually expressed proteins were pooled. Strep-tagged proteins or protein complexes were affinity-purified from these cell lysates and subjected to western blotting using multiclonal antibody against Streptag and HA-tag. Western blot analysis demonstrated that upon co-expression, the HA-tagged-F, $-F^{FCSmut}$ and -GP64 protein could be co-immunoprecipitated with the Strep-tagged version of the same protein (Fig. 5, Lane 3, 6 and 9). No coIP of HA-tagged proteins was seen when lysates of cells of individually expressed proteins were pooled prior to coIP analysis (Fig. 5, Lane 4, 8 and 12). The F proteins migrated as monomeric and oligomeric (presumably dimers and trimers) forms on the western blot. These oligomers were still visible when pre-treatment with *N*-ethylmaleimide (NEM) before cell lysis, which alkylates free thiol groups, preventing the formation of artificial interprotomer disulfide linkages (not shown). This implies the disulfide bonds might be involved in oligomerization of F.

Crystallization trial for pre-fusion form of soluble F protein

We set up the crystallization of sF^{FCSmut} and the acid-induced, trimeric trypsin-resistant core of the F protein, which we anticipate to resemble the post-fusion form of F. The sF_0 proteins produced from 293T cells were crystallized and diffracted up to 7Å as the highest resolution so far, which was not qualified for crystal structure analysis (Fig. 6). Glycosylation in mammalian cells results diverse and complex glycans (276). To increase the protein homogeneity we can either deglycosylate the heterogeneously glycosylated protein or express glycoprotein with high homogeneity in insect expression system. We are currently using Drosophila Schneider 2 cells expression system for further crystallography analysis.

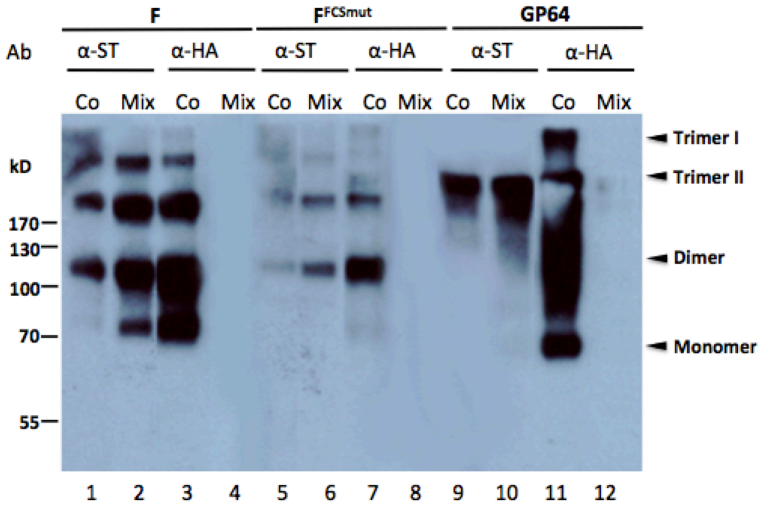


Fig. 5. Oligomerization of membrane-bound F protein. Strep-tagged F, F^{FCSmut} and GP64 proteins were co-expressed with the HA-tagged versions of the same protein in HEK-293T cells (Co). The cells were harvested at 24-hour post-transfection and lysed. As a control the cell lysates containing individually expressed strep-tagged and HA-tagged variants were mixed (Mix). Lysates were subjected to immunoprecipitation using Strep-tactin Sepharose. The precipitates were separated with 10% SDS-PAGE under non-reducing conditions and subjected to western blotting using antibodies against strep-tag (α-ST) or HA-tag (α-HA).

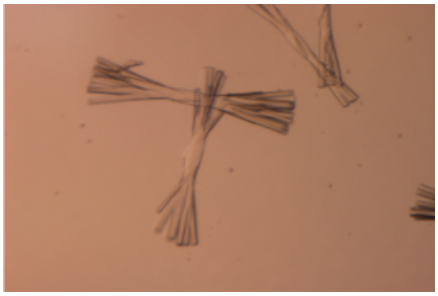


Fig. 6. Multi-mosaic crystals of sF_0 by vapor-diffusion method. The crystal diffracts up to 7 Å.

Structure prediction of the SeMNPV F protein

In attempt to get more structural information, we used the protein homology/analogy recognition engine (Phyre2) for the 3D structure prediction of the SeMNPV F protein based on homologous proteins of a library of known protein structures. The server can achieve high accuracy models based on very low sequence identities (15–25%). The Phyre2 program aligns close and remote sequence homologs, gathered by iterative searching using the PSI-BLAST algorithm. The secondary structure predicted from the alignment is then scanned against the library of known protein structures using a profile–profile alignment

algorithm. The highest scoring alignments, ranked by an *E*-value, are then used to construct 3D models of the query sequence (170).

The reliability of structure prediction based on known protein structures critically depends on the alignment accuracy of the query protein sequence and the known template structure. The PSI-BLAST search for remote homology yielded proteins containing an approximately 360 amino acid long domain of unknown function (DUF3609) which has been previously recognized in the F proteins of *baculoviridae*, in the coding sequences of *env* genes of endogenous insect retroviruses, and in cellular proteins in *Drosophila* and mosquito genomes (231, 319) (Fig. S1). It also reveals that the F proteins of *paramyxoviridae* as remote homologous of F proteins of *baculoviridae*, as recognized previously (262, 411). For example, respiratory syncytial virus (RSV) F and SeMNPV F display 20% sequence identity over a 358 residues long region as indicated (Fig. 7). We looked at the conserved residues among the F proteins of *paramyxoviridae* and *baculoviridae* and the consensus sequence of the 361-residues long DUF3609 domain (Conserved domain accession: pfam12259). Three cysteines (SeMNPV F C368, C377 and C392) and a single proline (SeMNPV F P316) were strictly conserved among the F proteins of *paramyxoviridae* as well as the proteins containing the DUF3609 domains, including the F proteins of *baculoviridae*. Moreover the spacing of the three cysteines and the proline in the polypeptide chain with respect to each other was highly conserved (222). Intriguingly when the F proteins of the baculovirus SeMNPV and paramyxovirus RSV were aligned at the position of the three conserved cysteines, the furin cleavage site, the fusion peptide and the downstream heptad repeat region of both F proteins were correspondingly aligned (Fig. 8). The region in between the furin cleavage site and the third conserved cysteine of RSV F and SeMNPV F are highly similar in length (SeMNPV vs. RSV: 243 and 246 residues, respectively) with a predicted high conservation of secondary structures (Fig. 8). Outside these regions there is more variability: The F2 region (SeMNPV vs. RSV: 149 and 136 residues, respectively) and most notably the region from C392 up to the transmembrane domain (SeMNPV vs. RSV: 188 and 144 residues, respectively).

Next, we extrapolated the disulfide connection based on the cysteine conservation among DUF3609 family members and the RSV F tertiary structure (Fig. 9). Within the DUF3609 domain, which spans residues 243-486 of the SeMNPV F protein, six out of the eight cysteines are conserved (SeMNPV F: C368, C377, C392, C410, C445 and C477). The C334 and C340 are missing in *Iris* elements of *Drosophila* species and the mosquito-infecting F protein in *Culex*

nigripalpus nucleopolyhedrovirus, a Deltabaculovirus (222, 319), which suggests that C334 and C340 are likely to pair. Based on the disulfide linkages within the RSV F crystal structure, it can be predicted that C368/C377 and C392/C410 within SeMNPV F make disulfide-linked pairs (366). In addition, the RSV F contains a disulfide bridge between two cysteines present in the F2 and the HRA region, respectively. Within SeMNPV F there is a single cysteine in the F2 region, which we predict to connect to either C218 or C229 located upstream or within the HR3 region, respectively.

```
SeMNPV  F  83   LRQNQTSAFDDCANVKYLKLEIDH----------MLSTVIPNLAQQHNLLDQKVPLTPSN  132
               +++N+ +  D  A VK +K E+D           ++ +  P   +    L + +  T +N
RSV F      64   IKENKCNGTD--AKVKLIKQELDKYKNAVTELQLLMQSTPPTNNRARRELPRFMNYTLNN  121

SeMNPV F   133  ATLTKATLSPTKRRSKRGLFNFMGHVDKYLFGIMDSDDAHELHMLANTTNSLNSQVKQLN  192
               A  T  TLS  K+R +R L   +G       G+  S   H    L    N + S +   N
RSV F      122  AKKTNVTLS--KKRKRRFLGFLLGVGSAIASGVAVSKVLH----LEGEVNKIKSALLSTN  175

SeMNPV F   193  DELIVLADYVDHEFHASRMRD----ADKQCRYIIENYNILCKQLDEVATLYNKLDLAVDN  248
                 ++ L++ V      S++ D     DKQ   I+   +     ++ V     K +  ++
RSV F      176  KAVVSLSNGVS--VLTSKVLDLKNYIDKQLLPIVNKQSCSISNIETVIEFQQKNNRLLEI  233

SeMNPV F   249  AKLNHLNSFVVSPERLLNEMNNVSGHLAGLSWPVPLTEKAMHVLIDNVINVH------VF  302
                +   +N+ V +P       N+    L  L   +P+T     ++ +NV  V       +
RSV F      234  TREFSVNAGVTTPVSTYMLTNS---ELLSLINDMPITNDQKKLMSNNVQIVRQQSYSIMS  290

SeMNPV F   303  VTAERKLLFIIEVPL---VSSEAFDVFHSIPLPYCDKSHKCAIMLPDSKYLGVSVDRRNY  359
               +  E  L +++++PL   + +  +   H+ PL   +      I L  +   G   D
RSV F      291  IIKEEVLAYVVQLPLYGVIDTPCWK-LHTSPLCTTNTKEGSNICLTRTDR-GWYCDNAGS  348

SeMNPV F   360  VR-LDDTTSCRMSDKVMLCFRPQIIYDVNQAKLCDVRIFMKNDKDIDYKRDCDVRVGR    416
               V       +C++    + C     +   ++  LC+V IF       + K DC +   +
RSV F      349  VSFFPQAETCKVQSNRVFCDTMNSLTLPSEINLCNVDIF-------NPKYDCKIMTSK    399
```

Score	Method	Identities	Positives	Gaps
28.9 bits(63)	Compositional matrix adjustment	72/358(20%)	139/358(38%)	46/358(12%)

Fig. 7. PSI-BLAST alignment of SeMNPV F (residues 83-416) and RSV F (Uniprot accession No.: P03420) (residues 64-399) and statistics.

DISCUSSION

We aim to study the structures of Baculovirus F protein at pre- and postfusion states using soluble F ectodomain of SeMNPV. It is therefore crucial to characterize the conformational features of F ectodomain prior to its structural study. In order to obtain homogenous proteins for protein crystallization we constructed F ectodomain with mutated furin cleavage site (sF^{FCSmut}), which prevents furin cleavage but is accessible to trypsin (Fig. 2). Other class I viral fusion proteins such as influenza virus HA and paramyxovirus F proteins showed that there were minor change in the core structure of the fusion proteins with or without proteolytic cleavage at neutral pH based on their determined structures

(Chen et al. 1998; Welch et al. 2012). The monomeric ectodomain of HA shows similar folded structure to each monomer in trimeric full length HA0 protein and is able to fold correctly and transported to the cell surface (333).

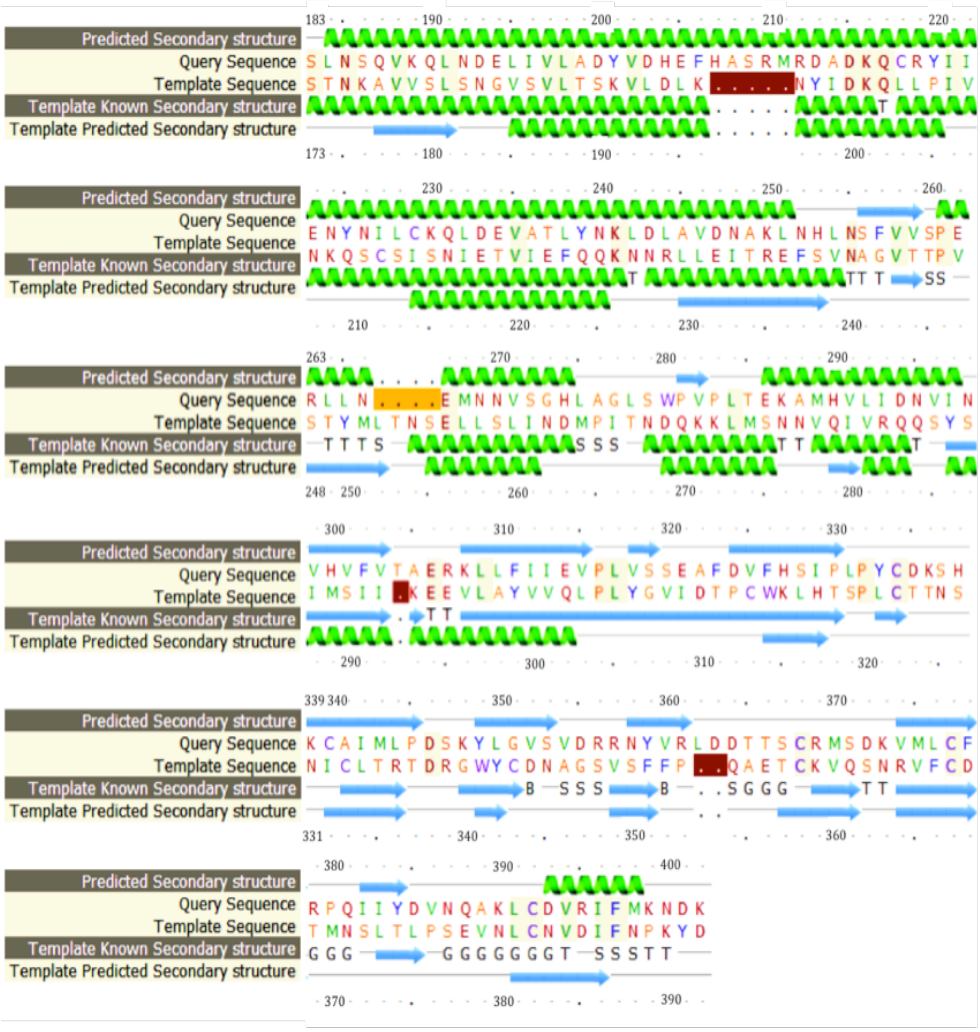


Fig. 8. Profile-to-profile alignment of SeMNPV F (Query sequence; residues 183-402) and RSV F (Template sequence; residues 173-392). Predicted and known secondary structural elements [helical regions (green helices) and beta-sheets (blue arrows)] are indicated.

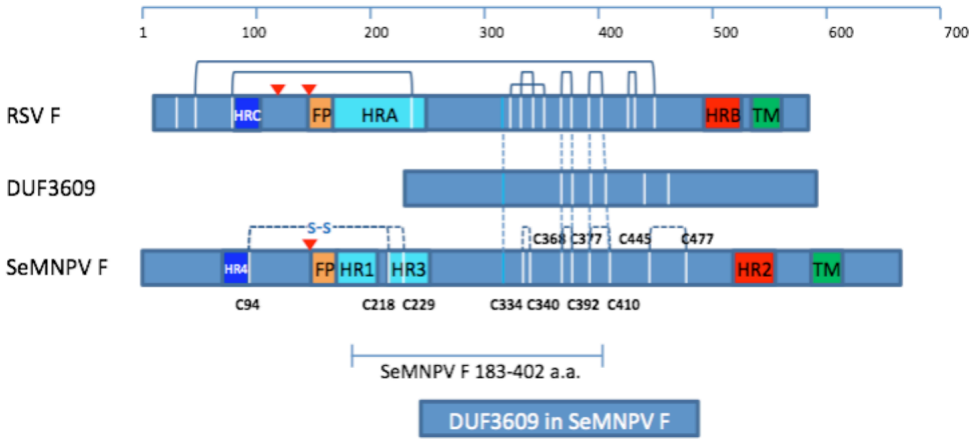


Fig. 9. Schematic representation of the F proteins of the paramyxovirus respiratory syncytial virus (RSV) and the baculovirus SeMNPV, and the DUF3609 domain aligned at conserved amino acids, indicated by the vertical dotted lines. Indicated are a conserved proline (blue vertical line), cysteines (white vertical lines, with corresponding amino acid positions for SeMNPV F shown underneath and above), known and predicted cysteine bridges (arches with solid respective dotted lines), heptad repeat (HR) regions, transmembrane domain (TM) and furin cleavage sites (113) (red arrowhead).

At neutral pH both cleaved and uncleaved sF^{FCSmut} are shown as monomers. Exposure of low pH triggers the transition from monomer to trimer and the oligomeric transition is more efficient upon proteolytic cleavage as at low pH uncleaved proteins appear to be a mixture of monomers and trimers, whereas trypsin cleaved proteins appear to be uniform trimers. Trypsin cleaved sF^{FCSmut} resists to aggregation upon low pH (Fig. 3) due to removal of fusion peptide by trypsin. On the contrary, F ectodomain (sF), which was cleaved by furin at only F_2-F_1 boundary, aggregated at a large extent at low pH as a result of exposure and clustering of fusion peptide in the absence of cellular membrane (Fig. S3), suggesting that cleaved F ectodomain is folded the fusion active structure. Furthermore, native-PAGE result shows that the migration patterns are the same for acidified protein with or without neutralization (Fig. 3), indicating that the conformational change activated by proteolytic cleavage is irreversible, which have been generally recognized in other class I fusion proteins (45, 201, 365, 366, 406). The conformational change of F ectodomain is triggered by low pH, which is consistent with requirement for transition of full length F protein from pre- to postfusion state (411). Taken together, the uncleaved sF^{FCSmut} at neutral pH and acidified, trypsin cleaved sF^{FCSmut} represent the pre- and post-fusion form of full length F protein, respectively.

Prior to fusion activation F ectodomain appears as monomer predominantly and upon activation the monomer undergoes refolding, resulting trimeric form (Fig. 3), which differs with full-length protein that is in trimeric form at both pre- and postfusion states (216) (Fig. 5). This indicates that transmembrane (TM) domain is important for protein trimerization at pre-fusion stage but not at post-fusion stage. A number of studies of other viral fusion proteins have observed inefficient trimerization without TM domain (45, 333, 341, 423). Equilibrium of monomers and trimers of VSV G ectodomain were observed in solution at high and neutral pH but undergo complete trimerization after crystallization due to lowering the energy level (7, 313), suggesting that trimeric interface in TM domain does not possess trimeric interface but may rather stabilize the trimeric prefusion structure. The crystal structures of HA reveal large difference between the pre- and postfusion trimeric interfaces (45, 418). The trimeric pre-fusion structure of HA is possibly regulated and stabilized by TM domain, which holds the N-terminal adjacent heptad repeat (HR) region in place through trimeric TM-TM interaction (340). Thus, the interprotomer contacts at prefusion form are weaker compared to postfusion form. It is generally believed that the after fusion activation structural transition of class I viral fusion proteins leads to exposure of fusion peptide and folding back of the C-terminal HR region of the protein (404). Without fusion peptide the acid-driven conformational changes still occurred in SeMNPV F ectodomain. The trimeric interface is possibly the hydrophobic interfaces of HR regions. We therefore propose that acid treatment of F ectodomain reveals the buried trimeric interface on protomers in the prefusion state, which strongly promote trimerization.

For paramyxovirus F protein TM domain is not only important for protein trimerization but also for protein stability. Mutation of an amino acid involved in TM-TM interactions decreased the stability of prefusion forms of paramyxovirus F proteins (341). Removal of TM domain in hPIV3 sF^{FCSmut} leads to spontaneous transition to postfusion form without fusion activation (429). On the other hand, SeMNPV sF^{FCSmut} only undergoes conformational changes upon low pH activation (Fig. 3). In addition, conformational changes of paramyxovirus occur at neutral pH by receptor binding (306), whereas those of baculovirus occur at low pH. These results provide the evidence for that the energy barrier to overcome to transit to postfusion form for paramyxovirus F protein is lower than for baculovirus F protein.

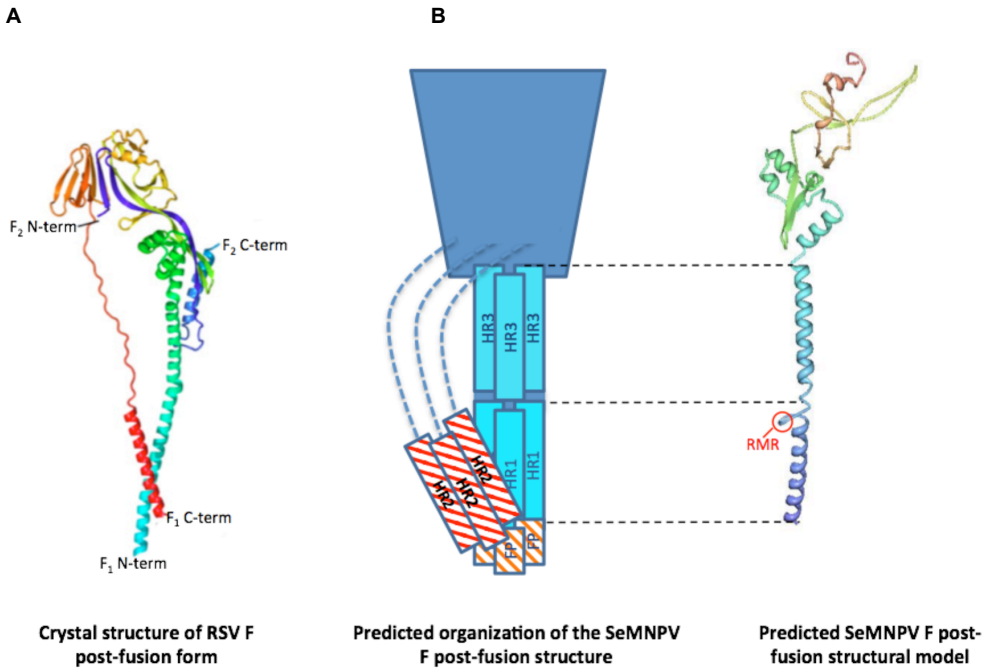


Fig. 10. Structural model of SeMNPV F based on RSV structure. (A) Ribbon diagram of RSV F post-fusion form. (B) Predicted organization of the SeMNPV F post-fusion structure (left) and ribbon diagram of structural model of the SeMNPV F (183-402 a.a.) (right) as predicted by Phyre2 program using the RSV F post-fusion structure as a template. The regions of predicted structural organization of SeMNPV F are color-coded in the same manner as Fig. 9. The HR2 and FP regions, which are not presented in the structural model, are represented as shaded boxes. The trypsin cleaved cluster RMR is indicated.

By searching with SeMNPV F protein sequences in Phyre2 protein fold recognition server using the profile–profile alignment the RSV F protein is identified as the highest scoring template. The server provides the 3D model for SeMNPV F based on RSV F crystal structure of post-fusion form (PDB accession ID: 3RKI, Fig. 10A). With the post-fusion structure of RSV F as a template, 212 residues (residues 183-402) of SeMNPV F were modeled (Fig. 10B) with an estimated precision score of 82.8%, indicating the probability that the query sequence and the template are homologous. The modeled structure includes part of HR1, HR3 and the conserved C368, C377 and C392 and P316 residues (Fig. 9). According to the class I fusion model, an antiparallel 6-helix bundle is formed during membrane fusion with a long inner trimeric coiled coil formed by the HR1 and HR3 regions and the HR2 region associated in an antiparallel fashion, bringing together the fusion peptide and transmembrane domain. Since the region connecting the HR2 region and the transmembrane domain is

considerably longer in SeMNPV F compared to that of RSV F (~28 and 7-10 residues, respectively), we predict that the helical HR2 region in the SeMNPV F post-fusion structure associates with the HR1/3 coiled coil more distal from the fusion peptide (e.g. with HR3) than seen for RSV F (Fig. 2A).

Currently we are investigating the postfusion structure of SeMNPV F protein using acidified, trypsin cleaved sF^{FCSmut} ectodomain. The cleaved sF^{FCSmut} ectodomain lacks of fusion peptide and HR1 region as protein sequencing and SDS-PAGE reveal that trypsin cleaves at two arginines (R↓MR↓DADKQ) in between the HR1 and HR3 regions and at the sF_1-sF_2 junction. The postfusion model shows the RMR cluster is located in the flexible hinge of HR3 and HR1, hence easily accessed and cleaved by trypsin. It is therefore necessary to prevent the trypsin cleavage by arginine mutations in the RMR cluster. Interestingly the hinge is absent at similar location in RSV F protein, suggesting the RSV F protein might adapt the sequence during the evolution to avoid undesirable proteolytic cleavage. Comparing the two postfusion structures of SeMNPV F with and without FP-HR1 region will help us to understand the influence of this flexible region to the structure.

Despite a low sequence homology between SeMNPV and RSV F proteins (sequence identity: 16.5%) (Fig. S2) the two proteins share significant structure similarities in large part of F1 subunit, which covers the longest stalk domain (SeMNPV F: HR1-HR3 vs. RSV F: HRA) and partial head domain. Previous studies indicate that the HRA and HRB regions in F1 subunits of paramyxovirus undergo a dramatic refolding to form hairpin structure (core trimer) upon fusion activation, which leads to fusion peptide anchoring to target cell membrane and merge of lipid bilayers (187) and therefore play the central role of membrane fusion. Due to the structural similarity the coiled coil regions of SeMNPV F protein may function the same way as RSV F protein. However, no other paramyxovirus F proteins were found structurally related to SeMNPV F. For example, PIV3 and PIV5 F proteins are structurally conserved with RSV F at head domain in F1 subunit but not like SeMNPV F conserved with RSV F at stalk domain (366). The other part of head domain of paramyxovirus is constituted by F2 subunit, which also shows no discernable structural relationship with baculovirus F2. The head domains of the two F proteins are likely bearing different functions. Baculovirus F2 has been reported to be implicated in receptor binding (400) and paramyxovirus F2 is indicated to regulate fusion states through interactions with HRA (103). The head domain of paramyxovirus F is also thought to contact directly with attachment protein

through receptor binding to trigger conformational change of F (196, 303) whereas the head domain of baculovirus F may have other functions since the conformational change is triggered by low pH. The difference in function may explain the dissimilarity of head domain structures of paramyxovirus and baculovirus F proteins.

SUPPLEMENTARY DATA

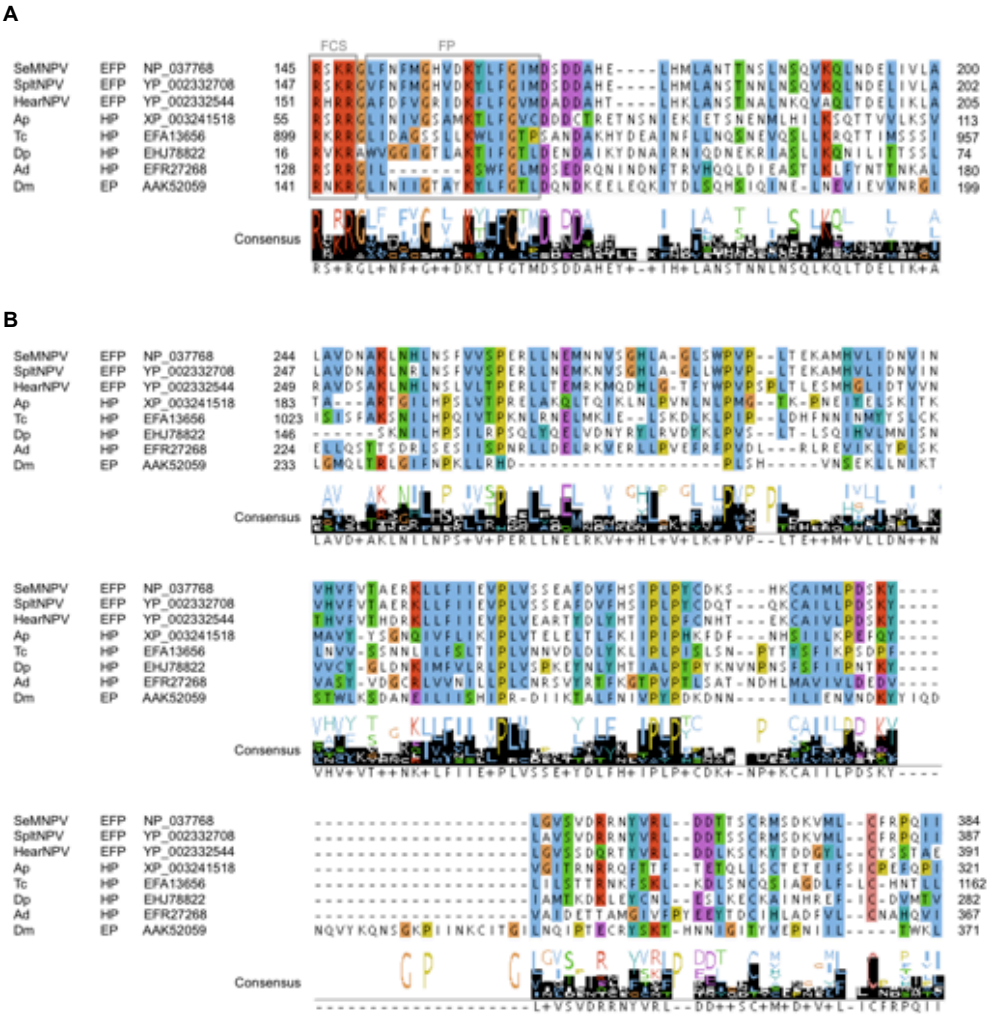


Fig. S1. Multiple sequence alignment of SeMNPV F and homologues at furin cleavage site (FCS), fusion peptide (FP) regions (boxed) (A) and DUF3609 domain (B). The sequences have been selected from the PSI-BLAST search with SeMNPV F protein sequence (GenBank accession no. AAF33539). The alignment was obtained with NCBI multiple sequence alignment tool (ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi) and deposited with Jalview program (jalview.org) and the conserved amino acids are colored with ClustalX and depicted at the bottom of sequence alignment. Abbreviations: EFP envelope fusion protein, EP envelope protein, HP hypothetical protein, SeMNPV *Spodoptera exigua* multiple nucleopolyhedrovirus, SpliMNPV *Spodoptera litura* nucleopolyhedrovirus II, HearNPV *Helicoverpa armigera* multiple nucleopolyhedrovirus, Ap: *Acyrthosiphon pisum*, Tc: *Tribolium castaneum*, Dp: *Danaus plexippus*, Ad: *Anopheles darlingi*, Dm: *Drosophila melanogaster*.

```
SeMNPV F          1 ----MLRFKVIVWLVAALTVEAKFAKDIVQVTPLPSTSGLYFQYINRMQF     46
                      :|:...|..::||:|:....:::|.:.....:.|.:...|::.:
RSV F             1 MELPILKTNAITTILAAVTLCFASSQNITEEFYQSTCSAVSKGYLSAL--     48

SeMNPV F         47 VQNIWH---FVIEMDHGSVFYRLQSIHQQAQKLQQSFISLRQNQTSAFDD     93
                     :..|:   ..||:.                          ::::|:.:..|
RSV F            49 -RTGWYTSVITIELS-----------------------NIKENKCNGTD-     73

SeMNPV F         94 CANVKYLKLEIDHMLSTVIPNLAQQHNLLDQKVPLTPS----------NA    133
                    |.||.:|.|:|.     ..|...:..||.|..|.|.:           |.
RSV F            74 -AKVKLIKQELDK-----YKNAVTELQLLMQSTPATNNRARRELPRFMNY    117

SeMNPV F        134 TL-----TKATLSPTKRRSKRGLFNFMGHVDKYLFGIMDSDDAH---ELH    175
                    ||     |..|||  |:|.:|.|...:|.......||..|...|   |::
RSV F           118 TLNNTKNTNVTLS--KKRKRRFLGFLLGVGSAIASGIAVSKVLHLEGEVN    165

SeMNPV F        176 MLAN---TTN----SLNSQVKQLNDELIVLADYVDHEFHASRMRDADKQC    218
                    .:.:   :||    ||::.|..|..:::.|.:|:             |||.
RSV F           166 KIKSALLSTNKAVVSLSNGVSVLTSKVLDLKNYI------------DKQL    203

SeMNPV F        219 RYIIENYNILCKQLDEVATLYNKLDLAVDNAKLNHLNSFVVSP-------    261
                    ..|:...:....::.|.....|.:..::..:...:|:.|.:|
RSV F           204 LPIVNKQSCSISNIETVIEFQQKNNRLLEITREFSVNAGVTTPVSTYMLT    253

SeMNPV F        262 -ERLLNEMNNVSGHLAGLSWPVPLTEKAMHVLIDNVINVH------VFVT    304
                     ..||:.:|:           :|:|.....::.:||..|.      :.:.
RSV F           254 NSELLSLIND-----------MPITNDQKKLMSNNVQIVRQQSYSIMSII    292

SeMNPV F        305 AERKLLFIIEVPLVSSEAFDVFHSIPLPYCDKSHKCAIMLPDSKYLGVSV    354
                    .|..|.:::::||        :..|..| |.|.|...:...::|. |.::
RSV F           293 KEEVLAYVVQLPL--------YGVIDTP-CWKLHTSPLCTTNTKE-GSNI    332

SeMNPV F        355 -----DRRNYVRLDDTTS---------CRMSDKVMLCFRPQIIYDVNQAK    390
                         ||..|  .|:..|         |::....:.|.....:...::..
RSV F           333 CLTRTDRGWY--CDNAGSVSFFPQAETCKVQSNRVFCDTMNSLTLPSEVN    380

SeMNPV F        391 LCDVRIFMKNDKDIDYKRDCDVRVGRFESELFYATSDYNNWLYVLQNDID    440
                    ||::.||       :.|.||.:...:                  .|:
RSV F           381 LCNIDIF-------NPKYDCKIMTSK--------------------TDV-    402

SeMNPV F        441 LNIQCIPSATITDGFGIAPVVLRAGVGIIHATGNDNCKLTTKKSRLTVHD    490
                           |::.:...|          .|:...|...|..:.|.
RSV F           403 -------SSSVITSLG----------AIVSCYGKTKCTASNKN-------    428

SeMNPV F        491 LYNNLNTVIEIPMGLSYNFTVALQDIDKISVDDMKINNDLEHTNLHELTS    540
                              .|:...|:....:...||.:.:.|.|.:.|..|..|
RSV F           429 ------------RGIIKTFSNGCDYVSNKGVDTVSVGNTLYYVNKQEGKS    466

SeMNPV F        541 ---------RLYD---------------LRRRINNNTVFSGSQVVDDNGD    566
                             ..||               :..:||.:..|     :..:.:
RSV F           467 LYVKGEPIINFYDPLVFPSDEFDASISQVNEKINQSLAF-----IRKSDE    511

SeMNPV F        567 IFAGMSSWFSSIGIDFHYVKIVMIWIVMAMLTLATVKIYRTCCSGACS-T    615
                    :...::..|:..|   .:..::|.|::.:|:|..|.:...|.:.:.. |
RSV F           512 LLHNVNAGKSTTNI---MITTIIIVIIVILLSLIAVGLLLYCKARSTPVT    558

SeMNPV F        616 L-------CNNFKICRGSEHTVVRREDRDMYYQTTLPKYKRGKKHVDSIF    658
                    |       .||.....
RSV F           559 LSKDQLSGINNIAFSS---------------------------------    574

SeMNPV F        659 DMEMEPM    665

RSV F           575 -------    574
```

Score	Method	Identities	Similarity	Gaps
155.0	Needleman-Wunsch alignment	125/757 (16.5%)	245/757 (32.4%)	275/757 (36.3%)

Fig. S2. Global pairwise sequence alignment of SeMNPV F and RSV F proteins using EMBOSS Needle (www.ebi.ac.uk/Tools/psa/emboss_needle).

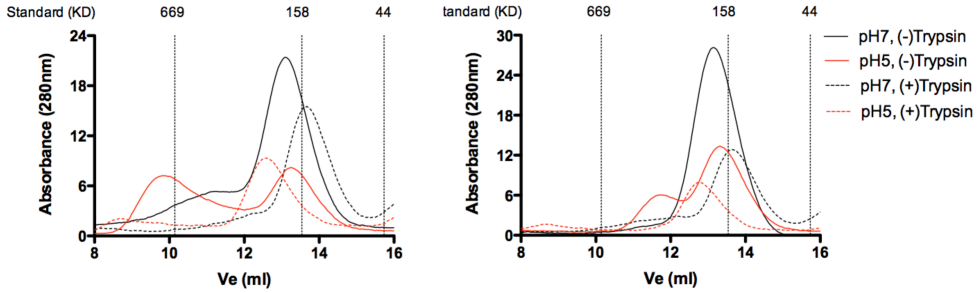


Fig. S3. Size exclusion chromatography analysis of sF (A) and sF^{FCSmut} (B) proteins

Chapter 3

The postfusion structure of baculovirus fusion protein F

Ieva Vasiliauskaite[1], Qiushi Wang[2, 3], Just M. Vlak[3], Peter J.M. Rottier[2], Berend Jan Bosch[2], Thomas Krey[1] and Felix A. Rey[1]

Structural Virology Unit, Department of Virology, Institut Pasteur, Paris, France[1]

Virology Division, Department of Infectious Disease and Immunology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, the Netherlands[2]

Laboratory of Virology, Wageningen University, Wageningen, the Netherlands[3]

Manuscript in preparation

ABSTRACT

Viral fusion proteins of all enveloped viruses mediate membrane fusion between host and viral membranes during cell entry. Most baculoviruses utilize the fusion protein (F) for low pH dependent membrane fusion. We report the crystal structure of the post-fusion trimer of a trypsin-truncated F fragment. This structure confirms previous predictions that baculovirus F protein adopts a class I fusion protein fold and is homologous to the paramyxovirus F protein. Baculovirus F is therefore the first class I fusion protein encoded by a DNA virus. The results support the hypothesis that F proteins may have a common ancestor and imply interesting evolutionary links between DNA and RNA viruses and their hosts.

INTRODUCTION

Baculovirus envelope fusion proteins

Baculoviruses are enveloped, circular double-stranded DNA viruses with a large genome (80-180 kbp) and are predominantly pathogenic to insects of the order *Lepidoptera*, *Hymenoptera* and *Diptera* (134). Based on phylogeny of baculoviruses and classifies its members into four genera: *Alphabaculovirus* (lepidopteran-specific NPVs), *Betabaculovirus* (lepidopteran-specific granulosis viruses), *Gammabaculovirus* (hymenopteran- specific NPV) and *Deltabaculovirus* (dipteran-specific NPV) (156).

The entry of baculovirus budded viruses (BVs) into their host cells is mediated by a specific envelope glycoprotein, either GP64 or F (32, 150). F proteins are not only more widespread within the *Baculoviridae* family; they are also more diverse (with amino acid identities of between 20-40% and >74% for F and GP64, respectively). Both proteins have similar biological functions and have been demonstrated to be involved in receptor binding, low-pH dependent fusion of virus and cellular membranes, and efficient budding (32, 130, 150, 216). However, the molecular basis of action of these proteins is significantly different. The three-dimensional structure of GP64 has been determined by X-ray crystallography, revealing that it belongs to the structural class III fusion proteins, which do not require proteolytic cleavage in order to be activated (161). The F protein has been suggested to perform a function analogous to GP64 based on the fact that the infectivity of a *gp64*-null *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) can be restored by introducing the F gene (223). Phylogenetic analyses imply that the F protein is likely to be an ancestral fusion protein of baculoviruses while GP64 has been incorporated into the

baculovirus genome relatively recently (157, 295). GP64 is only found in a subset of baculoviruses, the group I alphabaculoviruses. These viruses still encode a non-fusogenic F protein homologue (F-like protein) that most likely lost its fusogenic function due to the acquisition of GP64 during evolution (223, 398).

Characteristics of baculovirus F protein

One of the most studied F proteins comes from the group I alphabaculovirus *Spodoptera exigua* multicapsid nucleopolyhedrovirus (SeMNPV). Biochemical characterization of SeMNPV F protein has revealed that this protein contains the features characteristic to class I fusion proteins. SeMNPV F protein is secreted as a ~76kD large precursor called F0. The precursor is post-translationally cleaved by subtilisin-like endoprotease furin. The mutation of the furin cleavage site RSKR (amino acid residues 145-149) results in a loss of the fusogenic activity, indicating that the cleavage is essential to generate a fusion-competent protein (411). The furin cleavage yields two disulphide-linked subunits: a small N-terminal subunit F2 (theoretical molecular weight ~15kD), and a bigger membrane-anchored C-terminal subunit F1 (theoretical molecular weight ~59kD). The class I fusion proteins occur as homotrimers at the surface of a viral particle [reviewed in (415)]. The exact oligomeric state of SeMNPV F is unknown but it is predicted to be trimeric based on the fact that the F protein of closely related group II alphabaculovirus, *Helicoverpa armigera* multicapsid nucleopolyhedrovirus (HearMNPV), assembles as trimers on the virus envelope (216). The furin cleavage occurs upstream of the hydrophobic sequence located at the N-terminus of the membrane anchored F1 subunit. The stretch of the first 18 amino acid residues at the N-terminus of F1 subunit (150-GLFNFMGHVDKYLFGIMDS-168) has been suggested to represent a fusion peptide because it contains features that are previously described for viral fusion peptides: 1) it is hydrophobic, 2) it can form an amphipathic helix with conserved glycines at one side, and 3) it shows a high degree of conservation among baculovirus F proteins (414). However, some differences with vertebrate viral fusion proteins can be identified such as the absence of alanine residues and a higher number of polar residues. The deletion of amino acid residues 151-170 in F protein resulted in the loss of virus infectivity even though the protein was incorporated into viral particles and was cleaved by furin, supporting the role of this amino acid stretch as a fusion peptide (410).

The structural similarity between baculovirus F and paramyxovirus F proteins

The SeMNPV F protein has been predicted to share structural features with the paramyxovirus F protein. In contrast to baculoviruses, which are DNA viruses,

paramyxoviruses are negative-sense RNA viruses. The envelope glycoprotein F is responsible for virus fusion with the cellular membrane. Paramyxovirus F proteins are also synthesized as F0 precursors and cleaved in the trans-Golgi complex by furin into two subunits, F1 and F2, with the fusion peptide located at the newly generated N-terminus of the F1 subunit.

The important difference between the entry of baculoviruses and paramyxoviruses is that the latter fuse at the cell surface in a pH-independent manner, while baculoviruses mediate fusion in the endosomes in an acidic-pH environment. Moreover, paramyxoviruses use a separate attachment protein for binding to the cell surface, in contrast to baculoviruses, which have both membrane fusion and receptor binding activities in the same envelope glycoprotein F. The interaction of the paramyxovirus attachment protein with the cellular receptor triggers conformational changes in the prefusion F trimers, which eventually lead to the fusion of viral and cellular membranes [reviewed in (36, 155, 301)].

Of all paramyxoviruses, the F protein of human respiratory syncytial virus (RSV) was identified as having the highest amino acid identity (12%) to SeMNPV F (Chapter 2). Despite low amino acid sequence identity, these two proteins display conserved positioning of the furin cleavage sites, the secondary structure elements, and the transmembrane domains, which serve as indications of functional homology between their corresponding genes.

The crystal structures of F proteins in pre-fusion and/or post-fusion form have been determined for a number of paramyxoviruses: human parainfluenza virus 3 (hPIV3), human parainfluenza virus 5 (hPIV5), and New Castle disease virus (NDV) fusion protein. The crystal structures of RSV F in its post-fusion and pre-fusion forms are available (246, 365), revealing a class I fusion protein fold observed in previously solved F structures from other paramyxoviruses [human parainfluenza virus 3 (hPIV3), parainfluenza virus 5 (PIV5), Newcastle disease virus (NDV), Simian virus 5 (SV5) and metapneumovirus (MV)] (14, 57, 365, 406, 407, 429). The RSV F post-fusion trimer (Fig. 1) has an elongated shape with a globular head domain on the top and long stalk domain constituted by intertwined α- helices that form a stable 6-helix bundle (6HB) at the membrane proximal end of the molecule. The helices of the 6HB are composed by two HRs: HRA adjacent to the fusion peptide, and HRB located upstream from the transmembrane region. HRAs of three protomers form a central core of a triple-stranded coiled-coil creating three grooves into which the C-terminal HRBs pack antiparallel to the central core. In contrast, the most prototypical class I fusion

protein, influenza HA, forms just a small 6HB at the membrane distal end of the protein as it contains just an extended segment the C-terminal HR helix that packs into the grooves of the central N-terminal coiled-coil. Each protomer of RSV F in its post-fusion form is composed of three domains, termed DI, DII and DIII. The globular head domain is mostly composed of DI and DII. At the base of the head domain is DIII, which carries a long HRA helix that extends down and intertwines with HRA helices of the other two protomers to form a central coiled-coil of the 6HB. HRB helices extending from DII interact with the central coiled-coil to form the outer helices of a 6HB. The fusion peptide is located at the N-terminus of HRA while the transmembrane region (not present in the crystal structure) is positioned at the C-terminus of HRB. These two elements are located at the bottom of the stalk and are inserted into the cellular membrane in the full-length F.

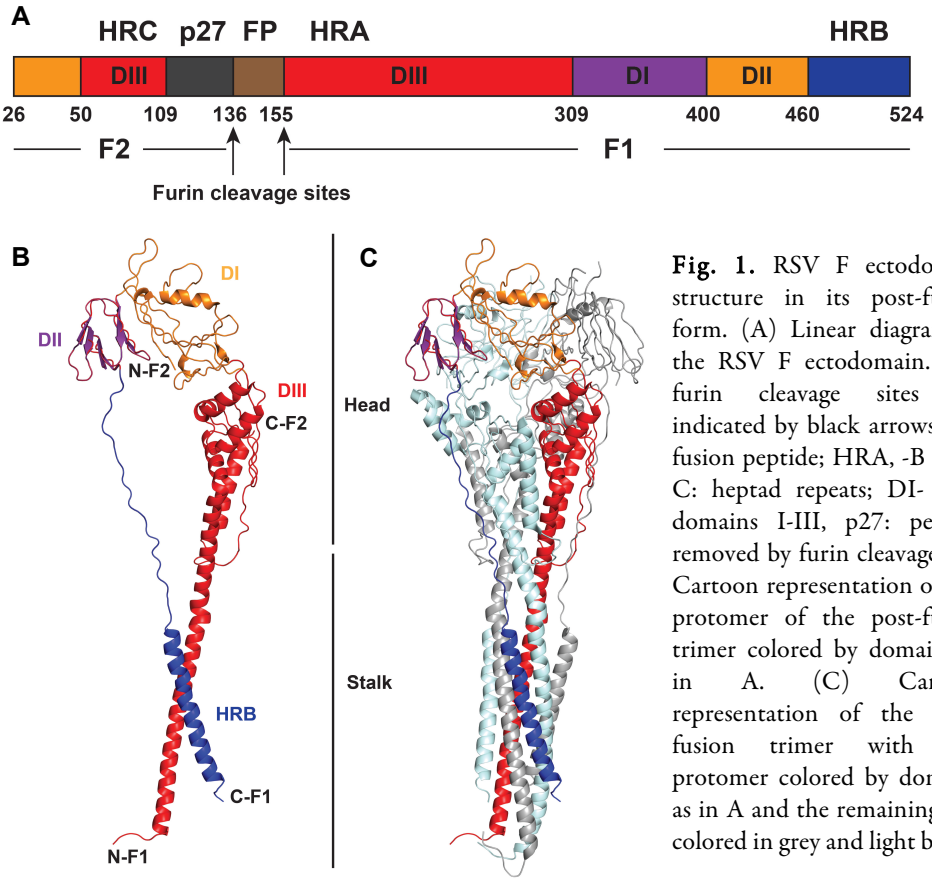


Fig. 1. RSV F ectodomain structure in its post-fusion form. (A) Linear diagram of the RSV F ectodomain. The furin cleavage sites are indicated by black arrows. FP: fusion peptide; HRA, -B and -C: heptad repeats; DI- DIII domains I-III, p27: peptide removed by furin cleavage. (B) Cartoon representation of one protomer of the post-fusion trimer colored by domains as in A. (C) Cartoon representation of the post-fusion trimer with one protomer colored by domains as in A and the remaining two colored in grey and light blue.

The corresponding putative HRs (HR1 and HR2) are also found in SeMNPV F protein. They are predicted to encompass amino acid residues 174-202 and 521-549 respectively. The three-dimensional structure of SeMNPV F is predicted to be similar to that of RSV F (262).

Receptor-binding function of baculovirus F

The insect cellular receptor used by baculoviruses has not yet been identified. Some studies indicate that GP64 and F proteins of baculoviruses interact with distinct insect cell receptors (130, 409, 416). The prototypic member of group II alphabaculoviruses *Autographa californica* (AcMNPV) containing GP64 is able to transduce a number of mammalian cell types through the endocytosis pathway, indicating that the cells carry the receptor recognized by GP64 at their surface. GP64, however, is also able to enter mammalian cells via direct fusion with the plasma membrane under low pH (82). The *gp64*-null AcMNPV pseudotyped with baculovirus F is unable to enter mammalian cells suggesting that the F protein receptor is not present on mammalian cells (432).

Relationship between insect retroviruses and baculoviruses

Retroelements with long-terminal repeats (LTRs) are found in the majority of eukaryotic genomes. The genomes of all cells contain a number of transposable elements integrated into their genomes. Exogenous retroviruses, endogenous retroviruses (ERVs), and LTR-retrotransposons are one of the sources of retroelements. These transposable elements have been described as being able to cross species barriers by horizontal transfer (158). The insect retroelements encoding an envelope (*env*) gene have been classified into the *Errantivirus* genus of the *Metaviridae* family and their *env* gene has been shown to share common ancestry with the gene encoding baculovirus F (232). The sequence similarity is the highest in the region that includes the furin cleavage signal and a predicted fusion peptide.

The baculovirus origin of *env* genes found in insect errantiviruses has been proposed based on the fact that errantiviruses and baculoviruses have the same hosts and LTR- retrotransposons have been previously found incorporated into baculovirus genomes (96). As a result, *env* genes could have been acquired via intragenomic recombination events that occurred after integration of LTR-retrotransposons into the genome of baculoviruses (232, 295). The *Drosophila melanogaster* genome encodes a number of errantiviruses with *gypsy* being the most studied retrovirus-like element in this organism. In contrast to retroviruses, errantiviruses are considered to be non-infectious. However, *gypsy* Env has been

demonstrated to localize to the cell membrane of insect cells and to possess fusogenic properties (173, 261, 345).

Cellular orthologs of baculovirus F protein

In addition to insect retroviruses, the F protein gene was identified in the genomes of four dipteran species: *Anopheles gambiae,* and the fruit flies *Drosophila melanogaster*, *Drosophila yakuba*, and *Drosophila pseudoobscura* (222). Phylogenetic studies have suggested that these F-like genes (named *Iris*) were incorporated into the *Drosophila* genome 25 million years ago from endogenous retroviruses. However, the Iris found in *Drosophila* species *(D.melanogaster, D. yakuba and D. pseudoobscura)* lack the structural elements of a fusion protein such as a predicted furin cleavage site, a predicted fusion peptide, and a coiled-coil domain. Moreover, a study that analyzed a possible membrane fusion activity of *D.melanogaster* Iris did not detect this protein as capable to mediate membrane fusion and showed that it localized to mitochondria and not to the cellular membrane as baculovirus F or gypsy Env (222). The *Anopheles gambiae* F protein has a potential furin cleavage site but the cleavage at this site has not yet been demonstrated.

The ectodomains of lepidopteran baculovirus F proteins contain 10 conserved cysteine residues, but only 6 of them (C5-C10) are found in Drosophila Iris and the F protein from mosquito baculovirus CuniNPV. In *Anopheles gambiae* mosquitoes, C9 is also missing. The spacing between those cysteine residues in lepidopteran baculovirus F proteins and insect cellular F protein is conserved. In addition, 6 highly conserved non-cysteine residues were identified between the cellular F and lepidopteran baculovirus F. Those conserved residues are located in the central and C-terminal portion of F and reside within the borders of so-called domain of unknown function (DUF3609) (conserved domain accession: pfam12259). This domain of ~360 amino acids in length has been previously recognized in eukaryotes and in viruses.

Although it has been proposed that *iris* was incorporated into the *Drosophila* genome from endogenous viruses, the opposite hypothesis should also be considered. It could be that the *f* gene in baculoviruses was acquired directly or indirectly from their insect hosts and evolved to acquire fusion activity. Although baculoviruses infecting the *Drosophila* and *Anopheles* species are not known, baculoviruses infecting other dipterans have been reported (21, 89). It could be that ancestral baculoviruses existed only as occlusion derived virions and were able to replicate only in the epithelial cells of the insect midgut (which is still the

case for sawfly baculoviruses). By acquiring a cellular *f* gene they were able to infect the insect hemocoel, which resulted in the evolution of a BV phenotype.

The indirect route of acquiring the *f* gene in baculoviruses might involve insect endogenous viruses. As mentioned earlier, insect endogenous viruses encode *f-like* genes that could potentially be of cellular origin. Combined with the fact that they can insert into the baculovirus genome, insect endogenous viruses may represent a source for the indirect transfer of the *f* gene (94, 232, 257).

Regardless, *f* gene homologues are very likely to be present in many insects including not only the four above-mentioned dipteran insects but also lepidopteran insects. Given the fact that the similarity of the F proteins of baculoviruses is rather low (less than 20% amino acid identity in some cases) it is possible that cellular F proteins are also divergent and still remain to be identified in insect genomes.

Structural characterization of a baculovirus fusion protein ectodomain

Amino acid sequence analyses have indicated that the F protein of baculoviruses displays a class I viral fusion protein fold and that it is related to the paramyxovirus fusion protein F (Chapter 2). Therefore, it is interesting to understand the organization of F protein counterparts from DNA viruses such as the baculoviruses, in order to provide insight into their evolution. Paramyxovirus F protein appears related to the spike protein of the coronaviruses with a large intervening domain between N-terminal and C-terminal HRs (363). These proteins are thus more distant from, and perhaps not true structural homologs to, other class I fusion proteins characterized to date such as those from retro-, filo-, arena- and influenza viruses. Importantly, there are no DNA viruses known to encode a class I fusion protein, and a crystal structure of SeMNPV F would provide important insight into evolutionary aspects relating class I viral fusion proteins from RNA and DNA viruses. Thus, we aim to obtain structural insight into the fusion process of the group II alphabaculovirus by determining the structure of the soluble F ectodomain of SeMNPV using X-ray crystallography.

MATERIAL AND METHODS

Constructs used or generated in the study

The gene encoding a soluble ectodomain of baculovirus F protein (SeFe)$^{\#}$ encoding amino acid residues 18-553 of the full-length SeMNPV F protein (UniProtKB accession number Q9J8C6) and with mutation at furin cleavage site

(RRSKR→SSSKK) was inserted into pT350 (described in Supplementary materials and methods). The residues 1-17 were excluded from the construct because they comprise the signal peptide. This construct additionally contained a triple Strep-tag. The stable cell line was generated by Scott Jeffers, former member of the Rey's laboratory.

SeFeΔ1 containing the deletion of the fusion peptide residues 150-GLFNFMGHV-158, SeFeΔ2 containing the deletion of the fusion peptide residues 150-GLFNFMGHVDKYLF-163 and SeFe-mut containing the hydrophobic residues within the fusion peptide replaced by the hydrophilic ones (150-GLFNFMGHV-158 and 150-GQTNSHGHN-158) were generated by site-directed ligase independent mutagenesis (SLIM) (63). In SeFeΔ1 and SeFeΔ2 constructs a short GGS linker was introduced at the same time instead of the deleted fusion peptide sequences.

Limited proteolysis

Purified SeFe protein in HNE buffer (0.15 NaCl, 0.1 M EDTA, 5 mM Hepes pH7) at 1 mg/ml was incubated with the indicated concentrations of TPCK (L-1- tosylamide-2-phenylmethyl chloromethyl ketone)-treated trypsin (Sigma) at a trypsin:SeFe ratio of 1:800 (w/w) for 30 min at 23°C Digestion was terminated by addition of PMSF to the final concentration of 0.2 mM (Sigma). The cleaved protein was subsequently analyzed by SDS-PAGE stained with Coomassie Brilliant Blue. Target protein bands were analyzed by N-terminal sequencing.

Acid-treatment of SeFe samples

Different variants of purified SeF ectodomain protein as well as the (SeFe, SeFe trypsin truncated fragment, SeFe-mut, SeFeΔ1 and SeFeΔ2) were exposed to low pH by adding NaAc pH5 to a final concentration of 30 mM and incubated over night at 4°C.

Multiangle light-scattering analysis

Multi-Angle Light Scattering (MALLS) is an analytical technique for determining absolute molar masses and the average size of particles in solution of all types of macromolecules including proteins by calculating the amount of scattered light at different angles. The sensitivity of the light scattering detector increases with molar mass of the analyte. Thus it is an excellent tool for detecting the oligomeric state of proteins. The higher the aggregation number, the more sensitive the detector becomes. MALLS detector is often coupled downstream to a chromatographic system allowing using SEC, which together provides means

for measuring the molar mass, size, and distribution (http://www.wyatt.eu/index.php?id=multi-angle-light-scattering).

The absolute molecular masses of different SeFe samples were determined by gel filtration combined with detection using MALLS and refractrometry (424). Purified protein (90 μg) was loaded onto a Superdex 200 10/300 column (GE Healthcare) connected to an MALLS instrument and an interferometric refractometer (DAWN HELEOS II, Wyatt Technologies, Santa Barbara, CA). The column was equilibrated either at pH5 (10 mM NaAc pH5, 150 mM NaCl, 0.1 mM EDTA) or pH7 (5 mM HEPES pH7, 150 mM NaCl, 0.1 mM EDTA), respectively. The absolute molecular masses were calculated using the ASTRA software (Wyatt Technology Corp., Santa Barbara, CA).

Deglycosylation of trimeric trypsin-resistant SeFe

Production of endodeglycosidases PNGase, EndoH and EndoD is described in Supplementary materials and methods. For crystallization trials, trimeric trypsin-resistant fragment of SeFe (SeFet) was deglycosylated with PNGase F using the ratio 1:16 of endoglycosidase: SeFet on a weight basis. For deglycosylation of SeFet with EndoD-EndoH, the ratio 1:4:28 (SeFet:EndoD:EndoH) on a weight basis was used. Deglycosylation reactions were carried out at 37°C. Deglycosylated SeFet was separated from PNGase by SEC on a Superdex 200 column (GE Healthcare). EndoD and EndoH were removed from the reaction mixture by Ni^{2+} ion affinity chromatography.

Removal of the Strep affinity tag

A C-terminal triple Strep tag preceded by an enterokinase recognition site was removed from the monomeric SeFe-mut, SeFeΔ1 and SeFeΔ2 prior trimerization at low pH by specific proteolytic cleavage with EKMax Enterokinase (Invitrogen, San Diego, USA). The detailed protocol is provided in Supplementary Materials and Methods.

Evaluation of accessibility of free cysteine residues in SeFet

The accessibility of the cysteine residues in SeFet was evaluated by DTNB- thiols assay. The detailed protocol is provided in Supplementary materials and methods.

Crystallization

Crystallization screening, crystal optimization and crystal cryo-protection techniques are described in detail in Supplementary materials and methods. The best-diffracting crystals of SeFet deglycosylated with PNGase (P1 crystal form) were grown at 2 mg/ml SeFet in 2.5 μl hanging drops (1:1:0.5 protein:reservoir solution:seed stock volume ratio) by vapor diffusion against a reservoir solution

containing 16-18% PEG 3350 and 200 mM ammonium chloride. The crystals of deglycosylated SeFet grown in 9% PEG 4000, 100 mM imidazole pH 8 and 30% 2-methyl-2,4-pentanediol (MPD) were used as a seed stock for microseeds. The P43212 crystal form grew under the same conditions when instead of the seed stock just a solution of 9% PEG 4000, 100 mM imidazole pH 8 and 30% 2-methyl-2,4- pentanediol (MPD) was used.

Dehydration of the crystals

The dehydration experiments were performed on SeFet crystals to improve the diffraction resolution of those crystals. Dehydration removes excess solvent from the crystals, which may result in tighter packing of protein molecules and a subsequent increase of the X-ray diffraction of the crystals. Two techniques were applied for the dehydration of SeFet crystals: 1) the cover slip with a hanging drop containing the crystals was transferred over a reservoir solution with a higher percentage of precipitant and allowed to equilibrate for 3 days, and 2) the crystals were directly transferred into a dehydrating solution consisting of the mother liquor with a higher percentage of precipitant, and dehydrated over a reservoir solution containing the same dehydrating solution for 3 days.

Soaking crystals in heavy atom solutions

In order to obtain the heavy atom derivative of SeFet crystals, the crystals were soaked in a number of different heavy atom compound solutions listed in Table 3 in the Results chapter. All compounds except hexatantalum tetradecabromide ($Ta_6Br_{12}^{2+}$) were available from Heavy atom screens (Hampton Research). $Ta_6Br_{12}^{2+}$ was kindly provided by Gérard Bricogne (Global Phasing Limited, Cambridge, UK). $Ta_6Br_{12}^{2+}$ has been reported in the literature to be a powerful derivatization agent.

For soaking experiments, 100 mM stock solutions of different heavy atom compounds were prepared in water. The heavy atom soak solution for soaking SeFet crystals was composed of mother liquor containing 20 % (v/v) glycerol and a heavy compound at a specific concentration. Various concentrations of the heavy atom compound (0.5-20 mM) as well as various soaking times (from 10 min to 18 h) were tried. 2 μl of the heavy atom soak solution was pipetted onto a new siliconized cover slip and the crystals were transferred from the mother liquor to the drops of the heavy atom soak solution. The cover slip with a drop was placed above a well containing mother liquor supplemented with 20% (v/v) glycerol.

SeFet crystals derivatized with $Ta_6Br_{12}^{+}$ were obtained by soaking crystals in a soak solution containing 1 mM $Ta_6Br_{12}^{+}$ for 18 h. The crystals were back-soaked by transferring the crystals into mother liquor containing 20% glycerol (v/v) and flash-frozen in liquid nitrogen.

Production of selenomethionine substituted SeFe

A selenomethionine (SeMet) substituted SeFe was produced in ESF-921 serum-free medium methionine-free medium (Expression Systems) supplemented with L-SeMet. Initially, 3 l of the S2-SeFe-expressing cells were grown in Insect Express medium until it reached the density of ~20×106/ml. The cells were collected by centrifugation for 5 min at 200 g and resuspended in 1.5 l of ESF-921 serum-free methionine-free medium supplemented with 0.8 g/l L-cysteine. After 4 h of starving, 300 mg/l L-SeMet (Acros Organics, Geel, Belgium) and 4 μM $CdCl_2$ were added. An extra 300 mg/l L-SeMet were added at day 3. The supernatant was harvested 5 days after induction. The substituted protein was purified in the same way as the native protein but the yields were 10-fold lower.

Structure determination of SeFet

Native as well as MAD (Multi-wavelength anomalous dispersion) and SAD (Single-wavelength anomalous diffraction) data sets were collected at the Synchrotron Soleil beamline Proxima 1, the Swiss Light source beamline PX I or European Synchrotron Radiation Facility (ESRF) beamlines ID23-1, ID14-4 and ID23-2 at 100 K. Typically, diffraction data were collected using Pilatus (Dectris, Baden, Switzerland) or charge-coupled device (CCD)-based detectors. Programs iMosflm (200, 307) or xdsme were used to determine the optimum orientation of the crystal for the complete data set collection. The best native data sets of X-ray diffraction data for P43212 and P1 crystal forms (2.9 Å and 2.7 Å resolution, respectively) were processed using XDS (160).

The initial experimental phases were obtained by the MAD method using the $Ta_6Br_{12}^{2+}$ derivative of P1 crystal forms. Data for this heavy-atom derivative were collected on a single crystal using an inverse beam data collection strategy with wedges of 10 degrees while cycling through three wavelengths (peak, remote inflection). This strategy was used in order to optimize the measurement of Friedel pairs. In addition, due to crystal sensitivity to radiation damage, the collection of MAD data sets was performed with reduced beam intensity. The data sets obtained at each wavelength were processed using XDS and scaled together using XSCALE (160). SHELX C/D within SHARP (42) was used to locate the heavy atom sites (327). Experimental phasing was performed with SHARP followed by solvent flattening, which provided initial experimental

phases. The obtained map was used to build an initial model in polyalanines. NCS averaging was applied to further improve the map. Better phases were obtained with a MAD experiment using the $Ta_6Br_{12}^{2+}$ derivative of the P43212 crystal form. This data set was collected using an inverse beam data collection strategy with wedges of 10 degrees while cycling through the two wavelengths of falling and mounting inflection (1.255070 (9879 eV), 1.254140 (9886eV). In the case of the P43212 crystal form $Ta_6Br_{12}^{2+}$ derivative, individual Ta atoms were identified using SHARP.

The phases in P43212 crystal form were combined with the phases of the initial model built in P1 crystal form. After applying density modification and multi-crystal averaging techniques, better quality maps were obtained for both crystal forms allowing the complete building of the model. NCS averaging and multi-crystal averaging was performed using the DM (CCP4 suite), and phenix.multi_crystal_average (372). The building of the model was performed manually in Coot (88) alternating between the electron density maps in different crystal forms. Manual building was supported by anomalous data for a SeMet derivative from the P43212 crystal form collected at the peak wavelength as well as the highly redundant Sulfur-SAD data set at a wavelength of 1.7995 Å from the native crystal in P1 form. The density for the anomalous scatterers (selenium and sulphur) was calculated using AnoDe (374). Refinement was performed using AutoBuster (42) against both the P1 and the P4 data sets imposing NCS restraints and TLS groups. Refinement was monitored following the Rfree Rwork and Rfree values. Rfree values were calculated for a random subset (5 %) of reflections omitted from refinement. Water molecules were added manually using Coot. Throughout the refinement, a structure-validation web service MolProbity (72) was used to monitor all-atom contact analysis as well as Ramachandran and rotamer distributions.

Crystal structure analysis

Electrostatic potentials were calculated using the adaptive Poisson Boltzmann solver (15). The protein interaction calculator (375) was used to identify the interactions between the protomers of SeFet. The surface area buried upon trimer formation was estimated using the PISA server (182). The topology diagram of SeFet protomer was generated using program PDBsum (192). SeFet structure comparison with other protein structures available in the PDB was carried out using DALI server (140). Sequence alignments were performed using MultAlin (http://multalin.toulouse.inra.fr/) (67) and formatted with ESPript 3 (http://espript.ibcp.fr/) (115). Sequence conservation among SeFe and F

proteins from different baculoviruses or different putative cellular F orthologs was analyzed by the program ConSurf (111). Figures of the crystal structures were prepared in the PyMOL Molecular Graphics System (www.pymol.org).

RESULTS

Production and purification of SeFe

We established an efficient expression system to obtain the soluble ectodomain of the SeMNPV F protein, called SeFe in *Drosophila S2* cells. The construct containing the SeFe encompasses residues A18-T553 (aa 1-17 comprise a signal peptide) of the full-length SeMNPV F protein (UniProtKB accession number Q9J8C6). It lacks the predicted transmembrane (TM) anchor domain (residues 580-602) and the C-terminal cytoplasmic tail (CT) domain (residues 603-665) present in the full-length protein (Fig. 2) in order to allow secretion from cells. Our first approach was to crystallize SeFe in its pre-fusion form. To stabilize the pre-fusion form, the furin cleavage site was mutated (RRSKR to SGSKK) to prevent proteolytic processing into F1 and F2 subunits by furin in the transfected S2 cells.

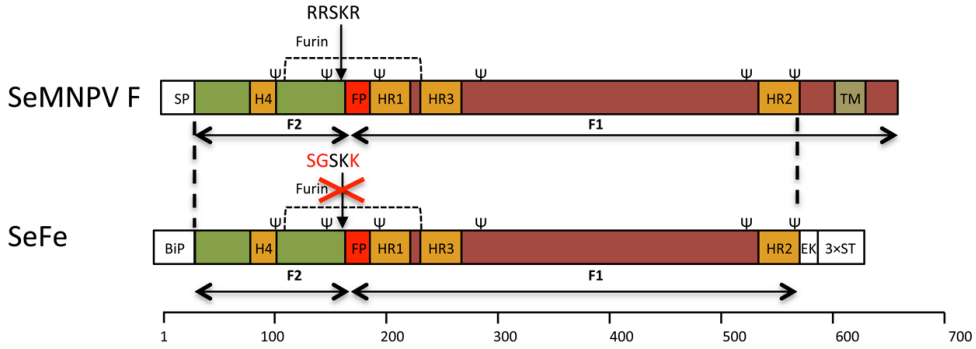


Fig. 2. Primary structure of SeMNPV F and the expression system of its ectodomain (SeFe) in *Drosophila melanogaster* S2 cells. The domains indicated are F2 and F1, the fusion peptide (FP), the signal peptide (SP), three heptad repeat regions (HR1, HR2, HR3 and H4), the transmembrane domain (TM), enterokinase cleavage site (EK), triple strep-tag (3×ST) and *Drosophila* secretion signal (BiP). Predicted *N*-glycosylation sites are marked by (Ψ), furin cleavage site in SEMNPV F (RRSKR) and the mutated furin cleavage site in SeFe (SGSKK) are indicated by black arrows. The disulfide bridge connecting domains F2 and F1 is shown as a thin dashed line. The thick dashed line indicates the borders of the F ectodomain cloned into the expression vector. The scale below corresponds to amino acid numbering of SeMNPV F (UniProtKB accession number Q9J8C6).

The large-scale expression and purification of SeFe was performed using standard procedures (see section 2). Isolation of pure SeFe employed a combination of Streptactin affinity column (Fig. 3A) and size exclusion chromatography using a

HiLoad 26/60 Superdex 200 gel filtration column (GE Healthcare) (Fig. 3B). The protein eluted from the gel filtration column as a single peak corresponding to monomeric SeFe (Fig. 3B). The final yield of SeFe was ~10 mg /l of culture supernatant.

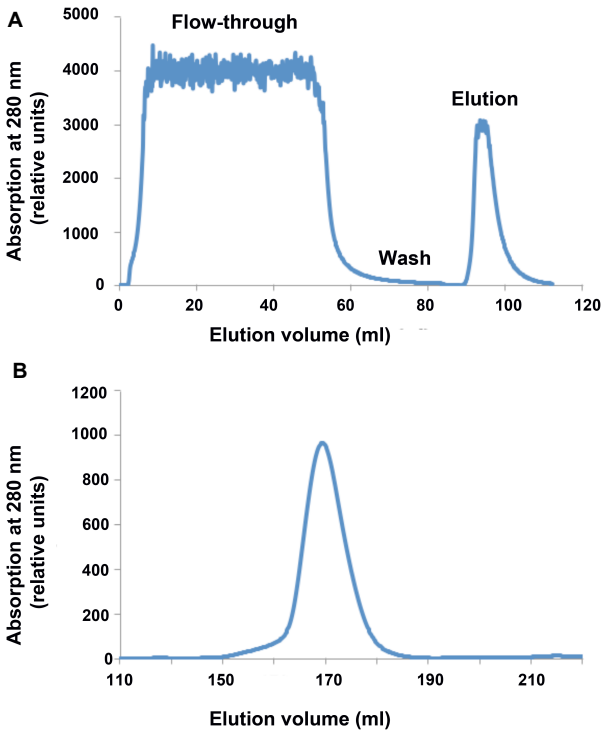


Fig. 3. Purification of SeFe. The concentrated supernatant was loaded on 8 ml Streptactin column (A). After a washing step, the Strep-tagged protein was eluted with 2.5 mM desthiobiotin. The fractions of the eluent were pooled and subjected to size exclusion chromatography (SEC) (B). Separation by SEC was performed using HiLoad Superdex 200 26/60 column (GE Healthcare) at a flow speed of 2 ml/min in 5 mM Hepes pH 7.0 150 mM NaCl and 0.1 mM EDTA. Protein elution was monitored by absorbance at 280 nm. Chromatogram B reveals a single major peak for SeFe corresponding to the monomeric protein.

Initially the monomeric SeFe was crystallized at neutral pH and crystals were obtained (most likely corresponding to the pre-fusion form of SeFe) but the crystals diffracted only to ~8 Å, which was not sufficient to determine the crystal structure. The crystallized protein was monomeric as judged by size exclusion chromatography, suggesting that it may represent a pre-fusion conformation. Crystallizing viral fusion proteins in their pre-fusion form is usually challenging because these forms are metastable. Based on the fact that the paramyxovirus F protein is a stable homotrimer in its post-fusion conformation (366), crystallization of the SeFe in its post-fusion form was performed as an alternative strategy.

Trypsin proteolysis of SeFe yields a trypsin-resistant fragment

The main problem in crystallizing the post-fusion conformation of viral fusion proteins is that the fusion peptide is exposed, usually resulting in aggregation of

the protein. In chapter 2 the SeFe was digested with various concentrations of trypsin and based on the results thereof the trypsin concentration necessary to obtain a trypsin-resistant fragment of SeFe was determined. Furthermore, it was also demonstrated that lowering the pH from pH7 to pH5 resulted in an altered oligomeric state of the trypsin-resistant fragment, suggesting the formation of a post-fusion trimer. N-terminal sequencing results, together with SDS-PAGE and Western blotting analysis of the trypsin-resistant SeFe trimer, revealed that:

1. Trypsin cleavage occurs in the region between the predicted HR1 and HR3 regions at residues R210 and R212 (210-RMRD-213) (based on N-terminal sequencing).

2. Trypsin also cleaves at the mutated furin cleavage site at residues K147 or K148 (SGSKK) (based on SDS-PAGE).

3. The Strep-tag can no longer be detected by anti-strep antibodies in the Western blot and is thus completely removed by trypsin cleavage.

4. The trypsin cleavage at the C-terminal end of the F1 subunit most likely occurs at one of the trypsin target sites located at either end of the HR2 domain.

Although trypsin cleavage results in the removal of the internal protein part composed of the fusion peptide and the HR1 domain, F1 and F2 should still remain linked by the disulphide bridge (Fig. 4).

The trypsin cleavage of SeFe shown in Chapter 2 was reproduced. The limited trypsin proteolysis with increasing concentrations of trypsin led to the formation of a predominant product, which was observed as a single band of ~50 kD in SDS-PAGE under non-reducing conditions (Fig. 5).

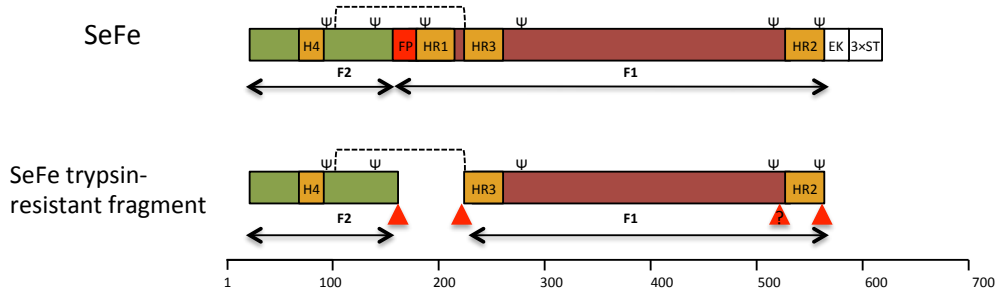


Fig. 4. Schematic representation of SeFe and its trypsin-resistant fragment. The trypsin cleavage sites are indicated with red arrowheads. The domains indicated are F2 and F1, the fusion peptide (FP), four heptad repeat regions (HR1, HR2, HR3 and HR4), enterokinase cleavage site (EK), and triple strep-tag (3×ST). Predicted *N*- glycosylation sites are marked by (Ψ). The disulfide bridge connecting domains F2 and F1 is shown as a thin dashed line.

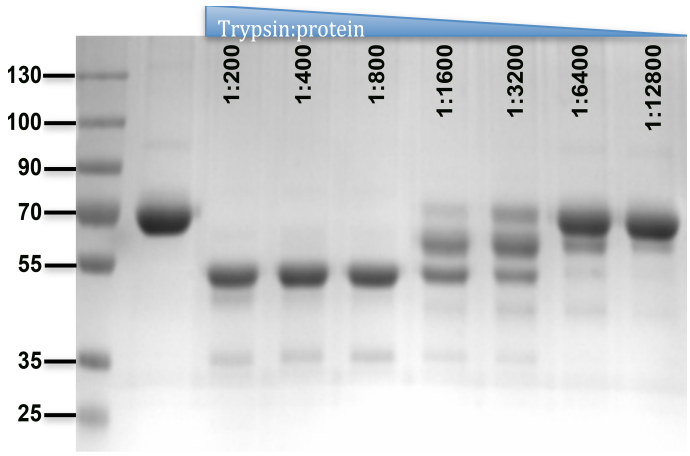


Fig. 5. Limited proteolysis of SeFe by trypsin. SeFe was mixed with trypsin at different w/w ratios, incubated at 22°C for 30 min and analyzed by SDS-PAGE under non-reducing conditions. The gel was stained with Coomassie blue. A predominant product of ~50 kD was observed by SDS-PAGE when using trypsin:SeFe ratio from 1:800 to 1:200. Lane 1: protein ladder.

On the intact virions, furin-cleaved baculovirus F proteins mediate membrane fusion upon exposure to acidic pH encountered during endocytosis (150). Therefore, we investigated the effect of acid treatment of the SeFe and the trypsin-resistant SeFe fragment. Both proteins were exposed to pH5.5 or pH7 and subsequently analyzed by SEC and multi- angle laser light scattering (MALLS). The trypsin resistant SeFe fragment exposed to pH5.5 eluted earlier from the SEC column than the SeFe trypsin resistant fragment at pH7, suggesting a change in its oligomeric state at acidic pH (Fig. 6).

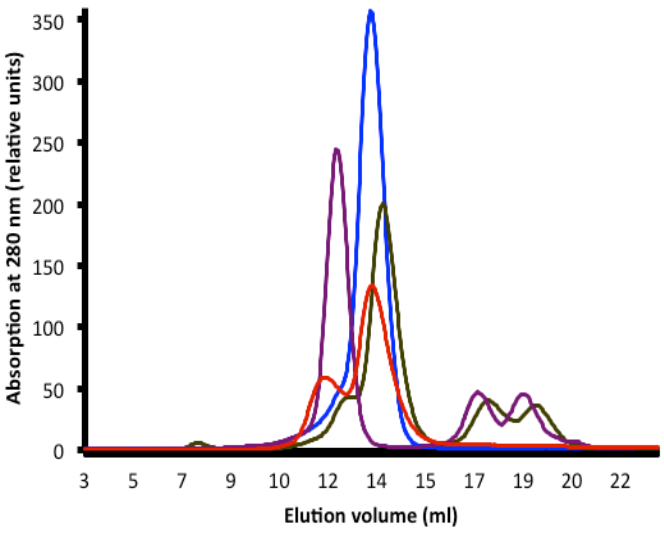


Fig. 6. Effect of acid treatment on SeFe and SeFe trypsin resistant fragment. Non-digested or trypsin digested SeFe was exposed to neutral (pH7) or acidic (pH5.5) pH and analyzed by SEC using Superdex 200 10/300 column (GE Healthcare). SeFe trypsin resistant fragment at pH5.5 eluted earlier from the SEC column than the SeFe trypsin resistant fragment at pH7, suggesting that there was a change in oligomeric state at acidic pH.

The analysis of the oligomeric state of the SeFe and SeFe trypsin resistant fragment by MALLS confirmed that the initially crystallized protein (SeFe, pH7) was a monomer while the trypsin resistant fragment forms a trimer at acidic pH (Table 1). The acid-induced trimerization of the trypsin resistant SeFe fragment was irreversible since the titration of pH back to neutral prior to MALLS analysis did not lead to trimer dissociation.

Table 1. MALLS analysis of the oligomeric state of SeFe and of the trypsin resistant SeFe fragment at neutral (pH7) and acidic (pH5.5) pH.

	Molecular weight [kD]	Mass fraction (%)	Oligomeric
- Trypsin pH7	95.4 (±5.5%)	100	Monomer
- Trypsin pH5	237.4 (±0.1%) and 91.3 (±0.2%)	36.4 and 63.6	Trimer and
+ Trypsin pH7	77.5(±0.2%)	100	Monomer
+ Trypsin pH5	188.0 (±1.9%)	100	Trimer

This protease resistant trimer was used for crystallization as it most likely represents the stable post-fusion conformation. A large-scale preparation of the SeFet was produced and purified by SEC on a Superdex 200 26/60 column, from which the protein eluted in a single symmetric peak corresponding to a SeFet trimer, as expected (Fig. 7)

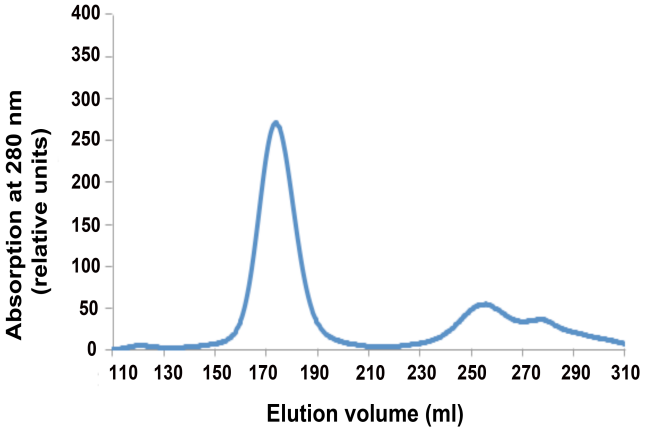


Fig. 7. Large scale SeFet formation. The trypsin proteolysis reaction of SeFe and acid induced trimerization were scaled up in a linear manner. SeFet was purified from the reaction mixture by SEC on Superdex 200 26/60 column (GE Healthcare). SeFet eluted from SEC column as a single symmetric peak corresponding to the trimeric SeFe trypsin resistant fragment.

Table 2. The crystals of SeFet subjected to X-ray diffraction experiments. The ability of the crystals to diffract X-rays were tested at Proxima I line, Synchrotron Soleil, France or ID 23 line European Synchrotron Radiation Facility (ESRF), France.

Crystallization condition	Concentration (mg/ml)	Resolution (Å)
18% PEG 6000, 100 mM Tris pH 8	6	6.5
16% PEG 4000, 100 mM Tris pH 7.5, 15% 2-propanol	6	12
18% PEG 6000, 100 mM Tris pH 8	6	8
14% PEG 8000, 100 mM Tris pH 7.5	6	7
23% PEG 6000, 100 mM Tris pH 7.5	6	9
12% PEG 6000, 10 mM NaAc, 40% ethanol	6	13
20% PEG 6000, 100 mM imidazole pH 8	6	10
26.5 % PEG 8000, 100 mM Hepes pH 7.5, 10% 2-propanol, 100 mM NaAc	6	20
20.5% PEG 10000, 100 mM Tris pH7.5	6	10
19% PEG 4000, 100 mM NaAc, 100 mM Hepes pH7.5	6	20
15% PEG 8000, 40 mM potassium phosphate monobasic	6	>30
16% PEG 6000, 100 mM NaCl, 100 mM Tris pH 7.5	6	25
19% PEG 6000, 200 mM NaCl, 100 mM Tris pH 8	6	9
19% PEG 6000, 100 mM NaCl, 100 mM Tris pH 8	6	10
16% PEG 8000, 170 mM NaAc, 90 mM sodium cacodylate pH 6.5	6	15
16% PEG 8000, 170 mM NaAc	6	9
14% PEG 8000, 100 mM Tris pH 7.5	6	
17.4% PEG 3350, 200 mM NaCl, 200 mM $MgCl_2$	6	>30

Crystallization of SeFet

The crystallization screening for SeFet was performed as described in Materials and Methods at protein concentration 6 and 10 mg/ml. The protein crystallized in a number of conditions that were further optimized in 24-well plates. The crystals that were subjected to X-ray diffraction analysis are listed in Table 2. Most of the crystals diffracted to 10-12 Å or lower resolution, with the exception of one crystal form that diffracted to 6.5 Å, which was sufficient for determination of the space group (P212121) and unit cell parameters (a=109.68, b=346.24, c=111.34). In addition, we observed a peak in self-rotation function at 120 degrees indicating that the crystallized molecule is a trimer. According to the Matthew's coefficient, the unit cell probably accommodated three trimers per asymmetric unit. These crystals of SeFet formed after three weeks in 2 µl hanging drops by vapor diffusion against a reservoir solution containing 18% PEG 6000 and 100 mM Tris pH8 (1:1 protein-to-reservoir-solution ratio) (Fig. 8) and reproducibly diffracted to 6.5-7 Å resolution.

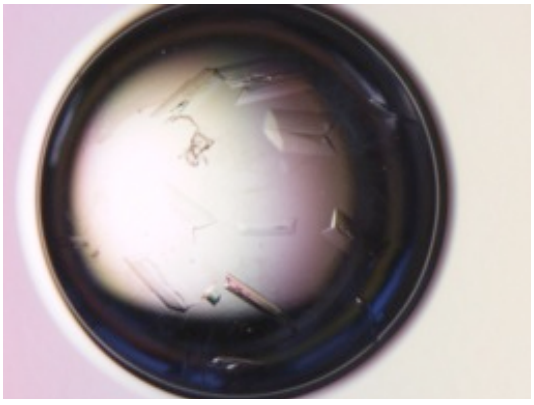


Fig. 8. Best diffracting SeFet crystals. These plate-shaped crystals of SeFet were formed after three weeks in 2µl hanging drops by vapor diffusion against reservoir solution containing 18% PEG 6000 and 100 mM Tris pH8 (1:1 protein-to-reservoir-solution ratio) and reproducibly diffracted to 6.5-7 Å resolution.

In order to obtain better diffracting crystals, we performed a number of optimization experiments (microseeding and streak seeding, crystallization at different protein concentrations, testing crystallization additives, crystallization at 4 ℃, crystallization of the protein after an extra purification step by ion-exchange chromatography and vapor diffusion dehydration). By vapor diffusion dehydration of the crystals for 2 days above the reservoir solution containing 23% PEG 6000, 100 mM Tris pH8, we were able to improve resolution and collect a complete dataset to ~5.5 Å.

Crystallization of the deglycosylated SeFet

SeFe contains six predicted N-linked glycosylation sites (N86, N132, N179, N270, N508 and N551) (Fig. 9). One of the advantages of the expression of the

recombinant proteins for crystallization in S2 cells is that N-linked glycosylation in *Drosophila* is less complex and more homogenous than in mammalian cells. Proteins expressed in S2 cells have high-mannose N-linked glycosylation and are not sialylated. In general, a deglycosylated protein might form a more rigid protein lattice than a protein containing all sugar chains and, thus, yield better diffracting crystals. Although one or two sugars are removed in trypsin resistant fragments because they are present in the parts of the protein cleaved off by trypsin, the rest of the sugars might still hinder the formation of a rigid protein lattice. As a result, we attempted to deglycosylate SeFet with different endoglysosidases for crystallization trials.

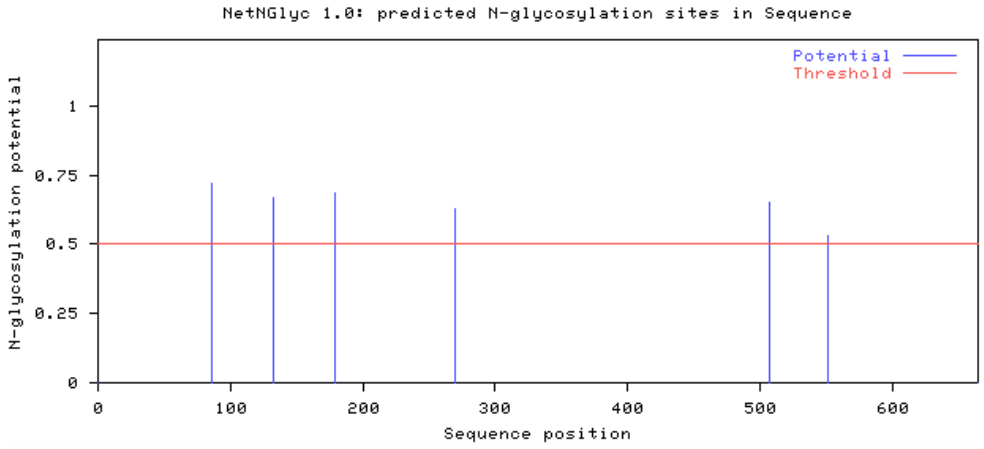


Fig. 9. Putative glycosylation sites in SeMNP F protein predicted by NetN Glyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc). The graph illustrates predicted N-glycosylation sites across the protein chain. Positions with potential (vertical lines) crossing the threshold (horizontal lines at 0.5) are predicted glycosylated.

The deglycosylases PNGase F, EndoH and EndoD were produced using periplasmic expression in *E.coli* and purified as described in Materials and Methods. The extent of deglycosylation of SeFet by different endodeglycosidases was assessed by mobility shift of the deglycosylated protein versus the intact glycoprotein on SDS-PAGE gels. A clear mobility shift of SeFet deglycosylated by PNGase F (Fig. 10) as well as EndoD/EndoH (not shown) was observed. Deglycosylation of SeFet by EndoH alone (not shown) did not have any evident effect on protein mobility on the SDS-PAGE gel, indicating that this endodeglycosidase most likely is not able to remove the sugars from the protein.

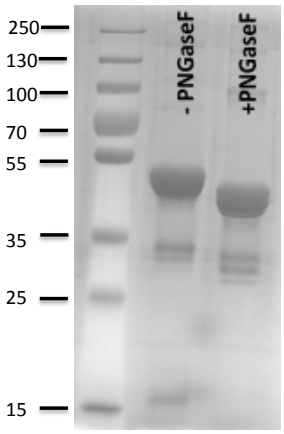


Fig. 10. Enzymatic deglycosylation of SeFet with PNGase F. SeFet was incubated with PNGase F overnight at 37°C using the ratio 1:16 of endoglycosidase:SeFet on a weight basis. A clear mobility shift of SeFet deglycosylated by PNGase F versus non-deglycosylated SeFet was observed. Lane 1: protein ladder, lane 2: non- deglycosylated SeFet, Lane 3: SeFet deglycosylated with PNGase F.

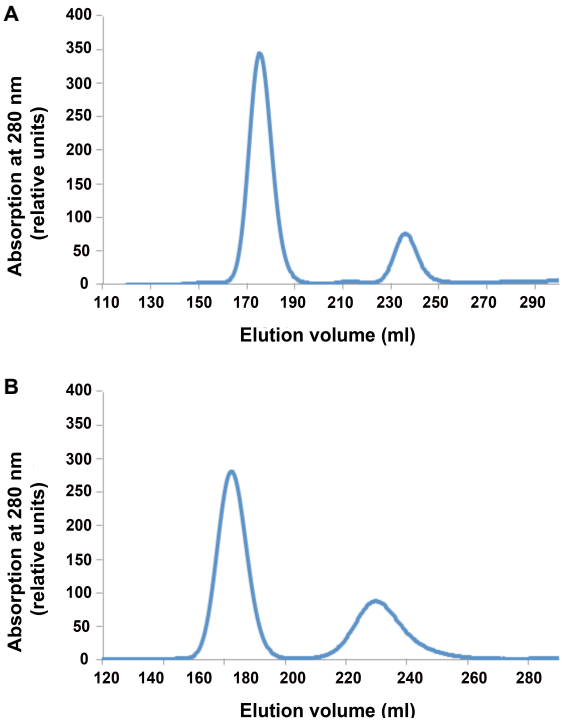


Fig. 11. Purification of deglycosylated SeFet. After deglycosylation with PNGase F, SeFet was separated from the endodeglycosidase by SEC (A). The first peak in the chromatogram corresponds to the deglycosylated SeFet, while the second peak corresponds to PNGase F. After deglycosylation with EndoD/EndoH, SeFet was separated from endodeglycosidases by Ni-affinity chromatography and subsequent SEC (B). The first peak in the chromatogram corresponds to the deglycosylated SeFet, while the second peak corresponds to the fraction of EndoH that was not fully removed by Ni-affinity chromatography.

Subsequently, a large-scale preparation of SeFet was deglycosylated by PNGase F. After deglycosylation, SeFet was separated from PNGase F by SEC (Fig. 11A). In addition, a large-scale preparation of SeFet deglycosylated by EndoD/EndoH was carried out. EndoD/EndoH were removed from the reaction mixture by Ni-affinity chromatography and subsequent SEC (Fig. 11B). The crystallization screening of SeFet deglycosylated with PNGase F and EndoD/EndoH was performed as described in Materials and Methods at protein concentrations of 4

and 6 mg/ml, respectively. Since the endodeglycosidase EndoD was obtained later during the project, the crystallization of SeFet deglycosylated with EndoD/EndoH will be described later.

SeFet deglycosylated with PNGase F crystallized under fewer conditions, which and were different from those of the crystals of non-deglycosylated protein. The crystallization conditions were further optimized in 24-well plates and the best crystals subjected to X-ray diffraction. However, none of the crystals of deglycosylated protein diffracted to higher than 8 Å resolution. In addition, the deglycosylated protein did not crystallize any more in the condition in which the best diffracting crystals of the sugar-containing SeFet were obtained (16% PEG 6000, 100 mM Tris pH8). Therefore, we tried to induce the crystal growth of deglycosylated protein under these conditions by seeding with the crystals of the glycosylated protein as a seed source. Although seeding helped to obtain the crystals of deglycosylated SeFet under the same conditions of the not-deglycosylated protein, the resulting crystals again only diffracted to about 8 Å resolution (Table 3).

Table 3. Crystals of the deglycosylated SeFet subjected to X-ray diffraction at Proxima I line, Synchrotron Soleil, France or PXI line, Swiss Light Source (SLS), Switzerland.

Crystallization condition	Concentration (mg/ml)	Resolution (Å)
21 % PEG 3350	4	30
19 % PEG 3350, 200 mM ammonium formate	3	3.8
19 % PEG 3350, 200 mM sodium formate	3	5
18 % PEG 3350, 200 mM ammonium chloride	2	2.7
18 % PEG 6000, 100 mM Tris pH8	4	8
20 % PEG 2000 mono-methyl polyethylene	4	10
22 % PEG 3350, 200 mM sodium nitrate	4	10
19 % PEG 3350, 200 mM ammonium fluoride	4	6
9% PEG 4000, 100 mM imidazole pH 8 and 30% MPD	4	10

To obtain different crystals of deglycosylated SeFet, we used microseed matrix screening to sample many more crystallization conditions. This is a seeding technique where crystals grown in one condition are ground to use as seeds, and are then distributed by a robotized procedure into hundreds of different crystallization conditions. The crystals of deglycosylated SeFet grown in 9% PEG 4000, 100 mM imidazole pH 8 and 30% 2-methyl- 2,4-pentanediol (MPD) were used as a seed stock for microseed matrix screening. This technique helped to

obtain crystals in different crystallization conditions that were further optimized in 24-well plates and subjected to X-ray diffraction analysis.

For one crystal form (Fig. 12), which grew in 18% PEG 3350, 200 mM ammonium chloride we were able to collect a complete native data set at 2.7 Å (space group P1, Fig. 13). The unit cell parameters (a=73.94, b=74.86, c=78.11, α=94.5, β=114.3, γ=114.2) could accommodate one trimer per asymmetric unit. The Matthews' coefficient (*V*m) was estimated to be 2.04 Å^3/Da (corresponding to a solvent content of 39.69%). The self-rotation function (Fig. 14) clearly showed the existence of a 3-fold non-crystallographic symmetry (NCS) axis. This crystal form was reproducible and grew in ~48 h.

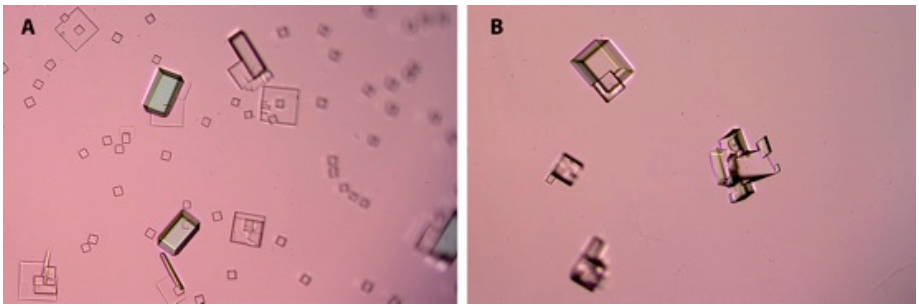


Fig. 12. Photographs of typical SeFet crystals obtained after seeding. (A and B). The crystals were grown at 2 mg/ml SeFet in 2.5 µl hanging drops (1:1:0.5 protein:reservoir solution:seed stock ratio) by vapor diffusion against reservoir solution containing 16-18% PEG 3350 and 200 mM ammonium chloride and diffracted to ~3 Å resolution. The crystallization drops usually used to contain different quality crystals. The crystals had a tendency to grow as clusters of multiple crystals (B).

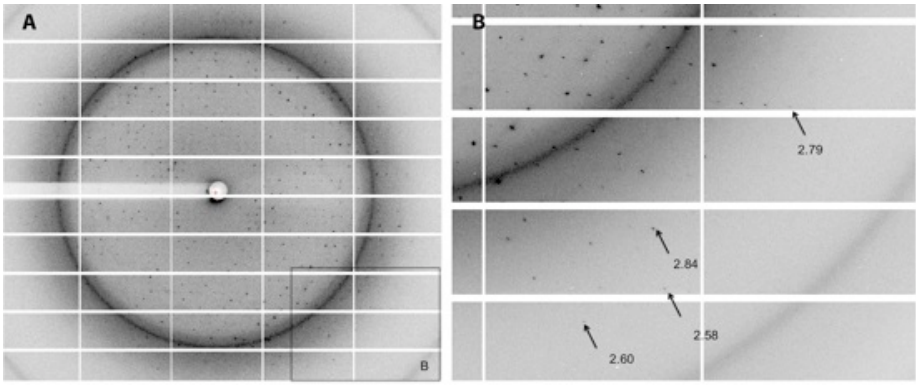


Fig. 13. An X-ray diffraction pattern from the best native deglycosylated SeFet crystal on a Pilatus detector (Dectris, Baden, Switzerland). (A) High resolution reflections of the same diffraction image and corresponding resolution (B). The data were collected on the PXI beam line at the SLS in Switzerland.

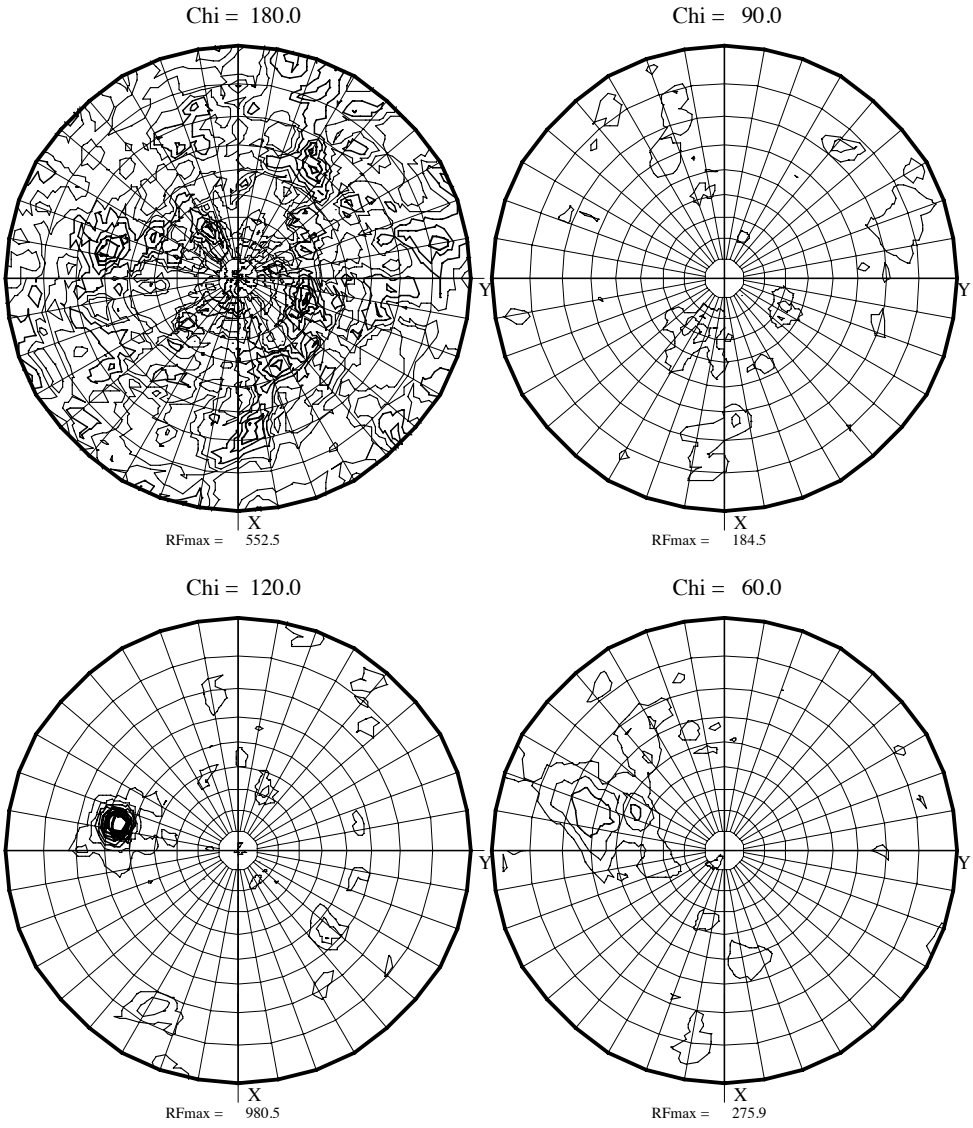


Fig. 14. Self-rotation function of SeFet P1 crystal form as determined from the program MOLREP (380). The peak on the sections (chi = 120) indicates the presence of non-crystallographic 3-fold axis.

The quality and morphology of the crystals were very tightly dependent on the seed stock dilution, protein concentration, and PEG 3350 concentration used.

Native crystals reproducibly diffracted to ~3 Å. All tested cryoprotectants (PEG 400, glycerol, MPD and ethylene glycol) were suitable for cryopreservation of the crystals. Screening for crystallization additives was performed, expecting to further improve the diffraction of the crystals but the identified additives did not enhance the diffraction of the crystals. It was possible to grow crystals using pH gradient from 6-8. However, pH did not have significant influence on crystal morphology or diffraction. Crystals growing at 4°C also did not help to improve the diffraction either.

Phasing of SeFet crystals

In order to obtain phase information, we resorted to experimental phasing given that the paramyxovirus F model was not sufficient to obtain accurate phases by molecular replacement. Several heavy atom derivatives of the crystals were tested by soaking the crystals in various heavy atom solutions. Derivatization depends on parameters such as the exposure of functional groups, local chemical environment, and ionization state. SeFe contains an odd number of cysteines so the first choice was soaking the crystals with different mercury compounds that are known to specifically react with free cysteine thiols. In addition, co-crystallization of SeFet in the presence of the same mercury compounds was attempted. Unfortunately, neither of the two approaches proved to be successful. The availability of free cysteine by a colorimetric assay was tested using 5,5′-Dithiobis (2-nitrobenzoic acid) (DTNB) reagent (described in section 2). A lower signal than expected for three free thiols indicated that the cysteines in the SeFet are difficult to access, which could explain the failure to derivatize SeFet crystals with mercury compounds. In parallel, a number of other heavy atom compounds was evaluated, using at least two different concentrations for screening, and also for testing different soak times. The heavy atom compounds tried in the derivatization of SeFet crystals are listed in Table 4.

Initially, we succeeded in obtaining heavy atom derivatives with potassium tetrachloroplatinate (II) (K_2PtCl_4) and sodium tetrachloraurate (III) ($NaAuCl_4$), but in both cases the derivatives diffracted only to low resolution (~5-6 Å) and the anomalous signal was not strong enough to obtain an initial set of phases. In addition, the derivatized crystals were highly non-isomorphous with the native crystals, excluding the multiple isomorphous replacement method as a possible phasing strategy. Furthermore, the fast decay and low symmetry made it difficult to determine accurate phases experimentally. The data sets were often not complete due to radiation damage, as data collection required a long period of time to obtain enough redundancy in this low symmetry P1 space group (at least

360° oscillation in inversed-beam collection mode). As a result, the low multiplicity of the measured intensities (Friedel pairs are measured only once if collected 360°) resulted in poor measurement precision. We therefore initiated collaboration with Pierre Legrand from the synchrotron Soleil who is very experienced in experimental phasing.

Table 4. The heavy atom compounds tried in the derivatization of SeFet crystals.

No	Heavy atom compound	Anomalous signal
1	Thimerosal	No
2	Ethyl Mercuric Phosphate	No
3	Methylmercury (II) chloride	No
4	Mercury (II) chloride	No
5	Mercury (II) cyanide	No
6	Mercury (II) acetate	No
7	Potassium tetracyanoplatinate (II) hydrate	No
8	Potassium tetrachloroplatinate (II)	Yes
9	Dipotassium hexachlororhenate	Maybe weak
10	Lead(II) acetate trihydrate10	No
11	Sodium tetrachloraurate (III) dihydrate	Yes
12	Gold (I) potassium cyanide	No
13	Potassium tetrachloraurate (III) hydrate	No
14	Neodymium trichloride, hexahydrate	No
15	Sammarium (III) acetate	Maybe weak
16	Hexatantalum tetradecabromide	Yes

First, we optimized the data collection strategy by attenuating the beam to reduce radiation damage, which allowed us to use the Multiple Anomalous Diffraction (MAD) method by collecting data in small wedges in the inverse beam mode while alternating between several wavelengths. We also collected a single-wavelength anomalous diffraction data set at 3.1 Å at the sulphur edge on native crystals, which allowed us to identify the locations of the disulfide bonds. But even after optimization of the data collection strategy, the anomalous signal of the crystals derivatized with K_2PtCl_4 or $NaAuCl_4$ were not strong enough to obtain an initial set of phases good enough to determine the structure.

Next we tried using the hexatantalum tetradecabromide ($Ta_6Br_{12}^{+}$) cluster compound, which is known to be a powerful derivatization reagent for phasing crystals at low resolution. The soaking of SeFet crystals in 1mM $Ta_6Br_{12}^{+}$ over

night led to the incorporation of the compound into the crystals, which was apparent due to the change of the crystal color. We applied the double-inflection MAD collection strategy using one crystal derivatized with $Ta_6Br_{12}^{+}$ and detected significant anomalous differences up to ~6 Å. This data set was used to locate the heavy atom sites in the unit cell of the crystal, and to calculate an initial set of phases, which resulted in an interpretable electron-density map with clearly identifiable helices resembling the six-helix bundle characteristic for class I fusion proteins (Fig. 15).

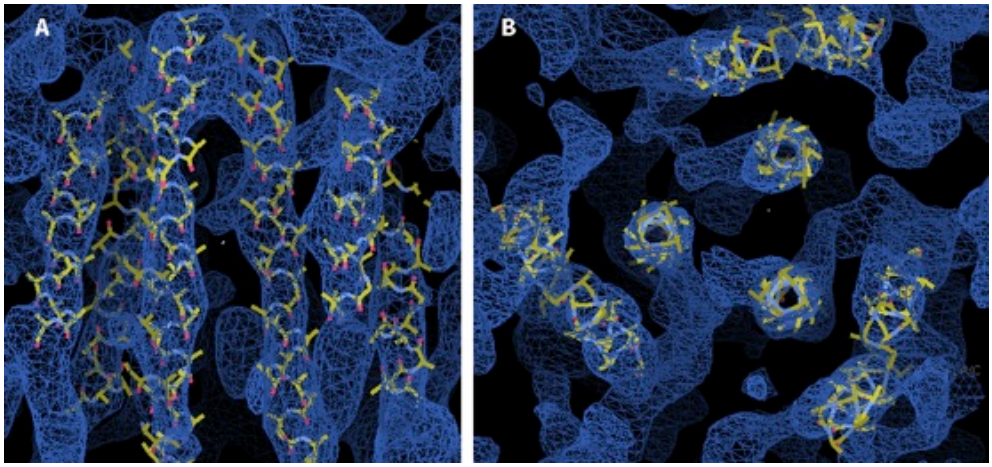


Fig. 15. The electron density map of SeFet with the first helices built in alanine residues. The side view of SeFet helices (A). The view along the trimer axis (B).

We then applied density modification techniques to further improve the map and started manual-building of the initial model as polyalanine chains within this map, along with iterative cycles of electron density modification [solvent flattening, non-crystallographic symmetry (NCS) averaging]. The initial maps indicated that the SeFet molecule is a very flexible trimer, with the angles between three protomers varying along the trimer axis. As a result, NCS averaging of the SeFet crystals required defining several masks and NCS operators for different parts of the molecule. In addition, the low solvent content of the crystals was not beneficial for solvent-flattening density modification, which is the more powerful the larger the solvent content. The phases were transferred to the native SeFe crystal (P1) by molecular replacement, which resulted in an electron density map at ~3 Å resolution and the model further improved. However, the quality of the map was still not good enough for building a full model of SeFet.

The main breakthrough in the experimental phasing of the SeFet crystals was obtaining a higher symmetry crystal form (space group P43212) with cell parameters (a=80.3, b=80.3, c=478.7, α=90, β=90, γ=90) that could accommodate one trimer of SeFe per asymmetric unit. This crystal form grew under identical conditions as the P1 crystal form. However, instead of the seed stock solution containing the seeds, the same solution (9% PEG 4000, 100 mM imidazole pH 8 and 30% MPD) without the seeds was used to set up the crystallization drops. Those crystallization drops were intended to be used for streak seeding, but self-nucleated crystals appeared after ~12 h (Fig. 16). Though these crystals were difficult to reproduce, and approximately only one out of 20 crystals diffracted, we managed to collect a complete native data set at 2.9 Å as well as double-inflection MAD data on a derivative with $Ta_6Br_{12}^{+}$ that diffracted to ~3.6 Å.

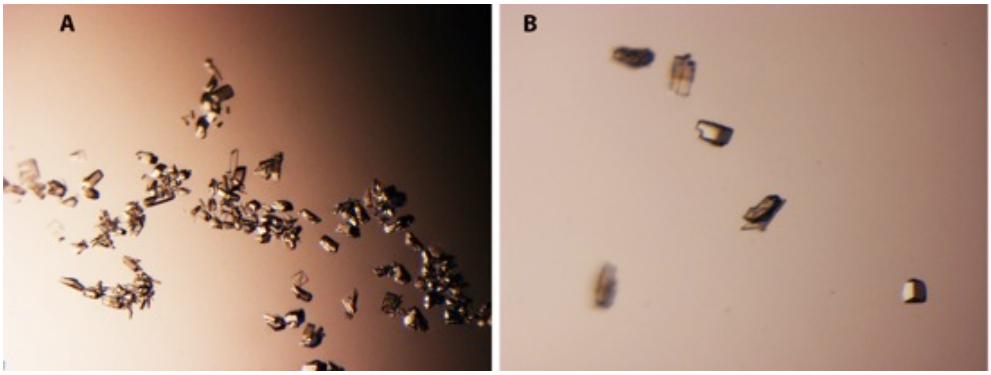


Fig. 16. Deglycosylated SeFet crystals of P43212 space group (A and B). Self-nucleated crystals appeared after ~12 h in 2.5 µl hanging drops (1:1:0.5 protein:reservoir solution:seed stock solution ratio) by vapor diffusion against reservoir solution containing 16-18% PEG 3350 and 200 mM ammonium chloride. Instead of seed stock solution containing the seeds the same solution (9% PEG 4000, 100 mM imidazole pH 8 and 30% MPD) without the seeds was used to set up the crystallization drops.

We were able to resolve the individual tantalum atoms in the cluster (using double-inflection MAD data for a crystal form P43212), which resulted in an improved anomalous signal in comparison with the anomalous signal of the cluster as a super-atom. We also applied multi-crystal averaging to further improve the experimental electron density map. The multi-crystal averaging was complicated due to difficulties finding the right symmetry operators between SeFet molecules in different crystal forms. Nevertheless, performing a few multi-crystal averaging cycles resulted in a better-quality map. The experimental

electron density map calculated from this MAD dataset was further used for model building. The phases were gradually extended to 2.7 Å (the best native data set in P1 crystal form).

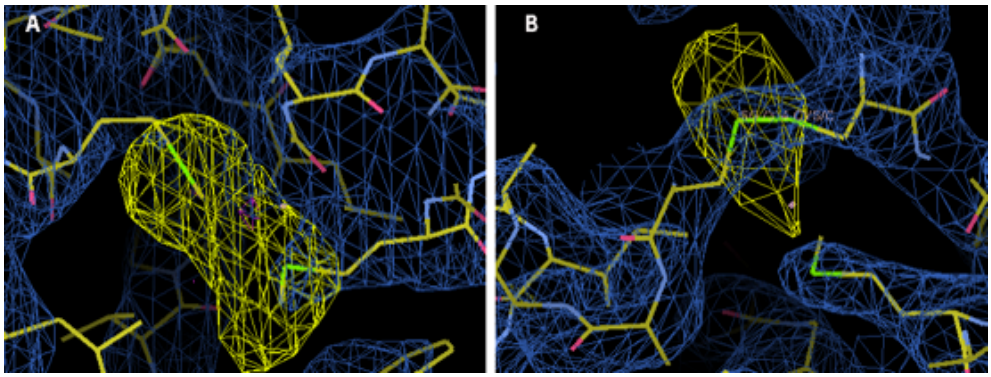


Fig. 17. Electron density for the anomalous scatterers (selenium and sulphur) calculated using AnoDe (374). (A) Electron density maps of SeFet showing the density (yellow) for selenium within selenomethionine residues (SeMet-SAD data set for P1 crystal form) and (B) for sulphur within the disulphide bond between two cysteine residues (S-SAD data set for P1 crystal form).

An anomalous scatterer was introduced into the SeFet cell culture by incorporating selenomethionine (SeMet) to replace methionine within the native protein. The incorporation of SeMet into proteins expressed in insect cells is difficult in comparison to proteins produced in *E. coli*, where the incorporation of SeMet is usually 100 %. Moreover, the yields of SeMet labeled protein are usually lower than those of a native protein due to SeMet toxicity to the cells. Although the expression levels of SeMet-labeled SeFet dropped significantly in comparison with the native protein (0.150 mg of labeled SeFet from 1 l of cell culture supernatant), it was enough to obtain crystals. One of these crystals diffracted to 3.2 Å and was used to collect a SAD data set. SeMet incorporated in SeFe gave a good anomalous signal with usable SAD phases to about 4 Å. In total, approximately 30 SeMet sites were identified (36 methionines in the SeFet trimer) meaning that SeMet incorporation into SeFe was more than 80%. This data set was used to identify the position of the methionines in the protein, which together with previously obtained information about the position of the disulphide bonds helped in assigning the correct amino acid sequence during model building (Fig. 17). Moreover, the SeMet SAD data provided us with another set of good experimental phases.

Table 5. Data collection and refinement statistics for those in the highest resolution bin.

Data collection	**SeFe (native)**	**SeFe (native)**
Space group	P1	$P4_32_12$
Cell dimensions		
a, *b*, *c* (Å)	73.670, 75.080, 78.420	80.230, 80.230, 477.860
a, b, c (°)	94.06, 114.52, 114.72	90.00, 90.00, 90.00
Resolution (Å)	41.38-2.70 (2.85-2.70)	48.78-2.9 (3.005-2.901)
Solvent content (%) (molecules per asymmetric unit)	39.69 (1)	42.41 (1)
R_{merge}	0.071 (0.227)	0.15 (0.222)
I / sI	9.5 (1.3)	9.91 (1.07)
Completeness (%)	89.3 (72.4)	99.87 (99.40)
Redundancy	2.7 (1.5)	4.2 (2.5)
Refinement		
Resolution (Å)	40.85-2.70 (2.79-2.70)	48.78-2.90 (2.98-2.90)
No. reflections	47116	36059
R_{work} / R_{free}		0.2144/ 0.2456
No. atoms		
Protein	9358	9442
Water	15	23
B-factors		
Wilson B-factor (Å^2)	77.18	106.48
Average B-factor (Å^2)	89.53	103.48
R.m.s. deviations		
Bond lengths (Å)	0.008	0.009
Bond angles (°)	0.99	1.08
Ramachandran statistics[#]		
Number of outliers	2	2

[#] Ramachandran statistic according to Molprobity server.

In addition, crystals of SeFet deglycosylated with EndoH/EndoD were obtained. One crystal form, which grew in 14% PEG 4000, 100 mM Tris pH8.5, 200 mM lithium sulphate diffracted X-ray to 3.4 Å resolution. The crystals belonged to P321 space group with cell parameters (a=66.92, b=66.92, c=180.79, α=90, β=90, γ=120). We tried to prepare heavy atom derivatives of this crystal form, but soaking the crystals in heavy atom solutions resulted in loss of diffraction.

The model building and refinement of SeFet was performed alternating between the electron density maps in different crystal forms. During refinement, NCS restraints and TLS groups were applied. The SeFet model was built and fully refined in the P1 crystal form (using the best native data set of 2.7 Å) and the P43212 crystal form (using the best native data set of 2.9 Å). Data collection and refinement statistics are summarized in Table 5.

Key features of SeFet crystal structure

The model of SeFet at 2.9 Å resolution built in the tetragonal crystal form is shown in Fig. 18. The SeFet crystal structure is consistent with the classification of this protein as a class I fusion glycoprotein as predicted from the primary sequence. The structure of SeFet possesses structural features characteristic to other class I fusion proteins in their post-fusion form. It is composed of three tightly intertwined protomers that form a "globular head domain", which is involved in receptor binding in many class I fusion proteins. Below the head, the trimer forms a "stalk region", which is known to assemble into a 6HB when the class I fusion proteins adopt the post-fusion structure during membrane fusion. The protomer of SeFet is wedge-shaped and consists of a globular, predominantly β-sheet-containing head domain, and a stalk region formed mostly by α-helices (Fig. 18). ~20% of the surface area of each protomer is buried in the trimer interface, resulting in a total buried surface area of 14370 Å^2. A number of inter-chain salt bridges stabilize the head region of the trimer, in addition to inter-chain hydrogen bonds and hydrophobic interactions. In contrast, the stalk region is mainly stabilized by inter-chain hydrogen bonds and hydrophobic interactions. Each protomer contains five disulphide bridges and one free cysteine residue (C229). A disulphide bond between residues C94 and C218 keeps the SeFe subunits F1 and F2 connected. Topology diagram of SeFet protomer in the post-fusion form is shown in Fig. 19.

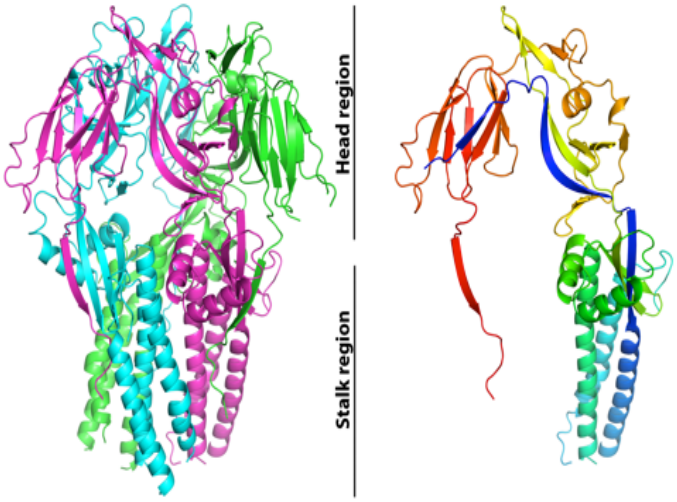


Fig. 18. Crystal structure of SeFet. (A) Cartoon representation of the crystal structure of SeFet with each subunit in different colors. (B) A single SeFe protomer ramp-colored from blue (N terminus) to red (C terminus), through cyan, green, yellow and orange. The head and stalk regions are indicated in the middle.

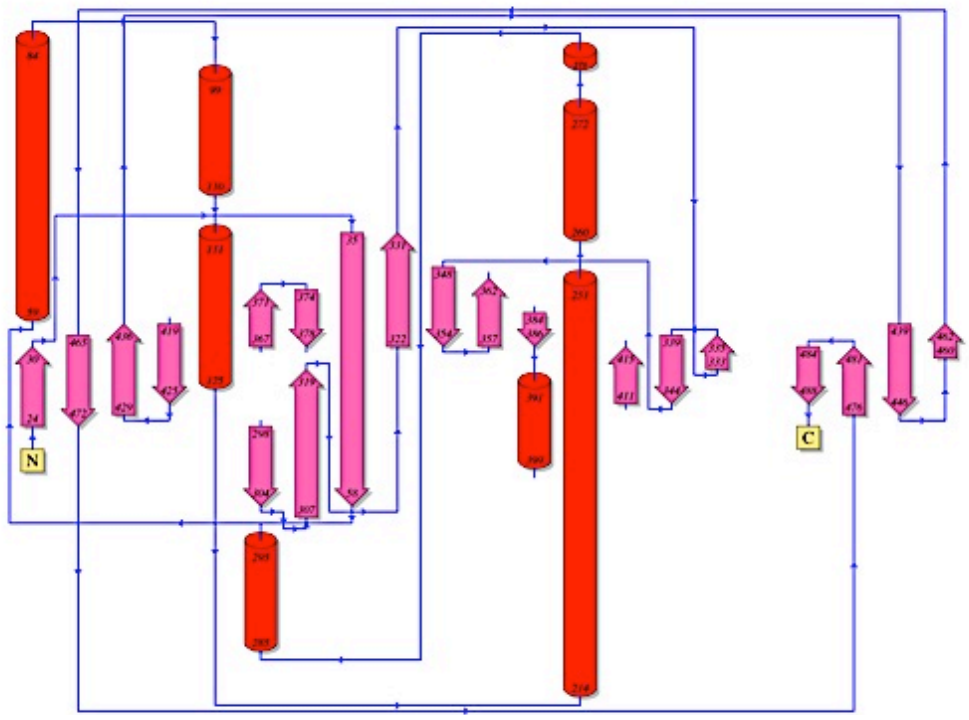


Fig. 19. Topology diagram of SeFet protomer in the post-fusion form. α helices are shown as red cylinders, and β strands are shown as pink arrows. Amino acid numbers at the boundaries of each secondary structure element correspond to the full length SeMNPV F numbering (including secretion signal). The topology diagram was generated using the program PDBsum (192) for chain C of the SeFet crystal structure.

Comparison of SeFet and RSV F

The SeFet structure was compared with other protein structures available in the PDB by using a DALI server (140). As expected, the DALI server assigned the highest Z-scores to the fusion proteins from paramyxoviruses with a Z-score of 11.7 for RSV F protein in its post-fusion conformation (Table 6). A Z-score above 2 indicates that structures have significant similarities, and have similar folds. The superposition of SeFet and RSV F indeed reveals that the proteins are very similar overall in domain organization (Fig. 20).

Table 6. Z-scores assigned by DALI server using SeFet as a query protein structure.

DALI Z-score	Protein	Reference
11.7	Respiratory syncytial virus F protein	(366)
11.1	Parainfluenza virus 3 F protein	(429)
11.1	Newcastle disease virus F protein	(365)
10.5	Parainfluenza virus 5 F protein	(430)
9.0	Human metapneumovirus F protein	(407)

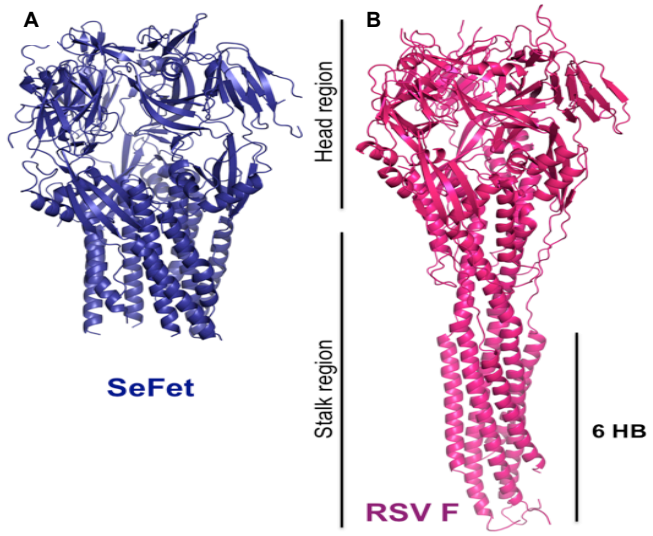


Fig. 20. Comparison of the RSV F and SeFet structures. (A) Structure of SeFet (B) Structure of the RSV F post-fusion trimer (PDB ID code 3RKI). Note the absence of the N-terminal region 6HB in SeFet structure which leads to a shortening of the stalk region. The SeFet structure displays close overall correspondence with the RSV F structure in the head and the first half of the stalk.

Comparison of the superposed post-fusion trimer of RSV F and SeFet side by side (Fig. 20) shows that the major difference between the two structures is located at the base of the stalk region. The RSV F structure extends further to reveal an intact 6HB formed by the HRA and HRB regions, while the structure of SeFet lacks this domain due to trypsin cleavage before low pH induced trimerization (Fig. 21). One of the missing regions in the SeFet structure, from amino acid 137 to 213, includes the fusion peptide and the HR1. Also missing is the region from amino acid 509 to 553, which includes the HR2 domain. As HR1 and HR2

correspond to the N-terminal part of HRA and HRB, respectively, in RSV F, this explains why the 6HB is not present in the crystal structures of trypsin treated low pH induced SeFet. In addition, no electron density was observed for residues 86 to 93, connecting two α-helices in the F2 subdomain. Those amino acid residues are likely to form a flexible loop that is disordered in the crystal structure. Fig. 21 shows missing regions in SeFet inserted as dashed lines. The HRA helix in RSV F extends further down the stalk by about 12 additional helical turns (45 residues), which corresponds approximately to the number of residues (42) between the fusion peptide and the N-terminal end of HR3 in SeFet. Thus, the stalk region in SeFet should be extended to approximately the same extent as in RSV F in a non-proteolysed SeFet structure.

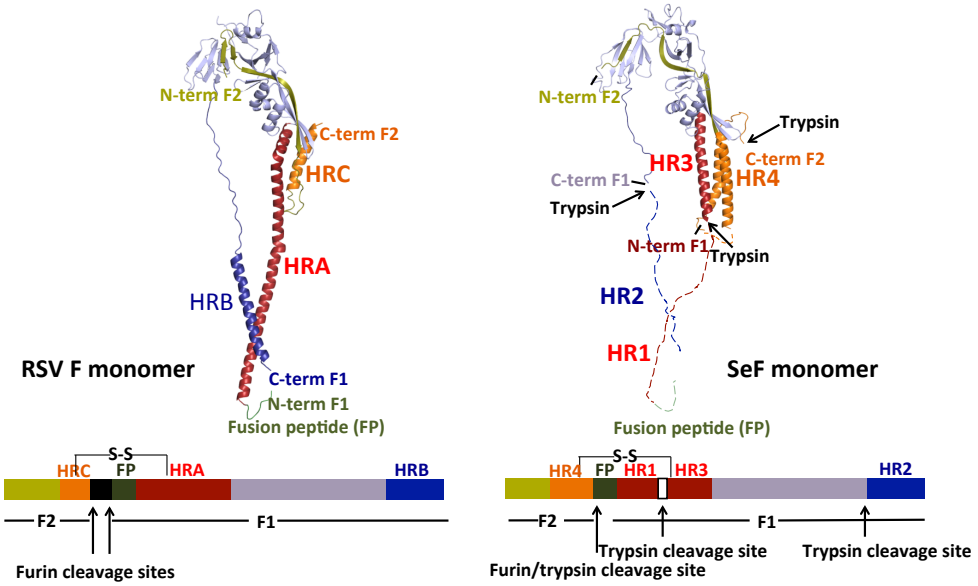


Fig. 21. Comparison of SeFet and RSV F monomer structures colored by key location within the primary sequences. The domains indicated are F2 and F1, the fusion peptide (FP), and the heptad repeat regions (HR1, HR2, HR3 and H4 in SeFe; and HRA, HRB and HRC in RSV F, respectively). Furin and trypsin cleavage sites are marked by black arrows. The missing regions in SeFet are drawn as dashed lines.

Another obvious difference between the superposed models of SeFet and RSV F is present at the beginning of the stalk region. SeFe has a larger F2 domain, which results in an additional α-helix at the C-terminus of the F2 domain which, in turn, packs against HR3 and HR4 helices (Fig. 22). Therefore, the upper part of the stalk region in SeFet is composed of 9 helices in total and is broader than the corresponding region of the RSV F trimer, which consists only of 6 helices (HRA and HRC from each monomer).

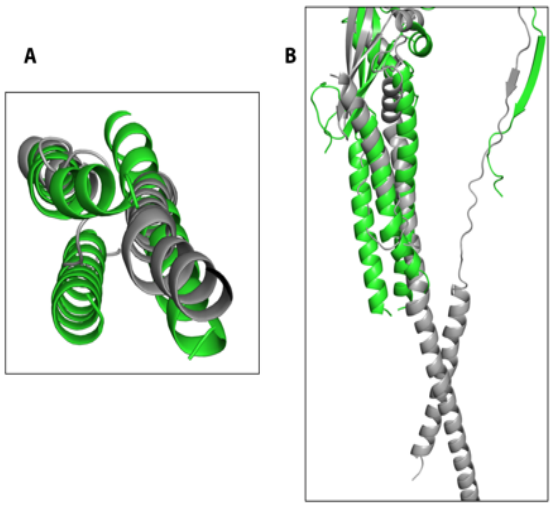
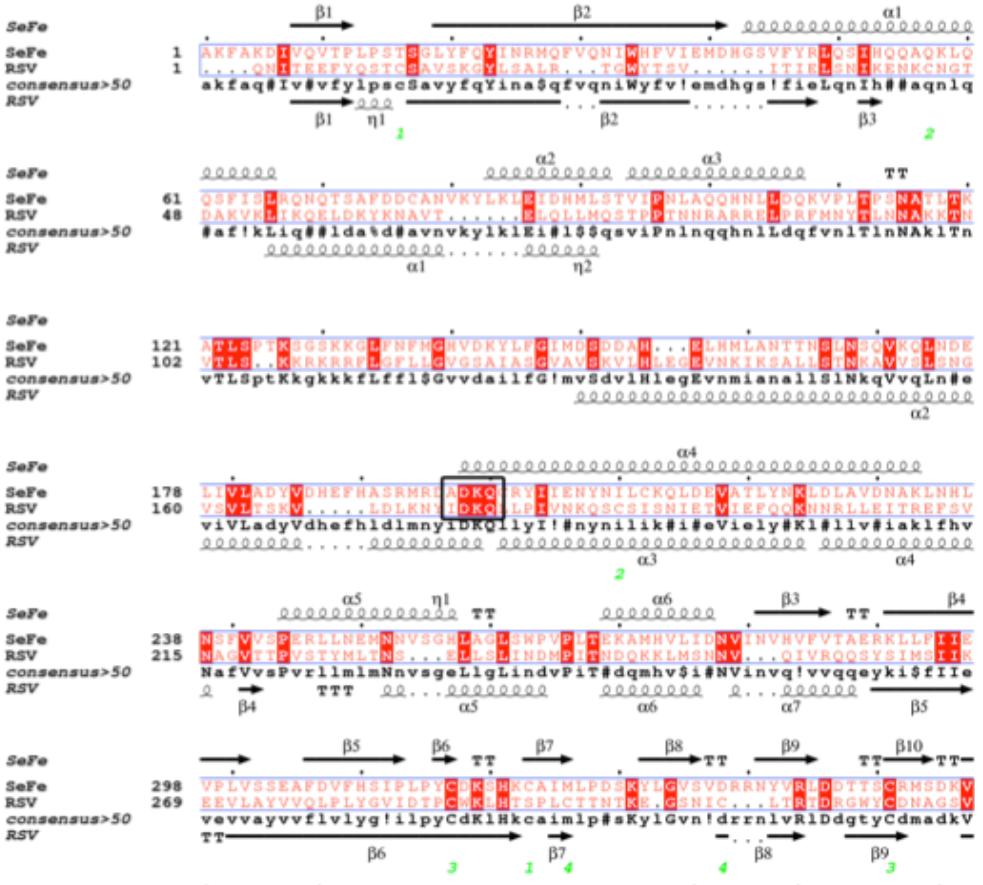


Fig. 22. SeFe and RSV F comparison. (A and B) Cartoon representation of the superposed SeFe (green) and RSV F (grey) monomers showing that SeFe contains an additional α-helix at the C-terminus of the F2 domain, which packs against the HR3 and HR4 helices. In the F2 domain of RSV F, the corresponding helix is missing. (A) Superposed SeFe and RSV F viewed along the three-fold axis from the bottom of the stalk, clipped by planes that cut at the N- and C-terminal ends of the HRA and HRC helices of SeFe. (B) Superposed SeFe and RSV F viewed perpendicular to the three-fold axis.



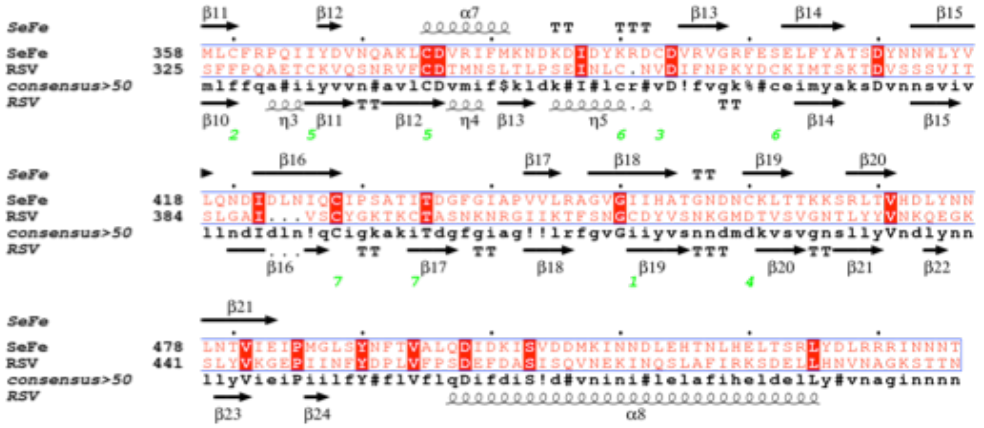


Fig. 23. Sequence alignment of the F ectodomains of SeMNPV and RSV. The sequences were aligned using MultAlin (http://multalin.toulouse.inra.fr/) (67) and formatted with ESPript 3 (http://espript.ibcp.fr/) (115). Numbering starts at the first residue of the ectodomains after the signal peptide removal. The conserved residues are highlighted in red. The conserved position of the x layer type stutter is framed by a black box. Secondary structure features of SeFe are indicated above the sequence, and secondary structure features of RSV F are indicated below the sequence. Green numbers designate residues that form disulfide bonds with the same number for each partner in a disulfide-linked pair.

The structures of the majority of the post-fusion viral fusogenic proteins containing trimeric coiled coils have a characteristic feature called an x-layer type stutter. The stutter is a perturbation of the heptad repeat pattern of the coiled coil due to the insertion of a four- residue motif "defg" between two "abcdefg" repeats (226). The stutter results in an opening of the coiled coil and has been shown to be a useful reference for superposing the central coiled coils of the fusion proteins from different viral families (148). Such superposition allows the comparison of the relative positions of the membrane- interacting elements (i.e. fusion loop and TM region).

The stutter position is also conserved in the *Paramyxoviridae* family with the residues 199-IDKQ-202 corresponding to the stutter in RSV F protein. The alignment of SeFe and RSV F sequences reveals that the RSV F stutter aligns with the residues 214-ADKQ-217 in SeFe (Fig. 23). All four residues 214-ADKQ-217 are present in the SeFet structure only in chain B because of the trypsin cleavage upstream (210-RMRDADKQ-217). Nevertheless, the residues 214-ADKQ-217 superpose with the residues 199-IDKQ-202 in RSV F (Fig. 24). Thus, the residues 214-ADKQ-217 in SeFe correspond to the stutter in the post-fusion form of the baculovirus F protein.

The surface of both molecules lack extensive positively or negatively charged areas (Fig. 25). The most obvious difference between the electrostatic profiles of SeFet and RSV F is that the top of the head domain of SeFet seems to be more negatively charged. The hydrophobicity of the surface of SeFet and RSV F prevents the identification of any distinctive features between the two molecules (Fig. 26).

Conserved residues in baculovirus F

It has been proposed that baculovirus F protein also acts as a receptor binding protein (409). Thus, it is possible that the F protein surface-exposed residues involved in receptor binding are evolutionarily conserved among different baculovirus F proteins. Although it is not known if the post-fusion form is still able to interact with a receptor, the residues interacting with the receptor in the post-fusion form of F protein are likely to stay exposed after the rearrangement of the protein in the post-fusion form.

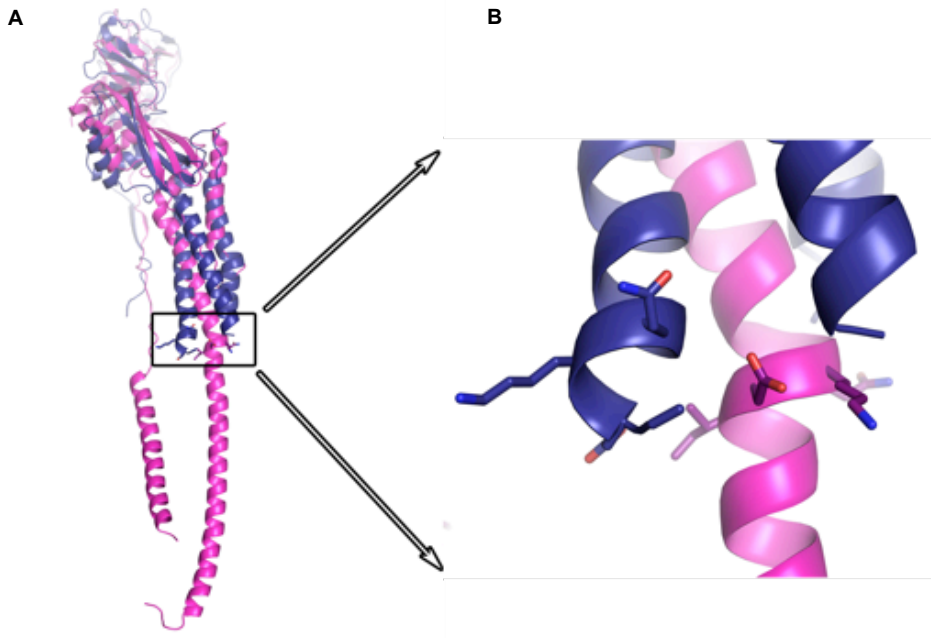


Fig. 24. Stutter conservation in RSV F (magenta) and SeMNPV F (blue). (A) The position of the stutter in the superposed RSV F and SeFe models is framed by the black box. (B) A closer view of the stutter in the superposed models. The side chains of the four residues composing the x-layer stutter are displayed as sticks.

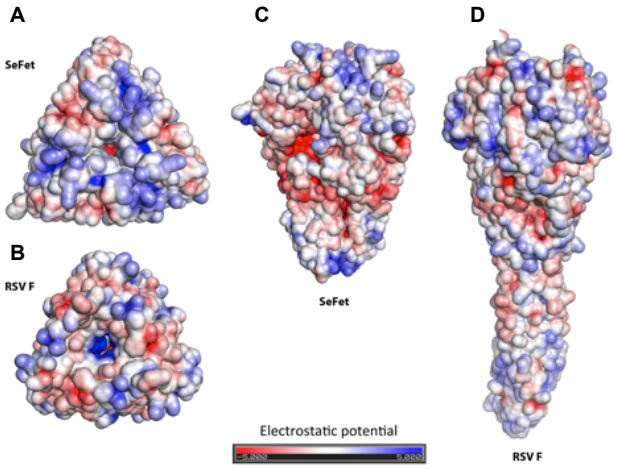


Fig. 25. Electrostatic properties of the surfaces of SeFet (A and C) and RSV F (B and D). The proteins are shown as solvent-accessible surfaces colored by electrostatic potential calculated using the adaptive Poisson-Boltzmann solver and contoured at ± 5 kT/e (red, acidic/negative; blue, basic/positive). (A and B) Surface representation viewed along the three-fold axis from the top of the head. (C and D) Surface representation viewed perpendicular to the three-fold axis.

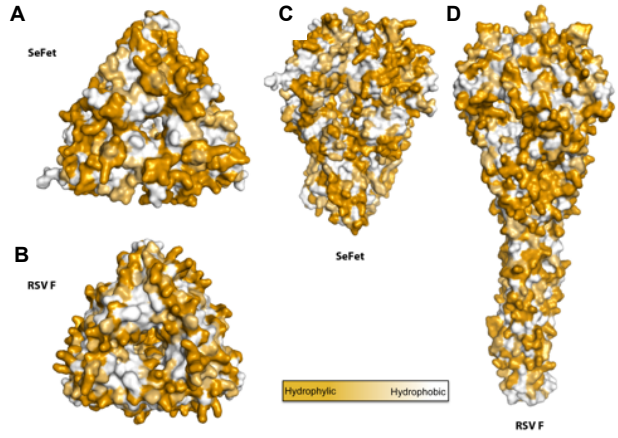


Fig. 26. Hydrophobicity of the SeFet (A and C) and RSV F (B and D) surfaces. The proteins are shown as solvent-accessible surfaces colored according to a normalized hydrophobicity scale from white (hydrophobic) to bright yellow (hydrophilic). (A and B) Surface representations viewed along the three-fold axis from the top of the head. (C and D) Surface representations viewed perpendicular to the three-fold axis.

In order to identify residues that potentially could be involved in the receptor binding, we analyzed which surface exposed residues of the F protein are conserved among the 39 different baculovirus F proteins. The conserved residues were mapped using ConSurf (111). The SeFet surface colored by residue conservation is shown in Fig. 27. The conserved residues are scattered around the whole protein. As a result, there is no easily identifiable highly conserved site. The residues around the axial canal of SeFet form the most distinct highly conserved patch. In addition, there are some conserved surface exposed residues on the side of the SeFet head region as well as the stalk region. The analysis yielded several clearly conserved residue patches on the SeFet surface that could potentially act in

receptor recognition. One pretty strong cluster is present at the top of the trimer axis (Fig. 27F). In order to further investigate the role of the conserved residues exposed on the surface of SeFet, mutagenesis of those residues combined with cell-binding/infectivity assays should be performed, but may be wise to wait for the structure of the pre-fusion form in order to carry out such an experiment.

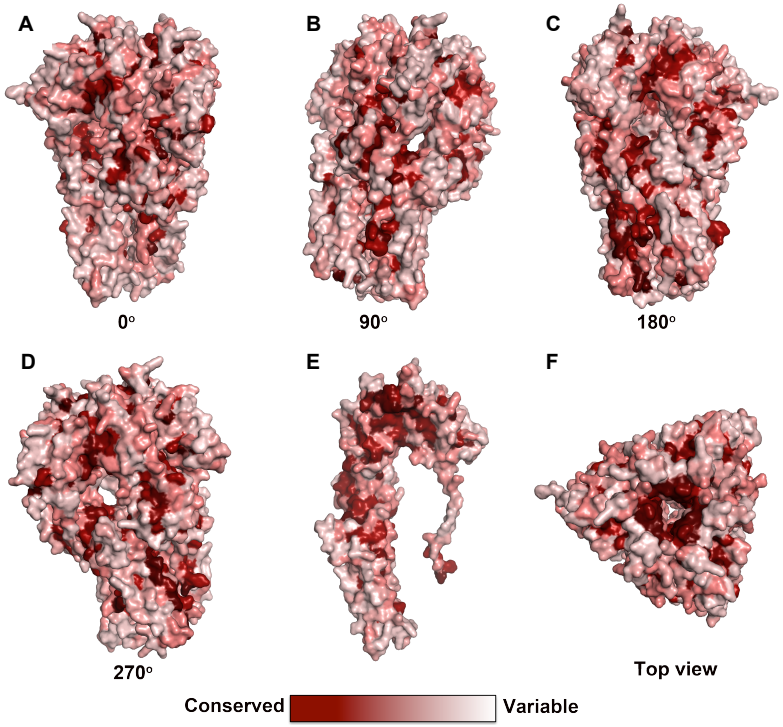


Fig. 27. Molecular surface of SeFe colored by sequence conservation among SeFe and 39 F proteins from different baculoviruses analyzed by the program ConSurf (111). The colors vary from dark red for highly conserved residues to white for residues with little conservation. (A, B, C and D) Surface representation of SeFet viewed perpendicular to the three-fold axis at angles 0, 90, 180 and 270°; (E) surface representation viewed along the three-fold axis from the top of the head; (F) surface representation of a monomer viewed perpendicular to the three-fold axis.

Cellular F homologues

The PSI-BLAST search for remote homology using the SeFe sequence (including only the residues present in 3D model of SeFet) yielded proteins containing an approximately 360 amino acid long domain of unknown function (DUF3609 has been previously found in the F proteins of baculoviruses as well as in the coding sequences of *env* genes of endogenous insect retroviruses (319). In addition, it was also identified in eukaryotic genomes of *Drosophila* and mosquito *Anopheles*

genomes encoding cellular proteins (231). However, the PSI-BLAST search also yielded the protein sequences from other eukaryotes that have not yet been proposed to display homology with baculovirus F proteins (Fig. 29). The highest degree of conservation is present within the sequences corresponding to the DUF3609 domain (Fig. 28). All those sequences were found in insect genomes with the exception of *Branchiostoma floridae* (Florida lancelet) and *Lottia gigantea* (Giant owl limpet). As a result, we looked for conserved surface exposed residues between the SeFe and cellular F-homologues within DUF3609 domain (Fig. 28). The analysis reveals that a number of highly conserved residues are exposed on the protein surface forming a few distinct patches. The functional and/or structural importance of these evolutionary conserved residues in DUF3609 domain still remains to be elucidated.

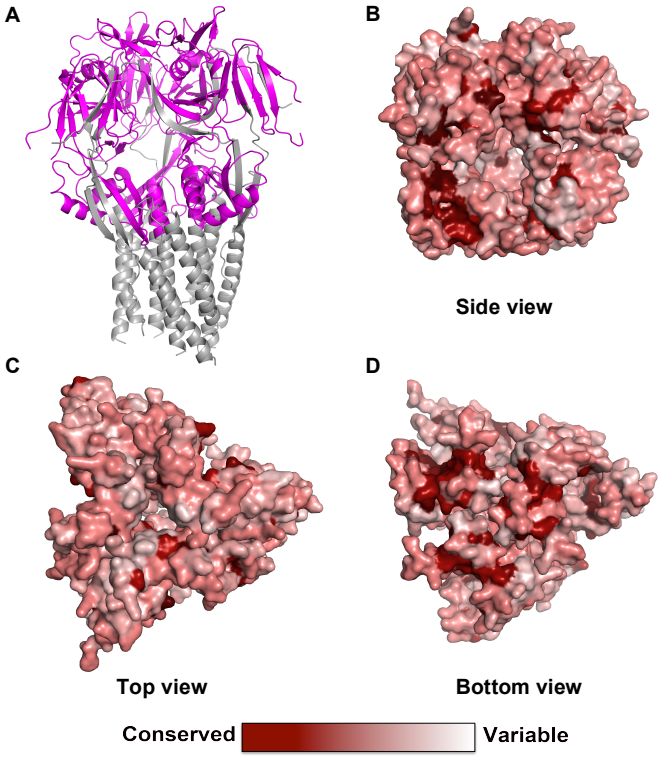
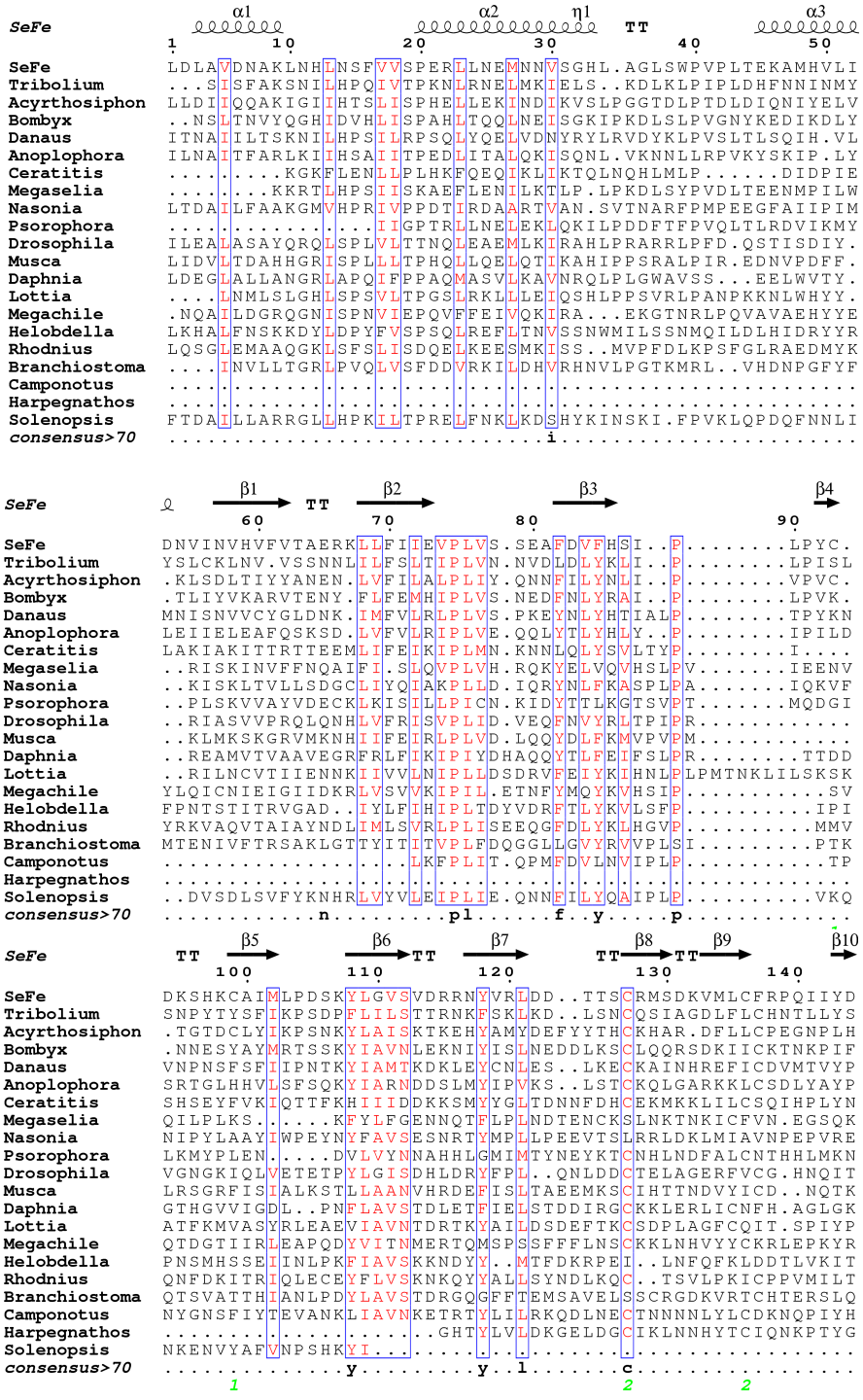


Fig. 28. DUF3609 domain. (A) Cartoon representation of SeFet with the sequence of DUF3609 domain colored in magenta. (B, C and D) Molecular surface of DUF3609 domain colored by sequence conservation among SeFe and 20 cellular F homologues analyzed by the program ConSurf (28). The colors vary from dark red for highly conserved residues to white for residues with little conservation.


SeFe
α1
α2
η1
TT
α3
1 10 20 30 40 50
SeFe LDLAVDNAKLNHLNSFVVSPERLLNEMNNVSGHL.AGLSWPVPLTEKAMHVLI
Tribolium ...SISFAKSNILHPQIVTPKNLRNELMKIELS..KDLKLPIPLDHFNNINMY
Acyrthosiphon LLDIIQQAKIGIIHTSLISPHELLEKINDIKVSLPGGTDLPTDLDIQNIYELV
Bombyx ..NSLTNVYQGHIDVHLISPAHLTQQLNEISGKIPKDLSLPVGNYKEDIKDLY
Danaus ITNAIILTSKNILHPSILRPSQLYQELVDNYRYLRVDYKLPVSLTLSQIH.VL
Anoplophora ILNAITFARLKIIHSAIITPEDLITALQKISQNL.VKNNLLRPVKYSKIP.LY
Ceratitis .....KGKFLENLLPLHKFQEQIKLIKTQLNQHLMLP......DIDPIE
Megaselia .....KKRTLHPSIISKAEFLENILKTLP.LPKDLSYPVDLTEENMPILW
Nasonia LTDAILFAAKGMVHPRIVPPDTIRDAARTVAN.SVTNARFPMPEEGFAIIPIM
Psorophora .....IIGPTRLLNELEKLQKILPDDFTFPVQLTLRDVIKMY
Drosophila ILEALASAYQRQLSPLVLTTNQLEAEMLKIRAHLPRARRLPFD.QSTISDIY.
Musca LIDVLTDAHHGRISPLLLTPHQLLQELQTIKAHIPPSRALPIR.EDNVPDFF.
Daphnia LDEGLALLANGRLAPQIFPPAQMASVLKAVNRQLPLGWAVSS...EELWVTY.
Lottia ....LNMLSLGHLSPSVLTPGSLRKLLLEIQSHLPPSVRLPANPKKNLWHYY.
Megachile .NQAILDGRQGNISPNVIEPQVFFEIVQKIRA...EKGTNRLPQVAVAEHYYE
Helobdella LKHALFNSKKDYLDPYFVSPSQLREFLTNVSSNWMILSSNMQILDLHIDRYYR
Rhodnius LQSGLEMAAQGKLSFSLISDQELKEESMKISS..MVPFDLKPSFGLRAEDMYK
Branchiostoma ....INVLLTGRLPVQLVSFDDVRKILDHVRHNVLPGTKMRL..VHDNPGFYF
Camponotus .....
Harpegnathos .....
Solenopsis FTDAILLARRGLLHPKILTPRELFNKLKDSHYKINSKI.FPVKLQPDQFNNLI
consensus>70 .....i.....
SeFe
β1
TT
β2
β3
β4
60 70 80 90
SeFe DNVINVHVFVTAERKLLFIIEVPLVS.SEAFDVFHSI..P.....LPYC.
Tribolium YSLCKLNV.VSSNNLILFSLTIPLVN.NVDLDLYKLI..P.....LPISL
Acyrthosiphon .KLSDLTIYYANEN.LVFILALPLIY.QNNFILYNLI..P.....VPVC.
Bombyx .TLIYVKARVTENY.FLFEMHIPLVS.NEDFNLYRAI..P.....LPVK.
Danaus MNISNVVCYGLDNK.IMFVLRLPLVS.PKEYNLYHTIALP.....TPYKN
Anoplophora LEIIELEAFQSKSD.LVFVLRIPLVE.QQLYTLYHLY..P.....IPILD
Ceratitis LAKIAKITTRTTEEMLIFEIKIPLMN.KNNLQLYSVLTYP.....I....
Megaselia ..RISKINVFFNQAIFI.SLQVPLVH.RQKYELVQVHSLPV.....IEENV
Nasonia ..KISKLTVLLSDGCLIYQIAKPLLD.IQRYNLFKASPLPA.....IQKVF
Psorophora ..PLSKVVAYVDECKLKISILLPICN.KIDYTTLKGTSVPT.....MQDGI
Drosophila ..RIASVVPRQLQNHLVFRISVPLID.VEQFNVYRLTPIPR.....
Musca ..KLMKSKGRVMKNHIIFEIRLPLVD.LQQQYDLFKMVPVPM.....
Daphnia ..REAMVTVAAVEGRFRLFIKIPIYDHAQQYTLFEIFSLPR.....TTDD
Lottia ..RILNCVTIIENNKIIVVLNIPLLDSDRVFEIYKIHNLPLPMTNKLILSKSK
Megachile YLQICNIEIGIIDKRLVSVVKIPIL.ETNFYMQYKVHSIP.....SV
Helobdella FPNTSTITRVGAD..IYLFIHIPLTDYVDRFTLYKVLSFP.....IPI
Rhodnius YRKVAQVTAIAYNDLIMLSVRLPLISEEQGFDLYKLHGVP.....MMV
Branchiostoma MTENIVFTRSAKLGTTYITITVPLFDQGGLLGVYRVVPLSI.....PTK
Camponotus .....LKFPLIT.QPMFDVLNVIPLP.....TP
Harpegnathos .....
Solenopsis ..DVSDLSVFYKNHRLVYVLEIPLIE.QNNFILYQAIPLP.....VKQ
consensus>70 .....n.....pl.....f..y.....p.....
SeFe
TT
β5
β6
TT
β7
TT
β8
TT
β9
β10
100 110 120 130 140
SeFe DKSHKCAIMLPDSKYLGVSVDRRNYVRLDD..TTSCRMSDKVMLCFRPQIIYD
Tribolium SNPYTYSFIKPSDPFLILSTTRNKFSKLKD..LSNCQSIAGDLFLCHNTLLYS
Acyrthosiphon .TGTDCLYIKPSNKYLAISKTKEHYAMYDEFYYTHCKHAR.DFLLCPEGNPLH
Bombyx .NNESYAYMRTSSKYIAVNLEKNIYISLNEDDLKSCLQQRSDKIICKTNKPIF
Danaus VNPNSFSFIIPNTKYIAMTKDKLEYCNLES..LKECKAINHREFICDVMTVYP
Anoplophora SRTGLHHVLSFSQKYIARNDDSLMYIPVKS..LSTCKQLGARKKLCSDLYAYP
Ceratitis SHSEYFVKIQTTFKHIIIDDKKSMYYGLTDNNFDHCEKMKKLILCSQIHPLYN
Megaselia QILPLKS.....KFYLFGENNQTFLPLNDTENCKSLNKTNKICFVN.EGSQK
Nasonia NIPYLAAYIWPEYNYFAVSESNRTYMPLLPEEVTSLRRLDKLMIAVNPEPVRE
Psorophora LKMYPLEN.....DVLVYNAHHLGMIMTYNEYKTCNHLNDFALCNTHHLMKN
Drosophila VGNGKIQLVETETPYLGISDHLDRYFPLL..QNLDDCTELAGERFVCG.HNQIT
Musca LRSGRFISIALKSTLLAANVHRDEFISLTAEEMKSCIHTTNDVYICD..NQTK
Daphnia GTHGVVIGDL..PNFLAVSTDLETFIELSTDDIRGCKKLERLICNFH.AGLGK
Lottia ATFKMVASYRLEAEVIAVNTDRTKYAILDSDEFTKCSDPLAGFCQIT.SPIYP
Megachile QTDGTIIRLEAPQDYVITNMERTQMSPSSFFFLNSCKKLNHVYYCKRLEPKYR
Helobdella PNSMHSSEIINLPKFIAVSKKNDYY..MTFDKRPEI..LNFQFKLDDTLVKIT
Rhodnius QNFDKITRIQLECEYFLVSKNKQYYALLSYNDLKQC..TSVLPKICPPVMILT
Branchiostoma QTSVATTHIANLPDYLAVSTDRGQGFFTEMSAVELSSCRGDKVRTCHTERSLQ
Camponotus NYGNSFIYTEVANKLIAVNKETRTYLILRKQDLNECTNNNNLYLCDKNQPIYH
Harpegnathos .....GHTYLVLDKGELDGCIKLNNHYTCIQNKPTYG
Solenopsis NKENVYAFVNPSHKYI.....
consensus>70 .....y.....y..l.....c.....
1 2 2

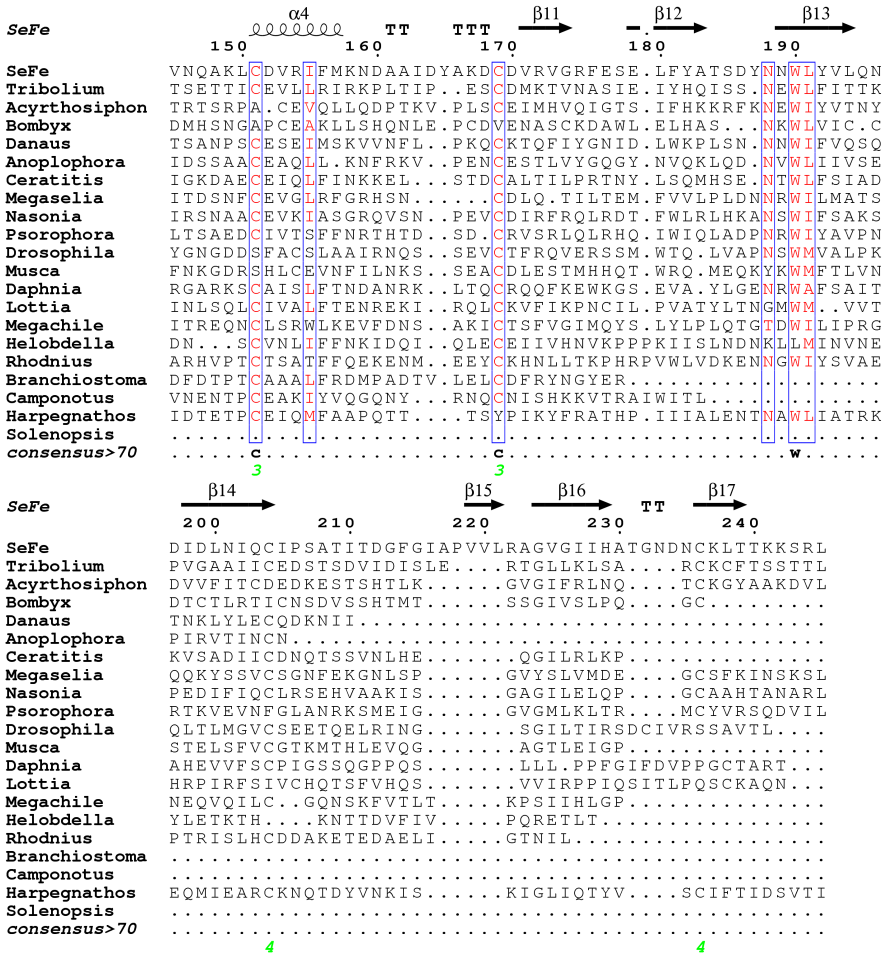


Fig. 29. Multiple sequence alignment of SeFe and cellular F homologues from different species. The aligned sequences: *Megachile rotundata* (alfalfa leaf cutting bee) UniRef90_UPI000258F7DF, *Helobdella robusta* (Californian leech) UniRef90_T1FA15, *Acyrthosiphon pisum* (Pea aphid) UniRef90_J9LX66_, *Megaselia scalaris* (Humpbacked fly) UniRef90_T1GHP5_, *Psorophora albipes* (Psorophora mosquitoe) UniRef90_T1DI06, Nasonia vitripennis (Parasitic wasp) UniRef90_K7JGM3, *Branchiostoma floridae* (Florida lancelet) UniRef90_C3ZJR6, *Drosophila lutescens* (Fruit fly) UniRef90_Q30CL5, *Tribolium castaneum* (Red flour beetle) UniRef90_D7GY13, *Harpegnathos saltator* (Jerdon's jumping ant) UniRef90_E2BQZ1, *Ceratitis capitata* (Mediterranean fruit fly) UniRef90_W8AJR2, *Daphnia pulex* (Water flea) UniRef90_E9HB17, *Bombyx mori* (Silk moth) UniRef90_H9JCV6, *Lottia gigantea* (Giant owl limpet) UniRef90_V4AHE8, *Danaus plexippus* (Monarch butterfly) UniRef90_G6CIW0, *Camponotus floridanus* (Florida carpenter ant) UniRef90_E2A5H8, *Anoplophora glabripennis* (Asian longhorn beetle) UniRef90_V5I841, *Musca domestica* (House fly) UniRef90_T1PDA3, *Rhodnius prolixus* (Triatomid bug) UniRef90_T1HQD7, *Solenopsis invicta* (Red imported fire ant) UniRef90_E9J0H3. Green numbers designate residues that form disulfide bonds, with the same number for each partner in a disulfide-linked pair.

Structural studies of SeF full-length ectodomain

In order to obtain structural information about the 6HB which is lacking in SeFet crystal structure, we obtained and crystallized the post-fusion trimer of the full-length SeF ectodomain. Previous observations have shown that low-pH triggered trimerization of the SeFe containing a wild type fusion peptide results in exposure of the hydrophobic fusion peptide, which leads to protein aggregation. Trypsin treatment of SeFe harbouring a mutated furin cleavage site yielded a proteolytic fragment of the protein lacking the fusion peptide, allowing successful acid-induced trimerization. It has been described in the literature that the post-fusion forms of viral fusion protein ectodomains can be obtained when the hydrophobic fusion peptide is deleted or its hydrophobic nature altered. Therefore, in order to achieve the trimerization of the full-length SeFe, we produced three new constructs of SeFe (all harboring a wild type furin cleavage site): 1) SeFeΔ1 containing the deletion of the fusion peptide residues 150-GLFNFMGHV-158 2) SeFeΔ2 containing the deletion of the fusion peptide residues 150-GLFNFMGHVDKYLF-163 and 3) SeFe-mut containing the hydrophobic residues within the fusion peptide replaced by hydrophilic ones (150-GLFNFMGHV- 158 and 150-GQTNSHGHN-158). In SeFeΔ1 and SeFeΔ2 constructs a short GGS linker was introduced instead of the deleted fusion peptide sequences. All proteins were expressed in *Drosophila* S2 cells yielding 6-10 mg/L of supernatant. SEC purification of SeFeΔ1 and SeFeΔ2 revealed that these proteins are monomeric at pH 8. In contrast, SeFe-mut formed both monomers and trimers at pH8. The SeFe-mut trimer was stable at pH 8 and presumably represented a pre-fusion trimer. This trimer was subjected to crystallization trials and crystals were obtained, albeit diffracting only to ~10 Å. The formation of post-fusion trimers was attempted with monomeric forms of SeFeΔ1, SeFeΔ2 and SeFe-mut by incubating them in 30 mM NaAc pH 5. Incubation of SeFeΔ2 at room temperature for two hours in 30 mM NaAc was sufficient to achieve a complete trimerization, while SeFeΔ1 and SeFe-mut required overnight incubation at room temperature for the formation of a post-fusion trimer.

Intensive crystallization trials with these proteins were performed, which resulted in obtaining crystals under different crystallization conditions. Unfortunately, all tested crystals diffracted to low resolution. The best diffraction (6.5 Å) was obtained for the SeFeΔ1 trimer after removal of the Strep-tag and deglycosylation of the protein with EndoD (space group I222; cell unit dimensions: a=111.13 b=117.09 c=348.37 $\alpha = \beta = \gamma = 90.000$ deg.). The packing of this new crystal

form provides an extra space for accommodating the extension of the trimer, which is lacking in the trypsin truncated crystal form. Moreover, the analysis of the crystal lattice shows that the crystal packing would not be stable without the extra chains joining the consecutive layers in the direction of the long axis (c=348.0) (Fig. 30).

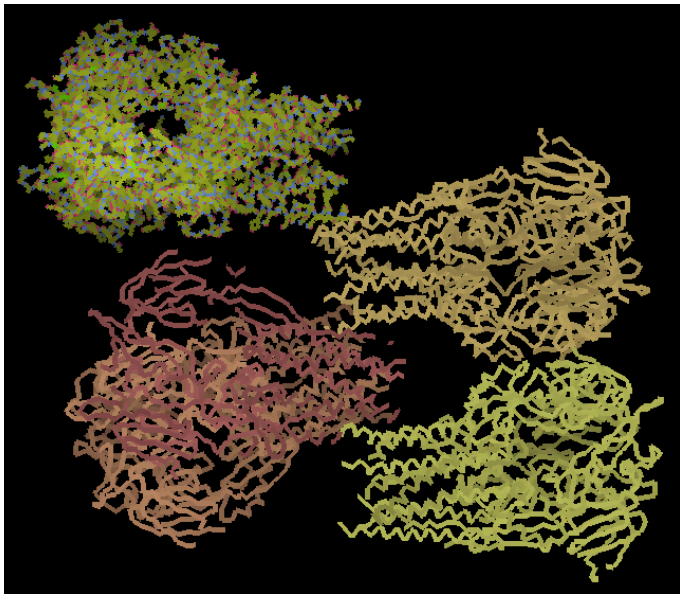


Fig. 30. Crystal packing of SeFeΔ1 timer. The model of SeFet (truncated post-fusion trimer) was used for molecular replacement. Crystal packing analysis reveals an extra space for extension of the molecule.

To further delineate the structure of the full-length SeFet post-fusion form, we performed electron microscopy (EM) studies on the SeFe-mut post-fusion trimer (Fig. 31). Docking SeFet crystal structure into the 30-50 Å resolution reconstruction map revealed an extra extensive density corresponding to the 6HB which is absent in the crystal structure (Fig. 32).

DISCUSSION

Evolutionary links revealed by baculovirus F protein structure

We characterized the trypsin-truncated SeMNPV F ectodomain (SeFe) and showed that low pH induces conformational changes within the protein leading to formation of a stable post-fusion trimer (Chapter 2 and 3). After intensive crystallization attempts, diffracting crystals of the post-fusion trimer of SeFe trypsin fragment (SeFet) were obtained. The structure determination of SeFet confirmed the previous predictions that baculovirus F protein exhibits a class I fusion protein fold.

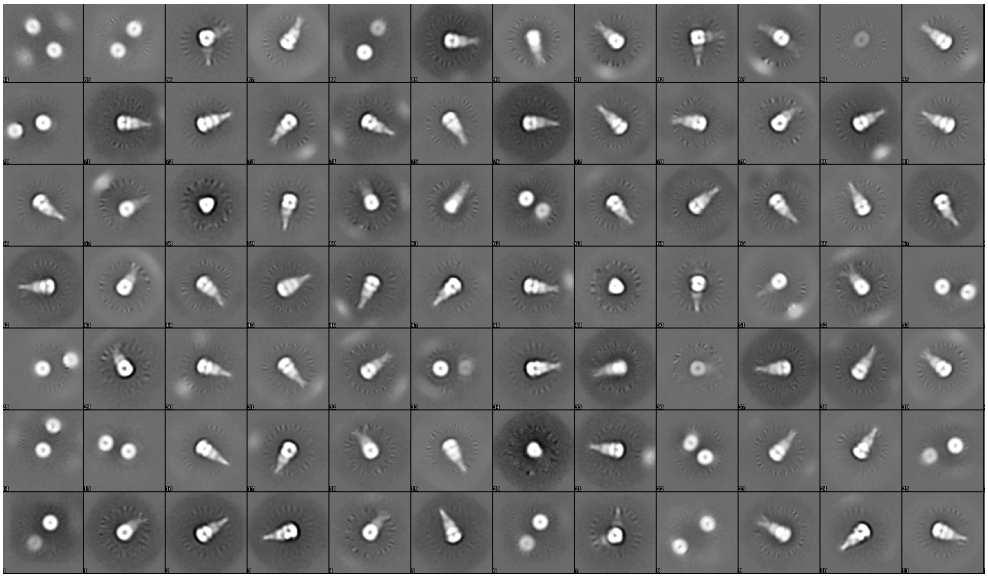


Fig. 31. Reference-free 2D class averages of negatively stained SeFe-mut post-fusion trimer.

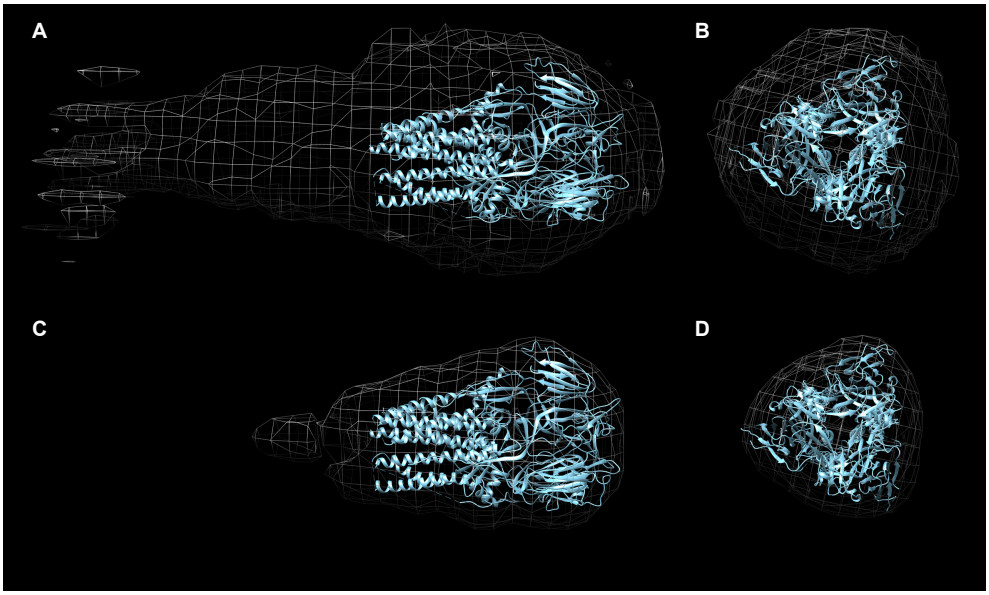


Fig. 32. EM reconstruction of SeFe-mut post-fusion trimer. The crystal structure of SeFet (displayed as cartoon) is fitted within the electron density. (A and B) Side view and top view, respectively, at contour level of the density 0.651. (C and D) Side view and top view, respectively, at contour level of the density 0.162.

The recombinant SeFe contained a mutated furin cleavage site in order to prevent furin cleavage and, as a result, stabilize the protein in its pre-fusion form. Based on the knowledge that furin cleavage occurs in the trans-Golgi network, the mutation was not expected to impede protein folding. In addition, available pre-fusion structures of other class I fusion proteins (influenza HA and paramyxovirus F) in their cleaved and uncleaved forms, are largely superimposable (56, 406). Proteolytic activation is required for the majority of class I fusion proteins to induce membrane fusion, i.e. for conformational changes to reach the postfusion state. When altering the furin cleavage site, two arginine residues were introduced within this site in order to allow proteolytic activation of the protein by trypsin cleavage. We have previously observed that the wild type SeFe cleaved by furin aggregates if incubated at low pH, which could be explained by exposure of the hydrophobic fusion peptides that cluster together in the absence of cellular membranes. As illustrated by examples of other class I fusion proteins, obtaining a soluble form of the post-fusion trimer usually requires removal of the fusion peptide (323, 366).

Trypsin cleavage of the mutated SeFe yielded a stable proteolytic fragment of the protein lacking the fusion peptide, which was able to form a trimer when exposed to low pH. This trimer was presumably corresponding to the postfusion form based on the observed low pH dependent fusion of the baculovirus F protein (411). The conformational changes induced in the SeFe trypsin fragment by low pH were irreversible, which is consistent with the transition of a metastable pre-fusion form to a more stable post-fusion conformation, which is a general feature of the majority of viral fusion proteins. SeFe with knocked-down furin cleavage site expressed in S2 cells was secreted as a monomer, which is in contrast to the observed trimeric state of the full-length protein (216). Both class I and class III fusion proteins lacking TM domain have been reported to trimerize inefficiently (7, 45, 333, 341, 423). TM-TM interactions have been suggested to be important for the stability of pre-fusion forms of paramyxovirus F proteins (341). The crystal structures of pre-fusion forms of parainfluenza virus 5 (PIV 5) F protein and RSV F protein trimers were obtained only when the TM domain was replaced with a trimerization domain (GCN4t or T4 fibritin, respectively) (246, 430). Trimeric interfaces of paramyxovirus F and influenza HA proteins are significantly different in their pre-and post-fusion forms (45, 418, 429, 430) revealing weaker interprotomer interactions in the pre-fusion forms. Postfusion forms of these proteins are stabilized via hydrophobic interactions of the HR regions forming the 6-helix bundle. The trypsin cleaved SeFe formed a stable

post-fusion trimer after acid-induced conformational changes, which also promoted transition from a monomeric to a trimeric state.

The crystal structure revealed that SeFet lacks HR1 and HR2 regions and, as a result, does not form a 6-helix bundle, which is believed to contribute the most to the stability of the post-fusion state of class I fusion proteins. Instead, the assembly and stabilization of SeFet seems to depend mainly on a trimeric coiled coil formed by the HR3 segments. We complemented our crystal structure with a negative stain EM reconstruction of the post-fusion form of the full-length SeFe ectodomain. The visualization of the molecule in the EM map revealed that the molecule has a shape and size characteristic of the post-fusion trimer of the paramyxovirus F protein and allowed us to confirm the presence of the six-helix bundle at the trimer axis. The trypsin-truncated SeFet crystal structure was docked into the EM map, clearly revealing extra density corresponding to the 6-helix bundle, which is absent in the crystal structure. Taken together, the EM density map and the X-ray crystallography data provide evidence that baculovirus F protein is a typical class I fusion protein.

SeFe requires low pH in order to undergo conformational changes from the pre-fusion to the post-fusion form. We did not observe formation of post-fusion trimeric forms at neutral pH for either trypsin-cleaved or uncleaved ectodomain. This is in contrast to F proteins from paramyxoviruses, which are able to catalyze the membrane fusion reaction at neutral pH. Thus, despite structural similarities between baculovirus F and paramyxovirus F proteins, conformational changes within these proteins leading to the fusion reaction are induced by different triggers. Conformational changes in paramyxovirus F proteins from PIV3 and PIV5 are triggered by the HN attachment protein and in others like RSV and SV5 F proteins upon receptor binding [reviewed in (36)]. The secreted hPIV3 F ectodomain with a knocked-down furin cleavage site has been shown to adopt primarily the post-fusion conformation instead of the expected pre-fusion form. This indicates that the energetic barrier for transition of paramyxovirus F ectodomain from pre-fusion to post-fusion form is rather low, once the stabilizing attachment protein is removed.

Remarkably, SeFe contains 20 histidine residues compared to only 3 found in the RSV F ectodomain. The protonation of histidine residues has been reported to trigger conformational changes of pH-dependent fusion proteins leading to membrane fusion (165). In the postfusion form of viral fusion proteins, a number of conserved histidine residues often tend to form salt bridges with negatively

charged residues that are supposed to have a stabilizing effect. Low numbers of histidine residues in RSV F ectodomain are in line with its low-pH independent fusion mechanism.

The SeFet crystal structure represents the first crystal structure of a class I fusion protein from a DNA virus. Moreover, it also confirmed the predicted similarities between baculovirus F and paramyxovirus F proteins, with the post-fusion form of RSV F protein (PDB accession code: 3RKI) having the highest DALI score. The conservation of the structural features between baculovirus F and paramyxovirus F proteins suggests that both proteins have a common ancestor, which has been already predicted from the sequence similarity (262). Remarkably, baculoviruses and paramyxoviruses are distant viral families. First of all they belong to different groups of viruses: group I (double-stranded DNA viruses) and group V (negative-sense single stranded RNA viruses), respectively (16). Second, they have different hosts: baculoviruses infect invertebrates, primarily insects, while paramyxoviruses infect vertebrates and have a very broad host range. Finally, these viruses use different entry mechanisms to their host cells: baculovirus F proteins function both as receptor binding and fusion proteins and induce fusion in the endosomes upon acidification while most of paramyxoviruses have separate receptor and fusion proteins and fuse directly with the cellular membranes at neutral pH. VSV G and herpes simplex virus glycoprotein B are other examples of homologous fusion proteins occurring in very different viruses, one RNA and the other one DNA (133, 311). Moreover, they do not infect the same hosts either. Therefore, the route by which baculoviruses and paramyxoviruses acquired the same common ancestral fusion protein is uncertain. The origin of RNA and DNA viruses is still debatable (92, 141); therefore, it is not clear if paramyxoviruses and baculoviruses acquired F proteins independently during evolution (probably from a cellular F ortholog) or if one of the two viral families was the source of the F protein.

Intriguingly, the DUF3609 domain found in baculovirus proteins has been also discovered in some cellular proteins and retroviral elements suggesting that that they evolved from the same ancestral protein, which could be either viral or eukaryotic origin (222). The SeFet crystal structure revealed that the DUF3609 domain resides within the globular head domain of SeFet. Therefore, it is unlikely that the DUF3609 domain is involved in membrane fusion. Presumably, it plays a role in cellular receptor recognition. Analysis of the conserved solvent- exposed residues in the DUF3609 domain revealed a few conserved patches on the surface

of the domain that could potentially represent a receptor binding site and/or some other unidentified functional site. The fusion activity of currently identified F cellular proteins from *Drosophila melanogaster, Drosophila yakuba, Drosophila pseudoobscura* and *Anopheles gambiae* have not yet been demonstrated. Although it is possible that the fusion activity was not observed due to assay conditions and these proteins do contain other domains than DUF3609 mediating fusion, it is also likely that cellular F proteins have completely different function.

The prevalence of *f-like* genes in the genomes of many insects suggests that baculovirus *f* gene or at least a DUF3609 domain could be acquired directly or indirectly from an insect host. The indirect route might include insect endogenous retroviruses encoding *f*-like genes as they have been reported to be able to insert into the baculovirus genome (232, 319). However, the direction of horizontal gene transfer is ambiguous given the fact that errantiviruses have been proposed to contain the *env* gene that they acquired from baculoviruses (232).

We further extended the search of F-like proteins in different organisms using the SeFe sequence. A PSI-BLAST search yielded previously unidentified F-like protein sequences in the genomes of a number of eukaryotes, mainly insects (phylum *Arthropoda*). Intriguingly, putative homologous proteins to baculovirus F were found in *Branchiostoma floridae* (Florida lancelet) and *Lottia gigantea* (Giant owl limpet) belonging to phyla *Chordata* and *Mollusca*, respectively. *Branchiostoma floridae* represents one of the closest living invertebrate relatives of vertebrates (76). Although lancelets split from vertebrates more than 520 million years ago, some studies revealed that vertebrates have employed the genes found in lancelets and changed their regulation or even function (76). If vertebrates also contain yet unidentified *f* genes, the *f* gene found in *Branchiostoma floridae* may represent a common ancestor from which they evolved. Our results indicate that *f-like* genes might be even more widespread than previously thought. It is important to recognize, however, that more detailed analysis is required to distinguish if these genes are not part of a previously integrated retrovirus genome. Determination of the structures of F-related cellular proteins would be helpful to elucidate the evolutionary link between viral and cellular F proteins. In addition, it might also help to understand the function of cellular F proteins.

Baculovirus F protein not only functions as a fusion protein but also interacts with unknown cellular receptors at the surface of host cells. In order, to identify residues that potentially could be involved in the receptor binding, we attempted

to identify surface exposed residues conserved among the 39 different baculovirus F proteins. The analysis revealed several well-defined conserved patches on the surface of the SeFet post-fusion trimer. Nevertheless, since the F protein should interact with the cellular receptor before the fusion, the structure of pre-fusion form of F would be more appropriate to identify the conserved surface residues potentially involved in receptor recognition.

Though our EM data of the full-length SeFe ectodomain post-fusion trimer revealed the presence of a 6-helix bundle in this molecule, the crystal structure of the full-length molecule would provide more insights into the fusion mechanism of baculovirus F protein. We have produced several constructs of SeFe allowing us to successfully form the full-length post-fusion trimer. The crystallization of these SeFet variants is underway. Though crystals of the full-length SeFet in its post-fusion form have already been obtained, they diffracted to low resolution (~7 Å). In addition, we observed that the full-length SeFe containing the fusion peptide in which the hydrophobic residues were mutated to hydrophilic ones yields a stable trimer at neutral pH in addition to a monomer. This trimer presumably corresponds to the pre- fusion trimer. Initial crystallization trials of the putative pre-fusion trimer have been performed and the crystals obtained, however, the crystals diffracted only to ~10 Å.

In summary, we have developed the protocols to form a stable trimer of post-fusion baculovirus F ectodomain and determined the crystal structure of trypsin truncated SeFet. The formation of a trimer is induced by low pH treatment, which induces irreversible conformational changes within the protein and the transition from monomeric pre-fusion state to a stable trimeric post-fusion state. In order to prevent the pH triggered aggregation of the protein, the fusion peptide must be removed or hydrophobic residues within the fusion peptide must be replaced by the hydrophilic ones. The crystal structure of trypsin-truncated SeFet confirmed the evolutionary ties between baculovirus and paramyxovirus F proteins. Our results also suggest that solving the structures of F homologues found in eukaryotes is important for a better understanding of the evolutionary link between viral and cellular F proteins.

SUPPLEMENTARY MATERIALS AND METHODS

Cell lines

The Drosophila Schneider 2 (S2) cell line was purchased from Invitrogen. This cell line was originally derived from a primary culture of late stage Drosophila melanogaster S2 embryos. Drosophila Schneider 2 (S2) cells were grown at 28°C in a normal atmosphere in a Schneider's Drosophila medium (Invitrogen, Carlsbad, USA) supplemented with 10% fetal bovine serum (FBS), 50 U/ml penicillin, and 50 μg/ml streptomycin. The cells were maintained applying standard protocols provided by Invitrogen. A remarkable feature of S2 cells is that they are able to integrate a high copy number of plasmids into their genome which renders them suitable for high-level expression of recombinant proteins. Stable cell lines expressing the recombinant proteins were generated by co-transfection of the expression and selection plasmids as described later. Stable cell lines were adapted to, and cultured in serum-free Insect Xpress media (Lonza, Basel, Switzerland) which was also used for protein production. S2 cells grow as a loose, semi-adherent monolayer in tissue culture flasks and are, thus, easily adapted to grow in suspension. For protein production, stable, suspension-adapted cells were propagated in spinner flasks of different sizes (from 1 to 4 l) under continuous agitation at 90 RPM.

Expression vectors

For expression of the recombinant proteins in Drosophila cells, a modified version of the pMT/BiP/V5-His vector (Invitrogen) plasmid (Invitrogen) designated pT350 was used. pT350 contains a double Strep-tag (IBA, http://www.iba-go.com/) with a linker region (GlyGlySer)4 in between for efficient purification of recombinant protein. An enterokinase cleavage site (DDDDK↓X) is introduced upstream of the Strep-tag to allow specific removal of the tag. The vector contains the inducible metallothionein promoter which is induced upon addition of divalent cations (Cu^{2+} or Cd^{2+}). A gene of recombinant protein is cloned in frame with *Drosophila* BiP signal sequence at the N-terminus which serves for translocation of recombinant protein into the endoplasmic reticulum lumen and directs the protein through the secretory pathway of S2 cells into the culture medium.

The modified versions of the pT350 vector were used for expression of recombinant Fab and scFv molecules as previously described (13, 110). The vectors are designated pMT-Fab-Strep and pMT-scFv- Strep, respectively.

Polymerase chain reaction

Polymerase chain reaction (PCR) for amplification of a specific insert for restriction cloning or blunt-cloning was performed using Phusion High-Fidelity DNA polymerase (Finnzymes) as recommended by the producer. Typically the PCR reaction contained: 0.5 μM of each primer (MWG Operon), 50 ng template plasmid DNA, 200 μM dNTPs, 1 U Phusion High-Fidelity DNA polymerase (Finnzymes), 3% DMSO, and 5×Phusion HF Buffer in a total volume of 50 μl.

DNA restriction digest, ligation and transformation

Purified PCR products and appropriate vectors were digested for 1–2 hours with the corresponding restriction enzymes (New England Biolabs, Ipswich, MA) according to the manufacturer's protocols. The digested DNA was verified by agarose gel electrophoresis and purified using Macherey Gel and a PCR Clean-up Kit (Macherey-Nagel). Linearized vectors were de-phosphorylated by Antarctic Phosphatase (NEB) applying a protocol provided by the manufacturer. The DNA vector and insert were ligated (3:1 molar ratio of insert DNA termini to vector DNA) for 1 h at 16–25°C using T4 DNA ligase (New England Biolabs). 2-5 μl of the ligation mixture was used to transform 50 μl of TOP10 DH5α chemically competent cells (Invitrogen) using standard procedures. 50-300 μl of cell suspension was spread on LB-agar plates containing an appropriate antibiotic and incubated for 16–20 h at 37°C. Typically one colony of the transformed DH5α cells was inoculated into 4 ml of LB supplemented with antibiotic and grown with agitation for 16-18 h at 37°C. The plasmid DNA was purified using a NucleoSpin Miniprep Kit (Macherey-Nagel) as described in the kit manual. To verify the presence of an insert the purified plasmid DNA was digested with appropriate restriction enzymes and analyzed by agarose gel electrophoresis. The plasmid DNA containing an insert of the expected size was sequenced for final validation (GATC Biotech AG, Constance, Germany or Sequencing Facility of Institut Cochin, Paris, France).

Production of recombinant proteins in *Drosophila S2* cells

For the large scale production of the recombinant proteins in *Drosophila S2* cells, the corresponding stable cell line was expanded at 28°C to a final volume of 1-4 l in spinner flasks depending on recombinant protein expression level. The cells were induced with 4 μM $CdCl_2$ when they reached the density >8×106 cells/ml. After 6-10 days the cells were pelleted by centrifugation for 15 min at 15 000 × g and the supernatant concentrated using a Vivaflow 200 flip filtration concentration system with 10 kD cutoff membrane (Sartorius Stedim Biotech

GmbH, Goettingen, Germany) to ~50 ml. The pH of the concentrated supernatant was adjusted by adding Tris pH 8 to a final concentration of 100 mM. Avidin was added to 15 μg/ml to prevent binding of the biotin present in the medium.

The supernatant was cleared by centrifugation at 20 000 × g for 30 min and filtered with 0.2 mm cutoff membrane and loaded onto a Streptactin Superflow column (IBA, Gottingen, Germany). The column was washed with 5 column volumes (CV) of 0.1 M Tris pH 8, 0.15 M NaCl, 1 mM EDTA, and a recombinant protein was eluted with 2 CV of the elution buffer containing 2.5 mM desthiobiotin, 0.1 M Tris pH 8, 0.15 M NaCl, 1 mM EDTA. The eluate was further purified by size exclusion chromatography using Superdex 200 or Superdex 75 column (GE Healthcare, Uppsala, Sweden). Typically, SEC was performed at 15 in 10 mM Tris pH 8 and 150 mM NaCl. In the case of SeFe, a monomer was purified by SEC in 5 mM Hepes, 150 mM NaCl, and 0.1 mM EDTA (HNE) buffer. Protein fractions were analyzed by SDS-PAGE. Based on the SDS-PAGE analysis fractions containing >95 % pure protein were pooled and concentrated to approximately 20 mg/ml using a Vivaspin centrifugal concentrator (Sartorius Stedim Biotech GmbH, Goettingen, Germany).

Production of endodeglycosidases and deglycosylation

For expression of PNGase, 3 l of LB medium containing 50 pg/ml ampicillin was inoculated 1/100 with an overnight culture of *E. coli* BL21 (DE3) strain transformed with pBlueScript (Agilent Technologies)-EndoH. After growing the cells at 37°C to an OD550 of 0.5-0.6 the culture was cooled to room temperature. Expression was induced with 0.5 mM isopropyl-I-thio-P-D-galactoside (IPTG) and the culture was incubated overnight at 30°C.
The cells were harvested by centrifugation at 6000 g for 15 min and EndoH was purified from the periplasm of the bacteria. The pelleted cells were resuspended in 5% of the original culture volume in ice cold 20% sucrose, 0.1 M Tris pH 8 and 1 mM EDTA and pelleted again by centrifugation at 12000g for 20 min. Then the cells were resuspended in equivalent volumes of ice-cold water+Protease Cocktail (Sigma) and incubated for 10 min on ice. After pelleting the cells at 16000g for 1h the supernatant was collected, adjusted to 50mM TRIS pH 8.0, 1mM $MgCl_2$, and filter through a 0.22 μM cut-off membrane. The supernatant was loaded onto 5 ml of pre- packed HP Nickel-sepharose (GE Healthcare) at 3.0 ml/min. After washing with 40 mM imidazole, 50 mM Tris 8.0, and 500mM NaCl, the protein was eluted with a two-step gradient with 75 mM imidazole, 50mM Tris 8.0, 500mM NaCl and 300 mM imidazole, 50mM Tris 8.0, 500mM

NaCl. The elution fractions containing EndoH were pooled and further purified by SEC on a Superdex 200 26/60 column (GE Healthcare).

Essentially the same procedure was followed for the expression of EndoH except that IPTG induction was performed at 20 hours at 20°C.

For expression of EndoD, 2 l of LB medium containing 50 pg/ml ampicillin were inoculated 1/30 with an overnight culture of *E. co*li BL21 (DE3) strain transformed with pET28a (Novagen)-EndoD. After growing the cells at 37°C to an OD550 of 1.0 the culture was cooled to room temperature. Expression was induced with 0.5 mM isopropyl-I-thio-P-D-galactoside (IPTG) and the culture was incubated for 20 h at 20°C. The cells were harvested by centrifugation at 4000 g for 15 min and EndoD was purified from the cytoplasm of the bacteria. The pelleted cells were resuspended in 25ml of 40 mM imidazole, 50 mM Tris 8.0, 500 mM NaCl buffer and homogenized by two passages through an Emulsiflex-C5 homogenizer (Avestin). The soluble fraction was separated by centrifugation at 35000 g for 30 min and filtered through 0.22 μM cut-off membrane. EndoD was purified from the supernatant applying the same protocol as for PNGase and EndoH.

Typically, deglycosylation of the native glycoproteins was carried out overnight (approximately 16 hours) at 37°C in 50 mM sodium citrate pH5.5 (Endo H and EndoD) or 50 mM Tris pH8. Initially, the amount of each endodeglycosidase used for deglycosylation of the target proteins was optimized in small-scale reactions at 37°C. The glycoprotein:endodeglycosylase ratio at which maximal deglycosylation was achieved as based on SDS-PAGE analysis was chosen and the reaction was upscaled in a linear manner.

Removal of the Strep affinity tag

A C-terminal Strep tag preceded by an enterokinase recognition site was removed from the recombinant proteins by specific proteolytic cleavage with EKMax Enterokinase (Invitrogen, San Diego, USA). To cleave the tag from the recombinant proteins, typically 1 unit of EKMax enterokinase was added to a protein solution at 0.66 mg/ml in 50 mM Tris pH 8, 50 mM NaCl, 1 mM $CaCl_2$ and the reaction mixture was incubated at 37°C. After 16 h 0.1 mM phenylmethylsulphonyl fluoride (PMSF) was added in order to inactivate the protease. To remove any residual molecules harboring Strep tag, the reaction mixture was loaded on to a Streptactin column and the flow-through containing the protein molecules without the double Strep tag collected. Subsequently, the collected flow-through was loaded onto the Superdex 200 column (GE

Healthcare, Uppsala, Sweden). The fractions containing purified protein without the Strep-tag were pooled and the protein was concentrated to approximately 25 mg/ml in 10 mM Tris pH 8 and 150 mM NaCl buffer.

SDS-PAGE and Western blot analysis

For electrophoresis under reducing and non-reducing conditions, samples in Laemmli sample buffer (125 mM Tris-HCl, 2% sodium dodecyl sulfate (SDS), 10% glycerol and 0.001% bromophenol blue, pH 6.8) with or without 100 mM DTT were denatured for 5 min at 98°C. Protein samples were subjected to SDS-PAGE electrophoresis using standard methods (Sambrook and Russell, 2000) and stained with Coomassie Blue (Bio-Rad) or subsequently transferred to PVDF membranes (Bio-Rad) by wet electrophoresis transfer (Bio-Rad).

The membrane was blocked for 1h at room temperature in PBS-T (PBS with the addition of 0.1% Tween 20) containing 5% dry milk. For detection of strep-tagged proteins, the membrane was first incubated for 1 h at room temperature with mouse- anti-Strep IgG (*Strep*MAB-Classic, IBA) for 1h at a dilution of 1:3000 in blocking buffer. After washing the membrane 3 times with PBS-T for 5 min, rabbit anti-mouse IgG conjugated with horseradish peroxidase was added at a dilution of 1:10000 in blocking buffer and the membrane was incubated for 1h. After washing the membrane three times for 15 min each time in PBS-T and once for 15 min in PBS, proteins were detected using the ECL kit (GE Healthcare).

DTNB-thiols assay

Free cysteine residues in proteins were detected using the DTNB-thiols assay, which measures sulfhydryl groups with 5-5dithiobis [2nitrobenzoic acid] (DTNB) reagent forming a measurable yellow-colored product when it reacts with sulfhydryls. 750 µM DTNB in 100 mM Tris pH8.0 was used as a working solution. 5 µl of L- cysteine standards (500 µM, 750 µM, 1 mM 1.25 mM, 1.5 mM, 1.75 mM, 2 mM, 2.5 mM, 3 mM and 3.5 mM) were mixed with 160 µl working solution and 35 µl of buffer (10 mM Tris, 150 mM NaCl) in microtiter wells and incubated at room temperature for 10 min in order to obtain a standard curve. Similarly, 40 µl of diluted protein sample was mixed with 160 µl working solution and incubated at room temperature for 10 min. After 10 min absorbance of the solutions was measured at 412 nm. Absorbance values of L-cysteine standards were plotted versus concentration to generate a standard curve. The concentration of the sulfhydryl groups in protein samples was estimated from the standard curve.

Screening for crystallization conditions

To assess the appropriate protein concentration for crystallization screens, a pre-crystallization experiment using PCT Pre-Crystallization Test kit (Hampton Research) was performed for each new protein sample. Initial crystallization trials were carried out in 96-well crystallization plates at 293 K using the sitting drop-vapor diffusion method. Liquid nanoliter scale pipetting was performed using a Mosquito robot (TTP LabTech Ltd, Royston, UK). Sitting drops contained 200 nl protein solution and 200 nl reservoir solution. To identify initial crystallization conditions screening initially was performed using the commercially available crystallization screening kits listed below:

- Crystal screen 1 and 2 (Hampton Research, HR2-110 and HR2-112)
- Wizard I and II (Jena Biosciences, EBS-WIZF)
- JBS Screen 1-8 bulk (Jena Biosciences, CS-101L-108L)
- Structure Screen 1 and 2 (Molecular Dimensions, MD1-01 and MD1-02)
- PEG/Ion Screen (Hampton Research, HR2-126)
- Crystal screen CRYO (Hampton Research, HR2-122)

In total, 576 different reservoir solutions were tested during initial screening. Crystallization plates were placed and sitting drops were regularly imaged using a Rock Imager 1000 (Formulatrix, MA, USA). Crystallization conditions for SeFe as well as DAO5 Fab and DAO5 scFv in complex with the peptides are provided in the Results chapter.

Optimization of crystallization conditions

Multiple, small or poorly X-ray diffracting crystals obtained in 96 well plates during initial screening were further optimized using the hanging drop method in 24- well VDX plates (Hampton Research). Typically, drops of 2 µl (containing 1 µl of protein solution and 1 µl of reservoir solution) were placed on siliconized cover slides (Hampton Research) over 0.75 ml well solution. Optimization was performed around the conditions that yielded crystals during initial screening. Optimization strategies included: 1) varying precipitant and/or other compounds concentration, 2) varying the pH, 3) screening with different protein concentrations, 4) performing crystallization at different temperatures, 5) using different compounds from Hampton Additive Screens 1–3 (Hampton Research) as additives to original crystallization condition (in 96 well plates), and 6) using seeding. Different methods of seeding are described in the following section.

Seeding techniques

It is believed that the optimal conditions for crystal nucleation can be different from the ones optimal for crystal growth (163). Introduction of the seeds into crystallization drops provides a preformed crystal surface which may further promote the ordered assembly of molecules at a lower degree of supersaturation than needed for nucleation (24). Seeding approach has been demonstrated to be efficient in crystal optimization or promoting crystal growth (179, 357).

In order to apply seeding techniques, seed stocks were prepared using the crystals from 4-6 crystallization drops. The crystals were transferred to a tube containing 50-100 μl of reservoir solution. Several 425-600 μm glass beads (Sigma) were added and the crystals were crushed by stirring for 5-10 min using a laboratory vortex. The seed-stocks were stored at 4°C. Three seeding techniques (direct seeding, streak seeding, and microseed matrix seeding) were applied.

1. In the case of direct seeding, serial 10-fold dilutions were prepared from the seed stock. The crystallization drops were set up by mixing 1 μl of reservoir solution, 0.5 μl of diluted seeds, and 1 μl of protein solution. The seed stock dilution yielding the best quality crystals was chosen for growing crystals for soaking or/and X-ray diffraction analysis.

2. Streak seeding was performed using a cat whisker on the drops pre-equilibrated for at least 8 h.

3. Microseed Matrix Screening (MMS) is a seeding technique where the seed of the crystals grown in one condition are seeded into new conditions as part of the screening procedure (70). This technique has been successfully applied to promote crystallization of different macromolecules (152) as well as antigen-antibody complexes (279). MMS can be further subdivided into self-seeding MMS (crystals of the same protein are used as seed-stock) and cross- seeding MMS (crystals of the related protein - a mutant or a homologous protein - are used as seed stock) (396). Typically, MMS was performed by adding seeds during screening with the commercial screening kits described above. The drops consisted of 200 nl protein solution, 200 nl reservoir solution, and 100 nl of solution containing seeds. The specific application of seeding techniques is described in a chapter Results.

Cryo-protection and freezing of crystals

Cryo-protection of protein crystals and cooling during data collection is necessary to reduce radiation damage when crystals are irradiated with high intensity X-ray sources [reviewed in (104)]. Good cryoprotectants prevent the formation of

molecules. The cryo-protection method was optimized depending on the crystal growth conditions. 20% glycerol (v/v) in mother liquor was used for cryoprotection of a majority of the crystals. Crystals growing in conditions containing 20-30% PEG400, MPD or glycerol was frozen using a mother liquor as a cryo-protectant solution. Cryo-cooling procedures typically included three steps: 1) crystals were transferred from the crystallization drops to the drops containing cryo-protectant solution using a nylon cryoloop, 2) the crystals were kept in the cryoprotectant solution for several seconds, and 3) the crystals were fished using a cryoloop and flash frozen by plunging into liquid nitrogen (-196 °C).

Chapter 4

Identification and characterization of a baculovirus F fusion protein homolog in *Anopheles darlingi*

Qiushi Wang[1,2], Just M. Vlak[2], Vera I. D. Ros[2], P. J. M. Rottier[1], Berend Jan Bosch[1]

[1] Virology Division, Department of Infectious Disease and Immunology, Faculty of Veterinary Science, Utrecht University, Utrecht, the Netherlands

[2] Laboratory of Virology, Wageningen University, Droevendaalsesteeg 1, 6708 PB, Wageningen, the Netherlands

Manuscript in preparation

ABSTRACT

Baculovirus F proteins are essential envelope proteins that mediate receptor binding and low pH-triggered membrane fusion enabling virus entry into host cells. Here we report on a novel cellular protein encoded by the mosquito *Anopheles darlingi* (family Culicidae) genome with amino acid sequence homology and possible functional similarity to baculovirus F proteins. The *Anopheles darlingi* cellular protein, named Ad-F contains an N-terminal signal peptide, a furin cleavage site, predicted coiled-coil domains, a conserved motif and a partly conserved fusion peptide. Protein structure modeling showed that Ad-F and baculovirus F congrue in a similar fashion with the respiratory syncytial virus F protein. Phylogenetic analysis and PSI-BLAST search result suggest that Ad-F has evolutionary links with baculovirus F proteins as well as insect endogenous retrovirus envelope proteins. Similar to baculovirus F proteins, Ad-F protein transiently expressed in mammalian cells was able to be cleaved by the Golgi-resident furin protease, indicative of its proper folding and transport through the exocytic route. The Ad-F protein was incorporated as oligomers into vesicular stomatitis virus lacking its own G protein encoding gene (VSVΔG). These Ad-F pseudotyped VSV particles appeared not infectious on mosquito cells; yet, syncytia were induced when overexpressing the Ad-F protein in baby hamster kidney cells. These observations provide the first indications for the domestication of a baculovirus F gene in mosquito, with preservation of the protein's fusion function.

INTRODUCTION

Baculoviruses are a large family of enveloped, double stranded DNA viruses that are mainly pathogenic to members of diverse orders within the class Insecta (134, 255). These viruses are classified into four genera: *Alphabaculovirus* (lepidopteran-specific), *Betabaculovirus* (lepidopteran-specific), *Gammabaculovirus* (hymenopteran-specific) and *Deltabaculovirus* (dipteran-specific). Baculoviruses have two structurally different, but genetically identical virion phenotypes, the budded virion (BV) and the occlusion derived virion (ODV). BV and ODV are responsible for the initial infection in the gut and for systemic infection, respectively. Both viruses gain entry into the host cell after fusion of the viral and cellular membranes. The fusion process of BV of most baculoviruses is mediated by the envelope fusion protein F, present on the virion surface. The F protein is a class I viral fusion protein that - similar to the fusion proteins from influenza virus, paramyxovirus and retrovirus – trimerizes upon folding and requires proteolysis by endogenous furin to prime the fusion function. The

cellular furin protease cleaves the F protein into two subunits, F_2 and F_1, which remain connected via a disulfide bond. Cleavage generates a hydrophobic N-terminus of the membrane anchored F_1 subunit, which is the fusion peptide (150, 216, 411). Upon fusion activation the fusion peptide inserts into the host cell membrane and the fusion protein undergo structural rearrangement. At the final stage of fusion the F protein forms a 6-helix bundle composed of a central trimeric N-terminal α-helical coiled coil surrounded by three C-terminal helices (415).

Genes encoding fusion proteins have been frequently transferred during evolution between viruses and between viruses and their hosts (232, 356). The best-studied example is a human endogenous retroviruses envelope gene, called *syncytin*, that, after its endogenization maintained its original membrane fusion activity and functions in syncytiotrophoblast formation during placenta morphogenesis (254). Baculoviruses belonging to the group II *Alphabaculovirus* genus and to the genera *Betabaculovirus* and *Deltabaculovirus* have an F protein as fusion protein. In group I of the Alphabaculovirus genus the F protein has been functionally replaced by the GP64 protein (225, 398). The GP64 proteins share high amino acid sequence identity (>74%) (295) and are closely related to the envelope fusion protein of a tick-borne orthomyxovirus, thogoto virus (~30% amino acid sequence identity) suggesting common ancestry (267). The baculovirus F proteins are relatively divergent with low amino acid sequence identity (20-40%), which may suggest they have a long evolutionary history (295). Intriguingly, baculovirus F homologs are found in endogenous retroviruses, retroviral elements and as cellular genes in insect hosts. Phylogenetic studies suggest that the baculovirus F proteins are closely related to *gypsy* retrovirus (also known as errantivirus) envelope proteins (*Bel/Pao* family in LTR retroelements) (210, 232, 319), which have maintained their fusogenic properties, as indicated by their ability to mediate low-pH induced cell-cell fusion (261). The *gypsy* retrotransposon *TED* has been found in the genome of baculovirus mutant FP-D (few polyhedra) from *Trichoplusia ni* (Lepidoptera) (96, 198, 289). This is an example of a recent acquisition of a host gene mediated by a retrotransposon and indicates how gene transfer of baculovirus *F* homologs may have occurred during evolution. In addition to the relatedness between the baculovirus F and errantivirus envelope proteins, a number of cellular homologs were identified in the *Drosophila* and mosquito genomes (231). The genes of these homologs (named *Iris* or *CG4715*) were argued to be domesticated from the *Kanga/roo* lineage of retrotransposons at least 25 million years ago and were stably maintained in the host genome ever since, suggesting a cellular function for

this domesticated gene. However, due to the absence of functionally important features found in F fusion proteins including a proteolytic cleavage site and a classical fusion peptide, the F homologs in *Drosophila* are thought to lack fusion activity (222).

During PSI-BLAST homology searches using the F protein of *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV, group II *Alphabaculovirus*) as a query we identified a cellular baculovirus F protein homolog from a recently genome-sequenced mosquito species, *Anopheles darlingi* (237), provisionally named Ad-F. We performed sequence analysis and phylogenetic studies to determine the interrelationships of Ad-F with baculovirus F, insect retrovirus envelope and Iris proteins. We characterized the expression, folding and furin cleavage of Ad-F protein after transient expression in a mammalian expression system. Furthermore, we investigated the fusogenic properties of the Ad-F protein by examining its potential to mediate cell-cell fusion and to function as a substitute for a genuine viral fusion protein in virus entry. Our results suggest that the Ad-*F* gene encodes a properly folded protein with possibly membrane fusion activity.

MATERIAL AND METHODS

Cells and viruses

Spodoptera frugiperda (Sf9) cells were cultured at 27°C in Sf-900 II SFM medium (Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Bodinco). Both C6/36 (isolated from *Aedes albopictus*) cells (ATCC) and Human Embryonic Kidney (HEK) 293T cells were cultured in DMEM (Lonza) supplemented with 10% FBS at 37°C with 5% CO_2. Baby hamster kidney (BHK) cells were cultured in Glasgow minimum essential medium (GMEM) (Life Technologies) supplemented with 4% tryptose phosphate broth, 1% minimum essential medium nonessential amino acids (MEM NEAA) (Life Technologies), and 5% FBS at 37°C with 5% CO_2. VSVΔG/GFP-G* (kindly provided by Dr. Y. Matsuura), generated and described previously (370) is a recombinant vesicular stomatitis virus (VSV), in which the coding sequence for the envelope fusion protein G has been replaced by a GFP sequence. This VSV vector was used to generate VSV virions pseudotyped with Ad-F, SeMNPV-F, *Autographa californica* nucleopolyhedrovirus (AcMNPV, group I *Alphabaculovirus*) GP64 or VSV-G on their surface (see below).

Plasmid construction

To construct plasmids expressing the Ad-F protein with a affinity tag in 293T and C6/36 cells the codon-optimized Ad-F cDNA (synthesized, Genscript)

appended with a 3'-end HA-tag was cloned into the pCAGGS (138) and pIB-GFP expression vectors separately. The constructed expression vectors were designated as pCAGGS-AdF-HA and pIB-AdF-HA. Plasmids expressing HA-tagged SeMNPV-F protein and AcMNPV-GP64 in 293T and BHK cells were constructed previously and designated as pCAGGS-SeMNPVF-HA and pCAGGS-GP64-HA, respectively (Chapter 2).

Production of pseudotyped vesicular stomatitis virus

The schematic representation of the genome structure and the production of pseudotyped vesicular stomatitis virus (VSVpv) are shown in Fig. 1. To generate VSVΔG viruses pseudotyped with VSV-G or Ad-F proteins 293T cells were transfected with pCAGGS plasmids encoding these fusion proteins using polyethylenimine (PEI, Sigma-Aldrich). At 48 h post transfection (p.t.) the transfected cells were infected with VSVG-pseudotyped viruses (VSVΔG/GFP-G*) at a multiplicity of infection (moi) of 0.1 plaque-forming unit per cell, in which the VSV-G ORF was replaced with the GFP ORF in the genome and the VSV-G protein was pseudotyped on the viral envelope. After the viruses harvested from the supernatant were adsorbed for 2 h at 37 ℃ the cells were washed three times with serum-free DMEM and subsequently incubated with DMEM supplemented with 1% FBS for 24 h at 37 ℃. The culture medium with generated virus particles was centrifuged to remove cell debris and used for inoculating of target cells.

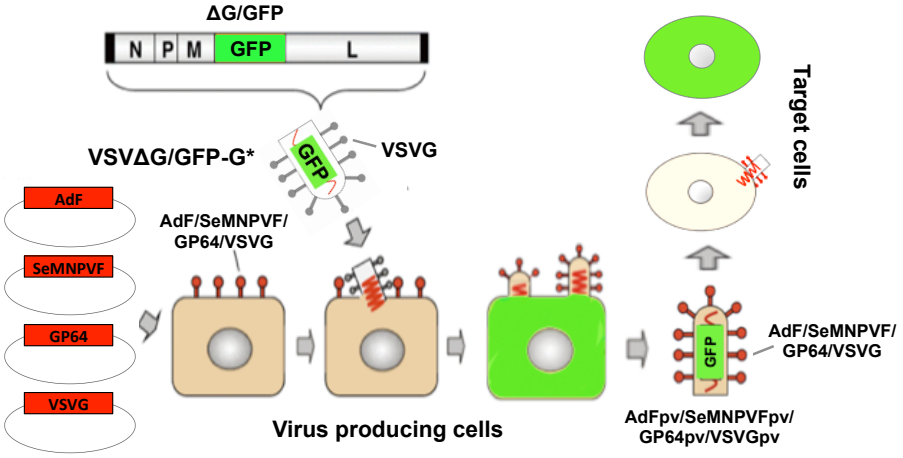


Fig. 1. Schematic representation of the production of pseudotyped VSV viruses. HEK 293T were transfected with an expression plasmid carrying the ORF of AdF or VSV-G, SeMNPV-F or GP64 and then infected with the pseudotyped VSVΔG/GFP-G*. The newly generated VSV pseudotyped viruses (VSVpv) released from the producer cells were examined for the ability to enter target cells by measuring GFP expression. [Figure modified from (166)].

Cell entry assay

To determine the fusogenic activity of Ad-F, various cells (C6/36, 293T and BHK) were inoculated with VSV pseudotyped with Ad-F (described above) for 2 h at 27°C for C6/36 cells and 37°C for 293T and BHK cells in serum-free DMEM medium. The cells were washed with fresh serum-free DMEM medium and then incubated with DMEM medium supplemented with 1% FBS for 24 h at 27°C for C6/36 cells and 37°C for 293T and BHK cells. Subsequently the cells were washed with PBS and suspended after incubating with FACS buffer (2% FBS, 5 mM EDTA and 0.02% NaN_3 in PBS) at room temperature for 10 min. The dispersed cells were fixed with 3.7% paraformaldehyde in PBS at 4°C for 10 min and washed twice with PBS for 10 min and once with FACS buffer at room temperature. Virus entry was measured by a FACSCantoII flow cytometer (BD Bioscience) provided with FACSDiva software (BD Bioscience).

Western blot analysis

For electrophoresis under reducing and non-reducing conditions, samples were taken up in Laemmli sample buffer (125 mM Tris-HCl, 2% sodium dodecyl sulfate (SDS), 10% glycerol and 0.001% bromophenol blue, pH 6.8) with or without 5% β-mercaptoethanol, respectively. Samples were denatured for 5 min at 95°C. Proteins were subjected to SDS-PAGE electrophoresis and subsequently transferred to nitrocellulose membranes (Bio-Rad) by wet electrophoresis transfer (Bio-Rad). The membrane was incubated for 1 h at room temperature with monoclonal anti-HA tag antibody (Sigma-Aldrich) raised from mouse at a dilution of 1:2000 in blocking buffer and subsequently with VRDye 490 Goat anti-Mouse IgG (LI-COR) at a dilution of 1:15,000 in blocking buffer with three times washing between the incubations of two antibodies. After washing the membrane once for 15 min in PBS-T and three times for 15 min in PBS, proteins were detected using the Odyssey Fc infrared image system (LI-COR) and analyzed using Image Studio 3.1 software (LI-COR).

Syncytium formation assay

BHK cells were seeded in 6-well plates and transfected with plasmids encoding Ad-F, AcMNPV-GP64 or SeMNPV-F using FuGENE 6 transfection reagent (Promega). After incubation at 37°C for 48 h, cells were exposed to PBS at pH 5.0 for 2 min and then incubated in growth medium for 4 h. To visualize the nuclei, the cells were permeablized with 96% ethanol and stained with DAPI (Life Technologies) at a dilution of 1:3000 in PBS at room temperature for 20 min. The cells were then washed once with 10% FBS in PBS-T (0.1% Tween 20 in PBS) at room temperature for 15 min. Cells were then incubated with

monoclonal anti-HA tag antibody (Mouse) at a dilution of 1:200 at room temperature for 1 h. Cells were washed three times with PBS-T for 15 min and incubated with Alexa Fluor488 goat anti-mouse IgG (Life Technologies) at a 1:200 dilution in PBS for 1 h. After washing three times with PBS, fluorescence was observed with an EVOS inverted fluorescence microscope (AMG).

Bioinformatic analysis

To obtain all homologous sequences to Ad-F we used PSI-BLAST analysis with four iterations of a search using Ad-F and SeMNPV-F as query sequences separately. The various homologous sequences obtained by PSI-BLAST were aligned using MAFFT (L-INS-i) version 7 (168) with BLOSUM 45 scoring matrix and a gap penalty of 1. The aligned sequences were trimmed to get a conserved domain based on sequence region selection in Malik and Henikoff, 2005 (231). Subsequently the protein alignment was converted to the corresponding codon alignment using PAL2NAL (364). PAUP* version 4.0b10 (58) was used to select the optimal evolution model, as described in Ros *et al*., 2009 (321). Maximum likelihood analysis (heuristic search, 100 bootstrap replicates) was performed in PAUP, using a submodel of the General Time Reversible Model with invariable sites and a gamma distribution of rate heterogeneity (GTR + I + G) with rate class 'abcadc'. Bayesian inference was conducted using MrBayes 3.2.2 (147), using the GTR + I + G model (default settings, six million generations, burn-in of 25%). Alignments are presented using BoxShade 3.21, written by K. Hofmann and M. Baron. Signal peptide prediction was performed with Signal P 4.1 server (300) and transmembrane domain prediction was performed with TMHMM server v 2.0 (183, 346). Coiled coil domain prediction was performed with the COILS program (227). Glycosylation sites are predicted in NetNGlyc server (cbs.dtu.dk/services/NetNGlyc). Fusion peptides were analyzed using WheelApp applet (cti.itc.virginia.edu/~cmg/Demo/wheel/wheelApp.html).

RESULTS

Domain conservation and phylogenetic analysis of F homologues

Homology searches using the PSI-BLAST algorithm with the SeMNPV-F protein as a query sequence yielded several non-baculovirus F homologs including cellular proteins encoded in the genomes of *Drosophila* and *Anopheles* species as well as *gypsy* endogenous retroviruses envelope proteins encoded in the genome of *Drosophila* and *Trichoplusia* species. The cellular proteins are encoded by *Iris* (also called CG4715) genes, which were indicated to be derived from retrovirus envelope genes but not associated with other retroviral elements (231). The genes

encoding the *gypsy* endogenous retrovirus envelope proteins are downstream of gag and/or pol genes in an insect pro-retrovirus genome (17, 198, 238), which includes *Cruiser*, *Tirant*, *17.6*, *297*, *Tom*, *Idefix*, *Gypsy* from *Drosophila* species and *TED* from *Trichoplusia ni*. We identified a cellular protein from *Anopheles darlingi* – designated as Ad-F – as a new F protein homolog. The alignment of SeMNPV-F and Ad-F proteins demonstrates 19% sequence identity with 65% query coverage. Using the Ad-F protein as a query for PSI-BLAST search we retrieved a cellular protein encoded in the *Aedes aegypti* genome with only 26% query coverage and 36% protein sequence identity. PSI-BLAST search using Ad-F as a query did not retrieve *gypsy* retrovirus envelope sequence but an envelope protein encoded in another endogenous retrovirus, *roo* (also named as *B104*, Bel/*Pao* family of LTR retroelements) from *Drosophila melanogaster* (210).

The Ad-F protein shares a domain of unidentified function – the DUF3609 motif (pfam accession no.: PF12259) – with the Iris, *roo* F homologs, and most baculovirus F proteins. This domain is only found in Arthropoda and Baculoviridae (1). The *gypsy* retrovirus envelope proteins lack a DUF3609 domain and contain a different conserved domain, *gypsy* (pfam accession no.: PF07253), which mainly presents in *gypsy* retroviruses. Yet, the DUF3609 and *gypsy* domains appear to be related with 26.6% amino acid identity, 38% similarity and an absolute identical motif (Fig. S1).

Surprisingly, the Ad-F gene in the genome is devoid of any introns, which suggests that the Ad-F gene is not an authentic gene in the *A. darlingi* genome and may have been domesticated more recently. The Ad-F ORF is not flanked by retroviral sequences and was not annotated as part of a retrotransposon in the genome-sequencing project of *A. darlingi* (237). In order to explore the relationship of the Ad-F coding gene with baculovirus F, *Iris* and insect retrovirus *env* genes we performed a phylogenetic analysis for a selection of F homologs listed in Table 1. In the resulting phylogenetic tree Ad-F has a clear linkage with *Drosophila Iris* genes and likely represents an *Iris* ortholog (Fig. 2). Furthermore, there are 12 residues of Ad-F protein that are highly conserved among all the lepidopteran baculovirus F proteins (higher panel, Fig. 3) and 34 residues of Ad-F that are fully conserved among *Iris* proteins (lower panel, Fig. 3), indicating a much higher level of sequence conservation of Ad-F with *Iris* compared to baculovirus F. Most of the conserved residues are concentrated in the DUF3609 domain; all the cysteines in this domain are conserved and this might allude to an important role of this domain in protein folding.

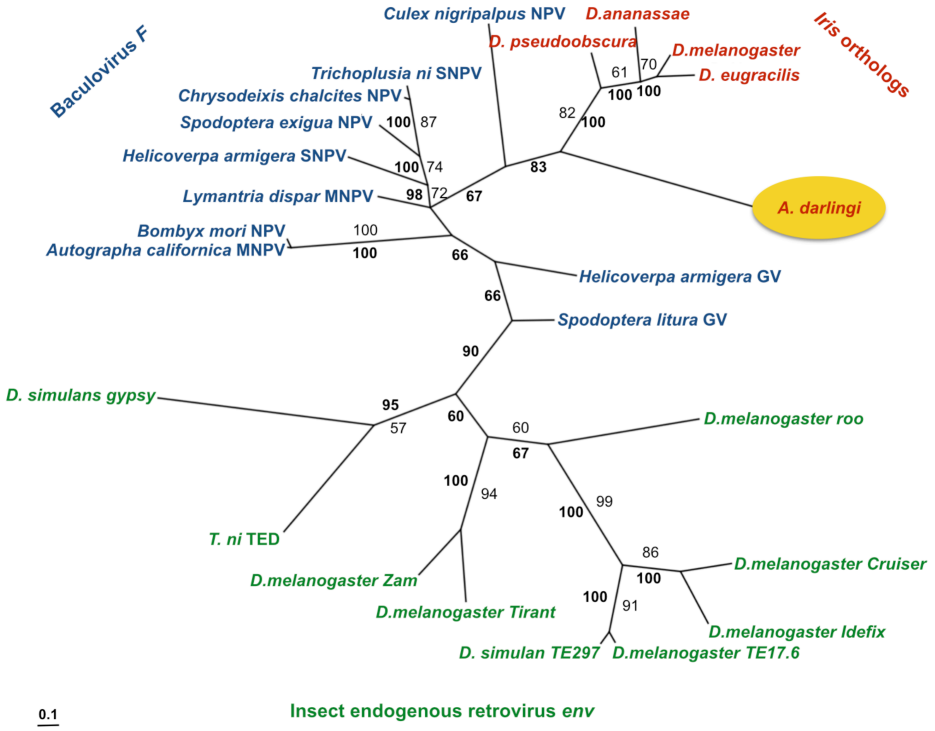


Fig. 2. Evolutionary relationships of Ad-F with other F-homologs. The core domains of F related genes were aligned and a Bayesian phylogenetic tree was generated. Numbers in bold indicate Bayesian posterior probabilities and plain numbers are maximum likelihood bootstrap values based on 100 replicates. Only values ≥ 50 are shown for both analyses. The bar at the bottom indicates a branch length of 10% distance. Three groups are color-coded: baculovirus *F* (blue), *Iris* orthologs in *Drosophila* and *Anopheles* genomes (red) and insect endogenous retrovirus envelope (green) genes. The Ad-*F* gene (indicated with a yellow oval) groups with *Drosophila Iris* sequences.

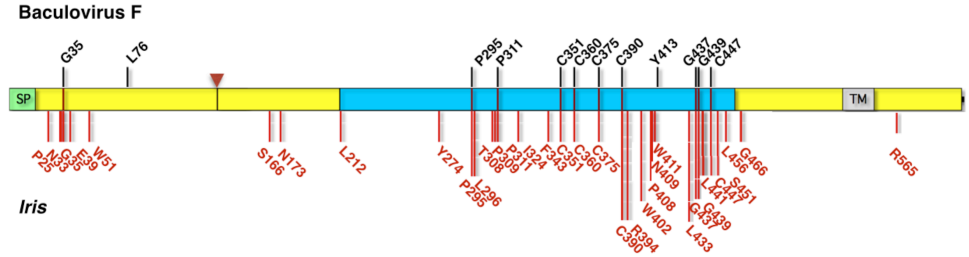


Fig. 3. Schematic view of conserved residues in Ad-F protein. Ad-F protein sequence (shown as linear domain map) was aligned with homologs of baculovirus F and Iris proteins respectively. Ad-F residues conserved in baculovirus F are indicated by black lines on top of the Ad-F linear map. Residues conserved in *Iris* proteins are presented as red lines beneath the linear map. The common conserved residues in Ad-F, baculovirus F and Iris are shown as dark red lines on the Ad-F map. The signal peptide (SP, green), potential furin cleavage site (red arrow head), DUF3609 (cyan-blue) and transmembrane (TM, gray) of Ad-F protein are indicated.

Table 1. Sequences used for analysis and their domain conservation.

Source[a]	Protein	Accession	Conserved domain	FCS	FP[b]
Baculovirus fusion proteins					
Spodoptera exigua MNPV	F	GI:9634229	DUF3609	+	+
Lymantria dispar MNPV	F	GI:566568075	DUF3609	+	+
Chrysodeixis chalcites NPV	F	GI:68304286	DUF3609	+	+
Trichoplusia ni SNPV	F	GI:74229828	DUF3609	+	+
Helicoverpa armigera SNPV	F	GI:209401195	DUF3609	+	+
Autographa californica MNPV	F	GI:9627765	-	-	-
Bombyx mori NPV	F	GI:393717207	-	-	-
Helicoverpa armigera GV	F	GI:164519221	-	-	+
Spodoptera litura GV	F	GI:148368845	DUF3609	+	+
Culex nigripalpus NPV	F	GI:15320898	DUF3609	+	-
Dipteran, *gypsy* retrovirus envelope proteins					
D. melanogaster	*Cruiser* Env	GI:14010626	Gypsy	+	+
D. melanogaster	*Tirant* Env	GI:60593116	Gypsy	+	+
D. melanogaster	*17.6* Env	GI:140836	Gypsy	+	+
D. melanogaster	*297* Env	GI:119381	Gypsy	+	+
D. ananassae	*Tom* Env	GI:422416	Gypsy	+	+
D. melanogaster	*Zam* Env	GI:2791287	Gypsy	+	+

D. melanogaster	*Idefix* Env	GI:4165195	Gypsy	+	+
D. subobscura	*Gypsy* Env	GI:119377	Gypsy	+	+
Dipteran, *roo* retrovirus envelope protein					
D. melanogaster	*roo* Env	GI:27368143[c]	DUF3609	+	+
Dipteran, *Iris* orthologs					
D. eugracilis	Iris	GI:76881719	DUF3609/DUF3610	-	-
D. ananassae	Iris	GI:76881713	DUF3609/DUF3610	-	-
D. pseudoobscura	Iris	GI:76881711	DUF3609/DUF3610	-	-
D. melanogaster	Iris	GI:77021956	DUF3609/DUF3610	-	-
A. darlingi	F	GI:568259328[d]	DUF3609	+	+/-
Lepidopteran, *gypsy* retrovirus envelope protein					
T. ni	TED Env	GI:102940	-	+	+

[a] Abbreviations of source names: MNPV, multiple nucleopolyhedrovirus; SNPV, single nucleopolyhedrovirus; GV, granulovirus; *D.*, *Drosophila*; *A.*, *Anopheles*; T., *Trichoplusia*.

[b] The presence of fusion peptide (FP) is predicted based on helical wheel structure (see materials and methods and Fig 6).

[c] The *roo* retrovirus envelope protein is part of a larger protein, covering residues 1787-2360.

[d] The protein includes an N-terminal tyrosine kinase and a C-terminal of Ad-F.

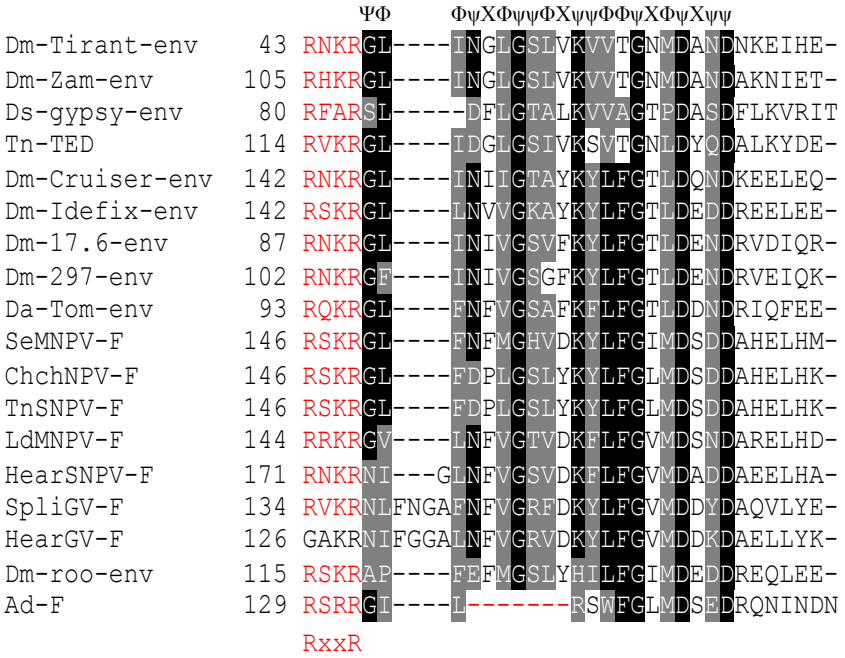


Fig. 4. Sequence alignment of the protease cleavage site and downstream fusion peptide of F protein homologs. The Iris, AcMNPV-F, BmNPV-F and CuniNPV-F proteins are not shown due to their large sequence divergence in the aligned region. Highly and moderately conserved residues in the fusion peptide alignment are indicated in black and grey, respectively. The RxxR furin cleavage site sequences are present in all the F homologs (indicated in red) with the exception of HearGV-F. Amino acid characteristics are shown above the alignment; Φ and ψ represent conserved hydrophobic and hydrophilic residues respectively, X represents any residue. Gap positions of Ad-F are shown as red dashes.

Ad-F contains a furin proteolytic cleavage site and a short, putative fusion peptide.

Next we looked into elements of the Ad-F protein that are highly conserved in baculovirus F proteins. Alignment of the homologs listed in Table 1 reveals that - like most baculovirus F and *Drosophila* retrovirus envelope proteins - Ad-F contains a furin cleavage site (FCS) sequence (RxxR, R=arginine and x=any amino acid) (Fig. 4) (39). Most baculovirus F proteins possess a hydrophobic 21 amino acid long segment, named fusion peptide (FP), located directly downstream of the FCS (49, 63, 66). Putative FPs were also found in *roo* and *gypsy* retrovirus envelope proteins based on sequence alignment (38) (Fig. 4). However, even though 12 out of 21 residues are conserved or with the same hydrophobicity at corresponding positions, the putative FCS contiguous sequence in Ad-F contains a deletion of 7 residues including 3 hydrophobic

residues compared to FPs in the majority of baculovirus F and *gypsy* retrovirus envelope proteins (Fig. 4).

Most fusion peptides of class I fusion proteins are characterized by an amphipathic helix. We examined the hydrophobic nature of the 18-amino-acid long region downstream of the FCS of the Ad-F and baculovirus F proteins using alpha helical wheel plotting (see Material and Methods). We found variation in the hydrophobicity pattern of helices for the fusion peptides in baculovirus F proteins (Fig. 5B, C and D). The hydrophobic residues in Ad-F (Fig. 5A) are concentrated on one side of the helical wheel, similarly as in SpliGV-F. However, the hydrophobic nature is relatively weak by the occurrence of three hydrophilic residues: Asp, Asn and Glu. The weak hydrophobicity and the occurrence of the 7-residue gap in the FP sequence, suggests the absence of a selective pressure on the FP characteristics during evolution.

Prediction of the Ad-F protein structure

The Ad-F protein contains a proteolytic cleavage site, a putative fusion peptide, coiled-coil domains and a transmembrane domain (Fig. 6). The protein structure may provide further insights into the Ad-F protein's function. Because the crystal structures of both Ad-F and baculovirus F protein were not available, we modeled the tertiary structures of Ad-F and SeMNPV-F proteins using PHYRE2 protein fold recognition server. Profile-to-profile alignment in PHYRE2 using Ad-F and SeMNPV-F as queries identified the respiratory syncytial virus F protein (RSV-F) as the highest scoring template for modeling these two proteins with high precision scores (AdF: 83% and SeMNPV-F: 82.8%). The modeled regions of Ad-F and SeMNPV-F cover the predicted heptad repeat (HR) region downstream of their furin cleavage site (Ad-F: HRA and SeMNPV-F: HR1-HR3) and the conserved DUF3609 domain (Fig. 6A) corresponding to the core domain of the postfusion RSV-F protein, including part of the stalk (HRA region) and head structure (region between HRA and HRB) (Fig. 6B). In the RSV-F protein HRA is adjacent to the fusion peptide (FP), and HRB is proximal to transmembrane domain (TM). The head structure connects the two HR regions. HR regions in RSV-F and other paramyxoviruses fusion proteins undergo dramatic structural rearrangements, which are modulated by the head structure after fusion activation. HRB folds back to HRA to form a hairpin-like structure (188, 246). This conformational change places the two membranes that FP and TM anchor into close proximity. Although the HR regions of Ad-F and SeMNPV-F corresponding to RSV-

HRB region are absent in their modeled structures a heptad repeat region was predicted at C-terminal region of each protein sequence, similar to the position to RSV-HRB (Ad-F: HRB and SeMNPV-F: HR2 in Fig. 6A). The high similarities of the structural models suggest that Ad-F much resembles baculovirus SeMNPV-F, calling for a functional analysis of Ad-F.

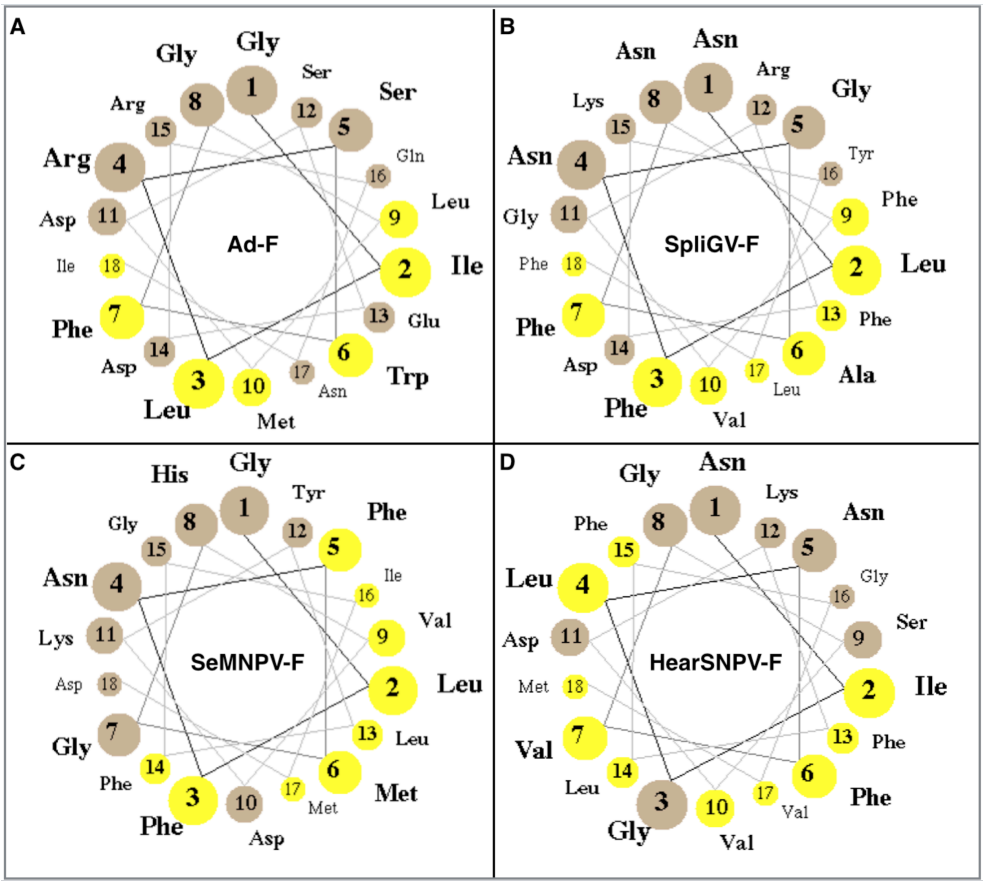


Fig. 5. Predicted helical structures of fusion peptides. The F homolog structures of *Anopheles darlingi* (Ad-F, A) and three baculoviruses, *Spodoptera litura* granulovirus (SpliGV-F, B), *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV-F, C) and *Helicoverpa armigera* single nucleopolyhedrovirus (HearSNPV-F, D) were determined using WheelApp applet (http://cti.itc.virginia.edu/~cmg/Demo/wheel/ wheelApp.html). The first 18 residues downstream of the furin cleavage site are plotted. Hydrophobic and hydrophilic residues are labeled in yellow and brown, respectively.

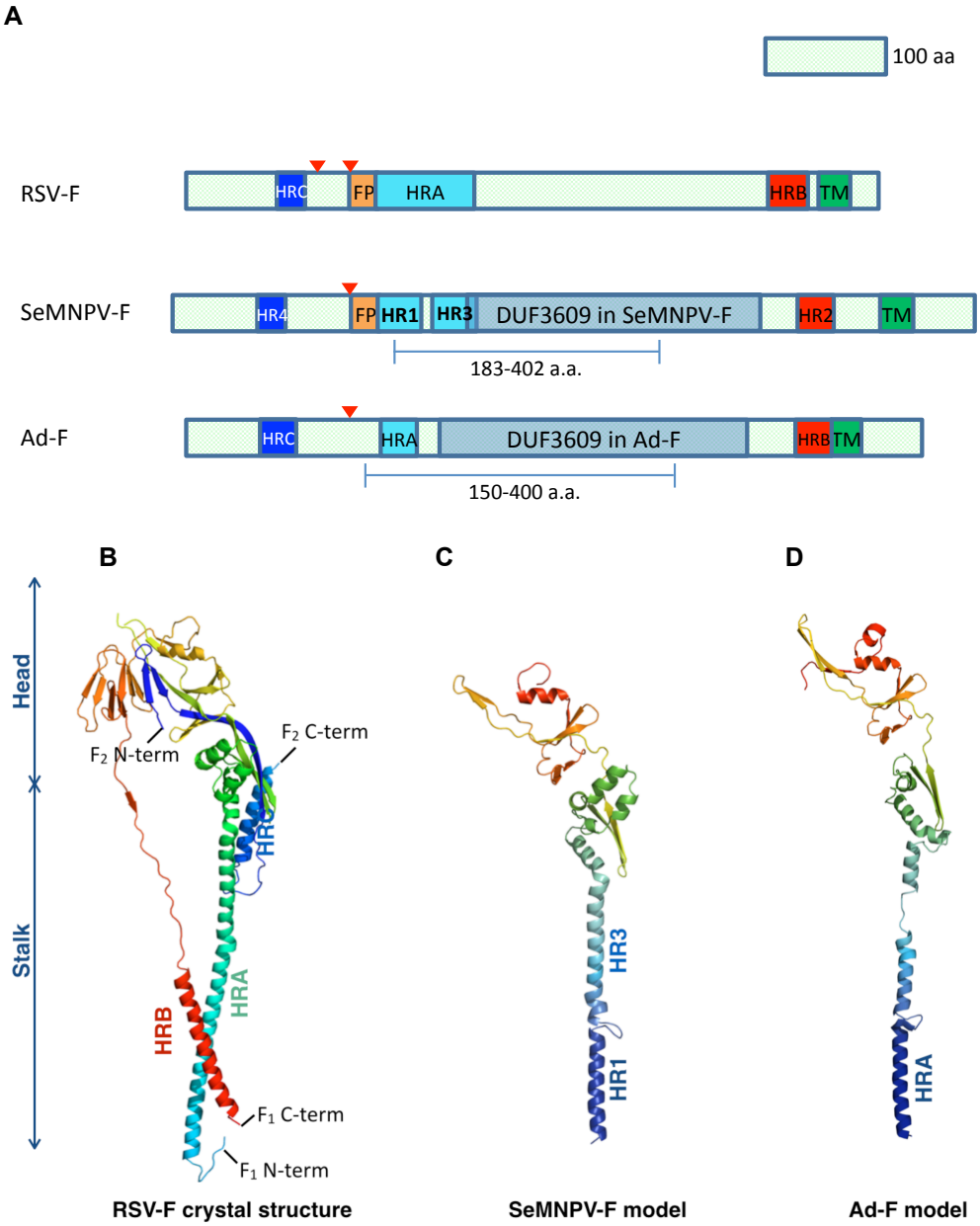


Fig. 6. Comparison of postfusion structural models between Ad-F (D, region 150-400 a.a.) and SeMNPV-F (C, region 183-402 a.a.) using the RSV-F postfusion crystal structure (B, PDB: 3RKI) as a basis. Schematic maps of RSV-F, SeMNPV-F and Ad-F proteins with positions of HR regions, TM, DUF3609 conserved domain were aligned at the furin cleavage site (red arrow heads). The heptad repeat (HR) regions of SeMNPV-F and Ad-F were predicted using the COILS program (ch.embnet.org/cgi-bin/COILS_form_parser). The HR regions and two furin cleavage sites of RSV-F were previously described (366).

Ad-F is proteolytically cleaved and oligomerizes in 293T cells and is incorporated into VSV pseudotyped particles.

To characterize the proper folding of Ad-F biochemically, HEK-293T cells were transiently transfected with an expression plasmid encoding C-terminally HA-tagged Ad-F. In addition, to assess its transport to the plasma membrane, the transfected cells were subsequently infected with VSVG protein pseudotyped VSVΔG after which incorporation of Ad-F into budding VSVΔG particles at the plasma membrane was determined (Fig. 1). Mock-infected cells and their supernatant were used as controls. Cell lysates and supernatant containing released viruses were analyzed by western blotting using an anti-HA-tag antibody. The HA antibody recognized three protein products of approximately 70 kD, 87 kD and 174 kD. The molecular weights of HA-tagged Ad-F_1 and Ad-F calculated from their amino acid sequence are 54.5 kD and 69.5 kD, respectively. Assuming each N-glycan side chain to contribute 2.5 kD (44) to the size of a glycoprotein, the 70 kD and 87 kD bands detected by western blotting correspond well with HA-tagged Ad-F_1 carrying 6 glycans and Ad-F carrying 7 glycans, respectively. The detection of Ad-F_1 indicated that cleavage occurred at the predicted furin cleavage site (Fig. 7A, red arrow head). The furin protease predominantly resides in the trans-Golgi network in mammalian cells (42, 56). Thus, furin cleavage of the Ad-F protein indicates that the protein is properly folded since it has passed the protein folding quality control in the ER required for ER exit. Both monomeric Ad-F_1 and Ad-F were detected in pseudotyped viruses and cells under reducing condition (Fig. 7B left panel, Lane 1 and 2), indicating that not all the Ad-F proteins were cleaved by furin and that a mixed population cleaved and uncleaved proteins was transported to plasma membrane and incorporated in pseudotyped virus particles. The relative amount of Ad-F_1 was significantly reduced in the cells (Fig. 7, Lane 2) and Ad-F_1 was not detected in the viruses (Lane 1) under non-reducing condition (Fig. 7B left panel), suggesting that the majority of the proteins occur as disulfide-linked Ad-F_1/F_2 complexes. The Ad-F_2 subunit does not contain the HA-tag and can hence not be detected with the anti-HA antibody. The most slowly migrating form of 174 kD confirms the presence of the Ad-F oligomer. Both monomeric and oligomeric forms were detected in the pseudotyped viruses under both reducing and non-reducing conditions (Fig. 7B left and right panels, Lane 2). However, the monomers in the viruses are more abundant under reducing condition while the oligomers are more abundant under non-reducing condition, suggesting disulfide bonding might be involved in protein oligomerization. The ratios of oligomers to monomers are similar in the cells under reducing and non-reducing conditions

(Fig. 7B left and right panels, Lane 1). However, a higher percentage of oligomerized Ad-F proteins are found in viruses than in the transfected cells. Furthermore, given its incorporation in VSV particles that bud from the plasma membrane, the Ad-F protein must have been transported to the plasma membrane. These biochemical data suggest that the expressed cellular baculovirus F homologue found in *Anopheles darlingi* is properly folded, oligomerized and transported through the secretory pathway to the plasma membrane.

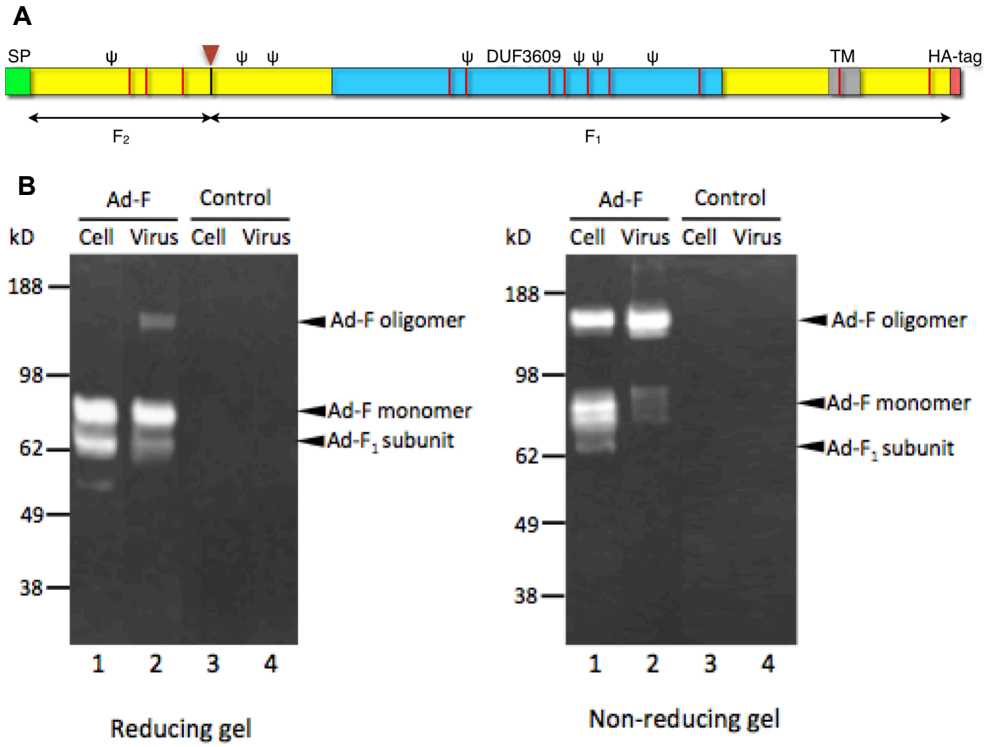
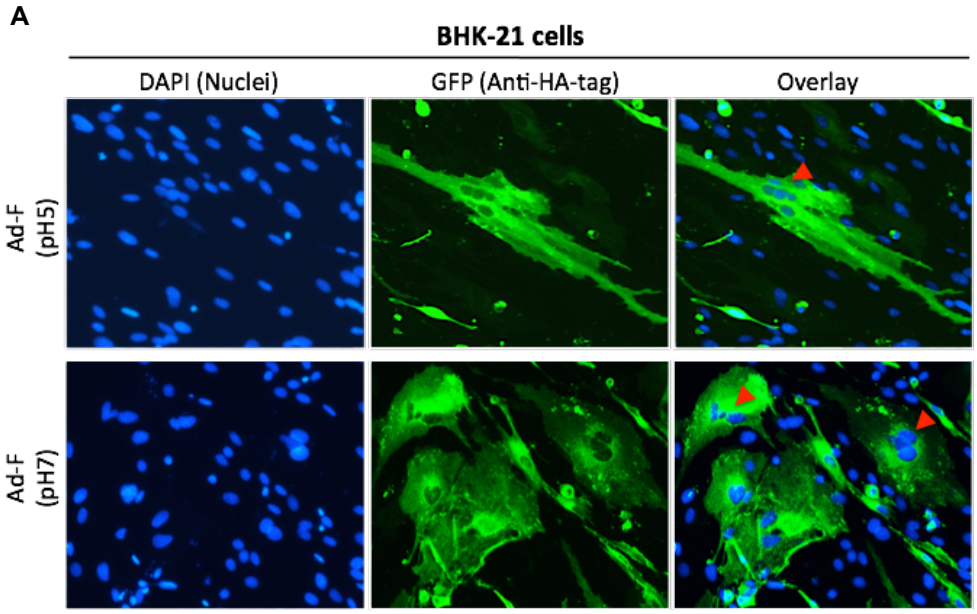


Fig. 7. (A) Schematic representation of Ad-F protein. The position of the SP (green bar), DUF3609 (blue bar) and TM domains (grey bar) are indicated as in Fig. 3. Additionally, the C-terminal HA-tag (red bar), cysteine residues (red lines) and predicted N-glycosylation sites (ψ) (NetNGlyc server: www.cbs.dtu.dk/services/NetNGlyc) are indicated. Similar to SeMNPV-F protein proteolytic cleavage site separates the protein into two subunits F_2 and F_1. (B) Western blot analysis of Ad-F protein. HEK-293T cells were transfected with 1 μg of plasmid DNA encoding Ad-F and harvested 48 h after transfection. In parallel, to generate VSVΔG/GFP-AdF pseudotyped viruses the transfected 293T cells were infected with VSVΔG/GFP-G viruses for 24 h. The non-transfected cells and non-pseudotyped viruses are used as controls (the two lanes on the left in both panels). The Ad-F protein levels of the transfected cells and generated VSVΔG/GFP-AdF pseudotyped viruses were analyzed under reducing (left panel) and non-reducing (right panel) conditions on SDS-PAGE followed by western blotting using the anti-HA-tag antibody to detect Ad-F.

Fusogenic activity of the Ad-F protein

To study the potential of Ad-F protein in mediating virus-cell fusion we assessed the virus entry of the generated Ad-F pseudotyped VSVΔG/GFP viruses on C6/36 mosquito cells by measuring fluorescence intensity of GFP expression upon virus entry. However no significant virus entry was observed (data not shown).

To determine the fusion capacity of Ad-F protein in another way we evaluated its cell-cell fusion activity of the Ad-F protein by expressing the Ad-F in BHK-21 cells. As controls, the AcMNPV-GP64 and SeMNPV-F proteins were taken along. These baculovirus fusion proteins, SeMNPV-F and AcMNPV-GP64, are activated by low pH, which then triggers the fusion process (32, 150). To assess the role of low pH in fusion activation of the Ad-F protein, BHK cells transfected with HA-tagged Ad-F (Fig. 8A), SeMNPV-F (Fig. 8B) or AcMNPV-GP64 (Fig. 8C) expression plasmids were briefly exposed to acidified or neutral buffer 48 h after transfection and analyzed for cell-cell fusion. The HA-tagged proteins expressed in the cells were detected with anti-HA-tag antibody and Alexa 488-conjugated secondary antibody. Overexpression of the Ad-F protein seems to result in syncytium formation independent of pH, suggesting an alternative fusion-triggering mechanism of the Ad-F protein. Yet, the relative low number of syncytia suggests that Ad-F mediated cell-cell fusion was not efficient.



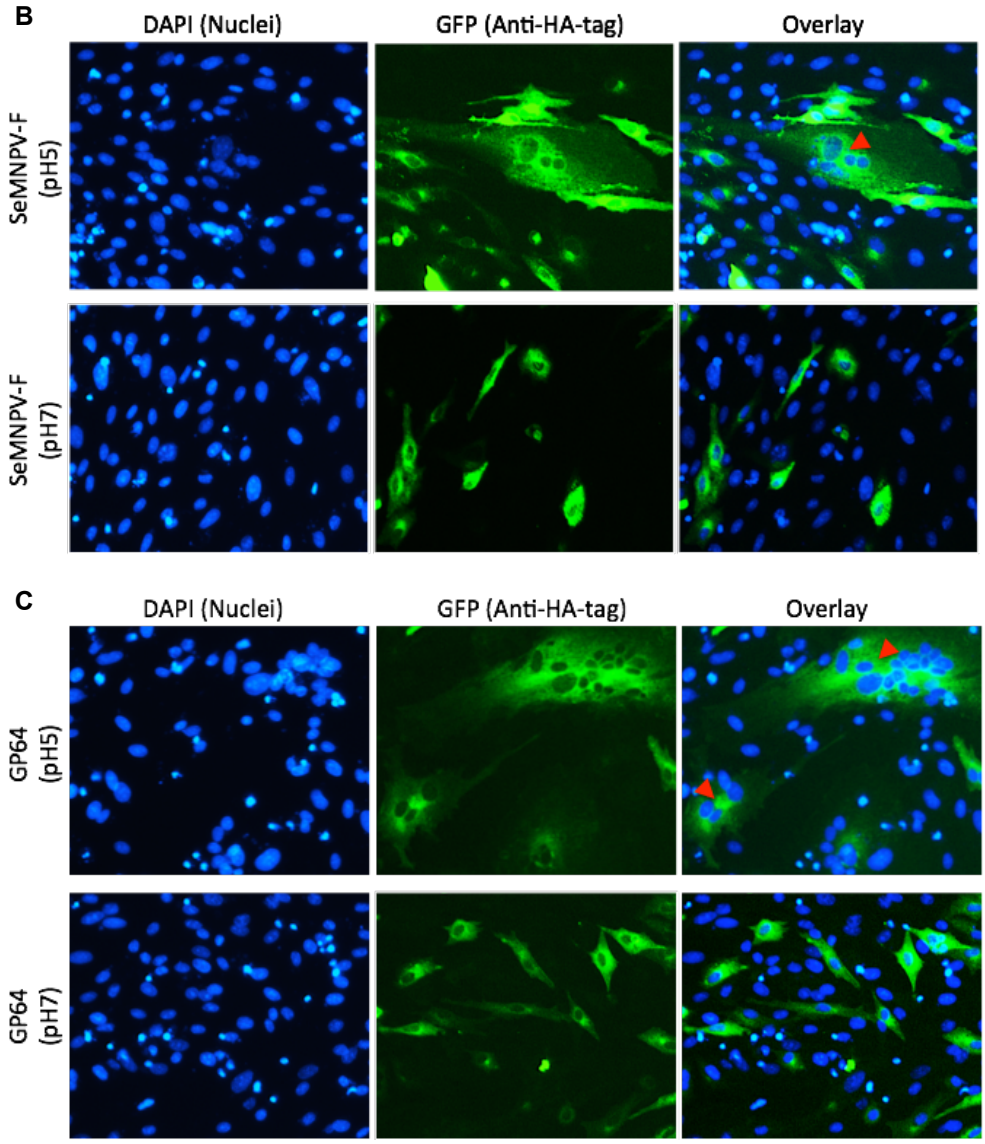


Fig. 8. Cell-cell fusion activity of Ad-F and GP64 proteins. BHK-21 cells were transfected with plasmids encoding F or GP64 proteins as indicated. At 48 h post transfection cells were incubated in PBS at pH 5 or pH7 for 2 min and then in DMEM medium supplemented with 10% FBS (pH 7.0-7.4) for 4 h. After fixation cells were stained with DAPI and the HA-tagged Ad-F and GP64 proteins were detected with monoclonal antibody against HA-tag and Alexa Fluor488-labeled secondary antibody. The stained cells and nuclei were visualized and photographed with an inverted fluorescence microscope using DAPI (A and B, left panel) and GFP channels (A and B, middle panel), respectively. The images recorded with these two channels are overlaid (A and B, right panel). Syncytia are indicated with red arrows in overlay images.

DISCUSSION

Genome sequencing of a rapidly increasing number of insect viruses and their hosts allows us to track their coevolution trajectory and to understand the interaction of host and virus. Previous phylogenetic analyses of the genes encoding the baculovirus F fusion proteins, insect endogenous retrovirus envelope proteins and cellular proteins in *Drosophila* (Iris) have suggested that the *gypsy* and *roo* retroviruses may have independently acquired their *env* gene from baculoviruses (232, 319), while the *Iris* gene may be the results of domestication of an *env* gene from *kanga* retroviruses, a sister lineage of the *roo* retroviruses (231). The possibility of horizontal gene transfer is supported by several pieces of evidence. First, long-terminal repeats (LTR) in retrotransposons have the ability to acquire additional open reading frames to become infectious [reviewed in (87, 128)]. Second, a viral gene can be domesticated by its host to encode a protein serving an important physiological function. The examples have been found throughout mammals, insects and nematodes with their pathogens (232). The best-studied example is a human endogenous retroviruses envelope gene, called *syncytin,* which maintained its original membrane fusion activity after domestication and mediates trophoblast differentiation fusion during the placenta formation process (254, 322). Although this is the case of vertebrate retroelement transposition, similar events could also have occurred between the insect retroviruses and their hosts. Indeed, the transposition of lepidopteran *gypsy* retrotransposon TED to baculovirus AcMNPV indicates that the *gypsy* retroelement is capable of integrating into the baculovirus genome (96). Like baculovirus F envelope fusion proteins, vertebrate retrovirus envelope proteins require a proteolytic cleavage to become functional and allow the release and exposure of a downstream hydrophobic fusion peptide (184). The fusogenic features found for baculovirus F proteins have also been found in insect retrovirus *env* proteins. Comparative sequence analyses have revealed that the *gypsy* and *roo* insect retroviruses possess a furin cleavage consensus motif (RxxR), a hydrophobic domain highly similar to baculovirus fusion peptide (Fig. 4) and a C-terminal transmembrane domain (262). It has been demonstrated that the *gypsy* retrovirus envelope protein is proteolytically processed in Sf9 cells and can mediate cell-cell fusion (261).

In the current study we identified and characterized a cellular protein from the mosquito *A. darlingi* (named Ad-F), which has a high degree of sequence homology to baculovirus F fusion proteins. Phylogenetic analysis of nucleotide sequences suggests that Ad-*F* is closely related to *Iris* and baculovirus *F* genes (Fig.

2). Each of the three F homologs has a signal peptide, a transmembrane domain, N-glycosylation sites and a highly conserved domain DUF3609 (approximately 200-360 a.a. in length). However, Ad-F protein appears to be more structurally and functionally related to the lepidopteran baculovirus F than to Iris. Structural modeling of Ad-F and baculovirus SeMNPV-F using RSV-F as a template yielded the same highest score (Fig. 6), suggesting Ad-F and SeMNPV-F are structural analogs. Our experimental data demonstrate that Ad-F protein is properly folded and transported to the plasma membrane (Fig.7B) where it is able to mediate cell-cell fusion not only at acidic but also at neutral pH though not as efficient as the baculovirus fusion proteins (Fig. 8A). The protein sequence homology as well as similarities in protein structure and function between SeMNPV-F and Ad-F suggest that divergent evolution through horizontal gene transfer has occurred between the baculovirus and mosquito. Furthermore, the discovery of lacking introns in the Ad-*F* gene indicates that this gene is not authentic in the mosquito genome and might have been endogenized from a viral source. All the evidence suggests that the Ad-F gene may have been domesticated from a baculovirus ancestor but not vice versa.

Baculoviruses are known to infect dipterans including several species of the mosquito genera *Aedes*, *Anopheles* and *Culex* [reviewed in (20)]. Therefore, it is well possible that the Ad-*F* gene was the outcome of direct or indirect gene transfer from a baculovirus ancestor infecting dipterans. Surprisingly, Ad-F shares higher amino acid sequence identity with F proteins from several lepidopteran NPVs including *Chrysodeixis chalcites* NPV (GI: 68304286, 19% identity), *Trichoplusia ni* SNPV (GI: 74229828, 20% identity) and *Helicoverpa armigera* SNPV (GI: 209401195, 20% identity) than with the F protein from dipteran baculovirus *Culex nigripalpus* NPV (CuniNPV, GI: 15320898, 15% identity) from a PSI-BLAST search with two iterations. Previous phylogenetic analysis based on various genes suggested that CuniNPV is divergent from lepidopteran baculoviruses, which is consistent with the position of the *F* gene of this virus in the phylogenetic tree of the *F* homologs (Fig. 2). The dipteran baculoviruses may either be more ancestral or have a different co-evolutionary speed compared to lepidopteran baculoviruses (5, 136, 269). Evidence suggested that some viruses (145, 344), including baculoviruses (316, 320) may become genetically divergent and evolve along with their hosts. If lepidopteran and dipteran baculoviruses diverged from a common baculovirus ancestor and have evolved separately, the *F* genes from dipteran baculoviruses may have been under different selective pressure e.g. the host immune responses and may therefore have undergone

sequence alteration at different pace. Possibly, the different levels of sequence identities of lepidopteran and dipteran NPV F proteins to Ad-F protein reflect the primordial separation of lepidopteran and dipteran NPVs due to host-dependent evolution.

To date, approximately 50 dipteran and over 100 lepidopteran species are sequenced (249, 291, 347). Among the baculovirus hosts with sequenced genomes, cellular F homologs have been found mostly in the genera *Drosophila*, *Anopheles* and *Aedes* (Diptera), but much less in Lepidoptera. We have identified only two lepidopteran cellular F homologs, one from *Danaus plexippus* (GI: 357631190, monarch butterfly) and one from *B. mori* (GI: 512912007, silkmoth), respectively, in the PSI-BLAST search with SeMNPV-F sequence as a query. This observation raised an interesting question: what may have caused *F* gene domestication in these baculovirus hosts? Due to the ability of acquiring additional ORFs the insect LTR retrotransposons and endogenous retroviruses are likely to play an important role in mediating the gene transfer between viruses and insects, based on phylogenetic analysis (232). In the genomes of *Drosophila* and mosquito species a large number of LTR retrotransposons are present. For example, there are 134 copies in *D. melanogaster* (164), 241 copies in *A. darlingi* and 28905 copies in *A. aegypti* (237); whereas, only 15 copies have been annotated in *Heliconius melpomene* (postman butterfly) (193) while 338 copies present in *B. mori* (425). Thus, the gene transfer may correlate to the number of LTR retroelements in the hosts.

Based on the structural features the viral envelope fusion proteins have been distinguished into three classes [reviewed in (415)]. Lepidopteran baculovirus F proteins have been recognized as class I viral fusion proteins based on the occurrence and location of structural domains including the heptad repeat region, the fusion peptide and the transmembrane domain in the protein (35) as well as on the biochemical properties such as proteolytic cleavability to prime fusion activity (216, 411). In the current study we expand the class I fusion protein range from viruses to eukaryotes by identifying the cellular F homolog in mosquito (Ad-F). At first sight the Ad-F protein is a cellular homolog of baculovirus F proteins. However, the Ad-F protein shows several unique features. First, the predicted fusion peptide is not as hydrophobic as those of other class I fusion proteins. Second, the Ad-F protein appeared to be incorporated into VSV particles as disulphide-linked oligomers, a situation similar to class III envelope fusion proteins (e.g. GP64) (287). In this study Ad-F protein appears to mediate

cell-cell fusion in a pH-independent manner; hence the trigger of Ad-F mediated cell-cell fusion remains to be resolved.

Our study provides evidence that Ad-F may have been domesticated from a baculovirus ancestor and preserved apparent fusion activity of baculovirus F proteins. However, this does not necessarily mean that the Ad-F protein actively mediates membrane fusion in the *A. darlingi*. As several studies have suggested, cellular genes homologous to the retrotransposon *gag* or *env* genes might be involved cell defense by encoding proteins that interact with cellular receptors and block the infection by exogenous viruses carrying a homologous envelope protein [reviewed in (275)]. Up to date this phenomenon has been only reported in mammals but not in insects. However, our current data demonstrated Ad-F was cleaved like other endogenous retrovirus envelope proteins, properly folded, transported to the plasma membrane and exhibit significant structural similarity with baculovirus F protein; therefore, the Ad-F protein may compete with other similar envelope proteins from exogenous viruses for receptor binding. Further investigation on whether this mosquito *F* homolog is indeed functional or junked from previous viral infection should be pursued.

ACKNOWLEDGMENTS

The authors thank Osvaldo Marinotti for valuable suggestions on genome annotation of *Anopheles Darlingi*. We are grateful to Gorben Pijlman for providing vector pIB-EGFP and to Corinne Geertsema for technical assistance. We also thank Harmit S. Malik for selection of conserved domain in aligned sequences and Erik de Vries for assistance with phylogenetic analysis.

SUPPLEMENTARY FIGURE

```
DUF3609            1 --------------------------------------------------      0

gypsy              1 ARITDYSHANYIPVDDGYVLVFDQRLYLRHSSNISEYRSVIDETDMLSDS     50

DUF3609            1 ----------------------------------LRLLGS-----------      6
                                                       |.:||:
gypsy             51 FPQSHMRKLLQVDTDHLRTMLSVLEVHHRIARSLDFLGTALKVVAGTPDA    100

DUF3609            7 ----RLKSSQEAILEA------VASAHQ---GKLSPLVLSIKQLEAEILK     43
                         ::|.::.:::||      :.|..|   .||:..:..|.|.....|.
gypsy            101 DDLNKIKNTEAQLVEANNNQIFINSETQKQINKLTDTINKILQSRKGDLV    150

DUF3609           44 ILGH-----LPRDRRLPFE----KFTISDIYRIASVVPRQLDNHIVFEIT     84
                     ...|     |.|:|.::.|    |:||:.. ::..|.|..|| |..:|..
gypsy            151 DTPHLYETLLARNRIIKEEIQNLKLTITLA-KLNIVNPTILD-HADLESL    198

DUF3609           85 V-------PLIDVEQFNLYRLTPIPRLNNGKIQLVDTETPYLGIN-----     122
                     |        |:::     |:..::|:.|....|..:..:.|.:...
gypsy            199 VEENKENTPIVE-----LLEAAKIKVLQSENIIHIIIKYPKVTFECNKVI    243

DUF3609          123 ----DHLDRYFPLQ--NLDDCLELAPERFIC-KPNQITY-----------     154
                         .|.|..:.|.  ::.:|.........| |.:.:|:
gypsy            244 IYPVSHQDTILRLKENTVAECENDIFAVKNCTKTTHVTFCERSSRETCAS    293

DUF3609          155 ----GN-----GDDSLPCSLAAIRNQTSKVCTFRQVERSSLWTPLVAPNS    195
                         ||     ||||||||||||||||||||||||||||||||||||||
gypsy            294 SLHAGNNASCHGDDSLPCSLAAIRNQTSKVCTFRQVERSSLWTPLVAPNS    343

DUF3609          196 WMVALTKELSLMGVCSGERQELKINGSGILSIRSDCIVRSTAVTLQGE--    243
                     ||||||||||||||||||||||||||||||||||||||||:   |..
gypsy            344 WMVALTKELSLMGVCSGERQELKINGSGILSIRSDCIVRSTT---QPSHL    390

DUF3609          244 PRKGIPSKQGYASLQLTNKSSRESDILE---------SFNQ-----LLEI    279
                     || ..|...|...:..:....|..|.:|         :|.:     ..|.
gypsy            391 PR-IEPVDDGIIVINDAHAHVRTDDNIEIHINGTYLITFERSATINETEF    439

DUF3609          280 VT-------QLKLDQEKLEAVE------DYPM-----------------    298
                     |:       |..:.:.:|..:|      .:|:
gypsy            440 VNLRKPLSKQPGIVRSPLLNIEGHDPVLSLPLLHRLNVNNLVSIRNLQIE    489

DUF3609          299 VFIAVCPAIVLIA-------LLISLGWLYRTHRRRQLRAAQRPVNEV---    338
                     |.:...|:|.::|       ::.||..::....||:..:.|:.::::
gypsy            490 VTVNRSPQIWFVAGAVLNSGIIGSLVLILMLRARRASHQKQKSIDPTESS    539

DUF3609          339 -DGLQDTKNETRTSNLPLLEKNEV--    361
                      ||.|             ||:..|
gypsy            540 EDGHQ-------------LEEGIVNN    552
```

Fig. S1. Pairwise alignment of the consensus sequences of DUF3609 and *gypsy* conserved domains using EMBOSS Needle (www.ebi.ac.uk/Tools/services/web_emboss_needle/toolform.ebi). The highly conserved motif is in bold.

Chapter 5

On the role of conserved histidines in the baculovirus fusion protein F

Qiushi Wang[1,2], Monique M. van Oers[2], Just M. Vlak[2],

P. Rottier[1], Berend Jan Bosch[1]

[1] Virology Division, Department of Infectious Disease and Immunology, Faculty of Veterinary Science, Utrecht University, Utrecht, the Netherlands

[2] Laboratory of Virology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

Manuscript in preparation

ABSTRACT

Budded baculovirus particles (BVs) deliver their nucleocapsid to the cytoplasm of host cells by membrane fusion. For most baculoviruses BV fusion is mediated by the viral fusion protein (F), which is believed to undergo conformational changes at low pH. To further explore the mechanism of such membrane fusion, we determined the fusion activation pH for the F protein of *Spodoptera exigua* multicapsid nucleopolyhedrovirus. The fusion activation pH (5.0-5.5) is lower than the pK_a of histidine (pK_a=6.0), suggesting that at the fusion activation pH certain histidines are likely to be protonated, which may destabilize the conformation of the F protein. To confirm the function of histidines as pH sensors we pretreated F-pseudotyped vesicular stomatitis virus (VSV) with diethylpyrocarbonate (DEPC) to prevent histidine protonation. This treatment potently reduced infectivity, indicating that histidine protonation is involved in F-mediated membrane fusion. Next, we studied the role of the conserved histidine residues using alanine substitution. Single substitution of conserved histidines in the F_2 subunit (H59 or H119), but not those in F_1, abolished syncytium formation of cells transiently expressing F protein. However, these two single substitutions in F_2 did not abrogate entry of pseudotyped VSV nor did they inhibit the low-pH triggered conformational change of the F ectodomain. However, these substitutions altered the postfusion structure of the F ectodomain. Our results suggest that the two conserved histidines H59 and H119 are important for membrane fusion by regulating the postfusion structure of F protein.

INTRODUCTION

Membrane fusion is a key step during the entry of enveloped viruses into host cells and is mediated by one or more viral envelope proteins including a viral fusion protein. Basic principles of the membrane fusion process are conserved among enveloped viruses. Upon virus attachment to the cell surface receptor binding and/or low pH in the endosome triggers a series of conformational changes in the viral fusion protein (117, 251, 306). First, the fusion protein - anchored in the viral envelope by a C-terminal transmembrane domain - extends and inserts a hydrophobic fusion peptide into the target cell membrane (302). Next, the extended fusion protein anchored in opposing membranes collapses, forming a hairpin structure. As a result the viral and cellular membranes are brought into close proximity, followed by merging of the outer leaflets of the two lipid bilayers (hemifusion state). Eventually the inner leaflets are mixed as well, forming a fusion pore (postfusion state) (127). To allow efficient delivery of the

viral nucleocapsid into the cytoplasm the fusion pore is further expanded (52, 65, 250, 309).

Membrane fusion is an energy dependent process (59) and kinetic barriers are found at the following stages: engaging and apposing of two bilayer membranes, merging of the lipid bilayers (fusion stalk formation or hemifusion state), fusion pore formation, and subsequent pore enlargement (302). Before the fusion event the viral fusion protein is in a metastable conformation. Upon fusion activation the conformational changes of the fusion protein as described above generate the free energy to compensate for the energy penalty caused by formation of the bilayer curvature and fusion stalk. Finally the fusion protein reaches a stable postfusion conformation, which leads to formation and enlargement of the fusion pore (34, 60, 65).

Three classes of viral fusion proteins have been recognized, based on structural features and the fusion strategies used. Class I fusion proteins which include influenza HA, HIV Env, paramyxovirus F and Ebola virus GP proteins, i) are trimeric in the pre- and postfusion states; ii) require proteolytic cleavage to prime and lower pH and/or receptor binding to trigger conformational change; and iii) form a central α-helical coiled-coil post-fusion structure. Class II fusion proteins - including flavivirus E, alphavirus E1 and bunyavirus Gc proteins - i) are dimeric in the prefusion state and trimeric in the postfusion state; ii) require chaperone proteins to regulate and a low pH to trigger the conformational transition of these proteins; and iii) form a β-sheet composed postfusion structure with fusion domain as loops. Class III fusion proteins which include herpes virus gB, rhabdovirus G, Thogotovirus 75K and baculovirus GP64 protein, have a combination of fusogenic characteristics: like class I fusion proteins they have similar oligomeric states of pre- and postfusion structures; like class II fusion protein they contain a fusion loop but in contrast to the first two classes they do not require proteolytic cleavage to prime the fusion function (302, 404, 415).

Notably, most viral fusion proteins are triggered by low pH, which is commonly around pH 5 to pH 6. The only amino acid that gets protonated at this pH range is histidine (pK_a=6.0). The proposed triggering mechanism suggests that at a pH above 6 histidine residues form hydrogen bonds with positively charged amino acids, whereas at low pH the histidine residues get protonated and hence positively charged, thereby destabilizing the hydrogen bonds formed in the pre-fusion state. The histidine residues are subsequently positioned in vicinity to negatively charged amino acids, forming new salt bridges (125, 165). There are

many examples of histidine protonation triggering structural changes at low pH (48, 162, 181, 205, 241). It is therefore expected that histidine residues generally play an important role in low pH-triggered conformational changes in membrane fusion.

Baculoviruses form a family of enveloped double stranded DNA viruses that are only pathogenic to arthropods (134). Two distinctive envelope fusion proteins, GP64 and F, have been identified in this family depending on the genus. GP64, only found in group I alphabaculoviruses, has been classified as a class III envelope fusion protein based on its postfusion crystal structure (13, 161). Structural data have recently confirmed that F is a class I envelope fusion protein (This thesis, Chapter 3). Previous studies on the F protein of *Spodoptera exigua* multicapsid nucleopolyhedrovirus (SeMNPV) have demonstrated that cleavage of a precursor protein F_0 by Golgi-resident furin (-like) proteases is required for fusion activity. This cleavage generates a large C-terminal membrane anchored subunit (F_1) and a smaller N-terminal subunit (F_2) that remain covalently linked via an intersubunit disulphide bond (411). The cleavage also unsheathes the fusion peptide, located in F_1 directly downstream of the furin cleavage site (150). The F_2 subunit harbors a receptor-binding domain (400), while four heptad repeat (HR) regions are present in the F_1 subunit (Chapter 3).

From functional experiments it became evident that the baculovirus F proteins are activated by acidic pH, which the virus encounters during endocytosis (221, 411). As many low pH triggered viral fusion proteins adopt histidine protonation as their basic mechanism for fusion activation, we examined the relevance of six highly conserved histidine residues in the SeMNPV F ectodomain for virus entry using mutagenesis in combination with functional and biochemical analyses.

MATERIAL AND METHODS

Cells and viruses

Spodoptera frugiperda (Sf9) cells were cultured at 27°C in Sf-900 II SFM medium (Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Bodinco BV). *S. exigua* (SeUCR) cells were cultured at 27°C in supplemented Grace's Insect medium (Life Technologies) with 10% FBS. HEK 293T cells were cultured at 37°C with 5% CO_2 in DMEM (Lonza) supplemented with 10% FBS. VSVΔG/GFP-G* (kindly provide by Dr. Y. Matsuura), generated and described previously (370) is a recombinant vesicular stomatitis virus (VSV), in which the coding sequence for the envelope fusion protein G has been replaced by the GFP

sequence. This VSV vector was used to generate VSV virions pseudotyped with F or histidine-mutated F protein (F^{Hismut}) on their surface (see below).

Plasmid construction

To construct plasmids expressing the SeMNPV F protein with an affinity tag, a strep-tag (ST) and a stop codon were fused at the 3'-end of the ORF for SeMNPV F (SeF) (149) excluding the stop codon in plasmid p166AcV5-Se8 (411) to generate an insect cell expression plasmid p166AcV5-SeF-ST containing the *Orgyia pseudotsugata* MNPV *gp64* early promoter. Strep-tagged SeF in p166AcV5-SeF-ST was subcloned into pCAGGS/MCS containing the chicken beta-actin promoter and the cytomegalovirus enhancer (138) to generate the mammalian cell expression plasmid pCAGGS-SeF-ST. In order to construct a plasmid expressing the soluble F (sF) the region encoding the F ectodomain without its signal peptide (residues 18 to 553) [GenBank Acc. no. AAQ11029) was PCR amplified with Pwo high-fidelity DNA polymerase (Roche) using plasmid p166AcV5-Se8 as a template. The PCR product was cloned into the pCD5 expression vector in frame with the DNA sequence coding for the CD5 signal peptide for efficient secretion by mammalian cells (435). At the 3'-end, the gene fragment encoding the F- ectodomain was cloned in frame with a C-terminal enterokinase cleavage site to have the possibility to remove downstream tags, followed by three strep-tags for detection and purification with streptactin. The resulting vector was designated as pCD5-sF-ST3. Plasmids containing PCR-amplified regions were confirmed by DNA sequencing.

Site-directed mutagenesis

Histidine-to-alanine mutations in SeMNPV F were generated using the PCR-based whole plasmid mutagenesis method (QuikChange II Site-directed mutagenesis kit, Agilent Technologies), with the plasmid p166AcV5SeF-ST as template. The primer sequences are listed in Table S1. All mutated plasmids were confirmed by PCR and DNA sequencing.

Production of pseudotyped vasicular stomatitis virus

The schematic representation of the genome structures and the production of pseudotyped vesicular stomatitis virus (VSVpv) is shown in Fig. 1. To generate VSV complemented either with the VSV G-protein (VSVG), wild type (wt) or histidine mutated SeMNPV F, 293T cells were transfected with pCAGGS plasmids encoding these fusion proteins using polyethylenimine (PEI, Sigma-Aldrich). (For clarity only VSVG and wt SeMNPV F pseudotyped viruses are shown in Fig.1). At 48 h post transfection the transfected cells were infected with VSVG-pseudotyped viruses (VSVΔG/GFP-G*) at an MOI of 0.1 $TCID_{50}$ units /

cell, in which the VSVG ORF was replaced with GFP ORF and was pseudotyped with VSVG. After the viruses were adsorbed for 2 h at 37℃ the cells were washed three times with serum-free DMEM and were subsequently incubated with DMEM supplemented with 1% FBS for 24 h at 37℃. The culture medium with generated virus particles was centrifuged to remove cell debris and used for inoculation of target cells.

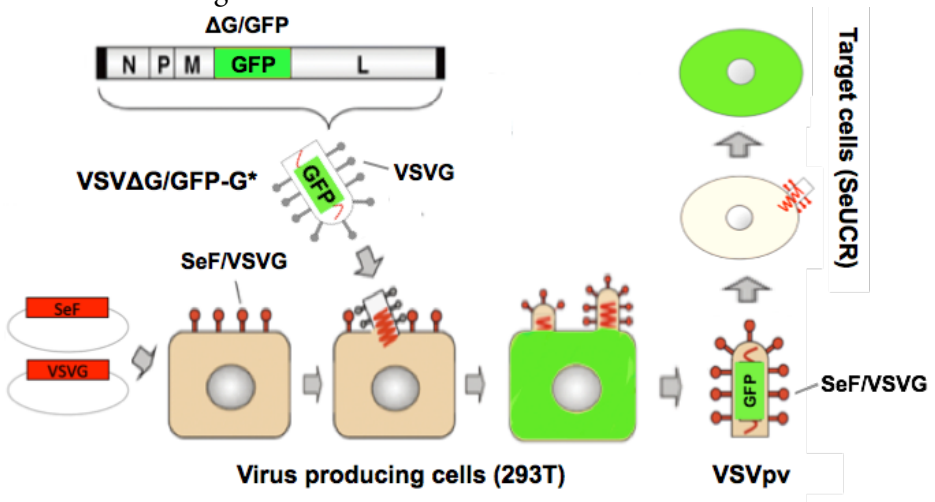


Fig. 1. Schematic representation of the genome structures and the production of pseudotyped VSV bearing SeMNPV F or VSVG. Mammalian cells (HEK 293T) were transfected with an expression plasmid carrying the SeMNPV F ORF (SeF) and then infected with the pseudotyped VSVΔG/GFP-G* virus. The newly generated VSV pseudotyped viruses (VSVpv) released from the producer cells were then tested for the ability to enter target insect cells (SeUCR) by measuring GFP expression. [Figure modified from (166)].

Measurement of optimum pH for fusion activation

SeUCR cells were incubated with VSVΔG/GFP-F* at an MOI of 0.1 $TCID_{50}$ units / cell on ice for 1 h in ice-cold Grace's Insect medium with 40 mM ammonium chloride (NH_4Cl) to allow virus binding to the plasma membrane without endocytosis. After incubation, the cells were treated with prewarmed PBS at pH 5.0, 5.5, 6, 6.5 or 7 for 2 min and then incubated in culture medium containing NH_4Cl for 24 h at 27°C. Incubation of virus and cells without NH_4Cl and pH treatment was considered as a control. Cells emitting fluorescence were scored using an EVOS inverted fluorescence microscope (AMG).

DEPC treatment of F-pseudotyped VSV viruses

Diethylpyrocarbonate (DEPC) solution (1 M) was prepared in 96% ethanol from a concentrated stock (Sigma-Aldrich) The VSV pseudotyped viruses in PBS were incubated with various concentrations of DEPC as indicated for 10 min at

22°C. SeUCR cells were inoculated with 10-fold serial dilutions of VSVpv and cultured at 27°C in Grace's Insect medium supplemented 10% FBS for 24 h and scored for virus entry (GFP-positive cells) by endpoint dilution assays.

Syncytium formation assay

Sf9 cells were seeded in 6-well plates and transfected with plasmids encoding F or F^{Hismut} using Cellfectine II (Invitrogen). After incubation at 27°C for 48 h, cells were exposed to PBS at pH 5.0 for 2 min and then incubated in Sf-900 II medium for 4 h. To visualize the nuclei, the cells were permeablized with 0.5% Triton X-100 in PBS and stained with DAPI (Invitrogen) at a dilution of 1:3000 in PBS at room temperature for 20 min. The cells were then washed once with 10% FBS in PBS-T (0.1% Tween 20 in PBS) at room temperature for 15 min. Cells were then incubated with polyclonal antibody directed against the SeMNPV F ectodomain (for antibody production see Chapter 2) at a dilution of 1:200 at room temperature for 1 h. Cells were washed three times with PBS-T for 15 min and incubated with Alexa Fluor488 goat anti-rabbit immunoglobulin G (IgG; Life Technologies) at a 1:200 dilution in PBS for 1 h. After washing three times with PBS, fluorescence was observed with an EVOS inverted fluorescence microscope (AMG). Relative levels of fusion activity were determined by dividing the total number of nuclei found in the syncytia by the total number of nuclei in the same field, normalized to the cell surface expression level of F or F^{Hismut}. A syncytium was defined as a cluster of cells containing at least 3 nuclei.

Cell surface detection of F protein by FACS

Sf9 cells transiently expressing wild type (wt) F or F histidine mutants were washed with PBS and harvested by incubating with 2% FBS, 5 mM EDTA and 0.02% NaN_3 in PBS at room temperature for 10 min. The cells were then fixed with 3.7% paraformaldehyde at 4℃ for 10 min and washed once with 2% FBS, 5 mM EDTA and 0.02% NaN_3 in PBS for 10 min at room temperature and blocked with 10% FBS in PBS-T for 15 min at room temperature. Subsequently, the cells were immunostained as described in syncytium formation assay. Finally, the cells were washed twice with PBS-T and surface expression was measured by a FACSCantoII flow cytometer (BD Bioscience) with FACSDiva software (BD Bioscience). To determine the total levels of wt and histidine mutated F, the fixed cells were permeablized using 0.5% Triton X-100 in PBS and stained with the antibodies as described above.

Western blot analysis of the full-length F protein

Sf9 cell lysate was mixed with either reducing loading buffer (125 mM Tris-HCl, 2% sodium dodecyl sulfate (SDS), 5% β-mercaptoethanol, 10% glycerol and

0.001% bromophenol blue, pH 6.8) or non-reducing loading buffer (125 mM Tris-HCl, 2% SDS, 10% glycerol and 0.001% bromophenol blue, pH 6.8). The samples were denatured for 5 min at 95°C. Proteins were electrophoresed in SDS-polyacrylamide gels. For Western blot analysis proteins were transferred to PVDF membranes (Bio-Rad) by wet electrophoresis transfer (Bio-Rad). The membrane was blocked overnight at 4°C in PBS-T containing 5% FBS, followed by incubation for 1 h at room temperature with horseradish peroxidase conjugated streptactin (IBA) at a dilution of 1:5000 in blocking buffer. After the membrane was washed three times for 15 min each time in PBS-T and once for 15 min in PBS, the signal was detected by ECL kit (GE Healthcare).

Protein expression and purification of the F ectodomain

Expression of the F ectodomain was performed by transient transfection of 293T cells with the expression vector using polyethylenimine (PEI). The supernatant was harvested 5 days after transfection and centrifuged for 3000 x g for 15 min to remove cells and cellular debris. The protein was subsequently affinity-purified from the culture medium by Streptactin sepharose beads (iBA) and eluted from the beads using a biotin elution buffer (2.5 mM biotin, 100 mM Tris (pH 8.0), 150 mM NaCl and 1 mM EDTA). Proteins were additionally purified by gel filtration (AKTA system, GE Healthcare) using the Superdex 200 10/300 column (Amersham Biosciences, GE Healthcare) with HNE buffer (0.15 M NaCl, 0.1 mM EDTA, 5 mM HEPES, pH 7.4) as a running buffer.

Trypsin digestion and acid-treatment of the F ectodomain

The purified F ectodomain in HNE buffer (see above) was incubated with TPCK-treated trypsin at an sF:trypsin ratio of 50:1 (w/w) and incubated for 30 min at room temperature. Digestion was stopped by adding soybean trypsin inhibitor (SBTI) at an SBTI-trypsin ratio of 5:1 (w/w). The cleaved protein was exposed to low pH by adding 0.5 M NaAc pH 5 at an HNE-NaAc ratio of 11:1 (v/v) for 1 h at room temperature.

SDS-PAGE, native-PAGE analysis and size exclusion chromatography of the F ectodomain

For SDS-PAGE analysis the purified protein was prepared in reducing or non-reducing condition and electrophoresed in SDS-polyacrylamide gels as described in western blot analysis. Subsequently the proteins were stained with GelCode Blue Safe Protein Stain (Thermo Scientific) according to its protocol.

For native-PAGE analysis the purified proteins were mixed with sample buffer (Life Technologies). Proteins were electrophoresed in NativePAGE Novex 3-

12% Bis-Tris gel (Life Technologies) in XCell SureLock Mini-Cell system (Life Technologies) and stained with GelCode Blue Safe Protein Stain (Thermo Scientific). The purified sF and sF^{Hismut} at various conditions were analyzed with size exclusion chromatography. Purified protein (90 μg) was loaded onto a Superdex 200 10/300 column equilibrated with HNE buffer (0.15 M NaCl, 0.1 mM EDTA, 5 mM HEPES, pH 7.4) as a running buffer. The elution profile was generated with UNICORN software (GE healthcare) and analyzed with GraphPad software (Prism).

The purified sF and sF^{Hismut} were analyzed at various conditions with size exclusion chromatography. Purified protein (90 μg) was loaded onto a Superdex 200 10/300 column equilibrated with HNE buffer (0.15 M NaCl, 0.1 mM EDTA, 5 mM HEPES, pH 7.4) as a running buffer. The elution profile was generated with UNICORN software (GE healthcare) and analyzed with GraphPad software (Prism).

RESULTS

The pH required for fusion

It has been reported that SeMNPV F protein-mediated fusion can be triggered at pH 5 (150). However, the pH range that allows fusion is not known. Knowing the upper limit at which membrane fusion can occur may indicate which amino acids are responsible for the actual fusion by comparing the pH required for fusion to the pK_a of their side chains. To determine the pH required for membrane fusion and viral entry mediated by SeMNPV F protein we analyzed the entry of SeMNPV F pseudotyped vesicular stomatitis virus (VSVΔG/GFP-F*) under different pH conditions following a protocol previously used for influenza virus (243). VSV entry is known to follow the endocytic pathway (239). In order to test membrane fusion extracellularly at an adjustable pH, the membrane fusion in the endosome is blocked by adding the lysosomotrophic agent ammonium chloride (NH_4Cl), which accumulates in acidic compartments including the endosome (305), and has been used before to inhibit membrane fusion in the endosome (130). To this aim, insect cells were incubated with VSV pseudotyped viruses at an MOI of 0.1 $TCID_{50}$ units / cell on ice in the presence of NH_4Cl. At low MOI and 0°C viruses bind to the cell surface, but are not internalized (243). The cells with prebound viruses were subsequently incubated at various extracellular pHs for a brief period. Virus entry occurs at the plasma membrane if the required pH conditions are met. The viral entry without NH_4Cl, hence undergoing endocytosis, was used as a control. The results show that the virus entry in the presence of NH_4Cl at pH 5.0 (Fig. 2A and B, +NH_4Cl

pH 5.0) is most comparable to that of endocytosed virus particles (Fig. 2A and B, -NH_4Cl or control). When the pH was elevated the virus entry decreased. At pH 6.0 the virus entry was dramatically reduced (below 50%), which is considered as the upper limit for efficient virus entry. This pH is closely related to the pK_a of histidine side chain, suggesting that SeMNPV F histidine protonation may play a role as pH sensor in membrane fusion.

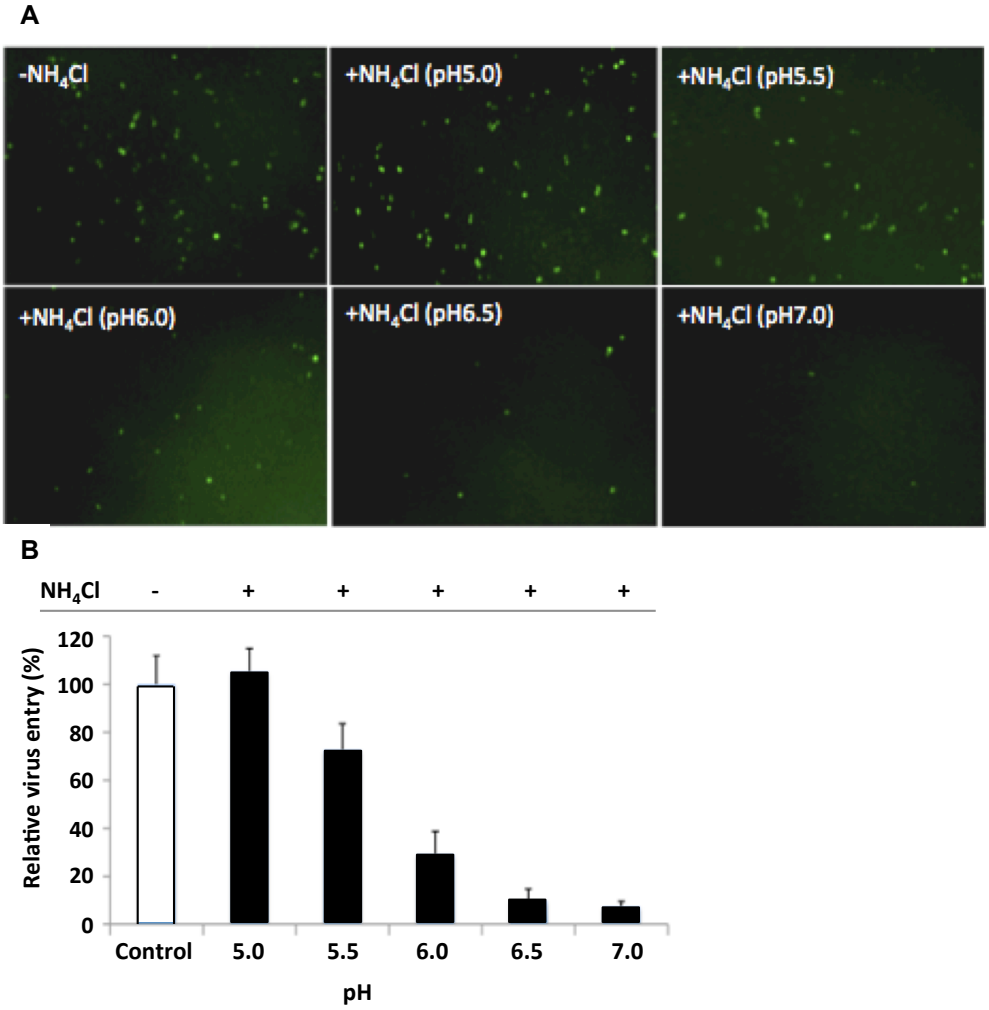


Fig. 2. Analysis of low-pH induced membrane fusion. In the presence of NH_4Cl, an inhibitor of endocytic acidification, SeUCR cells were inoculated with VSVΔG/GFP-F* on ice for 1 h to allow virus binding to the plasma membrane. The cells were then treated with prewarmed PBS at various pH for 2 min and recovered in culture medium containing NH_4Cl at 27°C for 24 h. The efficiency of virus entry was monitored as GFP positive cells using inverted fluorescence microscopy (A) and virus entry was quantified and compared to the control without NH_4Cl (B).

Inhibition of virus entry by chemical modification of F protein

The pH-sensing role of histidines has been identified for several viral fusion proteins (98, 205, 349). We utilized the histidine-modifying agent DEPC to prevent histidine protonation at the histidine side chain (46). The infectivity of DEPC-treated VSVΔG/GFP-F* virus dropped as the concentration of DEPC increased (Fig. 3). This result suggests that histidines in the F protein function as pH sensors.

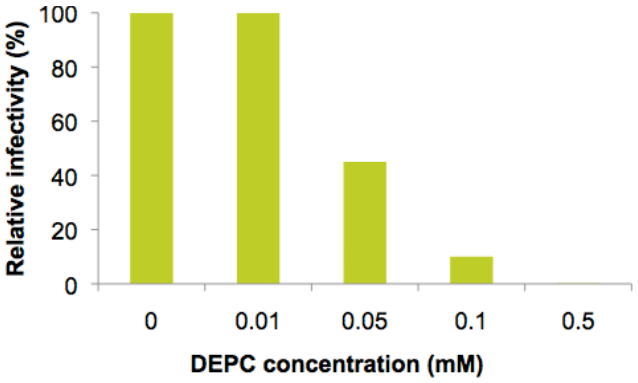


Fig. 3. Effect of DEPC on VSVΔG/GFP-F* virus. The pseudotyped VSV virus was incubated with DEPC at various concentrations for 10 min at 22°C. SeUCR cells were inoculated with the pre-treated viruses for 24 h and virus entry (relative to without DEPC) was determined using end-point dilution assays.

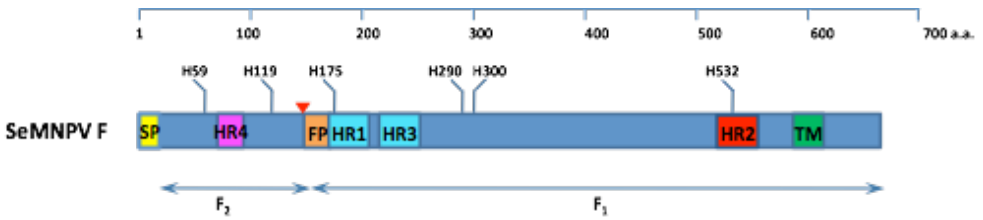


Fig. 4. Schematic representation of the SeMNPV F protein with indicated locations of conserved histidine residues. SP: signal peptide, FP: fusion peptide, TM: transmembrane domain, CT: cytoplasmic tail, HR: heptad repeat region. The furin cleavage site is indicated as a red arrowhead. The F_2 subunit is at N-terminal side and the F_1 subunit at the C-terminal side of the protein.

Table 1. Alignment of the histidine residues of SeMNPV F protein and its homologues using PRALINE (http://zeus.few.vu.nl). Conserved histidine residues are shown in black, and variable residues are shown in red. The histidine residues in SeMNPV F protein are shown in bold.

Genus	Name	Abbreviation	59	119	175	290	300	532
	***Spodoptera exigua* MNPV**	**SeMNPV**	**H**	**H**	**H**	**H**	**H**	**H**
	Spodoptera frugiperda MNPV	SfMNPV	H	H	H	H	H	H
	Agrotis ipsilon NPV	AgipNPV	H	H	H	H	H	H
	Spodoptera litura NPV	SpliNPV	H	H	H	H	H	H
	Mamestra configurata NPV	MacoNPV	H	H	H	H	H	H
	Lymantria xylina MNPV	LyxyMNPV	H	H	H	H	H	H
	Lymantria dispar MNPV	LdMNPV	H	H	H	H	H	H
	Apocheima cinerarium NPV	ApciNPV	H	H	H	H	H	H
	Hemileuca species NPV	HespNPV	H	H	H	H	H	H
Alphabaculovirus (Group II)	*Clanis bilineata* NPV	ClbiNPV	H	H	H	H	H	H
	Chrysodeixis chalcites NPV	ChchNPV	H	H	H	H	H	H
	Adoxophyes honmai NPV	AdhoNPV	H	H	H	H	H	H
	Adoxophyes orana NPV	AdorNPV	H	H	H	H	H	H
	Euproctis pseudoconspersa NPV	EupsNPV	H	H	H	R	N	H
	Leucania separate NPV	LeseNPV	Q	H	H	Y	H	H
	Trichoplusia ni SNPV	TnSNPV	H	H	H	F	H	H
	Helicoverpa armigera NPV	HearNPV	H	N	H	H	H	H
	Orgyia leucostigma NPV	OrleNPV	H	H	H	A	H	H
	Agrotis segetum NPV	AgseNPV	H	H	H	F	H	H
	Ectropis obliqua NPV	EcobNPV	H	H	H	H	H	Y
	Clostera anastomosis GV	CaLGV	H	H	Y	D	S	Y
	Clostera anachoreta GV	ClanGV	H	H	Y	D	S	Y
	Helicoverpa armigera GV	HearGV	H	H	H	H	H	L
	Choristoneura occidentalis GV	ChocGV	Y	H	Y	H	N	F
	Phthorimea operculell GV	PhopGV	Y	H	Y	H	F	Y
Betabaculovirus	*Pseudaletia unipuncta* GV	PsunGV	Y	H	Y	T	H	L
	Spodoptera litura GV	SpliGV	F	H	Y	Y	H	F
	Epinotia aporema GV	EpapGV	Y	H	Y	L	H	Y
	Andraca bipunctata GV	AnbiGV	Y	H	Y	T	S	Y
	Pieris rapae GV	PiraGV	Y	H	Y	T	N	F
	Adoxophyes orana GV	AdorNPV	Y	H	Y	A	D	F
Deltabaculovirus	*Culex nigripalpus* NPV	CuniNPV	A	C	M	H	P	S

Table. 2. Conservation of histidine residues among baculovirus F proteins with exclusion of the F-like proteins* in Group I alphabaculoviruses (156) (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons).

	% conservation of F proteins	
	Group II alphabaculoviruses	All baculoviruses except group I alphabaculoviruses*
H59	95	73
H119	94	91
H175	100	76
H290	80	45
H300	94	67
H532	96	64

*F-like proteins in group I alphabaculoviruses were left out. Gammabaculoviruses do not have F proteins.

Conservation of F histidine residues

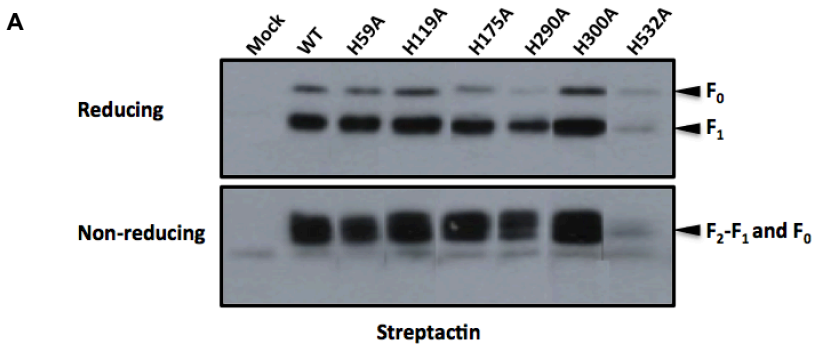
Sequence alignment of SeMNPV F with homologues of other group II alphabaculoviruses revealed six highly conserved histidines (≥80%). H59 and H119 are located in the F2 subunit, while H175 and H532 are located in the HR1 and HR2 regions, respectively. H290 and H300 are located between the HR3 and HR2 regions (Fig. 4). Alignment of group II alphabaculovirus F proteins and more distantly related homologues of betabaculoviruses and deltabaculoviruses shows variable conservation levels for the histidines (Table 2): H119 is highly conserved (91%) among all baculovirus F proteins; the other histidines are less well conserved (<70%). We used the ConSurf server for the prediction of the histidine residues important for protein structure or function by sequence alignment of group II alphabaculoviruses (http://consurf.tau.ac.il) (23). H59, H119, H175, H290 and H532 are predicted as exposed residues, for which H59, H119, H175 and H532 are highly conserved and with a predicted functional role. H300 is predicted as a highly conserved, but 'buried' residue of structural importance (Fig. S1).

Expression and cell surface distribution levels of F protein and histidine mutants

The role of the six conserved histidines of the F protein in virus entry was analyzed using alanine substitutions for each individual histidine. However, we realized that the alanine substitutions may affect protein production and transport of F. Therefore we first examined whether the mutant F proteins passed the quality control in the endoplasmic reticulum (ER) after synthesis (334, 377). Endoproteolytic cleavage by furin was considered as an indication of correct folding of F, since endogenous furin is located in the trans-Golgi network in both

mammalian and insect cells (176, 273), to which properly folded membrane and secreted proteins are transported from the ER (377). Sf9 cells transfected with plasmids encoding strep-tagged wt or histidine mutated F were analyzed via SDS-PAGE under reducing and non-reducing conditions followed by western blotting using streptactin (Fig. 5A). Like wt F protein all the histidine mutants were cleaved by furin as indicated on the blot with the majority being in the F_1 form. A small fraction was still in the F_0 form as can be seen in the reducing gel (Fig. 5A, upper panel). F_2 is not detectable on this blot as it does not have a strep-tag. Disulfide-linked F_1 and F_2 (indicated as F_2-F_1) were observed on the non-reducing gel (Fig. 5A, lower panel), in accordance with previous work (411). However, the ratio cleaved product to uncleaved protein (Fig. 5A, F_1 and F_0) for construct H532 is lower than for the wt construct, suggesting a larger percentage of F-H532A proteins were retained in the ER and thus failed to transport to the plasma membrane. These results show that protein folding of histidine mutants except F-H532A was apparently normal for all the histidine mutants since these proteins were transported to the trans-Golgi network.

Next we determined how much of the F protein was displayed on the cell surface. Transfected Sf9 cells were fixed without permeablization and subsequently incubated with SeMNPV F ectodomain polyclonal antibody and Alexa Fluor488-labeled secondary antibody at 48 h post transfection. The treated cells were analyzed with FACS. Most of the histidine mutants were detected on the cell surface at a similar level as wt F protein (Fig. 5B). A notable reduction was observed for H532 due to the low total expression level. However, all the proteins were easily detected again suggesting correct folding and routing to the cell surface.



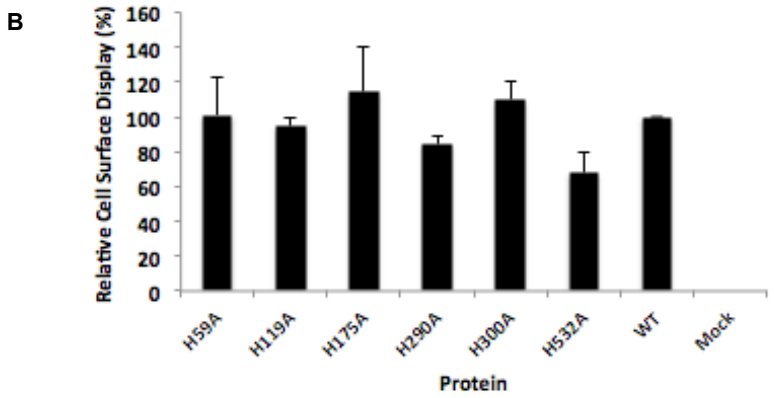


Fig. 5. Transient expression of F protein and histidine mutants. Sf9 cells were transfected with 1 μg of plasmid DNA encoding F or histidine mutants and harvested 48 h after transfection. The F protein levels of the cell lysates were analyzed under reducing (upper panel) and non-reducing (lower panel) conditions on SDS-PAGE followed by western blotting using HRP-conjugated streptactin (A). The uncleaved F_0 protein and the cleavage products F_1 under reducing conditions, and the uncleaved F_0 and disulfide linked cleavage product F_2- F_1 under non-reducing conditions are indicated on the upper and lower panels, respectively. The cell surface distribution levels of histidine mutants of F (relative to the wt F protein) were measured on fixed cells using polyclonal antibody against the SeMNPV F ectodomain. The graphical data (B) are the results of duplicate transfections. Error bars represent the standard deviation of the means.

Fusion activity of F protein and histidine mutants

Membrane fusion activity of the wt and histidine mutated F proteins was first evaluated in a syncytium formation assay. The cells transfected with F constructs were shortly exposed to acidified buffer 48 h after transfection and analyzed for the average percentage of nuclei in a syncytium. Single alanine replacement at position H59 or H119 showed severe abrogation of cell-cell fusion activity as no syncytia were observed (Fig. 6A). Alanine substitution of the other conserved histidines (H175, H290, H300 and H532) had no significant effect on syncytium formation.

Next, we examined the fusion activity of the histidine-modified F proteins by looking at the infectivity of VSV-GFP viruses pseudotyped with these F proteins. VSV virus lacking fusion proteins was used as a negative control. These pseudotyped viruses were used to inoculate SeUCR cells. At 24 h post infection, virus entry (GFP positive cells) was analyzed using inverted fluorescence microscopy. GFP signals were observed for all pseudotyped VSV viruses (Fig. 7), indicating that they were all able to enter the cells. Surprisingly, F proteins with histidine mutations at position of H59 or H119 were able to mediate virus entry of VSVpv (Fig. 7, H59A and H119A) despite their inability to induce syncytium

formation (Fig. 6A and B, H59A and H119A). Because the syncytium formation requires larger pore sizes than virus-cell fusion and the wt F protein was able to mediate syncytium formation (Fig. 6 and 7, WT) the defect of syncytium formation caused by H59A and H119A appears to relate to a problematic pore enlargement.

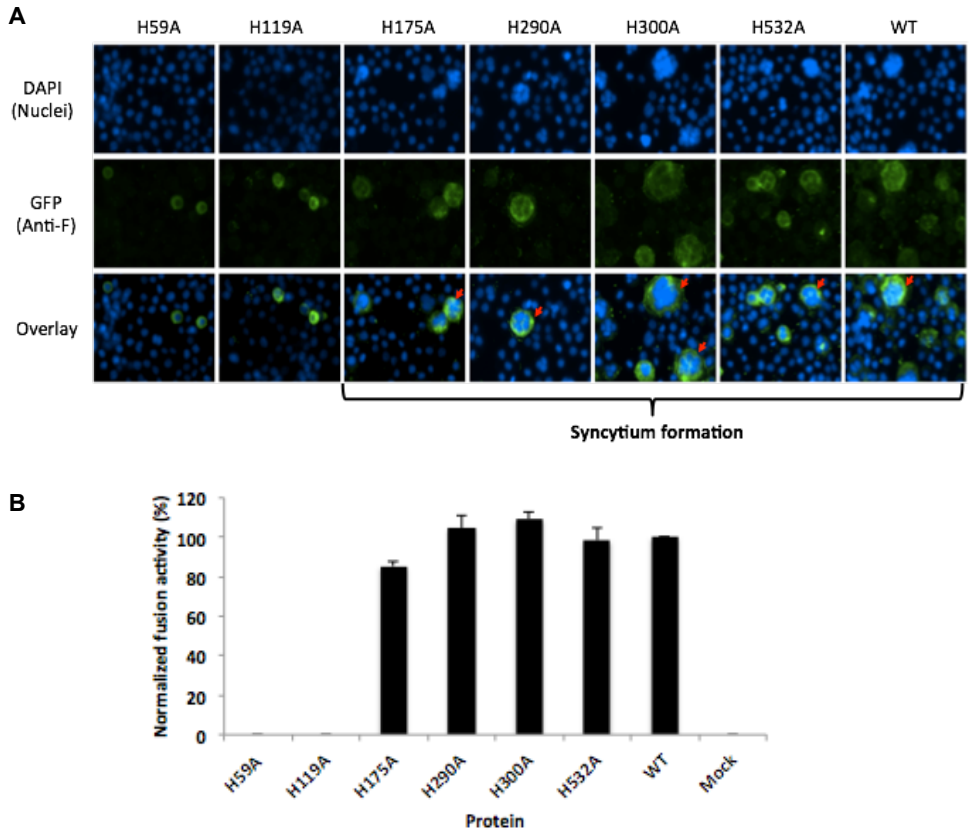


Fig. 6. Cell-cell fusion activity of wt F protein and histidine mutants. Sf9 cells were transfected with plasmids encoding F protein or histidine mutants as indicated. At 48 post transfection cells were incubated in PBS at pH 5.0 for 2 min and then in Sf900 II medium (pH 6.3) for 4 h. After fixation cells were stained with DAPI (A, upper panel) and the SeMNPV F ectodomain was detected with polyclonal antibody and Alexa Fluor488-labeled secondary antibody (A, middle panel). The stained cells and nuclei were visualized and photographed with an inverted fluorescence microscope using DAPI and GFP channels, respectively. Syncytia are indicated with red arrows in middle panel. The relative fusion activity was evaluated by measuring the relative percentage of nuclei found in syncytia compared to nuclei in all the cells, which was normalized for the cell surface display level of each F protein variant (Fig. 5B). For each transfection five microscopic fields were analyzed. The graphical data shown (B) are the results of duplicate transfections. Error bars represent the standard deviation of the means.

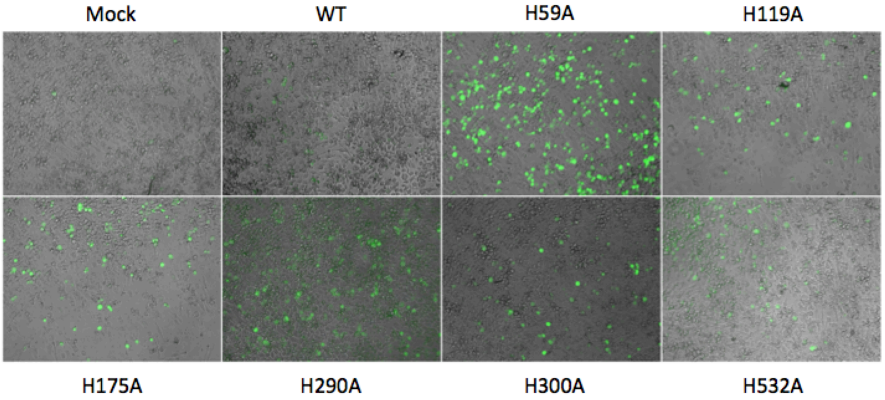


Fig. 7. Virus entry of VSV-GFP pseudotyped with wt or histidine mutated F protein. SeUCR cells were inoculated with the various VSV-GFP pseudotyped viruses as indicated. At 24 h p.i. microscopic images were recorded using a combination of bright field and GFP channel imaging.

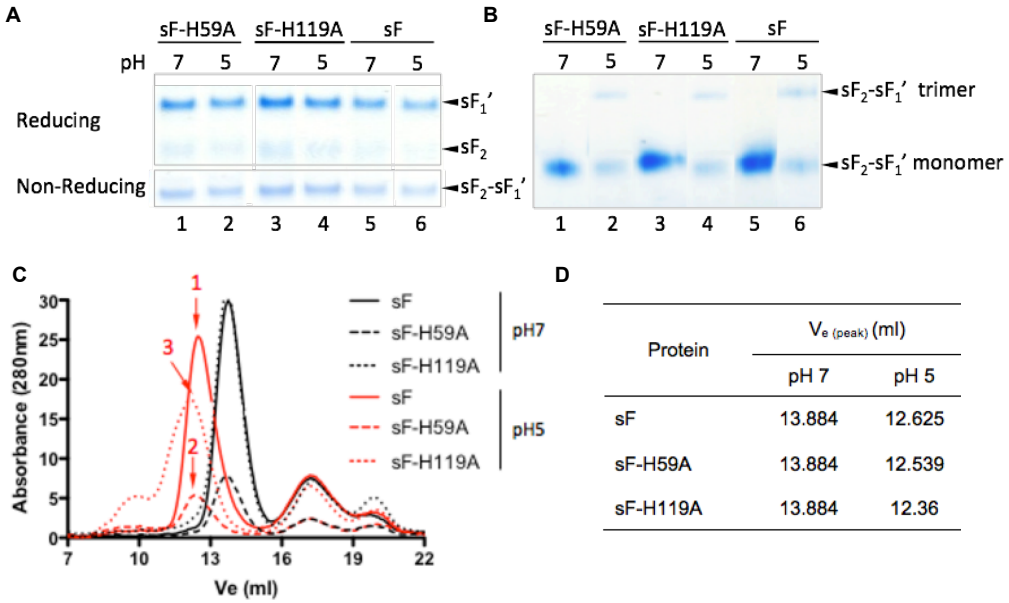
Low pH-triggered conformational changes of soluble F protein with histidine mutation

The results above showed that mutations H59A and H119A completely abolished syncytium formation, yet were still capable of rescuing virus entry of the G-null VSV virus. We therefore asked whether F protein containing either of the two mutations is able to undergo conformational changes triggered by low pH. Our previous biochemical studies showed that monomeric F ectodomain (soluble F protein, sF), representing the core structure of the F protein undergoes conformational arrangement and forms trimers upon low pH (Chapter 2). In the current experiment, the oligomerization of soluble F protein with alanine substitution at H59 or H119 (sF-H59A and sF-H119A) was studied in comparison to wt sF at low and neutral pH using native-PAGE and size exclusion chromatography (Fig. 7).

To prevent low pH triggered aggregation due to exposure of the hydrophobic fusion peptide the proteins were treated with trypsin at a desired concentration, which was determined by limited proteolysis (Fig. S2), to remove the fusion peptide. The F protein at either low or neutral pH was treated with the determined concentration of trypsin for comparison and trypsin cleavage was examined using SDS-PAGE. All F protein variants were cleaved at the furin cleavage site as indicated with separated sF_1' (sF_1' is FP-deficient sF as described above) and sF_2 polypeptides under reducing conditions, and disulfide-linked sF_2-sF_1' band under non-reducing conditions (411) (Fig. 7A).

Native-PAGE shows that trypsin cleaved sF-H59A and sF-H119A had similar oligomeric states as wt sF at pH 5 (trimer) and pH 7 (monomer), suggesting that

both histidine-mutated proteins were capable of responding to low pH with conformational rearrangement in the same manner as sF (Fig. 7B). However, size exclusion chromatography showed that histidine mutants sF-H59A and sF-H119A sF (peak fraction no. 2 and 3 in Fig. 8C) were eluted earlier than sF (peak fraction no. 1 in Fig. 8C) at low pH but not at neutral pH (Fig. 8C, pH 5 vs. pH 7 shown as red vs. black curves), suggesting that H59A and H119A are involved in the low pH-triggered conformation of soluble F protein. In summary, although sF-H59A and sF-H119A are low pH sensitive, their conformation at low pH exhibits differences with sF.



Protein	$V_{e\ (peak)}$ (ml)	
	pH 7	pH 5
sF	13.884	12.625
sF-H59A	13.884	12.539
sF-H119A	13.884	12.36

Fig. 8. Effects of acid treatment on trypsin cleaved wt and histidine mutated soluble F protein. Purified proteins in HNE buffer (pH 7) were treated with trypsin at an sF^{FCSmut}-trypsin ratio of 50:1 (w/w) and incubated for 30 min at room temperature. Digestion was stopped by adding soybean trypsin inhibitor (SBTI) at an SBTI-trypsin ratio of 5:1 (w/w). The trypsin-digested proteins were exposed to neutral (pH 7) or acidic pH (pH 5) and analyzed by SDS-PAGE under reducing (upper panel) and non-reducing (lower panel) conditions (A), native-PAGE (B), and size exclusion chromatography (C). The peak elution volumes are listed in (D). Oligomeric states of sF at neutral and low pH were studied previously (Chapter 2). sF: wt soluble F; sF-H59A: soluble F with alanine substitution at position H59; sF-H119A: soluble F with alanine substitution at position H119. The three peaks marked with 1, 2 and 3 represent the sF, sF-H59A and sF-119A proteins at pH 5, respectively.

DISCUSSION

In this study we explored the role(s) of six conserved histidine residues in the SeMNPV F protein in membrane fusion using single amino acid substitutions.

Two histidine mutations at H59 and H119 completely could still enter the cells but abolished the syncytium formation (Fig. 6A), which requires larger pore size than virus entry (Fig. 7). This suggests that these histidines may play an important role in fusion pore expansion, but not in initial pore formation. A similar role of histidine residues in AcMNPV GP64 has been suggested previously (205). Although native-PAGE analysis showed that monomeric F ectodomains with the H59A or H119A mutations responded to low pH and underwent trimerization like the wt F ectodomain (Fig. 7B), size exclusion chromatography indicated differences between these histidine mutated and wt F ectodomains in the low-pH triggered trimeric conformations (Fig. 8C). This suggests that H59 and H119 may regulate the conformation only in a charged state.

Based on our crystallography studies (Chapter 3), the protonated H119 is predicted to form hydrogen bonds with H59, L115 and V235 within the same protomer (Fig. 9). H119 and L115 are located in Helix A and H59 is in Helix B, both of which are short helices adjacent to and nearly parallel to the HR3 region. V235 is located in the HR3 region. Interestingly, H119 is at the "apex" position in the "tripod" interaction of these three hydrogen bonds (Fig. 10) formed by H119 with H59 on Helix A, with L115 on Helix B and with V235 on HR3, respectively, and therefore stabilizes HR3 through V235 in the postfusion structure with the aid of the other two neighboring residues H59 and L115. Since the HR3 region is directly connected to HR1 and to the fusion peptide, (not shown in the current postfusion structure) H119 as well as H59 might play an essential role in modulating the configuration of the fusion peptide. Thus, H59 and H119 may function in fusion pore expansion by regulating the movement of the coiled coil regions (HR3 and HR1) and the fusion peptide. Particularly H119 may play an important role in triggering the conformational changes related to pore expansion. Significant high-level conservation among the F proteins in the baculovirus family is a reflection of the important functionality of H119 (Table 2).

Mutation of the other conserved histidine residues H175, H290, H300 and H532 did not show a significant effect on membrane fusion ability as concluded from both the syncytium formation assay (Fig. 6A and B) and the virus entry experiment (Fig. 7). Based on the recently determined postfusion structure we also analyzed the location and hydrogen bonding interactions of H290 and H300 with neighboring residues. H290 forms a hydrogen bond with D294 within a small helix above HR3 in the 'head' region in the structure. The ionic interaction

between H290 and D294 is likely to stabilize the small helix, which may only have loose structural interaction with HR3 and neighboring peptides. H300 forms a hydrogen bond with I312 between two β-strands, which are connected via a β-turn in an antiparallel direction (Fig. 9). The β-strands are connected to the N-terminal neck region with very flexible loops. Therefore H300 and I312 may function in stabilizing the β-sheet structure of neck region, which are composed of these β-strands. H175 and H532 are located in HR1 and HR2, respectively, which are important components of the pH-sensitive core structure of class I fusion proteins (86, 189, 213, 248). Therefore these two histidines are most likely involved in pH sensing. Single substitution did not reveal the roles of these histidines; however, each of these two histidines might function as an independent pH sensor, which might be explained by evolutionary selection pressure and by constant mutation events in viruses, so that membrane fusion still can be triggered by the other histidine residue when one of the pH sensors is mutated. This is in line with other studies, suggesting the presence of multiple pH sensors in various viruses (48, 75, 101, 162, 205, 242).

Besides the 6 highly conserved histidines described above (>80% conservation in group II alphabaculovirus) there are 14 less conserved H residues in the F ectodomain, 9 of which are present in the part of the protein for which the crystal structure is currently solved. (Fig. S4 and Table S2) and all form hydrogen bonds with neighboring residues. One of these 9 histidines, H52 forms hydrogen bonds with L495 and T497 from a neighboring protomer (Fig. S3). H52 and the neighboring residues are located on paralleled β-strands in two different protomers, which may contribute to stabilize the trimeric conformation in the postfusion state. Four of the less conserved histidines H338, H327, H470 and H489 form hydrogen bonds with neighboring residues in adjacent β-strands or their connection loops in the head region of the F structure (Fig. S5). These four histidine and interacting neighboring residues may function in stabilizing the head region in the postfusion structure. H253, which is located in a loop directly connecting the HR3 region has a hydrogen bond with residue N248 located in the HR3 region, suggesting H253 may be involved in formation and stabilizing of the HR3 helix. Moreover 5 histidines of low conservation level in or proximal to the fusion peptide, the HR1 or HR2 regions (location marked in red in Fig. S4) are not present in the current postfusion crystal structure. Due to the structural importance of the fusion peptide and the heptad repeat regions and the accumulated evidence of pH triggers in these regions (11, 50, 162, 328, 426) the histidine residues in these regions require further study to reveal their functions.

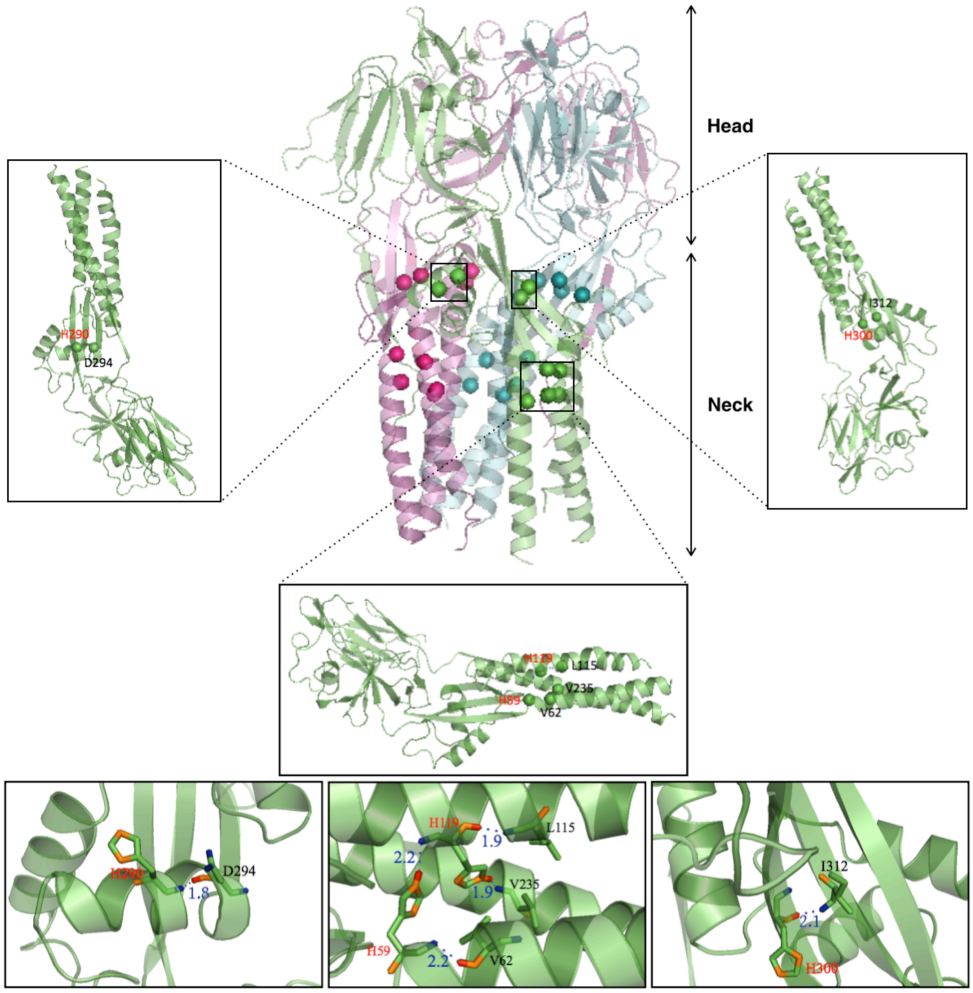


Fig. 9. The potential hydrogen bonds between highly conserved histidines and proximal residues in the postfusion structure (See chapter 3). The trimeric structure of F (protomer A, B and C in green, cyan and magenta, individually) is shown in the middle. The histidine and proximal residues with potential hydrogen bonds are shown as spheres and grouped with boxes. The monomer in green is viewed at different angles (left, bottom and right side to the trimeric structure) for zoomed-in hydrogen bonds. The hydrogen bonds with atom distance are displayed at the bottom panel. Nitrogen, oxygen and hydrogen atoms are colored in orange, blue and red. Names of histidine and neighboring residues are indicated as red and black respectively. Hydrogen bonds are indicated as blue dashed lines and distance of each bond is shown in angstroms. Protonated H290 forms a hydrogen bond with D294; H119 forms hydrogen bonds with H59, L115 and V235; H59 forms a hydrogen bond with V62; H300 forms a hydrogen bond with I312. Illustrated hydrogen bonds were predicted using WHAT IF web server (http://swift.cmbi.ru.nl/ (142)). The selection of hydrogen bonds was based on the probability score using optimal hydrogen bond calculation. The figures were generated with PyMOL molecular visualization system (version 1.3, DeLano Scientific LLC, 2009).

So far we examined the most highly conserved histidine residues by single mutation and showed the importance for membrane fusion of two conserved histidine residues in the F_1 subunit. However, we cannot exclude that other histidines are also involved in membrane fusion. In general the process of conformational rearrangements leading to membrane fusion involves: i) extension of fusion peptide; ii) hemifusion; iii) fusion pore formation and iv) fusion pore enlargement. Furthermore, the pH sensing may occur at different stages and may be conducted by different residues at each stage. From analysis of potential hydrogen bond formation and position of histidines in the postfusion structure the histidine residues in and near HR regions are most likely involved in triggering the initial conformational change. Our experimental results and structural analysis suggest that two of the highly conserved histidines H59 and H119 may be involved in fusion pore enlargement. More information on the role of histidines requires knowledge on the difference between the hydrogen bond networks in the pre-and postfusion states of F proteins.

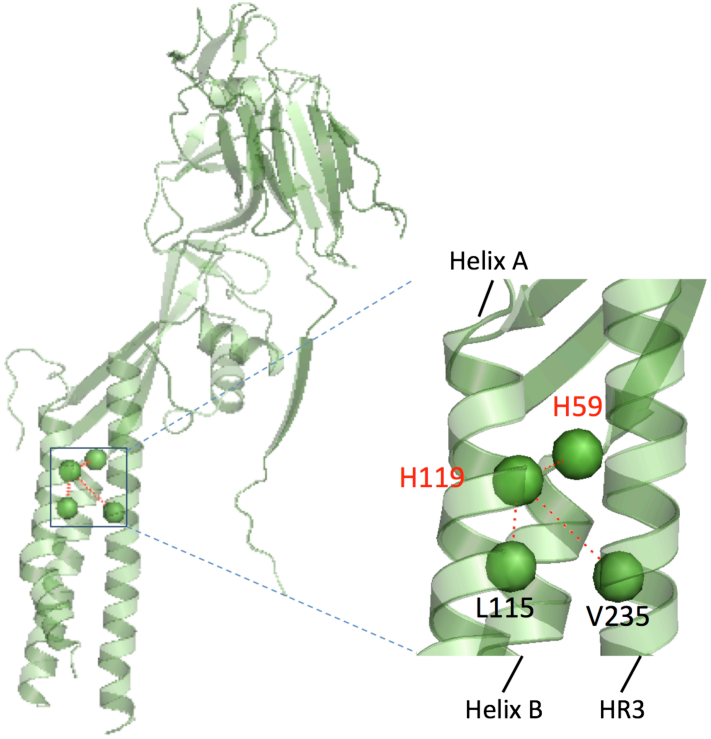


Fig. 10. The potential hydrogen bonding interactions of H119 and neighboring residues in one F protomers in postfusion structure. The hydrogen bonds are indicated as red dashed lines. Two short helices flanking to HR3 are named Helix A and Helix B. H119 and L115 are located on Helix A. H59 is located on Helix B. V235 is located on HR3 region.

SUPPLEMENTARY DATA

Table S1. Primers for histidine mutations.

Primer name	Sequence (5' to 3')
H59A_Forward	TGTCATCGAAATGGACGCCGGATCGGTGTTCTATCG
H59A_Reverse	CGATAGAACACCGATCCGGCGTCCATTTCGATGACA
H119A_Forward	ACCGAATCTAGCGCAGCAGGCCAATTTGCTCGATCAAAAAG
H119A_Reverse	CTTTTTGATCGAGCAAATTGGCCTGCTGCGCTAGATTCGGT
H175A_Forward	CGACGACGCTCACGAACTGGCCATGCTCGCCAACACCAC
H175A_Reverse	GTGGTGTTGGCGAGCATGGCCAGTTCGTGAGCGTCGTCG
H290A_Forward	TCTCACCGAAAAAGCGATGGCCGTTCTCATCGACAACG
H290A_Reverse	CGTTGTCGATGAGAACGGCCATCGCTTTTTCGGTGAGA
H300A_Forward	CGACAACGTCATCAATGTGGCTGTGTTTGTCACGGCGGAAC
H300A_Reverse	GTTCCGCCGTGACAAACACAGCCACATTGATGACGTTGTCG
H532A_Forward	CAACGATCTCGAAGCTACAAATCTTCACGAACTGACCAG
H532A_Reverse	CTGGTCAGTTCGTGAAGATTTGTAGCTTCGAGATCGTTG

Table S2. Conservation of histidine residues in F ectodomain with <80% conservation among group II alphabaculoviruses (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons).

	% conservation of F proteins		Presence in crystal structure
	Group II alphabaculoviruses	All baculoviruses except group I	
H52	55	50	Yes
H70	71	47	Yes
H106	42	37	Yes
H157	47	51	No
H172	54	45	No
H204	76	52	No
H207	45	36	No
H253	55	59	Yes
H274	64	51	Yes
H327	69	60	Yes
H338	41	37	Yes
H470	43	44	Yes
H489	46	46	Yes
H536	53	41	No

Fig. S1. Prediction of functional/structural related residues (ConSeq) using PRALINE sequence alignment of group II alphabaculovirus F proteins (http://www.ibi.vu.nl/programs/pralinewww) (23).

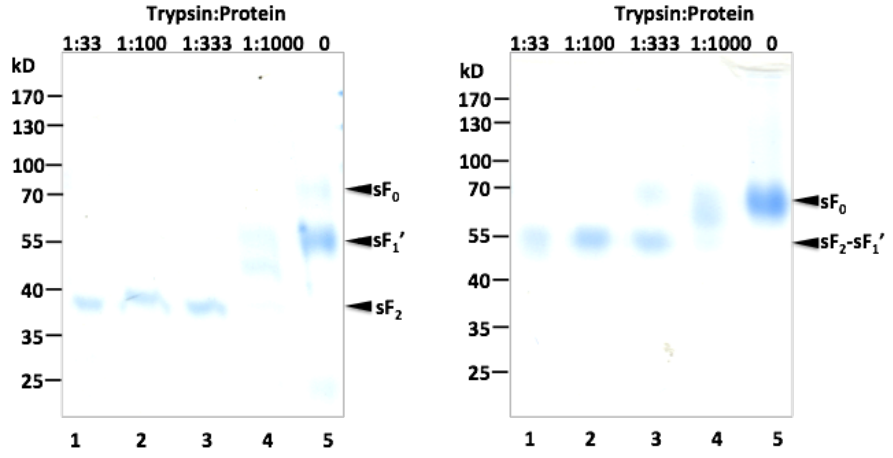


Fig. S2. Limited proteolysis of F ectodomain (sF). sF was mixed with trypsin in different ratio's as indicated and incubated at 22°C for 30 min and analyzed by SDS-PAGE under reducing (A) and non-reducing (B) conditions.

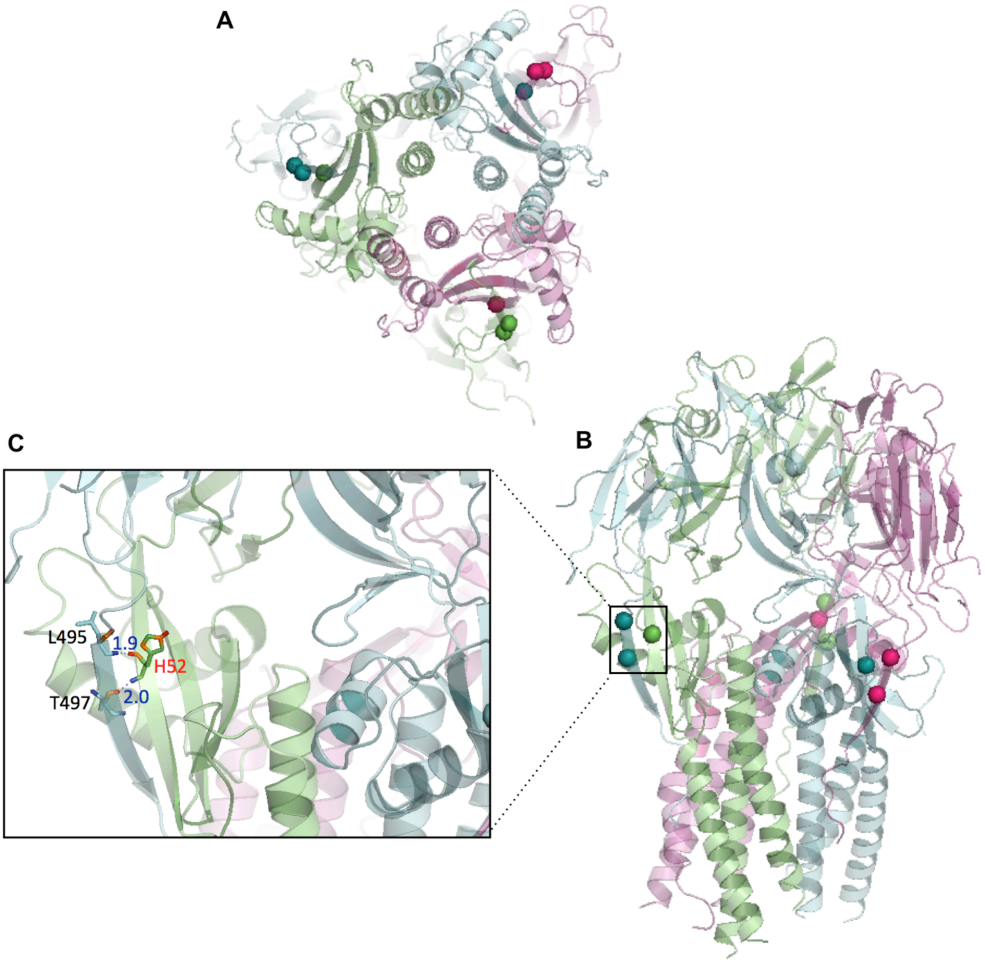


Fig. S3. The potential interprotomer hydrogen bonds between H52 and proximal residues. A) and B) Top view and side view of trimeric structure of postfusion F. Protomer A, B and C are shown in green, cyan and magenta, respectively. The histidine and proximal residues with potential hydrogen bonds are shown as spheres. C) Zoomed-in hydrogen bonds between protomer A (green) and B (cyan). Protonated H52 in protomer A forms hydrogen bonds with L495 and T497 in protomer B. Similarly, H52 in protomer B forms hydrogen bonds with the neighboring residues in protomer C and H52 in protomer C forms hydrogen bonds with L495 and T497 in protomer A. Nitrogen, oxygen and hydrogen atoms are colored in orange, blue and red. Histidine and neighboring residues are indicated as red and black respectively. Hydrogen bonds are indicated as blue dashed lines and distance of each bond is shown with dash lines in angstroms. Illustrated hydrogen bonds were predicted using WHAT IF web server (http://swift.cmbi.ru.nl/ (142)). The selection of hydrogen bonds was based on the probability score using optimal hydrogen bond calculation. The figures were generated with PyMOL molecular visualization system (version 1.3, DeLano Scientific LLC, 2009).

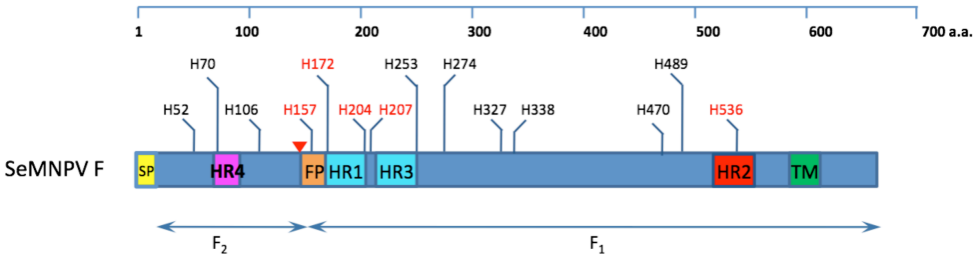
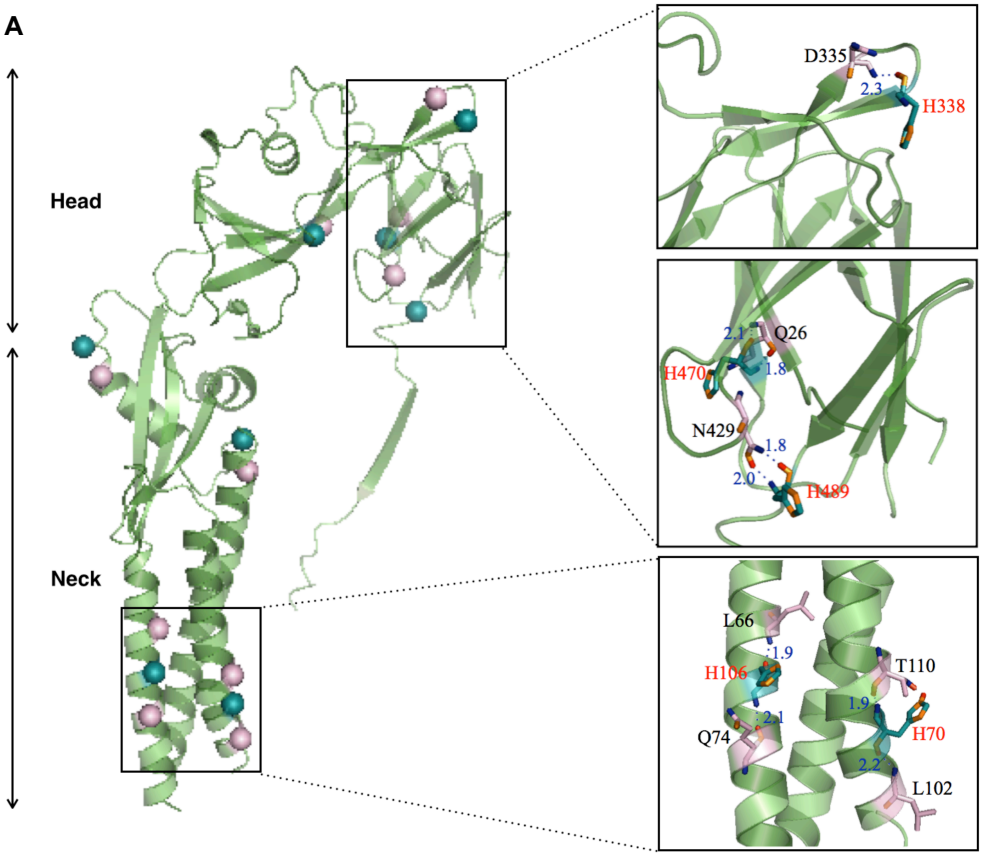


Fig. S4. Locations of histidine residues of low conservation (<80% conservation among group II alphabaculoviruses). The histidines absent in crystal structure of postfusion state are colored in red. SP: signal peptide FP: fusion peptide, TM: transmembrane domain, HR: heptad repeat region. Furin cleavage site is indicated as red arrowhead. F_2 subunit is at N-terminal side and F_1 subunit is at C-terminal side of protein.



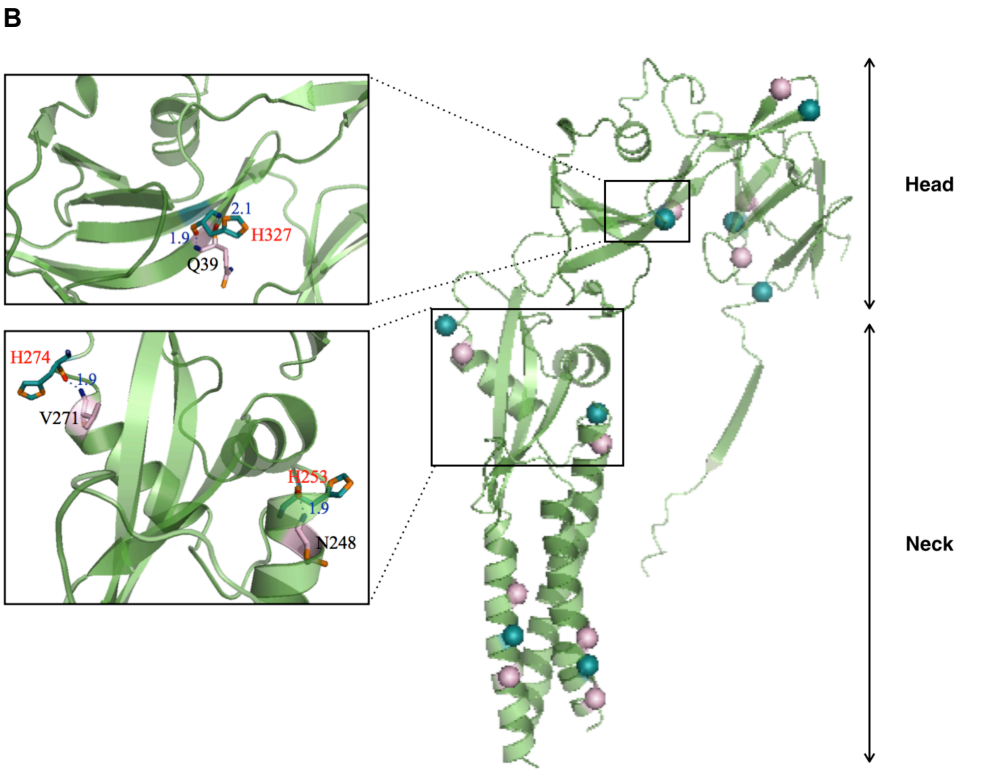


Fig. S5. The potential hydrogen bonds between less conserved histidines and proximal residues in one F protomer. In the overall protomer structure the histidine and proximal residues with potential hydrogen bonds are shown spheres in teal and magenta colors respectively and grouped with boxes. Next to the overview of protomer structure the grouped residues and hydrogen bonds are illustrated in detail. Nitrogen, oxygen and hydrogen atoms are colored in orange, blue and red. Names of histidine and neighboring residues are indicated as red and black respectively. Hydrogen bonds are indicated as blue dashed lines and distance of each bond is shown in angstroms. A) Protonated H338 forms a hydrogen bond with D335; H470 forms two hydrogen bonds with Q26; H489 forms two hydrogen bonds with N429; H106 forms hydrogen bonds with L66 and Q74; H70 forms hydrogen bonds with T110 and L102. B) Protonated H327 forms two hydrogen bonds with Q39; H274 forms a hydrogen bond with V271; H253 forms a hydrogen bond with N248. Illustrated hydrogen bonds were predicted using WHAT IF web server (http://swift.cmbi.ru.nl/ (142)). The selection of hydrogen bonds was based on the probability score using optimal hydrogen bond calculation. The figures were generated with PyMOL molecular visualization system (version 1.3, DeLano Scientific LLC, 2009).

Chapter 6

Budded baculovirus particle structure revisited

Qiushi Wang[1,2], Berend-Jan Bosch[2], Kalliopi Pervolaraki[2], Just M. Vlak[1], Monique M. van Oers[1], Peter J. Rottier[2], Jan van Lent[1]

[1] Laboratory of Virology, Wageningen University, Droevendaalsesteeg 1, 6708 PB, Wageningen, the Netherlands

[2] Virology Division, Department of Infectious Disease and Immunology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, the Netherlands

Journal of Invertebrate Pathology, submitted

ABSTRACT

Baculoviruses are a group of enveloped, double-stranded DNA insect viruses with budded (BV) and occlusion-derived (ODV) virions produced during their infection cycle. BVs are commonly described as rod shaped particles with a high apical density of protein extensions (spikes) on the lipid envelope surface. However, due to the fragility of BVs the conventional purification and electron microscopy (EM) staining methods considerably distort the native viral structure. Here, we use cryo-EM analysis to reveal the near-native morphology of two intensively studied baculoviruses, *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV) and *Spodoptera exigua* MNPV (SeMNPV), as models for BVs carrying GP64 and F as envelope fusion proteins on the surface. The now well-preserved AcMNPV and SeMNPV BV particles have a remarkable elongated, ovoid shape leaving a large, lateral space between nucleocapsid (NC) and envelope. Consistent with previous findings the NC has a distinctive cap and base structure interacting with the envelope. This interaction may explain the partial retaining of the envelope on both ends of the NC and the disappearance of the remainder of the BV envelope in the negative-staining EM images. Cryo-EM also reveals that the viral envelope contains two layers with a total thickness of ≈ 6-7 nm, which is significantly thicker than a usual biological membrane (< 4 nm). Most peplomers are densely clustered at the two apical ends of the virion although some envelope proteins are also found more sparsely on the lateral regions. The spikes on the surface of AcMNPV BVs are distinctly different from those of SeMNPV. Based on our observations we propose a new near-native structural model of baculovirus BVs.

INTRODUCTION

Baculoviruses are a family of enveloped, large double-stranded DNA viruses that infect arthropods. The genome of baculoviruses ranges from 80 to 180 kb in size and contains between 90 and 180 genes (134). The viral genome is packed into rod-shaped nucleocapsids (NC) of 30-70 nm in diameter and 200-400 nm in length (37, 156). Two phenotypes of infectious enveloped virions are produced during the infection cycle: the occlusion-derived viruses (ODV) that initiate infection of the midgut of the host upon oral ingestion and budded viruses (BV) that are responsible for cell-to-cell spread and further systemic infection. ODVs obtain their *de novo* synthesized envelopes in the nucleus while BVs obtain their envelope from the plasma membrane upon budding into the extracellular space or into neighboring cells (282). Although the two phenotypes contain identical viral genomes and nucleocapsids, they differ in the lipid and protein composition

and structure of their envelopes (339). The ODV envelope is more rigid than the BV envelope due to the presence of more saturated fatty acids (41, 339). The envelope-associated proteins of ODVs are mostly nonglycosylated, while those of BVs are glycosylated. Unlike the BVs, the ODVs have tegument proteins, which package nucleocapsids and create a matrix layer between the nucleocapsid and the envelope (339). One or multiple ODV virions are further embedded into a protein crystal called occlusion body (OB). Due to the highly compact structure and OB protection the intact ODVs are more easily isolated than BVs.

The baculovirus family is divided to four genera: *Alphabaculovirus* (lepidopteran hosts), *Betabaculovirus* (lepidopteran hosts), *Gammabaculovirus* (hymenopteran hosts) and *Deltabaculovirus* (dipteran hosts) on the basis of the genome phylogeny (134). The alphabaculovirus genus is subdivided into two groups, group I and group II according to the sequence and phylogenetic analysis of conserved genes (74, 156). The subdivision between these two groups is correlated with employment of two different BV envelope glycoproteins, i.e. GP64 and F, for virus-cell fusion and receptor binding (224). These two envelope glycoprotein types are highly distinct in sequence and structural features (161, 408). In group I alphabaculoviruses both GP64 and an F protein remnant are present; only GP64 mediates membrane fusion and the F protein remnant appears to have lost its fusion function (156, 225). In members of group II alphabaculoviruses, the betabaculoviruses and the deltabaculoviruses the major envelope protein is the F protein, which mediates membrane fusion (156, 224).

The F proteins differ from GP64 proteins not only in amino acid sequence, but also in their biochemical and structural properties, except that both proteins are activated at acidic pH (32, 150). F proteins require proteolytic cleavage to prime the membrane fusion (150). Based on the structural elements and organization, F protein and GP64 proteins were characterized as class I and class III viral fusion proteins, respectively (35, 161). Both classes of fusion proteins are composed of trimers with their ectodomains oriented perpendicular to the viral membrane. Two trimers assemble into peplomers or spikes on the surface of BV envelope. The fusogenic forms of class I and class III viral fusion proteins are of mainly α-helical structure or a mixture of α-helices and β-sheets, respectively [reviewed in (415)].

Previously, the baculovirus BV ultrastructure was studied using negative-staining electron microscopy (EM; (4, 19, 93, 123)). The BV envelopes of several group I and II alphabaculoviruses in baculovirus-infected cells and tissue of hosts were reported bulbous at one end. The surface of the envelope is serrated with notches,

which are concentrated at the bulbous end (4). These BV notches, called peplomers, on the surface were demonstrated to be composed of envelope proteins including GP64 and F protein (393).

BV and ODV structures have been studied by negative-staining EM (4, 19, 93, 123). However, due to the fragility and flexibility of BV particles, the procedure of negative-staining and subsequent drying severely impairs the structural integrity of the viral envelope and virions to collapse. In the current study we have used cryo-EM of vitrified BV suspensions to re-evaluate the morphology of virions of *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) and *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV) as representatives of group I (GP64) and group II (F) alphabaculoviruses to provide new insights on ultrastructural organization and assembly of BV virions.

MATERIALS AND METHODS

Virus production and purification of budded baculovirus

Hemolymph-derived SeMNPV BVs and cell culture-derived AcMNPV BVs were obtained from Marcel Westenberg and Els Roode. The BVs were subsequently used for virus production in cell culture. The titers of the SeMNPV and AcMNPV BV suspensions were first determined on Se301 and Sf9 cells, respectively, using end point dilution assays. Se301 cells were then infected with SeMNPV BVs at an MOI of 5 tissue culture infection dose 50% ($TCID_{50}$) units per cell in HyClone CCM3 insect culture medium (Thermo Scientific). Sf9 cells were infected with AcMNPV BVs at an MOI of 5 $TCID_{50}$ units per cell in Insect-XPRESS insect culture medium with L-glutamine (Lonza). At 48 h post-infection (p.i.) cell culture supernatants, where BVs were expected to present, were clarified at 3,000 × g for 15 min at 4°C to remove cells and cellular debris. This supernatant was directly used for cryo-EM.

Fractionation of NCs from budded baculoviruses

The clarified cell culture supernatant from the procedure above was incubated with 1% Nonidet P-40 (NP-40) in TE buffer (10 mM Tris-HCl at pH 7.4, 137 mM NaCl, 2 mM EDTA) for 30 min at room temperature to remove the viral envelope. Six ml of NP-40 treated BV suspension were layered onto 3 ml of 25% and 50% iodixanol solution in TE buffer and centrifuged at 16,000 × g for 3 h at 4°C in SW41Ti rotor (Beckman Coulter). The fractions were collected every 3 ml from the top to bottom. The NCs were most abundant in the third fraction from the top of the centrifuged tube upon examination with negative-staining EM as described below. The NC fraction was diluted in 30 ml phosphate-

buffered saline (PBS) at pH7.4 (Lonza) and centrifuged at 100,000 × g for 1 h at 4°C in an SW32Ti rotor (Beckman Coulter). The pellet was suspended in PBS.

Electron microscopy of budded baculovirus

For negative-staining EM, formvar/carbon coated 400-mesh copper grids were exposed to a glow discharge in air for 20 sec. Ten µl of a BV or NC suspension was placed on the grids and incubated for 2 min. Negative-staining was performed with 1% phosphotungstic acid (PTA, pH 7.2) for 1 min. The specimen were examined in a JEOL 1011 transmission EM equipped with an Olympus Keenview (1K × 1K) and Veleta (2K × 2K) digital camera.

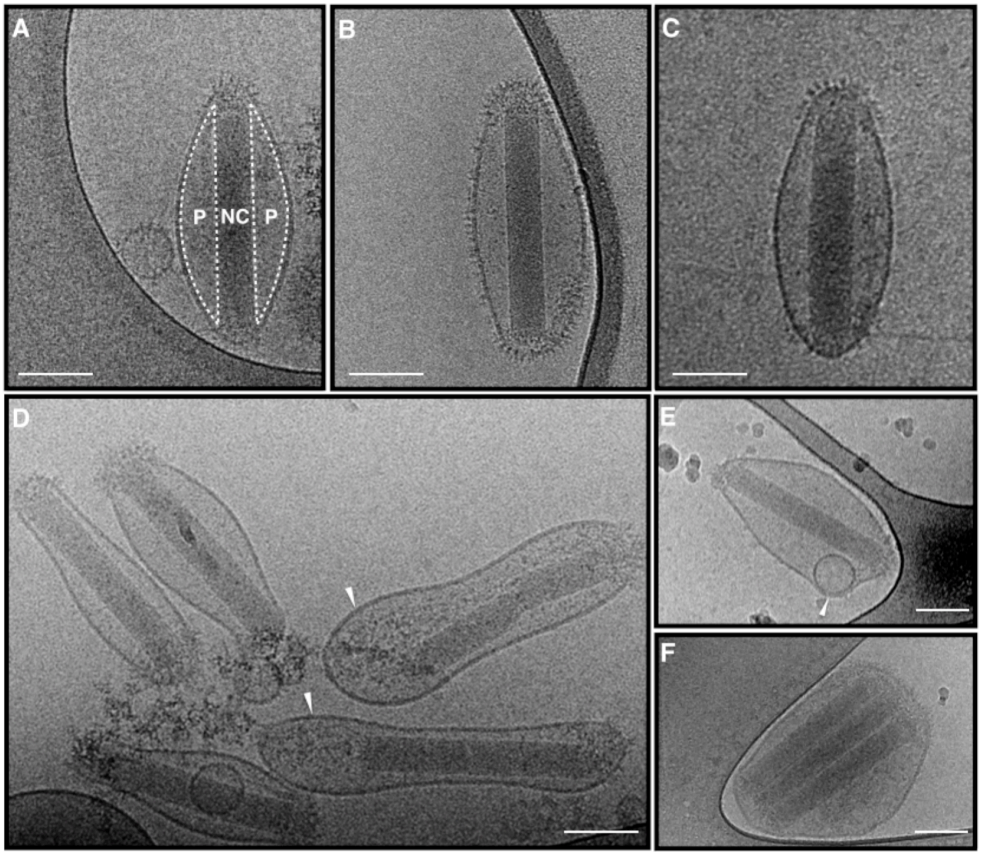
For cryo-EM, specimens were vitrified using a Vitrobot (FEI company). Four µl of a suspension containing virions or NCs was placed on a quantifoil carbon or lacey carbon grid and allowed to absorb for 30 sec. After blotting the grid was plunged into liquid ethane. The frozen specimen was examined at -180°C in a JEOL 2100 TEM equipped with a Gatan CT3500 cryo-holder and a Gatan US4000 (4K × 4K) digital camera. Images were recorded at low dose with DigitalMicrograph software (Gatan) and analyzed with iTEM platform (Olympus).

RESULTS

General architecture of baculovirus BVs

To study the native morphology of AcMNPV and SeMNPV BVs, virions were produced in Sf9 and Se301 cells respectively for 48 h, isolated from the supernatant and cryo-fixed by plunge freezing. Fig. 1 shows images of these BV particles, which exhibit remarkably different features compared to the previous knowledge on BV morphology. The majority (≈95%) of virions of both AcMNPV and SeMNPV with intact envelope show an extended "ovoid" shape (Fig. 1A-C) instead of a 'rod' shape described previously (4). The 'pocket' between NC and lateral envelope appears to be electron-lucent yet not empty because the electron density is darker than the ambient surround the BV particles. Small vesicles are apparently included in the lateral space of some virions (Fig. 1E), indicating that the pockets might be filled with soluble content. The interactions between NC and envelope appeared to be limited to the two ends of NC and therefore the virions tend to have heterogeneous shapes around the lateral area even with cryo-fixation (Fig. 1D). The virions marked with an arrowhead in Fig. 1D appear to get compressed at the inner lateral space, which generated high tension at one end of NC. As a result the envelope-NC interaction at this end was disrupted. In this case the distance between envelope and NC is increased as compared to the other end of virion (indicated as white

arrows in Fig. 1D). This envelope-NC interaction is more notable in the negative-staining EM micrographs of virions. The viral envelopes of negatively stained virions were disrupted to a great extent yet leaving the envelope part retained on the two ends of NCs (Fig. 1G and H). Furthermore, imaging of intact BVs revealed that the envelope proteins, which appear as spikes perpendicular to the viral envelope, are present at both ends of most AcMNPV and SeMNPV BVs (Fig. 1A and C). A small fraction (<1%) of virions is also distributed over the lateral region (Fig. 1B). In addition, a fraction of the BVs (≈10%) possess multiple NCs (Fig. 1F). The peculiar cases of lateral envelope protein distribution and multiple BV NCs per virion may be considered as outcomes of overwhelming virus infection with overexpression of envelope proteins and excessive number of NCs available for envelopment. By and large the overall structure of AcMNPV (group I alphabaculovirus) and SeMNPV (group II alphabaculovirus) as seen via cryo-EM is very similar.



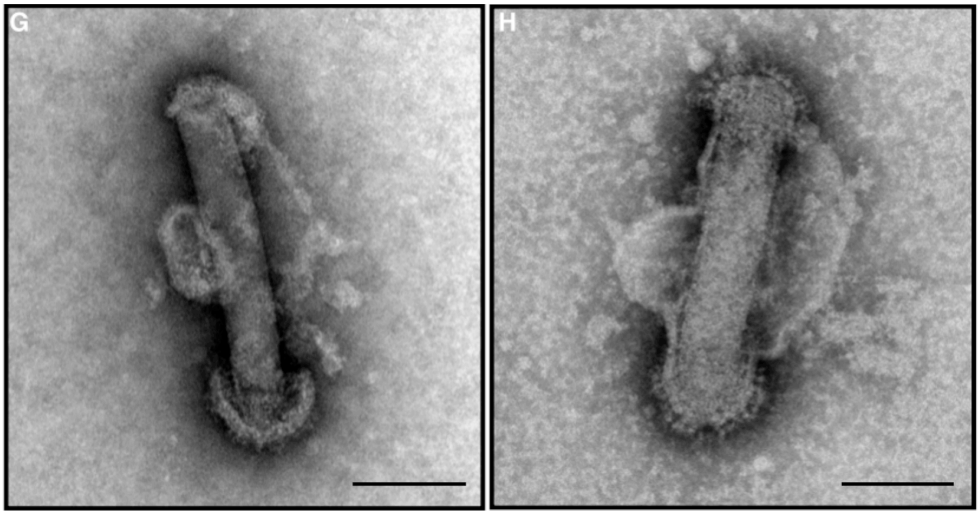


Fig. 1. Virus morphology of a group I and a group II alphabaculovirus BVs. (A-F) Cryo-EM images of AcMNPV and SeMNPV BVs. (A-C) Ovoid shaped of AcMNPV (A and B) and SeMNPV (C). NC, nucleocapsid; P, pocket at lateral side. (D) AcMNPV virions with stretched sites of viral envelopes of AcMNPV around NC ends indicated with arrows. (E) AcMNPV virion with vesicle-like inclusion. The inclusion is indicated with an arrow. (F) AcMNPV virion with multiple nucleocapsids. (G, H) Negative-staining EM images of AcMNPV (G) and SeMNPV (H), showing fragments of the envelope attached to both ends of the NC. Scale bars represent 100 nm.

Budded virus envelope morphology

The BVs of AcMNPV and SeMNPV, representing two groups of the alphabaculovirus genus, bear different envelope proteins. Both GP64 and truncated forms of the F proteins are present in AcMNPV. However, according to a proteomics studies, the GP64 protein is more abundant than the F protein in the AcMNPV BV (402). In contrast, F is the only major envelope protein in SeMNPV (150). Cryo-EM images show that the spikes or peplomers on AcMNPV and SeMNPV BVs appear to be 'Y' and 'burgee" -shaped, respectively (Fig. 2A and B). The two open arms of Y shaped GP64 peplomer appear to occupy more space than the 'banner' of burgee-shaped F spike. This is in good agreement with the distribution of the spikes, which are embedded denser on SeMNPV than on AcMNPV virus particle (Fig. 2A and B). However, the lengths of the majority of the spikes are similar on both baculoviruses (Fig. 2C). The envelopes of BVs are derived from the plasma membrane of host cells (30) and the lipid composition of the viral envelope is therefore similar to the plasma membrane (41, 236). Both AcMNPV and SeMNPV BV possess an envelope with a typical bilayer membrane structure (Fig. 2). These viral envelops with a

thickness of approximately 6 to 7 nm (Fig. 2D) compares to general lipid bilayer thickness of up to 4 nm (263).

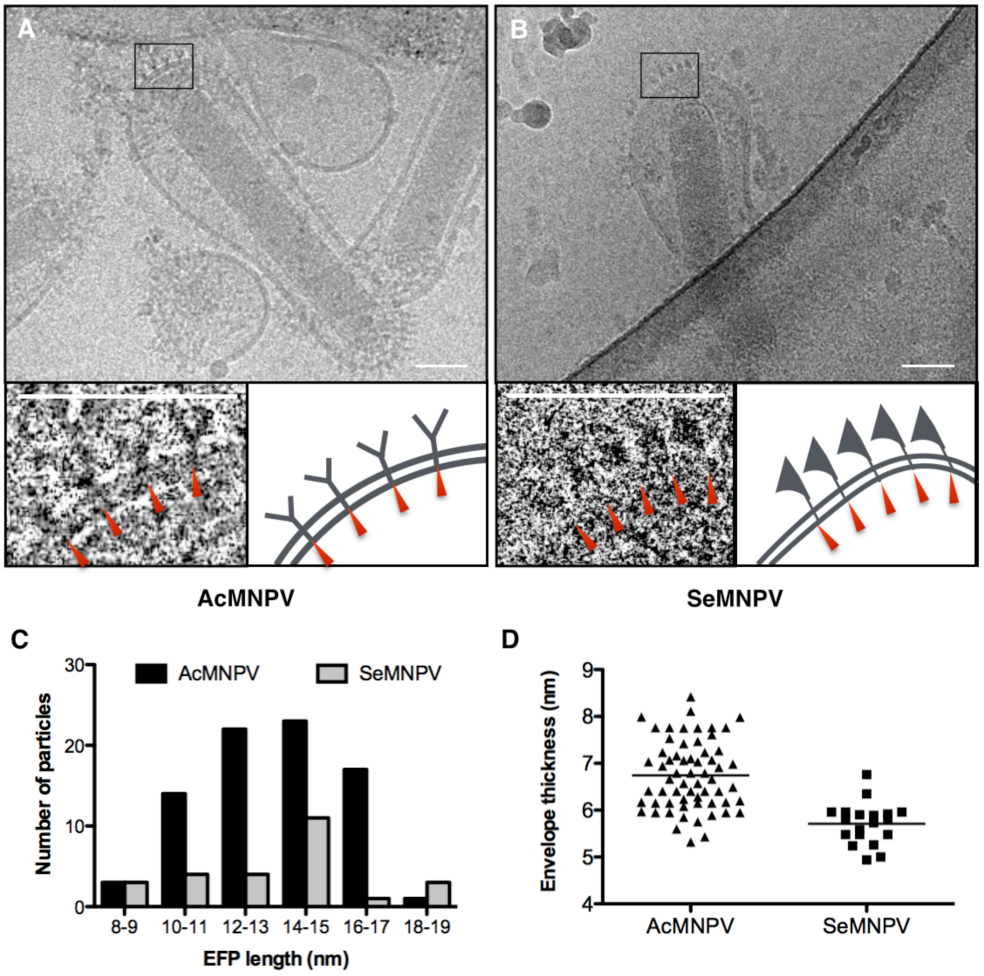


Fig. 2. Ultrastructural characterization of baculovirus envelope proteins. (A and B) Cryo-EM images of envelope proteins on budded AcMNPV (A) and SeMNPV (B). The squared regions on the viral envelope were further focused in the lower panels showing structural patterns. Scale bars represent 50 nm. (C) Comparative analysis of length distributions on envelope fusion proteins from AcMNPV (n = 79; mean = 15 nm, SD = 2.1 nm) and SeMNPV (n = 30; mean = 14 nm, SD = 3.8 nm). (D) Thickness distributions of viral envelopes from AcMNPV (n = 62; mean = 7 nm, SD = 0.7 nm) and SeMNPV (n = 30; mean = 6 nm, SD = 0.4 nm). Envelope proteins were excluded from membrane measurement.

NC structural organization

The ultrastructure of baculovirus NCs has only been studied from alkaline-dissolved ODVs with respect to organization of capsid and NC core (19, 44, 47).

We therefore investigated the structure of NC present in BVs in detail by cryo-EM. After cryo-fixation four distinct forms of NCs were observed from AcMNPV: (i) NCs with a high electron density core and polar ends of apical nipple-like and basal claw-like structures (Fig 3A); (ii) NC partially disassembled from both ends (Fig. 3B); (iii) empty capsids, in which NC cores are absent (Fig. 3C) and (iv) relaxed NCs with a markedly spiral structure (Fig. 3D). The relaxed NCs and empty capsids were also observed within intact envelopes (data not shown). Compared to native NCs the relaxed NCs show a more apparent structural organization. The diameter of the relaxed NCs (Fig. 3D) decreases approximately 10 nm as their length at the same time increases as compared with intact NCs (Fig. 3A), resembling a helical spring in the expanded and compressed states. The empty capsids did not collapse but instead appear to be a stiff shell of tubular shape and identical size as the native NCs (Fig. 3C), suggesting that the assembly of capsid sheath is independent of viral DNA packaging. Once the two ends of the capsid were open the viral DNA was extruded from capsid sheath as a nonsupercoiled form (Fig. 3B).

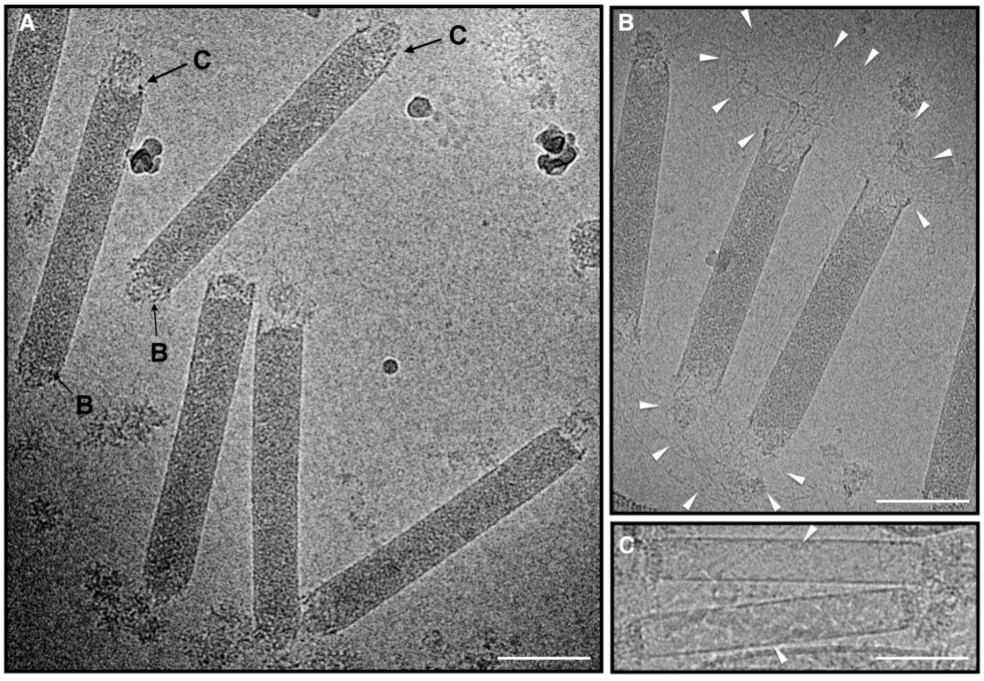
The genome sizes of SeMNPV (131.89 kb) (43) and AcMNPV (133.894 kb) (12) are nearly identical. However, the size of the NC in their BVs showed significant differences (Fig. 3E and F), indicating that the DNA packaging and assembly of these two types of BVs might undergo different processes.

DISCUSSION

In cryo-EM, baculovirus BV particles exhibit an ovoid shape with nearly symmetrical curvature at the apical and basal regions. This is different from previous reports on the BV structure where the envelope was loosely wrapped around the nucleocapsid (393). Interestingly, the envelope did not tightly surround the rod shaped NC at all but puffed up at the lateral region forming two large pockets (Fig. 1A-C), which appear to be electron lucent. The negative-staining EM micrographs of virions showed retained envelope parts on the two ends of NCs and disrupted envelope at lateral region (Fig. 1G and H) suggesting that there is an intimate interaction between envelope and capsids at the polar ends, but not at the lateral regions. Furthermore, the envelope proteins, visible as spikes perpendicular to the viral envelope, are present at both ends of most virions rather than only at the apical end of the virion as has been described previously (4, 126) (Fig. 1A and C). The polarized distribution of envelope proteins at apical and basal regions is a unique feature of baculovirus BVs. It is not known why baculoviruses have their envelope proteins polarized among all the elliptic enveloped viruses. Possibly, the polarized distribution of the envelope

proteins advantages BV entry since orienting the BV perpendicular to the cellular membrane minimizes the membrane fusion area. Moreover, the position of envelope proteins in a curved membrane could be favorable for membrane fusion, since membrane curvature causes membrane tension, which could make the membrane less stable and hence lower the energy barrier for fusion.

Although the envelope of both SeMNPV and AcMNPV exhibits a typical bilayer, the thickness of the envelope is 6-7 nm, which exceeds an average biological membrane thickness of ≈4 nm (263). The unusual thickness of viral envelopes suggests that some other envelope components such as proteins may contribute to the enhanced bilayer thickness. Although the current cryo-EM analysis was not able to resolve more details of the viral envelope, proteomic analysis of alphabaculovirus BVs revealed several other envelope proteins besides F and GP64, among which the most abundant one is ubiquitin (144, 402). Ubiquitin has been suggested to be attached to the inner membrane of the viral envelope by a type of phospholipid anchor (120). Therefore, we propose ubiquitin as a BV structural protein that among others extends an inner layer on the viral envelope. This may explain a thicker appearance of the envelope in cryo-EM. High-resolution cryo-tomography could confirm the presence of this third layer of the baculovirus BV envelope.



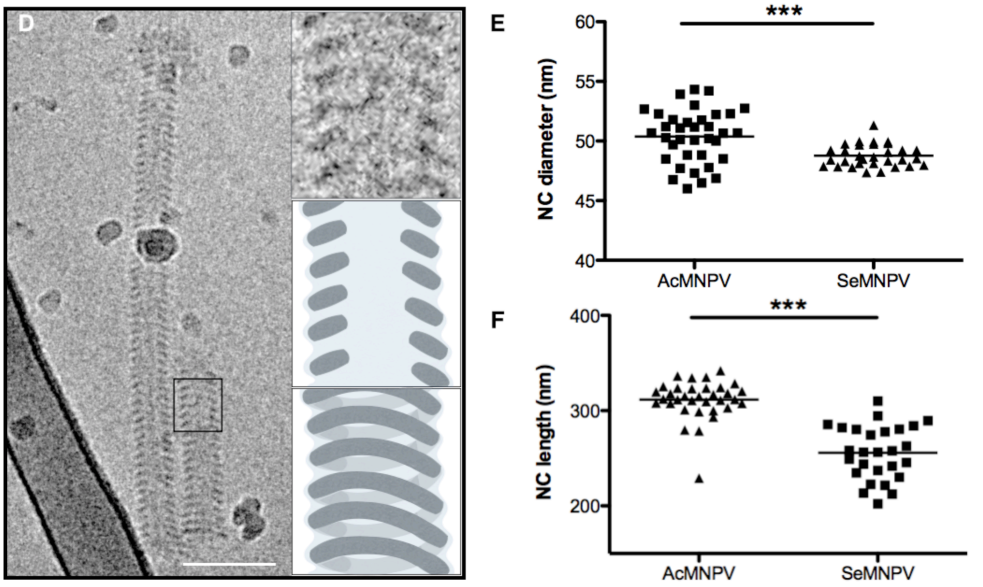


Fig. 3 Characterization of nucleocapsid structure of BVs. (A-D) Cryo-EM images of nucleocapsids of AcMNPV. (A) Native nucleocapsids. (B) Partially lysed nucleocapsids. The extruded fibrillar viral DNA is indicated with arrows. (C) Empty capsids without packaged viral DNA indicated with arrows. (D) "Relaxed" forms of nucleocapsids. The left panel shows the overall image of a relaxed NC. The panels at the right side show the detail of the spiral shapes of NC. (E) Distribution of the diameter of AcMNPV NC (n=36, mean=50.37 nm, SD=2.24 nm and SeMNPV NC (n=31, mean=48.78 nm, SD=0.93 nm). (F) Distribution of the length of AcMNPV NC (n=34, mean=311.5 nm, SD=20.39 nm) and SeMNPV NC (n=27, mean=255.7 nm, SD=28.31 nm). Significant differences between AcMNPV and SeMNPV are indicated (***, $P<0.0004$ for NC diameter; $P<0.0004$ for NC length). Scale bars represent 100 nm.

The 'Y'- and 'bungee'-shaped appearance of the peplomers for AcMNPV and SeMNPV, respectively, in cryo-EM (Fig. 2) differs from what is expected on the basis of the symmetric, trimeric structure of envelope fusion proteins in general. It could be that what we see here is a fusion protein intermediate in its extended conformation. This conformation may also be different between F (SeMNPV) and GP64 (AcMNPV). Since information on the appearance of envelope fusion proteins is derived from vertebrate viruses, it is possible that such structures are different in invertebrate envelope fusion proteins. Further analysis using known vertebrate F proteins, such as those from parainfluenza virus (class I) or thogoto virus (class III), as well as other baculovirus F proteins using cryo-EM should provide further support for our observations. In our cryo-EM micrographs of NCs the filled and empty capsids exhibit similar diameters and lengths (Fig. 3A and C). This suggests that the capsid assembly is independent on the DNA

packaging. Moreover, the two ends of NCs appear to be more fragile compared to the rest of the NC. These two ends are most likely composed with various structural proteins other than the major capsid protein VP39 [reviewed in (339)]. Once the ends are open, the viral DNA leaks out as an unstructured form (Fig. 3B). These results support the NC assembly model proposed by Fraser (1986). In this model, the viral genome inserts into a pre-assembled capsid sheath through the cap structure (93), which is consistent with the DNA packaging process for other large double-stranded (ds) DNA viruses such as bacteriophage lambda (100), T4 (99) and herpes virus (253). However, in case of the most extensively studied large dsDNA viruses including bacteriophage T4 and lambda, the terminase, an ATP-driven packaging motor protein is responsible for DNA insertion into their pre-assembled NCs (99, 288). Although packaging the large circular genome into a capsid shell is an energy dependent process, up to now no protein associated with the NC or DNA has been demonstrated to be responsible for viral genome packaging into preformed BV capsids. The essential structural proteins at both ends of NC, which are related to NC morphogenesis and maturation, are not homologous to the bacteriophage terminase protein sequence. However, deletion of VP80 (369) or 38K (421) results in half filled or empty NCs, respectively without affecting DNA replication, indicating these two proteins may be involved in DNA encapsidation. Furthermore, yeast two-hybrid assays demonstrated that the 38K protein interacts with VP80 (422). The latter protein interacts with both F-actin and DNA (234). In addition, actin has been indicated to be involved in the DNA packaging since actin is transported from cytoplasm to the nucleus (394) and interact with NC components (190) upon viral infection whereas, Inhibition of actin polymerization results in the empty NCs (390). Taken together, VP80, 38K and actin cytoskeleton may function collectively on the DNA encapsidation rather than one specific motor protein directly interacts with viral DNA.

Based on our cryo-EM visualization we propose a new model for baculovirus BV structure (Fig. 4). The model emphasizes the strong interactions of envelopes at polar ends with the NCs and the space at the lateral region between NC and envelope of the BVs. The major envelope protein GP64 mediating fusion between viral envelope and cellular membrane, are predominantly distributed on the envelope around the both ends of the NC suggesting a vertical orientation to the cellular membrane during virus entry. The capsid sheath has a compact helical structure.

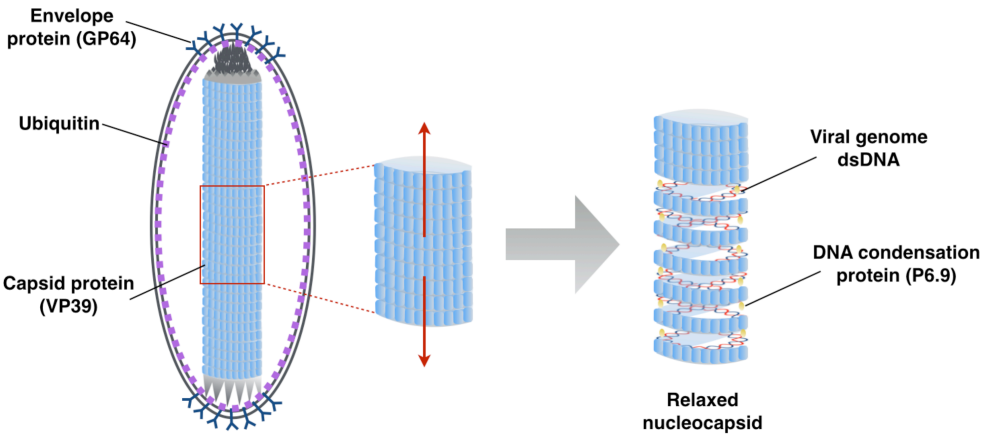


Fig. 4 Structural model of the baculovirus BV represented by AcMNPV. Major envelope proteins GP64 are distributed at the polar ends of the ovoid shaped envelope. The DNA genome is condensed by P6.9 protein and organized as super coil in the nucleocapsid. The capsid sheath containing major protein VP39 forms a spiral structure visualized by its relaxed state.

The fragile nature of BVs compared to rigidity of ODVs (339) denotes the baculovirus infection cycle. The BVs targets on cell-to-cell infection and therefore, are protected by the body fluid of host. The ODVs target *per os* infection of the host midgut epithelium; BVs do not have to survive the hostile conditions outside the host and hence this may permit a more fragile virion structure. Our understanding of baculovirus BV structure has long been hampered by the fragile nature of the virus and the limitations of conventional imaging techniques. Previous BV morphology has only been studied with negative-staining, which causes dramatic change of inner osmotic pressure of virions due to the air-drying procedure, which subsequently deform the viral envelope. In addition, high-speed centrifugal pelletation and prolonged virus production time in cell culture introduce disruption of viral envelope and heterogeneity of viral morphology (362, 436). All these artifacts have led to biased morphological view of BVs. Our cryo-EM analysis provides novel insight into the near-native structure and virus assembly of baculovirus BVs.

Chapter 7

Discussion and perspectives

Qiushi Wang

Virus-cell fusion is the means for enveloped viruses to deliver their genome into the host cell and initiate replication. This fusion process is mediated by one or several viral surface glycoproteins, in which the fusion protein plays a central role. Typically, upon activation the fusion protein undergoes a series of conformational changes that facilitate membrane fusion. Baculoviruses are a group of enveloped viruses, which also utilize fusion proteins to initiate their infection cycle. Studying the structures of baculovirus fusion proteins that are employed during the fusion process will enable us to understand the mechanism of membrane fusion between baculoviruses and their host cells. Baculovirus genome sequencing and functional analyses revealed two distinct fusion protein genes, i.e., *gp64* and *f* that encode the GP64 and F fusion proteins, respectively, and are responsible for fusion of the budded virion (BV) of baculoviruses. The GP64 proteins are the fusion proteins of BVs belonging to group I viruses within the genus *Alphabaculovirus*. The F proteins are present in all the baculovirus genera except *Gammabaculovirus*, and function as fusion proteins in BVs of group II alphabaculoviruses, betabaculoviruses and deltabaculoviruses.

In this dissertation a wide variety of assays was applied, including biochemical and functional assays, X-ray crystallography and cryo-electron microscopy (EM), to investigate the structure and function of the F fusion protein of the BVs of group II alphabaculoviruses The postfusion structure of the baculovirus F protein appeared to be highly similar to the class I fusion proteins of paramyxoviruses and distinct from the GP64 protein, which has been identified as a class III fusion protein. Since both pre- and postfusion structures of paramyxovirus fusion proteins are available we can infer the F-mediated fusion process of baculoviruses. Moreover the baculovirus F protein structure demonstrated the evolutionary ties between the F proteins of double stranded DNA baculoviruses, negative-stranded RNA paramyxoviruses and F-like proteins found in baculovirus hosts. In this chapter the implications of the current results will be discussed in a broader perspective.

VIRAL-CELL FUSION MEDIATED BY BACULOVIRUS F PROTEIN

Biogenesis of F ectodomain

The baculovirus F protein is considered as a class I viral fusion protein based on its structural domain organization (212). The membrane fusion reaction of the F protein is triggered by low pH (150, 411). Like most class I fusion proteins, the F protein requires priming through proteolytic cleavage. This cleavage is mediated

by the cellular endoprotease furin after synthesis during transport through the secretory pathway, separating the protein into membrane anchored subunit F1 and membrane distal subunit F2 that remain connected via an inter-subunit disulphide bond.

To understand the fusion process of the baculovirus F protein we aimed to gain insight into the structure of a baculovirus F protein. This was achieved by X-ray crystallography on a soluble version of the F protein of SeMNPV as a prototype for baculovirus F proteins. The soluble F protein (sF) - also called F ectodomain - was obtained by deletion of the transmembrane (TM) domain of the F protein. In Chapter 2 the sF protein was expressed in human embryotic kidney (HEK) 293T cells and biochemically characterized. The cleavage on sF occurs at the conventional furin cleavage site and is rather inefficient, resulting in a mixture of cleaved and uncleaved F ectodomains. To obtain a homogenous population of F ectodomains for X-ray crystallography the furin cleavage site of the F ectodomain was mutated in a way that would prevent cleavage by furin but still allow cleavage by trypsin. We confirmed that the mutated F ectodomain (sF^{FCSmut}: RRSKR→ SSSKK) was secreted as a homogenous, uncleaved, monomeric soluble protein. We next tested whether sF^{FCSmut} could be cleaved at the mutated furin cleavage site by trypsin *in vitro* and its response to acidic pH.

Our results show that sF^{FCSmut} could be cleaved at the mutated furin cleavage site by trypsin. Trypsin treatment of the sF also removed the fusion peptide and the HR1 and HR2 regions. Intriguingly, trypsin cleavage allowed the efficient conversion of sF^{FCSmut} from the monomeric to a trimeric form after exposure to acidic pH. The trypsin cleaved sF^{FCSmut} was designated sF_2-F_1'. The required pH for oligomeric transition of the baculovirus F ectodomain is similar to the pH for activating its full-length F protein. In Chapter 3 the crystal structure of the trimeric sF^{FCSmut} shows a high degree of similarity with known postfusion structures of class I fusion proteins; hence, representing the first postfusion structure of full-length baculovirus F protein.

The pre- and postfusion forms of class I viral fusion proteins are considered to be organized as trimers (415). However, the baculovirus F ectodomain was found to form monomers at neutral pH (prefusion), which is different from trimeric organization of the native, full-length baculovirus F protein (216). This suggests that at the prefusion state the interprotomer interactions of the ectodomains are too weak to form oligomers and indicates that the TM domain plays an important role in protein trimerization. Both paramyxovirus F and influenza

hemagglutinin (HA) proteins lacking the TM domain have also been shown to trimerize inefficiently (55, 333, 340, 341, 352). The TM interactions of the paramyxovirus F protein were suggested to be important to stabilize the trimeric prefusion form (340, 341). The monomeric cleaved baculovirus F ectodomain converted from a monomeric to a trimeric form upon acid treatment. This result strongly indicates that the trimeric interfaces of the prefusion and postfusion forms are different. The interprotomer interactions of ectodomains in the postfusion form are likely stronger than in the prefusion form. The exposure of interfaces in each protomer that allows stable trimer formation is triggered by low pH. In Chapter 3 the crystal structure of acidified, cleaved sF^{FCSmut} protein lacking HR1 and HR2 regions shows that the trimeric interactions are formed by HR3 regions. The negative stain EM reconstruction of the full-length F ectodomains confirms the presence of 6-helix bundles formed by HR1 and HR2 regions. The 6-helix bundle structure in paramyxovirus F and influenza HA proteins are known to stabilize their postfusion structures (45, 66). Intriguingly, the trypsin cleaved sF^{FCSmut} protein is lacking the HR1 and HR2 regions indicating that 6-helix bundle formation is not required for trimerization. Thus, the HR3 regions in the ectodomain appear to be important for trimerization and stabilization of the postfusion form of the baculovirus F protein, and formation of the HR1-HR2 6-helix bundle would only further stabilize such a trimeric structure. The absence of stable interprotomer interactions of the prefusion ectodomains may allow the necessary 'breathing' of the F prefusion trimer to undergo the conformational changes towards its postfusion form. The energy gained in the transition to the stable postfusion trimer form may be applied for the fusion of viral and cellular membranes.

Proteolytic cleavage of class I viral fusion proteins is thought to be essential to liberate the fusion peptide and enable the conformational changes of the fusion protein upon fusion activation, which leads to membrane fusion. The two best characterized class I viral fusion proteins, paramyxovirus F and influenza hemagglutinin (HA), were found to require cleavage for conformational changes at different levels. Most paramyxovirus F proteins are activated by host receptor engagement through a specialized viral attachment protein (HN/H/G, depending on the virus) (51, 155, 189). The receptor-induced conformational changes of the paramyxovirus F proteins only occur after the cleavage of the F protein (84). Distinct from the paramyxovirus F proteins, influenza HA proteins are activated by low pH (45, 85, 117) and the acid-induced conformational changes take place even without cleavage (38) although the uncleaved form is

fusion incompetent (413). The uncleaved HA ectodomain was also observed to undergo conformational changes at low pH, but in a less efficient and concerted manner compared to the membrane-anchored protein (333). The acidified, uncleaved HA ectodomain was unable to bind to liposomes indicating that the hydrophobic fusion peptide is not exposed at low pH. Hence, the cleaved and uncleaved HA proteins respond to acid differently and this may explain why uncleaved HA is not fusion active (38). Like influenza HA, the baculovirus F protein is activated by low pH (411). Similar to the uncleaved influenza HA ectodomain, the uncleaved F ectodomain was shown to undergo acid-induced conformational changes represented by the monomer-to-trimer transition, which is less efficient for uncleaved F (~35% of the protein population) than the cleaved F when exposed to low pH (Chapter 2). This indicates that uncleaved F ectodomain is acid-sensitive, yet the uncleaved status of the F ectodomain restrains acid-triggered conformational changes, at least to a certain extent.

The baculovirus F shares crucial features with influenza HA on top of the common characteristics of class I fusion proteins. First, these two proteins are acid-dependent. Second, both proteins contain N-terminal receptor binding and C-terminal fusion regions (338, 401). Third, the two subunits of the proteins are connected by a disulfide bond upon cleavage (336, 411). It is not clear why these proteins, in uncleaved form, are acid-sensitive. What we know is that both proteins are synthesized as trimers and only a fraction of the synthesized proteins can be cleaved by cellular furin (331). The fusion trimers assembled on the virus particles may contain both cleaved and uncleaved protomers. After the viruses are endocytosed, the conformational changes of the uncleaved protomers upon acidic pH in endosomes may expel the receptor-binding region to promote the transit of the cleaved protomers to a fusion-active conformation and to form a new trimeric interface.

Upon viral infection the viral fusion proteins are synthesized and transported through the acidic exocytic route in the host cell. Viruses carrying pH-sensitive fusion proteins utilize various strategies to avoid premature activation of the fusion protein during its transport through the acidic Golgi apparatus. The rhabdovirus G and group I alphabaculovirus GP64 proteins are characterized as class III fusion proteins based on their structural features; they can undergo reversible pH-dependent conformational changes (308, 312, 438). It has been suggested that G protein of rabies virus has at least three different conformations: the native (N) state in the newly assembled virus, a fusion-activated (A) state at the beginning of fusion process and a fusion-inactive (I) state with prolonged

exposure to low pH. The transition between the conformations takes place by shifting the pH-dependent equilibrium between these states (312). The newly synthesized G protein was observed to form the I state during transport in the Golgi apparatus and recover to the N state when reincubated above pH 7 (107, 312), which occurs on the viral surface. This strategy meant to avoid unspecific fusion during protein transport and maintain a native structure before the onset of the fusion process may also apply to GP64. Unlike class III fusion proteins, class I and II fusion proteins are thermodynamically more stable at the postfusion state and the transition from pre- to postfusion structure is irreversible (78, 172, 335, 353). Membrane fusion of alphaviruses and flaviviruses is mediated by the pH-dependent class II viral fusion proteins E1 and E, respectively [reviewed in (386)]. Both proteins form a complex with a second viral protein to prevent the conformational changes induced by low pH in trans-Golgi network during exocytosis (131, 348, 395). The pH threshold for fusion activation varies between different influenza virus HA proteins (229). The Avian influenza virus H7N1 HA protein has a high pH threshold (pH5.9) to trigger the fusion (367), which is close to the pH in the trans-Golgi network (TGN) (~pH6) (209, 292). It was found that the M2 protein ion channel activity plays a crucial role in stabilizing the native form of HA during exocytosis by transporting protons out of the lumen of the acidic TGN (367). At this stage the mechanism of stabilizing the native form of baculovirus F proteins in the acidic endocytic route is not known. In Chapter 2 the cleaved F ectodomain was shown to be acid-sensitive and to undergo irreversible, acid-triggered conformational changes. Most of the secreted F ectodomains with an intact furin cleavage site (sF) could be cleaved by Golgi-resident furin, indicating its transport through acidic TGN in the secretion pathway. While passing through the TGN the cleaved sF proteins could potentially undergo premature, irreversible conformational changes triggered by the acidic environment; however, they were secreted as monomers and converted only into postfusion trimers after exposure to pH5 (Chapter 4). The defined pH range for efficient fusion activation of baculovirus F protein (pH5-5.5) (Chapter 5) is lower than the pH in the acidic TGN (~pH6) (209, 292). Hence, this low pH threshold for fusion may prevent the premature activation of the baculovirus F protein during transport through the slightly less acidic TGN.

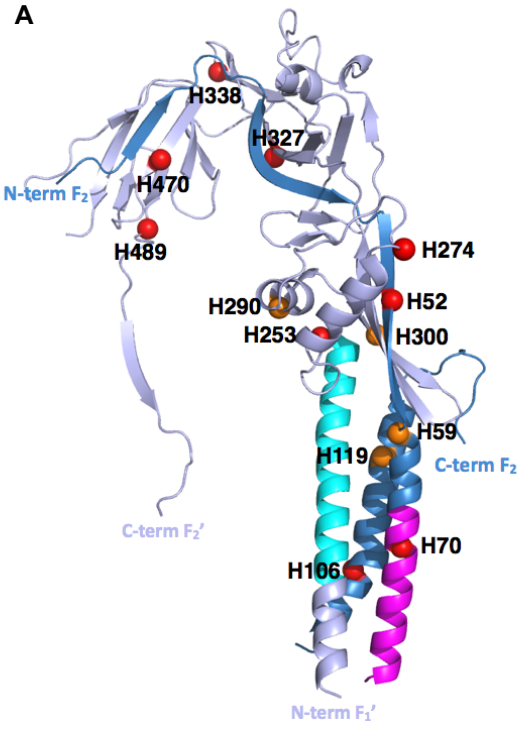
THE ROLE OF HISTIDINES IN THE F ECTODOMAIN

In Chapter 5 the fusogenic pH of SeMNPV F protein was determined to be 5-5.5, which is lower than the pK_a value of histidine. This F protein contains 23 histidine residues, six of which are located in the ectodomain and are highly

conserved in F proteins of the members of the baculovirus family. The histidine residue has been reported to act as a pH sensor in viral fusion proteins. Protonation of the histidine can trigger structural rearrangements of metastable fusion proteins during acid-induced membrane fusion (48, 98, 230, 270, 329). The current model of the histidine pH sensor suggests that the histidine residues become protonated at low pH and hence positively charged. Protonation of histidines may break existing interactions or create new ones, thereby lowering the kinetic barrier that prevents reorganization of the F protein into its most stable conformation (165). We investigated the role of the six conserved histidine residues of SeMNPV F protein by alanine substitution. The results showed that the H59A or H119A substitutions abolished syncytium formation of the cells expressing the mutated F proteins and altered the postfusion structure of F ectodomain. However, none of the six histidine-to-alanine substitutions abrogated entry of VSV viruses pseudotyped with mutated F proteins or inhibited the low-pH triggered conformational change of the F ectodomain. Since the trypsin-cleaved F ectodomain, which lacks the fusion peptide, HR1 and HR2 regions, underwent acid-induced irreversible conformational changes, the pH sensor(s) are likely located in the sF_2-F_1' fragment (Fig. 1B). The locations of histidine residues are shown in the postfusion structure of the F ectodomain in Fig. 1A. The conserved histidine residues H59, H119, H290 and H300 are present in the postfusion crystal structure of the acid-exposed sF_2-F_1' fragment. H119 is located in the stalk region of the F structure and forms hydrogen bonds with H59, L115 and V235. H59 forms a salt bridge with D123. H59 and H119 in the F_2 subunit may stabilize the postfusion structure of the F protein and pore expansion of plasma membrane to allow syncytium formation. Although H290 and H300 in the F_1 subunit individually form salt bridges with a neighboring negatively charged residue (H290-E286 and H300-E306), single substitution of each of two residues with alanine appeared to have no effect on cell-cell fusion.

While our results suggest that none of the individual, conserved histidines is critical in pH sensing, it is possible that several histidines act in concert. Multiple histidine residues that collectively orchestrate pH sensing have been found in the fusion proteins of VSV and AcMNPV (48, 205). Is this the case for SeMNPV F protein? Our structural studies indicate that acid-induced conformational changes of each protomer are coupled with trimerization of the SeMNPV F ectodomain. This may be caused by two possible triggering mechanisms in case histidine sensors are present in sF_2-F_1'. One is that multiple histidines are protonated and form salt bridges between protomers, which allows the formation

of the stable, trimeric post-fusion structure. However, inter-protein interaction analysis with Protein Interaction Calculator (375) revealed that none of the histidine residues in the sF_2-F_1' post-fusion structure is involved in inter-protomer salt bridges. The other possibility is that within a monomer of sF_2-F_1' the histidine sensors are protonated to destabilize the prefusion structure and trigger the conformational changes, exposing otherwise buried residues, thereby allowing interactions between monomers. Besides the conserved histidine residues H59, H290 and H300, the less conserved histidine H52 forms a salt bridge in the stalk region of the crystal structure. All these residues could possibly participate in pH sensing. At this stage it is not clear whether the salt bridges formed by the four histidines are formed upon acid-activation to stabilize the postfusion structure, since the prefusion structure of F ectodomain is unknown. Resolving this complexity requires further study.



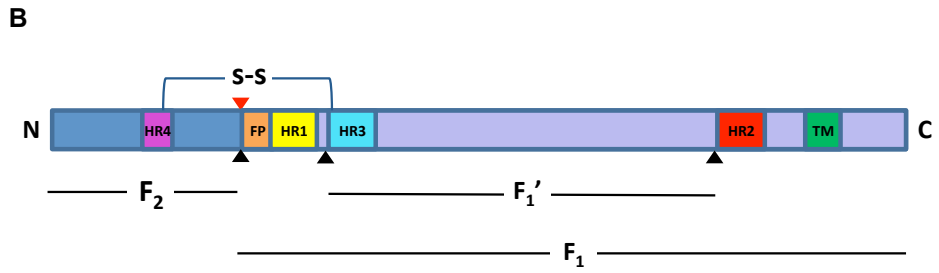


Fig. 1. Locations of histidine residues in SeMNPV F protein ectodomain. (A) Locations of histidine residues in the protomer, post-fusion structure of SeMNPV F ectodomain (sF_2-F_1'). The conserved histidine residues H59, H119, H290 and H300 are shown in orange and other histidine residues are shown in red. (B) Linear representation of the SeMNPV baculovirus F protein drawn to scale. Coloring of domains is similar as in (A). Note that the fusion peptide (FP), the HR1 region and the C-terminal region starting from HR2 are absent in the structure. The post structure of F ectodomain was prepared with MacPyMOL molecular graphics system (version 1.3, DeLano Scientific LLC, 2009).

F-LIKE PROTEINS IN VIRUSES AND THEIR HOSTS

Evolutionary implications from the structural similarity between class I viral fusion proteins

The structural features of SeMNPV F protein presented in Chapter 3 unambiguously show the presence of a class I viral fusion protein in a DNA virus family. The striking postfusion structural similarity and conserved positions of cysteine residues between RSV F protein and SeMNPV F protein suggest an evolutionary link between the fusion proteins of the viruses in the families *Paramyxoviridae* and *Baculoviridae*. The viruses in these two virus families clearly differ in genome type (negative-sense single-stranded RNA vs. double-stranded DNA), genome size and genomic organization. Therefore, it is tempting to speculate that, instead of diverging from a common ancestor virus, these class I fusion proteins may have been independently acquired from a common or related ancestral class I membrane fusion protein. This speculation of independent acquisition of fusion protein encoding genes is supported by the discovery that two distinct fusion proteins are found in viruses belonging to the same genus (*Alphabaculovirus* genus, *Baculoviridae* family). The group I alphabaculoviruses possess the class III fusion protein GP64, while group II alphabaculoviruses use the class I fusion protein F despite that these two groups are closely related in the phylogenetic tree of the *Baculoviridae* family (134). Although these two groups of alphabaculoviruses may have evolved from a common ancestor virus, their fusion proteins were clearly acquired from different sources, which could be either

different hosts or different viruses during co-infections of a host. The GP64 protein is closely related to the envelope protein GP encoded by Thogoto and Dhori viruses in the genus *Thogotovirus* (RNA viruses, mostly tick-borne) from the family *Orthomyxoviridae* based on its structural features and amino acid sequence (105, 295). This suggests that the gene encoding the envelope proteins had either been transferred between these viruses or acquired from a common or related source (278). The GP64 proteins of group I alphabaculoviruses share ~30% amino acid sequence identity with GP proteins of Thogoto and Dhori viruses. Such a high sequence identity has not been found for the F envelope protein of group II alphabaculoviruses with envelope proteins of other virus families. This indicates that the transfer of the gene encoding a class III fusion protein between members of the families *Baculoviridae* and *Orthomyxoviridae* is a relatively recent event (295). The independent acquisitions of GP64 and F proteins to mediate baculovirus-cell fusion is further supported by the observation, that alphabaculoviruses still encode a truncated, non-fusogenic version of an F protein in addition to a functional GP64 protein (225, 401).

Although the SeMNPV F and RSV F proteins exhibit a high degree of structural similarity, their amino acid sequence identity (16.5%) is relatively low (Fig. S2 in Chapter 2). Since the group II alphabaculoviruses and paramyxoviruses have no common hosts, direct gene transfer via recombination would be difficult to conceive. One of the possibilities could be that these two groups of viruses might have acquired their fusion genes during genesis independently from different eukaryotic host cells. Indeed, the intracellular vesicle fusion proteins in the SNARE family, which are broadly present in eukaryotes (2), share essential structural features with class I viral fusion proteins including α-helical coiled coils for zippering two opposing membranes (337) and α-helical fusion peptide for anchoring on the target membrane (80, 351, 368). The structural analogy of class I viral fusion proteins with SNARE proteins may provide evidence on the possible evolutionary path between insect viruses and mammalian viruses. Due to limited number of sequenced virus genomes we can not rule out the possibility that baculovirus and paramyxovirus both acquired their fusion protein-coding gene from a third virus during co-infection of a common host.

Evolutionary links between viruses and hosts

Our phylogenetic data together with sequence and structural analyses of an F-like gene of *Anopheles darlingi* shown in Chapter 4, suggested an endogenization event of a viral gene into an insect genome. The emerging studies on ancient viruses within the field of paleovirology have uncovered numerous viral elements

that are endogenized into a broad array of eukaryotic hosts from mammals to plants (62, 91, 143). Some endogenous viral elements are actively involved in physiological functions in their hosts. These include eukaryotic genes encoding fusogenic proteins with a viral origin. The best-characterized endogenous viral fusogenic gene *syncytin*, which plays an important role in cell-cell fusion during placental development, is the envelope gene of a human endogenous retrovirus, HERV-W. While the rest of the genes in the endogenous retrovirus genome show signs of defects, the envelope gene is preserved for its fusogenic function through positive selection during host evolution (175, 254). In addition, endogenized viral envelope proteins may be involved in triggering immune responses (139). The envelope proteins of endogenous retroviruses have been hypothesized to induce more avid host immune response to highly similar exogenous retroviruses (431). This indicates that viral genes may switch sides and get recruited by the host. In our study the mosquito cellular protein Ad-F was found to share several characteristic motifs and domains with SeMNPV-F including a preserved furin cleavage site (FCS), a transmembrane domain and several heptad repeat regions (Chapter 4). In addition, the structures of Ad-F and SeMNPV-F were predicted to have significant structural similarity based on structural modeling by the PHYRE2 protein fold recognition server. The evolutionary link between the Ad-*F* and baculovirus *F* genes was further strengthened by the lack of introns in the Ad-*F* gene since viral genes in general lack introns, suggesting that the Ad-*F* gene is an endogenized viral element. The hydrophobicity of the domain downstream of the furin cleavage site in Ad-F is weaker than observed for the fusion peptide of SeMNPV-F. Unlike the SeMNPV-F protein, our preliminary data indicate that Ad-F mediates cell-cell fusion in a pH-independent manner. It remains to be determined whether the Ad-F protein functions as a fusion protein in the mosquito. To study the function of Ad-F in *A. darlingi*, knowledge on its expression in *A. darlingi* and its localization at a cellular and organ level is required in the first place. Next, assessing the role of the Ad-*F* gene in *A. darlingi* by gene deletion or gene knock-down may help to understand the physiological function of Ad-F.

The possible involvement of transposable element (TE) in horizontal gene transfer

Searches for remote homologs of the SeMNPV-F and Ad-F proteins using the PSI-BLAST algorithm revealed that both proteins are homologous to a cellular *Drosophila* protein called Iris as well as to several envelope proteins of insect endogenous retroviruses (IERV), including errantiviruses (*17.6*, *297*, *Gypsy*,

Idefix, *Cruiser*, *ZAM*, *Tirant* and *TED*) and a semotivirus (*Roo*) (33, 350). Compelling evidence has accumulated that support the relatedness between baculoviruses and IERVs. Errantiviruses and semotiviruses belong to the *Metaviridae* family and exist as LTR retrotransposons, which are commonly found in eukaryotes (33). Previous phylogenetic analysis suggested that these IERV envelope genes were derived from an ancestral baculovirus envelope fusion gene *F* (232, 319). All the IERV envelope proteins homologous to baculovirus F proteins have a furin-like cleavage site (Chapter 4). One of the IERVs, the *Gypsy* retrovirus, has been reported to be infectious (174) and its envelope protein has fusogenic activity (261). For another errantivirus 17.6, it was demonstrated that by substituting 4 amino acids the predicted 21-amino-acid fusion peptide downstream of the conserved furin cleavage sequence in this errantivirus was able to render a fusion peptide-negative F mutant of a group II alphabaculovirus fusogenic (294).

Transposons have long been recognized for their influence in genome evolution through their potential for insertion, excision and recombination in their host genomes [reviewed in (146, 194)]. The errantivirus retrotransposon *TED* in *Trichoplusia ni*, which encodes Gag, Pol and Env proteins (96, 197, 290), was found in the genome of an AcMNPV mutant, FP-D (257). The baculovirus RNA polymerase, responsible for the transcription of late and very late baculovirus genes, recognizes the promoter sequence (A/G/TTAAG), which is found not only in baculovirus genome (30, 318) but also in the LTR sequence of *TED* (317). Upon baculovirus infection, the mRNA transcribed in the FP-D mutant, showed markedly higher level than in wild-type AcMNPV (97). This indicates that retrotransposons like *TED* can be transposed from the insect to the baculovirus genome, and presumably the other way around, and may by doing so carrying (new) genes with them. On the basis of the transposition capability of IERVs and the sequence similarity between baculovirus *F* and IERV envelope genes we could argue that the primordial IERVs might have obtained their envelope genes from a baculovirus and became infectious via DNA recombination. Sequence comparison reveals that F proteins of group II alphabaculoviruses and betabaculoviruses, the envelope protein of the *Roo* retrovirus and the Ad-F protein have a similar domain organization with respect to the presence and positioning of the furin-like cleavage site, the large conserved domain DUF3609 and transmembrane domain. This suggests that gene transfer from a baculovirus to the mosquito may have been mediated by an IERV. The relationship is further supported by our phylogenetic analysis showing that Ad-*F*

is most closely related to the *Drosophila* cellular gene *Iris*, which appeared to be derived from a sister lineage of the *Roo* retrovirus (named *Kanga*) (231). It was suggested that similar to errantiviruses the *Roo* retrovirus also had acquired its envelope gene from a baculovirus (232). Although the significant sequence and structural similarities of SeMNPV-F and Ad-F reviewed earlier suggest the possible gene transfer from a baculovirus to a mosquito host, the uncertainty of the evolutionary rate and the position of the root in the *F* phylogeny do not allow us to determine whether the gene transfer route from baculovirus to mosquito is direct or whether any of the other *F* homologs is involved as an intermediate for gene transfer.

FINAL REMARKS AND FUTURE PERSPECTIVES

The research described in this dissertation focused on the structure and functions of the budded baculovirus F fusion protein. Such knowledge is essential to understand the fusion mechanisms between baculoviruses and their hosts and provides insight into the evolutionary relationship between baculovirus F protein, eukaryotic proteins and other animal virus fusion proteins.

In this dissertation we describe the SeMNPV F postfusion structure, which represents the first crystal structure of a class I fusion protein from a DNA virus. This protein structure shows great similarity with the postfusion structure of the paramyxovirus F proteins. The pre- and postfusion structures of paramyxovirus F fusion proteins are well characterized and have led to a fusion model for paramyxoviruses. Given the high structural similarity in fusion proteins, the paramyxovirus fusion model can be extrapolated – at least in part - to the F-mediated baculovirus membrane fusion process. The baculovirus and paramyxovirus F proteins do however differ with respect to their fusion activation mechanism, with the paramyxovirus F protein being activated in a pH-independent manner, via a second companioning viral glycoprotein (188). The elucidation of the prefusion structure of the baculovirus F fusion protein will be important to identify the pH sensors that destabilize the prefusion structure and trigger conformational changes after protonation.

In respect to evolution, the similarity in overall structure of the baculovirus F and paramyxovirus F proteins strongly suggests that they may have a common ancestor. This is further supported by the sequence similarity in the fusion peptide (262), as well as the presence of highly conserved proline and cysteines in the baculovirus DUF3609 domain (Chapter 2). The similarities of sequence and functional features between Ad-F and SeMNPV-F proteins indicate that

horizontal transfer of an *F* gene may have occurred between baculoviruses and mosquitos (Chapter 4). The fact that the Ad-F gene lacks introns suggests that this cellular gene has a viral origin although it is not clear whether insect endogenous retroviruses are involved in this gene transfer. To further understand the role of Ad-F protein in the insect, deletion or down regulation of Ad-*F* gene in *Anopheles darlingi* may provide an indication.

Taken together, the research described in this dissertation, in which biochemical and structural studies were combined with evolutionary analyses, not only helps us to understand entry mechanism of baculovirus BV but also provides new insights in the evolutionary relationships between the F-like proteins found in viruses and eukaryotes.

References

1. Family: DUF3609 (PF12259): http://pfam.xfam.org/family/PF12259.
2. Family: SNARE (PF05739): http://pfam.xfam.org/family/SNARE.
3. **Abot, A. R., F. Moscardi, J. R. Fuxa, D. R. Sosa-Gómez, and A. R. Richter.** 1996. Development of resistance by *Anticarsia gemmatalis* from Brazil and the United States to a nuclear polyhedrosis virus under laboratory selection pressure. Biol. Control **7:**126-30.
4. **Adams, J. R., R. H. Goodwin, and T. A. Wilcox.** 1977. Electron microscopic investigations on invasion and replication of insect baculoviruses *in vivo* and *in vitro*. Biol. Cellulaire **28:**261-8.
5. **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:**11157-65.
6. **Albertini, A. A., E. Baquero, A. Ferlin, and Y. Gaudin.** 2012. Molecular and cellular aspects of rhabdovirus entry. Viruses **4:**117-39.
7. **Albertini, A. A., C. Merigoux, S. Libersou, K. Madiona, S. Bressanelli, S. Roche, J. Lepault, R. Melki, P. Vachette, and Y. Gaudin.** 2012. Characterization of monomeric intermediates during VSV glycoprotein structural transition. PLoS Pathog **8:**e1002556.
8. **Ames, R. S., and Q. Lu.** 2009. Viral-mediated gene delivery for cell-based assays in drug discovery. Expert Opin. Drug Deliv. **4:**243-56.
9. **Asser-Kaiser, S., E. Fritsch, K. Undorf-Spahn, J. Kienzle, K. E. Eberle, N. A. Gund, A. Reineke, C. P. W. Zebitz, D. G. Heckel, J. Huber, and J. A. Jehle.** 2007. Rapid emergence of baculovirus resistance in codling moth due to dominant, sex-linked inheritance. Science **317:**1916-18.
10. **Au, S., W. Wu, and N. Pante.** 2013. Baculovirus nuclear import: open, nuclear pore complex (NPC) sesame. Viruses **5:**1885-900.
11. **Aydin, H., B. M. Smrke, and J. E. Lee.** 2013. Structural characterization of a fusion glycoprotein from a retrovirus that undergoes a hybrid 2-step entry mechanism. FASEB J. **27:**5059-71.
12. **Ayres, M. D., S. C. Howard, J. Kuzio, M. Lopez-Ferber, and R. D. Possee.** 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. Virology **202:**586-605.
13. **Backovic, M., and T. S. Jardetzky.** 2011. Class III viral membrane fusion proteins. Adv. Exp. Med. Biol. **714:**91-101.
14. **Baker, K. A., R. E. Dutch, R. A. Lamb, and T. S. Jardetzky.** 1999. Structural basis for paramyxovirus-mediated membrane fusion. Mol. Cell **3:**309-19.
15. **Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon.** 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U.S.A. **98:**10037-41.
16. **Baltimore, D.** 1971. Expression of animal virus genomes. Bacteriolog. Rev. **35:**235-41.
17. **Bayev, A. A., Jr., N. V. Lyubomirskaya, E. B. Dzhumagaliev, E. V. Ananiev, I. G. Amiantova, and Y. V. Ilyin.** 1984. Structural organization of

transposable element mdg4 from *Drosophila melanogaster* and a nucleotide sequence of its long terminal repeats. Nucleic Acids Res. **12:**3707-23.

18. **Beas-Catena, A., A. Sánchez-Mirón, F. García-Camacho, A. Contreras-Gómez, and E. Molina-Grima.** 2014. Baculovirus biopesticides: an overview. J. Anim. Plant Sci. **24:**362-73.
19. **Beaton, C. D., and B. K. Filshie.** 1976. Comparative ultrastructural studies of insect granulosis and nuclear polyhedrosis viruses. J. Gen. Virol. **31:**151-61.
20. **Becnel, J.** 2007. Current status of deltabaculoviruses, cypoviruses and chloriridoviruses pathogenic for mosquitoes. Virologica Sinica **22:**117-27.
21. **Becnel, J. J., S. E. White, B. A. Moser, T. Fukuda, M. J. Rotstein, A. H. 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:**275-82.
22. **Belouzard, S., J. K. Millet, B. N. Licitra, and G. R. Whittaker.** 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. Viruses **4:**1011-33.
23. **Berezin, C., F. Glaser, J. Rosenberg, I. Paz, T. Pupko, P. Fariselli, R. Casadio, and N. Ben-Tal.** 2004. ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics **20:**1322-4.
24. **Bergfors, T.** 2003. Seeds to crystals. J. Struct. Biol. **142:**66-76.
25. **Berling, M., C. Blachere-Lopez, O. Soubabere, X. Lery, A. Bonhomme, B. Sauphanor, and M. Lopez-Ferber.** 2009. *Cydia pomonella* granulovirus genotypes overcome virus resistance in the codling moth and improve virus efficiency by selection against resistant hosts. Appl. Environ. Microbiol. **75:**925-30.
26. **Bianchi, F. J. J. A., N. N. Joosten, J. M. Vlak, and W. Van Der Werf.** 2001. The influence of greenhouse chrysanthemum on the interaction between the beet armyworm, *Spodoptera exigua*, and the baculovirus SeMNPV: parameter quantification for a process-based simulation model. J. Appl. Entomol. **125:**557-562.
27. **Black, B., L. Brennan, P. Dierks, and I. Gard.** 1997. Commercialization of baculoviral insecticides, p. 341-387. *In* L. K. Miller (ed.), The baculoviruses. Springer, U.S.A..
28. **Blissard, G. W.** 1996. Baculovirus-insect cell interactions. Cytotechnology **20:**73-93.
29. **Blissard, G. W., P. H. Kogan, R. Wei, and G. F. Rohrmann.** 1992. A synthetic early promoter from a baculovirus: roles of the TATA box and conserved start site CAGT sequence in basal levels of transcription. Virology **190:**783-93.
30. **Blissard, G. W., and G. F. Rohrmann.** 1990. Baculovirus diversity and molecular biology. Annu. Rev. Entomol. **35:**127-55.
31. **Blissard, G. W., and G. F. Rohrmann.** 1989. Location, sequence, transcriptional mapping, and temporal expression of the gp64 envelope glycoprotein gene of the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus. Virology **170:**537-55.
32. **Blissard, G. W., and J. R. Wenz.** 1992. Baculovirus gp64 envelope glycoprotein is sufficient to mediate pH-dependent membrane fusion. J. Virol. **66:**6829-35.
33. **Boeke, J. D., T. H. Eickbush, S. B. Sandmeyer, and D. F. Voytas.** 2012. Family *Metaviridae*, p. 359-367. *In* A. M. Q. King, M. J. Adams, E. B. Carstens, and E. J. Lefkowitz (eds.), Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses, 9th edition, Elsevier Academic Press.

34. **Borrego-Diaz, E., M. E. Peeples, R. M. Markosyan, G. B. Melikyan, and F. S. Cohen.** 2003. Completion of trimeric hairpin formation of influenza virus hemagglutinin promotes fusion pore opening and enlargement. Virology **316:**234-44.
35. **Bosch, B. J., R. van der Zee, C. A. de Haan, and P. J. Rottier.** 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J. Virol. **77:**8801-11.
36. **Bossart, K. N., D. L. Fusco, and C. C. Broder.** 2013. Paramyxovirus entry, p. 95-127, Adv. Exp. Med and Biol, vol. 790.
37. **Boucias, D. G., and J. C. Pendland.** 1998. Baculoviruses, p. 111–146, Principles of Insect Pathology. Kluwer Academic Publishers, Norwell.
38. **Boulay, F., R. W. Doms, I. Wilson, and A. Helenius.** 1987. The influenza hemagglutinin precursor as an acid-sensitive probe of the biosynthetic pathway. EMBO J. **6:**2643-50.
39. **Braunagel, S. C., P. A. Guidry, G. Rosas-Acosta, L. Engelking, and M. D. Summers.** 2001. Identification of BV/ODV-C42, an *Autographa californica* nucleopolyhedrovirus orf101-encoded structural protein detected in infected-cell complexes with ODV-EC27 and p78/83. J. Virol. **75:**12331-8.
40. **Braunagel, S. C., W. K. Russell, G. Rosas-Acosta, D. H. Russell, and M. D. Summers.** 2003. Determination of the protein composition of the occlusion-derived virus of *Autographa californica* nucleopolyhedrovirus. Proc. Natl. Acad. Sci. U.S.A. **100:**9797-802.
41. **Braunagel, S. C., and M. D. Summers.** 1994. *Autographa californica* nuclear polyhedrosis virus, PDV, and ECV viral envelopes and nucleocapsids: structural proteins, antigens, lipid and fatty acid profiles. Virology **202:**315-28.
42. **Bricogne, G., C. Vonrhein, C. Flensburg, M. Schiltz, and W. Paciorek.** 2003. Generation, representation and flow of phase information in structure determination: Recent developments in and around SHARP 2.0. Acta Crystallographica - Section D Biological Crystallography **59:**2023-30.
43. **Brown, S. E., J. E. Maruniak, and D. L. Knudson.** 1984. Physical map of SeMNPV baculovirus DNA: an AcMNPV genomic variant. Virology **136:**235-40.
44. **Bud, H. M., and D. C. Kelly.** 1980. An electron microscope study of partially lysed baculovirus nucleocapsids: the intranucleocapsid packaging of viral DNA. J. Ultrastruct. Res. **73:**361-8.
45. **Bullough, P. A., F. M. Hughson, J. J. Skehel, and D. C. Wiley.** 1994. Structure of influenza haemagglutinin at the pH of membrane fusion. Nature **371:**37-43.
46. **Burch, T. P., and M. K. Ticku.** 1981. Histidine modification with diethyl pyrocarbonate shows heterogeneity of benzodiazepine receptors. Proc. Natl. Acad. Sci. U.S.A. **78:**3945-9.
47. **Burley, S. K., A. Miller, K. A. Harrap, and D. C. Kelly.** 1982. Structure of the baculovirus nucleocapsid. Virology **120:**433-40.
48. **Carneiro, F. A., F. Stauffer, C. S. Lima, M. A. Juliano, L. Juliano, and A. T. Da Poian.** 2003. Membrane fusion induced by vesicular stomatitis virus depends on histidine protonation. J. Biol. Chem. **278:**13789-94.
49. **Carstens, E. B.** 2009. AcMNPV as a model for baculovirus DNA replication. Virol. Sin. **24:**243-67.

50. **Chanel-Vos, C., and M. Kielian.** 2004. A conserved histidine in the ij loop of the Semliki Forest virus E1 protein plays an important role in membrane fusion. J. Virol. **78:**13543-52.
51. **Chang, A., and R. E. Dutch.** 2012. Paramyxovirus fusion and entry: multiple paths to a common end. Viruses **4:**613-36.
52. **Chen, A., E. Leikina, K. Melikov, B. Podbilewicz, M. M. Kozlov, and L. V. Chernomordik.** 2008. Fusion-pore expansion during syncytium formation is restricted by an actin network. J. Cell Sci. **121:**3619-28.
53. **Chen, C.-Y., C.-Y. Lin, G.-Y. Chen, and Y.-C. Hu.** 2011. Baculovirus as a gene delivery vector: recent understandings of molecular alterations in transduced cells and latest applications. Biotechnol. Adv. **29:**618-31.
54. **Chen, C.-Y., H.-H. Wu, C.-P. Chen, S.-R. Chern, S.-M. Hwang, S.-F. Huang, W.-H. Lo, G.-Y. Chen, and Y.-C. Hu.** 2011. Biosafety assessment of human mesenchymal stem cells engineered by hybrid baculovirus vectors. Mol. Pharmaceut. **8:**1505-14.
55. **Chen, J., K. H. Lee, D. A. Steinhauer, D. J. Stevens, J. J. Skehel, and D. C. Wiley.** 1998. Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. Cell **95:**409-17.
56. **Chen, J., J. J. Skehel, and D. C. Wiley.** 1999. N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA(2) subunit to form an N cap that terminates the triple-stranded coiled coil. Proc. Natl. Acad. Sci. U.S.A. **96:**8967-72.
57. **Chen, L., J. J. Gorman, J. McKimm-Breschkin, L. J. Lawrence, P. A. Tulloch, B. J. Smith, P. M. Colman, and M. C. Lawrence.** 2001. The structure of the fusion glycoprotein of Newcastle disease virus suggests a novel paradigm for the molecular mechanism of membrane fusion. Structure **9:**255-66.
58. **Chernomordik, L. V., and M. M. Kozlov.** 2008. Mechanics of membrane fusion. Nat. Struct. Mol. Biol. **15:**675-83.
59. **Chernomordik, L. V., and M. M. Kozlov.** 2003. Protein-lipid interplay in fusion and fission of biological membranes. Annu. Rev. Biochem. **72:**175-207.
60. **Chernomordik, L. V., J. Zimmerberg, and M. M. Kozlov.** 2006. Membranes of the world unite! J. Cell Biol. **175:**201-7.
61. **Cheshenko, N., N. Krougliak, R. C. Eisensmith, and V. A. Krougliak.** 2001. A novel system for the production of fully deleted adenovirus vectors that does not require helper adenovirus. Gene Ther. **8:**846-54.
62. **Chiba, S., H. Kondo, A. Tani, D. Saisho, W. Sakamoto, S. Kanematsu, and N. Suzuki.** 2011. Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. PLoS Pathog. **7:**e1002146.
63. **Chiu, J., D. Tillett, I. W. Dawes, and P. E. March.** 2008. Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. J. Microbiol. Met. **73:**195-98.
64. **Choi, J., and L. A. Guarino.** 1995. A temperature-sensitive IE1 protein of *Autographa californica* nuclear polyhedrosis virus has altered transactivation and DNA binding activities. Virology **209:**90-8.
65. **Cohen, F. S., R. M. Markosyan, and G. B. Melikyan.** 2002. The process of membrane fusion: nipples, hemifusion, pores, and pore growth, p. 501-529. *In* S. A. Simon and T. J. McIntosh (eds.), Peptide-lipid Interactions, vol. 52.

66. **Connolly, S. A., G. P. Leser, H.-S. Yin, T. S. Jardetzky, and R. A. Lamb.** 2006. Refolding of a paramyxovirus F protein from prefusion to postfusion conformations observed by liposome binding and electron microscopy. Proc. Natl. Acad. Sci. U.S.A. **103:**17903-8.

67. **Corpet, F.** 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16:**10881-90.

68. **Cory, J., and D. L. Bishop.** 1997. Use of baculoviruses as biological insecticides. Mol. Biotechnol. **7:**303-13.

69. **Cox, M. M.** 2012. Recombinant protein vaccines produced in insect cells. Vaccine **30:**1759-66.

70. **D'Arcy, A., F. Villard, and M. Marsh.** 2007. An automated microseed matrix-screening method for protein crystallization. Acta Crystallographica Section D: Biological Crystallography **63:**550-4.

71. **Danquah, J. O., S. Botchway, A. Jeshtadi, and L. A. King.** 2012. Direct interaction of baculovirus capsid proteins VP39 and EXON0 with Kinesin-1 in insect cells determined by fluorescence resonance energy transfer-fluorescence lifetime imaging microscopy. J. Virol. **86:**844-53.

72. **Davis, I. W., L. W. Murray, J. S. Richardson, and D. C. Richardson.** 2004. MolProbity: Structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res. **32 suppl. 2):**W615-W619.

73. **Davis, T. R., K. M. Trotter, R. R. Granados, and H. A. Wood.** 1992. Baculovirus expression of alkaline phosphatase as a reporter gene for evaluation of production, glycosylation and secretion. Nat. Biotechnol. **10:**1148-50.

74. **de A. Zanotto, P. M., B. D. Kessing, and J. E. Maruniak.** 1993. Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations. J. Invertebr. Pathol. **62:**147-64.

75. **Degreve, L., and C. A. Fuzo.** 2013. Structure and dynamics of the monomer of protein E of dengue virus type 2 with unprotonated histidine residues. Genet. Mol. Res. **12:**348-59.

76. **Delsuc, F., H. Brinkmann, D. Chourrout, and H. Philippe.** 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature **439:**965-8.

77. **Dessau, M., and Y. Modis.** 2013. Crystal structure of glycoprotein C from Rift Valley fever virus. Proc. Natl. Acad. Sci. U.S.A. **110:**1696-701.

78. **Doms, R. W., A. Helenius, and J. White.** 1985. Membrane fusion activity of the influenza virus hemagglutinin. J. Biol. Chem. **260:**2973-81.

79. **Doms, R. W., A. Helenius, and J. White.** 1985. Membrane fusion activity of the influenza virus hemagglutinin. The low pH-induced conformational change. J. Biol. Chem. **260:**2973-81.

80. **Donald, J. E., Y. Zhang, G. Fiorin, V. Carnevale, D. R. Slochower, F. Gai, M. L. Klein, and W. F. DeGrado.** 2011. Transmembrane orientation and possible role of the fusogenic peptide from parainfluenza virus 5 (PIV5) in promoting fusion. Proc. Natl. Acad. Sci. U.S.A. **108:**3958-63.

81. **Dong, S., and G. W. Blissard.** 2012. Functional analysis of the *Autographa californica* multiple nucleopolyhedrovirus GP64 terminal fusion loops and interactions with membranes. J. Virol. **86:**9617-28.

82. **Dong, S., M. Wang, Z. Qiu, F. Deng, J. M. Vlak, Z. Hu, and H. Wang.** 2010. *Autographa californica* multicapsid nucleopolyhedrovirus efficiently infects Sf9 cells and transduces mammalian cells via direct fusion with the plasma membrane at low pH. J. Virol. **84:**5351-59.

83. **DuBois, R. M., M. C. Vaney, M. A. Tortorici, R. A. Kurdi, G. Barba-Spaeth, T. Krey, and F. A. Rey.** 2013. Functional and evolutionary insight from the crystal structure of rubella virus protein E1. Nature **493:**552-6.

84. **Dutch, R. E., R. N. Hagglund, M. A. Nagel, R. G. Paterson, and R. A. Lamb.** 2001. Paramyxovirus fusion (F) protein: a conformational change on cleavage activation. Virology **281:**138-50.

85. **Eckert, D. M., and M. S. Kay.** 2010. Stalking influenza. Proc. Natl. Acad. Sci. U.S.A. **107:**13563-4.

86. **Eckert, D. M., and P. S. Kim.** 2001. Mechanisms of viral membrane fusion and its inhibition. Annu. Rev. Biochem. **70:**777-810.

87. **Eickbush, T. H., and H. S. Malik.** 2002. Origins and evolution of retrotransposons, p.1111-44. *In* Mobile DNA ii, ASM Press, Washington, DC.

88. **Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan.** 2010. Features and development of Coot. Acta Crystallographica Section D: Biological Crystallography **66:**486-501.

89. **Federici, B. A.** 1980. Mosquito baculovirus: Sequence of morphogenesis and ultrastructure of the virion. Virology **100:**1-9.

90. **Fernandes, F., A. P. Teixeira, N. Carinhas, M. J. T. Carrondo, and P. M. Alves.** 2013. Insect cells as a production platform of complex virus-like particles. Expert Rev. Vaccines **12:**225-36.

91. **Feschotte, C., and C. Gilbert.** 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet **13:**283-96.

92. **Forterre, P.** 2006. The origin of viruses and their possible roles in major evolutionary transitions. Virus Res. **117:**5-16.

93. **Fraser, M. J.** 1986. Ultrastructural observations of virion maturation in *Autographa californica* nuclear polyhedrosis virus infected *Spodoptera frugiperda* cell cultures. J. Ultra. Mol. Struct. R. **95:**189-95.

94. **Fraser, M. J., J. S. Brusca, G. E. Smith, and M. D. Summers.** 1985. Transposon-mediated mutagenesis of a baculovirus. Virology **145:**356-361.

95. **Friesen, P. D.** 1997. Regulation of baculovirus early gene expression, p. 141-170. *In* L. K. Miller (ed.), The baculoviruses. Springer US.

96. **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:**3067-77.

97. **Friesen, P. D., W. C. Rice, D. W. Miller, and L. K. Miller.** 1986. Bidirectional transcription from a solo long terminal repeat of the retrotransposon TED: symmetrical RNA start sites. Molec. Cell. Biol **6:**1599-607.

98. **Fritz, R., K. Stiasny, and F. X. Heinz.** 2008. Identification of specific histidines as pH sensors in flavivirus membrane fusion. J. Cell Biol. **183:**353-61.

99. **Fuller, D. N., D. M. Raymer, V. I. Kottadiel, V. B. Rao, and D. E. Smith.** 2007. Single phage T4 DNA packaging motors exhibit large force generation, high velocity, and dynamic variability. Proc. Natl. Acad. Sci. U.S.A. **104:**16868-73.

100. **Fuller, D. N., D. M. Raymer, J. P. Rickgauer, R. M. Robertson, C. E. Catalano, D. L. Anderson, S. Grimes, and D. E. Smith.** 2007. Measurements of single DNA molecule packaging dynamics in bacteriophage lambda reveal high forces, high motor processivity, and capsid transformations. J. Mol. Biol. **373:**1113-22.

101. **Fuzo, C. A., and L. Degreve.** 2013. The pH dependence of flavivirus envelope protein structure: insights from molecular dynamics simulations. J. Biomol. Struct. Dyn. **32:**1563-74

102. **Garavaglia, M. J., S. A. B. Miele, J. A. Iserte, M. N. Belaich, and P. D. Ghiringhellia.** 2012. The *ac53*, *ac78*, *ac101*, and *ac103* genes are newly discovered core genes in the family *Baculoviridae*. J. Virol. **86:**12069-79.

103. **Gardner, A. E., and R. E. Dutch.** 2007. A conserved region in the F2 subunit of paramyxovirus fusion proteins is involved in fusion regulation. J. Virol. **81:**8303-14.

104. **Garman, E. F., and R. L. Owen.** 2006. Cryocooling and radiation damage in macromolecular crystallography. Acta Crystallographica Section D: Biological Crystallography **62:**32-47.

105. **Garry, C. E., and R. F. Garry.** 2008. Proteomics computational analyses suggest that baculovirus GP64 superfamily proteins are class III penetrenes. Virol. J. **5:**28.

106. **Gaudin, Y.** 2002. Reversibility in fusion protein conformational changes the intriguing case of rhabdovirus-induced membrane fusion, p. 379-408. *In* H. Hilderson and S. Fuller (ed.), Fusion of biological membranes and related problems, vol. 34. Springer US.

107. **Gaudin, Y., C. Tuffereau, P. Durrer, A. Flamand, and R. W. H. Ruigrok.** 1995. Biological function of the low-pH, fusion-inactive conformation of rabies virus glycoprotein (G): G is transported in a fusion-inactive state-like conformation. J. Virol. **69:**5528-34.

108. **Gaudin, Y., C. Tuffereau, D. Segretain, M. Knossow, and A. Flamand.** 1991. Reversible conformational changes and fusion activity of rabies virus glycoprotein. J. Virol. **65:**4853-9.

109. **Gibbons, D. L., M. C. Vaney, A. Roussel, A. Vigouroux, B. Reilly, J. Lepault, M. Kielian, and F. A. Rey.** 2004. Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. Nature **427:**320-5.

110. **Gilmartin, A. A., B. Lamp, T. Rümenapf, M. A. A. Persson, F. A. Rey, and T. Krey.** 2012. High-level secretion of recombinant monomeric murine and human single-chain Fv antibodies from *Drosophila* S2 cells. Prot. Engin., Design and Selection **25:**59-66.

111. **Glaser, F., T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz, and N. Ben-Tal.** 2003. ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics **19:**163-4.

112. **Goley, E. D., T. Ohkawa, J. Mancuso, J. B. Woodruff, J. A. D'Alessio, W. Z. Cande, L. E. Volkman, and M. D. Welch.** 2006. Dynamic nuclear actin assembly by Arp2/3 complex and a baculovirus WASP-like protein. Science **314:**464-7.

113. **Gonzalez-Reyes, L., M. B. Ruiz-Arguello, B. Garcia-Barreno, L. Calder, J. A. Lopez, J. P. Albar, J. J. Skehel, D. C. Wiley, and J. A. Melero.** 2001. Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion. Proc. Natl. Acad. Sci. U.S.A. **98:**9859-64.

114. **González-Reyes, L., M. B. Ruiz-Argüello, B. García-Barreno, L. Calder, J. A. López, J. P. Albar, J. J. Skehel, D. C. Wiley, and J. A. Melero.** 2001. Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion. Proc. Natl. Acad. Sci. U.S.A. **98:**9859-64.

115. **Gouet, P., X. Robert, and E. Courcelle.** 2003. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res. **31:**3320-3.

116. **Granados, R. R., and K. A. Lawler.** 1981. In vivo pathway of *Autographa californica* baculovirus invasion and infection. Virology **108:**297-308.

117. **Gray, C., and L. K. Tamm.** 1998. pH-induced conformational changes of membrane-bound influenza hemagglutinin and its effect on target lipid bilayers. Protein Sci. **7:**2359-73.

118. **Gross, K.** 1964. Studies on the mechanism of carcinogenesis (Fusion and life-cycles of mammalian cells). Oncology **18:**178-94.

119. **Grove, J., and M. Marsh.** 2011. The cell biology of receptor-mediated virus entry. J. Cell Biol. **195:**1071-82.

120. **Guarino, L. A., G. Smith, and W. Dong.** 1995. Ubiquitin is attached to membranes of baculovirus particles by a novel type of phospholipid anchor. Cell **80:**301-9.

121. **Haines, F. J., C. M. Griffiths, R. D. Possee, C. R. Hawes, and L. A. King.** 2009. Involvement of lipid rafts and cellular actin in AcMNPV GP64 distribution and virus budding. Virol. Sin. **24:**333-49.

122. **Hallenberger, S., V. Bosch, H. Angliker, E. Shaw, H. D. Klenk, and W. Garten.** 1992. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. Nature **360:**358-61.

123. **Harrap, K. A.** 1972. The structure of nuclear polyhedrosis viruses: II. The virus particle. Virology **50:**124-32.

124. **Harris, A., G. Cardone, D. C. Winkler, J. B. Heymann, M. Brecher, J. M. White, and A. C. Steven.** 2006. Influenza virus pleiomorphy characterized by cryoelectron tomography. Proc. Natl. Acad. Sci. U.S.A. **103:**19123-7.

125. **Harrison, J. S., C. D. Higgins, M. J. O'Meara, J. F. Koellhoffer, B. A. Kuhlman, and J. R. Lai.** 2013. Role of electrostatic repulsion in controlling pH-dependent conformational changes of viral fusion proteins. Structure **21:**1085-96.

126. **Harrison, R., and K. Hoover.** 2012. Baculoviruses and other occluded insect viruses, p. 73-131. *In* F. E. Vega and H. K. Kaya (eds.), Insect Pathology, 2nd ed. Elsevier Inc.

127. **Harrison, S. C.** 2008. Viral membrane fusion. Nat. Struct. Mol. Biol. **15:**690-8.

128. **Havecker, E. R., X. Gao, and D. F. Voytas.** 2004. The diversity of LTR retrotransposons. Genome Biol. **5:**225.

129. **Hefferon, K. L., and L. K. Miller.** 2002. Reconstructing the replication complex of AcMNPV. Eur. J. Biochem. **269:**6233-40.

130. **Hefferon, K. L., A. G. Oomens, S. A. Monsma, C. M. Finnerty, and G. W. Blissard.** 1999. Host cell receptor binding by baculovirus GP64 and kinetics of virion entry. Virology **258:**455-68.

131. **Heinz, F. X., K. Stiasny, G. Püschner-Auer, H. Holzmann, S. L. Allison, C. W. Mandl, and C. Kunz.** 1994. Structural changes and functional control of the

tick-borne encephalitis virus glycoprotein E by the heterodimeric association with protein prM. Virology **198:**109-17.

132. **Heldwein, E. E., H. Lou, F. C. Bender, G. H. Cohen, R. J. Eisenberg, and S. C. Harrison.** 2006. Crystal structure of glycoprotein B from herpes simplex virus 1. Science **313:**217-20.

133. **Heldwein, E. E., H. Lou, F. C. Bender, G. H. Cohen, R. J. Eisenberg, and S. C. Harrison.** 2006. Crystal structure of glycoprotein B from herpes simplex virus 1. Science **313:**217-20.

134. **Herniou, E. A., B. M. Arif, J. J. Becnel, G. W. Blissard, B. Bonning, R. Harrison, J. A. Jehle, D. A. Theilmann, and J. M. Vlak.** 2012. Family *Baculoviridae*, p. 163-173. *In* A. M. Q. King, M. J. Adams, E. B. Carstens, and E. J. Lefkowitz (eds.), Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses, 9th ed. Elsevier Academic Press.

135. **Herniou, E. A., J. A. Olszewski, J. S. Cory, and D. R. O'Reilly.** 2003. The genome sequence and evolution of baculoviruses. Annu. Rev. Entomol. **48:**211-34.

136. **Herniou, E. A., J. A. Olszewski, D. R. O'Reilly, and J. S. Cory.** 2004. Ancient coevolution of baculoviruses and their insect hosts. J. Virol. **78:**3244-51.

137. **Hitchman, R. B., R. D. Possee, and L. A. King.** 2009. Baculovirus expression systems for recombinant protein production in insect cells. Recent Patents on Biotechnol. **3:**46-54.

138. **Hitoshi, N., Y. Ken-ichi, and M. Jun-ichi.** 1991. Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene **108:**193-9.

139. **Holm, C. K., S. B. Jensen, M. R. Jakobsen, N. Cheshenko, K. A. Horan, H. B. Moeller, R. Gonzalez-Dosal, S. B. Rasmussen, M. H. Christensen, T. O. Yarovinsky, F. J. Rixon, B. C. Herold, K. A. Fitzgerald, and S. R. Paludan.** 2012. Virus-cell fusion as a trigger of innate immunity dependent on the adaptor STING. Nature Immunology **13:**737-43.

140. **Holm, L., and P. Rosenstrom.** 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res **38:**W545-9.

141. **Holmes, E. C.** 2003. Molecular clocks and the puzzle of RNA virus origins. J. Virol. **77:**3893-7.

142. **Hooft, R. W., C. Sander, and G. Vriend.** 1996. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. Proteins **26:**363-76.

143. **Horie, M., T. Honda, Y. Suzuki, Y. Kobayashi, T. Daito, T. Oshida, K. Ikuta, P. Jern, T. Gojobori, J. M. Coffin, and K. Tomonaga.** 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature **463:**84-7.

144. **Hou, D., L. Zhang, F. Deng, F. Wei, R. Wang, X. Liu, L. Guo, S. Rayner, X. Chen, H. Wang, and Z. Hu.** 2013. Comparative proteomics reveal fundamental structural and functional differences between the two progeny phenotypes of a baculovirus. J. Virol. **87:**829-39.

145. **Hu, Y. W., L. Rocheleau, B. Larke, L. Chui, B. Lee, M. Ma, S. Liu, T. Omlin, M. Pelchat, and E. G. Brown.** 2005. Immunoglobulin mimicry by Hepatitis C Virus envelope protein E2. Virology **332:**538-49.

146. **Hua-Van, A., A. Le Rouzic, T. S. Boutin, J. Filée, and P. Capy.** 2011. The struggle for life of the genome's selfish architects. Biol. Direct **6:**19.

147. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754-5.

148. **Igonet, S., M. C. Vaney, C. Vonhrein, G. Bricogne, E. A. Stura, H. Hengartner, B. Eschli, and F. A. Rey.** 2011. X-ray structure of the arenavirus glycoprotein GP2 in its postfusion hairpin conformation. Proc. Natl. Acad. Sci. U.S.A. **108:**19967-72.

149. **IJkel, W. F. J., E. A. van Strien, J. G. Heldens, R. Broer, D. Zuidema, R. W. Goldbach, and J. M. Vlak.** 1999. Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. J. Gen. Virol. **80:**3289-304.

150. **IJkel, W. F. J., 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.

151. **Inceoglu, A. B., S. G. Kamita, and B. D. Hammock.** 2006. Genetically modified baculoviruses: a historical overview and future outlook, p. 323-360. *In* K. M. Bryony C. Bonning and J. S. Aaron (eds.), Adv. Virus Res., vol. 68. Academic Press.

152. **Ireton, G. C., and B. L. Stoddard.** 2004. Microseed matrix screening to improve crystals of yeast cytosine deaminase. Acta Crystallographica Section D: Biological Crystallography **60:**601-605.

153. **Jahn, R., T. Lang, and T. C. Südhof.** 2003. Membrane fusion. Cell **112:**519-33.

154. **Jahn, R., and R. H. Scheller.** 2006. SNAREs--engines for membrane fusion. Nat. Rev. Mol. Cell Biol. **7:**631-43.

155. **Jardetzky, T. S., and R. A. Lamb.** 2014. Activation of paramyxovirus membrane fusion and virus entry. Curr. Opin. Virol. **5:**24-33.

156. **Jehle, J. A., G. W. Blissard, B. C. Bonning, J. S. Cory, E. A. Herniou, G. F. Rohrmann, D. A. Theilmann, S. M. Thiem, and J. M. Vlak.** 2006. On the classification and nomenclature of baculoviruses: a proposal for revision. Arch. Virol. **151:**1257-66.

157. **Jiang, Y., F. Deng, S. Rayner, H. Wang, and Z. Hu.** 2009. Evidence of a major role of GP64 in group I alphabaculovirus evolution. Virus Res. **142:**85-91.

158. **Jordan, I. K., L. V. Matyunina, and J. F. McDonald.** 1999. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. Proc. Natl. Acad. Sci. U.S.A. **96:**12621-5.

159. **Julien, J. P., A. Cupo, D. Sok, R. L. Stanfield, D. Lyumkis, M. C. Deller, P. J. Klasse, D. R. Burton, R. W. Sanders, J. P. Moore, A. B. Ward, and I. A. Wilson.** 2013. Crystal structure of a soluble cleaved HIV-1 envelope trimer. Science **342:**1477-83.

160. **Kabsch, W.** 2010. XDS. Acta Crystallographica Section D: Biological Crystallography **66:**125-132.

161. **Kadlec, J., S. Loureiro, N. G. A. Abrescia, D. I. Stuart, and I. M. Jones.** 2008. The postfusion structure of baculovirus gp64 supports a unified view of viral fusion machines. Nat. Struct. Mol. Biol. **15:**1024-30.

162. **Kalani, M. R., A. Moradi, M. Moradi, and E. Tajkhorshid.** 2013. Characterizing a histidine switch controlling pH-dependent conformational changes of the influenza virus hemagglutinin. Biophys. J. **105:**993-1003.

163. **Kam, Z., H. B. Shore, and G. Feher.** 1978. On the crystallization of proteins. J. Mol. Biol. **123:**539-55.

164. **Kaminker, J. S., C. M. Bergman, B. Kronmiller, J. Carlson, R. Svirskas, S. Patel, E. Frise, D. A. Wheeler, S. E. Lewis, G. M. Rubin, M. Ashburner, and S. E. Celniker.** 2002. The transposable elements *of the Drosophila melanogaster* euchromatin: a genomics perspective. Genome Biol. **3 (12):**research0084.1-20.

165. **Kampmann, T., D. S. Mueller, A. E. Mark, P. R. Young, and B. Kobe.** 2006. The role of histidine residues in low-pH-mediated viral membrane fusion. Structure **14:**1481-7.

166. **Kaname, Y., H. Tani, C. Kataoka, M. Shiokawa, S. Taguwa, T. Abe, K. Moriishi, T. Kinoshita, and Y. Matsuura.** 2010. Acquisition of complement resistance through incorporation of CD55/DAF into viral particles bearing baculovirus GP64. J. Virol. **84:**3210-9.

167. **Kato, T., M. Kajikawa, K. Maenaka, and E. Y. Park.** 2010. Silkworm expression system as a platform technology in life science. Appl. Microbiol. Biotechnol. **85:**459-70.

168. **Katoh, K., and D. M. Standley.** 2013. MAFFT multiple sequence alignment software Version 7: improvements in performance and usability. Mol. Biol. Evol. **30:**772-80.

169. **Kawanishi, C. Y., M. D. Summers, D. B. Stoltz, and H. J. Arnott.** 1972. Entry of an insect virus *in vivo* by fusion of viral envelope and microvillus membrane. J. Invertebr. Pathol. **20:**104-8.

170. **Kelley, L. A., and M. J. Sternberg.** 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. **4:**363-71.

171. **Kidd, I. M., and V. C. Emery.** 1993. The use of baculoviruses as expression vectors. Appl. Biochem. Biotechnol. **42:**137-59.

172. **Kielian, M., S. Jungerwirth, K. U. Sayad, and S. DeCandido.** 1990. Biosynthesis, maturation, and acid activation of the Semliki Forest virus fusion protein. J. Virol. **64:**4614-24.

173. **Kim, A., C. Terzian, P. Santamaria, A. Pélisson, N. Prud'homme, and A. Bucheton.** 1994. Retroviruses in invertebrates: The gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U.S.A. **91:**1285-9.

174. **Kim, A., C. Terzian, P. Santamaria, A. Pelisson, N. Prudhomme, and A. Bucheton.** 1994. Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U.S.A. **91:**1285-9.

175. **Kim, H. S., O. Takenaka, and T. J. Crow.** 1999. Isolation and phylogeny of endogenous retrovirus sequences belonging to the HERV-W family in primates. J. Gen. Virol. **80:**2613-9.

176. **Klenk, H. D.** 1996. Post-translational modifications in insect cells. Cytotechnology **20:**139-44.

177. **Kolodny-Hirsch, D. M., T. Sitchawat, T. Jansiri, A. Chenrchaivachirakul, and U. Ketunuti.** 1997. Field evaluation of a commercial formulation of the *Spodoptera exigua* (Lepidoptera: Noctuidae) nuclear polyhedrosis virus for control of beet armyworm on vegetable crops in Thailand. Biocontrol Sci. Technol. **7:**475-88.

178. **Kool, M., C. H. Ahrens, J. M. Vlak, and G. F. Rohrmann.** 1995. Replication of baculovirus DNA. J. Gen. Virol. **76:**2103-18.

179. **Korchin, Y., F. Frolow, O. Bogin, M. Peretz, A. J. Kalb, and Y. Burstein.** 1996. Crystalline alcohol dehydrogenases from the mesophilic bacterium *Clostridium beijerinckii* and the thermophilic bacterium *Thermoanaerobium brockii*: Preparation, characterization and molecular symmetry. Acta Crystallographica Section D: Biological Crystallography **52:**882-6.

180. **Kozlovsky, Y., L. V. Chernomordik, and M. M. Kozlov.** 2002. Lipid intermediates in membrane fusion: formation, structure, and decay of hemifusion diaphragm. Biophys. J. **83:**2634-51.

181. **Krishnan, A., S. K. Verma, P. Mani, R. Gupta, S. Kundu, and D. P. Sarkar.** 2009. A histidine switch in hemagglutinin-neuraminidase triggers paramyxovirus-cell membrane fusion. J. Virol. **83:**1727-41.

182. **Krissinel, E.** 2007. On the relationship between sequence and structure similarities in proteomics. Bioinformatics **23:**717-23.

183. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567-80.

184. **Kubo, Y., H. Hayashi, T. Matsuyama, H. Sato, and N. Yamamoto.** 2012. Retrovirus entry by endocytosis and cathepsin proteases. Adv. Virol. **2012:**640894.

185. **Kuhn, R. J., W. Zhang, M. G. Rossmann, S. V. Pletnev, J. Corver, E. Lenches, C. T. Jones, S. Mukhopadhyay, P. R. Chipman, E. G. Strauss, T. S. Baker, and J. H. Strauss.** 2002. Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. Cell **108:**717-25.

186. **Kwong, P. D., R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski, and W. A. Hendrickson.** 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature **393:**648-59.

187. **Lamb, R. A., and T. S. Jardetzky.** 2007. Structural basis of viral invasion: lessons from paramyxovirus F. Curr. Opin. Struct. Biol. **17:**427-36.

188. **Lamb, R. A., and T. S. Jardetzky.** 2007. Structural basis of viral invasion: lessons from paramyxovirus F. Curr. Opin. Struct .Biol. **17:**427-36.

189. **Lamb, R. A., R. G. Paterson, and T. S. Jardetzky.** 2006. Paramyxovirus membrane fusion: lessons from the F and HN atomic structures. Virology **344:**30-7.

190. **Lanier, L. M., and L. E. Volkman.** 1998. Actin binding and nucleation by *Autographa californica* M nucleopolyhedrovirus. Virology **243:**167-77.

191. **Lasa, R., I. Pagola, I. ibañez, J. E. Belda, T. Williams, and P. Caballero.** 2007. Efficacy of *Spodoptera exigua* multiple nucleopolyhedrovirus as a biological insecticide for beet armyworm control in greenhouses of southern Spain. Biocontrol Sci. Technol. **17:**221-232.

192. **Laskowski, R. A.** 2009. PDBsum new things. Nucleic Acids Res. **37:**D355-D359.

193. **Lavoie, C. A., R. N. Platt, 2nd, P. A. Novick, B. A. Counterman, and D. A. Ray.** 2013. Transposable element evolution in *Heliconius* suggests genome diversity within Lepidoptera. Mob. DNA **4:**21.

194. **Le Rouzic, A., T. Payen, and A. Hua-Van.** 2013. Reconstructing the evolutionary history of transposable elements. Genome Biol. Evol. **5:**77-86.

195. **Lee, J. E., M. L. Fusco, A. J. Hessell, W. B. Oswald, D. R. Burton, and E. O. Saphire.** 2008. Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. Nature **454:**177-182.

196. **Lee, J. K., A. Prussia, T. Paal, L. K. White, J. P. Snyder, and R. K. Plemper.** 2008. Functional interaction between paramyxovirus fusion and attachment proteins. J. Biol. Chem. **283:**16561-72.

197. **Lerch, R. A., and P. D. Friesen.** 1992. The baculovirus-integrated retrotransposon TED encodes *gag* and *pol* proteins that assemble into viruslike particles with reverse transcriptase. J. Virol. **66:**1590-1601.

198. **Lerch, R. A., and P. D. Friesen.** 1992. The baculovirus-integrated retrotransposon TED encodes *gag* and *pol* proteins that assemble into viruslike particles with reverse transcriptase. J. Virol. **66:**1590-601.

199. **Lescar, J., A. Roussel, M. W. Wien, J. Navaza, S. D. Fuller, G. Wengler, G. Wengler, and F. A. Rey.** 2001. The fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. Cell **105:**137-48.

200. **Leslie, A. G. W.** 2006. The integration of macromolecular diffraction data. Acta Crystallographica Section D: Biological Crystallography **62:**48-57.

201. **Li, F., M. Berardi, W. Li, M. Farzan, P. R. Dormitzer, and S. C. Harrison.** 2006. Conformational states of the severe acute respiratory syndrome coronavirus spike protein ectodomain. J. Virol. **80:**6794-800.

202. **Li, L., J. Jose, Y. Xiang, R. J. Kuhn, and M. G. Rossmann.** 2010. Structural changes of envelope proteins during alphavirus fusion. Nature **468:**705-8.

203. **Li, L., S. M. Lok, I. M. Yu, Y. Zhang, R. J. Kuhn, J. Chen, and M. G. Rossmann.** 2008. The flavivirus precursor membrane-envelope protein complex: Structure and maturation. Science **319:**1830-4.

204. **Li, S. C., and P. M. Kane.** 2009. The yeast lysosome-like vacuole: endpoint and crossroads. BBA-Mol. Cell Res. **1793:**650-63.

205. **Li, Z., and G. W. Blissard.** 2011. *Autographa californica* multiple nucleopolyhedrovirus GP64 protein: roles of histidine residues in triggering membrane fusion and fusion pore expansion. J. Virol. **85:**12492-504.

206. **Lin, L., J. Wang, R. Deng, J. Ke, H. Wu, and X. Wang.** 2009. *Ac109* is required for the nucleocapsid assembly of *Autographa californica* multiple nucleopolyhedrovirus. Virus Res. **144:**130-5.

207. **Lin, X., Y. Chen, Y. Yi, and Z. Zhang.** 2010. Baculovirus immediately early 1, a mediator for homologous regions enhancer function in trans. Virol. J. **7:**32.

208. **Liu, J., A. Bartesaghi, M. J. Borgnia, G. Sapiro, and S. Subramaniam.** 2008. Molecular architecture of native HIV-1 gp120 trimers. Nature **455:**109-13.

209. **Llopis, J., J. M. McCaffery, A. Miyawaki, M. G. Farquhar, and R. Y. Tsien.** 1998. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. Proc. Natl. Acad. Sci. U.S.A. **95:**6803-8.

210. **Llorens, C., A. Munoz-Pomer, L. Bernad, H. Botella, and A. Moya.** 2009. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. Biol. Direct **4:**41.

211. **Long, G.** 2007. Structure-function relationship of the baculovirus envelope fusion protein F. PhD thesis (150 pp). Wageningen University, Wageningen.

212. **Long, G., X. Pan, and J. M. Vlak.** 2007. Absence of N-linked glycans from the F2 subunit of the major baculovirus envelope fusion protein F enhances fusogenicity. J. Gen. Virol. **88:**441-9.

213. **Long, G., X. Pan, and J. M. Vlak.** 2008. Conserved leucines in N-terminal heptad repeat HR1 of envelope fusion protein F of group II nucleopolyhedroviruses are important for correct processing and essential for fusogenicity. J. Virol. **82:**2437-47.

214. **Long, G., X. Pan, M. Westenberg, and J. M. Vlak.** 2006. Functional role of the cytoplasmic tail domain of the major envelope fusion protein of group II baculoviruses. J. Virol. **80:**11226-34.

215. **Long, G., X. Y. Pan, R. Kormelink, and J. M. Vlak.** 2006. Functional entry of baculovirus into insect and mammalian cells is dependent on clathrin-mediated endocytosis. J. Virol. **80:**8830-33.

216. **Long, G., M. Westenberg, H. Wang, J. M. Vlak, and Z. Hu.** 2006. Function, oligomerization and *N*-linked glycosylation of the *Helicoverpa armigera* single nucleopolyhedrovirus envelope fusion protein. J. Gen. Virol. **87:**839-46.

217. **Lopez, S., J. S. Yao, R. J. Kuhn, E. G. Strauss, and J. H. Strauss.** 1994. Nucleocapsid-glycoprotein interactions required for assembly of alphaviruses. J. Virol. **68:**1316-23.

218. **Lu, A., and E. B. Carstens.** 1992. Nucleotide sequence and transcriptional analysis of the p80 gene of *Autographa californica* nuclear polyhedrosis virus: A homologue of the *Orgyia pseudotsugata* nuclear polyhedrosis virus capsid-associated gene. Virology **190:**201-9.

219. **Lu, A., and L. K. Miller.** 1995. The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. J. Virol. **69:**975-82.

220. **Lu, H. Y., Y. H. Chen, and H. J. Liu.** 2012. Baculovirus as a vaccine vector. Bioengineered Bugs **3**.

221. **Lua, L. H., and S. Reid.** 2000. Virus morphogenesis of *Helicoverpa armigera* nucleopolyhedrovirus in *Helicoverpa zea* serum-free suspension culture. J. Gen. Virol. **81:**2531-43.

222. **Lung, O., and G. W. Blissard.** 2005. A cellular *Drosophila melanogaster* protein with similarity to baculovirus F envelope fusion proteins. J. Virol. **79:**7979-89.

223. **Lung, O., M. Westenberg, J. M. Vlak, D. Zuidema, and G. W. Blissard.** 2002. Pseudotyping *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV): F proteins from group II NPVs are functionally analogous to AcMNPV GP64. J. Virol. **76:**5729-36.

224. **Lung, O., M. Westenberg, J. M. Vlak, D. Zuidema, and G. W. Blissard.** 2002. Pseudotyping *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV): F proteins from group II NPVs are functionally analogous to AcMNPV GP64. J. Virol. **76:**5729-36.

225. **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:**328-39.

226. **Lupas, A., S. Müller, K. Goldie, A. M. Engel, A. Engel, and W. Baumeister.** 1995. Model structure of the Ompα rod, a parallel four-stranded coiled coil from the hyperthermophilic *Eubacterium thermotoga maritima*. J. Mol. Biol. **248:**180-9.

227. **Lupas, A., M. Van Dyke, and J. Stock.** 1991. Predicting coiled coils from protein sequences. Science **252:**1162-4.

228. **Lyumkis, D., J. P. Julien, N. De Val, A. Cupo, C. S. Potter, P. J. Klasse, D. R. Burton, R. W. Sanders, J. P. Moore, B. Carragher, I. A. Wilson, and A. B. Ward.** 2013. Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. Science **342:**1484-90.

229. **Mair, C. M., K. Ludwig, A. Herrmann, and C. Sieben.** 2014. Receptor binding and pH stability — How influenza A virus hemagglutinin affects host-specific virus infection. BBA-Biomembranes **1838:**1153-68.

230. **Mair, C. M., T. Meyer, K. Schneider, Q. Huang, M. Veit, and A. Herrmann.** 2014. A histidine residue of the influenza virus hemagglutinin controls the pH dependence of the conformational change mediating membrane fusion. J. Virol. **88:**13189-200.

231. **Malik, H. S., and S. Henikoff.** 2005. Positive selection of *Iris*, a retroviral envelope–derived host gene in *Drosophila melanogaster*. PLoS Genet **1:**e44.

232. **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:**1307-18.

233. **Mao, Y., L. Wang, C. Gu, A. Herschhorn, A. Désormeaux, A. Finzi, S. H. Xiang, and J. G. Sodroski.** 2013. Molecular architecture of the uncleaved HIV-1 envelope glycoprotein trimer. Proc. Natl. Acad. Sci. U.S.A. **110:**12438-43.

234. **Marek, M., O. W. Merten, F. Francis-Devaraj, and M. M. van Oers.** 2012. Essential C-terminal region of the baculovirus minor capsid protein VP80 binds DNA. J. Virol. **86:**1728-38.

235. **Marek, M., O. W. Merten, L. Galibert, J. M. Vlak, and M. M. van Oers.** 2011. Baculovirus VP80 protein and the F-actin cytoskeleton interact and connect the viral replication factory with the nuclear periphery. J. Virol. **85:**5350-62.

236. **Marheineke, K., S. Grunewald, W. Christie, and H. Reilander.** 1998. Lipid composition of *Spodoptera frugiperda* (Sf9) and *Trichoplusia ni* (Tn) insect cells used for baculovirus infection. FEBS Lett. **441:**49-52.

237. **Marinotti, O., G. C. Cerqueira, L. G. de Almeida, M. I. Ferro, E. L. Loreto, A. Zaha, S. M. Teixeira, A. R. Wespiser, E. S. A. Almeida, A. D. Schlindwein, A. C. Pacheco, A. L. Silva, B. R. Graveley, B. P. Walenz, A. Lima Bde, C. A. Ribeiro, C. G. Nunes-Silva, C. R. de Carvalho, C. M. Soares, C. B. de Menezes, C. Matiolli, D. Caffrey, D. A. Araujo, D. M. de Oliveira, D. Golenbock, E. C. Grisard, F. Fantinatti-Garboggini, F. M. de Carvalho, F. G. Barcellos, F. Prosdocimi, G. May, G. M. Azevedo Junior, G. M. Guimaraes, G. H. Goldman, I. Q. Padilha, S. Batista Jda, J. A. Ferro, J. M. Ribeiro, J. L. Fietto, K. M. Dabbas, L. Cerdeira, L. F. Agnez-Lima, M. Brocchi, M. O. de Carvalho, M. Teixeira Mde, M. Diniz Maia Mde, M. H. Goldman, M. P. Cruz Schneider, M. S. Felipe, M. Hungria, M. F. Nicolas, M. Pereira, M. A. Montes, M. E. Cantao, M. Vincentz, M. S. Rafael, N. Silverman, P. H. Stoco, R. C. Souza, R. Vicentini, R. T. Gazzinelli, O. Neves Rde, R. Silva, S. Astolfi-Filho, T. E. Maciel, T. P. Urmenyi, W. P. Tadei, E. P. Camargo, and A. T. de**

Vasconcelos. 2013. The genome of *Anopheles darlingi*, the main neotropical malaria vector. Nucleic Acids Res. **41:**7387-400.

238. **Marlor, R. L., S. M. Parkhurst, and V. G. Corces.** 1986. The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell Biol. **6:**1129-34.
239. **Marsh, M., and A. Helenius.** 1989. Virus entry into animal cells. Adv. Virus Res. **36:**107-51.
240. **Martens, S., and H. T. McMahon.** 2008. Mechanisms of membrane fusion: disparate players and common principles. Nat. Rev. Mol. Cell Bio. **9:**543-56.
241. **Martín, C. S.-S., C. Y. Liu, and M. Kielian.** 2009. Dealing with low pH: entry and exit of alphaviruses and flaviviruses. Trends Microbiol. **17:**514-21.
242. **Mas, V., S. Herfst, A. D. Osterhaus, R. A. Fouchier, and J. A. Melero.** 2011. Residues of the human metapneumovirus fusion (F) protein critical for its strain-related fusion phenotype: implications for the virus replication cycle. J. Virol. **85:**12650-61.
243. **Matlin, K. S., H. Reggio, A. Helenius, and K. Simons.** 1981. Infectious entry pathway of influenza virus in a canine kidney cell line. J. Cell Biol. **91:**601-13.
244. **McCarthy, C. B., X. J. Dai, C. Donly, and D. A. Theilmann.** 2008. *Autographa californica* multiple nucleopolyhedrovirus *ac142*, a core gene that is essential for BV production and ODV envelopment. Virology **372:**325-39.
245. **McLachlin, J. R., and L. K. Miller.** 1994. Identification and characterization of vlf-1, a baculovirus gene involved in very late gene expression. J. Virol. **68:**7746-56.
246. **McLellan, J. S., M. Chen, S. Leung, K. W. Graepel, X. Du, Y. Yang, T. Zhou, U. Baxa, E. Yasuda, T. Beaumont, A. Kumar, K. Modjarrad, Z. Zheng, M. Zhao, N. Xia, P. D. Kwong, and B. S. Graham.** 2013. Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. Science **340:**1113-7.
247. **McLellan, J. S., Y. Yang, B. S. Graham, and P. D. Kwong.** 2011. Structure of respiratory syncytial virus fusion glycoprotein in the postfusion conformation reveals preservation of neutralizing epitopes. J. Virol. **85:**7788-96.
248. **McReynolds, S., S. Jiang, L. Rong, and M. Caffrey.** 2009. Dynamics of SARS-coronavirus HR2 domain in the prefusion and transition states. J. Magn. Reson. **201:**218-21.
249. **Megy, K., S. J. Emrich, D. Lawson, D. Campbell, E. Dialynas, D. S. Hughes, G. Koscielny, C. Louis, R. M. Maccallum, S. N. Redmond, A. Sheehan, P. Topalis, D. Wilson, and C. VectorBase.** 2012. VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics. Nucleic Acids Res. **40:**D729-34.
250. **Melikyan, G. B., R. J. O. Barnard, L. G. Abrahamyan, W. Mothes, and J. A. T. Young.** 2005. Imaging individual retroviral fusion events: From hemifusion to pore formation and growth. Proc. Natl. Acad. Sci. U.S.A. **102:**8728-33.
251. **Melikyan, G. B., R. J. O. Barnard, R. M. Markosyan, J. A. T. Young, and F. S. Cohen.** 2004. Low pH is required for avian sarcoma and leukosis virus Env-induced hemifusion and fusion pore formation but not for pore growth. J. Virol. **78:**3753-62.

252. **Merrihew, R. V., W. C. Clay, J. P. Condreay, S. M. Witherspoon, W. S. Dallas, and T. A. Kost.** 2001. Chromosomal integration of transduced recombinant baculovirus DNA in mammalian cells. J. Virol. **75:**903-9.
253. **Mettenleiter, T. C., B. G. Klupp, and H. Granzow.** 2006. Herpesvirus assembly: a tale of two membranes. Curr. Opin. Microbiol. **9:**423-9.
254. **Mi, S., X. Lee, X. Li, G. M. Veldman, H. Finnerty, L. Racie, E. LaVallie, X. Y. Tang, P. Edouard, S. Howes, J. C. Keith Jr, and J. M. McCoy.** 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature **403:**785-9.
255. **Miele, S. A., M. J. Garavaglia, M. N. Belaich, and P. D. Ghiringhelli.** 2011. Baculovirus: molecular insights on their diversity and conservation. Int. J. Evol. Biol. **2011:**379424.
256. **Mikhailov, V. S., and G. F. Rohrmann.** 2002. Binding of the baculovirus very late expression factor 1 (VLF-1) to different DNA structures. BMC Mol. Biol. **3:**14.
257. **Miller, D. W., and L. K. Miller.** 1982. A virus mutant with an insertion of a *copia*-like transposable element. Nature **299:**562-4.
258. **Miller, D. W., and L. K. Miller.** 1982. A virus mutant with an insertion of a *copia*-like transposable element. Nature **299:**562-4.
259. **Miller, L. K.** 1997. The Baculoviruses, 1st ed. Plenum Press, New York, NY.
260. **Miller, L. K.** 1988. Baculoviruses for foreign gene expression in insect cells, p. 457-465, Biotechnology, vol. 10.
261. **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:**3325-31.
262. **Misseri, Y., G. Labesse, A. Bucheton, and C. Terzian.** 2003. Comparative sequence analysis and predictions for the envelope glycoproteins of insect endogenous retroviruses. Trends Microbiol. **11:**253-6.
263. **Mitra, K., I. Ubarretxena-Belandia, T. Taguchi, G. Warren, and D. M. Engelman.** 2004. Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. Proc. Natl. Acad. Sci. U.S.A. **101:**4083-8.
264. **Modis, Y., S. Ogata, D. Clements, and S. C. Harrison.** 2003. A ligand-binding pocket in the dengue virus envelope glycoprotein. Proc. Natl. Acad. Sci. U.S.A. **100:**6986-91.
265. **Modis, Y., S. Ogata, D. Clements, and S. C. Harrison.** 2004. Structure of the dengue virus envelope protein after membrane fusion. Nature **427:**313-9.
266. **Molloy, S. S., P. A. Bresnahan, S. H. Leppla, K. R. Klimpel, and G. Thomas.** 1992. Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen. J. Biol. Chem. **267:**16396-402.
267. **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:**640-6.
268. **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:**640-6.

269. **Moser, B. A., J. J. Becnel, S. E. 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:**283-97.

270. **Mueller, D. S., T. Kampmann, R. Yennamalli, P. R. Young, B. Kobe, and A. E. Mark.** 2008. Histidine protonation and the activation of viral fusion proteins. Biochem Soc Trans **36:**43-5.

271. **Mueller, J., J. Pfanzelter, C. Winkler, A. Narita, C. Le Clainche, M. Nemethova, M. F. Carlier, Y. Maeda, M. D. Welch, T. Ohkawa, C. Schmeiser, G. P. Resch, and J. V. Small.** 2014. Electron tomography and simulation of baculovirus actin comet tails support a tethered filament model of pathogen propulsion. PLoS Biol. **12:**e1001765.

272. **Mukhopadhyay, S., W. Zhang, S. Gabler, P. R. Chipman, E. G. Strauss, J. H. Strauss, T. S. Baker, R. J. Kuhn, and M. G. Rossmann.** 2006. Mapping the structure and function of the E1 and E2 glycoproteins in alphaviruses. Structure **14:**63-73.

273. **Nakayama, K.** 1997. Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. Biochem. J. **327:**625-35.

274. **Narayanan, K., A. Maeda, J. Maeda, and S. Makino.** 2000. Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. J. Virol. **74:**8127-34.

275. **Nethe, M., B. Berkhout, and A. C. van der Kuyl.** 2005. Retroviral superinfection resistance. Retrovirology **2:**52.

276. **Nettleship, J. E., R. Assenberg, J. M. Diprose, N. Rahman-Huq, and R. J. Owens.** 2010. Recent advances in the production of proteins in insect and mammalian cells for structural biology. J. Struct. Biol. **172:**55-65.

277. **Noda, T., H. Ebihara, Y. Muramoto, K. Fujii, A. Takada, H. Sagara, J. H. Kim, H. Kida, H. Feldmann, and Y. Kawaoka.** 2006. Assembly and budding of *Ebolavirus*. PLoS Pathog. **2:**e99.

278. **Nuttall, P. A., M. A. Morse, L. D. Jones, and A. Portela.** 1995. Adaptation of members of the Orthomyxoviridae family to transmission by ticks, p. 416-425. *In* A. J. Gibbs, C. H. Calisher, and F. García-Arenal (eds.), Molecular basis of virus evolution. Cambridge University Press, New York, N.Y.

279. **Obmolova, G., T. J. Malia, A. Teplyakov, R. Sweet, and G. L. Gilliland.** 2010. Promoting crystallization of antibody-antigen complexes via microseed matrix screening. Acta Crystallographica Section D: Biological Crystallography **66:**927-33.

280. **Ohkawa, T., L. E. Volkman, and M. D. Welch.** 2010. Actin-based motility drives baculovirus transit to the nucleus and cell surface. J. Cell Biol. **190:**187-95.

281. **Ohki, S., and K. Arnold.** 2012. Molecular mechanisms of membrane fusion, p. 673-707, *In* Electrical Phenomena at Interfaces and Biointerfaces. John Wiley & Sons, Inc.

282. **Okano, K., A. L. Vanarsdall, V. S. Mikhailov, and G. F. Rohrmann.** 2006. Conserved molecular systems of the *Baculoviridae*. Virology **344:**77-87.

283. **Oker-Blom, C., D. L. Jarvis, and M. D. Summers.** 1990. Translocation and cleavage of rubella virus envelope glycoproteins: identification and role of the E2 signal sequence. J. Gen. Virol. **71:**3047-53.

284. **Olszewski, J., and L. K. Miller.** 1997. Identification and characterization of a baculovirus structural protein, VP1054, required for nucleocapsid formation. J. Virol. **71:**5040-50.

285. **Olszewski, J., and L. K. Miller.** 1997. Identification and characterization of a baculovirus structural protein, VP1054, required for nucleocapsid formation. J. Virol. **71:**5040-50.

286. **Oomens, A. G., and G. W. Blissard.** 1999. Requirement for GP64 to drive efficient budding of *Autographa californica* multicapsid nucleopolyhedrovirus. Virology **254:**297-314.

287. **Oomens, A. G., S. A. Monsma, and G. W. Blissard.** 1995. The baculovirus GP64 envelope fusion protein: synthesis, oligomerization, and processing. Virology **209:**592-603.

288. **Ortega, M. E., H. Gaussier, and C. E. Catalano.** 2007. The DNA maturation domain of gpA, the DNA packaging motor protein of bacteriophage lambda, contains an ATPase site associated with endonuclease activity. J. Mol. Biol. **373:**851-65.

289. **Ozers, M. S., and P. D. Friesen.** 1996. The *Env*-like open reading frame of the baculovirus-integrated retrotransposon TED encodes a retrovirus-like envelope protein. Virology **226:**252-9.

290. **Ozers, M. S., and P. D. Friesen.** 1996. The *Env*-like open reading frame of the baculovirus-integrated retrotransposon TED encodes a retrovirus-like envelope protein. Virology **226:**252-9.

291. **Papanicolaou, A., S. Gebauer-Jung, M. L. Blaxter, W. Owen McMillan, and C. D. Jiggins.** 2008. ButterflyBase: a platform for lepidopteran genomics. Nucleic Acids Res. **36:**D582-7.

292. **Paroutis, P., N. Touret, and S. Grinstein.** 2004. The pH of the Secretory Pathway: Measurement, determinants, and regulation. Physiology **19:**207-15.

293. **Passarelli, A. L., and L. A. Guarino.** 2007. Baculovirus late and very late gene regulation. Curr. Drug Targets **8:**1103-15.

294. **Pearson, M. N., and G. F. Rohrmann.** 2006. Envelope gene capture and insect retrovirus evolution: The relationship between errantivirus and baculovirus envelope proteins. Virus Res. **118:**7-15.

295. **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:**5301-4.

296. **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:**81-8.

297. **Pearson, M. N., R. L. Russell, and G. F. Rohrmann.** 2002. Transcriptional mapping of two genes encoding baculovirus envelope-associated proteins. J. Gen. Virol. **83:**937-43.

298. **Peng, K., J. W. van Lent, S. Boeren, M. Fang, D. A. Theilmann, M. A. Erlandson, J. M. Vlak, and M. M. van Oers.** 2012. Characterization of novel components of the baculovirus *per os* infectivity factor complex. J. Virol. **86:**4981-8.

299. **Pertel, P. E., A. Fridberg, M. L. Parish, and P. G. Spear.** 2001. Cell fusion induced by herpes simplex virus glycoproteins gB, gD, and gH-gL requires a gD receptor but not necessarily heparan sulfate. Virology **279:**313-24.

300. **Petersen, T. N., S. Brunak, G. von Heijne, and H. Nielsen.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8:**785-6.
301. **Plattet, P., and R. K. Plemper.** 2013. Envelope protein dynamics in paramyxovirus entry. mBio **4:**e413-13.
302. **Plemper, R. K.** 2011. Cell entry of enveloped viruses. Curr. Opin. Virol. **1:**92-100.
303. **Plemper, R. K., M. A. Brindley, and R. M. Iorio.** 2011. Structural and mechanistic studies of measles virus illuminate paramyxovirus entry. PLoS Pathog **7:**e1002058.
304. **Plemper, R. K., A. S. Lakdawala, K. M. Gernert, J. P. Snyder, and R. W. Compans.** 2003. Structural features of paramyxovirus F protein required for fusion initiation. Biochemistry **42:**6645-55.
305. **Poole, B., and S. Ohkuma.** 1981. Effect of weak bases on the intralysosomal pH in mouse peritoneal macrophages. J. Cell Biol. **90:**665-9.
306. **Porotto, M., M. Murrell, O. Greengard, and A. Moscona.** 2003. Triggering of human parainfluenza virus 3 fusion protein (F) by the hemagglutinin-neuraminidase (HN) protein: an HN mutation diminishes the rate of F activation and fusion. J. Virol. **77:**3647-54.
307. **Powell, H. R.** 1999. The Rossmann Fourier autoindexing algorithm in MOSFLM. Acta Crystallographica Section D: Biological Crystallography **55:**1690-5.
308. **Puri, A., J. Winick, R. J. Lowy, D. Covell, O. Eidelman, A. Walter, and R. Blumenthal.** 1988. Activation of vesicular stomatitis virus fusion with cells by pretreatment at low pH. J. Biol.l Chem **263:**4749-53.
309. **Richard, J. P., E. Leikina, and L. V. Chernomordik.** 2009. Cytoskeleton reorganization in influenza hemagglutinin-initiated syncytium formation. Biochim. Biophys. Acta **1788:**450-7.
310. **Ritchie, K., and A. Kusumi.** 2004. Role of the membrane skeleton in creation of microdomains. Subcell. Biochem. **37:**233-45.
311. **Roche, S., S. Bressanelli, F. A. Rey, and Y. Gaudin.** 2006. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. Science **313:**187-91.
312. **Roche, S., and Y. Gaudin.** 2002. Characterization of the equilibrium between the native and fusion-Inactive conformation of rabies virus glycoprotein indicates that the fusion complex is made of several trimers. Virology **297:**128-135.
313. **Roche, S., F. A. Rey, Y. Gaudin, and S. Bressanelli.** 2007. Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. Science **315:**843-8.
314. **Rodems, S. M., and P. D. Friesen.** 1995. Transcriptional enhancer activity of hr5 requires dual-palindrome half sites that mediate binding of a dimeric form of the baculovirus transregulator IE1. J. Virol. **69:**5368-75.
315. **Rohrmann, G. F.** 2013. The baculovirus replication cycle: effects on cells and insects, *In* Baculovirus molecular biology, 3rd ed, Bethesda (MD): National Center for Biotechnology Information (US).
316. **Rohrmann, G. F.** 1992. Baculovirus structural proteins. J. Gen. Virol. **73:**749-61.
317. **Rohrmann, G. F.** 2013. Baculoviruses, retroviruses, DNA transposons (piggyBac), and insect cells, *In* Baculovirus molecular biology, 3rd ed, Bethesda (MD): National Center for Biotechnology Information (US).
318. **Rohrmann, G. F.** 1986. Polyhedrin structure. J. Gen. Virol. **67:**1499-513.

319. **Rohrmann, G. F., and P. A. Karplus.** 2001. Relatedness of baculovirus and gypsy retrotransposon envelope proteins. BMC Evol. Biol. **1:**1.

320. **Rohrmann, G. F., M. N. Pearson, T. J. Bailey, R. R. Becker, and G. S. Beaudreau.** 1981. N-terminal polyhedrin sequences and occluded baculovirus evolution. J. Mol. Evol. **17:**329-33.

321. **Ros, V. I. D., V. M. Fleming, E. J. Feil, and J. A. J. Breeuwer.** 2009. How diverse is the genus *Wolbachia*? Multiple-gene sequencing reveals a putatively new *Wolbachia* supergroup recovered from spider mites (Acari: Tetranychidae). Appl. Environ. Microb. **75:**1036-43.

322. **Rote, N. S., S. Chakrabarti, and B. P. Stetzer.** 2004. The role of human endogenous retroviruses in trophoblast differentiation and placental development. Placenta **25:**673-83.

323. **Ruigrok, R. W. H., A. Aitken, L. J. Calder, S. R. Martin, J. J. Skehel, S. A. Wharton, W. Weis, and D. C. Wiley.** 1988. Studies on the structure of the influenza virus haemagglutinin at the pH of membrane fusion. J. Gen. Virol. **69:**2785-95.

324. **Russell, C. J., T. S. Jardetzky, and R. A. Lamb.** 2001. Membrane fusion machines of paramyxoviruses: capture of intermediates of fusion. EMBO J. **20:**4024-34.

325. **Russell, R. L., C. J. Funk, and G. F. Rohrmann.** 1997. Association of a baculovirus-encoded protein with the capsid basal region. Virology **227:**142-52.

326. **Sapir, A., O. Avinoam, B. Podbilewicz, and L. V. Chernomordik.** 2008. Viral and developmental cell fusion mechanisms: conservation and divergence. Dev. Cell **14:**11-21.

327. **Schneider, T. R., and G. M. Sheldrick.** 2002. Substructure solution with SHELXD. Acta Crystallographica Section D: Biological Crystallography **58:**1772-9.

328. **Schowalter, R. M., A. Chang, J. G. Robach, U. J. Buchholz, and R. E. Dutch.** 2009. Low-pH triggering of human metapneumovirus fusion: essential residues and importance in entry. J. Virol. **83:**1511-22.

329. **Schowalter, R. M., A. Chang, J. G. Robach, U. J. Buchholz, and R. E. Dutch.** 2009. Low-pH triggering of human metapneumovirus fusion: Essential residues and importance in entry. J. Virol. **83:**1511-22.

330. **Schultz, K. L., and P. D. Friesen.** 2009. Baculovirus DNA replication-specific expression factors trigger apoptosis and shutoff of host protein synthesis during infection. J. Virol. **83:**11123-32.

331. **Shapiro, J., N. Sciaky, J. Lee, H. Bosshart, R. H. Angeletti, and J. S. Bonifacino.** 1997. Localization of endogenous furin in cultured cell lines. J. Histochem. Cytochem. **45:**3-12.

332. **Simón, O., T. Williams, M. López-Ferber, and P. Caballero.** 2004. Virus entry or the primary infection cycle are not the principal determinants of host specificity of *Spodoptera* spp. nucleopolyhedroviruses. J. Gen. Virol. **85:**2845-55.

333. **Singh, I., R. W. Doms, K. R. Wagner, and A. Helenius.** 1990. Intracellular transport of soluble and membrane-bound glycoproteins: folding, assembly and secretion of anchor-free influenza hemagglutinin. EMBO J. **9:**631-9.

334. **Sitia, R., and I. Braakman.** 2003. Quality control in the endoplasmic reticulum protein factory. Nature **426:**891-4.

335. **Skehel, J. J., P. M. Bayley, E. B. Brown, S. R. Martin, M. D. Waterfield, J. M. White, I. A. Wilson, and D. C. Wiley.** 1982. Changes in the conformation of influenza virus hemagglutinin at the pH optimum of virus-mediated membrane fusion. Proc. Natl. Acad. Sci. U.S.A. **79:**968-72.

336. **Skehel, J. J., and M. D. Waterfield.** 1975. Studies on the primary structure of the influenza virus hemagglutinin. Proc Natl Acad Sci U.S.A. **72:**93-7.

337. **Skehel, J. J., and D. C. Wiley.** 1998. Coiled coils in both intracellular vesicle and viral membrane fusion. Cell **95:**871-4.

338. **Skehel, J. J., and D. C. Wiley.** 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu. Rev. Biochem. **69:**531-69.

339. **Slack, J., and B. M. Arif.** 2007. The baculoviruses occlusion-derived virus: virion structure and function. Adv. Virus Res. **69:**99-165.

340. **Smith, E. C., M. R. Culler, L. M. Hellman, M. G. Fried, T. P. Creamer, and R. E. Dutch.** 2012. Beyond anchoring: the expanding role of the Hendra virus fusion protein transmembrane domain in protein folding, stability, and function. J. Virol. **86:**3003-13.

341. **Smith, E. C., S. E. Smith, J. R. Carter, S. R. Webb, K. M. Gibson, L. M. Hellman, M. G. Fried, and R. E. Dutch.** 2013. Trimeric transmembrane domain interactions in paramyxovirus fusion proteins: Roles in protein folding, stability and function. J. Biol. Chem. **288:**35726-35.

342. **Smith, G. E., M. J. Fraser, and M. D. Summers.** 1983. Molecular engineering of the *Autographa californica* nuclear polyhedrosis virus genome: Deletion mutations within the polyhedrin gene. J. Virol. **46:**584-93.

343. **Smits, P. H., M. van de Vrie, and J. M. Vlak.** 1987. Nuclear polyhedrosis virus for control of *Spodoptera exigua* larvae on glasshouse crops. Entomol. Exp. Appl. **43:**73-80.

344. **Soeda, E., T. Maruyama, J. R. Arrand, and B. E. Griffin.** 1980. Host-dependent evolution of three papova viruses. Nature **285:**165-7.

345. **Song, S. U., M. Kurkulos, J. D. Boeke, and V. G. Corces.** 1997. Infection of the germ line by retroviral particles produced in the follicle cells: A possible mechanism for the mobilization of the gypsy retroelement of *Drosophila*. Development **124:**2789-98.

346. **Sonnhammer, E. L., G. von Heijne, and A. Krogh.** 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. **6:**175-82.

347. **St Pierre, S. E., L. Ponting, R. Stefancsik, P. McQuilton, and C. FlyBase.** 2014. FlyBase 102--advanced approaches to interrogating FlyBase. Nucleic Acids Res. **42:**D780-8.

348. **Stadler, K., S. L. Allison, J. Schalich, and F. X. Heinz.** 1997. Proteolytic activation of tick-borne encephalitis virus by furin. J. Virol. **71:**8475-81.

349. **Stauffer, F., J. De Miranda, M. C. Schechter, F. A. Carneiro, L. T. Salgado, G. F. Machado, and A. T. Da Poian.** 2007. Inactivation of vesicular stomatitis virus through inhibition of membrane fusion by chemical modification of the viral glycoprotein. Antiviral Res. **73:**31-9.

350. **Stefanov, Y., V. Salenko, and I. Glukhov.** 2012. *Drosophila* errantiviruses. Mob. Genet. Elem. **2:**36-45.

351. **Stein, A., G. Weber, M. C. Wahl, and R. Jahn.** 2009. Helical extension of the neuronal SNARE complex into the membrane. Nature **460:**525-8.

352. **Stevens, J., A. L. Corper, C. F. Basler, J. K. Taubenberger, P. Palese, and I. A. Wilson.** 2004. Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. Science **303:**1866-70.

353. **Stiasny, K., S. L. Allison, C. W. Mandl, and F. X. Heinz.** 2001. Role of metastability and acidic pH in membrane fusion by tick-borne encephalitis virus. J. Virol. **75:**7392-98.

354. **Stiasny, K., and F. X. Heinz.** 2006. Flavivirus membrane fusion. J. Gen. Virol. **87:**2755-66.

355. **Stieneke-Gröber, A., M. Vey, H. Angliker, E. Shaw, G. Thomas, C. Roberts, H. D. Klenk, and W. Garten.** 1992. Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. EMBO J. **11:**2407-14.

356. **Stoye, J. P.** 2012. Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nat. Rev. Microbiol. **10:**395-406.

357. **Stura, E. A., and I. A. Wilson.** 1991. Applications of the streak seeding technique in protein crystallization. J. Crystal Growth **110:**270-82.

358. **Südhof, T. C.** 2013. Neurotransmitter release: the last millisecond in the life of a synaptic vesicle. Neuron **80:**675-90.

359. **Südhof, T. C., and J. Rizo.** 2011. Synaptic vesicle exocytosis. Cold Spring Harb. Perspect. Biol. **3:**a005637.

360. **Südhof, T. C., and J. E. Rothman.** 2009. Membrane fusion: Grappling with SNARE and SM proteins. Science **323:**474-7.

361. **Summers, M. D.** 1971. Electron microscopic observations on granulosis virus entry, uncoating and replication processes during infection of the midgut cells of *Trichoplusia ni*. J. Ultra. Mol. Struct. R. **35:**606-25.

362. **Summers, M. D., and L. E. Volkman.** 1976. Comparison of biophysical and morphological properties of occluded and extracellular nonoccluded baculovirus from in vivo and in vitro host systems. J. Virol. **17:**962-72.

363. **Supekar, V. M., C. Bruckmann, P. Ingallinella, E. Bianchi, A. Pessi, and A. Carfi.** 2004. Structure of a proteolytically resistant core from the severe acute respiratory syndrome coronavirus S2 fusion protein. Proc. Natl .Acad. Sci. U.S.A. **101:**17958-63.

364. **Suyama, M., D. Torrents, and P. Bork.** 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. **34:**W609-12.

365. **Swanson, K., X. Wen, G. P. Leser, R. G. Paterson, R. A. Lamb, and T. S. Jardetzky.** 2010. Structure of the Newcastle disease virus F protein in the post-fusion conformation. Virology **402:**372-9.

366. **Swanson, K. A., E. C. Settembre, C. A. Shaw, A. K. Dey, R. Rappuoli, C. W. Mandl, P. R. Dormitzer, and A. Carfi.** 2011. Structural basis for immunization with postfusion respiratory syncytial virus fusion F glycoprotein (RSV F) to elicit high neutralizing antibody titers. Proc. Natl. Acad. Sci. U.S.A. **108:**9619-24.

367. **Takeuchi, K., and R. A. Lamb.** 1994. Influenza virus M2 protein ion channel activity stabilizes the native form of fowl plague virus hemagglutinin during intracellular transport. J. Virol. **68:**911-9.

368. **Tan, Y., L. Jiang, M. Wang, F. Yin, F. Deng, M. Liu, Z. Hu, and H. Wang.** 2008. Mutagenesis and nuclear magnetic resonance analyses of the fusion peptide of *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus F protein. J. Virol. **82:**8138-48.

369. **Tang, X. D., Y. P. Xu, L. L. Yu, G. J. Lang, C. H. Tian, J. F. Zhao, and C. X. Zhang.** 2008. Characterization of a *Bombyx mori* nucleopolyhedrovirus with *Bmvp80* disruption. Virus Res. **138:**81-8.

370. **Tani, H., Y. Komoda, E. Matsuo, K. Suzuki, I. Hamamoto, T. Yamashita, K. Moriishi, K. Fujiyama, T. Kanto, N. Hayashi, A. Owsianka, A. H. Patel, M. A. Whitt, and Y. Matsuura.** 2007. Replication-competent recombinant vesicular stomatitis virus encoding hepatitis C virus envelope proteins. J. Virol. **81:**8601-12.

371. **Tani, H., M. Nishijima, H. Ushijima, T. Miyamura, and Y. Matsuura.** 2001. Characterization of cell-surface determinants important for baculovirus infection. Virology **279:**343-53.

372. **Terwilliger, T. C.** 2002. Rapid automatic NCS identification using heavy-atom substructures. Acta Crystallographica Section D: Biological Crystallography **58:**2213-5.

373. **Thiem, S. M., and L. K. Miller.** 1989. Identification, sequence, and transcriptional mapping of the major capsid protein gene of the baculovirus *Autographa californica* nuclear polyhedrosis virus. J. Virol. **63:**2008-18.

374. **Thorn, A., and G. M. Sheldrick.** 2011. ANODE: Anomalous and heavy-atom density calculation. J. Appl. Crystallogr. **44:**1285-7.

375. **Tina, K. G., R. Bhadra, and N. Srinivasan.** 2007. PIC: Protein Interactions Calculator. Nucleic Acids Res. **35:**W473-6.

376. **Tjia, S. T., G. M. zu Altenschildesche, and W. Doerfler.** 1983. *Autographa californica* nuclear polyhedrosis virus (AcNPV) DNA does not persist in mass cultures of mammalian cells. Virology **125:**107-17.

377. **Trombetta, E. S., and A. J. Parodi.** 2003. Quality control and protein folding in the secretory pathway. Annu. Rev. Cell Dev. Biol. **19:**649-76.

378. **Tsai, H.-H. G., C.-M. Chang, and J.-B. Lee.** 2014. Multi-step formation of a hemifusion diaphragm for vesicle fusion revealed by all-atom molecular dynamics simulations. BBA-Biomembranes **1838:**1529-35.

379. **Tweeten, K. A., L. A. Bulla, and R. A. Consigli.** 1980. Characterization of an extremely basic protein derived from granulosis virus nucleocapsids. J. Virol. **33:**866-76.

380. **Vagin, A., and A. Teplyakov.** 2010. Molecular replacement with MOLREP. Acta Crystallographica Section D: Biological Crystallography **66:**22-5.

381. **Van Oers, M. M., D. Malarme, J. M. P. Jore, and J. M. Vlak.** 1992. Expression of the *Autographa californica* nuclear polyhedrosis virus p10 gene: effect of polyhedrin gene expression. Arch. Virol. **123:**1-11.

382. **Van Oers, M. M., and J. M. Vlak.** 1997. The baculovirus 10-kDa protein. J. Invertebr. Pathol. **70:**1-17.

383. **Vanarsdall, A. L., K. Okano, and G. F. Rohrmann.** 2004. Characterization of a baculovirus with a deletion of vlf-1. Virology **326:**191-201.

384. **Vanarsdall, A. L., K. Okano, and G. F. Rohrmann.** 2006. Characterization of the role of very late expression factor 1 in baculovirus capsid structure and DNA processing. J. Virol. **80:**1724-33.

385. **Vanarsdall, A. L., M. N. Pearson, and G. F. Rohrmann.** 2007. Characterization of baculovirus constructs lacking either the Ac 101, Ac 142, or the Ac 144 open reading frame. Virology **367:**187-95.

386. **Vaney, M. C., and F. A. Rey.** 2011. Class II enveloped viruses. Cell. Microbiol. **13:**1451-9.

387. **Vialard, J. E., and C. D. Richardson.** 1993. The 1,629-nucleotide open reading frame located downstream of the *Autographa californica* nuclear polyhedrosis virus polyhedrin gene encodes a nucleocapsid-associated phosphoprotein. J. Virol. **67:**5859-66.

388. **Villar, E., and I. M. Barroso.** 2006. Role of sialic acid-containing molecules in paramyxovirus entry into the host cell: a minireview. Glycoconjugate J. **23:**5-17.

389. **Volchkov, V. E., H. Feldmann, V. A. Volchkova, and H. D. Klenk.** 1998. Processing of the Ebola virus glycoprotein by the proprotein convertase furin. Proc. Natl. Acad. Sci. U.S.A. **95:**5762-7.

390. **Volkman, L. E.** 1988. *Autographa californica* MNPV nucleocapsid assembly: inhibition by cytochalasin D. Virology **163:**547-53.

391. **Volkman, L. E., and P. A. Goldsmith.** 1983. In vitro survey of *Autographa californica* nuclear polyhedrosis virus interaction with nontarget vertebrate host cells. Appl. Environ. Microbiol. **45:**1085-93.

392. **Volkman, L. E., and P. A. Goldsmith.** 1985. Mechanism of neutralization of budded *Autographs californica* nuclear polyhedrosis virus by a monoclonal antibody: Inhibition of entry by adsorptive endocytosis. Virology **143:**185-95.

393. **Volkman, L. E., P. A. Goldsmith, R. T. Hess, and P. Faulkner.** 1984. Neutralization of budded *Autographa californica* NPV by a monoclonal antibody: identification of the target antigen. Virology **133:**354-62.

394. **Volkman, L. E., S. N. Talhouk, D. I. Oppenheimer, and C. A. Charlton.** 1992. Nuclear F-actin: a functional component of baculovirus-infected lepidopteran cells? J. Cell Sci. **103:**15-22.

395. **Wahlberg, J. M., W. A. Boere, and H. Garoff.** 1989. The heterodimeric association between the membrane proteins of Semliki Forest virus changes its sensitivity to low pH during virus maturation. J. Virol. **63:**4991-7.

396. **Walter, T. S., E. J. Mancini, J. Kadlec, S. C. Graham, R. Assenberg, J. Ren, S. Sainsbury, R. J. Owens, D. I. Stuart, J. M. Grimes, and K. Harlos.** 2007. Semi-automated microseeding of nanolitre crystallization experiments. Acta Crystallographica Section F: Structural Biology and Crystallization Communications **64:**14-8.

397. **Wang, H., A. Alminaite, A. Vaheri, and A. Plyusnin.** 2010. Interaction between hantaviral nucleocapsid protein and the cytoplasmic tail of surface glycoprotein Gn. Virus Res. **151:**205-12.

398. **Wang, M., Y. Tan, F. Yin, F. Deng, J. M. Vlak, Z. Hu, and H. Wang.** 2008. The F-like protein *Ac23* enhances the infectivity of the budded virus of *gp64*-null *Autographa californica* multinucleocapsid nucleopolyhedrovirus pseudotyped with baculovirus envelope fusion protein F. J. Virol. **82:**9800-4.

399. **Wang, M., E. Tuladhar, S. Shen, H. Wang, M. M. van Oers, J. M. Vlak, and M. Westenberg.** 2010. Specificity of baculovirus P6.9 basic DNA-binding proteins and critical role of the C terminus in virion formation. J. Virol. **84:**8821-8.

400. **Wang, M., J. Wang, F. Yin, Y. Tan, F. Deng, X. Chen, J. A. Jehle, J. M. Vlak, Z. Hu, and H. Wang.** 2014. Unraveling the entry mechanism of baculoviruses and evolutionary implications. J. Virol. **88:**2301-11.

401. **Wang, M., J. Wang, F. Yin, Y. Tan, F. Deng, X. Chen, J. A. Jehle, J. M. Vlak, Z. Hu, and H. Wang.** 2014. Unraveling the entry mechanism of baculoviruses and its evolutionary implications. J. Virol. **88:**2301-11.

402. **Wang, R., F. Deng, D. Hou, Y. Zhao, L. Guo, H. Wang, and Z. Hu.** 2010. Proteomics of the *Autographa californica* nucleopolyhedrovirus budded virions. J. Virol. **84:**7233-42.

403. **Wang, Y., Q. Wang, C. Liang, J. Song, N. Li, H. Shi, and X. Chen.** 2008. *Autographa californica* multiple nucleopolyhedrovirus nucleocapsid protein BV/ODV-C42 mediates the nuclear entry of P78/83. J. Virol. **82:**4554-61.

404. **Weissenhorn, W., A. Hinz, and Y. Gaudin.** 2007. Virus membrane fusion. FEBS Lett. **581:**2150-55.

405. **Welch, B. D., Y. Liu, C. A. Kors, G. P. Leser, T. S. Jardetzky, and R. A. Lamb.** 2012. Structure of the cleavage-activated prefusion form of the parainfluenza virus 5 fusion protein. Proc. Natl. Acad. Sci. U.S.A. **109:**16672-7.

406. **Welch, B. D., Y. Liu, C. A. Kors, G. P. Leser, T. S. Jardetzky, and R. A. Lamb.** 2012. Structure of the cleavage-activated prefusion form of the parainfluenza virus 5 fusion protein. Proc. Natl. Acad. Sci. U.S.A. **109:**16672-7.

407. **Wen, X., J. C. Krause, G. P. Leser, R. G. Cox, R. A. Lamb, J. V. Williams, J. E. Crowe, Jr., and T. S. Jardetzky.** 2012. Structure of the human metapneumovirus fusion protein with neutralizing antibody identifies a pneumovirus antigenic site. Nat. Struct. Mol. Biol. **19:**461-3.

408. **Westenberg, M.** 2004. Functional analysis of a novel baculovirus envelope fusion protein. PhD thesis (119 pp). Wageningen University, Wageningen, the Netherlands.

409. **Westenberg, M., P. Uijtdewilligen, and J. M. Vlak.** 2007. Baculovirus envelope fusion proteins F and GP64 exploit distinct receptors to gain entry into cultured insect cells. J. Gen. Virol. **88:**3302-06.

410. **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:**6946-54.

411. **Westenberg, M., H. Wang, W. F. J. 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.

412. **Weyer, U., S. Knight, and R. D. Possee.** 1990. Analysis of very late gene expression by *Autographa californica* nuclear polyhedrosis virus and the further development of multiple expression vectors. J. Gen. Virol. **71:**1525-34.

413. **White, J., A. Helenius, and M.-J. Gething.** 1982. Haemagglutinin of influenza virus expressed from a cloned gene promotes membrane fusion. Nature **300:**658-9.

414. **White, J. M.** 1992. Membrane fusion. Science **258:**917-24.

415. **White, J. M., S. E. Delos, M. Brecher, and K. Schornberg.** 2008. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. Crit. Rev. Biochem. Mol. Biol. **43:**189-219.

416. **Wickham, T. J., M. L. Shuler, D. A. Hammer, R. R. Granados, and H. A. Wood.** 1992. Equilibrium and kinetic analysis of *Autographa californica* nuclear polyhedrosis virus attachment to different insect cell lines. J. Gen.l Virol. **73:**3185-94.

417. **Wickner, W., and A. Haas.** 2000. Yeast homotypic vacuole fusion: a window on organelle trafficking mechanisms. Ann.. Rev. Biochem., **69:**247-75

418. **Wilson, I. A., J. J. Skehel, and D. C. Wiley.** 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature **289:**366-73.

419. **Wilson, M. E., and K. H. Price.** 1988. Association of *Autographa californica* nuclear polyhedrosis virus (AcMNPV) with the nuclear matrix. Virology **167:**233-41.

420. **Wu, C., and S. Wang.** 2012. A pH-sensitive heparin-binding sequence from baculovirus gp64 protein is important for binding to mammalian cells but not to Sf9 insect cells. J. Virol. **86:**484-91.

421. **Wu, W., T. Lin, L. Pan, M. Yu, Z. Li, Y. Pang, and K. Yang.** 2006. *Autographa californica* multiple nucleopolyhedrovirus nucleocapsid assembly is interrupted upon deletion of the *38K* gene. J. Virol. **80:**11475-85.

422. **Wu, W. B., H. Q. Liang, J. S. Kan, C. Liu, M. J. Yuan, C. Liang, K. Yang, and Y. Pang.** 2008. *Autographa californica* multiple nucleopolyhedrovirus 38K is a novel nucleocapsid protein that interacts with VP1054, VP39, VP80, and itself. J. Virol. **82:**12356-64.

423. **Wurdinger, T., M. H. Verheije, K. Broen, B. J. Bosch, B. J. Haijema, C. A. de Haan, V. W. van Beusechem, W. R. Gerritsen, and P. J. Rottier.** 2005. Soluble receptor-mediated targeting of mouse hepatitis coronavirus to the human epidermal growth factor receptor. J. Virol. **79:**15314-22.

424. **Wyatt, P. J.** 1998. Submicrometer particle sizing by multiangle light scattering following fractionation. J. Colloid Interf. Sci. **197:**9-20.

425. **Xu, H. E., H. H. Zhang, T. Xia, M. J. Han, Y. H. Shen, and Z. Zhang.** 2013. BmTEdb: a collective database of transposable elements in the silkworm genome. Database (Oxford) **2013:**bat055.

426. **Xu, R., and I. A. Wilson.** 2011. Structural characterization of an early fusion intermediate of influenza virus hemagglutinin. J. Virol. **85:**5172-82.

427. **Yin, F., M. Wang, Y. Tan, F. Deng, J. M. Vlak, Z. Hu, and H. Wang.** 2013. Betabaculovirus F proteins showed different efficiencies when rescuing the infectivity of gp64-null *Autographa californica* nucleopolyhedrovirus. Virology **436:**59-66.

428. **Yin, F., M. Wang, Y. Tan, F. Deng, J. M. Vlak, Z. Hu, and H. Wang.** 2008. A functional F analogue of *Autographa californica* nucleopolyhedrovirus GP64 from the *Agrotis segetum* granulovirus. J. Virol. **82:**8922-6.

429. **Yin, H. S., R. G. Paterson, X. Wen, R. A. Lamb, and T. S. Jardetzky.** 2005. Structure of the uncleaved ectodomain of the paramyxovirus (hPIV3) fusion protein. Proc. Natl. Acad. Sci. U.S.A. **102:**9288-93.

430. **Yin, H. S., X. Wen, R. G. Paterson, R. A. Lamb, and T. S. Jardetzky.** 2006. Structure of the parainfluenza virus 5 F protein in its metastable, prefusion conformation. Nature **439:**38-44.

431. **Young, G. R., M. J. Y. Ploquin, U. Eksmond, M. Wadwa, J. P. Stoye, and G. Kassiotis.** 2012. Negative selection by an endogenous retrovirus promotes a higher-avidity CD4+ T Cell response to retroviral Infection. PLoS Pathog **8:**e1002709.

432. **Yu, I. L., Y. C. Lin, J. H. Robinson, and O. Lung.** 2009. Transduction of vertebrate cells with *Spodoptera exigua* multiple nucleopolyhedrovirus F protein-pseudotyped *gp64*-null *Autographa californica* multiple nucleopolyhedrovirus. J. Gen. Virol. **90:**2282-7.

433. **Yu, I. M., W. Zhang, H. A. Holdaway, L. Li, V. A. Kostyuchenko, P. R. Chipman, R. J. Kuhn, M. G. Rossmann, and J. Chen.** 2008. Structure of the immature dengue virus at low pH primes proteolytic maturation. Science **319:**1834-7.

434. **Yu, S. J.** 1991. Insecticide resistance in the fall armyworm, *Spodoptera frugiperda* (J. E. Smith). Pestic. Biochem. Phys. **39:**84-91.

435. **Zeng, Q., M. A. Langereis, A. L. W. van Vliet, E. G. Huizinga, and R. J. de Groot.** 2008. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. Proc. Natl. Acad. Sci. U.S.A. **105:**9065-9.

436. **Zherebtsova, E. N., L. I. Strokovskaya, and A. P. Gudz-Gorban.** 1972. Subviral infectivity in nuclear polyhedrosis of the great wax moth (*Galleria mellonella* L.). Acta Virol. **16:**427-31.

437. **Zhou, J., and G. W. Blissard.** 2008. Identification of a GP64 subdomain involved in receptor binding by budded virions of the baculovirus *Autographica californica* multicapsid nucleopolyhedrovirus. J. Virol. **82:**4449-60.

438. **Zhou, J., and G. W. Blissard.** 2006. Mapping the conformational epitope of a neutralizing antibody (AcV1) directed against the AcMNPV GP64 protein. Virology **352:**427-37.

Summaries

SUMMARY

Baculoviruses are enveloped, double-stranded DNA viruses that are pathogenic predominantly to insects. Originally, baculoviruses have been utilized as biological control agents and more recently as vectors to express recombinant proteins. Insects, such as caterpillars, become infected by oral feeding on baculovirus-contaminated plants. Baculoviruses survive outside their host in the form of large so-called occlusion bodies (OBs), containing the occlusion-derived virus (ODV) form of the virus. The ODV infects epithelial cells in the midgut of the host insect. In infected cells, a budded virus (BV) form of the virus is first produced, that is then responsible for the systemic infection of the host. In a later phase infected cells start to make new ODVs that are occluded in OBs. In this thesis I have investigated baculovirus envelope fusion proteins, in short F proteins that are present on the surface of BVs and mediate BV entry into target insect cells.

BVs deliver their genomes into the cytoplasm of host cells by membrane fusion upon exposure to the acidic environment encountered in endosomes. This low-pH triggered membrane fusion is mediated by a viral envelope fusion protein. Such fusion proteins are associated with all enveloped viruses. In general, these proteins contain two hydrophobic motifs - the transmembrane domain (TM) and fusion peptide (FP) - that interact with the viral envelope and the cellular membrane, respectively. Thereby, they connect both membranes and exert a pulling force during fusion. In the pre-fusion stage, the fusion proteins are in a metastable energy state with their FP sequestered in the center of the protein. Fusion is activated by a trigger (receptor binding and/or acidic environment), which lowers the energy barrier and allows the transition of the fusion protein to a stable state of low energy. Upon fusion activation the FP is projected to the cellular membrane followed by further conformational changes in the fusion protein that bring the viral envelope and cellular membrane into close contact. During transition towards the post-fusion stage, the two membranes merge with each other to form a fusion pore allowing the release of the nucleocapsid (DNA-protein core) into the cytoplasm.

Two distinct envelope fusion proteins, GP64 and F, are identified in the baculovirus family. Both proteins have a large, glycosylated ectodomain and a C-terminal transmembrane domain, which anchors the protein in the viral envelope. In contrast to the GP64 protein, the ectodomain of the F protein possesses regions with repeats of seven amino acids (heptad repeat (HR) regions

adjacent to the FP and TM motifs. These types of HRs have a strong propensity to form coiled-coil structures with other F proteins to form a trimer. Besides sequence and structural divergence to GP64, the F protein requires proteolytic cleavage by the cellular furin protease at furin cleavage site (FCS) as a post-translational modification to prime its fusion function (Chapter 1). The membrane fusion process is driven by the structural changes of fusion protein. This proteolytic cleavage results in a small fragment F_2, which remain connected to the large F_1 fragment by a disulfide bridge. Therefore, it is important to determine the structures of baculovirus fusion proteins to understand the mechanism of baculovirus-cell fusion. The postfusion structure of GP64 has been solved by others before this research started. However, the knowledge on the structure of the baculovirus F protein during the fusion event was still lacking. This research is important also because since F proteins primarily occur with RNA viruses, while baculovirus is a DNA virus. In this dissertation I focused on the F protein of *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV) as a model for baculovirus F proteins.

My first aim was to study the structures of the SeMNPV-F ectodomain in its pre- and post-fusion stage using X-ray crystallography. To this end, the oligomerization and conformational changes of two soluble F ectodomains, the F ectodomain (sF) and an F ectodomain variant containing a mutated furin cleavage site (sF^{FCSmut}), were produced and biochemically characterized (Chapter 2). In the absence of a transmembrane domain, both proteins were secreted as monomers. The majority of the furin-cleaved sF protein aggregated during purification (as shown by size exclusion chromatography) possibly due to exposure of the hydrophobic FP. The mutated FCS in sF^{FCSmut} was not cleaved by furin, which prevented FP exposure and subsequent aggregation. Proteolytic cleavage of the sF^{FCSmut} at the mutated FCS was this case achieved by limited trypsin proteolysis. In addition, trypsin also cleaved at two other positions in the F ectodomain. The trypsin resistant core fragment (sF_2-sF_1') consisted of an intact F_2 connected to a truncated F_1 via a disulphide bond. The truncated F_1 subunit lacks HR2 and the triple Strep-tags at the C-terminus of F_1, artificially added for purification, as well as FP and the adjacent HR1 downstream of the mutated FCS. Subsequently it was demonstrated that this monomeric, trypsin-resistant sF_2-sF_1' fragment was able to undergo an irreversible conformational rearrangement upon acidic pH treatment towards a trimeric form, which was anticipated to resemble the transition of the full-length F protein from its pre- to post-fusion stage.

The biochemical characterization of the SeMNPV soluble F ectodomain (SeFe) paved the way for X-ray crystallography (performed in the lab of Prof. Felix Rey, Pasteur Institute, Paris). Diffracting crystals were obtained from the acid-induced, trimeric trypsin-resistant F ectodomain core (Chapter 3). The crystal structure showed a trimeric F ectodomain in its post-fusion conformation constituted of a globular 'head' domain and an elongated 'stalk' region. The head domain is mainly composed of β sheets and β strands whereas the stalk region is formed by α helices. The α helices in the stalk region comprised the HR regions. The fusion peptide, the HR1 and HR2 regions are missing in the crystal structure as a result of the trypsin proteolysis. To obtain more information of the full-length F ectodomain, an F variant, in which the hydrophobic residues of FP were replaced by hydrophilic ones (sF^{FPmut}) was expressed. Electron microscopy (EM) of the acidified, trimeric sF^{FPmut} - which was presumed to represent a post-fusion form of F - revealed a structure with a similar globular head but an extended stalk region. The additional electron dense region of the stalk region likely corresponds to the HR1 and HR2 domains (missing in the crystal structure of SeFe) that form a six-helix bundle typical for class I fusion proteins. The SeFe postfusion crystal structure was compared with other available protein structures in Protein Data Bank (PDB) using the Dali server, showing high similarity of the baculovirus F protein with paramyxovirus F proteins. Superposition of the postfusion structures of SeFe and the F protein of the paramyxovirus respiratory syncytial virus (RSV-F) confirmed that the proteins are very similar in structure and domain organization.

Previous evolutionary studies indicated that the baculovirus F genes are related to endogenous insect retrovirus envelope genes and some cellular genes found in Drosophila and a few mosquito species In Chapter 4 a baculovirus F homolog was identified in the mosquito *Anopheles darlingi* (*Culicidae*), named Ad-F. The Ad-F protein was found to share amino acid sequence and domain structure with baculovirus F proteins, and possible functional similarity, to baculovirus F proteins (fusion activity). Phylogenetic analysis of the Ad-*f* gene highlighted the evolutionary relatedness of Ad-*f* and baculovirus *f* genes. The Ad-F protein displays conservation of functional and structural elements found in the baculovirus F protein, including an N-terminal signal peptide, a furin cleavage site, coiled-coil domains, a large conserved domain of unknown function (DUF3609), and a transmembrane domain. The structure of Ad-F modeled by the Phyre 2 server showed significant similarity with the structure of the SeMNPV-F protein. Similar to baculovirus F proteins, the transiently expressed

Ad-F protein was able to be cleaved by cellular furin and incorporated as oligomers into a pseudotyped vesicular stomatitis virus (VSV) particle. The Ad-F protein appeared to facilitate cell-cell fusion after expression in baby hamster kidney cells. These results provide an indication that an ancestral baculovirus F gene may have been domesticated by a mosquito host and adapted for a cellular function. Future studies such as gene deletion or down regulation of Ad-*f* in *Anopheles darlingi* may help us to understand the role of Ad-F protein in the insect host.

In Chapter 5 the mechanism on how low pH affects protein conformational changes of SeMNPV-F was explored. The pH threshold for fusion activation (pH ≤ 5.5) is lower than the pKa value of the amino acid histidine (6.03), suggesting that at the fusion activation pH certain histidines are likely to be protonated. This would result in destabilization of the conformation of the F protein leading to membrane fusion. Six highly conserved histidine residues were identified among the F proteins of group II alphabaculoviruses. To study the role of these highly conserved histidine residues in Se-F protein, F mutants were generated, in which each conserved histidine was individually substituted by an alanine. The function of all the mutants was examined for low-pH triggered cell-cell fusion activity after transient expression of each histidine mutant in *Spodoptera frugiperda* 9 cells. The single substitutions of histidines in the F_2 subunit (H59 or H119) abolished syncytium formation, but did not abrogate entry of a pseudotyped VSV virus nor did they inhibit the low-pH triggered conformational change of the F ectodomain. Yet, the H59A and H119A substitutions appear to affect the postfusion structure of the F ectodomain. These observations suggest that these highly conserved histidine residues are not readily crucial in pH sensing to trigger conformational changes.

In Chapter 6 the baculovirus BV morphology was examined using cryo-EM in comparison with previously performed 'negative staining EM' on baculovirus BVs. Cryo-EM provides a better image of the authentic form of BVs. AcMNPV and SeMNPV were chosen for this analysis as representatives of group I and group II alphabaculoviruses, as they differ in respective envelope fusion protein, GP64 and F. BV particles of both viruses have a remarkable elongated, ovoid shape leaving a large, lateral space between the nucleocapsid and the viral envelope, which appears to be filled with unorganized content. In consistence with previous findings the nucleocapsid has a distinctive cap and base structure interacting with the envelope. In addition, an unusual thickness of the viral

envelope was observed as compared to cell membranes, a difference which may be explained by the possible presence of an inner viral-membrane-linked ubiquitin protein in BV membranes. The envelope proteins were densely clustered as spikes at the two polar ends of the virion. The majority of the envelope proteins are fusion proteins, based on previous proteomics of group I and group II alphabaculoviruses. The fusion proteins mediate virus-cell fusion during virus entry and are transported to plasma membrane before virus assembly. The clustering of the envelope proteins at the two polar ends of the virion therefore may suggest a perpendicular orientation to the cellular membrane during baculovirus BV entry and exit. In addition, the fusion proteins of AcMNPV and SeMNPV exhibit distinct morphologies, which is in agreement with the previous studies on the spike structure of group I and group II alphabaculoviruses (class III vs. class I) (Chapter 3). On the basis of the observations using cryo-EM, a new structural model of baculovirus budded virions was proposed. The model emphasizes the ovoid-shaped virion with polar distributed envelope proteins and the 'empty' lateral space between the nucleocapsid and envelope of the BVs. This is in contrast to the current BV model known for the rod-shaped virion, which is bulbous only at one end and only there densely spaced with fusion proteins.

The research described in this dissertation provides novel structural details of the low-pH triggered baculovirus F protein in relation to its fusion mechanism. The gained knowledge on the molecular structure of baculovirus F proteins and other F-like proteins sheds new light on the evolutionary relationship of the baculovirus F proteins with homologs found in in insects and mammalian viruses.

SAMENVATTING

Baculovirussen zijn dubbelstrengige DNA-virussen, waarbij een DNA-eiwitkern is omhuld door een envelop. Deze virussen zijn pathogeen voor insecten. Oorspronkelijk werden baculovirussen gebruikt als biologische bestrijdingsmiddelen en meer recent als onderdeel van een expressiesysteem voor recombinante eiwitten. Insecten, zoals rupsen, raken geïnfecteerd door het eten van met baculovirus-besmette planten. Baculovirussen 'overleven' buiten hun gastheer in de vorm van grote, zogenaamde insluitlichamen of 'occlusion bodies' (OB's) genoemd, die zogeheten 'OB-derived' virusdeeltjes (ODV's) bevatten. Een ODV infecteert epitheelcellen in de middendarm van de insectengastheer. In geïnfecteerde cellen worden eerst zogeheten 'budded' virussen (BV's) geproduceerd, die verantwoordelijk zijn voor de systemische infectie van de gastheer. In een latere fase produceren geïnfecteerde cellen nieuwe ODV's die weer worden ingesloten in OB's. In dit proefschrift heb ik baculovirus envelop-fusie-eiwitten onderzocht, in het kort F-eiwitten genoemd, die aan het oppervlak van BV's aanwezig zijn en het binnentreden van de BV's in cellen faciliteren.

Na binnentreden wordt het BV blootgesteld aan de zure omgeving in endosomen en daardoor fuseert de envelop van het BV met het celmembraan, waarmee het virale genoom wordt afgeleverd in het cytoplasma van gastheercellen. Deze door lage pH teweeg gebrachte membraanfusie wordt uitgevoerd door een fusie-eiwit op de virusenvelop. Alle door een envelop omhulde virussen bevatten fusie-eiwitten. In het algemeen hebben deze eiwitten twee hydrofobe gebieden, het transmembraandomein (TM) en het fusiepeptide (FP), die interacties aangaan met respectievelijk de virale envelop en het celmembraan. Daarbij verbinden ze beide membranen en oefenen een trekkracht uit tijdens de fusie. In de prefusiefase bevinden de fusie-eiwitten zich in een metastabiele energietoestand, waarbij het FP zich in het midden van het eiwit bevindt en een vooruitgeschoven positie heeft. Fusie wordt geactiveerd door een externe impuls (receptorbinding en / of een lage zuurgraad), waardoor de energiebarrière verlaagd wordt en de overgang van het fusie-eiwit naar een stabiele toestand van lage energie mogelijk wordt maakt. Bij fusie-activering wordt het FP geprojecteerd op het celmembraan, gevolgd door verdere conformationele veranderingen in het fusie-eiwit, een proces waarbij de virale envelop en het celmembraan in nauw onderling contact worden gebracht. Tijdens de overgang naar de postfusiefase fuseren de twee membranen met elkaar en vormen een fusieporie, waardoor het nucleocapside (DNA-eiwit structuur) het cytoplasma kan ingaan.

Bij leden van de baculovirusfamilie zijn twee verschillende fusie-eiwitten, GP64 en F, geïdentificeerd op de BV envelop. Beide eiwitten hebben een groot, van suikers voorzien ectodomein en een C-terminaal TM-domein, dat het eiwit in de virale envelop verankert. In tegenstelling tot het GP64-eiwit bevat het ectodomein van het F-eiwit twee regio's met herhalingen van zeven aminozuren ('heptad repeats', HR's) in het ectodomein, dat grenst aan het FP en de TM motieven. Deze HR's hebben een sterke neiging om 'coiled-coil'-structuren te vormen met soortgelijke F-eiwitten, waardoor deze trimeren vormen. Naast verschillen in aminozuurvolgorde en in structuur tussen het F-eiwit en GP64, heeft het F-eiwit ook klieving nodig door een cellulair protease, furine genaamd. Deze modificatie na afloop van de synthese zorgt ervoor dat het F-eiwit fusie-activiteit krijgt (Hoofdstuk 1). Bij deze furinesplitsing ontstaat een klein fragment F_2, dat via een zwavelbrug is verbonden met het grote fragment van F (F_1). Het membraanfusieproces wordt gedreven door structurele veranderingen van het F-eiwit. Om het mechanisme van baculovirus-celfusie te kunnen begrijpen, is het nodig om de moleculaire structuur van baculovirusfusie-eiwitten te achterhalen. De postfusiestructuur van GP64 is opgehelderd door andere onderzoekers voordat mijn onderzoek begon. Kennis over de 3D-structuur van het baculovirus-F-eiwit ontbrak echter nog. Dit onderzoek is ook van belang, omdat homologe F-eiwitten verder alleen voorkomen bij RNA-bevattende virussen, terwijl het baculovirus een DNA-virus is. In dit proefschrift heb ik me gericht op het F-eiwit van het *Spodoptera exigua* kernpolyedervirus *(S. exigua* multiple nucleopolyhedrovirus, SeMNPV) als model voor baculovirus F-eiwitten.

Mijn eerste doel was om de structuur van het SeMNPV-F ectodomein te bestuderen in de pre- en postfusiefase via X-ray kristallografie. Hiertoe werden de oligomerisatie en conformatieveranderingen van twee oplosbare F-ectodomeinen biochemisch gekarakteriseerd, het normale F-ectodomein (SF) en een variant die een gemuteerde furineklievingsplaats (FCS) bevatte (sF^{FCSmut}), (Hoofdstuk 2). Beide eiwitten werden door cellen tot uitscheiding gebracht als monomeren door afwezigheid van een TM-domein. Het merendeel van het door furine gekliefde sF-eiwit aggregeerde tijdens de zuivering (zoals aangetoond via 'size exclusion chromatography'), mogelijk als gevolg van de vooruitgeschoven positie van het hydrofobe FP. De gemuteerde FCS in sF^{FCSmut} werd niet gekliefd door furine, waardoor het FP niet vooruitgeschoven kon worden en aggregatie werd voorkomen. Proteolytische klieving van sF^{FCSmut} in het gemuteerde FCS werd in dit geval bereikt door beperkte proteolyse via trypsine. Daarnaast klieft trypsine ook op twee andere plaatsen in het F-ectodomein. Het trypsineresistente

kernfragment (sF_2-sF_1') bestaat uit een intact F_2, dat verbonden is met een afgeknot F_1 via een zwavelbrug. De afgeknotte F_1-subeenheid mist HR2 en de toegevoegde drievoudige Strep-tags aan de C-terminus van F_1 (die was toegevoegd voor zuiveringsdoeleinden). Daarnaast ontbreken ook het FP en de aangrenzende HR1 regio, stroomafwaarts gelegen van de gemuteerde FCS. Vervolgens is aangetoond dat dit monomere, trypsine-resistente sF_2-sF_1' fragment een onomkeerbare conformationele verandering kan ondergaan naar een trimere vorm, waarvan wordt verondersteld dat deze de overgang van het F-eiwit van de pre- naar postfusiefase vertegenwoordigt.

De biochemische karakterisering van het oplosbare F-ectodomein van SeMNPV heeft de weg vrijgemaakt voor X-ray (röntgen)kristallografie, dat vervolgens is uitgevoerd in het laboratorium van dr. Felix Rey, Institute Pasteur, Parijs). Van de door zuurgeïnduceerde, trimere trypsine-resistente F-ectodomeinen werden kristallen gevormd, geschikt voor röntgendiffractie (Hoofdstuk 3). De kristalstructuur toonde een F-ectodomein in trimeervorm in zijn postfusieconformatie, bestaande uit een bolvormig 'hoofd'-domein en een langgerekte 'steel'-regio. Het 'hoofd'-domein bestaat hoofdzakelijk uit 'β-sheets' en β-strengen, terwijl de 'steel'-regio wordt gevormd door α-helices. De α-helices in de 'steel'-regio omvatten de HR-regio's. Het fusiepeptide en de HR1- en HR2-gebieden ontbreken in deze kristalstructuur als gevolg van de trypsinesplitsing. Om meer informatie te verkrijgen over het F ectodomein in volledige lengte werd een F-variant tot expressie gebracht, waarin de hydrofobe residuen van het FP zijn vervangen door hydrofiele residuen (sF^{FPmut}). Elektronenmicroscopie (EM) van de aangezuurde, trimere sF^{FPmut}, waarvan werd verondersteld dat deze postfusie vorm van F vertegenwoordigt, toonde een structuur met een soortgelijke bolvormige kop, maar met een langere 'steel'-regio. Het extra elektronendichte gebied van de 'steel'-regio komt waarschijnlijk overeen met de HR1- en HR2-domeinen, maar ontbreken in de uiteindelijke kristalstructuur van SeFe. Deze HR-domeinen vormen een 'zes-helix'-bundel, die typerend is voor klasse I fusie-eiwitten. De SeFe postfusiekristalstructuur is vergeleken met andere beschikbare eiwitstructuren in de Protein Data Bank (PDB) met behulp van de Dali-server. Deze analyse toonde hoge gelijkenis aan tussen het baculovirus-F-eiwit en paramyxovirus F-eiwitten. Superpositie van de postfusiestructuren van SeFe en het F-eiwit van het paramyxovirus respiratoir syncytieel virus (RSV-F) bevestigde dat de eiwitten qua moleculaire structuur en domeinorganisatie veel gelijkenis vertonen.

Eerdere evolutionaire studies gaven aan dat de baculovirus-F-genen gerelateerd zijn aan de envelopgenen, die bij endogene insectenretrovirussen worden aangetroffen en aan enkele cellulaire genen, gevonden in *Drosophila* en enkele muggensoorten. In Hoofdstuk 4 is een baculovirus F-homoloog geïdentificeerd in de mug *Anopheles darlingi* (Culicidae), genaamd Ad-F. Het Ad-F-eiwit bleek homoloog te zijn in aminozuurvolgorde en domeinstructuur met baculovirus F-eiwitten, en heeft mogelijk een overeenkomstige functie (fusie-activiteit). Fylogenetische analyse van het Ad-F-gen liet de evolutionaire verwantschap zien van Ad-F en baculovirus-F-genen. De Ad-F-eiwit toont conservering van functionele en structurele elementen van het baculovirus F-eiwit, waaronder een furineklievingsplaats, een N-terminaal gelegen signaalpeptide in F_1, 'coiled-coil'-domeinen, een TM-domein en een groot geconserveerd domein met onbekende functie (DUF3609). De structuur van Ad-F, gemodelleerd door de Phyre 2-server, heeft significante gelijkenis met de structuur van het SeMNPV-F-eiwit. Net zoals baculovirus-F-eiwitten kon het transiënt tot expressie gebrachte Ad-F-eiwit gekliefd worden door cellulair furine en opgenomen worden als oligomeren in een pseudo 'vesiculair stomatitis virus' (VSV) deeltje. Het Ad-F-eiwit faciliteerde cel-celfusie na expressie in babyhamsteriercellen. Deze resultaten leveren een aanwijzing dat een voorouderlijke baculovirus-F-gen kan zijn gedomesticeerd door een muggengastheer en kan zijn aangepast voor een cellulaire functie. Verdere studies zoals gendeletie of downregulering van Ad-F in *A. darlingi* kan helpen om de rol van Ad-F-eiwit in de insectengastheer te begrijpen.

In Hoofdstuk 5 is onderzocht hoe een lage pH een conformatieverandering van het SeMNPVF-eiwit tot stand brengt. De pH-drempelwaarde voor fusie-activatie ($pH \leq 5{,}5$) is lager dan de pKa-waarde van het aminozuur histidine (6,03), hetgeen suggereert dat bij pH-geïnduceerde fusie-activatie bepaalde histidines kunnen worden geprotoneerd (toevoeging van waterstof-ion). Dit zou resulteren in destabilisatie van de conformatie van het F-eiwit, wat dan leidt tot membraanfusie. Zes sterk geconserveerde histidineresiduen werden geïdentificeerd in de F-eiwitten van groep II alphabaculovirussen. Om de rol van deze sterk geconserveerde histidineresiduen in Se-F te bestuderen, werden F-mutanten gegenereerd, waarin elke geconserveerde histidine individueel werd vervangen door een alanine. Deze mutanten werden onderzocht op lage pH-geactiveerde cel-celfusie-activiteit na transiënte expressie van de histidinemutanten in *Spodoptera frugiperda*-9 cellen. De vervangingen van individuele histidines in de F_2-subeenheid (H59 of H119) verstoorde de

syncytiumvorming, maar niet de opname van F in de envelop van VSV, noch remden deze vervangingen de door lage pH veroorzaakte conformationele verandering van het F-ectodomein. Toch lijken de H59A- en H119A-substituties de postfusiestructuur van het F-ectodomein te beïnvloeden. Deze waarnemingen suggereren dat deze twee geconserveerde histidineresiduen niet al te cruciaal zijn voor het reageren op een lage pH om conformationele veranderingen teweeg te brengen.

In Hoofdstuk 6 is de morfologie onderzocht van het baculovirus BV als deeltje met behulp van cryo-electronenmicrosopie (cryo-EM) ter vergelijking met eerder uitgevoerde EM-studies, waarbij steeds negatief-gecontrasteerde BV deeltjes werden bekeken. Cryo-EM geeft een veel beter beeld van de authentieke vorm van deeltjes. Voor deze analyse werden AcMNPV en SeMNPV gekozen als vertegenwoordigers van groep I en II alphabaculovirussen, die onderling verschillen in hun respectievelijke fusie-eiwitten (GP64 en F) op het BV. BV-deeltjes van beide virussen hebben een opmerkelijke, langwerpige, ovale vorm, waarbij een grote ruimte lijkt te bestaan aan de laterale zijde van het deeltje tussen het nucleocapside en de virale envelop. Deze ruimte lijkt te zijn gevuld met ongestructureerde inhoud. In overeenstemming met eerdere bevindingen heeft het nucleocapside kenmerkende uiteinden met een basisstructuur die contact heeft met de envelop. Daarnaast werd waargenomen dat de virale envelop 'dikker' is dan het celmembraan. Dit zou kunnen worden verklaard door de aanwezigheid van een viraal-gecodeerd membraangebonden ubiquitine-eiwit in de BV-envelop. Bij de twee uiteinden van het virion werden uitsteeksels gevonden, waarvan de meeste hoogstwaarschijnlijk worden gevormd door fusie-eiwitten. Dit werd geconcludeerd op basis van eerdere proteomics-studies van groep I en groep II alphabaculoviruses. De fusie-eiwitten mediëren virus-cel-fusie tijdens het binnentreden van het virus in de cel en worden getransporteerd naar het celmembraan voordat virusdeeltjes gevormd worden. De clustering van de envelopeiwitten bij de twee polaire uiteinden van het virion suggereert daarom een loodrechte oriëntatie op de celmembraan tijdens het in- en uittreden van baculovirus BV via cellen. Daarnaast vertonen de fusie-eiwitten van SeMNPV en AcMNPV een verschillende morfologie, die in overeenstemming lijkt te zijn met eerdere studies naar de structuur van fusie-eiwitten van groep I en II alphabaculoviruses (klasse III vs. klasse I) (Hoofdstuk 3). Op basis van de waarnemingen via cryo-EM, is een nieuw model voor baculovirus BV's voorgesteld. Het model benadrukt de eivormige-structuur van het virusdeeltje met polair-verdeelde envelopeiwitten en de 'lege' laterale ruimte tussen het

nucleocapside en de envelop van het BV. Dit is in afwijking van het tot nu toe gebruikte BV-model voor staafvormige virions, welke bolvormig zijn aan één einde met alleen daar een kleine afstand tot het fusie-eiwit.

Het in dit proefschrift beschreven onderzoek heeft nieuwe details opgeleverd over de rol van een lage pH bij het fusiemechanisme van baculovirus-F-eiwitten. De opgedane kennis over de moleculaire structuur van baculovirus F-eiwit en andere, F-achtige eiwitten werpt nieuw licht op de evolutionaire relatie van baculovirus-F-eiwitten met homologen, gevonden in insecten en in zoogdiervirussen.

摘要

杆状病毒是一类有囊膜的双链DNA病毒， 主要感染昆虫。最初，杆状病毒被用作生物杀虫剂，最近用于重组蛋白质的表达载体。毛虫等昆虫在吸食了被杆状病毒感染的植物而被感染。杆状病毒以包涵体（OB）的形式宿主体外生存，含有包涵体病毒（ODV）形式的病毒。ODV感染宿主昆虫的中肠上皮细胞。在感染细胞中，首先产生出芽病毒（BV）形式的病毒 - 即负责寄主的系统性感染。在稍后阶段被感染的细胞开始产生新一批被包裹在OB中的ODV。本论文研究了杆状病毒包膜融合蛋白，简称F蛋白，它分布在BV的表面，介导BV进入靶细胞。

BV 在内吞体中的酸性条件下，通过自身的包膜和内吞体膜融合把其基因组释放到宿主细胞的细胞质中。这种低 pH 值诱发膜融合是由病毒包膜上的融合蛋白介导的。这种融合蛋白存在于所有包膜病毒中。通常，这些蛋白含有两个疏水基序 - 跨膜结构域（TM）和融合肽（FP），它们分别与病毒包膜和细胞膜相互作用才能够在融合过程中连接两膜并施加拉力。在融合前的阶段，融合蛋白处于亚稳的能量状态， FP 隐藏在融合蛋白的中心。融合由受体结合， 酸性 pH 或两者共同激活，这降低了能量势的障碍从而达到融合蛋白的低能量的稳定状态。融合激活时 FP 被发射到细胞膜随后通过融合蛋白进一步的构象变化使得病毒包膜和细胞膜紧密接触。在向融合后阶段过渡的过程中，两个膜合并而形成融合孔从而能够将核衣壳（核酸－蛋白质中心）释放到细胞质中。

在杆状病毒中鉴定出GP64和F这两种截然不同的包膜融合蛋白。这两种蛋白具有大型的糖基化的胞外结构域和把蛋白固定在病毒包膜的C端跨膜结构域。与GP64蛋白不同，F蛋白的胞外域具有七肽重复序列（HR）， 与FP和TM域相邻，具有与其他F蛋白形成卷曲螺旋结构的强烈趋势， 从而形成三聚体。与GP64除了在序列和结构方面不同以外，F蛋白在被翻译后需要经过细胞里的furin蛋白酶在furin酶切位点（FCS）的酶切，以保证其融合功能（第1章）。由这个酶切得到一个小片段F2和大片段F1, 两个片段由二硫键连接。由于膜融合过程由融合蛋白的结构改变来驱动，因此为了解杆状病毒细胞融合的机制， 确定杆状病毒融合蛋白的结构是很重要的。 GP64的融合后结构在本研究开始之前就已经被破解了。但是，关于杆状病毒F蛋白的融合过程的结构的知识仍不清楚。此项研究重要也是因为F蛋白主要存在于RNA病毒，而同样具

有F蛋白的杆状病毒却是一种DNA病毒。本论文把重点放在一个典型杆状病毒F蛋白－斜纹夜多蛾核型病毒（SeMNPV）的F蛋白上。

我的原有目标是利用X射线晶体学技术研究SeMNPV的F胞外域在融合前后的结构。为此，我得到并分析了F胞外域（sF）和含有一个突变的furin蛋白酶切割位点（sF^{FCSmut}）一个F胞外域变体这两种可溶性F胞外结构域的聚合和构象的变化（第2章）。在缺少跨膜结构域的情况下，这两种蛋白被分泌为单体。多数经过furin蛋白酶酶切的sF蛋白在纯化时高度聚集（正如尺寸排阻色谱法所展示）这可能是由于疏水性的FP的被暴露在外。突变的FCS中的sF^{FCSmut}不再被furin蛋白酶酶切，从而防止FP暴露和随后的高度聚合。sF^{FCSmut}中突变的FCS却可被限量胰蛋白酶酶切。此外，胰蛋白酶还酶切了F胞外区其他两个位置。胰蛋白酶抗性核心片段（sF_2-sF_1'）组成经二硫键连接一个完整的F_2和截短的F_1。截短的F_1亚基缺失了HR2和在C端的F_1人为添加三个用于纯化的Strep-tag标记，以及FP和与突变的FCS相邻的下游HR1。随后，这种单体的胰蛋白酶抗性sF_2-sF_1'片段被证明是能够通过酸性pH处理不可逆地重排成一个三聚体形式构象，这个过程是被认为与全长F蛋白的融合相似。

对SeMNPV的可溶性F胞外区（SeFe）生化特性的测定为其X射线晶体解析（在巴黎巴斯德研究所Felix Rey教授领导的实验室中进行）铺平了道路。衍射的晶体是由酸诱导处理过的三聚胰蛋白酶抗性F胞外域核心（第3章）形成的。该晶体结构呈现了一个由球状“头”区和一个细长的“茎”区构成的融合后F胞外域三聚体结构。头区主要由β折叠和β链，而茎区由α螺旋形成。在茎区的α螺旋由HR域组成。该结构缺少融合肽，HR1和HR2区域乃是胰蛋白酶酶切的结果。为了获得全长F胞外域的更多信息，需要表达一个FP的疏水残基被替换为亲水性的F突变体（sF^{FPmut}）被表达。 这个酸化，三聚sF^{FPmut}被推定为代表的F融合后形式。电子显微镜 （EM）揭示出其结构具有类似的球状头但延长的茎区。茎区的附加电子致密区域可能恰恰对应于形成 I类融合蛋白中典型六螺旋束的HR1和HR2结构域（在晶体结构中缺失）。通过使用Dali服务器把SeFe融合后的晶体结构与在蛋白质数据库（PDB）其他可用的蛋白质结构进行比较，显示出杆状病毒F蛋白的与副粘病毒的F蛋白高度相似性。叠加SeFe和副粘病毒呼吸道合胞病毒F蛋白（RSV-F）的融合后结构证实了它们的整体结构及结构域排列上都非常相似。

以往的有关进化的研究表明杆状病毒的F基因与内源性昆虫逆转录病毒包膜基因以及在果蝇和蚊发现一些细胞基因相关。在第4章杆状病毒的F同源物在蚊子*Anopheles darlingi*（蚊科），命名为Ad-F。所述的Ad-F蛋白被发现具有与杆状病毒的F蛋白的氨基酸序列和结构的同源性并且在功能（融合性）上也可能具有相似性。Ad-*f*基因的系统发育分析突出了Ad-*f*和杆状病毒*f*基因在进化上相关。所述的Ad-F蛋白保留了在杆状病毒F蛋白上发现的功能和结构元件，包括N端信号肽，一个furin蛋白酶酶切位点，卷曲螺旋结构域，功能未知的保守结构域（DUF3609），和一跨膜结构域。由Phyre 2服务器生成的Ad-F模型结构与SeMNPV-F蛋白的结构显示出显著相似度。与杆状病毒F蛋白相似，在瞬时表达的Ad-F蛋白能够被细胞内furin蛋白酶酶切并以低聚物的形式与假水泡性口炎病毒（VSV）结合。Ad-F蛋白似乎可以在幼仓鼠肾细胞表达后促进细胞和细胞之间的融合。这些结果首次提供的证据表明，一个祖先杆状病毒*f*基因可能已经被内化到宿主蚊子中并具有其一种细胞功能。未来的研究，如在*Anopheles darlingi*体内缺失或下调Ad-*f*基因可能有助于我们理解Ad-F蛋白在昆虫宿主中的作用。

第5章对SeMNPV-F蛋白的低pH值触发机制进行了探讨。融合激活的pH阈值（pH为≤5.5）比氨基酸组氨酸（6.03）的pKa低，这表明在融合活化pH值某些组氨酸有可能被质子化。这将导致的F蛋白融合的构象不稳定， 导致膜融合。在组II alphabaculovirus的F蛋白中确定了六个高度保守的组氨酸残基。为了研究这些高度保守的组氨酸残基在SeMNPV-F蛋白的作用，我们生成了六种F单突变体，其中，每个保守的组氨酸由丙氨酸分别单独取代。我们检查了所有各组氨酸单突变体在Sf-9细胞瞬时表达后由低pH值的引发的细胞与细胞融合活性。单独取代F_2亚基中的组氨酸（H59或H119）废除了合胞体形成，但并没有废除假型的VSV病毒的入侵，也没有抑制所述低pH值触发在F胞外域的构象变化。然而，取代H59A和H119A似乎影响了F胞外域融合后的结构。这些观察表明，这些高度保守的组氨酸残基在感受pH值而触发的蛋白构象变化上并不关键 。

第6章通过Cryo-EM对杆状病毒BV形态进行观察，并与先前采用的负染EM电镜进行了比较。Cryo-EM能够更好地呈现BV的真实的结构。AcMNPV的和SeMNPV被选为分析组I和II组alphabaculovirus的代表。两种病毒的BV颗粒呈现显著的细长卵形形状， 在核衣壳和病毒包膜之间

有巨大的侧面空间，其中似乎充满无组织的内含物。与先前的调查结果一致， 核衣壳有着鲜明的上盖和底座结构并与包膜相互作用。此外，所观察到的该病毒包膜的不寻常的厚度可能与内病毒膜联泛素蛋白有关。包膜蛋白密集地聚集在病毒粒子的两端。基于以前的蛋白质组学对组I和组II alphabaculovirus的分析，大多数的包膜蛋白为融合蛋白。融合蛋白在病毒进入的过程中介导病毒和细胞融合并且在病毒装配之前被运输到质膜。包膜蛋白聚集在病毒体的两个极端也因此暗示了杆状病毒BV以垂直细胞膜的方向进出。另外，AcMNPV和SeMNPV的融合蛋白表现出不同的形态，这与以往关于I和II组alphabaculovirus（III类与I类）的融合蛋白的结构研究结果是一致的 （第3章）。在使用冷冻电镜观察的基础上，我们提出了杆状病毒出芽病毒体的新结构模型。该模型强调的卵形形状的病毒粒子具有极性分布包膜蛋白以及BV核衣壳和包膜之间“空”的侧面空间，与以往所知BV模型中所描述的一端呈球形并且其上紧密排列着融合蛋白的棒状病毒体不同。

本论文中描述的研究提供了杆状病毒 F 蛋白在低 pH 值触发的融合机制下新的结构方面的细节。所获得的结构方面的信息揭示了杆状病毒 F 蛋白与在哺乳动物病毒及昆虫中发现的同系物的进化关系。

Acknowledgements

During the past five years of my PhD study I met a number of people who raised me up in my research and life. My advancement would never happen without them. Now I have reached the end of this extraordinary research experience with success. Here I take the opportunity to acknowledge all these people.

First, I acknowledge my promoter Prof. Just M. Vlak. Dear Just, you trusted my potential and gave me the opportunity to work on the joint research project between your laboratory and Prof. Peter J. M. Rottier's lab in Utrecht. Thank you for sending me to Utrecht to get new expertise and pursue this PhD research in the past four years. Every time when I came to your lab for discussion you always gave me warm welcome. I still remember that at the beginning of my PhD research you paid a visit to Prof. Rottier to check how I was doing in the new lab and backed me up mentally. In the final phase of my thesis you have been very efficient with coordination of my supervisory team and thesis planning. Once you said to me "Qiushi, I always say that you are my 'crystal' lady." Such words always lingered in my ears and gave me so much strength whenever my experiments did not go smoothly. Your passion and devotion in science and particularly in baculovirology has such a great impact on me and will continue to encourage me for the rest of my life.

I also acknowledge my second promoter Prof. Peter J. M. Rottier for allowing me to work in your lab. Dear Peter, I benefited a lot from the excellent academic environment you created. You were very active in every meetings of my research project with so much enthusiasm and contribution. Without you and Prof. Vlak my research would not have been initiated and would not come this far.

I would like to express my sincere appreciation to my third promoter Prof. Monique M. van Oers. Dear Monique, you truly care for all the students from the bottom of your heart. You were always there to help me when I was stuck. During the last five years of my PhD research you gave a lot of strategic advice on research and self-development. In my PhD thesis-writing phase you made quite some arrangement to provide me with a new office in your lab and read and comment on my manuscripts no matter how busy you were.

I am indebted to my co-promoter Dr. Berend Jan Bosch for coaching me to start and carry out my PhD research. Berend, you were always interested, knowledgeable and enthusiastic with the F venture. Without your encouragement and forbearance, it would be difficult if at all to accomplish the

whole protein crystallization and structural analysis of the baculovirus F proteins. We had countless meetings to analyze and discuss various problems throughout my research. You were always patient and helpful yet kept challenging my scientific thinking. Your supervision inspired my passion for science and sharpened my research mind. I appreciated all your input in my scientific research even when you were quite busy with multiple projects.

Besides my supervisory team, several people have made valuable contributions to the success of my PhD research. First, I would like to thank Prof. Felix Rey (Structural Virology unit, Institut Pasteur) for your continuous interest in solving baculovirus glycoprotein structure. Together with Dr. Thomas Krey, Scott Jeffers and Ieva Vasiliauskaite in your team we cracked the nutshell! Next, I must thank my EM mentor Dr. Jan van Lent. Dear Jan, we have spent quite a lot of time working side by side and your ample EM experience and humorous character made the work really pleasant and fruitful. I appreciate for your amendments and suggestions for my manuscript on the baculovirus ultrastructural study. I would like to thank Dr. Vera I. Ros for her help in building a phylogenetic tree and revising one of my manuscripts. I have learned a lot from your excellent evolutionary expertise. You are not only a talented scientist but also a kind friend. I wish you a lot more scientific achievement and a wonderful life in the future. I thank Dr. Gary Blissard for advice on liposome flotation assays and for being one of my thesis defence opponents. I thank Dr. Yuling Bai for checking the Chinese summary of my thesis.

I also like to acknowledge all participants in our project in the framework of the Programme Strategic Scientific Alliances (PSA) between China and the Netherlands. In particular, I would like to thank Prof. Ineke Braakman, Prof. Zhihong Hu, Dr. Basil M. Arif and Dr. Mali Wang for invaluable discussion and advices. I would also like to thank my PSA fellows Ke, Stineke, Xin and Jingfang. As PhD students working in the same consortium we were connected and exchanged our ideas in the meetings. In this way we helped each other and advanced together. I cherish your friendship and support during my PhD study. The valuable experience in PSA will definitely become an unforgettable memory.

I thank all the colleagues and students in Rottier's lab. First I would like to mention my wonderful office mates Marne, Oliver and Christine for their support and friendship during my PhD study. Marne, you showed precision, persistence and confidence in your research and are very friendly in person. Oliver, your curiosity and critical mind made you stand out of the lab. Christine,

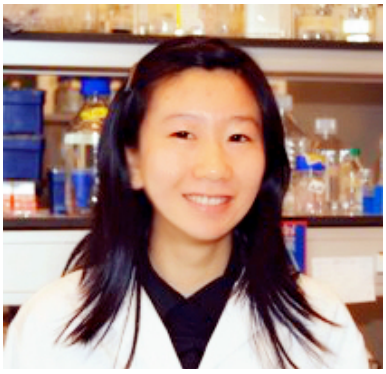
your fast-learning and multilingual skills are fascinating. I have learned so much from you guys! I was lucky to share the same office with you and I look forward to seeing you again elsewhere in the world! Next, I thank Willem for his experimental support and company in the lab. Willem, it was a pleasure to work next to you in the "red" lab! Without you it would have been difficult to start my protein expression in mammalian cells. I also like to thank Dr. Erik de Vries for your scientific discussion and kind social support. You always have many great research ideas, which we often discussed met in the train on the way to work. I thank Matthijn for showing me the protein purification procedure and sharing your academic experience with me. I thank Arno, Nancy, Martijn, Robbert, Iryna, Hui-wen, Ivy, Yong (Zou), Zhen, Chuihua, Huihui, Harry, Kazuya, Qian, Wentao, Hongbo, Meiling and Laura for all your great support and friendship in the lab. I thank my students Mike and Tom for their contribution in my research and good company.

I also acknowledge all members in the Van Oers' laboratory. First, I would like to express my sincere gratitude to the mentor of my master thesis, Fang Xu. Fang, since I started my master thesis under your supervision you are always a true friend and role model to look up to. I have learned a lot from you not only about experimental methodologies but also positive attitude to life. Thank you for guiding me through at the beginning of my academic career and further encouragement during my PhD research. Very special thanks to my paranymphs, Yue Han and Els Rood. Han, as soon as I settled in my new office desk in Wageningen I got to know there was another PhD student working on baculoviruses and also from Hebei Province and even sitting right opposite of me. So many common aspects made us quickly become very good friends. You are a friend I can count on. Els, we knew each other since seven years ago when I started to do my master thesis in the lab. I also thank you for your continuous help in my research. I am so honored to have you as one of my paranymphs. Dr. Gorben Pijlman, thank you for your interest and discussion of my research and kindly provide me with C6/36 cells and the pIB-EGFP plasmid. I would also like to thank Henry Kariithi for always being supportive in both science and life. Athos, your dry humor is so natural that I always found it is difficult not to smile. Thanks for your company during my thesis writing time in the office. Corien, you always helped me patiently with Scientific Dutch language. I appreciate you a lot for amending my Dutch summary in my thesis. Last but certainly important I would like to thank Mia, Amaya and Giel for your support and company in the lab.

I thank all the friends I met in Utrecht and Wageningen. Aileen, Xiaoqian, Weidong, Shuang (Jin), Xueqing, Yao (Liu), Qingyang, Jiefei, Xu, Chunxu, Wei (Qin), Wenfeng, Fengfeng, Nadia, Lamia, Flávia, Elina, Elske and Karst. You have colored my life besides work and I always cherish your friendship.

Finally, I would like to express my deepest gratitude to my family from both China and the Netherlands. To my parents, I appreciate your unconditional love and tolerance to me. You gave me the freedom to explore and develop. When I feel down both of you are always there for me. I am deeply indebted to you both. To my parents in law and families, you made me feel at home in the Netherlands ever since we met for the first time. You have opened my eyes to the Dutch culture so that I could quickly adapt. Most gratitude goes to my husband Paul Bakker. You are my best friend and a wonderful man. I thank you for supporting me all these years and truly appreciating my work.

About the author



Qiushi Wang （王秋实）was born October 2nd, 1983 in Zhumadian, China. In 2004 she started her study at Wageningen University (WU) in the Netherlands, where she obtained a bachelor degree in Biotechnology and a master degree (MSc) in Molecular and Cellular Biotechnology. She performed her bachelor thesis "Optimization of terminal-restriction fragment length polymorphism analysis for starch-consuming microbial communities in human colon using stable-isotope probing in the Laboratory of Microbiology and a master thesis 'Identification of nuclear localization signal in baculovirus *Chrysodeixis chalcites* nucleopolyhedrovirus (ChchNPV) photolyases' at the Laboratory of Virology, WU.

After MSc graduation she worked at the Laboratory of Molecular Biology, WU, as a research assistant for five months. She studied the mechanism of root nodule formation during symbiosis of rhizobium bacteria and legume plants. This research broadened her research interest and provided good training of experimental approaches in molecular biology. In December 2009 she started her PhD research on structural and functional aspects of the baculovirus envelope fusion protein F. This research was a joint venture between the Virology Division of the Department Infectious Disease and Immunology, Utrecht University and the Laboratory of Virology, WU, and was embedded in the collaborative Sino-Dutch Programme Strategic Scientific Alliances (PSA) between the Chinese Academy of Sciences and the Royal Dutch Academy of Arts and Sciences of the Netherlands under the theme entitled 'Baculovirus-insect cell interactions: keys for tailoring viral entry, processing and release'. She performed most of her experiments in Utrecht and collaborated with Dr. Jan van Lent of the Wageningen Electron Microcopy Center on an ultrastructural study of baculovirus budded virions and with Prof. Felix Rey and Dr. Ieva Vasiliauskaite at the Pasteur Institute on the crystallization of the baculovirus F protein. This thesis is the result of her PhD research.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

Review of literature (6 ECTS)

- The evolutionary linkage of an insect-virus membrane fusion protein with orthologs of mammalian viruses and eukaryotes

Writing of project proposal (4.5 ECTS)

- Cell entry mechanism of baculoviruses mediated by their envelope fusion protein F

Post-graduate courses (3.3 ECTS)

- Baculovirus-insect cell interactions: keys for tailoring viral entry, processing and release; WUR and CAS Wuhan Institute of Virology (2011)
- Bioinformatics; EPS, WUR (2013)

Deficiency, refresh, brush-up courses (3 ECTS)

- Electron microscopy; Bimolecular Imaging, Department Biology, Utrecht University (2011)

Competence strengthening / skills courses (3.4 ECTS)

- PhD Competence assessment; WUR (2010)
- Effective behaviour in your professional surroundings; WUR (2010)
- Scientific writing; WUR (2012)

PE&RC Annual meetings, seminars and the PE&RC weekend (0.9 ECTS)

- PE&RC Introduction weekend, first year (2010)

Discussion groups / local seminars / other scientific meetings (10.2 ECTS)

- Seminar of strategic infection biology; Department Infectious Disease and Immunology, Utrecht University (2010-2013)
- Dutch Annual Virology Symposium (2010-2013)
- PhD Evenings; Institute and Graduate School of Biomembrane (2010-2013)

International symposia, workshops and conferences (5 ECTS)

- Annual meeting of Society of Invertebrate Pathology; oral presentation (2012)
- Annual meeting of American Society for Virology; oral presentation (2013)

Supervision of 2 MSc students

- Functional study of alphabaculovirus F fusion protein
- Fusion activation mechanism of alphabaculoviruses

The research described in this thesis was financially supported by the Program Strategic Scientific Alliances of the Royal Dutch Academy of Sciences and the Chinese Academy of Science.

Printing: CPI–KONINKLIJKE WÖHRMANN, Zutphen, the Netherlands.

Propositions

1. Baculovirus F proteins are evolutionarily related to F proteins of vertebrate paramyxoviruses.
   (this thesis)

2. The bipolar distribution of spikes in the baculovirus budded virion is unique as compared with virus particles of members of other envelope virus families.
   (this thesis)

3. The statement in Espinosa and Fornoni (2006) that host resistance is favorable over host tolerance for selection of pathogens is dependent on host and pathogen properties. (Espinosa, E. G. and Fornoni, J. 2006, *Host tolerance does not impose selection on natural enemies*, New Phytologist, 170, 609-614)

4. To conclude that telomere shortening leads to aging as stated in Shammas et al. (2011) requires further evidence. (Shammas, M. A. 2011, *Telomeres, lifestyle, cancer, and aging*, Current Opinion in Clinical Nutrition and Metabolic Care, 14, 28-34)

5. In the Netherlands beer is an indispensable buffer and catalyzer for social activities.

6. Time passing tends to be perceived faster with unpredictability of the goals.

Propositions belonging to the thesis

Structural and functional studies of the baculovirus envelope fusion protein F

Qiushi Wang

Wageningen, 18th November 2015