Expression and characterization of spike protein complexes Gp2/Gp3/Gp4 and Gp5/M of the Arterivirus

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy
of Freie Universität Berlin

by

Balaji Chandra Sekhar Sinhadri

Born in Avanigadda (India)

March 2015

This work was accomplished between July 2010 and February 2015 at the Institut für Virologie**,** Freie Universität Berlin under the supervision of PD Dr. Michael Veit.

Date of disputation: 19.05.2015

Disputation committee:

Prof. Dr. Udo Heinemann (Head of the doctoral committee)
Priv.-Doz. Dr. Michael Veit
Prof. Dr. Helge Ewers
Prof. Dr. Volker A. Erdmann
Priv.-Doz. Dr. Matthias Peiser

TABLE OF CONTENTS

2 Materials and methods

List of Figures

Abbreviations

Asp	Asparagine
BHK-21	baby hamster kidney
C	Carboxyl terminus
Cys	Cysteine
cDNA	Complementary DNA
CD169	Sialoadhesin
DNA	Deoxyribonucleic acid
DMV	Double membrane
CHO-K1	Chinese hamster ovary cells
E	Envelope protein
EAV	Equine arteritis virus
ER	Endoplasmatic reticulum
ERGIC	ER-Golgi intermediate compartment
GAGs	Sulphate glycosaminoglycans
Gp2	Glycoprotein 2
Gp3	Glycoprotein 3
Gp4	Glycoprotein 4
Gp5	Glycoprotein 5
HA	Haemagglutinin
HCMV	Human cytomegalovirus
HR	Hydrophobic region
His	Histidine
kDa	Kilo Dalton
LDV	Lactate dehydrogenase-elevating virus
LC-MS	Liquid chromatography–mass spectrometry
M	Membrane protein
m.o.i.	Multiplicity of infection
mRNA	Messenger RNA
N	Nucleocapsid protein

N	Amino terminus
NCBI	National Center for Biotechnology Information
nsp	Non-structural protein
ORF	Open reading frame
p.i.	Post infection
PBS	Phosphate-buffered saline
PNGase F	Peptide-n-glycosidase
pp	Polyprotein
PRRSV	Porcine reproductive and respiratory syndrome virus
PVDF	Polyvinylidene difluoride
RK 13	Rabbit kidney
RNA	Ribonucleic acid
RTC	Replication and transcription complex
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
sgRNA	Subgenomic RNA
SHFV	Simian hemorrhagic fever virus
SP	Signal peptide
SPase	Signal peptidase
TM	Transmembrane
TMD	Transmembrane domain
TMR	Transmembrane region
VERO	African green monkey kidney
VLPs	Virus-like particles
US11	Unique short 11 protein
WB	Western blot
wt	Wild type

Zusammenfassung

Das equine Arteritisvirus (EAV) sowie das porzine reproduktive und respiratorische Syndrom Virus (PRRSV) sind umhüllte RNA Viren, die zu der Familie der Arteriviren gehören. Während eine Infektion mit EAV in Pferden zu Fehlgeburten und respiratorischen Symptomen führen kann, kommt es bei PRRSV zu persistenten Infektionen von Schweinen und somit zu einem der größten ökonomischen Verluste in Schweinemastbetrieben. Während der Infektion sind die Glykoproteinkomplexe Gp2/3/4 und Gp5/M essentiell für den Zelleintritt und das Abknospen neuer Viruspartikel.

Die Ektodomänen des Gp2/3/4-Komplexes von PRRSV sowie EAV wurden koexprimiert, um ihre Komplexbildung in Insektenzellen zu untersuchen. Die erzielten Resultate zeigen, dass die PRRSV Gp2/3/4 Proteine als ein Komplex in den Zellkulturüberstand abgegeben wurden, der über Disulphidbrücken miteinander verbunden sind. Im Gegensatz dazu wurde Gp3 von EAV zwar exprimiert, bildete jedoch keinen Komplex mit Gp2/4. Die Gp2/3/4 Proteine von PRRSV und EAV sind durch ihre hydrophobe C-terminale Domäne in der Membran verankert, wohingegen sie sekretiert wurden, wenn man diese Domäne entfernte. Dieses Ergebnis lässt zu dem Schluss kommen, dass alle Proteine bis auf das Gp3 von EAV, wessen C-terminale Domäne die Membran nicht durchspannt, eine Typ 1 Membrantopologie aufweisen.

Massenspektrometrische Untersuchungen konnten die Signalpeptid-Schnittstelle für Gp2, Gp3 und Gp4 von PRRSV sowie EAV genau bestimmen. Zudem konnte die Analyse der Glykosylierungen der Proteine zeigen, dass alle vorausgesagten N-Glykosylierungsstellen in allen drei Glykoproteinen der beiden Viren genutzt werden. Gp3 besitzt, exprimiert in Säugerzellen, ein ungeschnittenes Signalpeptide, wohingegen meine Ergebnisse in Insektenzellen zeigen, dass das Signalpeptide unter diesen Bedingungen jedoch abgeschnitten wird. Der Größenunterschied zwischen dem in Säuger- oder Insektenzellen hergestellten, unglykosylierten Gp3 von EAV indiziert, dass die Abtrennung durch eine bisher unbekannte spezifische Prozessierung in Insektenzellen erfolgt.

Die EAV Proteine GP2 und vier waren in der Lage Komplexe auszubilden, die durch Disulphidbrücken verbunden sind, jedoch zudem auch zur Ausbildung von fehlerhaften Multimeren führen können. Meine erzielten Resultate belegen, dass die Cysteinreste in der GP4-Ektodomäne keine Einfluss auf diese Multimerformation haben und dass jeder Cysteinrest in der Gp4 Ektodomäne die Disulphidbrücke zu Gp2 kompensieren kann.

Die massenspektrometrische Untersuchung von PRSSV Gp5 ergab, dass das Signalpeptid an beiden Stellen abgeschnitten wird, die durch die Nutzung von bioinformatische Software vorhergesagt wurden. Dies deutet darauf hin, dass der größte Anteil der Gp5-Moleküle kein “decoy” Epitop besitzt, von dem angenommen wird, dass es das Immunsystem von dem neutralisierenden Epitop ablenkt.
Die Mutation des einzigen Cysteinrestes nach der Signalpeptidschnittstelle in der Gp5-Ektodomäne konnte zeigen, dass dieses für die Ausbildung der Disulphidbrücke von Gp5 und M verantwortlich ist.

Die in dieser Arbeit erzielten Ergebnisse über die Haupt- und Nebenglykoproteinkomplexe von EAV und PRRSV könnten hilfreich für weitere strukturelle und funktionale Studien sein, um die Rolle der Komplexe während des Zelleintritts besser zu verstehen.

Summary

Equine arteritis virus (EAV) and Porcine reproductive and respiratory syndrome virus (PRRSV) are enveloped RNA viruses belonging to the family Arteriviridae. EAV infection leads to abortion and respiratory illness in horses. PRRSV causes persistent infections in pigs, which is one of the main reasons for the economic losses in the swine industry. The glycoprotein complex Gp2/3/4 and Gp5/M are essential for cell entry and budding, respectively.

PRRSV and EAV Gp2/3/4 ectodomains were co-expressed to determine their complex formation in insect cells. The results revealed that PRRSV Gp2/3/4 proteins secreted into the cell culture supernatant as a disulphide linked complex. In contrast, Gp3 of EAV was expressed, but not associated with Gp2/4. The Gp2/3/4 proteins of both PRRSV and EAV are attached to the membrane through their C-terminal hydrophobic transmembrane region as they were secreted when it is removed. This result showed that all the proteins contain type I membrane topology except EAV Gp3 since the C-terminus does not span the membrane.

Mass spectrometry results demonstrated the exact signal peptide cleavage sites for Gp2, Gp4, and Gp3 of both PRRSV and EAV. The analysis of glycosylation showed that all predicted N- linked glycosylation sites are used in both EAV and PRRSV Gp2/3/4. EAV Gp3 possesses an uncleavable signal peptide. But, my results showed that the signal peptide is cleaved off from the insect cell expressed Gp3. The difference in size between mammalian and insect cell expressed EAV Gp3 (deglycosylated) indicated that the signal peptide cleavage is due to an unknown Gp3 specific processing difference in insect cells. EAV Gp2/4 proteins were able to form the disulphide linked complex but exist in aberrant multimeric forms. My results showed that cysteine residues in the Gp4 ectodomain have no influence on multimer formation, and any cysteine residue in the Gp4 ectodomain can compensate the disulphide linkage with Gp2.

Mass spectrometry result of PRRSV Gp5 demonstrated that the signal peptide is cleaved at two sites, which are also predicted by the bioinformatics software. This

finding indicates that the majority of the Gp5 molecules lack the decoy epitope, which is believed to mask the immune response from its neutralising epitope. Mutating the only cysteine residue in the Gp5 ectodomain after the SP cleavage demonstrated that it was responsible for the disulphide linkage between Gp5/M.

The data produced in this study on minor and major glycoprotein complexes of EAV and PRRSV may be useful for future structural and functional studies to understand their role in cell entry.

Publications:

Parts of this PhD thesis have been published in the following manuscripts.

Thaa, B., Sinhadri, B.C., Tielesch, C., Krause, E. and Veit, M. 2013. Signal peptide cleavage from GP5 of PRRSV: a minor fraction of molecules retains the decoy epitope, a presumed molecular cause for viral persistence. PLoS One 8(6), e65548. doi:10.1371/journal.pone.0065548

Veit M., Matczuk AK., Sinhadri BC., Krause E., Thaa B.. 2014. Membrane proteins of arterivirus particles: Structure, topology, processing and function. Virus Res 19;194:16-36. http://dx.doi.org/10.1016/j.virusres.2014.09.010

1. Introduction

Equine arteritis virus (EAV) and Porcine reproductive and respiratory syndrome virus (PRRSV) are single-stranded, positive-sense RNA viruses of the family *Arteriviridae* (Meulenberg, 2000, Balasuriya *et al.,* 2013). Other members of the *Arteriviridae* family are simian haemorrhagic fever virus (SHFV) and murine lactate dehydrogenase elevating virus (LDV) (Snijder & Meulenberg, 1998, Snijder, E. J. & Kikkert, M. 2013). PRRSV is a major causative agent of persistent infections in pigs and sometimes causes severe disease that leads to respiratory problems, poor growth, and weight loss. PRRSV is one of the main causes for severe economic losses in the swine industry (Snijder *et al.,* 2014, Thaa *et al.,* 2013, Veit *et al.,* 2014). EAV is a critical pathogen in horses; the infection signs include abortions and respiratory illness with flu-like symptoms, which sometimes leads to death in young animals (MacLachlan and Balasuriya, 2006). *Arteriviridae* family belongs to order *Nidovirales* together with *Coronaviridae*, *Toroviridae,* and *Roniviridae* based on their genome organisation and replication strategy (Snijder & Meulenberg, 1998, Snijder, E. J. & Kikkert, M, 2013)

1.1 Virion structure

Arteriviruses are coated with a smooth envelope with a diameter of 50-60 nm (Spilman *et al.,* 2009, Snijder *et al.,* 2013). Nucleocapsid protein (N-protein) encapsulates the genome that forms a core packaged in a lipid envelope. The envelope of arteriviruses consists of membrane proteins and major and minor glycoproteins (de Vries *et al.,* 1992) (Fig.1.1). Gp5 and M are major envelope proteins, and Gp2, Gp3, Gp4 and E are minor envelope proteins (de Vries *et al.,* 1992). Minor envelope proteins are essential for infectivity, whereas the major envelope proteins Gp5 and M are essential for budding of the virions (Tian *et al.,* 2012, Snijder *et al.,* 2014).

1.2 Cell tropism and life cycle of the virus

Arteriviruses have very restricted cell tropism (Plagemann & Moennig, 1992, Veit *et al.,* 2014). EAV is an exception in this regard: it infects and replicates in primary equine macrophages, endothelial cells lining, and small blood vessels. EAV also infects a variety of cell lines such as baby hamster kidney (BHK-21), rabbit kidney (RK-13), African green monkey kidney (VERO), rhesus monkey kidney (LLC-MK2), and MARC-145 (Maess *et*

al., 1970, Konishi *et al.,* 1975, Lu *et al.,* 2012, Balasuriya U.B.R., *et al.,* 2013). PRRSV has restricted cell tropism and replicates in porcine alveolar macrophages in the infected animal (Snijder *et al.,* 2014; Veit *et al.,* 2014).

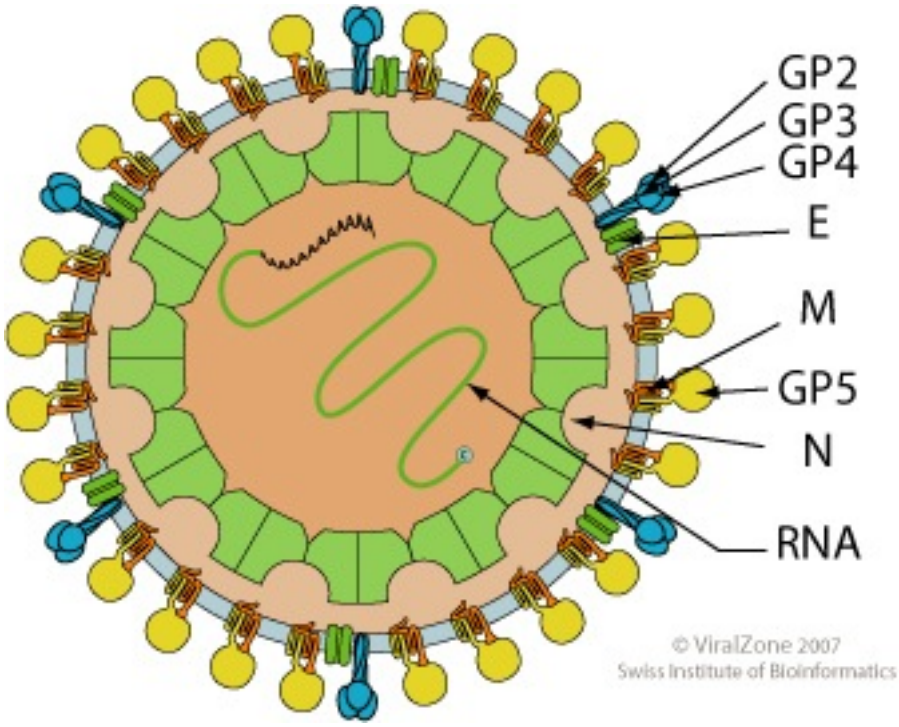


Figure 1.1 Structure of the Arterivirus virion.
Figure used with authorisation from Swiss Institute of Bioinformatics. Picture Source: ViralZone: www.expasy.org/viralzone

The genome of arterivirus is 12-16 kb long, presumably 5′ capped and 3′- polyadenylated (den Boon *et al.,* 1991). The 5′ region of the genome is occupied by Open Reading Frame 1a (ORF1a) and ORF1b. ORF1a and ORF1b translate into replicase polyproteins (PP) 1a and polyprotein 1b, which are precursors for at least 13 non-structural proteins (nsp) (Fang *et al*., 2010). The 3′ proximal region of the arterivirus genome has a tight organisation and translates into 8 to 12 genes (Pasternak *et al.*, 2006). These genes overlap with neighbouring genes and are expressed from a nested set of segmented mRNAs (Sawicki *et al.*, 2007) (Fig.1.2).

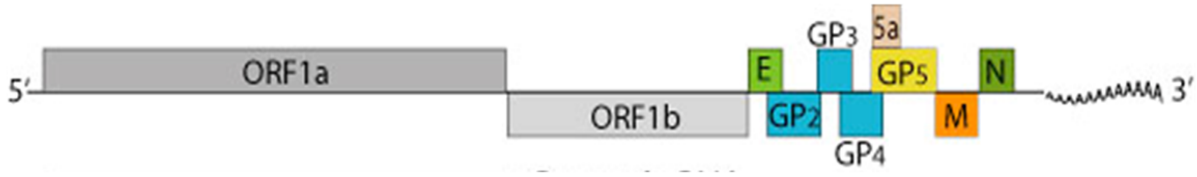


Fig 1.2: Genome structure of Arterivirus: The replicase ORF1a and ORF1b codes for non-structural proteins, which form the Replication and Transcription complex (RTC). The structural genes that are transcribed into six subgenomic RNAs are encoded by 3′ end. Figure used with authorisation from Swiss Institute of Bioinformatics. Picture Source: ViralZone: www.expasy.org/viralzone

The polyprotein 1a and 1b are cleaved by cellular proteases, and the NSPs (nonstructural protein) derived from them are assembled into a Replication and Transcription Complex (RTC) (Snijder *et al*., 2001). RTC engages in both full-length and subgenomic-length RNA strand synthesis. The subgenome-length RNA strand serves as a template for the synthesis of Sg mRNAs, which are essential for expression of structural proteins (Pasternak *et al.*, 2006). A nucleocapsid is formed when the newly synthesised genome is encapsulated by budding through Endoplasmic Reticulum-Golgi Intermediate Compartment (ERGIC), which contains the membranes with the viral envelope proteins (Wada *et al.*, 1995). The newly formed virions mature in the Golgi complex and are released from the infected cells via the exocytic pathway (Fig.1.3).

1.3 Role of the membrane proteins for virus replication

Arteriviruses contain seven structural proteins: the nucleocapsid protein N, glycoproteins Gp2, Gp3, Gp4, Gp5, and the non-glycosylated M-protein (Wieringa *et al*., 2003a, Das *et al.,* 2010, Wissink *et al*., 2005). Gp5 forms a disulphide linkage to M, and Gp2/Gp3/Gp4 forms a disulphide-linked heterotrimeric complex. A recently identified membrane protein called ORF5a is synthesised from an open reading frame that also encodes Gp5 (Firth *et al*., 2011, Johnson *et al*., 2011). The ORF5a protein is incorporated as a minor component into virus particles. The ORF5a protein is essential for PRRSV replication; in the case of EAV, the virus grows in the absence of ORF5a, although the viral titres are low, and the plaque size is tiny (Sun *et al.,* 2013).

All the structural proteins have been proven to be essential for replication of EAV and PRRSV (Veit *et al.,* 2014, Snijder *et al.,* 2014). No virus particles have been released from the infected cells in the absence of GP5 or M. Hence, the Gp5 and M complex is essential for virus budding. Virus-like particles (VLPs) are formed in the absence of Gp2, Gp3, or Gp4, but they are non-infectious. This observation indicates that the minor glycoprotein complex is critical for cell entry (Snijder *et al.,* 2014). In addition, exchanging the ectodomains of EAV Gp2/Gp3/Gp4 with the respective domains of PRRSV alter the cell tropism, whereas exchanging the ectodomains of Gp5 and M between EAV and PRRSV does not result in any change in cell tropism. Thus, the minor GP complex Gp2/Gp3/Gp4 might be essential for virus entry into cells, i.e. binding to a receptor and catalysing membrane fusion (Tian *et al.,* 2012; Snijder *et al.,* 2014).

The Gp5/M complex is also expected to be involved in virus entry in addition to its role in virus assembly and budding (Van Breedam *et al.,* 2010). Gp5/M initiates the initial interactions of PRRSV with the cell entry mediators heparin sulphate glycosaminoglycans (GAGs) and sialoadhesin (CD 169) (Van Breedam *et al.,* 2010).

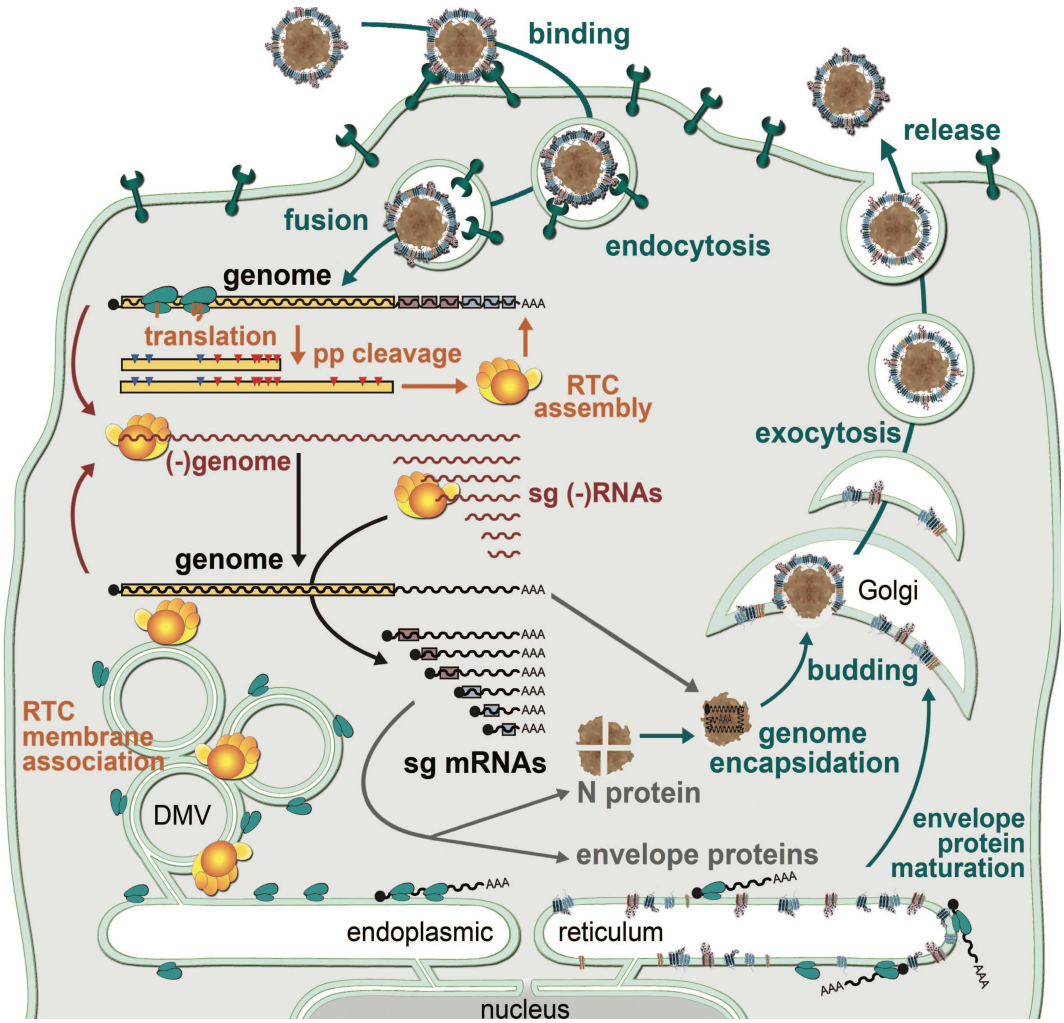


Fig 1.3: Arterivirus replication cycle. RTC – replication and transcription complex; sg mRNA – sub-genomic messenger RNA; pp – polyprotein; DMV – double membrane; N protein-Nucleocapsid protein. (Used with permission from Snijder *et al.*, J Gen Virol 94 (2013) 2141–2163).

1.4 Structure and processing of the membrane proteins of Arteriviruses

1.4.1 Gp5/M complex

M protein is around 160-170 amino acids in length and is conserved among the Arteriviruses. It consists of three putative transmembrane regions and a hydrophilic cytoplasmic tail, which is around 70-80 amino acids long (Fig.1.4). M protein lacks an N-terminal signal peptide and glycosylation sites for the attachment of N-linked glycans. (Snijder *et al*., 2003). The only cysteine residue at position 8 in the ectodomain of EAV M protein (Bucyrus strain) is responsible for the disulphide linkage with Cysteine 34 of Gp5; it is also present in M proteins of all PRRSV strains (Dobbe *et al*., 2001, Verheije *et al*., 2002, Tian *et al*., 2012) (Fig.1.4). Gp5/M heterodimerization is crucial for the transport of Gp5/M from the ER to the Golgi complex. Only accurately assembled Gp5/M complexes can pass the quality control system of ER. Exchange of cysteine 8 in M completely eradicates the release of infectious virus particles (Snijder *et al*., 2003). It was also reported that Gp5-M disulphide linkage is essential for lactate dehydrogenase elevating virus (LDV) infectivity (Veit *et al.,* 2014).

Gp5 monomers rapidly form disulphide bonds with M and all Gp5 monomers synthesized form heterodimers. In contrast, the association of M protein with Gp5 occurs slowly, and a fraction of the M protein ends up in cells as monomers (Veit *et al*., 2014). Monomer of M protein also forms homodimers by covalent linkage (de Vries *et al*., 1995a).

Gp5 is around 200 amino acids in length in PRRSV and 225 amino acids in EAV. Gp5 contains an N-terminal cleavable signal peptide, an ectodomain of approximately 30 amino acids in PRRSV, and 90 amino acids in EAV. The ectodomain contains at least one N-glycosylation site (Dokland, 2010, Veit *et al*., 2014). The region between amino acids 65 and 127 (PRRSV) or between 115 and 180 (EAV) is hydrophobic in nature and considered to span the membrane three times (Fig.1.4). The C-terminal part ends up in the virus interior (de Vries *et al*., 1992; Dokland, 2010). Gp5 of PRRSV only shows 50% amino

acid homology between European and North American strains; it is the most variable structural protein of Arteriviruses (Meng, 2000, Shi *et al*., 2010, Stadejek *et al.,* 2013, Veit *et al*., 2014)

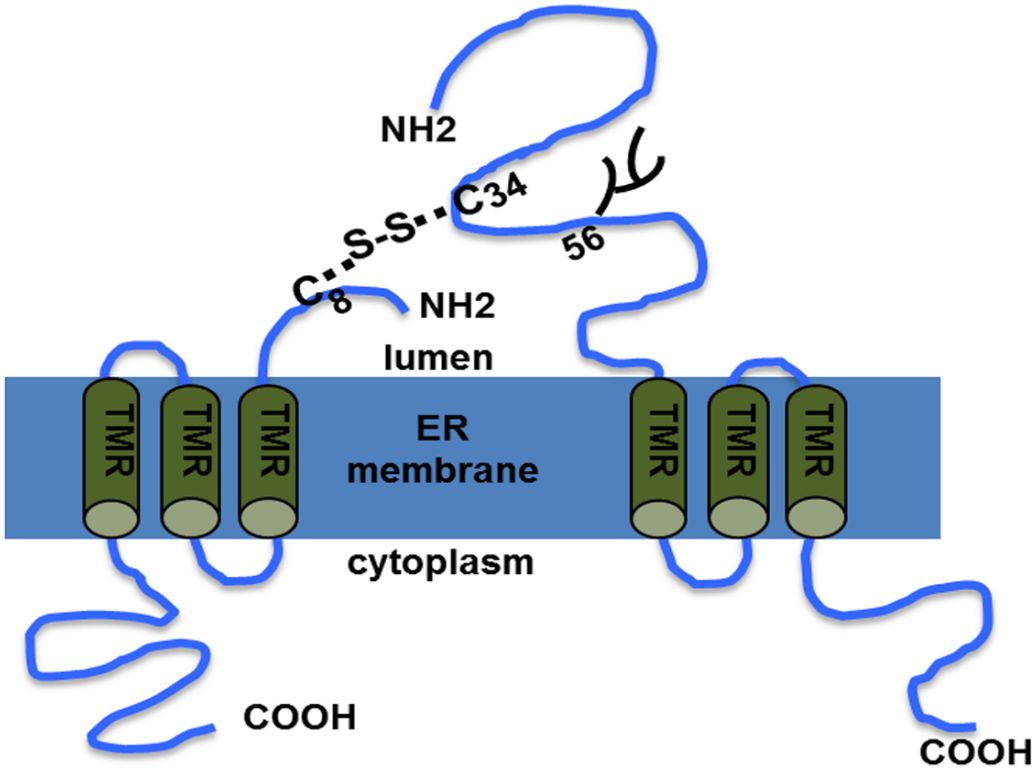


Fig.1.4. Schematic representation of the Gp5/M complex of EAV Bucyrus strain. Cysteine residues involved in the disulphide linkage are represented as 'C' with their respective amino acid number. S-S indicates the disulphide linkage between cysteine 8 of M with cysteine 34 in Gp5. N-linked glycans are shown as a branch with the number of asparagine residue below. TMR: putative transmembrane regions.

1.4.2 PRRSV Gp5 and the antigenic epitopes:

Gp5 is presumed to be the main target for neutralising antibodies in arteriviruses (Lopez and Osorio, 2004). Two epitopes, epitope A and epitope B, were identified in PRRSV North American genotype (strain VR-2332) (Fig. 1.5). Epitope A (residues 27-31) induces a strong and early immune response, but it cannot block the virus infection (non-neutralising). Epitope B (residues 37-44) stimulates neutralising immune response, but the antibodies appear late after infection (Lopez and Osorio, 2004). A 'decoy hypothesis' has

been proposed based on these findings suggesting that the epitope A masks immune response towards neutralising epitope B. However, epitope A is present in the potential signal peptide (Fig.1.5).

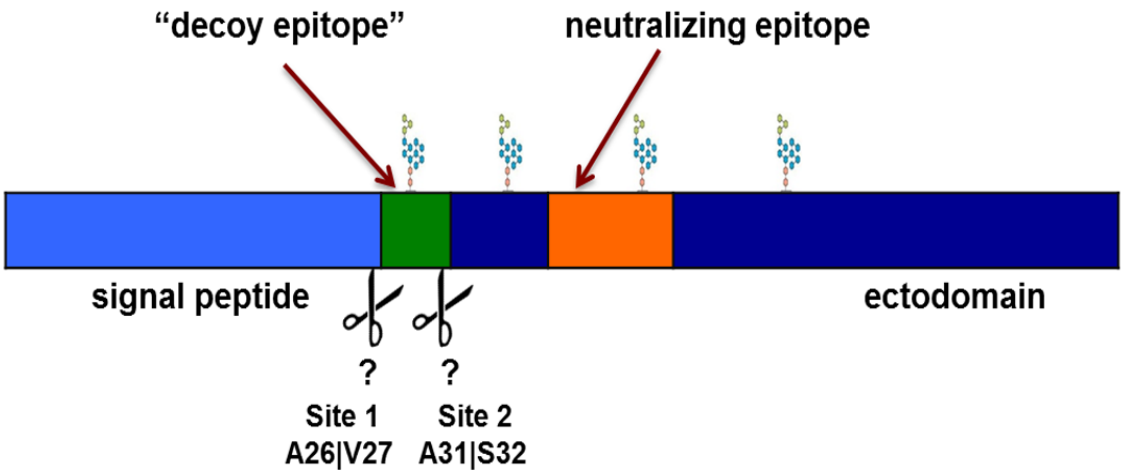


Figure 1.5. Scheme of PRRSV GP5 ectodomain and the antigenic epitopes: Signal peptide cleavage predicted to occur either at site 1 (A26|V27) or site 2 (A31|S32) as indicated in the scheme. The predicted signal peptide is represented in light blue colour, and the cleavage sites are shown as site 1 and site 2. The 'decoy epitope' (green) is not present in the mature protein by the signal peptide cleavage at site 2. The neutralising epitope is represented in orange colour, and the ectodomain is in dark blue colour. Glycans are depicted as branches. This scheme represents the Gp5 of PRRSV type 2-reference strain VR-2332.

1.5 Gp2b/Gp3/Gp4 complex

1.5.1 Gp2

Gp2 is 227 amino acids in length in EAV and 249 in PRRSV, and is composed of a cleavable signal peptide at the N-terminus, a large ectodomain (around 150 amino acids in EAV and 170 amino acids in PRRSV), one hydrophobic transmembrane region (around 25 amino acids in PRRSV and 30 amino acids in EAV), and a short cytoplasmic tail (around 20 amino acids in PRRSV and 30 in EAV) (Dokland, 2010, Veit *et al.,* 2014). Gp2 of the majority of the PRRSV strains has two glycosylation sites (N178 and N184) that are situated at about 25 amino acids distance from the transmembrane region (Wissink *et al*., 2005, Murtaugh *et al*., 2010). Gp2 of EAV Bucyrus strain contains two potential

glycosylation sites, and out of two functional sites, only one N-glycosylation sequon is used.

1.5.2 GP3

Gp3 is 163 amino acids long in EAV and 265 amino acids in PRRSV (Dokland, 2010). Gp3 consists of an N-terminal signal peptide, an ectodomain containing potential N-glycosylation sites, a hydrophobic region, and a hydrophilic domain. An *in vitro* translation experiment with the microsomes confirms that the signal peptide is not cleaved from Gp3, and it is speculated that it might serve as a membrane anchor (Faaberg and Plageman, 1997; Wieringa *et al*., 2002). Based on these observations, it was proposed that the GP3 might be a Type II (N-terminal membrane anchor and C-terminal ectodomain) or Type III (hairpin structure by spanning the membrane with both N- and C- terminal regions) (Wieringa et al., 2003a) (Fig. 1.6B)

The highly reliable bioinformatics tool signalP 4.1 predicts with high confidence that the signal peptide is cleaved from Gp3. Gp3 of EAV contains an overlapping sequon NNTT nearby the signal peptide cleavage site, and it was demonstrated that both glycosylation sites were efficiently used. This is a new observation in glycobiology (Matczuk *et al.,* 2014). Gp3 of EAV contains six potential glycosylation sequons, and all six glycosylation sites are utilised (Matczuk *et al.,* 2014).

1.5.3 Gp4

Gp4 protein is 152 amino acids in length in EAV and 178 amino acids in PRRSV. Gp4 contains an N-terminal signal peptide, an ectodomain that is around 100 amino acids in EAV and 130 amino acids in PRRSV, a short hydrophobic transmembrane region, and a short cytoplasmic tail that is 11 amino acids in EAV. The cytoplasmic tail is absent in PRRSV Gp4 (Dokland, 2010, Veit *et al.,* 2014). Gp4 of PRRSV American type 2 reference strain VR-2332 has four N-glycosylation sites (N37, N84, N120, N 130) and all four glycosylation sites are glycosylated (Dokland, 2010; Veit *et al*., 2014). EAV Gp4 comprises four potential N-glycosylation sequons (N33, N55, N65 and N90), out of which only three are used. The glycosylation sequon NPT at N65 is most likely not used because of the presence of proline, which prevents the N-glycosylation (Veit *et al.,* 2014).

1.5.4 E protein

E is a small (69-73 amino acids long, 8kDa) non-glycosylated protein of Arterivirus, which is encoded by ORF2a in EAV and ORF2b in PRRSV. E protein composed of a N-terminal domain with no predicted signal sequence, a hydrophobic transmembrane region, and a hydrophilic C-terminal region (Thaa *et al*., 2009; Veit *et al*., 2014). The E protein of Arteriviruses presumably has a proton channel function. The role of E protein as an ion channel is supported by the requirement of arterivirus entry into cells through clathrin-dependant virus uptake and exposure to lower pH acidic compartments (Nauwynck *et al.,* 1999, Lee and Yoo, 2006, Nitschke *et al.,* 2008; Veit *et al.,* 2014). Another fact that supports the role of E protein as a proton channel is that its structure resembles well-studied viroporins, such as the M2 protein of influenza virus (Veit *et al.,* 2014). However, this has to be proven with further experiments.

The E protein also has an effect on assembly of the Gp2/Gp3/Gp4 complex. Incorporation of Gp2/Gp3/Gp4 complex into the virus particles can be entirely averted in the absence of E protein. Likewise, severe reduction in the amount of E protein was registered in the absence of the minor glycoproteins Gp2, Gp3, or Gp4 (Wieringa *et al.,* 2004). Thus, E protein interacts with Gp2/Gp3/Gp4, either via a direct, physical interaction or indirectly, for example by adjusting the ionic milieu in the ER and/or Golgi lumen to favour Gp2/3/4 complex formation.

1.5.5 Disulphide linkage between Gp2/Gp4 and Gp3:

The building of Gp2/Gp3/Gp4 in EAV starts with the intramolecular disulphide bond formation in Gp2, which is followed by intermolecular association of Gp4 with Gp2 to form a Gp2/Gp4 dimer (Wieringa *et al*., 2003a) (Fig.1.6). Gp2 of EAV contains three cysteine residues in its ectodomain, and they are located at amino acids 48, 102, and 137. Mutagenesis experiments indicate that cysteine 48 forms an intramolecular disulphide linkage with cysteine 137, and cysteine 102 forms a covalent linkage with an unidentified cysteine in Gp4 (Wieringa *et al.,* 2003a) (Fig.1.6). EAV virus particles were totally non-infectious when cysteine 102 of Gp2 protein was mutated to serine. A 3-log difference in the infectivity was observed between Gp2 cysteine 48 or 137 mutants and the wild type virus (Wieringa *et al*., 2003a).

The association of EAV Gp3 with the already formed Gp2/Gp4 dimer occurs later and is a unique process since Gp3 associates with the Gp2/Gp4 only after the release of virus

particles (Wieringa *et al*., 2003b). Since disulphide linkage requires PDI and other folding factors, it is unclear how Gp3 is linked to the Gp2/Gp4 dimer outside the lumen of ER. (Wieringa *et al*., 2003b). It was proposed that Gp3 is a non-structural protein in PRRSV, and is secreted out from the infected cells. However, recent studies demonstrate that Gp3 is a structural component of the mature PRRSV virus particles and forms covalent linkage with Gp2/Gp4 (de Lima *et al*., 2009). Gp2/Gp3/Gp4 trimer formation happens in a time-dependent manner, and trimer formation can be enhanced with oxidative environment and alkaline pH (Wieringa *et al.,* 2003b). Gp2 and Gp4 also form homodimers (Gp2-Gp2 and Gp4-Gp4) and various disulphide linked multimers of Gp2/Gp3 and Gp4 when expressed in cells (Wieringa *et al.,* 2003b).

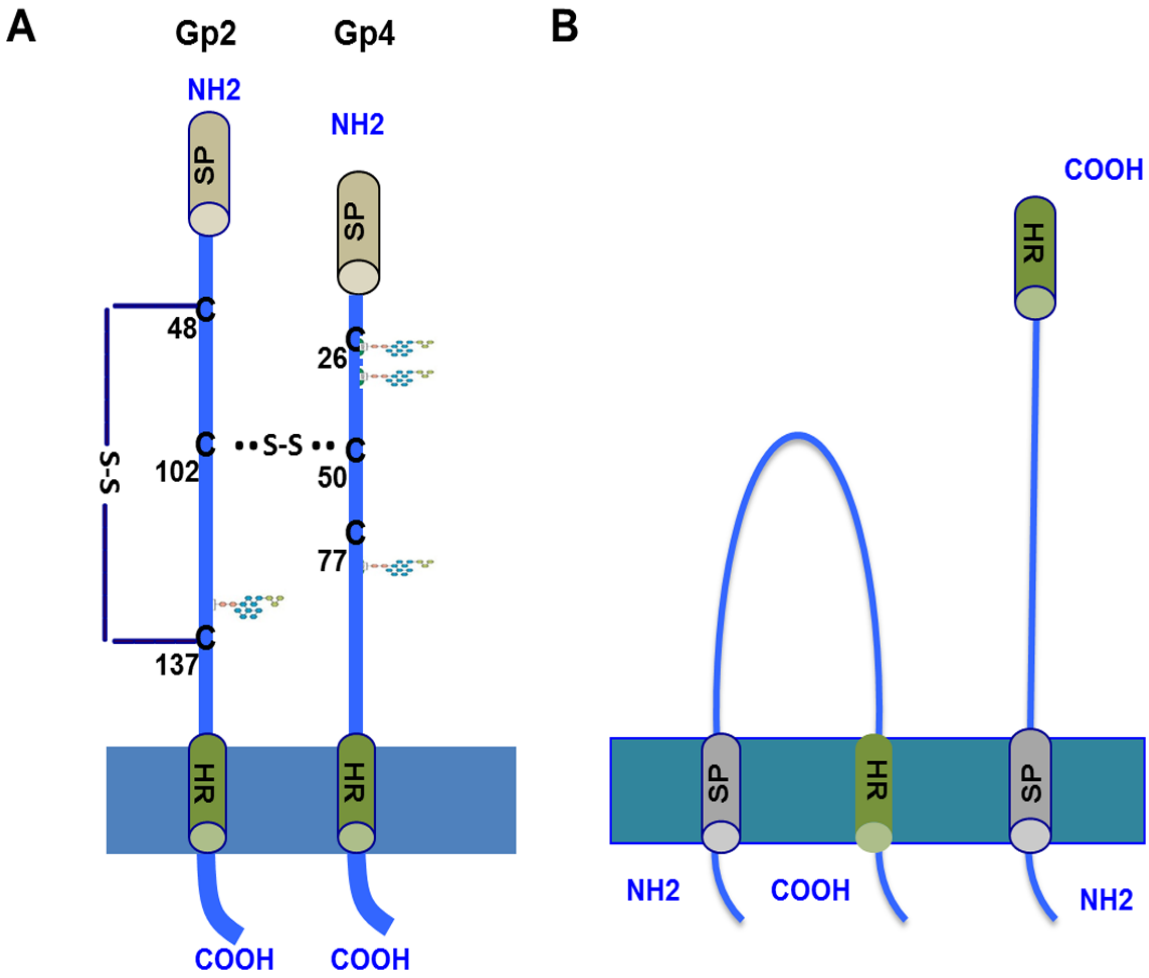


Fig. 1.6. Schematic representation of EAV Gp2/Gp3/Gp4 complex. Cysteine residues within the ectodomains are indicated with the letter 'C', and N-glycosylation sites are

represented with branches. S-S and dotted lines indicates a disulphide linkage between cysteine 102 of Gp2 and an undetermined cysteine in Gp4. Gp2 also forms an intramolecular disulphide bond between cysteines 48 and 137. B) Two proposed Gp3 membrane topologies (Wieringa *et al*., 2003a). HR: hydrophobic region, SP: signal peptide.

1.6 Function of co-translational modifications of proteins

1.6.1 N-glycosylation

N-linked glycosylation occurs at the membrane of ER and it is performed by oligosaccharyl transferase (OST) complex, which catalyse the attachment of a high mannose core to the side chain amide of an asparagine residue in the context of the consensus sequence N-X-S/T (Zielinska *et al*., 2010, Markus Aebi, 2013). The efficiency of the glycosylation depends on the residue in the sequon (e.g. NXT is more efficiently used compared to NXS) (Yan and Lennarz, 2005, Kelleher and Gilmore, 2006, Markus Aebi, 2013). Glycosylation is often essential for protein folding. Glycans prevent the exposure of hydrophobic regions on the surface of the folding protein, its misfolding, and aggregation (Markus Aebi, 2013; Veit *et al.,* 2014). Removal of potential glycosylation sequons by site-directed mutagenesis often results in misfolded proteins or affects their intracellular transport (Braakman and Bulleid, 2011).

1.6.2 Disulphide bond formation

Disulfide bond formation results in a covalent bond (-S-S-) between two cysteine residues by oxidation of sulfhydryl (-SH) groups. Any two cysteine residues in close proximity can form disulphide linkage in the oxidising environment of the ER lumen (Veit *et al.,* 2014). Disulphide bonds stabilise the native protein structure and play a role in folding and secretion. Protein Disulphide Isomerase (PDI) can reduce or reform the disulphide bonds (Ellgaard and Ruddock, 2005, Oka and Bulleid, 2013).

1.6.3 Signal peptide cleavage

Signal peptides are usually N-terminal domains of a polypeptide and direct the protein to the ER. The size of a signal peptide ranges between 20-30 amino acids, but in some cases it might be up to 80 amino acids. The structure of a signal peptide comprises an N-region composed of basic amino acids, a middle hydrophobic h-region, and a polar c-region

(Auclair *et al.*, 2012). Uncharged and small amino acids, such as Ala, Gly, Ser, and Cys, are required at position -1 and -3 upstream of the cleavage site (Veit *et al.,* 2014). The cleavage of a signal peptide is usually very rapid after it enters into the lumen of the ER, but in some instances cleavage occurs a few minutes after the synthesis of the polypeptide (Land et *al*., 2003, Braakman and Bulleid, 2011). Uncleaved signal peptides with a long hydrophobic –h region can span the membrane and act as signal anchor domains. Proteins, which possess a cleavable signal peptide and also lack other hydrophobic regions are secreted by the cell (Veit *et al.,* 2014).

Signal peptides of majority of the proteins are cleaved off by serine protease, a translocon-associated SPase, which forms a complex known as the signal peptidase complex (SPC) (Nyathi *et al.*, 2013). SPC is a hetero-oligomeric membrane protein, which is situated on the ER membrane luminal side (Shelness *et al.*, 1993). Cleaved signal peptides have post-cleavage functions in some viruses (e.g. they help in protein maturation in Lassa virus glycoprotein) (Eichler *et al*., 2003). Cleaved signal peptides are processed by signal peptide peptidase (SPP), which is an intermembrane aspartyl protease. SignalP 4.1 can predict the cleavage of the signal peptide from the mature protein.

1.6.4 Protein folding and oligomerisation

The membrane protein folds in the lumen of the ER with the assistance of molecular chaperones such as calnexin, calreticulin, BiP, and heat shock proteins. These chaperones assist the protein folding by protecting the newly synthesised protein and shielding its hydrophobic patches in order to prevent their aggregation (Braakman and Bullied, 2011). Co-translational modifications such as glycosylation and disulphide bond formation stabilise the newly synthesised proteins and thus play a key role in protein folding. Subsequent oligomerisation of viral glycoproteins is a complex process; the glycoproteins that fail to fold properly cannot oligomerise into viral spikes (Veit *et al.*, 2014). Folding and oligomerisation of hetero-oligomers depends on each other and also on the translation rate (Braakman and Bullied, 2011).

Aims of the study

The Gp2/Gp3/Gp4 of EAV and PRRSV are known to form a disulphide linked complex, but very limited structural information is available about the minor glycoproteins of arteriviruses despite their important role in the cell entry. Hence, in this study I aimed to express the minor glycoprotein complex Gp2/Gp3/Gp4 of EAV and PRRSV in insect cells to identify the exact signal peptide cleavage sites, used N-linked glycosylation sites, and also the location of the disulphide-bonds that determine oligomerisation of the Gp2/Gp3/Gp4 complex. The goal was to purify the functional complex (if it was to be obtained) to study its secondary and, if feasible, tertiary structure.

The major envelope proteins Gp5/M also form a disulphide-linked complex, but the cysteine residue involved was determined experimentally only for Gp5/M of EAV (Snijder *et al.,* 2003). Therefore, I aimed to determine the cysteine residue involved in the oligomerisation of Gp5/M for the PRRSV virus. Gp5 is considered as one of the major targets for neutralising antibodies. However, the neutralising immune response is only seen late in infection, which is one of the reasons for the persistent viral infection. This delay was attributed to the presence of a 'decoy epitope' near the hypervariable region, which induces a strong non-neutralising immune response (Lopez and Osorio, 2004) (Fig.1.5). However, the 'decoy epiotope' is present in a potential signal peptide region and its presence depends on whether and where the signal peptide is cleaved. Therefore, in this study I aimed to determine the Gp5 signal peptide cleavage site from the Gp5 expressed in insect cells.

2 Materials and methods

2.1 Materials

2.1.1 Eukaryotic cells

- Sf9 cells (Life Technologies, Carlsbad, USA, #11496-015), derived from the pupal ovarian tissue of the fall army worm, *Spodoptera frugiperda,* suspension adapted to Sf-900™ II SFM (Serum-Free Media).
- TriEx™ Sf9 Cells (EMD Millipore, Massachusetts, USA # 71023), derivative of Sf9 cells adapted to serum free medium for the improved protein expression.

2.1.2 Bacteria

- *Escherichia coli* XL-1 blue (Stratagene/Agilent, Waldbronn): Genotyp *rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* [F´ *pro*AB *lac*I^qZ ΔM15 Tn*10* (Tet^r)]
- DH10MultiBac transformed with MultiBac bacmid (ATG biosynthetics, Merzhausen Germany)
- *Escherichia coli* pirHC (express pir gene required for propagation of vector with R6Kg origin)

2.1.3 Reagents and buffers

- Equilibration buffer: 20 mM sodium phosphate, 0.5 M NaCl, pH 7.4
- Wash buffer: 20 mM sodium phosphate, 0.5 M NaCl, 0 to 30 mM imidazole, pH 7.4
- Elution buffer: 20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, pH 7.4
- Buffer W (wash buffer): 100 mM Tris/HCl, pH 8.0, 150 mM NaCl , 1 mM EDTA
- Buffer E (elution buffer): 100 mM Tris/HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin
- TfB I buffer: 30 mM potassium acetate, 10 mM $CaCl_2$. $2H_2O$, 15% (W/V) Glycerin, 100 mM $RbCl_2$, 50 mM $MnCl_2$.4 H_2O, pH 6.2
- TfB II buffer: 10 mM MOPS, 75 mM $CaCl_2$. $2H_2O$, 10 mM $RbCl_2$, 15% (W/V) Glycerin, 1 N KOH pH 7,0

- TYM medium: for 1 litre 2% Bacto ™ Tryptone 20 g, 0.5% Bacto ™ Yeast extract 5 g, 0.1 M NaCl 5.8 g, 10 mM $MgSO_4$ 2.5 g
- Lysis buffer 1: 20 mM sodium phosphate, 0.5 M NaCl, pH 7.4, 1% NP-40 or 1% Triton X-100
- Lysis buffer II: 20 mM sodium phosphate, 0.5 M NaCl, pH 7.4, 1% n-dodecyl-ß-D-maltoside (DDM)
- Lysis buffer III: 50 mM $NaPO_4$, pH 8.0, 500 mM NaCl, 0.7% CHAPS, 0.3 % NP-40 and Roche Complete protease inhibitors).
-

2.1.4 Antibodies

- Monoclonal Anti-polyHistidine antibody produced in mouse (Sigma-Aldrich, Taufkirchen # H1029), recognises synthetic polyHistidine tagged protein, recommended dilution for western blotting is 1:3000

- Anti-HA tag antibody - ChIP Grade (Abcam, Cambridge, United Kingdom, #ab9110), recognises amino acids YPYDVPDYA (influenza hemagglutinin-HA-epitope) tagged to target proteins, recommended dilution for western blotting is 1:5000

- StrepMAB-Classic, HRP conjugate (IBA GmbH, Göttingen, Germany, # 2-1509-001), recognises *Strep*-tag® II tagged proteins

- Secondary antibody for western blotting: Anti-Mouse HRP conjugated (Sigma-Aldrich, Taufkirchen), recommended dilution for western blotting is 1:5000

- Goat Anti-Rabbit IgG H&L HRP conjugated (Abcam, Cambridge, United Kingdom, #ab6721), recommended dilution for western blotting is 1:5000.

2.1.5 Primers used in this study

Table 1. Sequences of the primers used in this study

Primer name	Primers for PCR with restriction sites for restriction enzymes	**Target**
pBAC1-Gp3-His FW	AGCTGCGGATCCGCCGCCACCATGGGTC GTGCCTACAGTG (*BamHI*)*	EAV Gp3

pBAC1-Gp3-His RV	ACTGCTAAGCTTTCAATGGTGATGGTGATG GTGATGTGAACCTCCAGCCTTGTCATCGTC ATCACCGCTACCTCCCCGTAGATCTTCATC TACAAAGG (*HindIII*)*	EAV Gp3
pIDK-GP2b-FW	ACTAGTCTCGAGGCCACCATGCAGCGCT TTTCTTTCTC (*XhoI*)*	EAV Gp2
pIDK-GP2b-HA- RV	CTGGAGGGTACCTTAGGCGTAATCAGGC ACGTCATAAGGGTAGCCACTTCCACCCG CTTTATCATCATCATCGCTGCCAGAGGA AGAAGCATAATAGTGCAA (*KpnI*)*	EAV Gp2
pIDS-GP4- FW	ACTAGTCTCGAGGCCACCATGAAGATCT ACGGCTGCATA (*XhoI*)*	EAV Gp4
pIDS-GP4-StrepII-RV	CTGGAGGGTACCTTACTTCTCGAACTGA GGATGTGACCAGCCACTTCCACCCGCTT TATCATCATCATCGCTGCCGTAAGGATG CCTCTTGTCCTC (*KpnI*)*	EAV Gp4
POLYH-FW	AAATGATAACCATCTCGC	Polyh Promoter
P10-FW	GACCTTTAATTCAACCCAACAC	P10 Promoter
Gp4-C26S-FW	TGCACTTTCTACCCAAGCCACGCTGCAG AGGC	EAV Gp4
Gp4-C26S-RV	GCCTCTGCAGCGTGGCTTGGGTAGAAAG TGCA	EAV Gp4
Gp4-C55S-FW	CACGGTCATGAGGGGAGTAGGAATTTTA TTA	EAV Gp4
Gp4-C55S-RV	TAATAAAATTCCTACTCCCCTCATGACC GTG	EAV Gp4
Gp4-C77S-FW	CGGCTATAACTCATTCTTTACTTCTGGTT CT	EAV Gp4
Gp4-C77S-RV	AGAACCAGAAGTAAAGAATGAGTTATA GCCG	EAV Gp4

PRRSV GP5-HIS-FW	ACTAGTGGATCCATGTTGGAGAAATGCTTGACC (*BamHI*)*	PRRSV Gp5
PRRSV GP5-HIS-RV	CTGTATAAGCTTTTAGTGATGGTGGTGATGGTGATGACTACCAGGACGACCCCATTGTTC (*HindIII*)*	PRRSV Gp5
pIDC-PRRSV M-HA- FW	ACGATCGGATCCATGGGGTCGTCCTTAGATGA (*BamHI*)*	PRRSV M
PRRSV M-HA-RV	CAGTCTCTGCAGTTAGGCATAATCAGGCACGTCATAAGGATAACTACCTTTGGCATATTTGACAAGGTTT (*PstI*)*	PRRSV M
PRRSV-GP5 DET- FW	ATGTTGGAGAAATGCTTGACC	PRRSV Gp5 detection primer
PRRSV GP5 DET- RV	AGGACGACCCCATTGTTCC	PRRSV VR-2332 GP5 detection primer
pIDK-GP2a-HA- FW	ACTAGTCTCGAGGCCGCCACCATGAAATGGGGTCCATGC (*XhoI*)*	PRRSV VR-2332 GP2A Protein
pIDK-GP2a-HA- RV	CTGGAGGGTACCTTAGGCGTAATCAGGCACGTCATAAGGGTAGCCACTTCCACCCGCTTTATCATCATCATCTCCGCTGCCTTGCTGAAAATCATGAAG CTTTG (*KpnI*)*	PRRSV VR-2332 GP2A Protein
pBAC1-GP3-HIS- FW	AGCTGCGGATCCGCCGCCACCATGGTTAATAGCTGTACATTCCTCCA (*BamHI*)*	PRRSV VR-2332 GP3 Protein
pBAC1-GP3-HIS- RV	ACTGCTAAGCTTTCAATGGTGATGGTGATGGTGATGGTGATGTGAACCTCCAGCCTTGTCATCGTCATCACCGCTACCTCCTTCTAGGTG AAACCAATTGCC (*HindIII*)*	PRRSV VR-2332 GP3 Protein
pIDS-GP4 Onestrep- FW	ATCAGTGCTAGCCTCGAGGCCGCCACCATGGCTTCGTCCCTTCTTTTC (*XhoI*)*	PRRSV VR-2332 GP4 Protein

pIDS- GP4-Onestrep - RV	ACTTCTGGTACCTCACTTCTCGAATTGAGG GTGT (*KpnI*)*	PRRSV VR-2332 GP4 Protein
GP5-C48S-FW	CAACCTGACCCTGAGCGAGCTGAACGG	PRRSV VR-2332 GP5 Mutagenic primer
GP5-C48S-RV	CCGTTCAGCTCGCTCAGGGTCAGGTTG	PRRSV VR-2332 GP5 Mutagenic primer

*Restriction enzyme sequences are underlined and the name of the restriction enzyme is mentioned in the brackets

2.1.6 Protein sequences of expression constructs used in this study

EAV ΔC-Gp2b-HA Construct sequence for soluble expression

```
MQRFSFSCYL HWLLLLCFFS GSLLPSAAAW WRGVHEVRVT DLFKDLQCDN
LRAKDAFPSL GYALSIGQSR LSYMLQDWLL AAHRKEVMPS NIMPMPGLTP
DCFDHLESSS YAPFINAYRQ AILSQYPQEL QLEAINCKLL AVVAPALYHN
YHLANLTGPA TWVVPTVGQL HYYASSSGSD DDDKAGGSGG YPYDVPDYA* stop
```

EAV ΔC-Gp4-strep II Construct sequence for soluble expression

```
MKIYGCISGL LLFVGLPCCW CTFYPCHAAE ARNFTYISHG LGHVHGHEGC
RNFINVTHSA FLYLNPTTPT APAITHCLLL VLAAKMEHPN ATIWLQLQPF
GYHVAGDVIV NLEEDKRHPY GSDDDDKAGG SGWSHPQFEK* stop
```

EAV ΔC-Gp3-HIS Construct sequence for soluble expression

```
MGRAYSGPVA LLCFFLYFCF ICGSVGSNNT TICMHTTSDT SVHLFYAANV
TFPSHFQRHF AAAQDFVVHT GYEYAGVTML VHLFANLVLT FPSLVNCSRP
VNVFANASCV QVVCSHTNST TGLGQLSFSF VDEDLRGGSG DDDDKAGGSG
HHHHHHH* stop
```

EAV Gp3-His full-length construct sequence

```
MGRAYSGPVA LLCFFLYFCF ICGSVGSNNT TICMHTTSDT SVHLFYAANV
TFPSHFQRHF AAAQDFVVHT GYEYAGVTML VHLFANLVLT FPSLVNCSRP
VNVFANASCV QVVCSHTNST TGLGQLSFSF VDEDLRLHIR PTLICWFALL
LVHFLPMPRC RGSGDDDDKA GGHHHHHHH* stop
```

PRRSV ΔC-Gp2a-HA construct sequence for soluble expression

```
MKWGPCKAFL TKLANFLWML SRSSWCPLLI SLYFWPFCLA SPSPVGWWSF
```

ASDWFAPRYS VRALPFTLSN YRRSYEAFLS QCQVDIPTWG TKHPLGMLWH
HKVSTLIDEM VSRRMYRIME KAGQAAWKQV VSEATLSRIS SLDVVAHFQH
LAAIEAETCK YLASRLPMLH NLRMTGSNVT IVYNSTLNQV FAIFPTPGSR
PKLHDFQQGS GDDDDKAGGS GYPYDVPDYA* stop

PRRSV ΔC-Gp3-HIS construct sequence for soluble expression

MVNSCTFLHI FLCCSFLYSF CCAVVAGSNT TYCFWFPLVR GNFSFELTVN
YTVCPPCLTR QAATEIYEPG RSLWCRIGYD RCGEDDHDEL GFMIPPGLSS
EGHLTGVYAW LAFLSFSYTA QFHPEIFGIG NVSRVYVDIK HQLICAEHDG
QNTTLPRHDN ISAVFQTYYQ HQVDGGNWFH LGGSGDDDDK AGGSHHHHHH
HHH*stop

PRRSV ΔC-Gp4-one strep construct sequence for soluble expression

MASSLLFLVV GFKCLLVSQA FACKPCFSSS LADIKTNTTA AASFAVLQDI
SCLRHRDSAS EAIRKIPQCR TAIGTPVYVT ITANVTDENY LHSSDLLMLS
SCLFYASEMS EKGFKVVFGN VSGIVAVCVN FTSYVQHVKE FTQRSLVVDH
VRLLHFGGSS GPISGPFEDD DDKAGWSHPQ FEKGGGSGGG SGGGSWSHPQ
FEK*stop

PRRSV VR-2332 Gp5-His construct sequence

MLEKCLTAGC CSRLLSLWCI VPFCFAVLAN ASNDSSSHLQ LIYNLTLCEL
NGTDWLANKF DWAVESFVIF PVLTHIVSYG ALTTSHFLDT VALVTVSTAG
FVHGRYVLSS IYAVCALAAL TCFVIRFAKN CMSWRYACTR YTNFLLDTKG
RLYRWRSPVI IEKRGKVEVE GHLIDLKRVV LDGSVATPIT RVSAEQWGRP
GSHHHHHHH*stop

PRRSV VR-2332 M-HA construct sequence

MGSSLDDFCH DSTAPQKVLL AFSITYTPVM IYALKVSRGR LLGLLHLLIF
LNCAFTFGYM TFAHFQSTNK VALTMGAVVA LLWGVYSAIE TWKFITSRCR
LCLLGRKYIL APAHHVESAA GFHPIAANDN HAFVVRRPGS TTVNGTLVPG
LKSLVLGGRK AVKQGVVNLV KYAKGSYPYD VPDYA*stop

* Affinity tag sequence is underlined

2.1.7 Plasmids constructed for this study

Table 2. Overview of vectors cloned for this study.

Plasmid name	Properties
pIDS Gp2b ΔC –HA[a]	pIDS backbone : 2231 bp Gp2b ΔC -HA : 531 bp Ectodomain of EAV Gp2b with HA tag Spectinomycin resistance
pACEBAC1 **ΔC**-Gp3-His [a]	pACEBAC1 backbone : 2904 bp Gp3 ΔC-His : 540 bp

	Ectodomain of EAV Gp3 with His tag Gentamycin resistance
pACEBAC1 **ΔC**-Gp3 N28/29S-His [a]	pACEBAC1 backbone : 2904 bp Gp3 ΔC-His : 540 bp Ectodomain of EAV Gp3 with His tag Aspargine 28/29 mutated to serine Gentamycin resistance
pACEBAC1 Gp3 full-His [a]	pACEBAC1 backbone : 2904 bp Gp3 full-His : 1441 bp EAV Gp3 full length gene Gentamycin resistance
pACEBAC1 **ΔC**-Gp3-His PRRSV [b]	pACEBAC1 backbone : 2904 bp ΔC-Gp3 -His : 624 bp Gentamycin resistance
pIDS **ΔC**-Gp4-strep II [a]	pIDS backbone : 2231 bp ΔC-Gp2b-HA : 360 bp Ectodomain of EAV Gp4 with strep II tag Spectinomycin resistance
pIDK **ΔC**-Gp2b–HA PRRSV [b]	pIDK backbone : 2281 bp PRRSV ΔC-Gp2b -HA : 702 bp Kanamycin resistance
pIDK **ΔC**-Gp4-one strep PRRSV [b]	pIDK backbone : 2231 bp ΔC-Gp4-one strep : 612 bp Ectodomain of PRRSV Gp4 with one strep tag Kanamycin resistance
pACEBAC1-Gp5-His [b]	pACEBAC1 backbone : 2904 bp Gp5-His : 630 bp EAV Gp5 with His tag Gentamycin resistance
pIDC M-HA [b]	pIDC backbone : 2346 bp M-HA : 558 bp PRRSV M with HA tag Chloramphenicol resistance
pIDS- **ΔC**-Gp4 StrepII -C48S [a]	pACEBAC1 backbone : 2904 bp EAV Gp4-Strep II : 360 bp Gentamycin resistance Cysteine 48 was mutated to serine
pIDS- **ΔC**-Gp4 StrepII-C26S [a]	pIDS backbone : 2231 bp Gp4 ΔC –strep II : 360 bp Spectinomycin resistance Cysteine 26 was mutated to serine
pIDS- **ΔC**-Gp4 StrepII-C50S [a]	pIDS backbone : 2231 bp Gp4 ΔC – strep II : 360 bp

	Spectinomycin resistance Cysteine 50 was mutated to serine
pIDS- **ΔC**-Gp4 StrepII-C77S [a]	pIDS backbone : 2231 bp Gp4 ΔC - strep II : 360 bp Spectinomycin resistance Cysteine 77 was mutated to serine
pIDS- **ΔC**-Gp4 StrepII mut -C50/77S [a]	pIDS backbone : 2231 bp Gp4 ΔC Strep II : 360 bp Spectinomycin resistance Cysteine 50 and 77 was mutated to serine
pIDS- **ΔC**-Gp4 StrepII mut -C26/77S [a]	pIDS backbone : 2231 bp Gp4 ΔC Strep II : 360 bp Spectinomycin resistance Cysteine 26 and 77 was mutated to serine
pIDS- **ΔC**-Gp4 StrepII mut -C26/50S [a]	pIDS backbone : 2231 bp Gp4 ΔC Strep II : 360 bp Spectinomycin resistance Cysteine 26 and 50 was mutated to serine

a : PRRSV VR2332 strain b : EAV Busyrus strain

2.2 Methods

2.2.1 Culture conditions for *E.coli* cells

The culture of *E. coli* bacteria was grown in liquid LB medium (Luria-Bertani medium, containing 10 g/l Bacto Tryptone, 5 g/l Bacto Yeast Extract, and 10 g/l NaCl at pH 7, obtained from Life Technologies, Carlsbad, USA, in a powder form). 35 grams of the LB powder were dissolved and autoclaved at 121°C for 15 minutes. LB-agar plates were made by adding 1.5% agarose (Life Technologies, Carlsbad, USA) to the LB media before autoclaving. Bacterial cultures were grown overnight at 37 °C with 220 rpm shaking in a shaking incubator (Sartorius, Göttingen, Germany). *E. coli* glycerol stocks were prepared for long-term storage by adding sterile (v/v) glycerol (VWR, Germany) to 10% final concentration and stored at -80°C.

2.2.2 Preparation and transformation of *E.coli*

E.coli cells for heat shock transformation were prepared using a high competence protocol (Inoue *et al.,* 1990). A single colony of *E.coli* was inoculated into 50 ml of TYM medium and cultured overnight at 37 °C with 180 rpm shaking. 5 ml of overnight culture were inoculated with 500 ml of TYM medium and incubated at 37 °C with 180 rpm shaking

until OD^{600} reached 0.4. The culture was cooled down quickly by placing the flask in cold water for 10 minutes and centrifuged for 10 minutes with 1000x g with slow deceleration in an ultracentrifuge (Thermo Scientific, USA). The pellet was resuspended in 45 ml cold Tfb I buffer and incubated on ice for 10 minutes and centrifuged for 8 minutes at 1500 x g at 4°C. The pellet was resuspended in 10 ml ice-cold Tfb II solution. The cells were then flash-frozen in liquid nitrogen and stored at -80°C.

Transformation of DNA into *E. coli* cells was performed by the heat shock method (Sambrook and Russell 2001). 5 µl of the plasmid DNA (5 ng/µl) were added to 100 µl of chemically competent *E. coli.* The cells were incubated for 30 minutes on ice and then exposed to a heat shock for 45 seconds at 42 ^{0}C. After the heat shock, the suspension was immediately placed on ice for five minutes, and 900 µl of SOC was added. This culture was incubated for 60 minutes at 37 ^{0}C on shaker and plated on appropriate LB agar antibiotic plates after one hour.

2.2.3 Isolation of plasmid DNA from *E.coli*

For small-scale plasmid DNA preparation, Invisorb® Spin Plasmid Mini Two kit (STRATEC Molecular GmbH, Berlin, Germany) was used, and plasmid DNA was isolated using the manufacturer's protocol. 1.5 ml of the overnight *E. coli* culture were transferred into a 2.0 ml microcentrifuge tube, and the bacterial cells were pelleted by centrifuging at 16000 X g for 1 min. Supernatant was removed, and the pellet was resuspended in 250 µl of solution A. 250 µl of solution B was added, and the tube was gently inverted 4 to 5 times. 250 µl of solution C was added, and the tube was immediately inverted five to ten times and centrifuged at 16,000 x g for 5 minutes. The clarified supernatant containing the plasmid DNA was transferred to the spin filter and incubated for 1 min. The spin filter was centrifuged at 11,000 X g for 1 minute, and the flow-through was discarded. 750 µl of wash solution were added and centrifuged for one minute at 16,000 X g, and the flow-through was discarded. The spin filter was again centrifuged for 3 minutes at 16,000 X g to remove the residual ethanol. The plasmid DNA was eluted with 75 µl of elution buffer by centrifuging at 11,000 x g for 1 minute.

Large-scale plasmid DNA purification was done by using PureYield™ Plasmid Maxiprep System (Promega, Wisconsin, USA) according to the manufacturer's protocol. 200 ml of transformed *E. coli* was grown overnight and pelleted by centrifugation at 5000x g for 10 minutes. Cell pellet was resuspended in 12 ml of resuspension buffer, followed by 12 ml of lysis buffer. The tube was gently inverted 3 to 5 times to mix and incubate at room

temperature for 3 minutes. After this incubation, 12 ml of neutralisation buffer was added and inverted 15 times to mix. The supernatant was separated from the cell debris by centrifuging the lysate at 14,000 x g for 20 minutes. The clarified lysate was applied to the PureYield™ Maxi Binding Column, and the lysate was cleared by applying vacuum. The column was washed with 5 ml of the endotoxin removal buffer, and the solution was pulled through the column by applying vacuum. After this step, the column was washed with 20 ml of wash buffer, and the wash buffer was pulled by vacuum. The column was dried by applying vacuum for 5 minutes, and the plasmid DNA was eluted with 1 ml of nuclease-free water.

2.2.4 Culture conditions for insect and mammalian cells

Suspension-adapted *Sf9* (Life Technologies, Carlsbad, USA) and *TriEx Sf9* (EMD Millipore, Massachusetts,USA) cells were routinely cultured in Erlenmeyer flasks (BD, New Jersey, USA) in Sf-900 II and Sf-900 III serum-free medium (Life Technologies, Carlsbad, USA) in an orbitally-shaking 120 rpm incubator (Sartorius, Göttingen, Germany) at 27 ^{0}C.

2.2.5 Polymerase chain reaction (PCR)

PCR reactions for the amplification of DNA sequences such as open reading frames (ORF) and site-directed mutagenesis was carried out by the following protocols. A high-fidelity Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs, Massachusetts, USA) was used to perform the ORF amplification and site-directed mutagenesis; for the analytical PCR reaction low-fidelity one Taq polymerase (New England Biolabs, Massachusetts, USA) was used. PCR reactions were set up according the following protocol (Table 3). PCR program 1 (Table 4) was used for the Q5® Hot Start High-Fidelity 2X Master Mix, and PCR program 2 (Table 5) was used for one Taq polymerase.

Table 3: Components for the PCR reaction for the amplification of DNA sequences

Component	50 µl reaction	Final concentration
Q5 High-Fidelity 2X Master Mix	25 µl	1 X
10 µM Forward Primer	2.5 µl	0.5 µM

10 µM Reverse Primer	2.5 µl	0.5 µM
Template DNA	4 µl (10 ng/ µl)	40 ng
Nuclease-Free Water	To 50 µl	

Table 4: PCR program 1 for the Q5® Hot Start High-Fidelity 2X Master Mix

step	[^{0}C]	[hh.mm.ss]	Number of cycles
Initial Denaturation	98	00.00.30	1
Denaturation Annealing Extension	98 70 70	00.00.10 00.00.30 30 sec/kb	] 27
Final extension	72	00.02.00	
hold	8	∞	

Table 5: PCR program 2 for the one Taq polymerase

step	[^{0}C]	[hh.mm.ss]	Number of cycles
Initial Denaturation	94	00.00.30	1
Denaturation Annealing Extension	94 60 68	00.00.30 00.00.45 60 sec/kb	] 29
Final extension	68	00.05.00	
hold	8	∞	

2.2.6 Restriction digestion

Restriction digestion is used to generate compatible ends from PCR products or genomic DNA, which has recognition sites for the corresponding enzymes that are capable of being ligated together. 1 to 4 µg of DNA were digested in 50 µl reactions by incubating at 37 ^{O}C for 2 hours. After the digestion, the DNA was run on 1% agarose gel to separate the desired DNA fragments from the other fragments. The essential DNA fragment was excised from the gel and purified using Invisorb Spin DNA Extraction (STRATEC Molecular GmbH, Berlin, Germany) according to the manufacturer's protocol.

2.2.7 Cloning of Gp2b, Gp3 and Gp4 ectodomains of EAV

The coding regions of minor glycoproteins of EAV were amplified from full-length clone pEAV 211, which contains the full genome of EAV by polymerase chain reaction (PCR) and cloned into Multibac vectors (ATG biosynthetics, Merzhausen Germany). HA tag (YPYDVPDYA) was attached to the carboxy-terminus of Gp2b by PCR using primers pIDK-Gp2b-HA-FW and pIDK-Gp2b-HA-RV (table.1), and cloned into pIDK vector under the control of p10 promoter. A 9X His tag (HHHHHHHHH) was attached to the carboxy-terminus of Gp3 by PCR using primers pBAC1-Gp3-His-FW and pBAC1-Gp3-His-RV (table.1), and cloned into pACEBAC1 vector under the control of polyhedrin promoter. A Strep II tag (WSHPQFEK) was attached to the carboxy-terminus of Gp4 by PCR using primers pIDS-Gp4-strep II-FW and pIDS-Gp4-strep II-RV (table.1), and cloned into pIDS vector under the control of p10 promoter

2.2.8 Cloning of minor and major glycoproteins of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV)

The coding regions of minor glycoprotein ectodomains of PRRSV strain VR-2332 were amplified from full length clone pVR-V7, (Han *et al.,* 2007, *J Virol*) which contains the full genome of PRRSV by polymerase chain reaction (PCR) and cloned in Multibac vectors (ATG Biosynthetics, Merzhausen Germany). HA tag (YPYDVPDYA) was attached to the carboxy-terminus of Gp2a by PCR using primers pIDK-GP2a-HA-FW and pIDK-GP2a-HA-RV (Table 1), and cloned into pIDK vector under the control of p10 promoter. A 9X His tag (HHHHHHHHH) was attached to the carboxy-terminus of Gp3 by PCR using primers pBAC1-GP3-His-FW and pBAC1-GP3-His-RV (Table 1) and cloned into pACEBAC1 vector under the control of polyhedrin promoter. One Strep-tag (WSHPQFEK) was attached to the carboxy-terminus of Gp4 by PCR using primers pIDS-

GP4-onestrep-FW and pIDS-GP4-onestrep-RV (Table 1) and cloned into pIDS vector under the control of p10 promoter. A 7X His tag (HHHHHH) was attached to the carboxy-terminus of Gp5 by PCR using primers pBAC1-Gp5-His-FW and pBAC1-Gp5-His-RV (Table 1) and cloned into pACEBAC1 vector under the control of polyhedrin promoter. HA tag (YPYDVPDYA) was attached to the carboxy-terminus of M by PCR using primers pIDC-M-HA-FW and pIDC-M-HA-RV (Table 1) and cloned into pIDC vector under the control of p10 promoter.

2.2.9 Agarose gel electrophoresis

DNA fragments such as PCR products and gene fragments, cut out from plasmids after the restriction digestion, are separated in an agarose gel according to their size. Samples were treated with loading buffer, loaded onto agarose gel (0.8% and 1.0% (w/v) for small fragments <1,500) in 1X TAE buffer with 0.5 ug / mL ethidium bromide, and electrophoresed at a constant voltage of 100 to 120 volts. 1 Kb plus ladder (Thermo Fisher Scientific, Massachusetts, USA) was used as a size marker. The DNA bands on the gels were visualised with a UV lamp (excitation wavelength: 302 nm). The required DNA bands were cut with a scalpel. The DNA was extracted from the gel fragments by Invisorb Spin DNA Extraction according to the supplier's protocol and eluted with 20 µl elution buffer provided in the kit.

2.2.10 Ligation

Cohesive (sticky) end ligation was performed by adding the components mentioned in the Table 6.

Table 6: Components of ligation mixture

Component	20 µl reaction
Vector DNA	50 ng
Insert DNA	20 ng
10 X T4 DNA ligase buffer	2 µl
T4 DNA ligase	1 µl
Nuclease free water	To 20 µl

Ligation mixture was prepared by mixing the components mentioned in the Table 6 and incubated at room temperature for an hour. 100 µl of competent cells were transformed with 5 µl of ligated mixture.

2.2.11 Site directed mutagenesis of GP4-strep II

Site directed mutagenesis was performed using GeneArt® Site-Directed Mutagenesis System (Life Technologies, Carlsbad, USA) according to the manufacturer's protocol. pIDS-Gp4-strep II vector was used to generate the Gp4-strep II cysteine to serine mutations. Gp4-strepII-C26S mutation was generated by using mutagenic primers Gp4-C26S-FW and Gp4-C26S-RV (Table 1). Gp4-strepII-C50S mutation was generated by using mutagenic primers Gp4-C50S-FW and Gp4-C50S-RV (Table 1) and Gp4-strepII-C77S mutation was generated by using mutagenic primers Gp4-C77S-FW and Gp4-C77S-RV (Table 1).

2.2.12 Fusion of acceptor and donor vectors

Acceptor and donor vectors that harbour the gene of interest are fused with Cre-LoxP recombination. For 20 µl reaction, 1.5 µg of each vector were mixed in equal amounts, the total volume was adjusted to 16 µl, and then 2 µl of 10 X cre buffer and 2 µl of cre recombinase (New England Biolabs, Frankfurt am Main, Germany) were added. The reaction was then incubated at 37°C for an hour, and 15 µl of this reaction were transformed into 200 µl of *E.coli* by heat shock method. After the transformation, 1 ml of SOC medium was added to the transformed cells and incubated overnight at 37 °C for the recovery. The recovered cell suspension was plated on LB agar plates containing preferred combination of antibiotics and incubated overnight. Resultant fusion vector DNA was isolated, and the fusion of acceptors and donor vectors was verified by the sequencing.

2.2.13 Transposition of bacmid DNA

Approximately 50 ng of sequenced pACEBac1 vector or pACEBac1 fusion vector with the donor vectors were added to 100 µl DH10*Multi*Bac*Turbo* cells and transformed using heat shock method. Cells were incubated overnight at 37°C, and the resultant cell suspension was plated onto LB agar plates containing kanamycin (50 µg/ml), gentamycin (7 µg/ml), tetracyclin (10 µg/ml), X-Gal (100 µg/ml), and IPTG (40 µg/ml). Additional antibiotics were included in the plates depending on the antibiotic resistance of the acceptor-donor vector combination.

2.2.14 Isolation of bacmid DNA

Recombinant bacmid DNA was isolated using BACMAX ^{TM} DNA purification kit (EPICENTER, Madison, USA) according to the manufacturer’s protocol.

A single white colony was inoculated into 10 ml of LB media supplemented with kanamycin (50 µg/ml), gentamycin (7 µg/ml), tetracyclin (10 µg/ml), and additional antibiotics depending on the acceptor-donor combination used. Cells were pelleted from the overnight culture by centrifuging at 5000 X g for 8 min at 4°C. The cell pellet was dissolved in 1 ml of chilled BACMAX solution 1, and 2 ml of BACMAX solution 2 were added to the tube for the cell lysis. After the lysis step, 1.5 ml of BACMAX solution 3 were added and the tube was inverted immediately four times. The suspension was centrifuged at 15,000 X g for 15 minutes at 4°C after incubation for 15 min on ice. 2.7 ml of room temperature isopropanol was added to the recovered supernatant, and the nucleic acids were precipitated by centrifuging the supernatant at 15000 X g for 15 min at 4°C. Isopropanol was removed, and the pellet was air-dried at room temperature for 5 minutes. The pellet was resuspended in 250 µl of TE buffer and incubated at 37 °C for 30 minutes after adding 6 µl of RiboShredder RNase blend. After this digestion, an equal volume of the suspension was transferred to two different tubes, and 250 µl of chilled BACMAX solution 4 was added. This solution was incubated on ice and centrifuged at 15000 X g for 15 minutes at 4 °C. The supernatant was transferred to a new tube. The bacmid DNA was precipitated by adding 1 ml of absolute ethanol to the supernatant and centrifuging at 15000 X g for 15 minutes at 4 °C. The resultant DNA pellet was air-dried at room temperature for 5 minutes and resuspended in 50 µl of TE buffer.

2.2.15 Transfection of *Sf9* cells with bacmid

Sf9 cells were transfected with recombinant bacmid DNA by using cellfectin II (Life Technologies, Carlsbad, USA) according to the supplier's protocol.

*Sf*9 insect cells (8×10^5 cells) were seeded in a six-well plate in 2 ml of plating media (8.5% unsupplemented Grace's insect media 1.5% Grace's insect medium supplemented with 10% FBS) and allowed to attach for 15 min at room temperature. 8 µl of cellfectin II reagent were diluted in 100 µl unsupplemented Grace's insect media,; similarly, 2 µg of bacmid DNA were diluted in 100 µl of unsupplemented Grace's insect media. Diluted DNA was combined with diluted cellfectin-II solution and incubated at room temperature for 30 minutes. The DNA/transfection reagent mixture (~210 µl) was added to the cells drop-wise and incubated at 27 °C for 5 hours. The transfection media was replaced with sf-900 II growth media (Life Technologies, Carlsbad, USA) after the incubation. The recombinant baculoviruses (P0 stock) were harvested three days post-transfection and stored at 4 °C.

2.2.16 Baculovirus titration using qRT PCR

Baculoviruses were titrated using baculoQUANT ALL-IN-ONE DNA extraction and quantification kit (Oxford Expression Technologies, Oxford, UK) according to the manufacturer's protocol.

Viral DNA was isolated from 80 µl of recombinant budded virus stock. The virus was pelleted by centrifugation at 16,000 X g for five minutes, and the supernatant was carefully removed. The pellet was dissolved in 20 µl of lysis buffer by vortexing, and the viral DNA was extracted using the following lysis program in the thermal cycler:

Step 1: 65 ^{o}C for 15 min.

Step 2: 96 ^{o}C for 2 min.

Step 3: 65 ^{o}C for 4 min.

Step 4: 96 ^{o}C for 1 min.

Step 5: 65 ^{o}C for 1 min.

Step 6: 96 ^{o}C for 30 sec.

Step 7: 20 ^{o}C hold

The virus internal standard (~10^8 pfu/ml) was diluted in a serial log (1 in 10) dilution up to 10^3 pfu/ml and the viral DNA was extracted by similar procedure mentioned above.

QPCR master mix was prepared according to the protocol mentioned in the Table 7.

Table 7. Components of qRT-PCR master mix

Reagent	1 QPCR reaction
QPCR low ROX mix	12.5 µl
RNase free water (QPCR)	7.5 µl
Probe + primers solution	3 µl
Total volume	23 µl

23 µl of master mix were aliquoted into each PCR tube, and 2 µl of purified Viral DNA that were added to perform titration. The DNA amplification was done in Bio-Rad iQ5 QPCR system using the program shown below (Table 8)

Table 8. PCR program for the qRT-PCR

	Temperature	Time	Number of cycles
Enzyme activation	95 ^{o} C	10 min	1 cycle
Denaturation	95 ^{o} C	15 sec	40 cycles
Annealing/Extension	60 ^{o} C	60 sec	

A standard curve was generated using virus internal standard Ct values vs. pfu/ml. Unknown virus Ct values were plotted against the standard curve, and a pfu/ml value was obtained for the each of the unknown virus.

2.2.17 Amplification of virus stocks and harvesting

For the amplification of recombinant baculovirus, 2 x 10^6 *TriEx Sf9* cells were infected with P0 virus with multiplicity of infection (M.O.I) of 0.01, and the cells were cultured at 27 ^{o}C with 120 rpm shaking for 3 days. Three days post-infection the supernatant containing the recombinant virus stock (P1 stock) was harvested by centrifuging at 1500X g for 10 minutes. Virus titration was performed by using baculoQUANT ALL-IN-ONE DNA extraction and quantification kit.

2.2.18 Expression of protein complexes in TriEx *Sf9* cells

2 x 10^6/ ml TriEx *Sf9* cells were infected with M.O.I (multiplicity of infection) 2 and the cells were cultured at 27 ^{o}C with 120 rpm shaking for three days. Three days post-infection the supernatant was harvested (for secreted proteins) by centrifuging the cells at 1500 rpm for 10 minutes. For the intracellular proteins, the cells were pelleted and stored at -80 ^{o}C until further use.

2.2.19 Intracellular protein extraction from insect cells

Cell number was determined prior to the lysis, and cells were harvested by centrifugation at 800 x g for five minutes at room temperature. Cells were washed with PBS and

centrifuged for 5 minutes at 800 x g. One ml of lysis buffer per 2 x 10^7 cells was added. The cells were resuspended by pipetting and incubated on ice for 20 minutes. After the incubation cells were centrifuged at 15,000 X g at 4 ^{o}C for 20 minutes, and the supernatant containing the soluble protein was transferred to a fresh tube and stored at -80 ^{o}C until further use.

2.2.20 Enrichment of strep-tagged proteins

The Strep-Tactin Superflow column (IBA GmbH, Göttingen, Germany) was equilibrated by adding 2 column volumes of buffer W, and the sample that contains strep-tagged protein was applied to the column. The column was washed with 5 column volumes of buffer W, and the protein was eluted using six column volumes of buffer E.

2.2.21 Enrichment of His tagged proteins

Secreted his-tagged proteins were enriched from supernatant by using HisTrap™ Excel (GE Healthcare, Uppsala, Sweden) according to the manufacturer's protocol.

The column was equilibrated with five column volumes (CV) of equilibration buffer, and the sample containing his-tagged protein was applied to the column. Column was then washed with 20 CV of wash buffer, and the target protein was eluted with five CV of elution buffer.

2.2.22 Removal of N-linked glycans with PNGase F

Glycoproteins were digested with the PNGase F (New England Biolabs, USA) to remove the glycans, and the glycan digestion was performed according to the manufacturer's protocol. Protein sample was mixed with 10% NP-40 to a final concentration of 1.0% and boiled for 10 min at 100 °C for 10 minutes. The sample was cooled down to room temperature, and G7 buffer (0.05 M sodium phosphate, pH 7.5) was added to a final concentration of 1X. 1 µl of PNGase F was added to the reaction and incubated at 37 °C for 2 hours. After the digestion, 1X SDS-PAGE sample buffer was added to the sample and boiled for 5 min.

2.2.23 SDS-PAGE and Western blot

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed using Bolt® 4-12% Bis-Tris Plus Gel for western blot (Invitrogen, Carlsbad, USA) and 4-20% Tris-Glycine Gels for Coomassie staining (Pierce/Thermo, Bonn, Germany). After

SDS-PAGE, gels were either stained with Imperial Protein Stain (Pierce/Thermo, Bonn, Germany) or blotted onto polyvinylenedifluoride (PVDF) membrane (GE Healthcare, Freiburg im Breisgau, Germany) using standard protocol. After blocking membranes in the blocking solution (5% skim milk powder in PBS with 0.1% Tween-20) for 1 h at room temperature, primary antibodies were put in blocking solution for 1 hour at room temperature. Rabbit-anti-HA tag antibody (ab9110, Abcam, Cambridge, UK, 1:5,000) was used to detect HA-tagged proteins, and mouse-anti-His tag antibody (H1029, Sigma-Aldrich, Taufkirchen, Germany, 1:3,000) was employed for His tagged proteins. After washing (2×10 min with PBST), horseradish peroxidase-coupled secondary antibody anti-rabbit for HA tag and anti-mouse for His tag (Sigma-Aldrich, Taufkirchen, Germany, 1:5,000) was put on for 1 hour at room temperature. After three washes for 10 minutes, signals were detected by chemiluminescence using the ECL plus reagent (Pierce/Thermo, Bonn, Germany) and a Fusion SL camera system (Peqlab, Erlangen, Germany). Strep-tagged proteins were detected by using StrepMAB-Classic, HRP conjugate antibody (IBA GmbH, Goettingen, Germany). The western blot was performed according to the supplier's protocol.

2.2.24 Protein sample preparation for mass spectrometry

After SDS-PAGE separation of proteins, protein bands were excised and washed with 50% (v/v) acetonitrile in 50 mM ammonium bicarbonate/$(NH_4)HCO_3$. The protein bands were shrunk by dehydration in acetonitrile and dried in a vacuum centrifuge. Disulphide bonds were reduced by incubation in 60 µL of 10 mM DTT in 50 mM $(NH_4)HCO_3$ for 45 min at 56 °C. Alkylation was performed by replacing the DTT solution with 55 mM iodoacetamide in 50 mM $(NH_4)HCO_3$. The gel pieces were shrunk by dehydration in acetonitrile, dried in a vacuum centrifuge, re-swollen in 20 µL of 50 mM $(NH_4)HCO_3$ containing 100 ng trypsin (Promega, Madison, WI, USA), and incubated at 37°C overnight. In the case of chymotrypsin, excised protein bands were incubated with 110 ng of enzyme (Roche Diagnostics, Penzberg, Germany) in 20 µL of 50 mM $(NH_4)HCO_3$ for 20 h at 25 °C. Peptides were extracted using 20 µL of 0.5% (v/v) trifluoroacetic acid (TFA) in acetonitrile, and the separated liquid was dried under vacuum. The samples were reconstituted in 6 µL of 0.1% (v/v) TFA and 5% (v/v) acetonitrile in water.

2.2.25 Liquid chromatography-tandem mass spectrometry

LC-MS/MS analyses were performed on an Orbitrap Elite hybrid mass spectrometer (Thermo Fisher, Bremen, Germany) equipped with an UltiMate 3000 LC (Dionex). LC separations were performed on a capillary column (Acclaim PepMap100, C18, 2 µm, 100 Å, 150 mm×75 µm i.d., Dionex) at an eluent flow rate of 200 nL/min using a linear gradient of 3–30% eluent B for 33 min with further increase to 80% B at 40 min. Mobile phase A contained 0.1% (v/v) formic acid in water, and mobile phase B contained 0.1% (v/v) formic acid in acetonitrile. MS data were acquired in a data-dependent strategy selecting MS/MS fragmentation events based on the precursor abundance in the MS scan. LTQ MS/MS spectra were acquired with a target value of 20,000 ions. The maximum injection time for MS/MS was 300 ms, and the dynamic exclusion time was 30 s. MS and MS/MS spectra were used to search against a custom-made database containing all proteins of the SwissProt 2010_7 database (521,024 sequences; 183,901,752 residues). In addition, the database contains all possible *N*-terminally truncated sequences of the target proteins resulting from signal peptide cleavage site prediction in the range between residue 20 and 40. Asn/Asp amino acid exchanges were used as variable modifications. For identification of target protein peptides, the processed MS/MS spectra were compared with the theoretical fragment ions of target protein peptides using the MASCOT server version 2.2.2 (Matrix Science Ltd., London, UK). The maximum of two and six missed cleavages was allowed for tryptic and chymotryptic peptides, respectively. The mass tolerance of precursor and sequence ions was set to 10 ppm and 0.35 Da, respectively.

3. Results

3.1 Expression and characterisation of Gp2/Gp3/Gp4 of EAV and PRRSV for structural and functional studies

3.1.1 Expression of Gp2, Gp3, and Gp4 complex from EAV in insect cells

Minor glycoproteins Gp2/Gp3/Gp4 exist as trimeric complex in EAV virions. This complex is demonstrated to be essential for virus infectivity (Wieringa *et al.,* 2003). Hence, I wanted to express Gp2/Gp3/Gp4 complex and characterise the proteins for structural and functional studies. I tried to express Gp2/Gp3/Gp4 as trimeric complex in insect cells using baculovirus expression system. The hydrophobic regions of all three glycoproteins (Gp2/Gp3/Gp4) at the C-terminus were removed to enable their secretion into the cell culture supernatants. A C-terminal Strep II tag was added to Gp4, a HA tag to Gp2, and a His tag to the Gp3 to detect the protein expression by immunoblotting and to purify the proteins by affinity chromatography (Fig. 3.1).

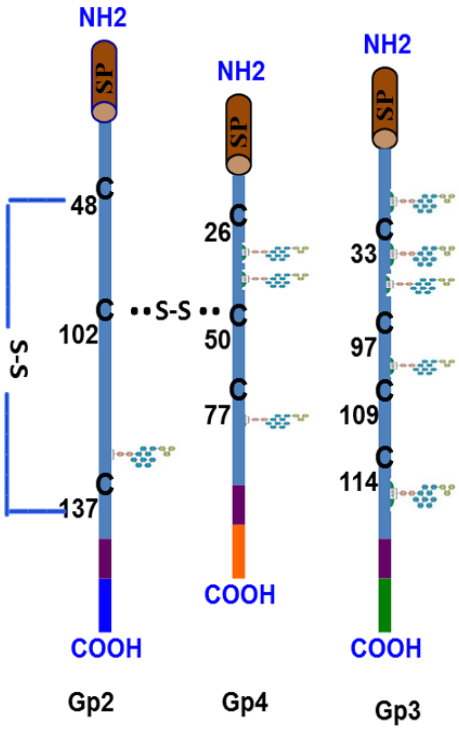


Figure 3.1: Schematic diagram of EAV Gp2, GP4, and Gp3 constructs. The cysteine residues and their respective amino acid numbers are indicated next to the constructs. Disulphide bonds are indicated with dotted lines and S-S symbol. C102 of Gp2 forms disulphide bond with an undetermined cysteine in Gp4. Cysteine 48 and 137 of Gp2 form an intramolecular disulphide bond (Wieringa *et al.,* 2003). Light blue - ectodomain of EAV minor glycoprotein; red - signal peptide; dark purple - short glycine-serine linker; blue - HA tag; orange - Strep II tag; green - His tag; N-linked glycans are shown as branches.

A baculovirus expressing ΔC-Gp2-HA, ΔC-Gp4-Strep, and ΔC-Gp3-His was generated; TriEx™ Sf9 cells were infected to express the above-mentioned complex. I have analysed the expression and secretion of ΔC-Gp2-HA, ΔC-Gp4-strep, and ΔC-Gp3-His by a western

blot of cell culture supernatants using the antibodies against the C-terminal affinity tags. The estimated molecular weights of fully glycosylated ΔC-Gp2-HA, ΔC-Gp4-strep, and ΔC-Gp3-His are 21 kDa, 19 kDa, and 23 kDa, respectively. The size of all three proteins corresponds to the estimated molecular weight under reducing conditions (Fig 3.2A). Under non-reducing conditions, I detected a band at around 42 kDa with anti-HA and anti-strep, which corresponds to the size of ΔC-Gp4-strep and ΔC-Gp2-HA dimer (Fig 3.2A). I detected a band at ~ 45 kDa with ΔC-Gp3-His under non-reducing conditions, which corresponds to the size of ΔC-Gp3-His homodimer or ΔC-Gp3-His linked to either ΔC-Gp4-strep or ΔC-Gp2-HA (Fig 3.2A). I also observed covalently linked multimers above 45-kDa band. Digestion with PNGase F showed that all three proteins (ΔC-Gp2-HA, ΔC-Gp4-strep, and ΔC-Gp3-His) were expressed and secreted into the cell culture media in glycosylated form (Fig 3.2 B

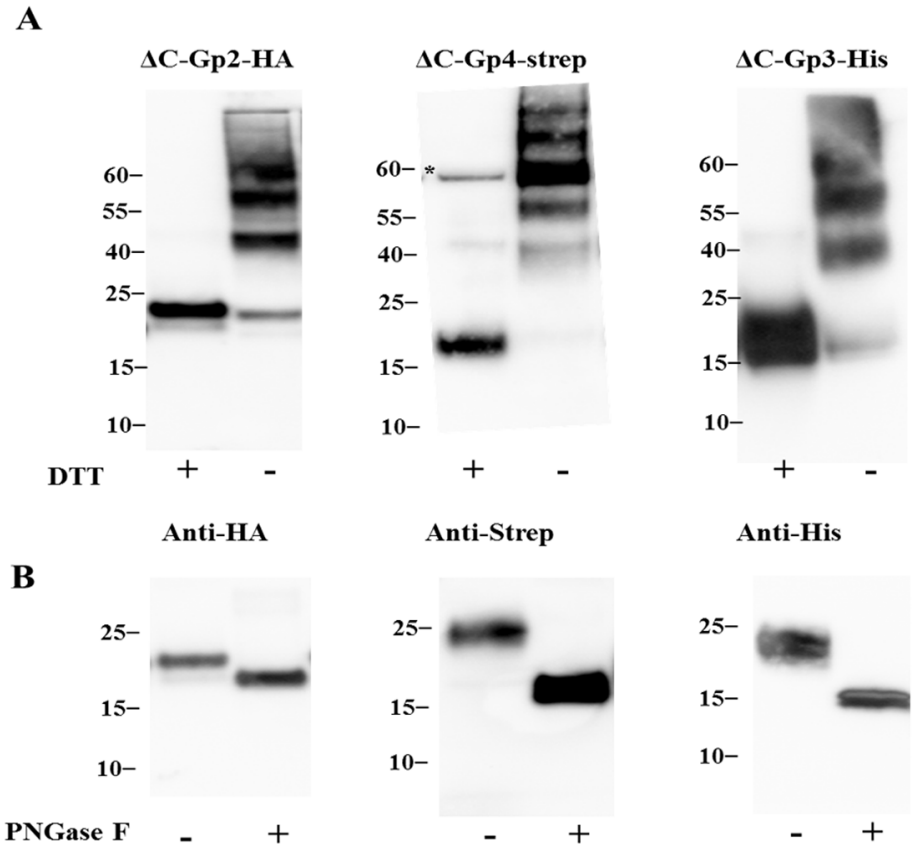


Figure 3.2. Co-expression of EAV minor Glycoproteins Gp2/Gp3/Gp4. A) Cell culture supernatants were subjected to SDS-PAGE and western blot with anti-HA, anti-strep, and anti-His under reducing (+DTT) and non-reducing conditions (-DTT). B) Proteins were

subjected to PNGase F digestion before SDS-PAGE and western blot with respective antibodies. The left lane shows the proteins prior to PNGase F digestion, and the right lane shows the protein after PNGase F digestion. Construct names are shown at the top of the figure, and names of the antibodies are shown in the middle. The protein band marked with an asterisk (*) is most likely an unspecific band. Molecular weight markers in kDa are shown on the left side of the blots

3.1.2 The composition of multimers

Gp2 and Gp4 proteins are known to form various disulfide-linked homo-oligomers (Wieringa *et al*., 2003). In order to show that the multimers contain both ΔC-Gp2-HA and ΔC-Gp4-strep, and not disulfide-linked homo-oligomers (Gp2-Gp2 and Gp4-Gp4), I pulled down ΔC-Gp4-Strep using Strep-Tactin Superflow column and performed western blot using HA tag (tagged to Gp2) antibody under reducing and non-reducing conditions. A prominent band was observed at ~ 22 kDa under reducing conditions, which corresponds to ΔC-Gp2-HA, and a 42 kDa band under non-reducing conditions, which corresponds to Gp2-HA/Gp4-strep heterodimer. The multimeric bands above 42 kDa were also observed under non-reducing conditions (Fig. 3.3). This result demonstrates that oligomers contain both ΔC-Gp2-HA and ΔC-Gp4 Strep. Even though ΔC-Gp3-His was expressed in glycosylated form, I could not detect any ΔC-Gp3-His from the sample enriched with the Strep-tag column (data not shown). This result implies that ΔC-Gp3-His is not oligomerised with ΔC-Gp2-HA/ ΔC-Gp4-strep complex when co-expressed in cells.

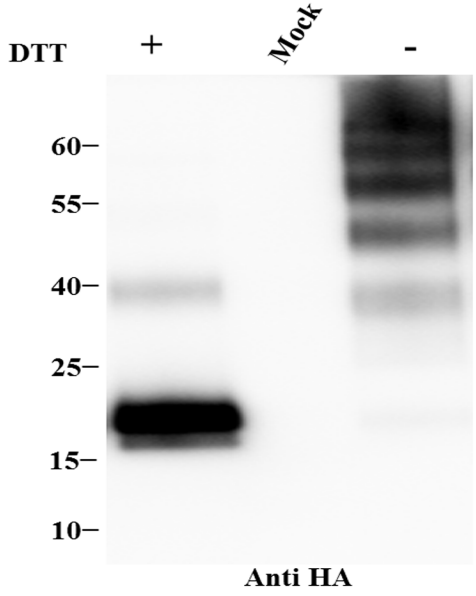


Figure 3.3. Analysis of Gp2/Gp3/Gp4 complex formation. ΔC-Gp4-strep and its binding partners were enriched with the supernatant using Strep-tag column. Immunoblotting was performed using HA tag antibody under reducing (+DTT) and non-reducing conditions (-DTT). Molecular weight markers in kDa are shown on the left side of the blot.

3.1.3 N-glycosylation site usage in EAV Gp4

I analysed a number of used glycosylation sites in ΔC-Gp4-strep by performing limited digestion with serially diluted PNGase F (Peptide-*N*-Glycosidase F), which cleaves N-linked glycans and gives a ladder-like pattern that allows to count the number of carbohydrates. Gp4 sequence contains four potential N-glycosylation sequons (Fig.3.4). Since one of the four predicted glycosylation sequons has a proline next to the asparagine, only three out of four glycosylation sequons were used.

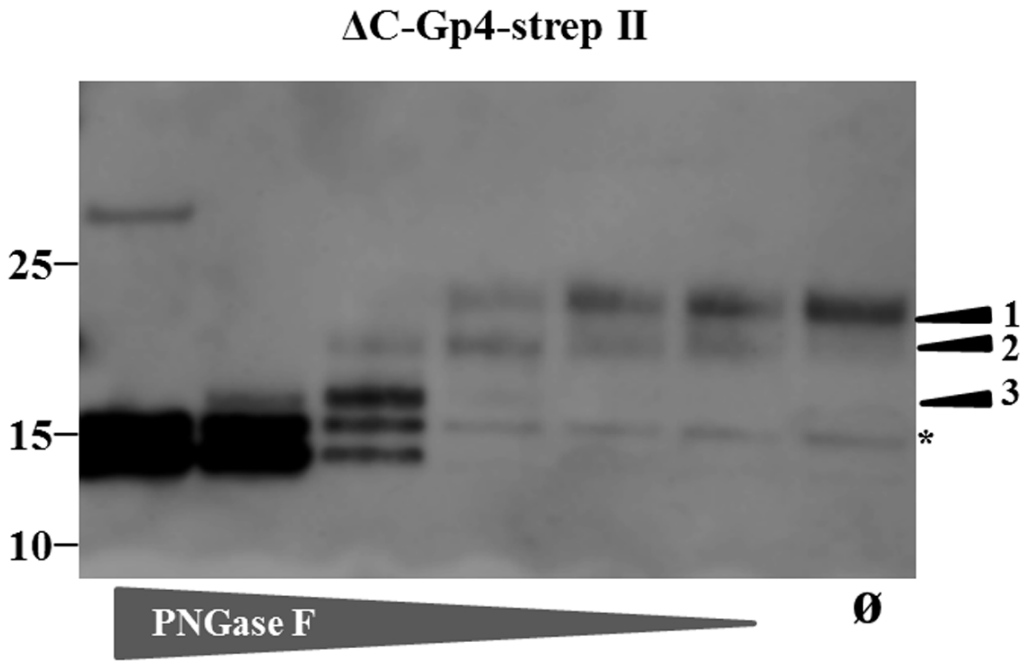


Figure 3.4. ΔC-Gp4-strep comprises three N-linked glycans. Glycans were digested using serially diluted PNGase F prior to SDS PAGE. The western blotting was performed using anti-strep antibody. The numbered arrows indicate the number of glycans remaining on the protein bands. Molecular weight markers in kDa are shown on the left side of the blot. The protein band marked with an asterisk (*) is most likely an unspecific band. (⌀) Symbol indicates no PNGase F digestion.

3.1.4 Signal peptide cleavage prediction from Gp2b and Gp4 from different members of the family *Arteriviridae*

Signal peptide cleavage can be predicted using reliable bioinformatics software programs such as SignalP 4.1 (www.cbs.dtu.dk/services/SignalP/). I predicted the signal peptide cleavage sites for both Gp2 and Gp4 of EAV using SignalP 4.1. In both proteins signal

peptide cleavage is predicted with high certainty, which is represented by the D-score (Table 3.1). I also predicted the signal peptide cleavage of Gp2 and Gp4 from two different porcine reproductive and respiratory syndrome virus (PRRSV) strains, other members of family *Arteriviridae.* The signal peptides were predicted to be cleaved with high confidence (Table 3.1).

Table 3.1 Signal peptide cleavage predictions for Gp2 and Gp4 of Arteriviruses equine arteritis virus (EAV) and porcine reproductive and respiratory syndrome virus (PRRSV).

Virus + Protein	N-terminal amino acid sequence (1–50)	D score
EAV Gp2	**MQRFSFSCYLHWLLLLCFFSGSLLPSAAA↓** WWRGVHEVRVTDLFKDLQCDN	0.872
PRRSV 1 Gp2	**MQWGHCGVKSASCSWTPSLSSLLVWLILP** **FSLPYCLG↓** SPSQDGYWSFFSE	0.849
PRRSV 2 Gp2	**MKWGPCKAFLTKLANFLWMLSRSSWCPL** **LISLYFWPFCLA↓**SPSPVGWWSF	0.769
EAV Gp4	**MKIYGCISGLLLFVGLPCCWC↓**TFYPCHAA EARNFTYISHGLGHVHGHEGC	0.814
PRRSV 1 Gp4	**MAAATLFFLAGAQHIMVSEAFA↓**CKPCFST HLSDIETNTTAAAGFMVLQDI	0.528
PRRSV 2 Gp4	**MASSLLFLVVGFKCLLVSQAFA↓**CKPCFSSSL ADIKTNTTAAASFAVLQDI	0.808

N-terminal amino acid sequences (1–50) of Gp2 and Gp4 of the Arterivirus family members: EAV Bucyrus strain, (GenBank accession number; ABI64076.1), PRRSV type 1/European, strain Lelystad (AAA46278.1), and type 2/North American, strain VR 2332 (AAD12129.1). The predicted signal peptide sequence (in bold) and the 'D value' for the possible cleavage site (vertical inverted arrow↓) predicted by SignalP 4.1 are indicated. The D value is a measure for cleavage probability; threshold value: 0.45. A D value above 0.45 indicates cleavable signal peptide.

3.1.5 Analysing signal peptide cleavage of Gp2 and Gp4 from a recombinant GP2–Gp4 dimer by mass spectrometry

Previous studies indicate that the signal peptides from Gp2 and Gp4 are cleaved off from the mature protein (Wieringa *et al.,* 2002*).* In order to determine the exact signal peptide cleavage site, I used mass spectrometry to identify the protein fragments. Partially enriched ΔC-Gp2-HA and ΔC-Gp4-strep proteins were digested with PNGase F to remove the

glycans, a procedure that is required to avoid the heterogeneity in protein fragments due to the presence of differentially processed glycan side chains. The PNGase F digested protein sample was separated by SDS PAGE and stained with Coomassie to visualize proteins. Bands corresponding to the size of deglycosylated ΔC-Gp2-HA (running ~19 kDa) and ΔC-Gp4-Strep (running ~13 kDa) were excised from the gel (Fig. 3.5b). The proteins were digested using chymotrypsin for ΔC-Gp2-HA and elastase for ΔC-Gp4-Strep. The resulting fragments were extracted from the gel slices and analysed by liquid chromatography-tandem mass spectrometry (LC MS/MS). The LC-MS analysis was done by Dr. Eberhard Krause, Head of the Mass Spectrometry Unit at the Leibniz-Institut für Molekulare Pharmakologie (FMP) in Berlin/Buch.

ΔC-Gp2-HA and ΔC-Gp4-Strep were identified with high confidence in all samples. The ΔC-Gp2-HA and ΔC-Gp4-Strep peptides were the major hits in the MASCOT search, where protein fragments identified by mass spectrometry are matched with the database entries to conclude protein identity. In ΔC-Gp2-HA samples digested with chymotrypsin, the peptide RGVHEVRVTDLF was identified. This peptide matches the amino acids 32–43 of ΔC-GP2-HA. All other possible peptides were identified, but not the ones corresponding to the signal peptide. This result implies that the signal peptide from the ΔC-Gp2-HA is cleaved off from the mature protein. SignalP predicted signal peptide cleavage at site A21|W30. However, ΔC-Gp2-HA was digested with chymotrypsin, which cuts C-terminal to the amino acids F, Y, W, and L if the next residue is not a P. Thus, amino acids W30|31W are cleaved during the chymotrypsin digestion, leaving peptides starting from R32 in ΔC-Gp2-HA (Fig. 3.5B).

After the removal of glycans with PNGase F, ΔC-Gp4-Strep exists as two bands (Fig. 3.5A). I excised the two bands (both upper and lower) separately and analysed them by mass spectrometry to identify the exact signal peptide cleavage site. ΔC-Gp4-Strep samples were digested with elastase. The fragment TFYPCHAAEARNFT starting with T22 was identified in both upper and lower bands of ΔC-Gp4-Strep excised from the gel. In both cases, all other possible peptides were identified except the peptides from the N-terminal signal peptide region. The same peptides were identified in both upper and lower bands; hence, the two-band pattern might be due to some protein modification but not by glycosylation. (Fig. 3.5C).

In order to be detected by the signal peptidase, the amino acid residues at -3 and -1 upstream to the signal peptide cleavage site have to be neutral and small. The well-studied signal peptidase I (Spase I) recognises and cleaves the signal sequence with the c -region "Ala-X-Ala" consensus motif (Renu Tuteja, 2005). However, eukaryotic signal peptides exhibit flexibility at the Ala at -1 and -3 positions and accept the neutral amino acids alanine, glycine, serine, threonine, cysteine, isoleucine, leucine, and valine (Renu Tuteja, 2005). EAV Gp2 has alanine at -1 and -3, and Gp4 has cysteine at -1 and -3. Thus, the identified cleavage sites comply with the signal peptidase specificity.

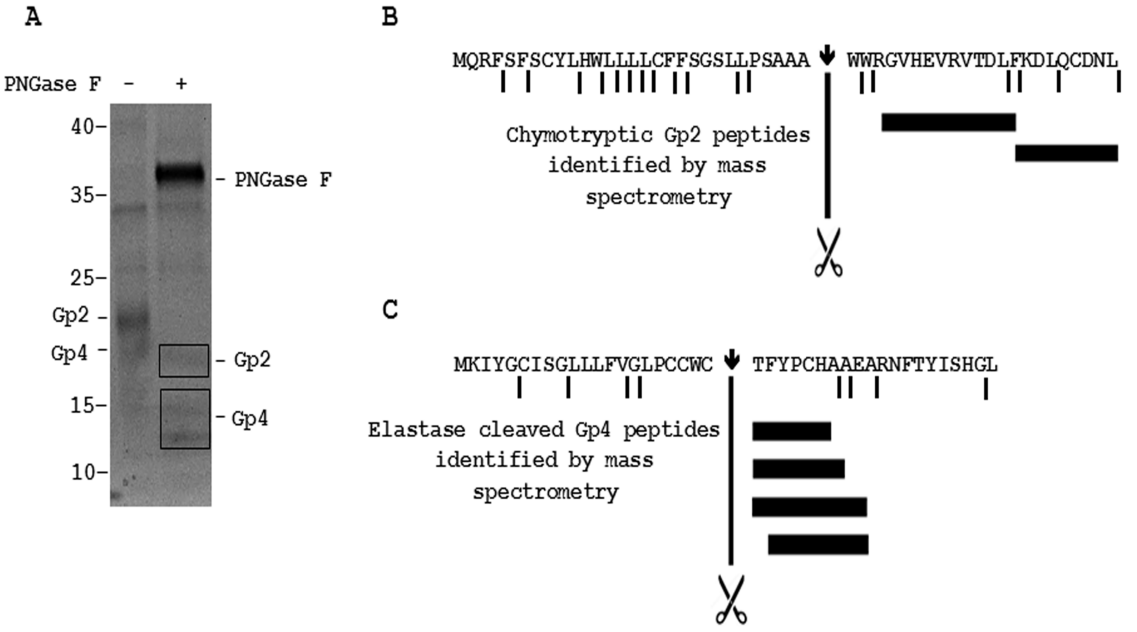


Figure 3.5. Identification of the signal peptide cleavage sites of recombinant Gp2 and Gp4 by mass spectrometry. (A) The ectodomains of Gp2 from EAV Bucyrus strain (amino acids 1–177 fused to a His-tag) and Gp4 (amino acids 1–120 fused to a Strep-tag) were co-expressed in TriEx™ Sf9 cells by infection with recombinant baculovirus. Gp2-Gp4 complexes were enriched from the cell culture supernatants using a Strep-Tactin Superflow column (IBA, Göttingen, Germany). The eluted proteins were digested with PNGase F and subjected to the reducing SDS-PAGE and Coomassie staining. One band with the size of deglycosylated Gp2-HA and a double band for Gp4-Strep (black boxes) were excised from the gel, digested with chymotrypsin or elastase, and analysed by LC–MS/MS. The left lane shows the purified proteins prior to PNGase F digestion. In both cases peptides corresponding to Gp2 and Gp4, respectively, were the predominant peaks in

the MS spectra. (B) An illustrative result from the mass spectrometry of Gp2. The first 51 residues of Gp2-HA are shown with the predicted chymotrypsin cleavage sites are marked with black lines, and the signal peptide cleavage site is indicated by an arrow and scissors. The identified chymotryptic peptides are denoted by black bars. (C) Illustrative results from the mass spectrometry of Gp4. The first 41 residues of Gp4 are shown with the predicted elastase cleavage sites and are marked with black lines. The signal peptide cleavage site is indicated by an arrow and scissors. The identified peptides resulting from digestion with elastase are denoted by black bars.

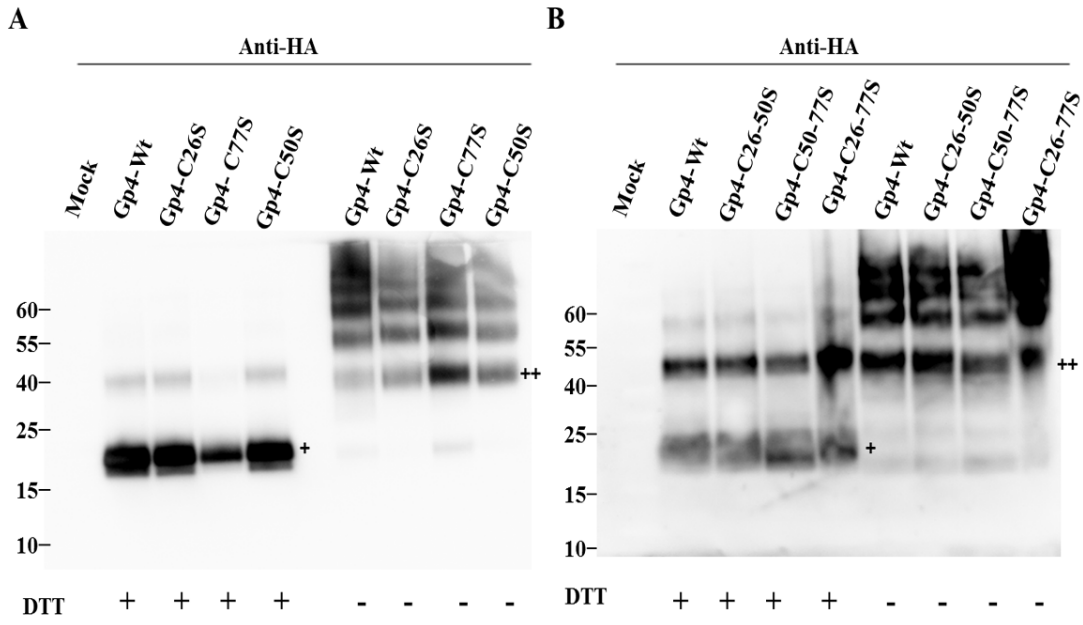
3.1.6 Effect of cysteine mutations in Gp4 ectodomain on disulphide linkage with Gp2

I showed experimentally that the Gp2-HA and Gp4-strep ectodomains could form a disulphide-linked dimer (Fig. 3.3). The signal peptide from Gp4 ectodomain expressed in insect cells is cleaved between amino acids 21 and 22 (Fig. 3.5C). Hence, I eliminated the possibility of the cysteine residues in the signal peptide to be involved in disulphide linkages. Therefore, I mutated the other cysteine residues, i. e. C26, C50, and C77 of Gp4 to serine, present in the ectodomain, to determine the cysteine residue involved in the disulphide linkage between Gp2-HA and Gp4-Strep. I infected the insect cells with baculovirus expressing Gp4-wt, Gp4-C26S, Gp4-C50S, and Gp4-C77S mutants. The secreted proteins were enriched from supernatants using Strep-Tactin column. In all cases, Gp4 is tagged to Strep II tag in the C-terminus, and all four viruses are co-expressing ΔC-Gp3-His and ΔC-Gp2b HA. The western blot was performed with Strep-tag enriched samples using anti-HA antibody, and I could detect Gp2-HA in Gp4-wt and all the Gp4 single cysteine mutants (Fig 3.6A). This result shows that mutated single cysteine residues could not abolish the formation of the disulphide linkage between Gp2 and Gp4 ectodomains.

Among the three cysteines in the Gp4 ectodomain, one cysteine is assumed to form disulphide linkage with Gp2 and one with Gp3. However, Gp3 is not oligomerised with the Gp2/4 complex when expressed in cells. Thus, I hypothesised that the cysteine that is supposed to form disulphide bond with Gp3 might cause multimer formation. Therefore, I wanted to test the effect of double cysteine mutations in Gp4 on multimer formation. I expressed Gp4-wt, Gp4-C50-77S, Gp4-C26-77S, and Gp4-C26-50 in insect cells. Gp4 is tagged to Strep II tag in the C-terminus, and ΔC-Gp2b-HA and ΔC-Gp3-His are co-

expressed in all four cases. I pulled Gp4-strep from all the four combinations to test its binding partners. In every double mutant Gp4 is disulphide linked to ΔC-Gp2-HA. I detected multimers in all combinations (Fig. 3.6B). This result indicates that even double cysteine mutations in Gp4 do not have any influence on multimer formation, and any of the three cysteines in the Gp4 ectodomain can form disulphide linkage with Gp2 (Fig. 3.6B). It is unclear how the multimers can still form because the only cysteine that is left on the Gp4 ectodomain must bind to the Gp2. There were no cysteines in the strep II tag, and also no peptides were identified from the SP region that might contribute to the disulphide linkage (Fig 3.5C).

Figure 3.6. Cysteine mutations in Gp4 do not influence ΔC-Gp4-Strep and ΔC-Gp2-HA dimer formation. ΔC-Gp4-Strep constructs with single cysteine mutations (A) and double cysteine mutations (B) were expressed in insect cells, and the secreted protein was enriched using Strep-Tactin column. Enriched samples were subjected to SDS-PAGE



under reducing (+DTT) and non-reducing (-DTT) conditions and western blotting with HA tag antibody. In both A) and B) sections, molecular weight markers in kDa are shown on the left side of the blots. Monomers (+) and dimers (++) are indicated by plus marks.

3.1.7 Expression of PRRSV Gp2/Gp3/Gp4 ectodomains in insect cells

Gp2, Gp3, and Gp4 proteins exist as a disulfide-linked complex in PRRSV virus particles (Dokland, 2010; Wissink *et al.,* 2005). In EAV, the association of the Gp3 with Gp2/Gp4

dimer only happens after the release of virus particles in a unique manner. Therefore, I wanted to examine whether the Gp2/Gp3/Gp4 of PRRSV forms disulphide-linked hetero-oligomers or not. The C-terminal hydrophobic regions of glycoproteins Gp2, Gp3, and Gp4 were deleted to enable secretion of proteins. A Twin Strep-tag was added to Gp4, an HA tag to Gp2, and a His-tag to Gp3, in each case to the C-terminus of the protein to detect protein expression and also for the purification of the proteins (Fig. 3.7).

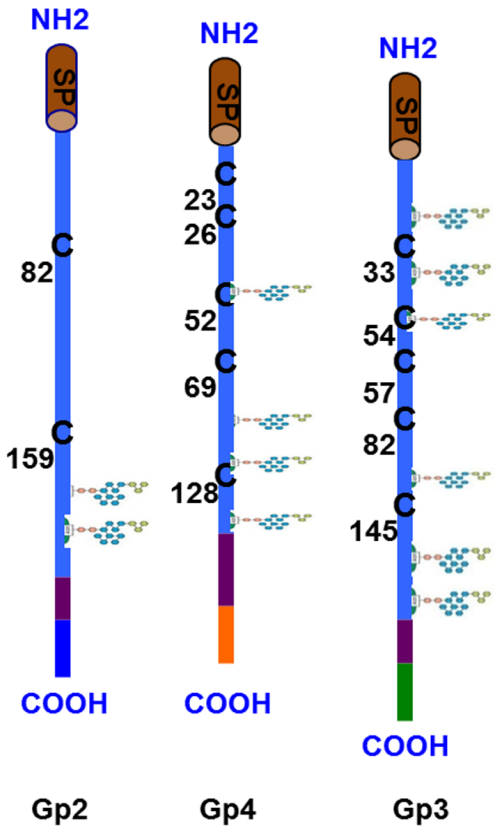


Figure 3.7: **Schematic diagram of PRRSV Gp2, GP4, and Gp3 constructs.** The cysteine residues in the ectodomain and their respective amino acid numbers are mentioned next to the constructs. Light blue - ectodomain of EAV minor glycoprotein; red - signal peptide; dark purple - short GGSG linker; blue - HA tag; orange - Twin-Strep tag; green - His tag; N-linked glycans are shown as branches.

Insect cells were infected using a baculovirus expressing ΔC-Gp2-HA, ΔC-Gp4-Twin strep, and ΔC-Gp3-His. Expression and secretion of the proteins was analysed by western blot with cell culture supernatants. I observed a band at ~24 kDa under reducing conditions with HA tag antibody, which corresponds to the size of glycosylated ΔC-Gp2-HA (Fig.3.8B), a 25 kDa band with strep II tag antibody, which corresponds to the size of glycosylated ΔC-Gp4-Twin strep (Fig. 3.8B), and a band at ~32 kDa that corresponds to glycosylated ΔC-Gp3-His (Fig. 3.8C). The digestion of the proteins with PNGase F indicates that all three proteins were secreted in a glycosylated state (Fig. 3.8 A, B and C).

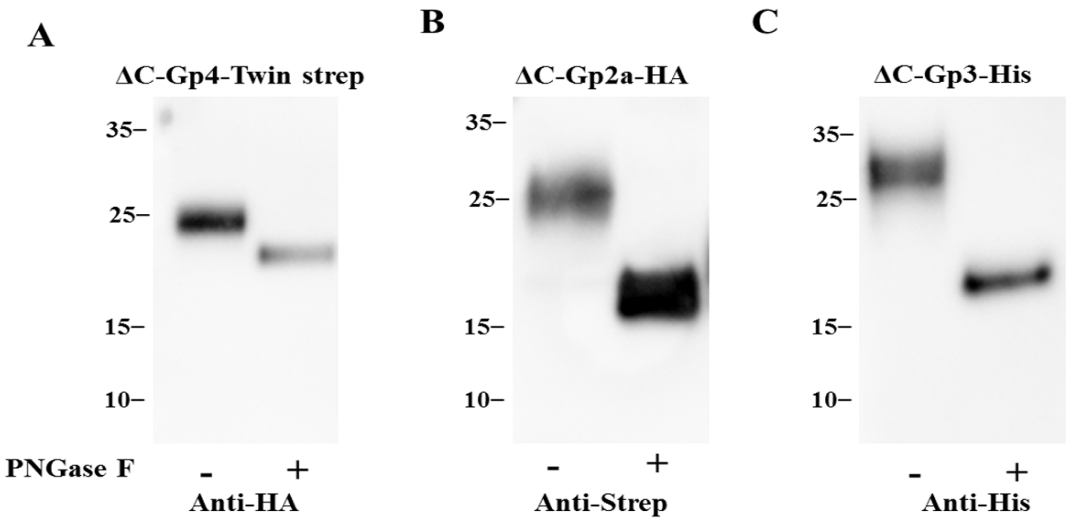


Figure 3.8. Expression of PRRSV minor glycoproteins Gp2/Gp3/Gp4. ΔC-Gp3-His, ΔC-Gp4-Twin strep, and ΔC-Gp2-HA proteins were expressed in TriEx™ Sf9 cells. Proteins in the supernatant were subjected to SDS-PAGE, and western blot was performed using respective antibodies. The left lane shows the proteins with no PNGase F digestion, and the right lane shows the proteins after the PNGase F digestion. Construct names are shown at the top of the figure, the names of the antibodies are mentioned at the bottom. Molecular weight markers in kDa are shown on the left side of the blots.

3.1.8 Complex formation between PRRSV Gp2/Gp3/Gp4

Under non-reducing conditions, ΔC-Gp2-HA, ΔC-Gp3-His, and ΔC-Gp4-Twin strep exist as various disulphide-linked multimers (data not shown). In order to show that the multimers contain disulphide linked ΔC-Gp2-HA, ΔC-Gp3-His, and Gp4-Twin strep (but not disulphide-linked homo-oligomers (Gp2-Gp2, Gp4-Gp4 or Gp3-Gp3)), I pulled down ΔC-Gp4-Twin strep using Strep-Tactin Superflow column and performed western blot using HA tag (tagged to Gp2), His tag (tagged to Gp3), and Strep-tag (tagged to Gp4) antibodies under reducing and non-reducing conditions. I observed a band 25 kDa under reducing conditions with anti-strep antibodies in the western blot, which corresponds to ΔC-Gp4-Twin strep and several disulphide-linked multimers above 50 kDa under non-reducing conditions, which might be ΔC-Gp4-Twin strep linked to one or more copies of ΔC-Gp2-HA/ ΔC-Gp3-His proteins (Fig. 3.9A). Immunoblotting with anti-HA antibody

shows a band ~25 kDa under reducing conditions that match the size of ΔC-Gp2-HA and several disulphide-linked multimers above 50 kDa under non-reducing conditions. They might be ΔC-Gp2-HA linked to one or more copies of ΔC-Gp3-His / ΔC-Gp4-twin strep proteins (Fig. 3.9B). I have observed a band around 30 kDa under reducing conditions with anti-His antibodies in western blot, which corresponds to ΔC-Gp3-His, and several disulphide-linked multimers above 50 kDa under non-reducing conditions, which might be ΔC-Gp3-His linked to one or more copies of ΔC-Gp2-HA/ ΔC-Gp4-Twin strep proteins (Fig. 3.9C). This result shows that both ΔC-Gp2-HA and ΔC-Gp3-His are co-purified with ΔC-Gp4-twin strep, which means they are covalently linked to each other through a disulphide linkage. This result demonstrates that the Gp2/Gp3/Gp4 of PRRSV is secreted as a heterooligomeric complex into the cell culture supernatant. The resultant complexes could be either a trimer or Gp2-Gp4 and Gp4-Gp3 dimers or a mixture of all combinations.

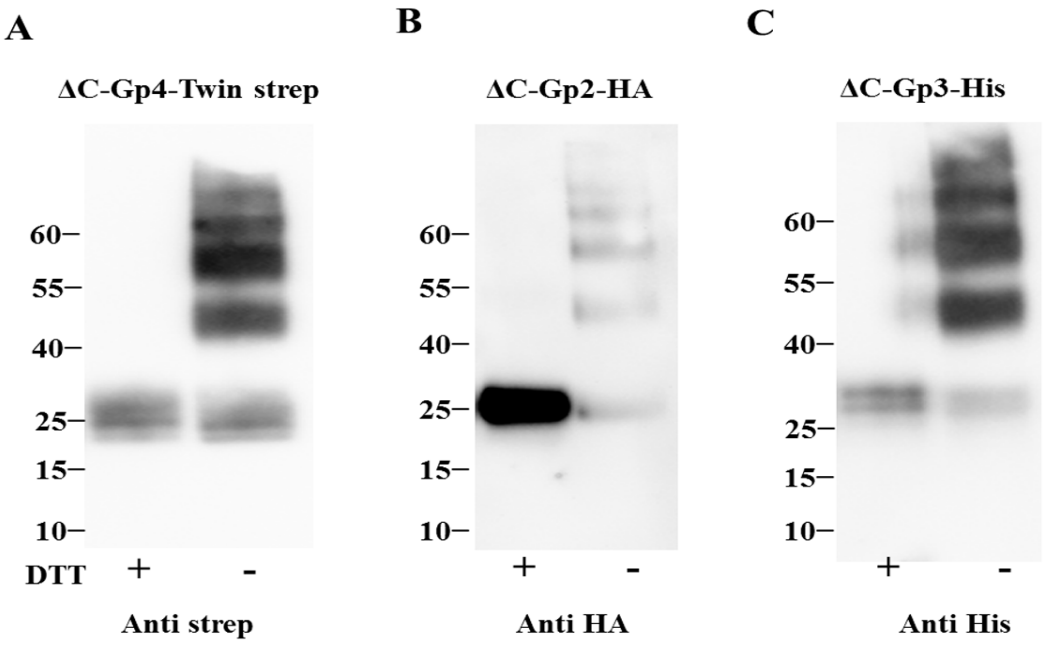


Figure 3.9 ΔC-Gp2-HA/ ΔC-Gp4-Twin strep/ ΔC-Gp3-His proteins were co-expressed in insect cells; secreted proteins were enriched using Strep-Tactin column, subjected to SDS-PAGE and western blot with anti-HA, anti-strep, and anti-His antibodies, respectively, under reducing (+DTT) and non-reducing conditions (-DTT). Construct names are shown at the top of the figure; the names of the antibodies are mentioned at the bottom. Molecular weight markers in kDa are shown on the left side of the blots.

3.1.9 N-glycosylation site usage in PRRSV Gp2, Gp3, and Gp4

I analysed the number of glycosylation sites used in PRRSV Gp2, Gp3, and Gp4 ectodomains. A stepwise reduction of PNGase F amount results in a ladder-like pattern, which allows to count the glycosylation sites. ΔC-Gp3-His has seven predicted potential glycosylation sites, and limited digestion results showed that only six N-glycosylation sequons were used (Fig. 3.10A). The glycosylated PRRSV Gp3 exists as two bands in SDS-PAGE, whereas after glycan removal it exists as one band. Thus, the reason for the existence of two bands is the partial usage of one of the six N-glycosylation sites. Gp2 contains two potential N-glycosylation sites, and both of them are efficiently used (Fig. 3.10B). Gp4 sequence contains four potential N-glycosylation sequons, and all four N-glycosylation sequons were used (Fig. 3.10C

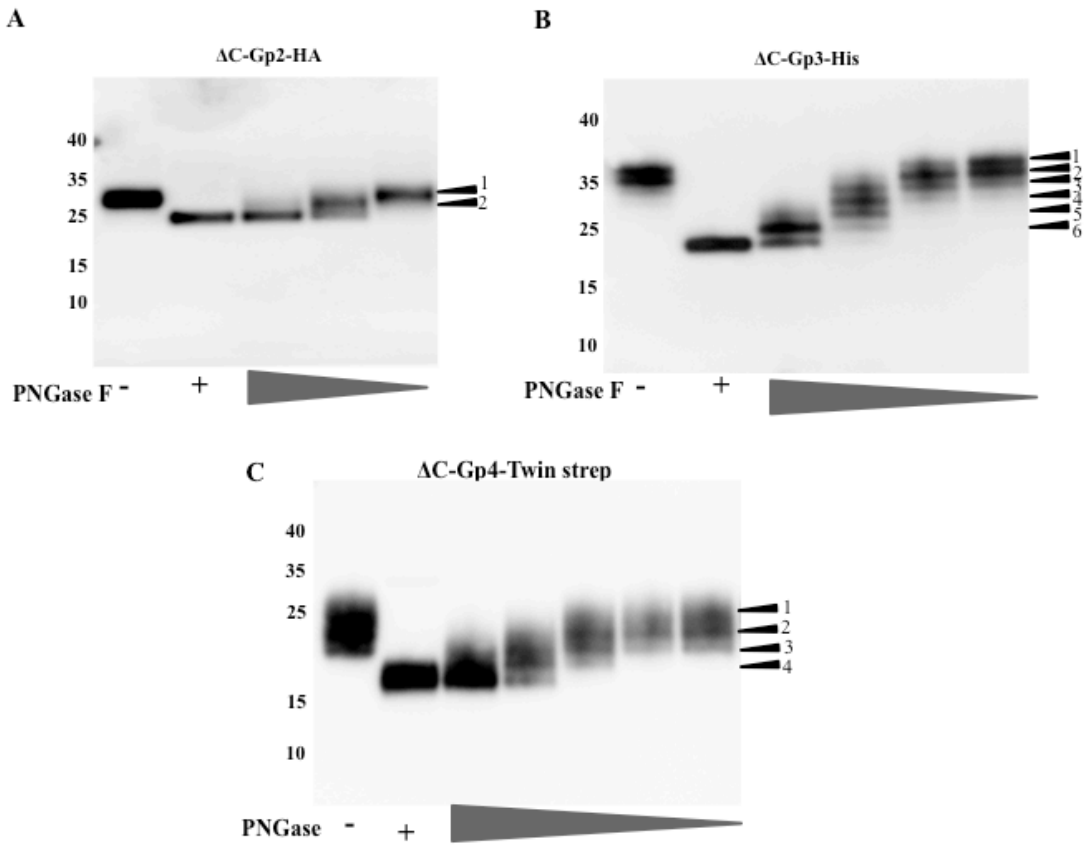


Figure 3.10. All predicted N-glycosylation sites are used in PRRSV Gp2/Gp3/Gp4. ΔC-Gp2-HA, ΔC-Gp3-His, and ΔC-Gp4-Twin strep proteins were digested by stepwise reduction of PNGase F before the SDS PAGE and western blotting was performed using anti-HA, anti-His, and anti-strep antibodies.The numbered arrows indicate the number of

glycans remaining on the protein bands. Molecular weight markers in kDa are shown on the left side of the blots.

3.2 Expression of glycoprotein 3 of EAV and PRRSV and the role of N-glycosylation in signal peptide processing

3.2.1 Signal peptide cleavage prediction of Gp3 from different members of *Arteriviridae* by SignalP 4.1

Signal peptide cleavage sites of EAV and PRRSV Gp3 were predicted using bioinformatics software SignalP 4.1. In both EAV and PRRSV, the signal peptide cleavage is predicted with high confidence, which is signified by the D-score (Table 3.2.1).

Table 3.2. Signal peptide cleavage prediction for GP3 of Arteriviruses equine arteritis virus (EAV) and porcine reproductive and respiratory syndrome virus (PRRSV).

Virus + Protein	N-terminal amino acid sequence (1–50)	**D score**
EAV Gp3	**MGRAYSGPVALLCFFLYFCFICG↓SVG↓**SNN TTICMHTTSDTSVHLFYAANV	0.872
PRRSV 1 Gp3	**MAHQCARFHFFLCGFICYLVHSALA↓S**NSS STLCFWFPLAHGNTSFELTIN	0.854
PRRSV 2 Gp3	**MVNSCTFLHIFLCCSFLYSFCCA↓**VVAGSNTT YCFWFPLVRGNFSFELTVN	0.685

N-terminal amino acid sequences (1–50) of Gp3 of the Arterivirus family members: EAV Bucyrus strain, (GenBank accession number; ABI64076.1), PRRSV type 1/European, strain Lelystad (AAA46278.1), and type 2/North American, strain VR 2332 (AAD12129.1). The predicted signal peptide sequence (in bold) and the D value for the possible cleavage site (vertical inverted arrow↓) are indicated according to bioinformatics prediction tool SignalP 4.1. The D value is a measure of cleavage probability, and the threshold value is 0.45. A D value above 0.45 indicates a cleavable signal peptide. For EAV Gp3, two possible signal peptide cleavage sites were predicted (designated with vertical bar↓).

3.2.2 Expression of soluble EAV Gp3 in insect cells

In vitro translation experiments with microsomes demonstrated that the signal peptide of Gp3 from EAV is not cleaved from the mature protein, and the uncleaved signal peptide might act as a membrane anchor (Wieringa *et al.,* 2003). However, SignalP 4.1 predicted that the Gp3 signal peptide is cleaved off with high probability. Hence, I wanted to express the Gp3 in insect cells to determine by mass spectrometry whether and where the signal peptide is cleaved (LC MS/MS). The C-terminal hydrophobic region from both

Gp3-wt and Gp3-N28/29S mutant were removed by PCR amplification, and a 9X histidine tag was added to the C-terminal region.

TriEx™ Sf9 insect cells were infected with the baculoviruses expressing ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S, and the expression and secretion of proteins into the cell culture supernatant was detected by western blot using His-tag antibody. I detected protein bands at 24 kDa and 21 kDa with Δ Gp3 wt-His and ΔC-Gp3-His-N28/29S, respectively (Fig. 3.11). The reduction in the size of ΔC-Gp3-His-N28/29S is due to the loss of carbohydrates by the removal of potential N-glycosylation sites N28/29 near the signal peptide cleavage site. PNGase F digestion result showed that both Δ-Gp3 wt-His and ΔC-Gp3-His-N28/29S are expressed and secreted in glycosylated form, and exist as two bands before and after PNGase F digestion for (Fig. 3.11)

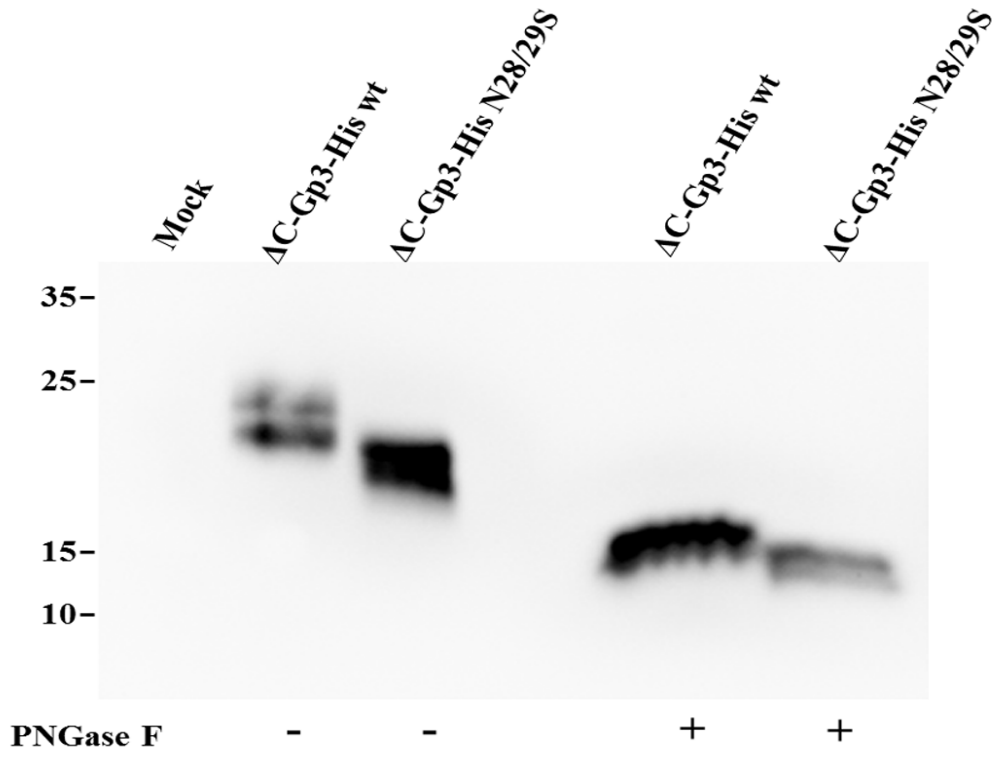


Figure 3.11 Expression of EAV Gp3 in insect cells. Cell culture supernatants were digested with or without PNGase F, and subjected to SDS-PAGE and western blot using anti-His antibody. The construct names are shown at the top of the figure. Mock indicates that cells were infected with non-recombinant baculovirus. Molecular weight markers in kDa are shown on the left side of the blots.

3.2.3 Overlapping sequon NNTT near the signal peptide is glycosylated at only one asparagine in insect cells

Gp3 of EAV has two overlapping N-glycosylation sequons near the signal peptide cleavage site. Such overlapping sequons are usually rare in proteins and, if present, either only one asparagine is glycosylated or both asparagines are not used (Reddy *et al*., 1999; Karamyshev *et al*., 2005). However, in the case of Gp3 from EAV it was shown that the two overlapping sequons are efficiently used in mammalian cells (Matczuk *et al.,* 2013). Therefore, I wanted to determine whether this overlapping sequon NNTT is also glycosylated in insect cells. I digested insect cell expressed ΔC-Gp3-His wt and ΔC-Gp3-His-N28/29S with a decreasing amount of PNGase F. This results in a ladder-like band pattern, which allows to count the number of glycans remaining on the proteins. The result showed that only five out of six potential N-glycosylation sites are used in ΔC-Gp3–His-wt (Fig. 3.12A) and that all four N-glycosylation sites are used in the case of ΔC-Gp3-His N28/29S (Fig. 3.12B). This result indicates that only one N-glycosylation site from the overlapping sequon NNTT is used in insect cells. The two-band pattern of Gp3 still exists even after PNGase F digestion. Thus, the two bands are not due to inefficient glycosylation of overlapping sequon at only one site, but due to some other unknown protein modification of Gp3.

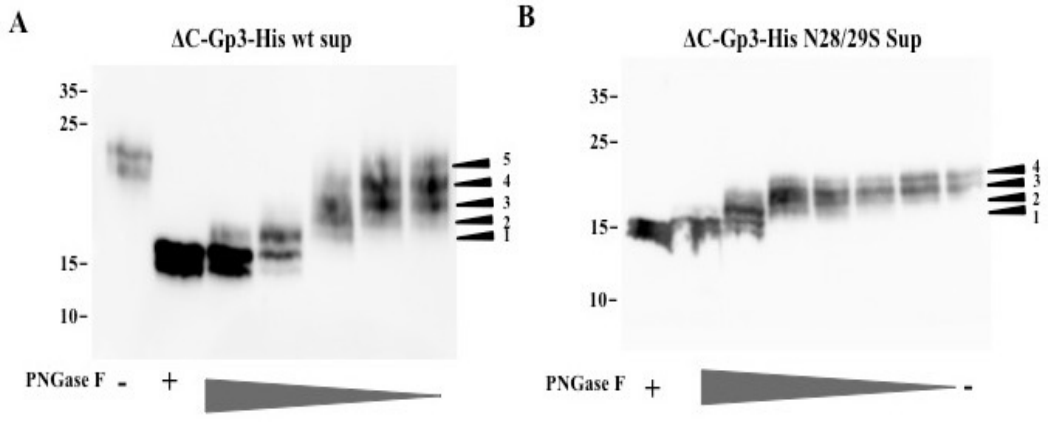


Figure 3.12. Limited PNGase F digestion of EAV Gp3 shows that the overlapping sequon NNTT is glycosylated at only one asparagine. ΔC-Gp3-His and ΔC-Gp3-His-N28/29S were expressed in TriEx™ Sf9 cells. Glycans were digested by a stepwise

reduction in PNGase F amount before the SDS-PAGE and western blotting were performed using anti-His antibody. The numbered arrows imply the number of glycans remaining on the protein bands. Molecular weight markers in kDa are shown on the left side of the blots.

3.2.4 Signal peptide is cleavage from insect cell-expressed Gp3

Gp3 signal peptide is predicted to be cleaved from the mature protein at two sites: between amino acids G23|S24 with a minor probability and between G26|S27 with a higher probability. It was shown that in EAV the Gp3 signal peptide cleavage is prevented by the attachment of glycans at the overlapping sequon NNTT nearby the signal peptide cleavage site (Matczuk *et al.,* 2014). However, the exact signal peptide cleavage site is not known. Therefore, I wanted to analyse the signal peptide cleavage by mass spectrometry when the overlapping sequon was mutated. ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S proteins were enriched by cell culture supernatants using HisTrap column. The glycans were digested with PNGase F. The protein samples were run on 15% tris-glycine gel and stained with Coomassie stain. After PNGase F digestion, both ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S exist as two bands (Fig. 3.13B). I carefully excised the two bands (both upper and lower) separately and analysed by mass spectrometry to identify the exact signal peptide cleavage site. The proteins were digested using chymotrypsin. The subsequent digested fragments were extracted from the gel slices, and peptides were analysed by liquid chromatography-tandem mass spectrometry (LC MS/MS).

ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S peptides were identified with high confidence in all samples, and they are the major hits in the MASCOT search. In ΔC-Gp3-His-wt samples digested with chymotrypsin, the peptides starting with SVGSNNTTICMHTTSDTSVHLF (S24) were identified as a minor population, whereas the ones starting with SNNTTICMHTTSDTSVHLF (S27) were identified as a major population. This sequence matches to the amino acids 24–45 and 26-45 of ΔC-Gp3-His-wt (Fig. 3.13C). In ΔC-Gp3-His-N28/29S samples, the same peptides were identified as in the case of ΔC-Gp3-His-wt (Fig. 3.12D). All other possible peptides were identified, but not the ones corresponding to the signal peptide in both ΔC-Gp3-His-wt and ΔC-Gp3-His- N28/29S samples (Fig. 3.13C and D). There is no difference between the identified

peptides in both upper and lower bands from ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S samples. This result matches with the software prediction and suggests that the signal peptide from insect cell-expressed Gp3 ectodomain is cleaved off from both the wild type and the mutant. This result also indicates that there is no influence of the overlapping sequon glycosylation on signal peptide cleavage of insect cell-expressed Gp3.

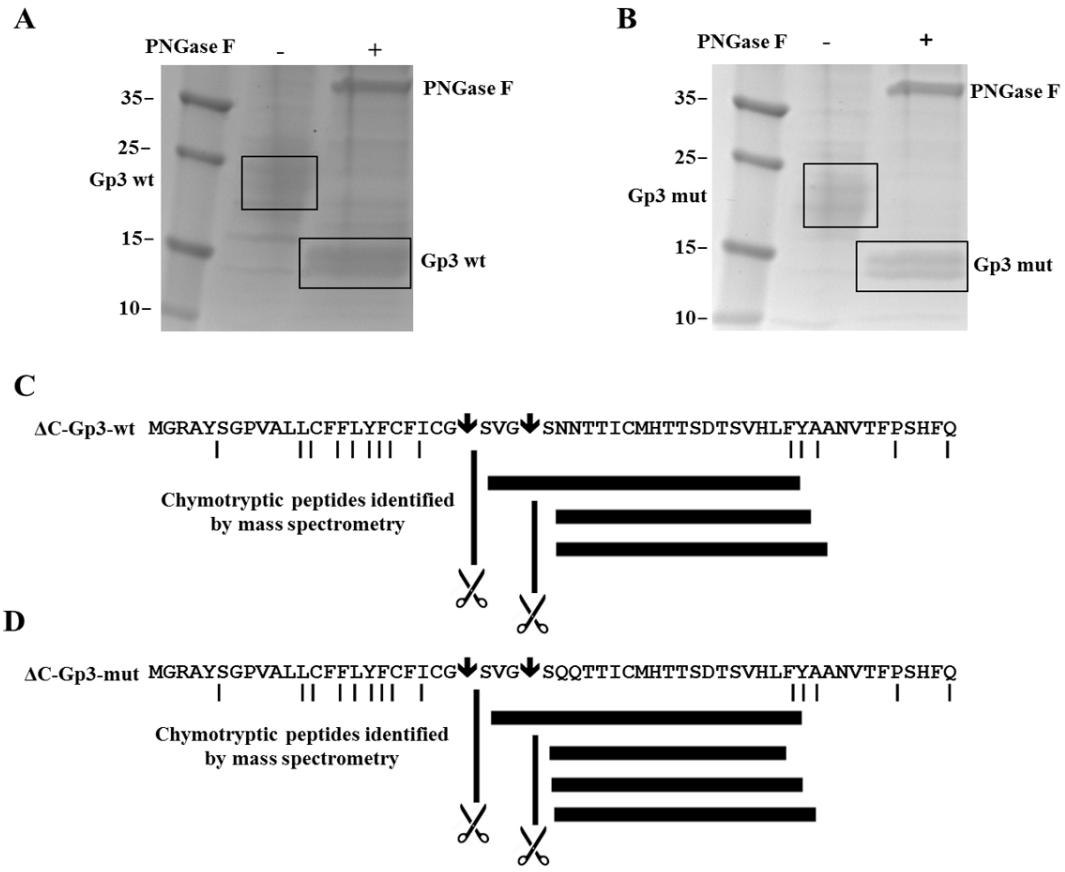


Figure 3.13. Identification of the signal peptide cleavage site of recombinant ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S by mass spectrometry. (A) ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S from EAV Bucyrus strain (amino acids 1–136 fused to a His-tag) were expressed in TriEx™ Sf9 insect cells by infection with recombinant baculovirus. ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S proteins were enriched by the cell culture supernatants using HisTrap column (GE healthcare*, Uppsala, Sweden*). The eluted proteins were digested with PNGase F and subjected to reducing SDS-PAGE and Coomassie staining. Band corresponding to the size of deglycosylated ΔC-Gp3-His-wt and Gp3-His-N28/29S were excised from the gel, digested with chymotrypsin, and analysed by LC–MS/MS. The left lane shows the purified proteins prior to PNGase F

digestion. In both cases, peptides corresponding to ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S were the predominant peaks in the MS spectra. (C) An illustrative result from the mass spectrometry of ΔC-Gp3-His-wt. The first 57 residues of ΔC-Gp3-His-wt are shown with the predicted chymotrypsin cleavage sites marked with black lines, and the signal peptide cleavage site is indicated by an arrow and scissors. The identified chymotryptic peptides are denoted by black bars. (D) Illustrative results from mass spectrometry of ΔC-Gp3-His-N28/29S. The first 57 residues of ΔC-Gp3-His-N28/29S are shown with the predicted chymotrypsin cleavage sites marked with black lines, and the signal peptide cleavage site is indicated by an arrow and scissors. Identified peptides resulting from digestion with chymotrypsin are denoted by black bars. Molecular weight markers in kDa are shown on the left side of the blots.

3.2.5 Glycosylation and signal peptide cleavage from intracellular ΔC-Gp3

I assumed that the ΔC-Gp3-His fraction with the signal peptide might remain inside the cells by anchoring to the membrane by its signal peptide. To determine this, I examined both intracellular and secreted ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S for the presence of the signal peptide and the number of glycosylation sites used. I enriched the intracellular ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S by the cell lysates by the nickel affinity chromatography. Both intracellular and secreted ΔC-Gp3-His-wt and ΔC-Gp3-His- N28/29S were digested with gradually reduced amounts of PNGase F to remove the glycans. The proteins were separated by SDS-PAGE, and the western blot was performed using anti-His antibody. The result showed that only five N-glycosylation sites and four N-glycosylation sites were used in intracellular ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S, respectively (Fig. 3.14A and B). Intracellular and secreted samples of both ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S were digested with PNGase F and analyzed next to each other to compare the signal peptide cleavage difference. The western blot results shows that the secreted versions of both ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S are of same size after the PNGase F digestion (Fig. 3.14C). This result demonstrates that there is no difference between signal peptide cleavage of intracellular and secreted ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S.

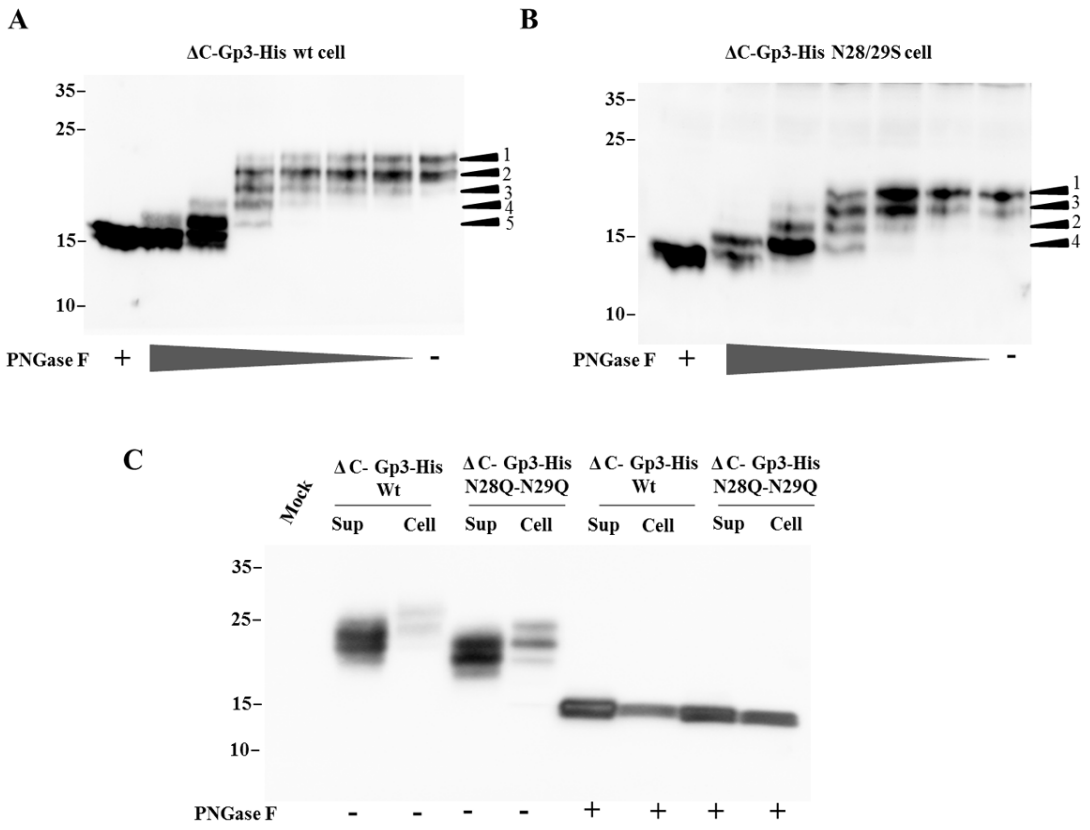


Figure 3.14. Limited PNGase F digestion of EAV ΔC-Gp3 from cells and signal peptide cleavage from both secreted and intracellular ΔC-GP3. A) and B) ΔC-Gp3-His-wt and ΔC-Gp3-His-N28/29S proteins were enriched by cell lysate using Ni-NTA agarose. Glycans were digested using serially digested PNGase F prior to the SDS-PAGE. The western blotting was performed using anti-his antibody. The numbered arrows imply the number of glycans remaining on the protein bands. C) ΔC-Gp3-His and ΔC-Gp3-His N28/29S from cells secreted before and after PNGase F digestion. Sup - supernatant; Cell - intracellular. Mock indicates that cells were infected with a non-recombinant baculovirus. Molecular weight markers in kDa are shown on the left side of the blots.

3.2.6 Expression of full-length Gp3-His and N-glycosylation site determination

My result showed that the signal peptide is cleaved from the ΔC-Gp3–His, and only one site on the double glycosylation sequon NNTT is used. Thus, I wanted to examine the influence of C-terminal transmembrane region on signal peptide cleavage and N-glycosylation. I made a full-length Gp3-His construct with a 9X histidine tag at the C-terminus and expressed in insect cells. Expression of Gp3 was confirmed by western blotting with His-tag antibody from the cell lysates. The protein was enriched by the cell

lysates using Ni- NTA agarose beads. Limited digestion with PNGase F revealed that only five glycosylation sites were used in the full-length Gp3 expressed in insect cells (Fig. 3.15).

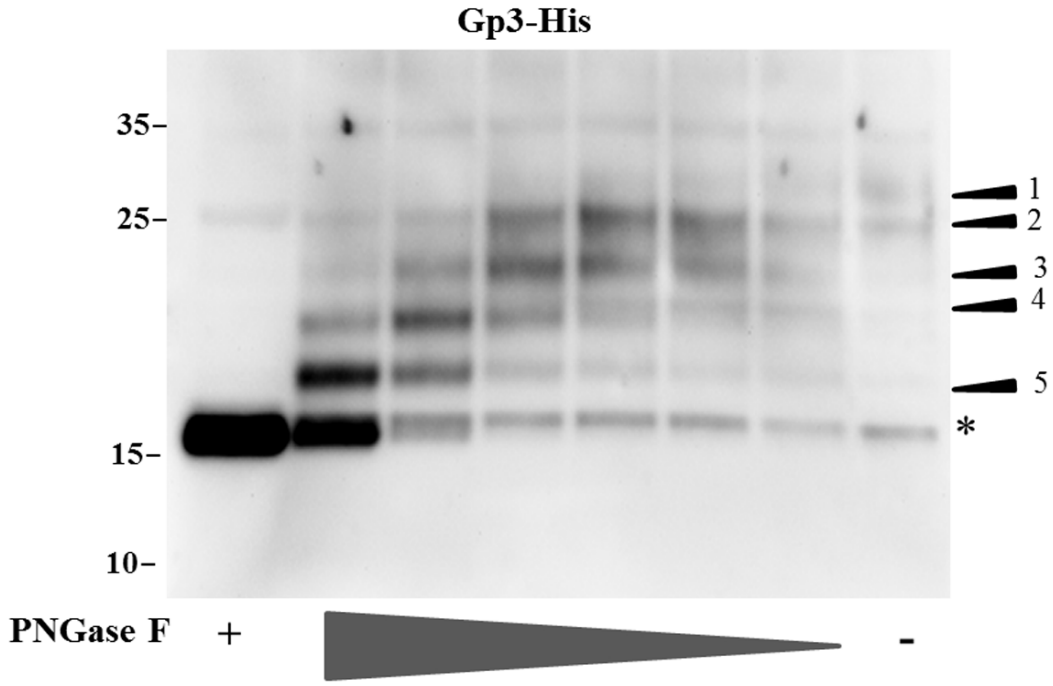


Figure 3.15. Limited PNGase F digestion of EAV Gp3 from cells. A full-length Gp3-His protein was expressed in TriEx™ Sf9 cells. The glycans were digested using serially digested PNGase F prior to the SDS-PAGE, and western blotting was performed using anti-His antibody. The numbered arrows imply the number of glycans remaining on the protein bands. Molecular weight markers in kDa are shown on the left side of the blots. * symbol indicates an unspecific band.

3.2.7 Signal peptide cleavage from EAV full-length Gp3-His

Previous results indicated that the C-terminal region of the viral protein might influence the signal peptide processing (Rehm *et al.,* 2001). Therefore, I wanted to determine the signal peptide cleavage from a full-length Gp3-His**.** Deglycosylated Gp3-His protein samples were run in 15% tris-glycine gel and stained with Coomassie to visualise proteins. The band corresponding to the size of deglycosylated Gp3-His (running at 16 kDa) was excised from the gel (Fig 3.16A), and the proteins were digested using

chymotrypsin. The subsequent digested fragments were extracted from the gel slices. Peptides were analysed by liquid chromatography-tandem mass spectrometry (LC MS/MS). Peptides corresponding to Gp3-His-wt were identified with high sequence coverage. In Gp3-His samples digested with chymotrypsin, the peptide SNNTTICMHTTSDTSVHLF was identified (Fig. 3.16C). This sequence matches with amino acids 26-45 of a full-length Gp3, and it is the same site as determined for the secreted Gp3 (Fig. 3.13). All the other peptides corresponding to the Gp3 were identified, except for the peptides from the N-terminal signal peptide sequence. This result suggests that the signal peptide is cleaved from the full length Gp3-His, and the C-terminal transmembrane region has no influence on its cleavage in the insect cell-expressed Gp3-His.

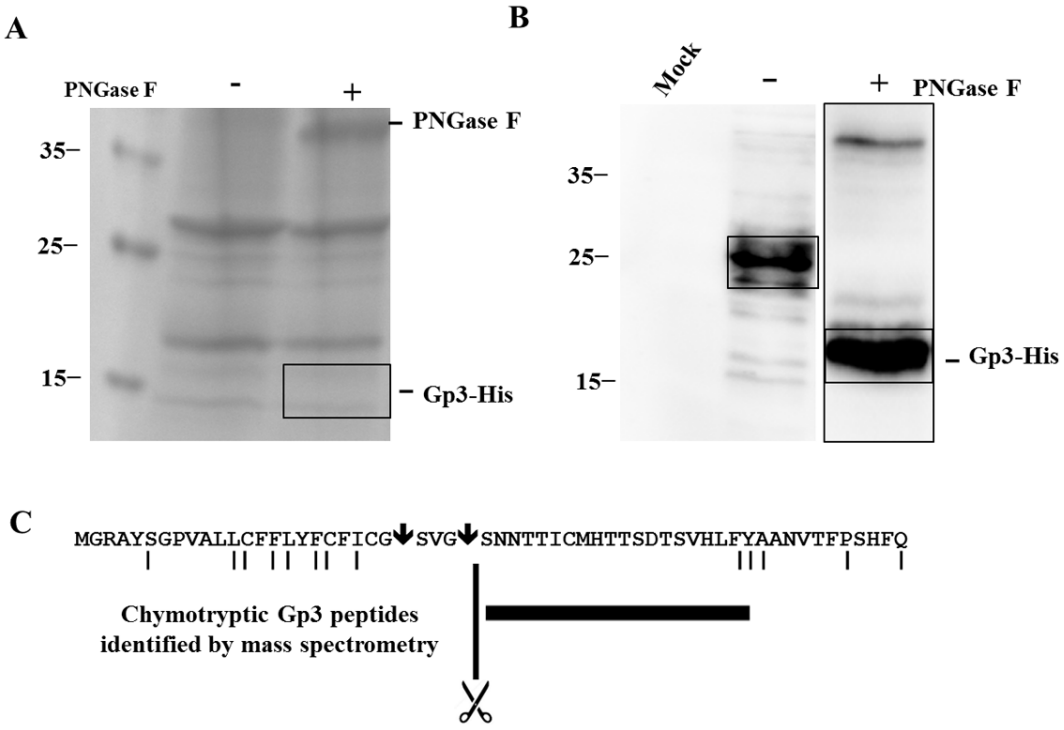


Figure 3.16. Identification of the signal peptide cleavage site of recombinant Gp3-His by mass spectrometry. (A) Gp3-His protein from EAV Bucyrus strain (amino acids 1–163 fused to a His-tag) was expressed in TriEx™ Sf9 insect cells. The protein was enriched by the cell lysate using a Ni- NTA agarose beads. The left lane shows the purified proteins prior to PNGase F digestion. The eluted proteins were digested with PNGase F and subjected to a reducing SDS-PAGE and Coomassie staining. The left lane shows the purified proteins prior to PNGase F digestion. The band corresponding to the

size of deglycosylated Gp3-His (black box) was excised from the gel, digested with chymotrypsin, and analysed with LC–MS/MS; B) The right lane shows the band corresponding to the size of deglycosylated Gp3-His (black box) in the western blot with anti-His antibody and the left lane shows the partially purified protein prior to PNGase F digestion; C) An illustrative result from the mass spectrometry of Gp3-His. The first 57 residues of Gp3-His are shown with the predicted chymotrypsin cleavage sites marked with black lines and the signal peptide cleavage site indicated by an arrow and scissors. Black bars denote the identified chymotryptic peptides.

3.2.8 The Gp3 is anchored to the membrane by its C-terminal transmembrane region

Previous studies indicated that the EAV Gp3 might be a type-II membrane protein, which means it anchors to the membrane by its N-terminal signal peptide (Wieringa *et al*., 2003a). Gp3 contains an 18 amino acid long hydrophobic region, and I assumed that Gp3 is most likely anchored to the membrane by its C-terminal hydrophobic region. To confirm this hypothesis, I expressed both full length Gp3-His and ΔC-Gp3-His in insect cells. Cells were lysed with a lysis buffer to extract the intracellular protein. I performed a western blot with anti-His antibodies using both secreted and intracellular protein samples of full-length Gp3-His and ΔC-Gp3-His, and performed western blot with the anti-his antibody. I was able to detect Gp3 protein in both secreted and intracellular samples in the case of ΔC-Gp3-His (Fig. 3.17A). In the case of full-length Gp3-His, I could only detect the protein in the intracellular fraction, and no Gp3-His was detectable from the secreted fraction (Fig. 3.17B). This result indicates that Gp3 is only secreted when the C-terminal hydrophobic region is deleted, which means it anchors to the membrane by its C-terminus hydrophobic region.

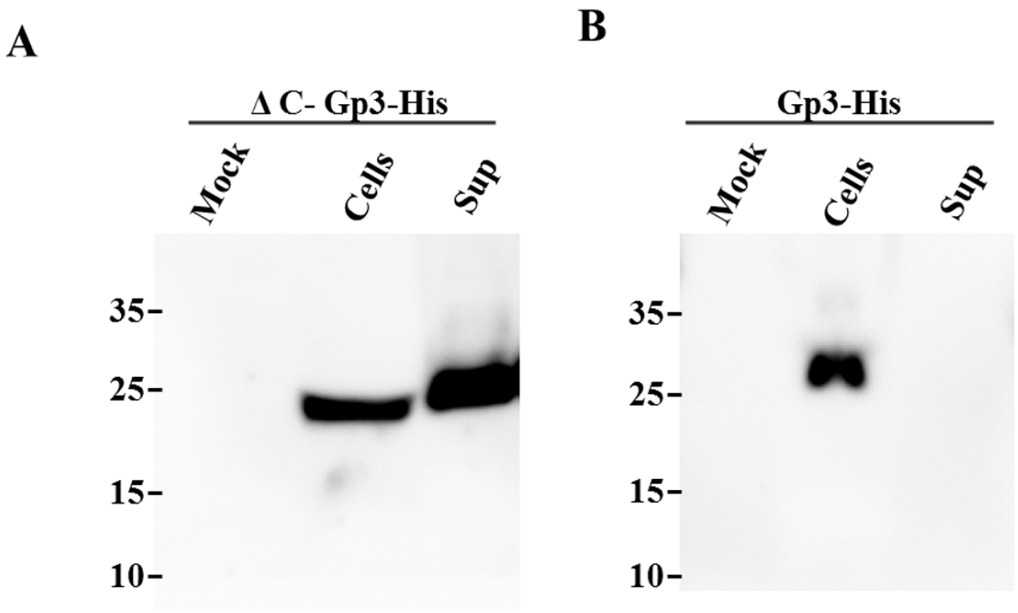


Figure 3.17. Deletion of C-terminal hydrophobic region of Gp3 enables secretion. A) ΔC-Gp3-His cell lysate (Cell) cell culture supernatant (Sup) was subjected to western blotting using anti-His antibody B) Gp3-His cell lysate (Cell) and cell culture supernatant (Sup) was subjected to western blotting using anti-His antibody. Mock indicates that cells were infected with a non-recombinant baculovirus. Molecular weight markers in kDa are shown on the left side of the blots.

3.2.9 Comparison of mammalian and insect cell-expressed Gp3-His

It has been demonstrated that the Gp3-wt signal peptide is not cleaved in CHO cells (Matczuk *et al.,* 2014). However, my results with an insect cell-expressed Gp3 showed that the signal peptide is cleaved, and the overlapping sequon NNTT is only glycosylated at one asparagine. Hence, I wanted to test the processing difference between the insect- and mammalian-expressed Gp3-His. To perform this, I expressed the same Gp3-His construct in CHO cells, which I used for the insect cell expression. I digested the glycans using PNGase F from both insect- and mammalian-expressed Gp3 and run them side-by-side to compare the size difference. The western blot result showed that the Gp3 expressed in CHO cells is running slightly higher compared to the Gp3-His expressed in TriEx™ Sf9 cells (Fig. 3.18). This indicates that the Gp3 expressed in insect cells is different from the mammalian cell-expressed Gp3.

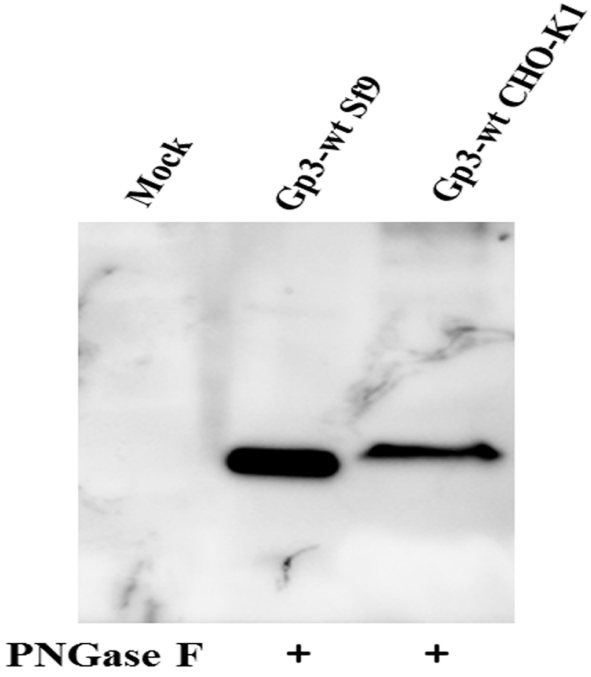


Figure 3.18. Insect cell-expressed Gp3 runs lower than the mammalian-expressed Gp3. Gp3-wt was expressed in both Sf9 and CHO cells. The protein was digested with PNGase F and subjected to SDS-PAGE and western blotting with anti-his antibody. Mock indicates that cells were infected with a non-recombinant baculovirus. Molecular weight markers in kDa are shown on the left side of the blots.

3.2.10 Signal peptide cleavage from PRRSV Gp3 ectodomain expressed in insect cells

PRRSV Gp3 is predicted to be cleavable from the mature protein (Table 3.2). In EAV, it was demonstrated that the Gp3 signal peptide is not cleaved and that the double N-glycosylation sequon nearby the signal peptide prevents its cleavage (Matczuk *et al.,* 2014). In PRRSV Gp3, there is only one N-glycosylation sequon near the signal cleavage site. Therefore, I wanted to determine whether PRRSV Gp3 signal peptide is cleaved or not. I expressed the ΔC-Gp3-His together with ΔC-Gp2-HA and ΔC-Gp4-Twin Strep as a complex in TriEx sf9 cells using baculovirus expression system and partially purified the ΔC-Gp3-His, ΔC-Gp2-HA, and ΔC-Gp4-twin Strep complex from cell culture supernatant using StrepTactin Superflow column. Glycans were removed using PNGase F digestion. The protein sample was separated by SDS PAGE and stained with Coomassie stain to visualise the protein bands. The band corresponding to the size of deglycosylated ΔC-Gp3-His (running at 19 kDa) was excised from the gel (Fig. 3.19A), and the proteins were digested using elastase. The resulting digested fragments were extracted from the gel slices and analysed by liquid chromatography-tandem mass spectrometry (LC MS/MS).

ΔC-GP3-His was identified with high certainty in all samples analysed. In ΔC-Gp3-His samples digested with elastase, the peptide VVAGSNTTYCFWFPLVR was identified. This sequence matches the amino acids 23–39 of ΔC-Gp3-His. All other possible peptides were identified, but not the ones corresponding to the signal peptide. This result indicates that the signal peptide from the insect cell-expressed PRRSV ΔC-Gp3-His is cleaved off from the mature protein (Fig. 3.19B). This result exactly matched the signal peptide cleavage site predicted by signal 4.1(Table 3.2). The peptides corresponding to the Gp4 were identified with good sequence coverage. However, the result is not conclusive due to presence of both SP-cleaved and uncleaved Gp4 population (data not shown).

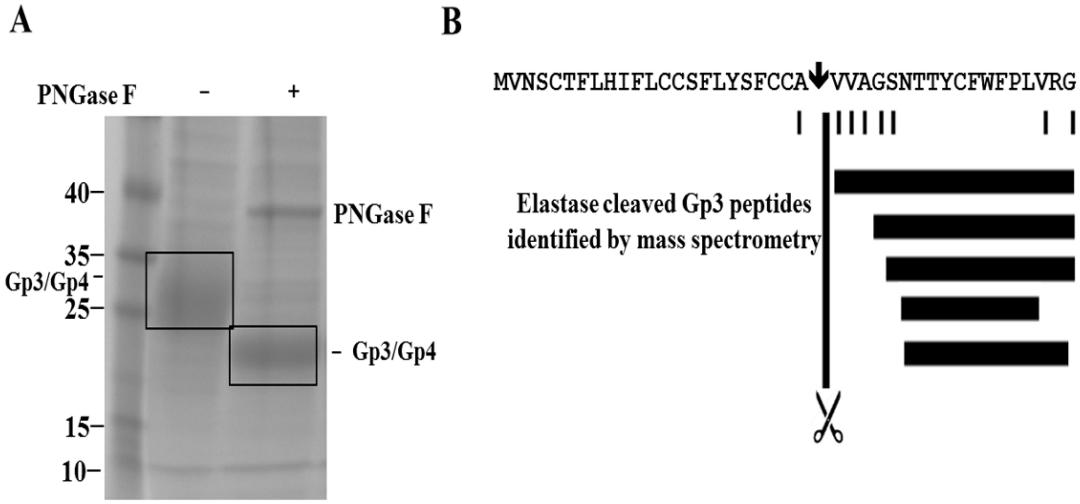


Figure 3.19. Identification of the signal peptide cleavage sites of recombinant ΔC-Gp3-His by mass spectrometry. (A) ΔC-Gp3-His protein from PRRSV VR-2332 strain (amino acids 1–181 fused to a His-tag) was expressed in insect cells by infection with a recombinant baculovirus. ΔC-Gp3-His protein was enriched by the supernatant using a HisTrap column (GE Healthcare*, Uppsala, Sweden*). The eluted proteins were digested with PNGase F and subjected to reducing SDS-PAGE and Coomassie staining. A) The band corresponding the size of deglycosylated ΔC-Gp3-His (Black box; coinciding with ΔC-Gp4-onestrep) was excised from the gel, digested with elastase, and analysed by LC–MS/MS. The left lane shows the purified proteins prior to PNGase F digestion; B) An illustrative result from mass spectrometry Gp3-His-wt. Peptides corresponding to ΔC-Gp3-His are the predominant peaks in the MS spectra. The first 57 residues of Gp3-His-wt are shown with the predicted elastase cleavage sites marked with black lines, and the

signal peptide cleavage site is indicated by an arrow and scissors. Elastase-cleaved peptides that were identified are denoted as black bars.

3.3 Heterodimerization of PRRSV Gp5-M and signal peptide cleavage of Gp5

3.3.1 Disulphide linkage between PRRSV Gp5 and M protein in insect cells

The major envelope proteins Gp5 and M of EAV form a disulphide-linked heterodimer when co-expressed. Cys 48 of Gp5 and Cys 8 of M protein are essential for the heterodimerisation (Snijder *et al.,* 2003). In this study, I wanted to examine the heterodimerisation of Gp5-M of PRRSV. Therefore, I cloned full-length Gp5 and M nucleotide sequences into insect cell vectors. A C-terminal histidine tag to the Gp5 and HA tag to the M were added for the detection of the protein in the western blot. I made baculovirus-expressing Gp5-His and M-HA proteins and infected TriEx™ Sf9 cells to express the Gp5-His/M-HA complex. Three days post-infection I lysed the cells using lysis buffer. The proteins were enriched using Ni-NTA agarose beads and then separated by SDS-PAGE. The western blot was performed using anti-His and anti-HA antibodies under reducing (+DTT) and non-reducing (-DTT) conditions. I observed a band at 26 kDa with anti-His antibody under reducing conditions, which corresponds to the glycosylated Gp5-His. I identified a band at 41 kDa under non-reducing conditions, which corresponds to the Gp5-His/M-HA dimer (Fig. 3.20A). The Western blot with HA tag showed a band at 19 kDa under reducing conditions, which corresponds to the M protein. Under non-reducing conditions, I detected a band at 41 kDa, which matches the size of M-HA/Gp5-His (Fig. 3.20A). These results indicate that Gp5-His/M-HA can form a heterodimer when co-expressed in insect cells

3.3.3 Determining cysteine residue involved in the heterodimerization between Gp5 and M

Disulphide linkage between Gp5/M is crucial for the infectivity of the arteriviruses (Snijder *et al*., 2003). We experimentally demonstrated that Gp5/M can form disulphide linked dimer when co-expressed (Fig 3.20A). Hence, I wanted to determine the cysteine residue involved in the Gp5/M heterodimerisation. After signal peptide cleavage, there is only one cysteine (C48) remaining in the ectodomain of Gp5 and this is the most likely cysteine that can form the disulphide linkage with M. Therefore, I mutated cysteine 48 of

Gp5-His to serine. I made a baculovirus-expressing Gp5 C48S-His (tagged with His tag at C-terminus) and M protein (tagged with HA tag C-terminus) and infected TriEx™ Sf9 cells. I detected a band around 19 kDa with anti-HA antibody under reducing conditions and a band around 26 kDa when probed with anti-His antibody. No band was observed at 41 kDa under non-reducing conditions either with anti-HA or anti-His (Fig. 3.20B). This result indicated that the heterodimerisation between Gp5/M is abolished when cysteine 48 of GP5-His is mutated to serine. This result confirms my hypothesis that cysteine 48 is responsible for the disulphide linkage of Gp5 with M.

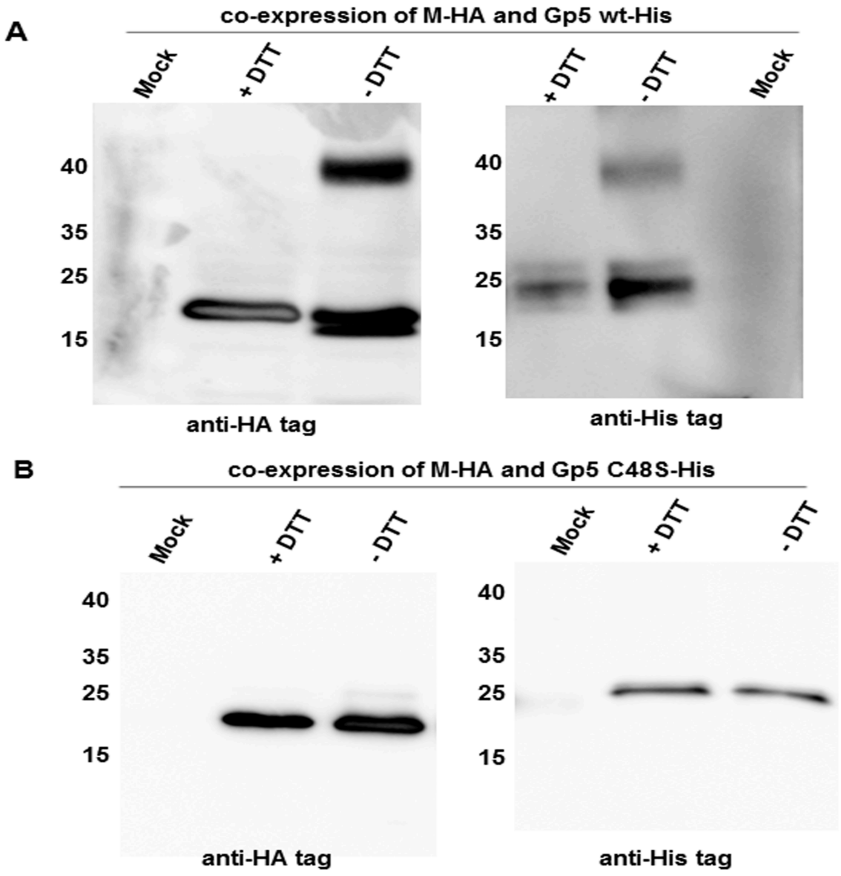


Figure 3.20. Cysteine 48 of PRRSV-Gp5 forms a disulphide bond with the M protein. Full-length genes of Gp5 from PRRSV VR-2332 (fused via the linker GS to a 7×His-tag) and M (fused via the same linker to the HA-tag) were co-expressed with baculovirus in insect cells. The cells were lysed with reducing (+DTT) or non-reducing (−DTT) lysis buffer, and proteins were separated by SDS-PAGE (equal amounts were loaded onto the gel). The Western blot was performed with anti-HA and anti-His antibodies. (A) Western blot of cells expressing Gp5-His wt and M-HA. (B) Western blot of cells expressing Gp5-His C48S with M-HA. The numbers on the left side of each blot

indicate the mobility of molecular weight markers in kDa. Mock indicates that cells were infected with a non-recombinant baculovirus.

3.3.4 The analysis of the signal peptide cleavage from a recombinant Gp5 expressed in insect cells

Prediction result from SignalP 4.1 showed that the Gp5 signal peptide cleaved at two sites site 1 (A26|V27) with minor probability and site 2 (A31|S32) with major probability. Therefore, I wanted to determine whether and where the signal peptide is cleaved. I expressed the Gp5-His together with M-HA as a complex in TriEx™ Sf9 cells; the Gp5-His/M-HA complex was enriched using Ni-NTA beads from the cell lysate. The Gp5-His was then digested with PNGase F to remove the glycans. The PNGase F-digested protein sample was separated by SDS-PAGE and stained with Coomassie stain. The band corresponding to the size of deglycosylated Gp5-His (running at 16 kDa) was excised from the gel (Fig. 3.21B), and the proteins were digested using chymotrypsin. The resulting fragments after the digestion were extracted from the gel slices. The consequent peptides were analysed by LC MS/MS.

GP5-His was identified with high certainty in all samples, and the Gp5-His was the major hit with high confidence in the MASCOT search. In the samples subjected to chymotrypsin digest, the fragment SDDSSSHLQLIYDLTL-CELDGTDWLANK was identified. This matches the residues 32–59 of GP5, which corresponds to the site 2 (A31|S32) of the prediction result. Fragments derived from GP5 starting with V27 (VLANASDDSSSHLQ) were also detected, indicating that cleavage also occurred at site 1 (A26|V27) (Fig. 3.21A). There were no Gp5-His peptides that matched the signal peptide identified from the N-terminus. This result suggests that the signal peptide from the insect cell-expressed PRRSV Gp5-His ectodomain is cleaved off from the mature protein. This result matches the prediction result of SignalP 4.1.

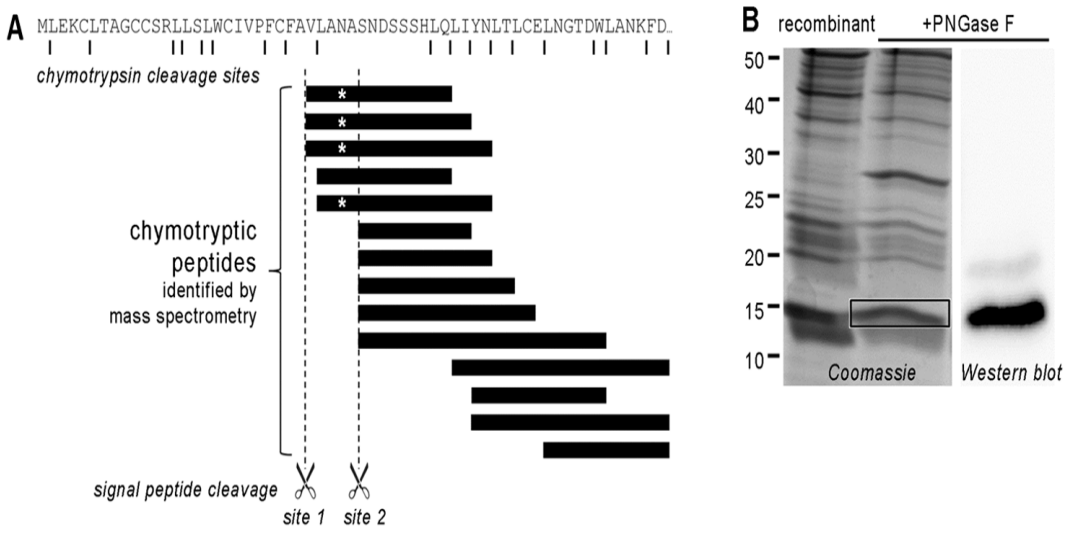


Figure 3.21. Identification of the signal peptide cleavage site of insect cell-expressed GP5 by mass spectrometry. A) PRRSV GP5 (with His tag) and M (with HA tag) were co-expressed in insect cells. Following cell harvesting and lysis, GP5–M was enriched using Ni-NTA agarose (binding to GP5–His). The eluted protein was left untreated or digested with PNGase F and subjected to reducing SDS-PAGE and Coomassie staining (left-hand side) or Western blot (anti-His tag antibody, right-hand side). The deglycosylated GP5 band (black box; coinciding with M) was cut out, digested with chymotrypsin, and analysed by LC-MS/MS. B) A demonstrative result from mass spectrometry of GP5. The first 61 residues of GP5 are shown with the positions of the predicted chymotrypsin cleavage sites (black lines) and the putative signal peptide cleavage sites (broken lines). The identified chymotryptic peptides are represented as black bars.

4. Discussion

Arteriviridae family members EAV and PRRSV cause significant economic losses in equine and swine industry. Arteriviruses contain glycoprotein Gp5 (Major) and Gp2/3/4 (Minor) on the virion surface. Previous studies on arteriviruses mainly described their replication and pathogenesis (Snijder *et al.*, 2013; Balasuriya UBR *et al*., 2013). Minor glycoproteins Gp2/3/4 are demonstrated to be essential for cell tropism (Wissink *et al.,* 2005; Tian *et al.,* 2012; Snijder *et al.*, 2013). Minor glycoproteins have not been well studied, and very limited information is available so far despite their essential role in virus entry and infectivity. Therefore, in this thesis I studied the Gp2/3/4 complex formation in insect cells and characterised the minor glycoproteins for signal peptide cleavage, N-linked glycosylation, and disulphide linkage. I also determined the signal peptide cleavage site of Gp5 and disulphide linkage site between Gp5 and M of PRRSV North American strain VR-2332.

4.1 Gp2, Gp3, and Gp4 complex formation in insect cells

Gp2, Gp3, and Gp4 are reported to form a disulphide linked complex and exist as a trimer in the virions (Wieringa *et al.,* 2003; Wissink *et al.,* 2005). In this thesis, I studied complex formation between EAV and PRRSV Gp2, Gp4, and Gp3 ectodomains in insect cells. My results demonstrate that EAV Gp2 and Gp4 form a disulphide-linked complex and are secreted in glycosylated form when the C-terminal hydrophobic regions are deleted (Fig. 3.2 &3.3). However, Gp3 is not associated with the Gp2/4 complex when co-expressed in insect cells, but it is expressed and secreted in glycosylated form. My results confirm a previous finding that Gp3 does not associate with Gp2/4 in cells and it only happens in virus particles (Wieringa *et al.,* 2003). A significant observation from my results is that Gp2/4 ectodomains can form a heterodimer when co-expressed in cells.

My results from the experiments involving co-expression of PRRSV Gp2/3/4 ectodomains reveal that they can form a disulphide linked complex and are secreted in glycosylated form (Fig. 3.8 and 3.9). Unlike Gp3 from EAV, Gp3 of PRRSV formed a covalent-linkage with Gp2/4 (Fig. 3.9C). A significant outcome of this study is that the Gp2/3/4 ectodomain of PRRSV can form a hetero-oligomeric complex when co-expressed in insect cells.

Under non-reducing conditions, I observed several disulphide-linked multimeric forms with EAV and PRRSV minor glycoproteins Gp2/3/4 (Fig. 3.3 and 3.9). Similar observations have been reported that the minor glycoproteins of EAV form multimeric complexes when co-expressed in mammalian cells by vaccinia expression system (Wieringa *et al.,* 2003b). The EAV Gp4 multimers are unquestionably caused by disulphide linkage as they were not present under reducing conditions. Previous data indicated that envelope protein E interacts with Gp2/3/4 complex and influences the integration of EAV minor glycoproteins into the virions (Wieringa *et al.,* 2004). Hence, it is possible that the E protein might play a role in a stable Gp2/3/4 trimer formation. It was also reported that Gp4 of PRRSV interacts with Gp5 (Das *et al*., 2010). Therefore, it is possible that a homogeneous Gp2/3/4 trimer formation might require its interaction with Gp5 or E protein. Further experiments are needed to co-express the Gp5 or E protein with the Gp2/3/4 complex in order to explore their influence on complex formation of minor glycoproteins through covalent linkage.

4.2 The effect of cysteine mutations in EAV Gp4 on aberrant multimer formation

I observed various disulphide-linked Gp2/4 hetero-oligomers under non-reducing conditions (Fig. 3.2A). I assumed that the free cysteine residues in Gp2/4 might be causing the multimer formation. It was previously reported that the C102 of Gp2 forms a disulphide linkage with Gp4, and the remaining two cysteines in Gp2 ectodomian form an intra-molecular linkage (Wieringa *et al.,* 2003) (Fig. 3.1). Hence, there are no free cysteine residues available in Gp2 ectodomain that can form covalent linkages. Therefore, I mutated the three cysteine residues present in the ectodomain of Gp4 after signal peptide cleavage between aa C21 and T22 and investigated their influence on disulphide linkage and multimer formation (Fig. 3.5C). My experiments with single and double cysteine mutations in Gp4 were unable to abolish the formation of Gp2/4 heterodimers and multimers (Fig. 3.6 A and B). These data indicate that any of the cysteines in Gp4 could build the Gp2/4 dimer, and it is not clear which cysteine residues are involved in the Gp2/4 multimer formation. This is a surprising result because, according to the previously proposed trimer model, one among three cysteines present in the Gp4 ectodomain is supposed to form one disulphide linkage with Gp2 and one with Gp3 (Wieringa *et al.,* 2003). Thus, there should be only one free cysteine available, which can contribute to covalent bonds. Since the Gp3 association with Gp2/4 dimer is not observed in the cells,

one can speculate that the cysteine residue responsible for covalent linkage between Gp3 and Gp4 can also be involved in the artificial disulphide linkage. However, my experiments with double cysteine mutations in Gp4 showed no effect on Gp2/4 oligomerisation and rule out this possibility. Due to formation of aberrant multimers, a homogeneous Gp2/4 complex expression is not feasible, which is essential for structural studies. Thus, structural studies are not possible with these proteins.

EAV virus particles were not infectious when C26 and C77 of Gp4 were mutated, and the plaque size of the C50S mutant virus was smaller when compared to wild type (Veit *et al.,* 2014). Therefore, it is most likely that all the cysteines in Gp4 are essential for the virus infectivity. A similar mutation in the proteins expressed in cells has no effect on dimer formation. One could speculate that the cysteines from uncleaved SP might be involved in the disulphide linkage. Nonetheless, the mass spectrometry results undoubtedly showed that there were no peptides identified from the SP region. Therefore, it is very unlikely that the cysteines from SP can play a role. Overall, my results suggest that Gp2/4 might be covalently linked with two disulphide linkages, which is contradictory to the previous results. However, it is also possible that the cysteine in Gp2 ectodomain causes these aberrant multimers if the intramolecular disulfide bond between Cysteine 48 and 137 is not formed when expressed in cells. Further experiments are needed to identify either one or two cysteines of EAV Gp4 that are involved in the disulphide linkage with Gp2 and to understand the complex covalent association between minor glycoproteins Gp2, Gp3, and Gp2.

4.3 The cleavage of EAV Gp2 and Gp4 signal peptides

It has been shown by a previous study that the signal peptide (SP) is cleaved from EAV Gp4 (Wieringa *et al.*, 2002). Also, SignalP 4.1 calculated that Gp2 and Gp4 contain cleavable SPs. My results from mass spectrometry reveal that the resultant cleavage sites of Gp2/4 are exactly the same as the software prediction (Fig. 3.5 and Table 4.1). All the SP cleavage sites identified in my study contain small and neutral amino acids at -3 and -1 position upstream to the cleavage site. Therefore, the cleavage complies to the rule that the signal peptidase requires small amino acids at positions -1 and -3 (Renu Tuteja, 2005) (Table 4.1). Secretion of EAV Gp2 and Gp4 when the C-terminal region is removed and

the lack of SP from the secreted proteins indicate that both Gp2 and Gp4 anchor to the membrane with their C-terminal region.

Table. 4.1. SP cleavage sites determined by mass spectrometry

Protein	Predicted SP cleavage site	Identified SP cleavage site by LC-MS
EAV Gp2	A29↓W30	A21↓W30
EAV Gp4	C21↓T22	C21↓T22
EAV Gp3	G22↓S24 (Site 1) G26↓S27 (Site 2)	G22↓S24 (Site 1) G26↓S27 (Site 2)
PRRSV Gp3	A22↓V23	A22↓V23

4.4 N-glycosylation of EAV and PRRSV Gp2/3/4

The minor envelope glycoproteins of EAV and PRRSV are heavily glycosylated. My results reveal that all predicted N-glycosylation sites in EAV and PRRSV Gp2/3/4 ectodomains are efficiently used (Fig. 4.1 and Fig. 3.10). Glycosylated EAV and PRRSV Gp3 exist as two bands in the western blot. The two-band pattern of EAV Gp3 is due to the partial usage of one of the two N-glycosylation sites in the overlapping sequon NNTT near the signal peptide cleavage site (Matczuk *et al.,* 2014). Gp3 of PRRSV does not contain any overlapping sequon. Hence, it is most probable that one among the six potential N-glycosylation sites is partially used.

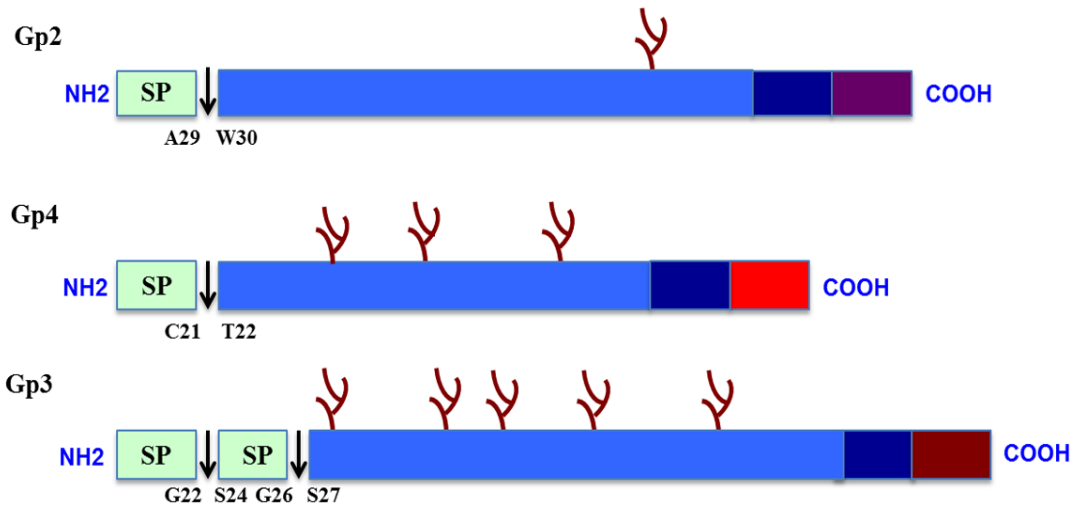


Figure. 4.1 Schematic of EAV Gp2, GP4 and Gp3 SP cleavage and N-glycosylation. Signal peptide cleavage sites determined in this study are denoted with an arrow mark (↓) with the respective amino acid abbreviation and number. Identified N-linked glycans are shown as branches. SP - signal peptide; light green - signal peptide, light blue - ectodomain of EAV minor glycoprotein; dark blue - short glycine-serine linker; dark purple - HA tag; red - Strep II tag; dark red - His tag

4.5 Signal peptide is cleaved from EAV and PRRSV Gp3 in insect cells, but not from mammalian cell-expressed Gp3

Gp3 is a heavily glycosylated protein of EAV and PRRSV (Wieringa et al., 2003; de Lima et al., 2009). Both EAV and PRRSV Gp3 are predicted to have a cleavable SP at the N-terminus and a hydrophobic C- terminal region with which it is attached to the membrane. *In vitro* translation experiments with canine microsomes have demonstrated that the SP of EAV Gp3 is not cleaved (Wieringa *et al.*, 2002). It was assumed that the Gp3 might attach to the membrane through its uncleaved SP (Type II membrane topology) or by both N and C terminus (Type III) (Wieringa *et al.*, 2003a).

SignalP 4.1 calculated that the SP of EAV Gp3 is cleaved with high probability (Table 3.2 and Fig. 4.2). This result is contradictory to the previous findings. However, bioinformatics software tools only consider the peptide sequence and do not take other co-translational

protein modifications into consideration. Gp3 of EAV harbours an overlapping sequon NNTT two amino acids away from the predicted SP cleavage site. Previous data showed that both asparagines in the overlapping sequon are N-glycosylated, and their mutation leads to the cleavage of SP in mammalian cells (Matczuk *et al.,* 2014).

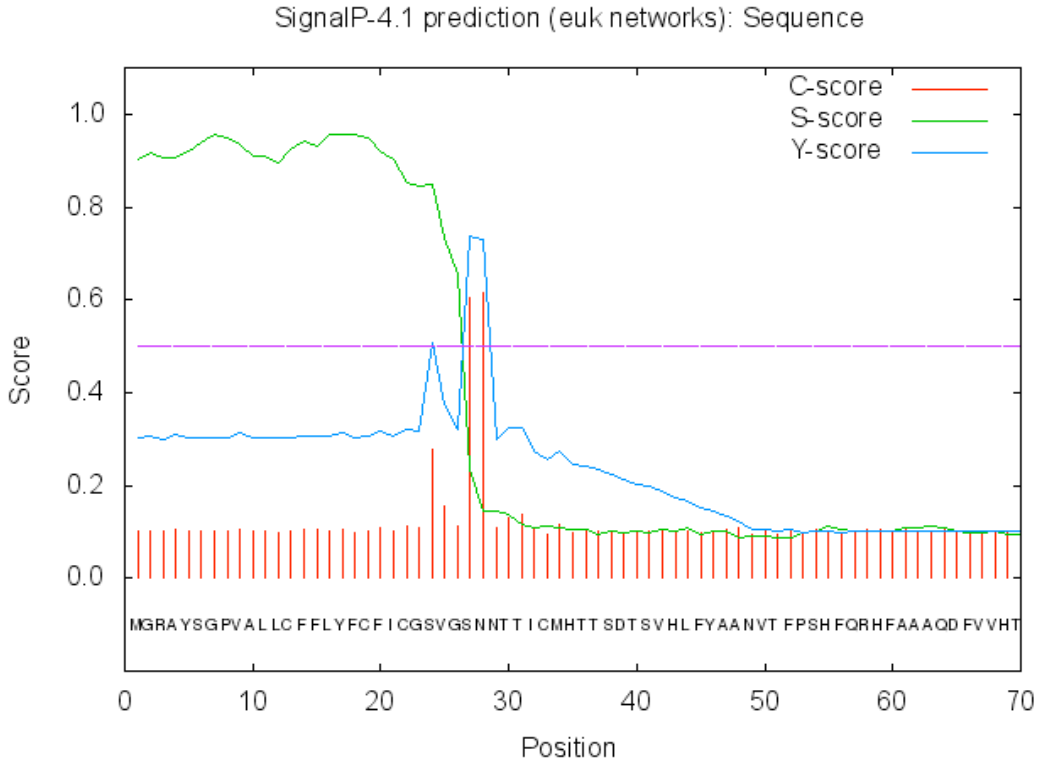


Figure 4.2. Result from SignalP 4.1 prediction software for Gp3 of the EAV (Bucyrus strain). Red vertical lines show C-score (raw cleavage site score), which is high after the signal peptide cleavage site. Green line shows S-score (signal peptide score), which distinguishes the signal peptide from the mature part of the protein. Position - amino acid position; Y-score (blue line) - geometric average of C and S scores, which shows most probable signal peptide cleavage site than the raw C score.

Surprisingly, my results from mass spectrometry analysis demonstrate that the SP from insect cell-expressed Gp3-wt and Gp3-double mutant (asparagines 28 and 29 to serine) is cleaved at two sites (between amino acids G23|S24 and G26|S27) (Fig. 4.1). My results with insect cell-expressed PRRSV Gp3-ectodomain (North American strain VR-2332) also show that the SP is cleaved from the mature protein (Fig. 3.18). These results match with the software predictions (Table 4.1) but are contradictory to the previous findings where

the SP from mammalian cell-expressed Gp3 is shown to be not cleaved (Wieringa *et al.*, 2002; Matczuk *et al.,* 2014).

However, the insect cell-expressed Gp3 is different from mammalian-expressed Gp3 (Table 4.2). The overlapping sequon of EAV Gp3 is only glycosylated at one asparagine in insect cells (Fig. 3.12 and 3.15). In mammalian cells, even one glycan at the overlapping sequon is sufficient to block the SP cleavage, which is not the case in insect cells. Insect cell-expressed EAV Gp3 runs as two bands after deglycosylation (Figure 3.11). First, I thought this could be due to the presence of Gp3 population with cleaved and uncleaved SP. However, mass spectrometry results revealed that both bands lack SP, indicating that they are due to some insect cell-specific unknown protein modification of Gp3. In mammalian cells, this appearance of two bands is due to partial glycosylation of overlapping sequon NNTT, and only one band was observed after the glycans were digested (Matczuk *et al.,* 2014).

My results demonstrate that mammalian-expressed Gp3 is slightly larger in size than the insect-cell expressed Gp3, and this size difference is most likely due to the presence of a SP in mammalian-expressed Gp3 (Fig. 3.18). These observations suggest that the Gp3 expressed from insect and mammalian cells are different (summarised in Table 4.2). It is unclear why Gp3 SP is processed differently in insect cells. To explain these contradictory results, it is necessary to determine the SP cleavage site from a virus-derived Gp3. But, Gp3 is a minor component of the virion, and there are no perfect methods to purify the arteriviruses. Thus, it is not possible to obtain enough virus-derived Gp3 protein for the mass spectrometry.

Table 4.2. Summary of differences in SP cleavage and N-glycosylation of Gp3 expressed in mammalian and insect cells.

Protein + expression host	Signal peptide cleaved	Number of N-glycans attached	Number of bands observed before deglycosylation	Number of bands observed after deglycosylation
EAV Gp3 insect	Yes	Five	Two	Two
EAV Gp3 mammalian	No	Six	Two	One

PRRSV Gp3 insect	Yes	Six	Two	One
PRRSV Gp3 mammalian	Unknown	Six	Two	One

4.6 Hydrophobic C-terminus has no influence on EAV Gp3 SP cleavage but acts as a transmembrane anchor

A previous study proposed that deletion of the C-terminal transmembrane region delays the SP cleavage of the human cytomegalovirus (HCMV) US11 protein (Rehm *et al.,* 2001). This observation suggests that the C-terminus of viral transmembrane proteins might play a role in their SP cleavage. My results reveal that Gp3 SP is cleaved with or without C-terminus, and no difference was observed in glycosylation (Fig. 3.13 and 3.16). Thus, the C-terminus has no influence on Gp3 SP cleavage and N-glycosylation. Previous studies indicated that Gp3 is either anchored to the membrane with its uncleaved SP (type II topology) or anchored with both C and N terminus (type III topology) (Fig. 1.6B). My results show that Gp3 is secreted only when the C-terminal hydrophobic region is deleted. Similar results were obtained with Gp3 expressed in mammalian cells (Matczuk *et al.,* 2014). These observations from my study indicate that EAV Gp3 is anchored to the membrane by its C-terminal hydrophobic region (Fig. 3.17). These are significant findings, which have not been reported before.

4.7 Signal peptide cleavage and presence of the decoy epitope in PRRSV Gp5

PRRSV is an important pathogen in the swine industry. Despite its economic impact, the molecular information on PRRSV structural proteins is limited. The major glycoprotein Gp5 ectodomain harbours a neutralising epitope, which has significant immunological importance. The neutralising antibodies appear late after infection, which has been reported as one of the leading causes of persistent PRRSV infection in infected animals. It was hypothesised that the Gp5 of PRRSV North American strain VR 2332 possesses a non-neutralising immunodominant 'decoy epitope', which is situated upstream of the neutralising epitope. SignalP 4.1 predicted that the Gp5 SP is cleaved at two sites, cleavage site 1 (A26|V27) and 2 (A31|S32). The 'decoy epitope' is only present in the mature protein if the GP5 SP is cleaved at site 1. Therefore, the presence of a decoy epitope depends on whether and where the SP is cleaved.

I answered this question in this study by determining the SP cleavage site from a recombinantly-expressed Gp5 by mass spectrometry (Thaa *et al.,* 2013). The result reveals that the PRRSV Gp5 (American type 2 reference strain VR-2332) SP is cleaved at both site 1 and site 2 (Fig. 3.21 & 4.2). The exactly same result was observed with Gp5 from insect cells and virions (Thaa *et al.,* 2013). Since it is not possible to quantify the ratio of cleavage at site 1 and site 2 with mass spectrometry, no conclusions are possible regarding the abundance of Gp5 fraction with 'decoy epitope'. However, biochemical experiments have demonstrated that the significant portion of Gp5 SP cleavage happened at site 2 (Thaa *et al.,* 2013). Although the mass spectrometry results showed the presence of Gp5 molecules with 'decoy epitope', the biochemical data suggest that the abundance of this Gp5 population is minute. Based on our findings, it is possible that only a minor fraction of Gp5 molecules with 'decoy epitope" are incorporated into the mature virions.

It is hard to understand how such a small portion of Gp5 species with 'decoy epitope' hinders the antibody response from the neutralising epitope. The fact that the mature Gp5 lacks the 'decoy epitope' could be established by infecting the pigs with recombinant viruses containing a homogenous population of GP5 molecules that either are cleaved only at site 1 (which harbours the 'decoy epitope') or entirely at site 2 such that the mature Gp5 lacks the 'decoy epitope'. If the viruses that harbour these mutations (which enable the cleavage exclusively at either site 1 or site 2) show similar growth kinetics to the wild type, the pigs can be infected with the resultant mutant viruses. The animals infected with these recombinant PRRSV viruses would provide an answer for the role of 'decoy epitope' in the persistent infections. Recombinant PRRSV viruses completely lacking the 'decoy epitope' might stimulate a robust neutralising antibody response and therefore could be strong vaccine candidates. Designing such mutations might be complicated as they interfere with the expression of ORF5a, which encodes membrane protein 5a.

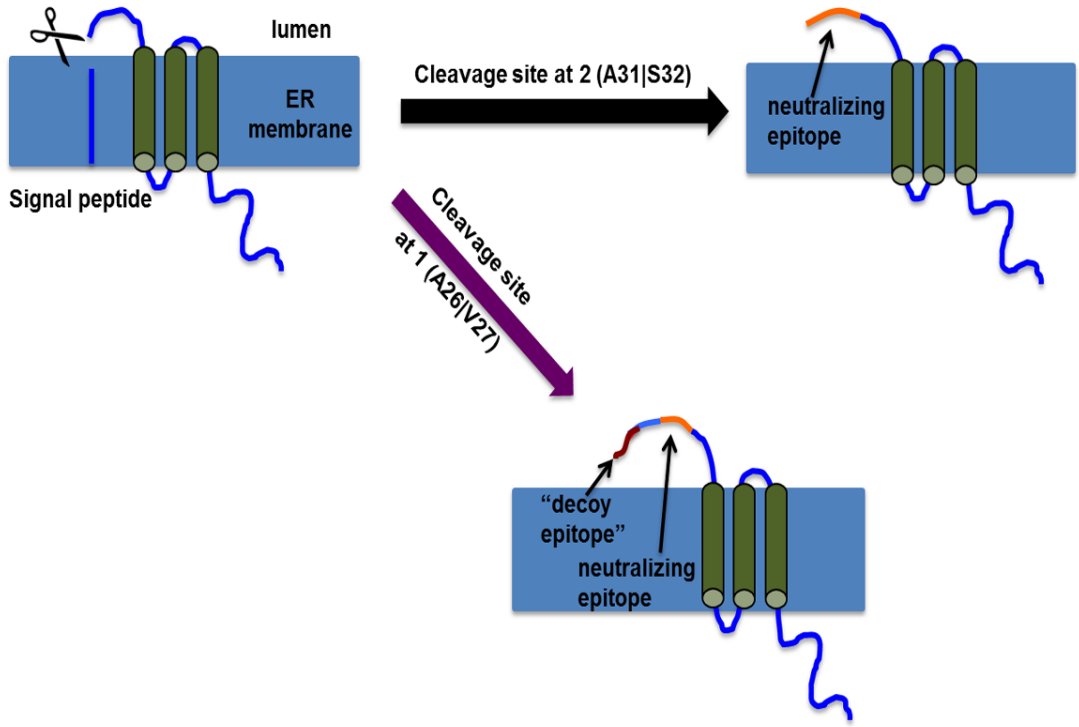


Figure 4.3. Schematic representation of GP5 fractions as demonstrated in this study. The signal peptide of GP5 is cleaved at site 1 (A26|V27, dark purple arrow) and site 2 (A26|V27; black arrow). The 'decoy epitope' (dark red) is preserved only in GP5 cleaved at site 1. The neutralising epitope (orange) is present in either case.

4.8 Gp5 is covalently linked to M protein by Cysteine 48

Gp5-M exists as disulphide-linked dimer in PRRSV and EAV particles. The cysteine residues involved in EAV Gp5/M heterodimerisation have already been determined and shown to be essential for viral infectivity (Snijder *et al.,* 2003). Cysteine residues involved in PRRSV Gp5 (American type 2 reference strain VR-2332) and M have not been identified so far. My results indicate that Gp5/M forms a disulphide-linked heterodimer when co-expressed in insect cells (Fig. 3.20A). Mutating the only cysteine residue, Cys-48 of Gp5 remaining in the ectodomain after SP cleavage, completely abolished heterodimerization of Gp5/M (Fig. 3.20 B). PRRSV type 1 (European, strain/Lelystad strain) contains a cysteine residue at the position 50. It is the only cysteine available in the ectodomain if the signal peptide is cleaved. Therefore, one can assume that this is the most probable cysteine that forms the disulphide linkage with M protein in comparison to Gp5 of type 2 strain.

My results also show that the lack of heterodimerization did not affect the N-glycosylation of the Gp5. Further experiments are needed to explore the effect of the Gp5 Cysteine 48 mutation and the lack of heterodimerization of PRRSV Gp5/M on virus infectivity and budding.

Bibliography

Aebi, M. 2013. N-linked protein glycosylation in the ER. Biochim Biophys Acta 1833(11), 2430-7

Auclair, S.M., Bhanu, M.K., Kendall, D.A., 2012. Signal peptidase I: Cleaving the way to mature proteins. Protein Science 21, 13-25.

Balasuriya UB, Go YY, MacLachlan NJ.,2013. Equine arteritis virus. Vet Microbiol. 29;167(1-2):93-122

Braakman, I. and Bulleid, N.J., 2011. Protein folding and modification in the mammalian endoplasmic reticulum. Annu Rev Biochem 80, 71-99.

Das, P. B., Dinh, P. X., Ansari, I. H., de Lima, M., Osorio, F. A. & Pattnaik, A. K., 2010. The minor envelope glycoproteins GP2a and GP4 of porcine reproductive and respiratory syndrome virus interact with the receptor CD163. J Virol 84, 1731–1740.

den Boon, J. A., Snijder, E. J., Chirnside, E. D., de Vries, A. A. F., Horzinek, M. C. & Spaan, W. J. M., 1991. Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. J Virol 65, 2910–2920.

de Vries, A.A., Chirnside, E.D., Horzinek, M.C. and Rottier, P.J., 1992. Structural proteins of equine arteritis virus. J Virol 66(11), 6294-303.

de Vries, A.A., Post, S.M., Raamsman, M.J., Horzinek, M.C. and Rottier, P.J., 1995a. The two major envelope proteins of equine arteritis virus associate into disulfide-linked heterodimers. J Virol 69(8), 4668-74.

de Lima, M., Ansari, I.H., Das, P.B., Ku, B.J., Martinez-Lobo, F.J., Pattnaik, A.K. and Osorio, F.A., 2009. GP3 is a structural component of the PRRSV type II (US) virion. Virology 390(1), 31-6.

Delrue, I., Delputte, P. L. & Nauwynck, H. J. (2009). Assessing the functionality of viral entry-associated domains of porcine reproductive and respiratory syndrome virus during inactivation procedures, a potential tool to optimize inactivated vaccines. *Vet Res* 40, 62.

Dobbe, J.C., van der Meer, Y., Spaan, W.J. and Snijder, E.J., 2001. Construction of chimeric arteriviruses reveals that the ectodomain of the major glycoprotein is not the main determinant of equine arteritis virus tropism in cell culture. Virology 288(2), 283-94.

Dokland, T. 2010. The structural biology of PRRSV. Virus Res 154(1-2), 86-97.

Ellgaard, L., Ruddock, L.W., 2005. The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep* 6, 28-32.

Eichler, R., Lenz, O., Strecker, T., Eickmann, M., Klenk, H.D. and Garten, W., 2003. Identification of Lassa virus glycoprotein signal peptide as a trans-acting maturation factor. EMBO Rep 4(11), 1084-8.

Entry of porcine reproductive and respiratory syndrome virus into porcine alve- olar macrophages via receptor-mediated endocytosis. J. Gen. Virol. 80 (Pt 2), 297–305.

Faaberg, K.S., Even, C., Palmer, G.A. and Plagemann, P.G., 1995. Disulfide bonds between two envelope proteins of lactate dehydrogenase-elevating virus are essential for viral infectivity. J Virol 69(1), 613-7.

Faaberg, K.S. and Plagemann, P.G., 1997. ORF 3 of lactate dehydrogenase-elevating virus encodes a soluble, nonstructural, highly glycosylated, and antigenic protein. Virology 227(1), 245-51.

Fang, Y. & Snijder, E. J., 2010. The PRRSV replicase: exploring the multifunctionality of an intriguing set of nonstructural proteins. Virus Res 154, 61–76.

Firth, A.E., Zevenhoven-Dobbe, J.C., Wills, N.M., Go, Y.Y., Balasuriya, U.B., Atkins, J.F., Snijder, E.J. and Posthuma, C.C., 2011. Discovery of a small arterivirus gene that overlaps the GP5 coding sequence and is important for virus production. J Gen Virol 92(Pt 5), 1097-106.

Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J. and Snijder, E.J., 2006. Nidovirales: evolving the largest RNA virus genome. Virus Res 117(1), 17-37.

Johnson, C.R., Griggs, T.F., Gnanandarajah, J. and Murtaugh, M.P., 2011. Novel structural protein in porcine reproductive and respiratory syndrome virus encoded by an alternative ORF5 present in all arteriviruses. J Gen Virol 92(Pt 5), 1107-16.

Kelleher, D.J. and Gilmore, R., 2006. An evolving view of the eukaryotic oligosaccharyltransferase. Glycobiology 16(4), 47R-62R.

K.H., Kang, S.M., Kim, M.C. and Choi, I.S., 2013. Immune responses in mice vaccinated with virus-like particles composed of the GP5 and M proteins of porcine reproductive and respiratory syndrome virus. Arch Virol 158(6), 1275-85.

Konishi, S., Akashi, H., Sentsui, H., Ogata, M., 1975. Studies on equine viral arteritis. I. Characterization of the virus and trial survey on antibody with Vero cell cultures. Nihon Juigaku Zasshi 37, 259–267.

Land, A., Zonneveld, D. and Braakman, I,. 2003. Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. FASEB J 17(9), 1058-67.

Lee, C., Yoo, D., 2006. The small envelope protein of porcine reproductive and respi- ratory syndrome virus possesses ion channel protein-like properties. Virology 355 (1), 30–43.

Lopez,O.J. and Osorio, F.A., 2004. Role of neutralizing antibodies in PRRSV protective Immunity. Vet Immunol Immunopathol 102(3), 155-63.

Lu, Z., Zhang, J., Huang, C.M., Go, Y.Y., Faaberg, K.S., Rowland, R.R., Timoney, P.J., Balasuriya, U.B., 2012. Chimeric viruses containing the N-terminal ectodomains of GP5 and M proteins of porcine repro- ductive and respiratory syndrome virus do not change the cellular tropism of equine arteritis virus. Virology 432, 99–109.

Pasternak, A.O., Spaan, W.J.M., Snijder, E.J., 2006. Nidovirus transcription: how to make sense...? *J Gen Virol* 87, 1403–1421.

Plagemann, P. G. W. & Moennig, V., 1992. Lactate dehydrogenase- elevating virus, equine arteritis virus, and simian hemorrhagic fever virus: a new group of positive-strand RNA viruses. Adv Virus Res 41, 99–192.

MacLachlan, N.J., Balasuriya, U., 2006. Equine Viral Arteritis. In: Perlman, S., Holmes, K. (Eds.), The Nidoviruses, Vol. 581. Springer US, pp. 429–433.

Maess, J., Reczko, E., Bohm, H.O., 1970. Equine arteritis virus: multi- plication in BHK 21-cells buoyant density and electron microscopical demonstration. Arch. Gesamte Virusforsch. 30, 47–58.

Matczuk, A.K., Kunec, D., Veit, M., 2013. Co-translational processing of glycoprotein 3 from equine arteritis virus: N-glycosylation adjacent to the signal peptide prevents cleavage. J. Biol. Chem. 288 (49), 35396–35405.

Meng, X.J., 2000. Heterogeneity of porcine reproductive and respiratory syndrome virus: implications for current vaccine efficacy and future vaccine development. Vet Microbiol 74(4), 309-29.

Murtaugh, M.P., Stadejek, T., Abrahante, J.E., Lam, T.T., Leung, F.C., 2010. The ever-expanding diversity of porcine reproductive and respiratory syndrome virus. Virus Res. 154, 18-30.

Oka, O.B. and Bulleid, N.J., 2013. Forming disulfides in the endoplasmic reticulum. Biochim Biophys Acta 1833(11), 2425-9.

Nauwynck, H.J., Duan, X., Favoreel, H.W., Van Oostveldt, P., Pensaert, M.B., 1999. Entry of porcine reproductive and respiratory syndrome virus into porcine alve- olar macrophages via receptor-mediated endocytosis. J. Gen. Virol. 80 (Pt 2), 297–305.

Nam, H.M., Chae, K.S., Song, Y.J., Lee, N.H., Lee, J.B., Park, S.Y., Song, C.S., Seo, K.H., Kang, S.M., Kim, M.C., Choi, I.S., 2013. Immune responses in mice vaccinated with virus-like particles composed of the GP5 and M proteins of porcine reproductive and respiratory syndrome virus. Arch. Virol 158 (6), 1275–1285.

Nitschke, M., Korte, T., Tielesch, C., Ter-Avetisyan, G., Tunnemann, G., Cardoso, M.C., Veit, M., Herrmann, A., 2008. Equine arteritis virus is delivered to an acidic com- partment of host cells via clathrin-dependent endocytosis. Virology 377 (2), 248–254.

Nyathi, Y., Wilkinson, B.M., Pool, M.R., 2013. Co-translational targeting and translocation of proteins to the endoplasmic reticulum. *Biochim Biophys Acta* 1833, 2392–2402.

Rehm A, Stern P, Ploegh HL, Tortorella D, 2002, Signal peptide cleavage of a type I membrane protein, HCMV US11, is dependent on its membrane anchor. EMBO J. 2;20(7):1573-82.

Sawicki, S.G., Sawicki, D.L., Siddell, S.G., 2007. A Contemporary View of Coronavirus Transcription. *J Virol* 81, 20–29.

Shelness, G.S., Lin, L., Nicchitta, C.V., 1993. Membrane topology and biogenesis of eukaryotic signal peptidase. *J Biol Chem* 268, 5201–5208.

Spilman, M. S., Welbon, C., Nelson, E. & Dokland, T., 2009. Cryo- electron tomography of porcine reproductive and respiratory syndrome virus: organization of the nucleocapsid. J Gen Virol 90, 527–535.

Snijder, E. J. & Meulenberg, J. J. M., 1998. The molecular biology of arteriviruses. J Gen Virol 79, 961–979.

Snijder, E.J., Dobbe, J.C. and Spaan, W.J., 2003. Heterodimerization of the two major envelope proteins is essential for arterivirus infectivity. J Virol 77(1), 97-104.

Snijder, E.J., van Tol, H., Roos, N. and Pedersen, K.W., 2001. Non-structural proteins 2 and 3 interact to modify host cell membranes during the formation of the arterivirus replication complex. J Gen Virol 82(Pt 5), 985-94.

Snijder, E. J. & Kikkert, M., 2013. Arteriviruses. In Fields Virology, 6th edn, pp. 859–879. Edited by D. M. Knipe & P. M. Howley. Philadelphia, PA: Lippincott Williams & Wilkins.

Snijder EJ., Kikkert M., Fang Y., 2014. Arterivirus molecular biology and pathogenesis. J Gen Virol. 94(Pt 10):2141-63.

Stadejek, T., Stankevicius, A., Murtaugh, M.P. and Oleksiewicz, M.B. 2013. Molecular evolution of PRRSV in Europe: current state of play. Vet Microbiol 165(1-2), 21-8.

Sun, L., Li, Y., Liu, R., Wang, X., Gao, F., Lin, T., Huang, T., Yao, H., Tong, G., Fan, H., Wei, Z. and Yuan, S., 2013. Porcine reproductive and respiratory syndrome virus ORF5a protein is essential for virus viability. Virus Res 171(1), 178-85.

Thaa, B., Sinhadri, B.C., Tielesch, C., Krause, E. and Veit, M. 2013. Signal peptide cleavage from GP5 of PRRSV: a minor fraction of molecules retains the decoy epitope, a presumed molecular cause for viral persistence. PLoS One 8(6), e65548.

Thaa, B., Kabatek, A., Zevenhoven-Dobbe, J.C., Snijder, E.J., Herrmann, A., Veit, M., 2009. Myristoylation of the arterivirus E protein: the fatty acid modification is not essential for membrane association but contributes significantly to virus infectivity. J. Gen. Virol. 90 (Pt 11), 2704–2712.

Tian, D., Wei, Z., Zevenhoven-Dobbe, J.C., Liu, R., Tong, G., Snijder, E.J. and Yuan, S. 2012. Arterivirus minor envelope proteins are a major determinant of viral tropism in cell culture. J Virol 86(7), 3701-12.

Van Breedam, W., Delputte, P.L., Van Gorp, H., Misinzo, G., Vanderheijden, N., Duan, X. and Nauwynck, H.J., 2010. Porcine reproductive and respiratory syndrome virus entry into the porcine macrophage. J Gen Virol 91(Pt 7), 1659-67.

Veit M., Matczuk AK., Sinhadri BC., Krause E., Thaa B.. 2014. Membrane proteins of arterivirus particles: Structure, topology, processing and function. Virus Res 19;194:16-36.

Wieringa, R., de Vries, A. A. F., Post, S. M. & Rottier, P. J. M., 2003a. Intra- and intermolecular disulfide bonds of the GP2b glycoprotein of equine arteritis virus: relevance for virus assembly and infectivity. J Virol 77, 12996–13004.

Wieringa, R., de Vries, A.A. and Rottier, P.J., 2003b. Formation of disulfide-linked complexes between the three minor envelope glycoproteins (GP2b, GP3, and GP4) of equine arteritis virus. J Virol 77(11), 6216-26.

Wieringa, R., de Vries, A.A., Raamsman, M.J. and Rottier, P.J., 2002. Characterization of two new structural glycoproteins, GP(3) and GP(4), of equine arteritis virus. J Virol 76(21), 10829-40.

Wieringa, R., de Vries, A.A., van der Meulen, J., Godeke, G.J., Onderwater, J.J., van Tol, H., Koerten, H.K., Mommaas, A.M., Snijder, E.J., Rottier, P.J., 2004. Structural protein requirements in equine arteritis virus assembly. *J Virol* 78, 13019-13027.

Wissink, E. H. J., Kroese, M. V., van Wijk, H. A. R., Rijsewijk, F. A. M., Meulenberg, J. J. M. & Rottier, P. J. M., 2005. Envelope protein requirements for the assembly of infectious virions of porcine reproductive and respiratory syndrome virus. J Virol 79, 12495– 12506.

Yan, A. and Lennarz, W.J., 2005. Unraveling the mechanism of protein N-glycosylation. J Biol Chem 280(5), 3121-4.

Zielinska, D.F., Gnad, F., Wiśniewski, J.R., Mann, M., 2010. Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* 141, 897-907.

Acknowledgements

I would like to thank my supervisor Dr. Michael Veit for providing the opportunity to work in his lab for my Ph.D. He gave me good support during my Ph.D. and helped me with the new ideas and approaches for solving the scientific problems. I am thankful to Professor Udo Heinemann for being my Ph.D. supervisor and reviewing my thesis.

My sincere thanks to Dr. Eberhard Krause for analysing the protein samples with mass spectrometry. I would like to thank all my colleagues at the Institute for Virology, Freie University Berlin for their support. My special thanks to all the Ph.D. students and post-docs, Katharina, Maren, Ania, Steffi, Mingyang, Bastian and Ludwig for their help during my work in the laboratory.

Finally, I would like to thank my family for their support throughout my life and education.