

Electronic Thesis and Dissertation Repository

---

8-17-2021 2:00 PM

## Production of self-assembling protein nanocages and virus-like particles displaying porcine reproductive and respiratory syndrome virus epitopes in *Nicotiana benthamiana*

Jordan T. VanderBurgt, *The University of Western Ontario*

Supervisor: Menassa, Rima, *The University of Western Ontario, Agriculture and Agri-Food Canada*

Co-Supervisor: Kohalmi, Susanne E., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biology

© Jordan T. VanderBurgt 2021

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Biotechnology Commons](#)

---

### Recommended Citation

VanderBurgt, Jordan T., "Production of self-assembling protein nanocages and virus-like particles displaying porcine reproductive and respiratory syndrome virus epitopes in *Nicotiana benthamiana*" (2021). *Electronic Thesis and Dissertation Repository*. 7979.  
<https://ir.lib.uwo.ca/etd/7979>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact [wlsadmin@uwo.ca](mailto:wlsadmin@uwo.ca).

## **ABSTRACT**

Livestock diseases are major hurdles facing the agriculture industry. One major illness affecting the swine industry is the viral porcine reproductive and respiratory syndrome (PRRS). Current methods of controlling its spread are inadequate, so there is urgent need for better prevention tools. Vaccines are the most effective way to control viral diseases because viruses cannot be controlled by antibiotics. Protein nanoparticle-based subunit vaccines provide safe, effective disease prevention. In this study, an epitope derived from portions of the two most abundant PRRS viral surface proteins, M and GP5, was fused to three self-assembling proteins for use as vaccine candidates. Their expression was examined without the epitope using transient expression in *Nicotiana benthamiana*, and all three proteins were detected on Western blots. Self-assembling M-GP5 fusion proteins were also expressed at similar or higher levels than the self-assembling proteins alone. These findings provide groundwork for future research investigating these PRRS vaccine candidates.

## **Keywords**

porcine reproductive and respiratory syndrome, PRRS, subunit vaccine, protein nanocages, virus-like particles, *Nicotiana benthamiana*, transient expression

## **SUMMARY FOR LAY AUDIENCE**

One of the major diseases affecting the swine industry around the globe is the porcine reproductive and respiratory syndrome (PRRS). This disease leads to issues with carrying fetal piglets to term in pregnant sows, as well as respiratory and developmental problems in growing piglets. This disease is caused by a virus and cannot be treated by antibiotics. The main method to control the spread of viral diseases is through the use of vaccines, however, those currently available against PRRS are inadequate. Because of this, there is an urgent need for a safer and more effective vaccine against this disease. Therefore, the goal of this research was to produce and examine several vaccine candidates against PRRS. In this project portions of two viral surface proteins, the M-GP5 epitope, were fused to three carrier proteins. Because this was the first time these carrier proteins were expressed in plants, they were first produced on their own to investigate their soluble protein accumulation levels in this expression platform. They were then each fused to the M-GP5 epitope and their accumulation levels were examined again. All of these proteins had moderate to high soluble accumulation levels, demonstrating their successful expression in plants. This study provides the groundwork for further development of these vaccine candidates so that one day PRRS will no longer be one of the major concerns affecting the global swine industry.

## ACKNOWLEDGEMENTS

I wish to begin by expressing my deepest gratitude for my supervisor, Dr. Rima Menassa. My experience as part of her team has been nothing short of exceptional, even through the unique and difficult times we faced. Her continuous planning, guidance, and support provided the foundation for this research project and all of my successes. Words cannot fully convey my thanks for her tireless effort and patience.

I am also indebted to my co-supervisor, Dr. Susanne Kohalmi, whose thoughtful feedback has helped me grow both as a researcher and writer. I am especially appreciative of her willingness to welcome me into her lab for months during the pandemic when I had nowhere else to work.

I would like to thank my advisors, Dr. Chris Garnham and Dr. Kathleen Hill, for their interest and support. I always looked forward to their insight and feedback throughout my project, bringing new ideas and avenues to explore.

This work would not have been possible without the immense guidance I received from my lab technicians Angelo Kaldis and Hong Zhu. Their expertise and advice made invaluable contributions to this research project.

I wish to acknowledge my lab members and friends who have been at my side through this amazing experience. Their support means the world, and has helped get me where I am today.

Finally, I would like to pay special regards to my family; Marg, Rick, Austin, and Liam. None of this would have been possible without their love and constant encouragement to follow through with my passions.

# TABLE OF CONTENTS

ABSTRACT.....	ii
SUMMARY FOR LAY AUDIENCE .....	iii
ACKNOWLEDGEMENTS.....	iv
TABLE OF CONTENTS.....	v
LIST OF FIGURES .....	viii
LIST OF APPENDICES.....	ix
LIST OF ABBREVIATIONS.....	x
1 INTRODUCTION.....	1
1.1 Porcine reproductive and respiratory syndrome virus (PRRSV) .....	1
1.1.1 Disease and impacts.....	1
1.1.2 PRRSV infection.....	1
1.1.3 Viral genome.....	3
1.1.4 Pathogenesis of PRRSV .....	4
1.1.5 PRRSV epitopes for vaccine development.....	6
1.2 Traditional vaccine design for veterinary medicine.....	8
1.2.1 Killed virus vaccines.....	8
1.2.2 Live attenuated virus vaccines .....	9
1.2.3 Subunit vaccines .....	10
1.2.4 PRRSV vaccine attempts .....	11
1.3 New vaccine concepts .....	12
1.3.1 Virus-like particles.....	13
1.3.2 Protein nanocages .....	14

1.4	Nanoparticle platforms of interest.....	15
1.4.1	<i>Aquifex aeolicus</i> lumazine synthase.....	15
1.4.2	<i>Helicobacter pylori</i> ferritin.....	17
1.4.3	Tobacco mosaic virus coat protein .....	18
1.5	Expression platforms.....	19
1.5.1	Bacterial expression .....	19
1.5.2	Yeast expression .....	20
1.5.3	Insect cell expression .....	20
1.5.4	Mammalian cell expression .....	21
1.5.5	Plant expression .....	22
1.6	Rationale and goals .....	24
1.7	Objectives.....	25
2	MATERIALS AND METHODS .....	26
2.1	Plant expression vectors.....	26
2.1.1	Original vectors.....	26
2.1.2	Modification of vectors.....	29
2.2	Construct sequences .....	30
2.2.1	Self-assembling protein sequences .....	30
2.2.2	PRRSV epitope sequences.....	30
2.3	Cloning and transformations .....	30
2.4	Infiltrations and sampling.....	31
2.4.1	Plant growth and care.....	31
2.4.2	Infiltration and sampling of <i>Nicotiana benthamiana</i> .....	31

2.5	Recombinant protein quantification .....	35
2.6	Statistical modelling .....	36
3	RESULTS .....	37
3.1	Modification of plant protein expression vectors .....	37
3.2	Self-assembling protein construct design .....	44
3.3	Expression analysis of self-assembling proteins .....	45
3.4	Epitope fusion construct design .....	50
3.5	Expression analysis of fusion proteins .....	53
3.6	Quantification and statistical analysis of accumulation .....	57
4	DISCUSSION .....	60
4.1	Protein nanoparticles .....	60
4.2	Livestock vaccines .....	63
4.3	Protein mobility and assembly in gels .....	66
4.4	N-linked glycosylation .....	67
5	CONCLUSIONS AND FUTURE STUDIES .....	70
6	REFERENCES .....	72
7	APPENDICES .....	84
8	CURRICULUM VITAE .....	89

## LIST OF FIGURES

Figure 1. Porcine reproductive and respiratory syndrome virus structural proteins.....	2
Figure 2. M and GP5 co-localize on the PRRSV envelope.....	5
Figure 3. Assembly of ALS, Fer, and TMVc nanoparticles.....	16
Figure 4. Original pCLGG plant expression vectors.....	28
Figure 5. Agroinfiltration of <i>N. benthamiana</i> leaves for transient protein expression....	33
Figure 6. Sampling of infiltrated <i>N. benthamiana</i> leaf tissue.....	34
Figure 7. Modification pCLGG plant expression vectors to remove the <i>Xpress</i> tag.....	38
Figure 8. Schematic for primer binding to pCLGG and pCLGG-X plant expression vectors.....	40
Figure 9. Modification of pCLGG-ER-ELP vector .....	42
Figure 10. Modified pCLGG-X plant expression vectors.....	43
Figure 11. Schematic for self-assembling proteins.....	46
Figure 12. ALS, Fer, and TMVc accumulate in all three subcellular compartments.....	49
Figure 13. Self-assembling proteins fused to the M-GP5 epitope.....	51
Figure 14. Self-assembling proteins and their M-GP5 fusions accumulate well in plants.....	55
Figure 15. Recombinant soluble protein accumulation of ALSm, Ferm, TMVc, and their M-GP5 fusion constructs.....	59
Figure 16. N-linked glycan chains from plants and mammals.....	68



## **LIST OF APPENDICES**

Appendix 1. Nucleotide sequences of codon-optimized genes for plant expression.....	84
Appendix 2. Amino acid sequences of all proteins produced recombinantly.....	86
Appendix 3. Nucleotide sequences for primers.....	88
Appendix 4. Results of Games-Howell post-hoc test from R.....	88

## LIST OF ABBREVIATIONS

\* Standard SI units not listed

ALS	<i>Aquifex aeolicus</i> lumazine synthase
ALSm	modified <i>Aquifex aeolicus</i> lumazine synthase
APO	apoplast
BLS	<i>Brucella</i> lumazine synthase
CYT	cytoplasm
CHL	chloroplast
CTPP	C-terminal propeptide
dpi	days post-infiltration
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ELP	elastin-like polypeptide
ER	endoplasmic reticulum
Fer	<i>Helicobacter pylori</i> ferritin
Ferm	modified <i>Helicobacter pylori</i> ferritin
Fuc	fucose
FW	fresh weight
Gal	galactose
Glc	glucose
GlcNAc	N-acetylglucosamine
GLS	generalized least squares
GP	glycosylated protein
HA	hemagglutinin
HBc	hepatitis B virus core antigens
HPV	Human Papillomaviruses
kan	kanamycin
KV	killed viral vaccine
LAV	live attenuated virus
LB	Luria-Bertani
Man	mannose

MES	2-(N-morpholino)ethanesulfonic acid
ncFill	non-coding filler sequence
NeuAc	sialic acid
nsp	non-structural protein
ORF	open reading frame
PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
PRRS	porcine reproductive and respiratory syndrome
PRRSV	porcine reproductive and respiratory syndrome virus
PVDF	polyvinylidene difluoride
PVPP	polyvinylpyrrolidone
RE	restriction enzyme
rif	rifampicin
RTC	replication/transcription complex
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sg mRNA	subgenomic mRNA
TBS	Tris-buffered saline
tCUP	tobacco cryptic upstream promoter translational enhancer
TMV	tobacco mosaic virus
TMVc	tobacco mosaic virus coat protein
VAC	vacuole
VLP	virus-like particle
Xyl	xylose

# 1 INTRODUCTION

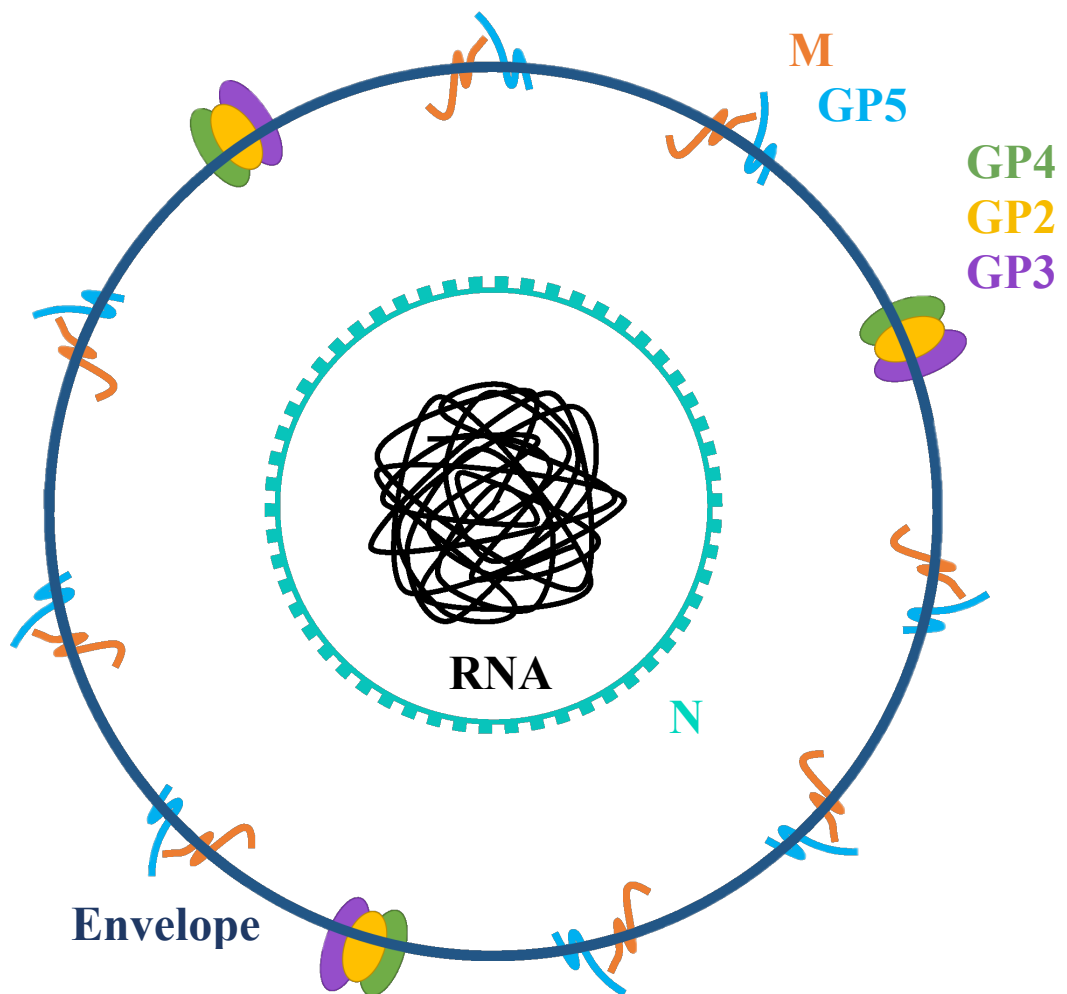
## 1.1 Porcine reproductive and respiratory syndrome virus (PRRSV)

### 1.1.1 Disease and impacts

Diseases in livestock production affect animal welfare and bear a burden on the economy. The porcine reproductive and respiratory syndrome virus (PRRSV) is a small enveloped RNA virus from the family *Arteriviridae* within the order *Nidovirales* (Figure 1). This virus infects pigs, leading to the development of porcine reproductive and respiratory syndrome (PRRS). This disease presents as a reproductive disease in sows causing spontaneous abortions and stillbirths. It also causes a respiratory disease in growing pigs leading to sneezing, pneumonia, stunting of growth, and lowered life expectancy (Charentantanakul, 2012). PRRS is prevalent worldwide and has major economic impacts costing over \$650 million to the US swine industry alone in 2013, which was an increase of approximately \$100 million since 2005 (Holtkamp *et al.*, 2013). PRRS has also been described as one of the top five most researched diseases affecting swine production around the world along with *Salmonella* spp., *Escherichia coli*, and influenza (VanderWaal and Deen, 2018). While continued research has improved the knowledge surrounding PRRS, this disease continues to be a pressing issue with insufficient tools to control its spread.

### 1.1.2 PRRSV infection

PRRSV infection generally occurs at mucosal tissues such as those within the respiratory and reproductive tracts, however, transmission can also occur directly into the bloodstream (Pileri and Mateu, 2016). PRRSV infection can be broken down into three



**Figure 1. Porcine reproductive and respiratory syndrome virus structural proteins.**

The most abundant proteins in the PRRSV envelope (dark blue outer circle) are the M and GP5 proteins (orange and light blue, respectively) which co-localize. The nucleocapsid (N: cyan) monomers form dimers once phosphorylated to interact with and encapsulate the viral RNA genome (black). The minor structural proteins GP2, GP3, and GP4 (yellow, purple, and green, respectively) form heterotrimers in the PRRSV envelope. Image modified from Harper (2019).

stages: initial infection, persistent infection, and eventually viral extinction within the host. The first stage has viral replication predominantly localized to the lung tissues mainly within alveolar macrophages and, to a lesser extent, dendritic cells (Duan *et al.*, 1997; Loving *et al.*, 2007). The presence of virions within the bloodstream, or viremia, occurs between 6-12 h post-infection and lasts for several weeks (Lunney *et al.*, 2016). During the second stage of PRRSV infection, replication decreases and may be undetectable within the bloodstream. Lower levels of replication continue to take place within the host lymphatic system and individuals continue to be infectious during much of this period (Rowland *et al.*, 2003). Eventually replication continues to diminish until there are no more virions present within the host, marking the final stage of PRRSV infection. Viral extinction may not occur until after 250 days post-infection in some cases (Wills *et al.*, 2003), which is the approximate lifespan of pigs within the swine industry making this a lifelong infection.

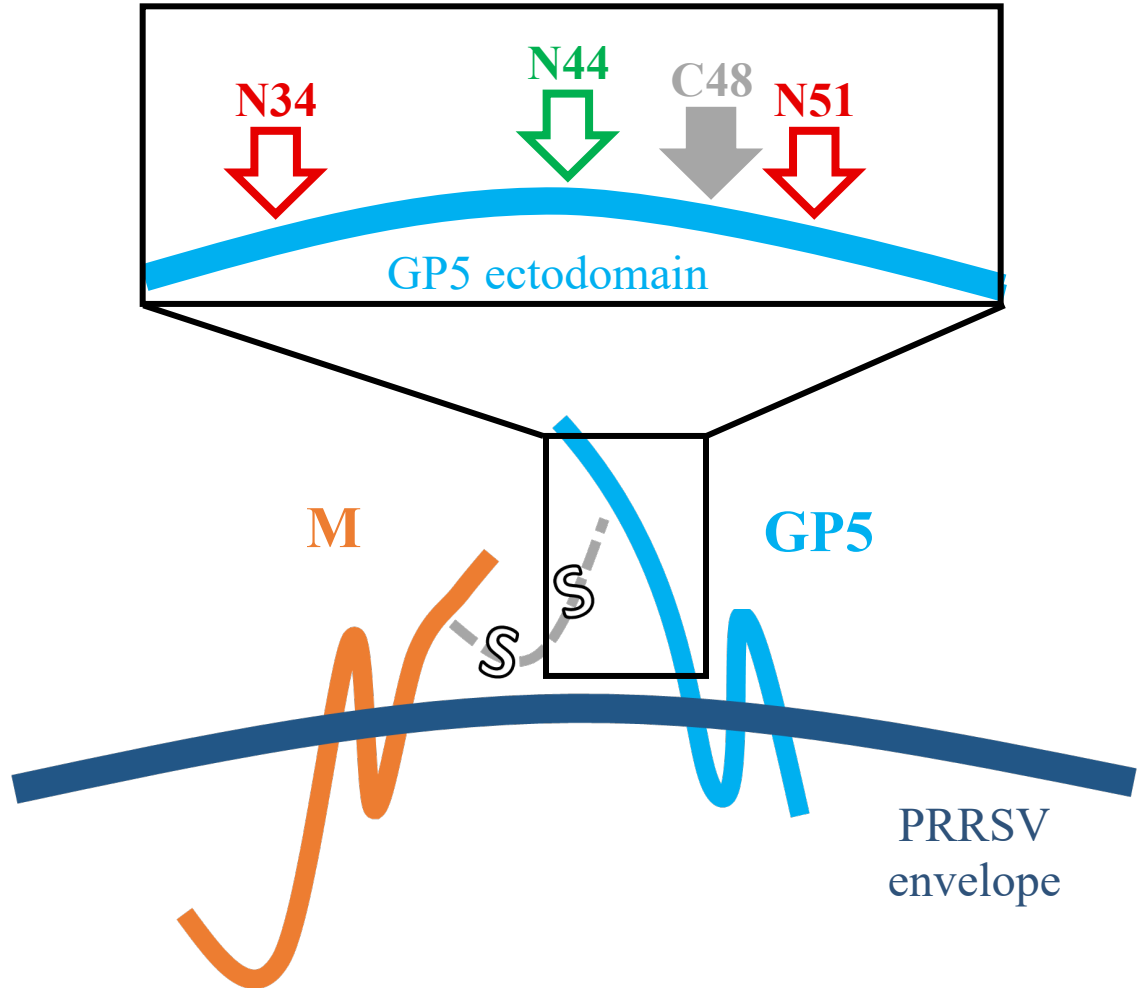
### 1.1.3 Viral genome

The PRRSV is a small icosahedral virus with a positive-stranded RNA genome approximately 15 kb in length. The genome's 5' end encodes open reading frames (ORFs) 1a and 1b which encode polyproteins 1a and 1ab (Ziebuhr *et al.*, 2000). These polyproteins undergo extensive autoproteolytic processing to produce fourteen non-structural proteins (nsps) responsible for the autoproteolysis, regulation of subgenomic mRNA (sg mRNA) synthesis, RNA-dependent RNA polymerization, helicase activity, and more (Lunney *et al.*, 2016; Ziebuhr *et al.*, 2000). A subset of these nsps are also responsible for the formation of replication/transcription complexes (RTC) (Lunney *et al.*, 2016).

The 3' end of the PRRSV genome contains eight ORFs encoding structural proteins of the virus which can be broken into two groups: the glycoproteins (GPs) and the unglycosylated proteins. The two major structural proteins, GP5 and the matrix protein (M), are the most abundant proteins on the viral envelope (Figure 1) and interact via a disulfide bond on their N-terminal ectodomains (Figure 2) (Mardassi *et al.*, 1996; Zhang *et al.*, 2021). This interaction between M and GP5 is crucial for viral infectivity, with the heterodimer playing a major role in virus assembly and budding within host cells (Snijder *et al.*, 2013; Wissink *et al.*, 2005; Zhang *et al.*, 2021). The minor structural proteins GP2, GP3, and GP4 form heterotrimers on the viral envelope (Figure 1) and are involved in viral attachment and entrance into host cells (Wissink *et al.*, 2005). The other minor structural proteins, ORF5a and E, have been less studied, however, they were found to be essential for viral viability and infectivity, respectively (Lunney *et al.*, 2016). Another PRRSV major structural protein is the nucleocapsid protein (N), which is the only structural protein not integrated into the viral envelope. This protein is unglycosylated, however, it undergoes phosphorylation then forms homodimers to interact with and encapsulate the viral genome (Mardassi *et al.*, 1996; Snijder *et al.*, 2013).

#### 1.1.4 Pathogenesis of PRRSV

Upon PRRSV infection, proteins in the viral envelope interact with receptors in host cell membranes. Research has pointed specifically towards two interactions: GP2 and GP4 with CD163 (Das *et al.*, 2010) as well as M and GP5 with sialoadhesin (Van Breedam *et al.*, 2010), however, the latter interaction has been found nonessential for cell entry (Prather *et al.*, 2013). Upon binding, the virus is internalized via receptor-mediated endocytosis. Following entry and uncoating, the positive-sense viral



**Figure 2. M and GP5 co-localize on the PRRSV envelope.** The ectodomains of M (orange) and GP5 (light blue) interact via a disulfide bond (grey) between the C9 and C48 residues of M and GP5, respectively. GP5 has three potential glycosylation sites on its ectodomain, all of which are glycosylated on the native PRRS virus. The glycosylation of one of these sites, N44 (green), was shown to be necessary for viral function and vital for eliciting antibodies capable of neutralizing wild-type virions. The other two sites, N34 and N51 (red), were shown to shield the neutralizing antibody binding site on GP5's ectodomain and their removal improves the immunogenicity of this epitope. The co-expression of M with GP5 was also shown to increase the resulting immune response. Image modified from Harper (2019).



RNA genome is released into the cytoplasm where the replication cycle is initiated by host translational machinery. Some nsps associate with the endoplasmic reticulum (ER) membrane forming double membraned vesicles, the RTC, which act as the site of viral reproduction within host cells (Lunney *et al.*, 2016).

Within the RTC, a full length viral antigenome (negative-strand) is synthesized by the RNA-dependent RNA polymerase which subsequently serves as a template for production of new genomic RNA and sg mRNAs (Pasternak *et al.*, 2006). The sg mRNAs code for the eight structural proteins encoded within the PRRSV genome. The structural proteins other than N are co-translationally embedded into the ER membrane where the glycoproteins acquire N-linked glycan chains, and disulfide bridges form between the M and GP5 ectodomains (Mardassi *et al.*, 1996; Zhang *et al.*, 2021). The replicated viral genome, encapsulated in N proteins, interacts with the ER membrane where the other structural proteins are co-localized. This allows for budding of the membrane to occur, forming immature virions within the ER lumen decorated with PRRSV envelope proteins (Wissink *et al.*, 2005). As the immature virions proceed through the ER and Golgi apparatus, the glycoproteins acquire more complex N-linked glycan chains (Li *et al.*, 2015). Eventually, the mature PRRS virions are released into the extracellular space via exocytosis.

### 1.1.5 PRRSV epitopes for vaccine development

The most effective way to control the spread of diseases like PRRS is with the use of vaccines, especially because it is viral in origin and therefore unaffected by antibiotics for treatment and prevention. One goal of vaccines is to stimulate the production of antibodies within the host system capable of neutralizing the native virus. Only antibodies

corresponding to specific parts of the virus will be capable of this neutralizing effect, and therefore these particular epitopes must be identified and used for vaccine production. The PRRSV has been previously shown to have neutralizing antibody binding sites on several of its structural proteins including GP2, GP3, GP4, and GP5 (Popescu *et al.*, 2017; Stoian and Rowland, 2019). Among these, however, anti-GP5 antibodies appear most efficient at neutralizing PRRSV to stop infection (Gonin *et al.*, 1999; Popescu *et al.*, 2017; Stoian and Rowland, 2019), making GP5 a promising target for vaccine development.

GP5 is one of the most abundant proteins in the PRRSV envelope along with the M protein (Figure 1). GP5 has an N-terminal ectodomain, three transmembrane domains, and a C-terminal endodomain (Figure 2). The first 2-3 dozen amino acids of GP5, depending on the strain, correspond to a signal peptide that is removed from the mature protein (Popescu *et al.*, 2017). The remaining ectodomain contains at least one PRRSV neutralizing epitope (Kim *et al.*, 2013; Popescu *et al.*, 2017). This region also contains varying numbers of potential N-linked glycosylation sites, generally between 2 and 4, depending on the strain. On mature PRRSV, these glycosylation sites display the complex glycans conferred by passage through the late Golgi when exiting the host cells (Li *et al.*, 2015). At least one of these glycosylation sites appear to shield or mask the epitope(s) around it (Ansari *et al.*, 2006), however, one of the glycosylation sites (N44) has been shown necessary for eliciting antibodies capable of neutralizing wild-type virions (Figure 2) (Wei *et al.*, 2012). Another important site on GP5's ectodomain is a cysteine residue (C48) that forms a disulfide bond with a cysteine present on the N-terminal ectodomain of M (Mardassi *et al.*, 1996; Zhang *et al.*, 2021). In fact, while M itself contains no neutralizing antibody binding sites, its co-expression alongside GP5 was shown to

increase the immunogenicity of GP5 (Jiang *et al.*, 2006). Based on this information, the co-expression of GP5 and M, or at least their ectodomains, is an important pairing for the production of vaccines against the PRRSV.

## **1.2 Traditional vaccine design for veterinary medicine**

When bacteria and viruses infect an individual, the body produces an immune response in attempt to eradicate the invader. After the infection is eliminated, the body retains memory of the foreign invader as a part of its adaptive immune system. This allows for subsequent encounters with the same pathogen to be recognised more quickly, either preventing disease development or reducing its severity (Bartlett *et al.*, 2009; Karch and Burkhard, 2016). Vaccination is the process of exposing the immune system to inactive, modified, or antigenic parts of disease-causing invaders to induce this acquired immunity without actual infection. This process has saved the lives of countless humans and animals against detrimental diseases, especially ones caused by viruses which are unaided by antibiotic treatments. While there have been immense advances in the field of vaccinology, including novel design approaches, more traditional vaccines are still commonly used in veterinary medicine today.

### **1.2.1 Killed virus vaccines**

One method of vaccine production is the use of killed viruses (KV). This involves inactivating viruses using chemicals (such as formaldehyde or hydrogen peroxide), heat, ultraviolet light, or other methods (Bartlett *et al.*, 2009). The resulting inactivated, non-infectious viral particles can be used for vaccination.

While KV are praised for their safety, a major concern is that they do not confer as much protection as they should. The inactivation processes can damage viral proteins and they may not mimic native infection very well, leading to a reduced immune response and therefore less robust protective immunity (Vetter *et al.*, 2018). Another concern, although uncommon, is the risk of incomplete inactivation of the viruses. This can leave some live viral particles within the vaccine resulting in potential infection of the “vaccinated” individual. This has occurred several times in the past century, including a major incident in 1955 (Offit, 2005). After hundreds of thousands of children were vaccinated with incompletely inactivated poliovirus batches about 40,000 got sick, over 50 were paralyzed, and 5 lost their lives.

### 1.2.2 Live attenuated virus vaccines

A second method of vaccine production, using live attenuated viruses (LAV), relies on stimulating the immune system with modified viruses that do not cause major disease symptoms for the host (Bartlett *et al.*, 2009). LAV induce a host immune response resulting in protective immunity against any subsequent encounters from the original, native virus. LAV are modified from their native form through selective passage in non-host-species cell lines leading to improved infection in these cell cultures and, more importantly, decreased replication within the original host species (Bartlett *et al.*, 2009; Vetter *et al.*, 2018). The reduced ability to replicate in their original host’s cells is the result of modifications in the viral genome, such as mutations, and subsequent alteration of the encoded proteins. LAV have been shown to provide strong, lasting immunity because they stimulate comprehensive immune responses in a similar manner to the native virus (Francis, 2018).

The major drawback with the use of LAV revolves around their safety. While they cannot replicate very well in healthy hosts, there is a potential for greater reproduction and damage to immunocompromised individuals and fetuses (Pirofski and Casadevall, 1998; Vetter *et al.*, 2018). Another concern is the possibility of mutation of the LAV within the vaccinated individual causing a reversion to virulence and spreading the disease instead of protecting from it (Vetter *et al.*, 2018). This reversion to virulence may happen due to spontaneous mutation, however, this can also occur through recombination with wild strains (Wang *et al.*, 2019).

### 1.2.3 Subunit vaccines

Subunit vaccines are a third method of vaccination. These involve presenting whole or fragments of viral proteins to the immune system to stimulate a response and induce immunity (Bartlett *et al.*, 2009). Originally these proteins were derived from breaking down pathogens themselves (Francis, 2018), but advances in genetic engineering have allowed for the recombinant production of these proteins. However, many proteins, especially those that reside within membranes, can prove difficult to produce in their entirety, while smaller fragments of proteins can be troublesome as they are often degraded quickly. Moreover, purified subunit vaccines often provide less immunity compared to other vaccine types due to weaker interactions with the host immune system and rapid turnover once administered. Therefore, adjuvants and booster immunizations are often required (Francis, 2018; López-Sagaseta *et al.*, 2016).

#### 1.2.4 PRRSV vaccine attempts

There have been many attempts to create vaccines for PRRS since its discovery. These attempts have had varying degrees of success and some, including both KV and LAV vaccines, have become commercially available (Zuckermann *et al.*, 2007). While these vaccines have helped to prevent the spread of PRRS around the globe, they are not an adequate solution to the problem. PRRSV KV vaccines have been shown unable to provide protective immunity in immunized pigs (Zuckermann *et al.*, 2007) and are no longer used in some places.

LAV vaccines for PRRS are much more effective than the KV vaccine at immunizing pigs, however, this appears to apply only to homologous strains with minimal protection against heterologous challenges (Okuda *et al.*, 2008). Another concern with PRRSV LAV vaccines is with their safety because they are known to revert to virulence in rare cases, as mentioned previously. PRRSV LAV vaccines have been shown to do this at least several independent times in different parts of the world, and the actual rate of occurrence is likely higher than what is documented (Kimman *et al.*, 2009). One record was on Danish farms in 1996, where they were having no issues with PRRS until a few months after vaccination with a PRRSV LAV vaccine (Nielsen *et al.*, 2001). It was then found that the vaccine had accumulated mutations facilitating its reversion to virulence, spreading the disease not only to the vaccinated individuals but also unvaccinated pigs. More recently, Wang *et al.* (2019) found that recombination was occurring between the LAV used and a field strain, thereby contributing to the spread of PRRS among pigs on the studied farm. Eclercy *et al.* (2019) also demonstrated that recombination between two PRRSV LAV led to an outbreak at a French pig farm.

Because the currently available PRRSV vaccines do not provide sufficient protection against the spread of this devastating disease and it continues to be a prevalent issue in the global swine industry, a safer and more effective vaccine is needed.

### **1.3 New vaccine concepts**

The use of nucleotide-based vaccines has gained momentum and popularity recently, especially since the approval and use of two RNA-based vaccines to control the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) during the global pandemic (Meo *et al.*, 2021). Nucleotide-based vaccine strategies involve the administration of encapsulated DNA or RNA into individuals. The nucleotide sequences enter cells and, using the host transcription and translation machinery, produce protein(s) from the pathogen of interest (Brisse *et al.*, 2020; Noor, 2021). These proteins are identified as foreign by the host immune system, eventually resulting in protective immunity in a similar manner as traditional vaccine approaches. DNA-based vaccines often have reduced effectiveness due to insufficient antigen expression within the vaccinated individual (Brisse *et al.*, 2020). In contrast, RNA-based vaccines appear effective as demonstrated by their current use against SARS-CoV-2 (Meo *et al.*, 2021; Noor, 2021). Nucleotide-based vaccines are more expensive than some other vaccines, and keeping costs low is an important economic factor for livestock vaccines. Therefore, DNA- and RNA-based vaccines are unlikely to be adopted for use in veterinary medicine.

Another new approach to vaccine development has moved towards an enhancement of subunit vaccine design. Instead of producing full or partial antigenic proteins on their own, they can be fused to larger molecules with the hope of making them more stable, immunogenic, and easier to produce. As mentioned, a major drawback of traditional

subunit vaccines is their inefficiency at producing protective immunity, generally caused by a weak interaction with the host immune system (Francis, 2018; López-Sagaseta *et al.*, 2016). When antigenic proteins or epitopes are attached to larger molecules, such as protein nanoparticles, they have been shown to have a stronger interaction with the immune system and to elicit better protective immunity than the protein subunits alone (Bachmann and Jennings, 2010; López-Sagaseta *et al.*, 2016). Two inexpensive and increasingly promising platforms for antigen presentation include the use of virus like particles (VLPs) and protein nanocages.

### 1.3.1 Virus-like particles

There are two main types of VLPs with their main difference being the presence or absence of a lipid envelope. Enveloped VLPs involve the expression of one or more outer protein(s) of an enveloped virus (Latham and Galarza, 2001; Ward *et al.*, 2021). These result in budding of intracellular membranes, like the plasma membrane or ER, to produce vesicles decorated with the viral proteins. The other type of VLP is exclusively composed of viral proteins lacking lipid integration, and these proteins self-assemble into stable structures (Mohsen *et al.*, 2017). Both types of VLPs are safe as they contain no genetic material and are therefore incapable of infection or replication (López-Sagaseta *et al.*, 2016; Mohsen *et al.*, 2017). Examples of VLP vaccines currently commercialized are the Human Papillomaviruses (HPV) vaccines Gardasil<sup>®</sup> and Cervarix<sup>®</sup> (Wang and Roden, 2013). Recombinant expression of HPV's major capsid protein, L1, produces self-assembling VLPs used for effective vaccination against this virus.

Many non-enveloped VLPs can display foreign peptides on their surfaces in a repetitive and organized manner, which can be used for antigen display against other



pathogens. These are sometimes referred to as chimeric VLPs as they are a combination of protein sequences from two distinct sources. An example is a recent vaccine candidate against human toxoplasmosis, which is caused by *Toxoplasma gondii* infection. Guo *et al.* (2019) fused several different epitopes previously shown to provide protection against this pathogen to hepatitis B virus core antigens (HBc). These HBc-epitope fusions were expressed in *E. coli* and, upon purification, it was found that the proteins self-assembled into stable VLP structures similar to that of native HBc proteins. Upon testing in mice, these VLPs produced strong immune responses and provided protection against *T. gondii* infection (Guo *et al.*, 2019).

### 1.3.2 Protein nanocages

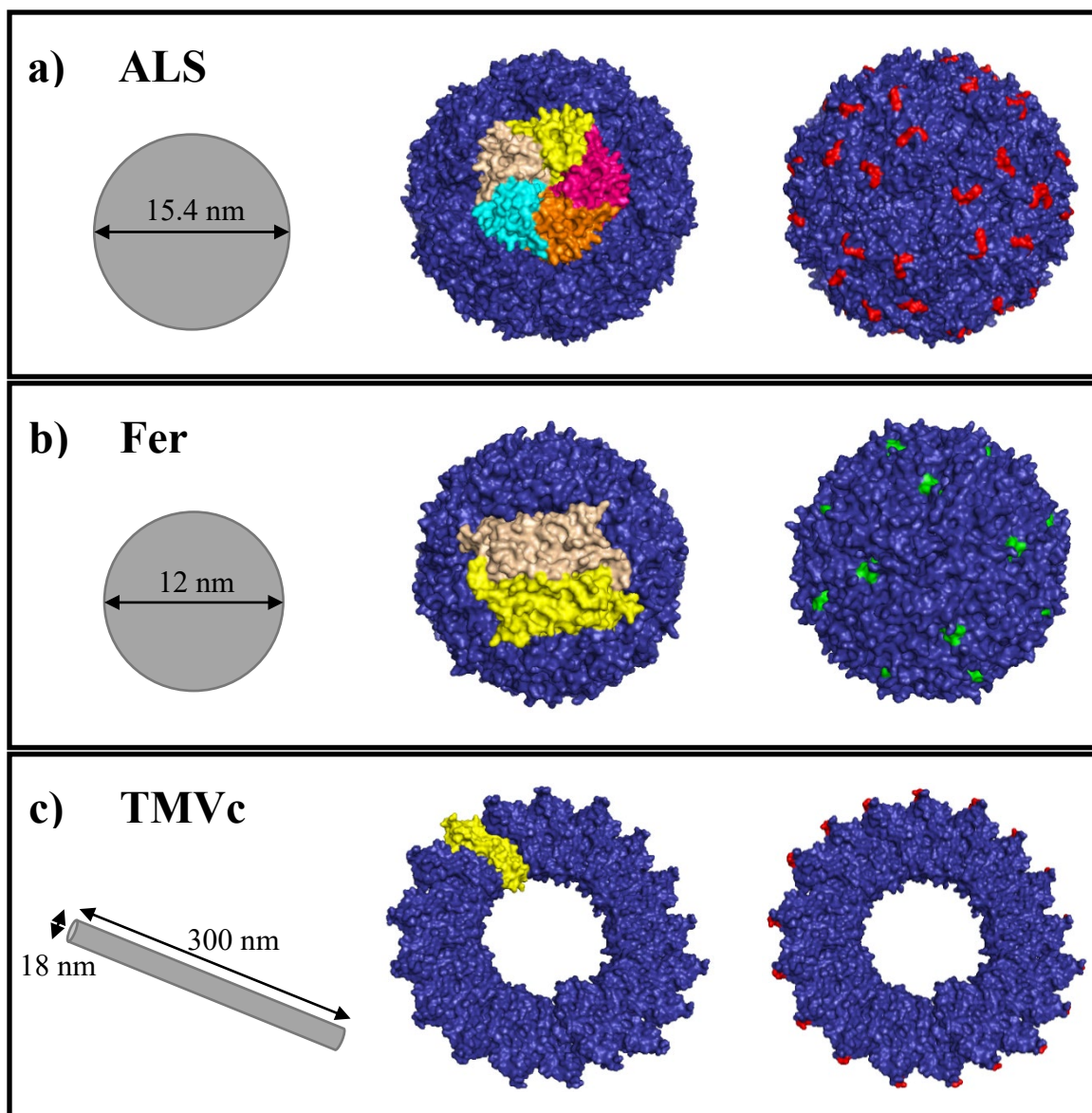
Bacteria often produce self-assembling proteins that vary in shape and size (Diaz *et al.*, 2018). While they serve functions in their host organisms, such as enzymatic activity, many of these nanocages are capable of displaying foreign peptides such as epitopes or full proteins on their outer and/or inner surface, while others are able to encapsulate molecules like drugs or nucleic acids (Bhaskar and Lim, 2017). A major benefit of fusing antigenic peptides to a protein nanocage is that the peptides can be displayed on the nanoparticle surface in an organized, multivalent, repetitive manner reminiscent of a viral particle. This is thought to be a major factor in their improved interaction with the host immune system (López-Sagaseta *et al.*, 2016; Zhao *et al.*, 2014). Presentation of multiple antigens on a particle to host immune cells was shown to improve recognition and immune responses (López-Sagaseta *et al.*, 2016). The antigens can be displayed in a manner similar to that of their native pathogen, inducing protective immune responses.

An example of a nanocage-based vaccine candidate was developed by Qi *et al.* (2018), who set out to create an influenza vaccine by fusing a surface-exposed influenza epitope from its matrix protein 2 onto human ferritin subunits and expressing them recombinantly in *E. coli*. They found the resulting proteins assembled into stable, twenty-four subunit structures displaying the influenza epitope on their surface. Upon intranasal vaccination of mice, the nanoparticle fusions induced high antibody titers and provided complete protection against both homologous and heterologous influenza challenge (Qi *et al.*, 2018). These very promising results demonstrate the viability of using protein nanoparticles as scaffolds for vaccine development.

## **1.4 Nanoparticle platforms of interest**

### **1.4.1 *Aquifex aeolicus* lumazine synthase**

*Aquifex aeolicus* is a hyperthermophilic bacterium found within underwater hot springs (Deckert *et al.*, 1998). A protein of particular interest from *A. aeolicus* is its lumazine synthase (ALS). Lumazine synthases catalyze the penultimate reaction in the riboflavin biosynthesis pathway (Bacher and Mailänder, 1978). Five ALS protein subunits assemble as pentamers, twelve of which self-assemble into sixty subunit nanocages about 15.4 nm in diameter (Figure 3a) (Min *et al.*, 2014; Zhang *et al.*, 2001). The ability of ALS nanocages to display foreign peptides on their outer surface has been demonstrated (Min *et al.*, 2014; Zhang *et al.*, 2020). The addition of foreign peptides to an external loop or either terminus of ALS monomers leads to the formation of uniform structures near-identical to those of native ALS, displaying the foreign peptides on their surface.



**Figure 3. Assembly of ALS, Fer, and TMVc nanoparticles.** **a)** *Aquifex aeolicus* lumazine synthase nanoparticle structure. Initial pentameric assembly is shown in the middle, and surface-exposed C-termini are in red on the right. **b)** *Helicobacter pylori* ferritin nanoparticle structure. Initial dimeric assembly is shown in the middle, and surface-exposed N-termini are in green on the right. **c)** Tobacco mosaic virus coat protein helix and nanorod structure. One monomer (yellow) is shown on a cross section of a single helix in the middle, and surface-exposed N-termini are in red on the right. Protein structures were produced in PyMOL.

One reason this protein is of particular interest is its high thermostability. ALS was found to have a melting temperature of approximately 120°C giving it one of the highest melting temperatures of any protein analyzed to date (Zhang *et al.*, 2006). Another reason ALS is of interest is its large size. Lumazine synthases from other species form different quaternary structures such as from *Brucella abortus*, which forms decameric structures (Zylberman *et al.*, 2004), or from *Saccharomyces cerevisiae*, which forms pentamers (Meining *et al.*, 2000). A larger quaternary structure allows for ease in imaging of the particles and more subunits to display the epitope per nanocage. ALS can have up to sixty copies of the epitope regularly displayed on its surface with the hope of greater interaction with the host immune system. Therefore, due to its size and stability, ALS is a promising display platform for recombinant vaccines.

#### 1.4.2 *Helicobacter pylori* ferritin

Ferritin is a very conserved protein found in a diverse range of organisms. It functions by binding iron, thus controlling its solubility, uptake, and homeostasis within an organism. *Helicobacter pylori* is a bacterium commonly found within the human gut and it has been linked to causing ulcers and other gastrointestinal problems (Suerbaum and Michetti, 2002). The ability to internalize extracellular iron is an important factor for the growth and survival of *H. pylori*, which is a process mediated by ferritin (Waidner *et al.*, 2002). In *H. pylori*, ferritin (Fer) subunits self-assemble into stable twenty-four subunit nanocages (Figure 3b) (Cho *et al.*, 2009). Fer from *H. pylori* was shown to display foreign peptides on its N-terminus and within certain internal loops without disrupting its quaternary structure and nanocage formation. This has been demonstrated several times with foreign peptides of various sizes. For example,

Kanekiyo *et al.* (2015) demonstrated that peptides larger than the Fer subunits themselves could be fused to the nanocage without disrupting its assembly. Domains of the Epstein-Barr virus' glycoprotein 350/220 were fused onto Fer subunits and showed that the protein fusions not only assembled into stable nanoparticles, but they were also able to elicit strong neutralizing antibodies against the native virus in both mouse and non-human primate trials. More recently, Zhang *et al.* (2020) displayed SARS-CoV-2 spike proteins on these Fer subunits for production in a mammalian expression system. They found high expression and self-assembly of nanoparticles displaying the spike proteins in a repetitive and organized manner on their surface.

#### 1.4.3 Tobacco mosaic virus coat protein

The tobacco mosaic virus (TMV), a member of the *Tobamovirus* genus, is a virus that targets and infects a wide range of plant species. The TMV virion consists of a long, rigid nanorod structure approximately 300 nm x 18 nm in size composed of over two thousand identical coat protein subunits (TMVc) (Figure 3c) surrounding the positive-sense viral RNA genome (Klug, 1999; Lomonosoff and Wege, 2018). These nanorod structures are stable under extracellular conditions, however, upon virion entry into host cells the TMVc subunits disassemble to release the RNA genome and initiate the viral replication process. Producing these TMVc proteins in the absence of viral infection and replication, or even the viral genome, allows for the production of virus-like particles. In early experiments, the TMVc subunits only formed smaller ring structures in the absence of its genome due to the instability of their quaternary structures at neutral pH and lower ionic conditions (Durham *et al.*, 1971; Lomonosoff and Wege, 2018). It was later confirmed that repulsive carboxylate groups on TMVc subunits are the cause of this

instability, and modifying specific amino acids (E50Q and D77N) produces more stable nanorod structures under a wider range of conditions (Lu *et al.*, 1996). These modified TMVc can display foreign peptides on their nanorod surface in an organized and repetitive manner (Brown *et al.*, 2013).

An early use of TMVc as a display platform for vaccine development involved fusing an antigenic poliovirus epitope to the C-terminus of TMVc subunits (Haynes *et al.*, 1986). Not only were TMV-like nanorods observed, but these chimeric proteins were also able to induce neutralizing antibodies in rats against the native poliovirus. In more recent years, TMVc epitope fusions have been used for the production of several vaccine candidates. A successful example is involved in the production of a vaccine for feline parvovirus by displaying two of its epitopes separately on TMVc proteins (Pogue *et al.*, 2007). The recombinant vaccine produced strong protective immunity when challenged and greatly reduced clinical signs of illness when infection did occur. All of this together indicates that TMVc can be a promising platform for recombinant vaccine production.

## **1.5 Expression platforms**

### **1.5.1 Bacterial expression**

One expression system frequently used to produce recombinant proteins, including vaccines, is bacteria. Bacteria are relatively simple, single-celled organisms and there has been an abundance of research into certain species, such as *E. coli*, leading to a wide range of tools available for their modification (Francis, 2018). *E. coli* are also relatively easy to transform and have the capacity to quickly produce large quantities of recombinant proteins. A disadvantage to using this expression platform is that bacteria are prokaryotic

and therefore have different protein processing mechanisms than eukaryotes, such as generally lacking machinery for post-translational modifications (Barbosa Viana *et al.*, 2012). This can lead to issues with protein folding or assembly (Lomonossoff and Wege, 2018) and epitope immunogenicity (Wei *et al.*, 2012), depending on the target downstream applications. Because of this, a eukaryotic expression system is preferred for proteins requiring post-translational modifications such as glycosylation.

### 1.5.2 Yeast expression

Another platform for the expression of recombinant proteins is yeast. Different species of yeast have been used, such as *Pichia pastoris*, but the most commonly used is baker's yeast, *S. cerevisiae*. Yeast are also single-celled organisms and have the associated benefits of being inexpensive to grow and produce large amounts of recombinant proteins (Francis, 2018). Unlike bacteria, they are eukaryotic and have more complex protein processing machinery. These post-translational modifications are similar to those of higher eukaryotic organisms, possibly improving protein folding of eukaryotic recombinant proteins (Francis, 2018). A downside of yeast, however, is that the glycosylation patterns added to proteins will be distinct from those produced by plant or mammalian cells, generally being simpler (Barbosa Viana *et al.*, 2012). This can affect folding and epitope immunogenicity of some proteins.

### 1.5.3 Insect cell expression

Insect cells have been developed more recently as an expression platform for recombinant proteins, such as the ovarian cells from *Spodoptera frugiperda* (Francis, 2018). Insertion of a transgene into these cells can be accomplished using the baculovirus

*Autographa californica* containing a sequence of interest, resulting in high levels of recombinant protein production (Zhang *et al.*, 2018). As a more complex eukaryotic expression system, the recombinant proteins undergo post-translational modifications such as glycosylation, phosphorylation, and disulfide bond formation (Francis, 2018), leading to a closer match to their natural counterparts. Insect cell expression platforms also have their downsides, such as the need for bioreactors to grow larger quantities of insect cells which can be very costly to purchase and operate. Also, insect cells are known to have glycosylation patterns that are different from mammalian cells with less complex sugars added (Rychlowska *et al.*, 2011), which can be important for protein folding and immunogenicity when trying to produce recombinant proteins naturally expressed in mammals.

#### 1.5.4 Mammalian cell expression

Another expression platform for recombinant proteins is cultured mammalian cells. This can be ideal for production of authentic post-translational modifications of proteins naturally expressed within mammals as the increased similarity to the native pathogen may improve the level of protective immunity conferred upon vaccination. One downside of using mammalian cells as an expression platform is that they tend to have lower recombinant protein accumulation than other systems (Francis, 2018). Also, as with insect cell expression, it can be more difficult and expensive to use mammalian cells as there are fewer molecular tools available, bioreactors are required for scaling up production, and they require more expensive growth media.

In addition to the technical problems, a risk of using mammalian cells is the possibility of infection with human pathogens. This can completely shut down production



of potentially lifesaving pharmaceuticals, as was the case with Genzyme's Cerezyme, an enzyme replacement therapy for Gaucher disease patients (Mor, 2015). The production of this human  $\beta$ -glucocerebrosidase replacement in Chinese hamster ovary cells was halted when a vesivirus 2117 infection was found in the company's bioreactors (Mor, 2015). Protalix, a company working on a recombinant form of human  $\beta$ -glucocerebrosidase, stepped up to compensate for the loss in production of this lifesaving medicine. This recombinant enzyme from Protalix was produced in carrot cell suspension cultures and targeted to the vacuoles for the addition of proper N-linked glycan chain structures (Shaaltiel *et al.*, 2007). The viral infection of Genzyme's mammalian cell culture bioreactors provided this opportunity for Protalix's enzyme replacement therapy for Gaucher disease to become the first plant-derived protein therapy approved by the US Food and Drug Administration (Mor, 2015).

### 1.5.5 Plant expression

Plants are an expression system that has emerged in the past thirty years and picked up momentum in the past ten years (MacDonald *et al.*, 2015). Plants are easy and inexpensive to grow, requiring only soil, light, water, and fertilizer to thrive. They act as their own bioreactors and different expression techniques have been developed to allow for recombinant protein expression in all or select plant tissue (Marsian and Lomonosoff, 2016). These methods include stable expression in transgenic plants or cell lines, transplastomic plants, and transient expression using agroinfiltration (Lico *et al.*, 2012). In recent years there has been a shift towards increased use of transient expression as it can be quickly and easily scaled up to produce large amounts of recombinant proteins (Margolin *et al.*, 2020). Transient expression also allows for the same plants to be grown

and used for different purposes by changing the *Agrobacterium* cultures infiltrated, thus allowing production to be altered or repurposed more quickly.

Plants have inexpensive growth requirements and relatively easy infiltration methods, allowing for easy scale up of production (Lai and Chen, 2012; Margolin *et al.*, 2020) compared to some of the other expression platforms, such as insect and mammalian cells. Several promising vaccines and candidates have been produced in plant systems, including the influenza VLP designed by Medicago Inc. (Quebec, QC) (D'Aoust *et al.*, 2008; Ward *et al.*, 2020). Transiently expressed H5N1 and H1N1 hemagglutinin in *Nicotiana benthamiana* produced enveloped VLPs able to induce complete protection from both homologous and heterologous viral challenges. There are increasing numbers of vaccine candidates using plant expression progressing through clinical trials, highlighting the increased interest in this platform. The Medicago SARS-CoV-2 vaccine produced in *N. benthamiana* is the first plant-derived vaccine against this disease to reach clinical trials (Ward *et al.*, 2021). Enveloped VLPs with the viral spike protein were produced using transient expression in *N. benthamiana*, resulting in VLPs similar in size and shape to the SARS-CoV-2 virus displaying a high density of spike proteins on their surface. Phase I clinical trials in humans found it to be safe and highly immunogenic, especially when paired with an adjuvant, producing neutralizing antibody levels similar to those of individuals recovering from the disease itself (Ward *et al.*, 2021).

Plants are eukaryotic and therefore have more complex protein processing machinery (Strasser *et al.*, 2021). A specific benefit of plant expression is that proteins can be targeted to a wider range of subcellular compartments than other expression platforms, such as vacuoles or chloroplasts. Proteins targeted to different subcellular compartments

acquire different N-linked glycosylation patterns (Faye *et al.*, 2005; Strasser *et al.*, 2021). Because of this, researchers can match the absence, presence, and type of glycan chains on recombinant proteins to their native counterparts including proteins from prokaryotic origins where no glycosylation is desired. It is important to note, however, that the glycosylation patterns from plants may not always perfectly match those from mammalian cells, particularly the more complex modifications of the late Golgi (Margolin *et al.*, 2020). Research is being performed aiming to “humanize” plant glycosylation, matching it more closely to those of mammalian cells. This involves avoiding the addition of plant-specific sugar residues on the N-linked glycan chains and adding pathways for the inclusion of several mammalian-specific sugar residues (Margolin *et al.*, 2020). Results indicate that plant expression of proteins with mammalian-type N-linked glycan chains can be accomplished.

## **1.6 Rationale and goals**

One goal of this project is to investigate recombinant protein accumulation levels of plant-produced ALS, Fer, and the modified TMVc. This is important as previous work has only examined the synthesis of these proteins in prokaryotic systems and this is the first attempt at expressing them in plants.

The second goal of this project is to investigate whether and how the genetic fusion of the PRRSV M-GP5 epitope to these three self-assembling proteins alters their soluble protein accumulation levels upon expression in plants. Previous work with ALS, Fer, and other protein nanocages found that upon expression in *E. coli*, fusion of a PRRSV epitope greatly decreased their soluble protein accumulation levels (Harper, 2019). Insoluble protein fractions are largely inaccessible without more laborious and costly unfolding and

refolding steps therefore soluble proteins, which tend to be correctly folded, are preferred. Previous work has also demonstrated that some proteins are more soluble when expressed in plant systems (Hong Zhu and Rima Menassa, personal communications), which is the basis for this research goal.

## **1.7 Objectives**

1. To express ALS, Fer, and modified TMVc on their own in *N. benthamiana* and examine their soluble protein accumulation in three subcellular compartments of the secretory pathway.
2. To express the three self-assembling proteins genetically fused to the M-GP5 epitope in *N. benthamiana* and compare their soluble protein accumulation levels to each other and their self-assembling proteins alone.

## 2 MATERIALS AND METHODS

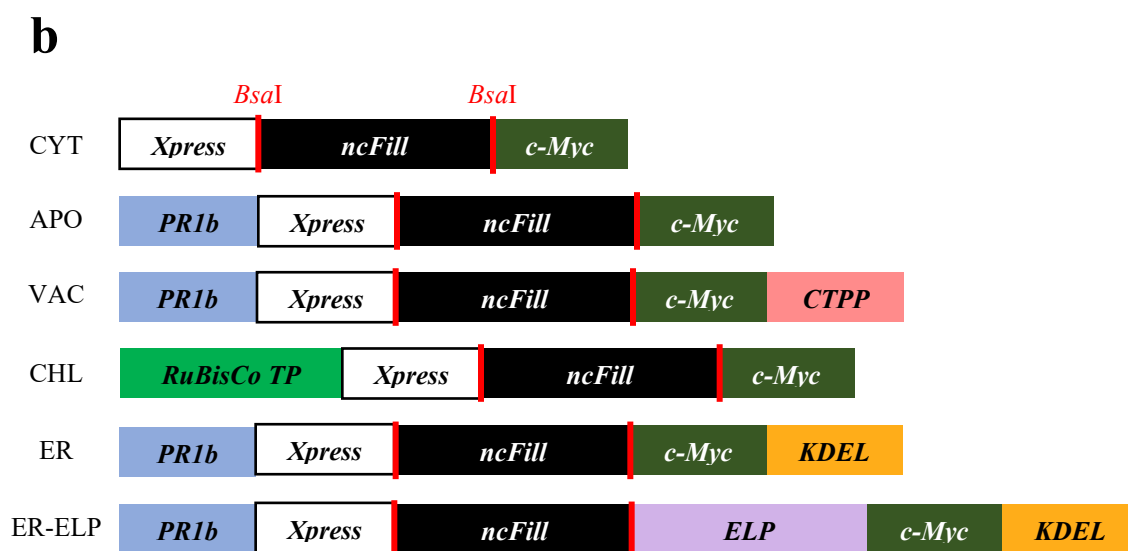
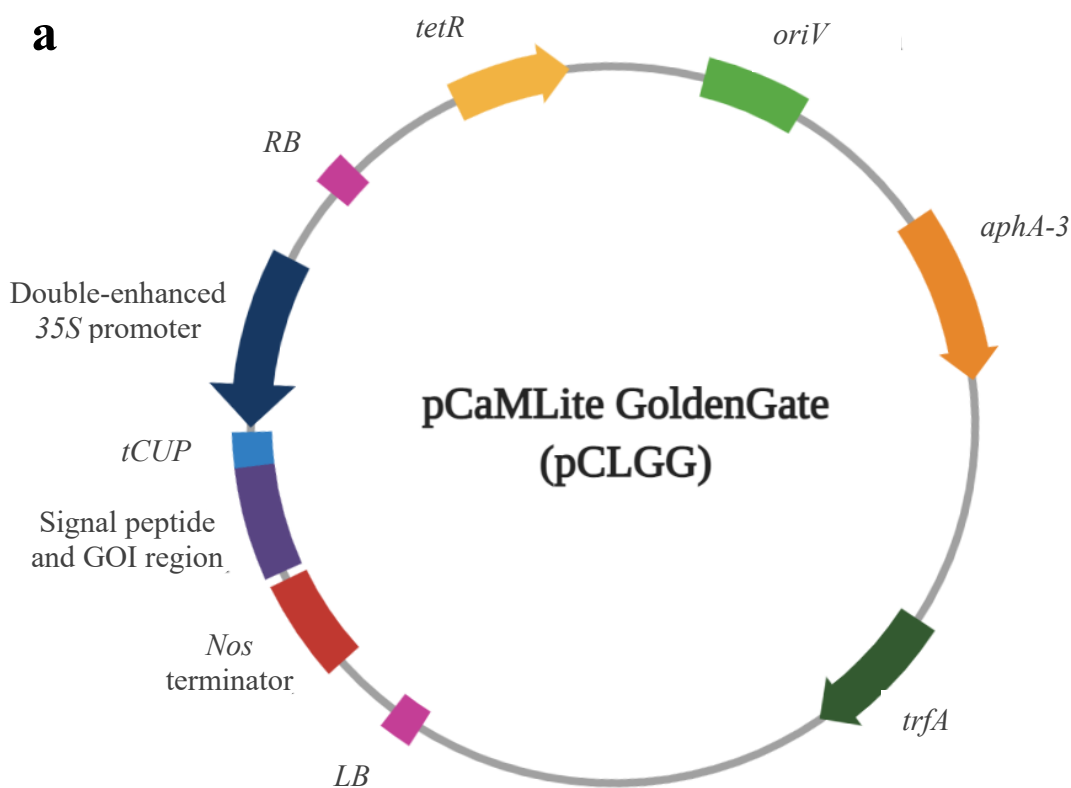
### 2.1 Plant expression vectors

#### 2.1.1 Original vectors

The pCaMLite GoldenGate (pCLGG) plant expression vectors allow for proteins to be expressed in plants (Figure 4a). Five of these pCLGG vectors were designed to target expressed proteins to different subcellular locations within plant cells including the apoplast (APO), vacuole (VAC), ER, cytoplasm (CYT), and chloroplast (CHL). The sixth vector also targets proteins to the ER but has the addition of an elastin-like polypeptide (ELP) tag that causes increased protein accumulation to ER-targeted proteins (Conley *et al.*, 2009) (Figure 4b).

All six of these plant expression vectors contain a double-enhanced 35S promoter (Kay *et al.*, 1987) and tobacco cryptic upstream promoter translational enhancer (*tCUP*) (Wu *et al.*, 2001) for high expression, the *Nos* terminator (Bevan *et al.*, 1983), a *c-Myc* tag for detection and purification of the encoded proteins, and an *Xpress* tag for increased protein accumulation (Figure 4b). A PR1b signal peptide from tobacco is encoded in the APO, VAC, ER, and ER-ELP pCLGG vectors for targeting proteins to the secretory pathway (Cutt *et al.*, 1988). The ER and ER-ELP vectors also encode a KDEL sequence for retrieval of the proteins to the ER (Pelham, 1990). The vacuole targeting vector contains the coding sequence for a C-terminal propeptide (CTPP) vacuolar sorting signal from the tobacco chitinase gene (Vitale and Raikhel, 1999). The transit peptide from the tobacco RuBisCo small subunit is encoded in the CHL vector to target proteins to the chloroplast (Figure 4b) (Gnanasambandam *et al.*, 2007).

**Figure 4. Original pCLGG plant expression vectors.** **a)** Vector map of pCaMLite GoldenGate (pCLGG) plant expression vectors. The *oriV* is the bacterial origin of replication (light green), and *trfA* (dark green) encodes the initiator protein for *oriV*. The *aphA-3* gene (orange) encodes an aminoglycoside phosphotransferase, which confers kanamycin resistance to bacteria and allows for selection of transformed colonies. The tetracycline resistance gene, *tetR* (yellow), encodes a tetracycline resistance regulatory protein. The left (*LB*) and right (*RB*) borders (pink) of the T-DNA flank the sequence transferred to the plant cell. The double-enhanced *35S* promoter (dark blue), *tCUP* translation enhancer (light blue), and *Nos* terminator (red) facilitate expression. The signal peptide and gene of interest (GOI) region (purple) contains the only differences between the six pCLGG vectors. **b)** Signal peptide and GOI regions for the pCLGG vectors. The signal peptides encoded in these vectors target proteins to the cytoplasm (CYT), apoplast (APO), vacuole (VAC), chloroplast (CHL), and ER (ER and ER-ELP). The non-coding filler sequences (*ncFill*: black) is where the sequences of interest are inserted using *BsaI* restriction enzyme sites (red lines) that flank the *ncFill* for GoldenGate cloning. The targeting peptides encoded in these vectors include the PR1b secretory-pathway signal peptide (light blue), the C-terminal propeptide (CTPP) from the tobacco chitinase for vacuolar targeting (pink), a KDEL for ER retrieval (orange), and the transit peptide from the small subunit of RuBisCo for targeting to the chloroplast (light green). The *c-Myc* tag (dark green) is on the 3' end of every inserted gene to act as a detection and purification tag on the produced proteins. The *Xpress* tag (white box) is on the 5' end of every inserted gene. The elastin-like polypeptide (*ELP*: violet) encodes a large tag that causes increased protein accumulation to ER-targeted proteins. Schematics not drawn to scale.



### 2.1.2 Modification of vectors

To remove the *Xpress* tag, all six pCLGG plant expression vectors were modified by identifying unique restriction enzyme (RE) recognition sequences flanking the regions to be modified. New sequences were designed i) lacking the *Xpress* tag and ii) replacing it with codons for two glycine residues. These sequences were synthesized by BioBasic.

Digests were performed on the modified sequences and original vectors separately using the selected RE pairs (New England Biolabs) per manufacturer's instructions. The restriction fragments were size separated using agarose gel electrophoresis and the fragments of interest were extracted from the gel using the QIAGEN QIAquick Gel Extraction Kit per manufacturer's instructions. Ligations were performed between the vectors and inserts using a T<sub>4</sub> DNA ligase (New England Biolabs) per manufacturer's instructions. These modified vectors, denoted as pCaMLite GoldenGate-Xpress (pCLGG-X), were transformed into NEB 10- $\beta$  *E. coli* using heat shock, plated onto Luria-Bertani (LB: 10 g tryptone, 5 g yeast extract, and 10 g NaCl for 1 L of broth, pH-adjusted to 7.0) plates (15 g agar per litre of LB) with 50  $\mu$ g/ml kanamycin (LB<sup>kan</sup>), and incubated overnight at 37°C. Individual colonies were selected and screened for both successful transformation and modification using colony polymerase chain reaction (PCR) with the 35S forward or Xpress forward primer paired with the Nos reverse primer (Appendix 3). PCRs were run at 95°C for 5 min, then 30 cycles of 95°C for 30 sec, 62°C for 30 sec, and 72°C for 2 min, followed by 72°C for 7 min. Cultures were incubated overnight from the selected colonies at 37°C in liquid LB<sup>kan</sup> media. The plasmids were isolated from *E. coli* cultures using the QIAGEN QIAprep Spin Miniprep Kit per manufacturer's instructions, and sequenced at Eurofins Genomics.



## 2.2 Construct sequences

### 2.2.1 Self-assembling protein sequences

The *ALS* sequence corresponds to the entire ALS monomer (accession number: WP\_010880027) and the *Fer* sequence corresponds to the entire ferritin monomer (accession number: WP\_000949210). The *TMVc* sequence was based on the entire TMVc monomer from Brown *et al.* (2013), containing two modifications compared to wild type (E50Q and D77N). All genes were synthesized by BioBasic flanked by *BsaI* sites for GoldenGate cloning. See Appendix 1 for nucleotide sequences. Protein nanoparticle structures were visualized with PyMOL software (Schrödinger Inc.).

### 2.2.2 PRRSV epitope sequences

The PRRSV epitope was composed of amino acids 2-18 of the M protein (accession number: AAO13197.1) and amino acids 30-54 of GP5 (accession number: AAO13196.1) from PRRS virus strain VR-2332 (Harper, 2019). All corresponding nucleotide sequences (Appendix 1) were synthesized by BioBasic.

## 2.3 Cloning and transformations

All genes were cloned into pCLGG-X vectors targeting proteins to the apoplast, vacuole, or ER using GoldenGate Cloning methodology (Marillonnet and Werner, 2015). All vectors were then transformed into *E. coli*, plated, and screened using colony PCR with 35S forward and Nos reverse primers (Appendix 3). Glycerol stocks were produced by mixing 750  $\mu$ l of cell culture with 750  $\mu$ l of sterile 30% glycerol and storing at  $-80^{\circ}\text{C}$ . The plasmids were then isolated from 5 ml *E. coli* liquid cultures.

Isolated vectors were transformed into *Agrobacterium tumefaciens* EHA105 cells (Hood *et al.*, 1993) using electroporation (Wise *et al.*, 2006), plated on LB plates containing 50 µg/ml kanamycin and 10 µg/ml rifampicin (LB<sup>kan+rif</sup>), and incubated two days at 28°C. Individual colonies were screened for successful transformation using colony PCR, and positive colonies were grown overnight at 28°C in LB<sup>kan+rif</sup> media and stored in glycerol as described for *E. coli*. Plasmids were isolated from the *Agrobacterium* cultures as described for *E. coli* and sequenced at Eurofins Genomics or the London Regional Genomics Centre.

## 2.4 Infiltrations and sampling

### 2.4.1 Plant growth and care

Wild-type *N. benthamiana* plants were grown at 22°C with 65% humidity on a light:dark cycle of 16:8 h in a walk-in growth chamber. Plants were grown in 4” pots with PRO-MIX BX soil, and were watered as necessary using 20:8:20 fertilizer (nitrogen:phosphorus:potassium) at 0.25 g/L of water.

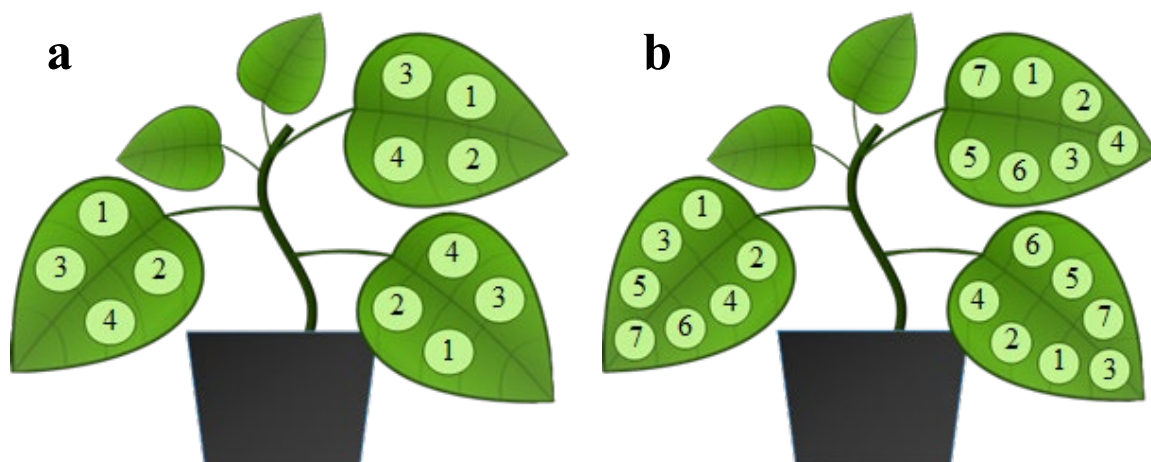
### 2.4.2 Infiltration and sampling of *Nicotiana benthamiana*

For infiltrations, transformed *Agrobacterium* cultures were grown in LB<sup>kan+rif</sup> media overnight at 28°C. These cultures were then used to inoculate infiltration culture medium (LB, 10 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 5.6, 100 µM acetosyringone, 50 µg/ml kanamycin, and 10 µg/ml rifampicin) and grown at 28°C to an OD<sub>600</sub> of between 0.5-1.2 overnight. *Agrobacterium* cells were then pelleted and resuspended in Gamborg’s solution (3.2 g/L Gamborg’s B5 + vitamins, 20 g/L sucrose, 10 mM MES (pH 5.6), and 200 µM acetosyringone), followed by incubation with gentle agitation at room temperature for one hour (Miletic *et al.*, 2016).

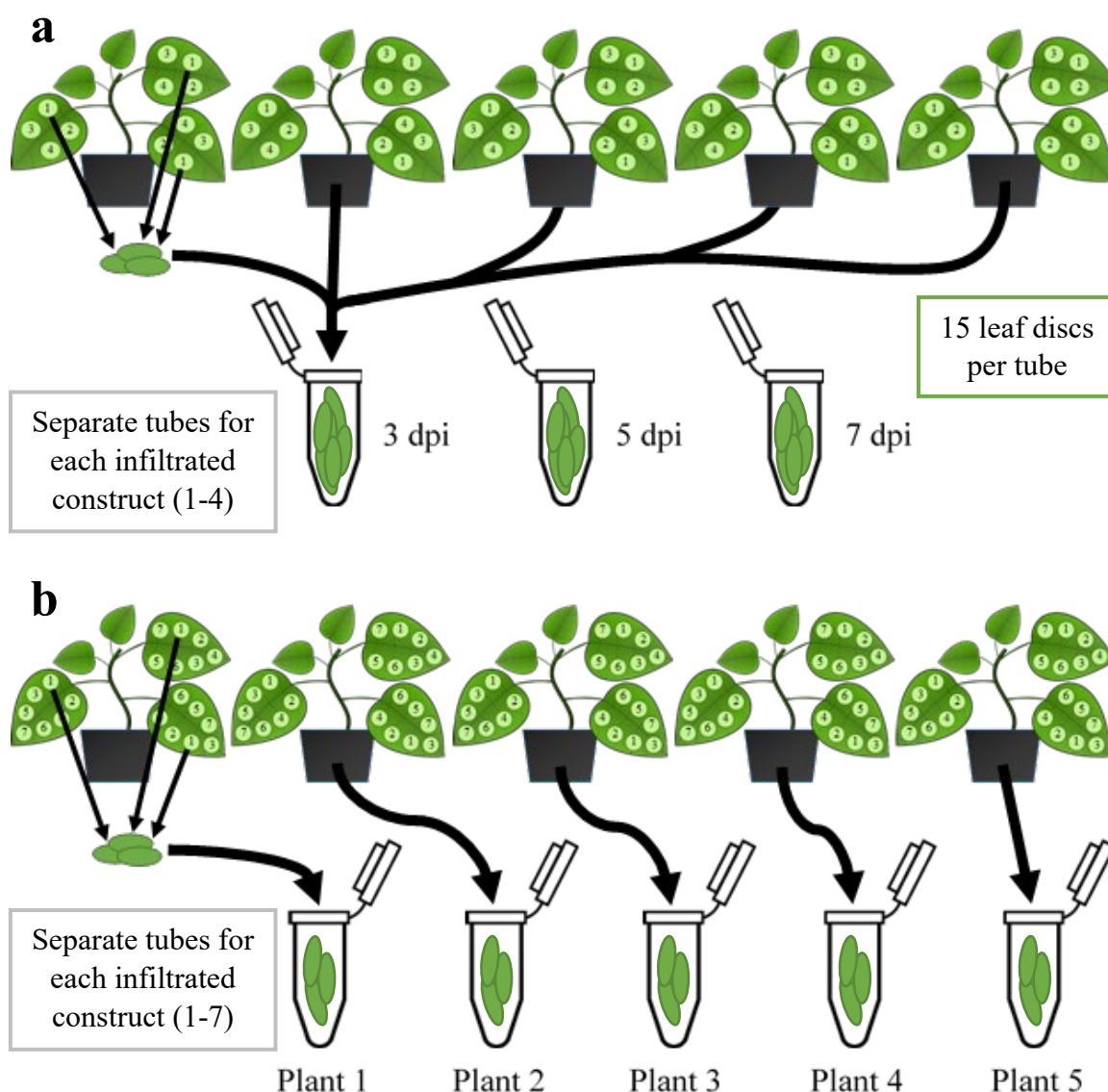
Transient expression in *N. benthamiana* plants was achieved by infiltrating *Agrobacterium* cultures into the abaxial leaf surface of 4-6 week old *N. benthamiana* plants using a needleless syringe (Miletic *et al.*, 2016). For each infiltration, three leaves on five plants were co-infiltrated with cells carrying *p19*, as well as a *p19*-only negative control. *p19* is a suppressor of posttranscriptional gene silencing, a plant defense mechanism, from the *Cymbidium* ringspot virus (Silhavy *et al.*, 2002).

Each self-assembling protein construct (ALS, Fer, and TMVc without the epitope) was infiltrated into each three leaves per plant using three of the pCLGG-X vectors targeting proteins to the apoplast, vacuole, and ER, as well as *p19* alone (Figure 5a). Separate sets of five plants were used for each self-assembling protein. For sampling, one leaf disc was taken from every infiltrated area on each of days 3, 5, and 7 post-infiltration (dpi). Leaf discs were pooled by sample and by day from the five plants, resulting in a pool of 15 leaf discs for each construct and each subcellular localization at each day (Figure 6a). These transient expression experiments were repeated twice.

Infiltrations of fusion constructs included each self-assembling protein construct and its respective M-GP5 fusion in the pCLGG-X-APO vectors. All six constructs and *p19* were infiltrated into three leaves per plant, and five plants were used as biological replicates (Figure 5b). For sampling, one leaf disc was taken at 7 dpi from every infiltrated area with discs pooled by plant, resulting in 3 leaf discs per tube (Figure 6b). Each plant was analyzed separately and considered as a biological replicate. This transient expression experiment was repeated three times. All plant tissue samples were weighed, flash-frozen in liquid nitrogen, and stored at -80°C until processed.



**Figure 5. Agroinfiltration of *N. benthamiana* leaves for transient protein expression.** *Agrobacterium* cultures for each construct were co-infiltrated with equal parts p19 culture into the abaxial side of leaves. **a)** Infiltrations of the self-assembling protein constructs were performed separately for ALS, Fer, and TMVc. Three leaves on five plants were infiltrated with the APO (1), VAC (2), and ER (3) targeted constructs as well as a p19 alone negative control (4). **b)** For the protein fusion constructs, three leaves on five plants were infiltrated with APO-targeted ALSm (1), Ferm (2), TMVc (3), and their M-GP5 fusions (4-6). A p19 negative control was also included (7).



**Figure 6. Sampling of infiltrated *N. benthamiana* leaf tissue.** One leaf disc was taken from each infiltrated area on each day sampled. **a)** Sampling of the self-assembling proteins (ALS, Fer, TMVc) was performed on days 3, 5, and 7 post-infiltration (dpi). For each construct each dpi, leaf discs were pooled across all five plants (3 leaves on 5 plants = 15 discs/tube). **b)** Sampling of the protein fusions was performed on 7 dpi. For each construct, leaf discs were pooled by plant allowing for each plant to act as a biological replicate. All tissue samples were weighed, flash-frozen, and stored at  $-80^{\circ}\text{C}$ .

## 2.5 Recombinant protein quantification

To isolate total soluble proteins from the plant tissue, frozen leaf discs were homogenized using the TissueLyser II (QIAGEN) twice for one minute at 30 Hz, then extraction buffer (0.1% Tris-buffered saline (TBS) with Tween 20, 2% polyvinyl-pyrrolidone (PVPP), 1 mM ethylene-diaminetetraacetic acid (EDTA), 1 mM phenyl-methylsulfonyl fluoride (PMSF), 1  $\mu$ g/ml leupeptin, and 100 mM sodium-ascorbate) (Miletic *et al.*, 2016) was used at a 1:3, 1:5, or 1:10 ratio of plant mass (mg) to buffer volume ( $\mu$ l) and samples were vortexed at maximum speed for 30 seconds. After centrifugation at 23,500 x g for 15 minutes at 4°C, the supernatants containing total soluble proteins were transferred to new tubes. Samples were prepared by mixing protein extracts with 1x TBS and 5x loading dye (0.3 M Tris-Cl pH 8.0, 5% sodium dodecyl sulfate (SDS), 50% glycerol, 100 mM dithiothreitol (DTT), 0.05% phenol red, and water). The samples were then denatured at 90°C for 10 minutes and separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using SurePAGE pre-cast 4-20% gradient gels (GenScript, Piscataway, NJ, USA). The proteins were transferred to a polyvinylidene difluoride (PVDF) membrane and blocked using a 5% solution of skim milk in TBS for at least one hour. For immunoblotting, 1  $\mu$ l of mouse anti-hemagglutinin (HA) primary antibody (Sigma-Aldrich) was mixed with 10  $\mu$ l of WB-1 solution from the ONE-HOUR Western Basic Kit (Mouse) from GenScript (containing anti-mouse horseradish peroxidase-conjugated secondary antibodies) and incubated at room temperature for at least 40 minutes. This solution was then mixed with 5 ml WB-2 (2% bovine serum albumin in TBS) and used to hybridize to the membrane, resulting in a 1:5000 dilution of primary antibody. Recombinant protein accumulation was visualized

using Clarity Western ECL Blotting Substrates (Bio-Rad) and quantified compared to a standard curve of known concentrations of a protein standard using GelQuant software (DNR Bio-Imaging Systems Ltd.). The protein standard, eGEHK, is a synthetic protein designed with multiple detection tags by Angelo Kaldis at AAFC London.

## **2.6 Statistical modelling**

Soluble protein accumulation levels of the recombinant self-assembling proteins and their M-GP5 fusions were calculated in mg/g of fresh plant tissue. This was accomplished by considering the quantification values from the blots, the volume of plant protein extraction buffer used, the mass of plant tissue per sample, and dilution of samples loaded on gels (i.e. volume of extract loaded per lane).

The resulting soluble protein accumulation data ( $n = 15$  per protein: 5 biological replicates per experiment and 3 transient expression experiments) was examined using R 4.0.3 (R Core Team, 2020). The analysis began with checking the data for outliers, zero inflation, balance of categorical covariates, and collinearity. A linear model was used originally, however, it was clear the assumption of equal variance was violated so a generalized least squares (GLS) model was applied as it does not assume equal variance. Soluble protein accumulation was modelled as a function of the protein, replicate number (transient expression experiment 1-3), and their interaction. A variance of identity residual structure was also used to allow for separate variance for each protein. The GLS model was validated by checking the residual distribution, residuals vs fitted values, and residuals against covariates. A two factor ANOVA was performed, followed by a Games-Howell post hoc test (PMCMRplus package) which does not assume equal variance.

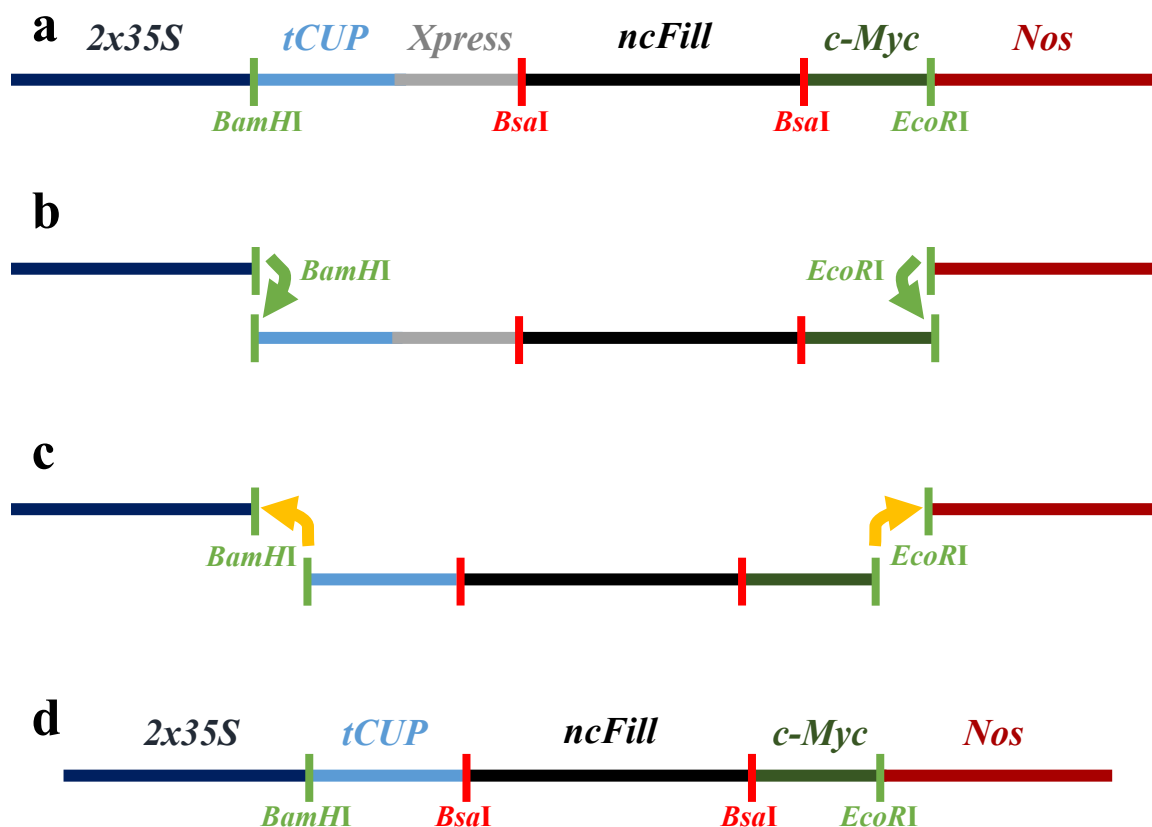
## 3 RESULTS

### 3.1 Modification of plant protein expression vectors

When the pCLGG plant expression vectors were produced, an *Xpress* tag was initially added to be used as an N-terminal protein detection tag. This was not found to be effective, however, it appeared to increase the protein accumulation levels so the tag was maintained in the vectors for this purpose (Conley et al., 2009). To reduce possible interference with nanoparticle assembly or masking of the epitope displayed on their surface, I decided to remove the *Xpress* tag from the pCLGG vector series. In this project this was especially relevant for the M-GP5-Ferm fusion proteins as the *Xpress* tag and the epitope are both on the N-terminus. Removal of the *Xpress* tag was also important for other projects, especially those working with self-assembling proteins that cannot tolerate additional amino acids on either of their termini. The C-terminal c-Myc tag can be easily excluded by incorporating a stop codon before its coding sequence, but this is not an option for the current N-terminal *Xpress* tag which is placed between the signal peptide sequences and the sequence of interest. The newly modified pCLGG-X plant protein expression vectors avoid this issue, benefiting this project and others.

The initial step of removing the *Xpress* tag from the pCLGG plant expression vectors was to identify unique RE sites flanking the area to be modified. For the APO, VAC, ER, CYT, and CHL vectors *Bam*HI and *Eco*RI restriction sites were selected (Figure 7). Following design and synthesis of the modified insert regions (lacking the *Xpress* tag) and their digestion in parallel with the original pCLGG vectors, agarose gel electrophoresis showed the digestions were successful and fragments of predicted sizes were present (data not shown).





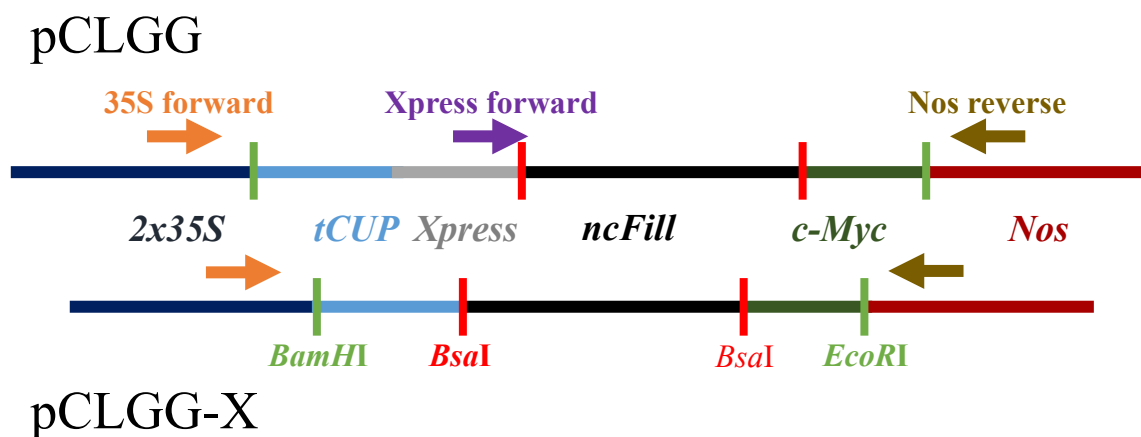
**Figure 7. Modification pCLGG plant expression vectors to remove the *Xpress* tag.**

**a)** All pCLGG vectors (pCLGG-CYT shown) contain a double-enhanced 35S promoter (dark blue), a *tCUP* translation enhancer (light blue), an *Xpress* tag (grey), a *c-Myc* tag (dark green), and *Nos* terminator (dark red). They also contain the *ncFill* sequence (black), which is where the sequence of interest is inserted using the *Bsa*I cloning sites (light red). **b)** Digestions for the APO, CYT, CHL, ER, and VAC vectors were performed using *Bam*HI and *Eco*RI restriction enzyme cut sites (light green), removing the region containing the *Xpress* tag. **c)** Modified insert regions without the *Xpress* tag were designed, synthesized, and digested with the same restriction enzymes. These inserts were ligated into the digested original vectors and transformed into *E. coli*. **d)** The resulting pCLGG-X plant protein expression vectors are nearly identical to the original vectors, but lack the *Xpress* tag.

The empty pCLGG vector fragments and the modified *Xpress*-less region fragments were excised from the gels, purified, and ligated (Figure 7). Following ligation, the newly modified vectors were transformed into *E. coli*. For each vector, two colony PCR reactions were performed using either the 35S forward or *Xpress* forward primers paired with the Nos reverse primer (Figure 8) using several individual colonies as templates. Each sample amplified one PCR fragment of expected size with the 35S forward and Nos reverse primers (data not shown). At least one colony for each vector did not have a fragment present in the *Xpress* primer lane, demonstrating the *Xpress* tag was absent.

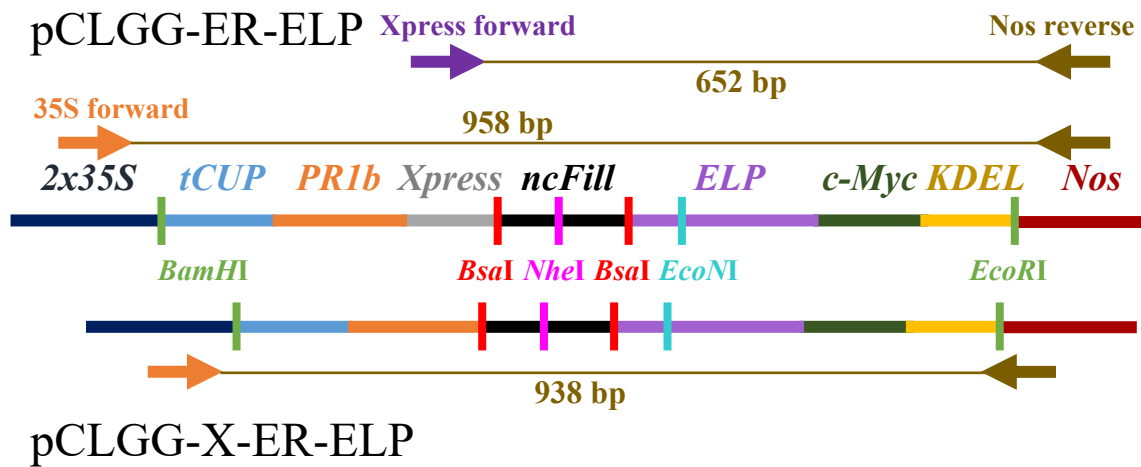
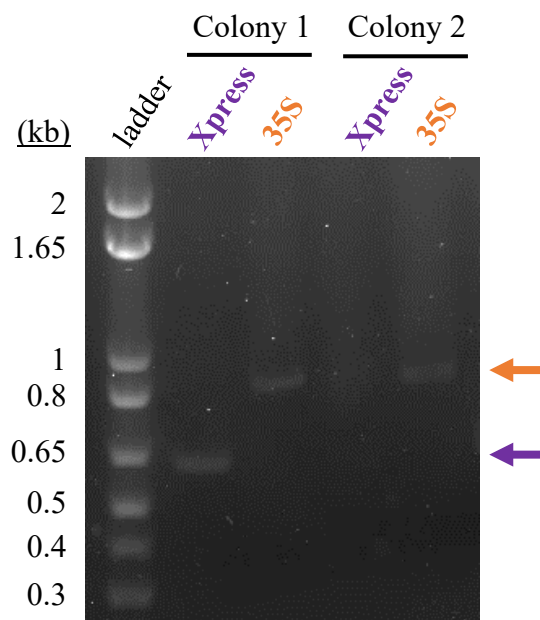
The ER-ELP vector was digested with *Eco*NI instead of *Eco*RI with *Bam*HI to avoid synthesizing the *ELP* tag (Figure 9a). The *ELP* tag is difficult to synthesize as it is long and repetitive, therefore an *Eco*NI site within the 5' end of the *ELP* tag was selected. However, attempts at ligating the ER-ELP insert and empty vector were unsuccessful. Therefore, several nucleotides within the *ncFill* sequence were mutated to create a new, unique RE site (*Nhe*I) on the pCLGG-ER-ELP vector (Figure 9a). A new modified insert lacking the *Xpress* tag was also synthesized having the corresponding *Bam*HI and *Nhe*I sites. When these were digested, ligated, and transformed into *E. coli*, colony PCR results showed successful cloning (band present using 35S forward primer with Nos reverse) and no *Xpress* tag (no band using *Xpress* forward primer with Nos reverse) (Figure 9b).

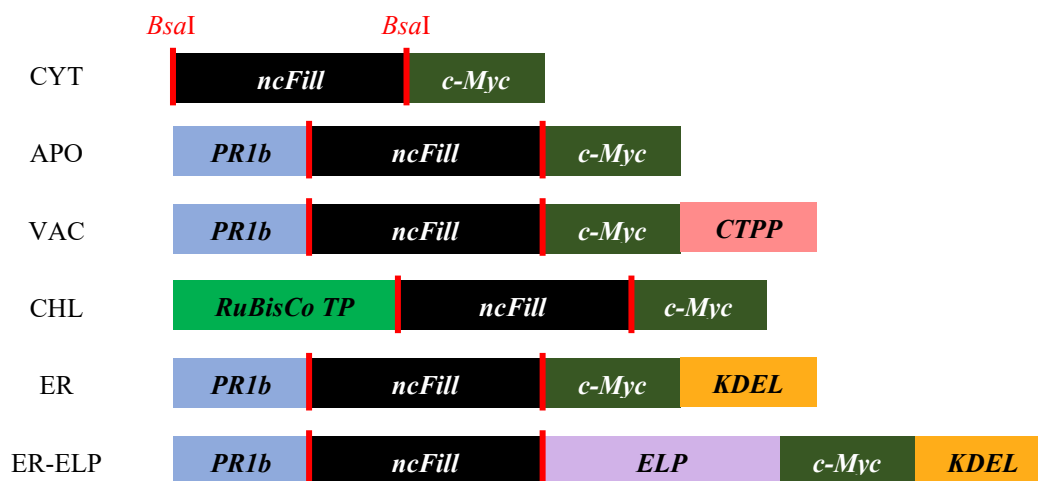
For each modified vector (Figure 10), DNA was isolated from a positive colony and sent for sequencing, which showed the new vectors lacked the *Xpress* tag as expected (data not shown).



**Figure 8. Schematic for primer binding to pCLGG and pCLGG-X plant expression vectors.** Original pCLGG region is on top, with the modified pCLGG-X region on the bottom (CYT vectors shown). Colony PCRs were performed using either the 35S forward (orange) or Xpress forward (purple) primers paired with the Nos reverse primer (brown).

**Figure 9. Modification of pCLGG-ER-ELP vector.** **a)** The pCLGG-ER-ELP vector contains a double-enhanced *35S* promoter (dark blue), *tCUP* translation enhancer (light blue), *PR1b* transit peptide nucleotide sequence (orange) for protein targeting to the secretory pathway, and *Xpress* tag (grey). The *ncFill* (black) is replaced with the sequence of interest using the *BsaI* cloning sites (light red). The *ELP* (violet), *c-Myc* tag (dark green), and *KDEL* (yellow) precede the *Nos* terminator (dark red). The pCLGG-X-ER-ELP vector is nearly identical, lacking the *Xpress* tag. The *BamHI* and *EcoRI* restriction enzyme (RE) sites (light green) were used for modifying the other pCLGG vectors. The *EcoNI* site (teal) was originally used with the *BamHI* site to modify the ER-ELP vector, however, an *NheI* site (pink) was later introduced for use instead. Colony PCR was performed with either the *35S* forward primer (orange arrow) or the *Xpress* forward primer (purple arrow) with the *Nos* reverse primer (brown arrow). The amplified fragments and lengths are shown in brown. **b)** Transformed *E. coli* were screened by colony PCR using the two primer pairs (*35S* with *Nos*, or *Xpress* with *Nos*). The expected size for the modified ER-ELP vector with the *35S* primer pair is 938 bp (orange arrow), and for the *Xpress* primer pair it is 652 bp (purple arrow). Colony 1 shows a band of approximately expected size in each lane, meaning it was successfully transformed but its vector still contained the *Xpress* tag. Colony 2 only shows a band in the *35S* primer lane, meaning it was successfully transformed with a vector lacking the *Xpress* tag. Colony 2 therefore contains the modified pCLGG-X-ER-ELP vector.

**a****b**



**Figure 10. Modified pCLGG-X plant expression vectors.** These vectors target proteins to the cytoplasm (CYT), apoplast (APO), vacuole (VAC), chloroplast (CHL), and ER (ER and ER-ELP). The *ncFill* (black) is where the genes of interest are inserted. This is accomplished using *Bsa*I restriction enzyme sites (red lines) for GoldenGate cloning that flank the *ncFill* region. The signal peptides encoded in these vectors include the *PR1b* secretory-pathway targeting tag (light blue), the C-terminal propeptide (*CTPP*) for vacuolar targeting (pink), *KDEL* for ER retention (orange), and *RuBisCo* transit peptide for targeting to the chloroplast (light green). The *c-Myc* tag (dark green) is on the 3' end of every inserted gene to act as a detection and purification tag on the produced proteins. The elastin-like polypeptide (*ELP*: violet) encodes a large tag that causes increased protein accumulation of ER-targeted proteins.

### 3.2 Self-assembling protein construct design

Nanoparticle-based subunit vaccines are a promising tool for developing safer, more effective methods of controlling livestock diseases such as PRRS. Previous work found that when self-assembling proteins, including ALS and Fer, were expressed in *E. coli*, fusion of PRRSV epitopes greatly decreased their solubility (Harper, 2019). Their best candidate was a *Brucella* lumazine synthase (BLS) fused to an M-GP5 epitope, however, most of the protein accumulation was still insoluble. They also found M-GP5-BLS assembled into nanoparticle structures similar in shape and size to BLS on its own. Recent work in the Menassa lab found evidence of a similar BLS and M-GP5 fusion remaining soluble with high accumulation when expressed in plants (Hong Zhu and Rima Menassa, personal communications). Therefore, plants appear to be a better expression platform for soluble production of these nanoparticle-based vaccine candidates. Because of these results I decided to explore the plant-production of some larger nanoparticles examined by Harper (2019), including ALS and Fer. While BLS forms decamers, ALS and Fer form 60 and 24 subunit complexes, respectively, allowing for the display of more epitopes per nanoparticle. Also, ALS and Fer were selected for this project as neither were examined in plants previously.

Wild-type TMVc has been studied extensively in both plants and bacteria, however, a modified sequence shown to have improved nanoparticle assembly and stability was used in this project. This modified TMVc sequence from Brown *et al.* (2013) was investigated in *E. coli* and had not been expressed in a plant system until now. Thousands of TMVc subunits are expected to assemble into very large nanorod structures, thereby displaying thousands of copies of the epitope on the nanorod surface.

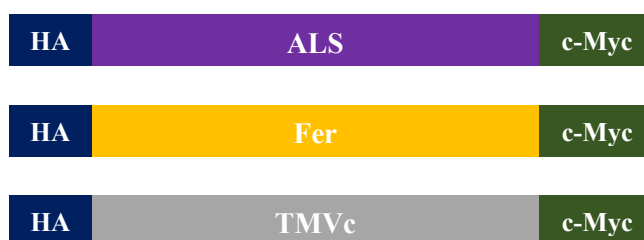
Design of these genes included the addition of the coding sequence for an HA tag for detection on the proteins' N-termini and a c-Myc tag from the pCLGG-X vectors on their C-termini (Figure 11). See Appendix 2 for full protein sequences.

### **3.3 Expression analysis of self-assembling proteins**

To determine if the designed genes express and the encoded proteins accumulate in plants, I used transient agroinfiltration of the constructs containing the sequences for the self-assembling proteins without the PRRSV epitope. This was the first time wild type ALS, Fer, and the modified TMVc were expressed in plants. The recombinant proteins were targeted to organelles in the secretory pathway (apoplast, vacuole, and ER) because in its native host, the pig, the PRRSV is secreted and GP5 acquires complex-type glycans (Li *et al.*, 2015). Targeting the recombinant proteins to the plant apoplast is preferred because the resulting N-linked complex glycans on the GP5 epitope are expected to most closely match those of the native PRRS virus (Li *et al.*, 2015; Strasser *et al.*, 2021). However, some proteins do not accumulate well in the plant apoplast due to proteolytic degradation along the secretory pathway (Margolin *et al.*, 2020), and thus the vacuole and ER were used as alternatives in case apoplast accumulation was inadequate. The accumulation of some proteins rises steadily with time, while others decrease with time. Therefore, a time-course was conducted to investigate the accumulation profile of each protein for one week after infiltration. This determined the best time to sample the leaf tissue for maximal accumulation of recombinant protein.

Following infiltration of ALS, Fer, and TMVc targeted to the three subcellular compartments and sampling on days 3, 5, and 7 post-infiltration, total soluble proteins were extracted from the leaf samples and examined using SDS-PAGE and Western blotting.





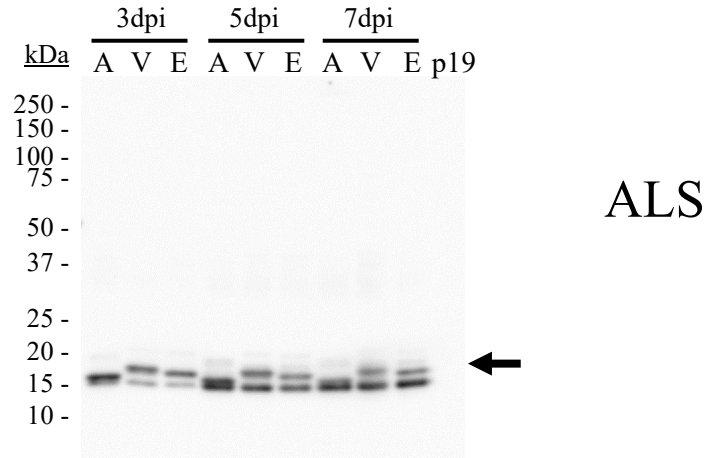
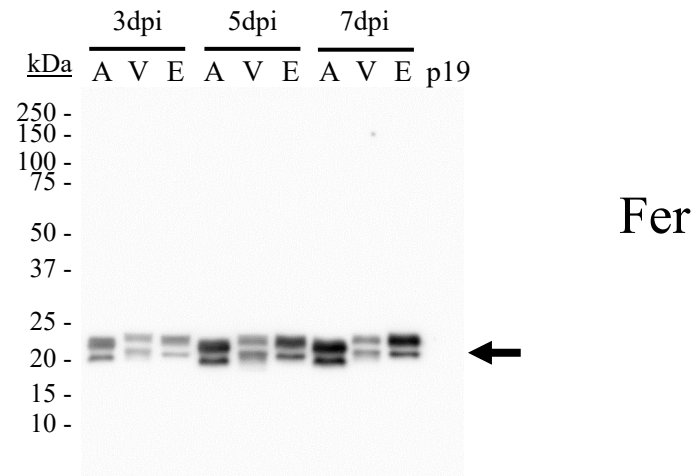
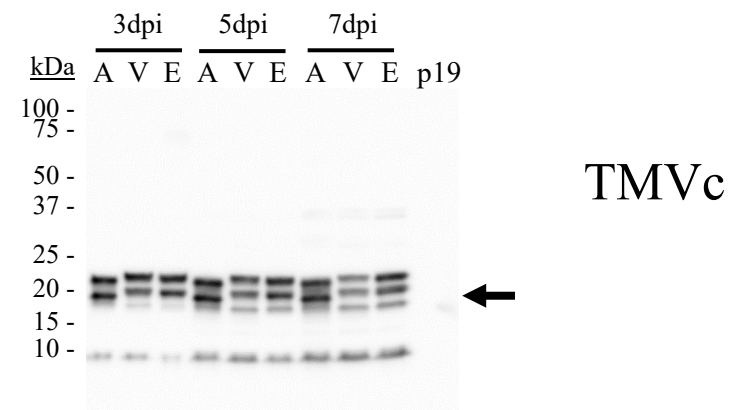
**Figure 11. Schematic for self-assembling proteins.** Schematic for ALS (purple), Fer (yellow), and TMVc (grey) for the self-assembling protein analyses. The C-terminal c-Myc tags (dark green) are encoded in the pCLGG-X plasmids, whereas the N-terminal HA tags (dark blue) were synthesized with the accompanying genes. Sequence lengths not to scale.

This experiment was repeated twice, and the blots of both infiltrations for each self-assembling protein appear very similar in their banding patterns and relative intensity of bands, therefore only the second replicate is shown (Figure 12).

The doublet of bands detected in each ALS and Fer lane (Figure 12a and b) can be explained by glycosylation at the putative glycosylation site identified by *in silico* analysis using the NetNGlyc 1.0 Server (<http://www.cbs.dtu.dk/services/NetNGlyc/>). The lower band for each is likely the unglycosylated monomer, and the upper band the mono-glycosylation of ALS and Fer. Expected sizes are in the figure legend. The triplet of bands observed for TMVc (Figure 12c) can also be explained by glycosylation as *in silico* analysis found two putative N-linked glycosylation sites on the protein sequence. The lower band is likely the monomer, with the upper bands being due to single and double glycosylation of the proteins. The APO-targeted TMVc appears to only have the upper two bands of the triplet present. This is likely the result of efficient glycosylation occurring and therefore no band for the unglycosylated monomer being detectible. TMVc also shows a smaller band around 10 kDa in each lane which is likely a cleavage degradation product of the monomer.

In each blot, the bands ran slightly lower in the APO targeted samples than in the ER and VAC samples (Figure 12). Proteins targeted to all three compartments contain a PR1b signal peptide which is cleaved once they enter the ER (Lund and Dunsmuir, 1992). While the APO-targeted proteins have no other targeting sequence, the ER-targeted proteins have an additional C-terminal KDEL for retrieval and retention in the ER which adds about 0.5 kDa to the protein size. The VAC-targeted proteins have a C-terminal propeptide for sequestration and retention within the vacuole (Vitale and Raikhel, 1999),

**Figure 12. ALS, Fer, and TMVc accumulate in all three subcellular compartments.** Proteins were targeted to the apoplast (A), vacuole (V), and ER (E) and were sampled on 3, 5, and 7 dpi. The p19 samples act as negative controls. Black arrows denote expected size of unglycosylated monomers. **a)** ALS. The expected size for the unglycosylated ALS monomer is 18.9 kDa. In each lane 4  $\mu$ l of soluble plant protein extract were loaded. **b)** Fer. The expected size for the unglycosylated Fer monomer is 21.4 kDa. In each lane 20  $\mu$ l of soluble plant protein extracts were loaded. **c)** TMVc. The expected size for the unglycosylated TMVc monomer is 19.8 kDa. In each lane 20  $\mu$ l of soluble plant protein extract were loaded. For all three blots, detection was performed using anti-HA primary antibodies at a 1:5000 dilution. Every 1  $\mu$ l of soluble plant protein extract contained 0.33 mg of plant tissue fresh weight. Each blot is representative of two transient expression experiments.

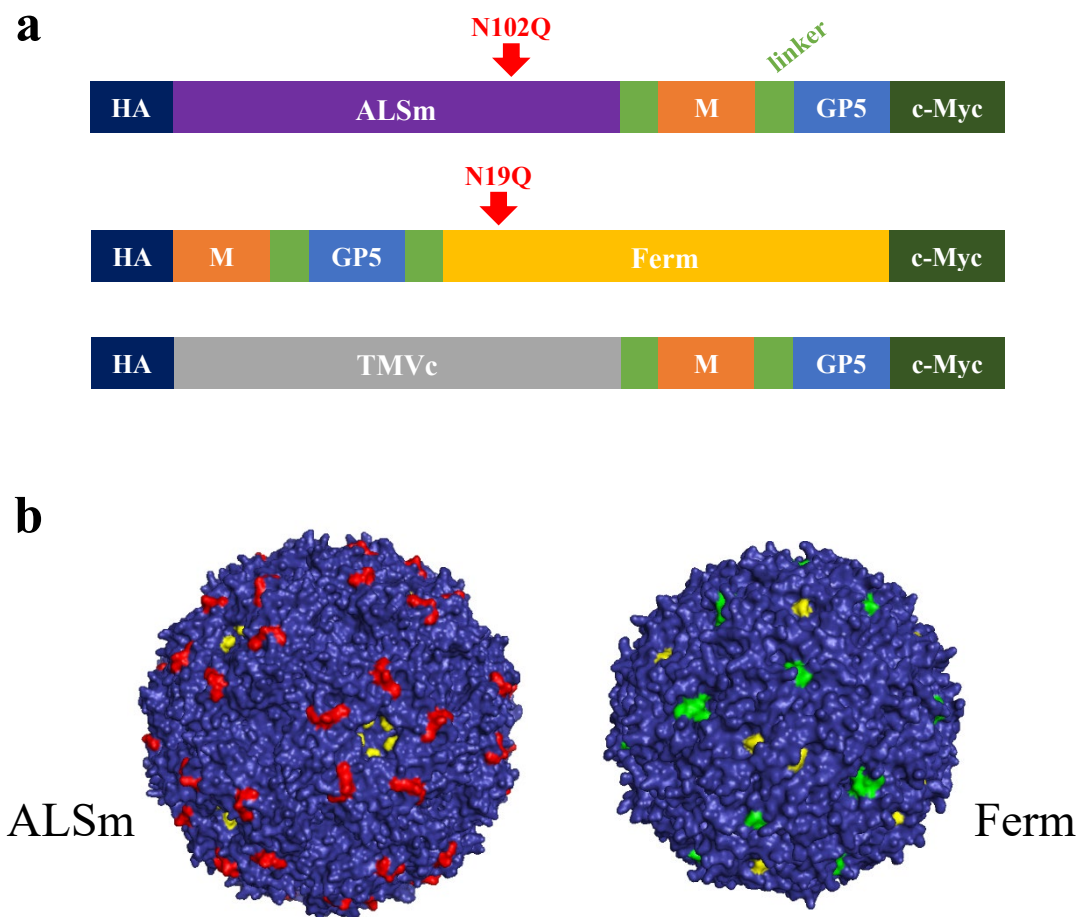
**a****b****c**

which adds about 0.9 kDa if not cleaved. Therefore, these extra targeting/retrieval peptides are likely responsible for the observed size difference between the APO, VAC, and ER samples.

While no quantification or data analysis was performed for these blots, visual inspection of band intensity indicated that sampling on 7 dpi and apoplast targeting are suitable for further analysis of each protein. Seven dpi was selected because its samples had similar or slightly higher protein accumulation and it provides more time for the proteins to possibly self-assemble into nanoparticle structures. Apoplast targeting allows for the N-linked glycan chains to most closely match those of the native PRRSV. As the accumulation of each APO targeted protein was similar or slightly higher than those in the ER and VAC on 7 dpi, it was also selected for further use.

### **3.4 Epitope fusion construct design**

Expression of the self-assembling proteins in plants found evidence that ALS and Fer were being glycosylated (Figure 12a and b), and *in silico* analysis showed that these proteins each contain one potential N-glycosylation site. Both ALS and Fer are bacterial in origin and therefore do not undergo glycosylation in their host organisms. Because of this, there was concern the glycosylation observed upon their expression in plants may hinder their ability to fold or assemble correctly. To avoid this, the asparagine residues predicted to be potential glycosylation sites on ALS (N102) and Fer (N19) were substituted with a glutamine (Q) residue to prevent N-linked glycosylation. The modified proteins were labelled ALSm and Ferm, respectively, and were used for the PRRSV epitope fusions (Figure 13). Data also indicates TMVc being glycosylated at both of its two potential N-linked glycosylation sites (Figure 12c). TMVc was not modified to prevent



**Figure 13. Self-assembling proteins fused to the M-GP5 epitope.** **a)** The self-assembling protein scaffolds are in purple (ALSm), yellow (Ferm), and grey (TMVc). The M (orange) and GP5 (light blue) epitope regions are separated from each other and the self-assembling proteins by (GGG)<sub>4</sub> flexible linkers (light green). The C-terminal c-Myc tags (dark green) are encoded in the pCLGG-X plasmids, whereas the coding sequences of the N-terminal HA tags (dark blue) were synthesized with the accompanying genes. Sequence lengths not to scale. Red arrows denote approximately where the ALSm and Ferm sequences were modified to avoid N-linked glycosylation. **b)** Assembled ALSm and Ferm nanoparticles. The surface exposed ALSm C-termini are in red, and the surface exposed Ferm N-termini are in green. The locations of each N→Q substitution for avoiding N-linked glycosylation are shown in yellow. Protein structures were produced in PyMOL.

glycosylation as the tobacco mosaic virus naturally replicates in plant systems, and therefore its glycosylation status was not of concern.

The previous work in *E. coli* (Harper, 2019) examined several different PRRSV epitopes fused to the *Brucella* lumazine synthase (BLS) self-assembling protein. The GP5 ectodomain on its own was tested first, which greatly decreased the protein solubility. To increase protein solubility, the M and GP5 ectodomains were fused. Both orientations of M and GP5 sequences were tested (ie. M-GP5-BLS and GP5-M-BLS). Only M-GP5-BLS resulted in increased the soluble accumulation, and therefore the M-GP5 orientation was used for this study (Figure 13a).

For the fusion proteins, the GP5 sequence I designed has a two amino acid substitution (N33A, N51A) as described by Ansari *et al.* (2006). They found glycosylation at these two sites appeared to shield GP5's antibody binding site from the immune system, and this modification increased its immunogenicity. The M and GP5 ectodomain sequences were separated by a flexible linker (GGG)<sub>4</sub> to allow folding over and interaction of these peptides, and possibly forming the disulfide bridge characteristic of the native viral proteins. This mimics the antigenic structure of the M and GP5 ectodomains together on the PRRS viral envelope. For the fusion proteins, the epitope was genetically fused to the C-terminus of the ALSm and TMVc monomers and the N-terminus of the Ferm monomer (Figure 13a) as previous work found that ALS and TMVc have a surface-exposed N-terminus and Fer has a surface-exposed C-terminus (Kanekiyo *et al.*, 2015; Min *et al.*, 2014; Pogue *et al.*, 2007). A second flexible (GGG)<sub>4</sub> linker was also added between each self-assembling protein and the epitope, allowing for flexibility of the epitope on the nanoparticle's surface. See Appendix 2 for full protein sequences.

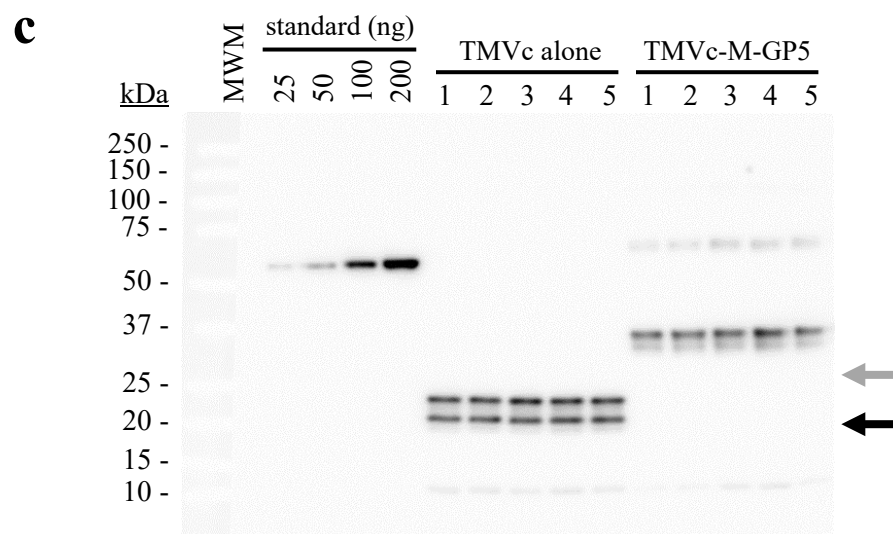
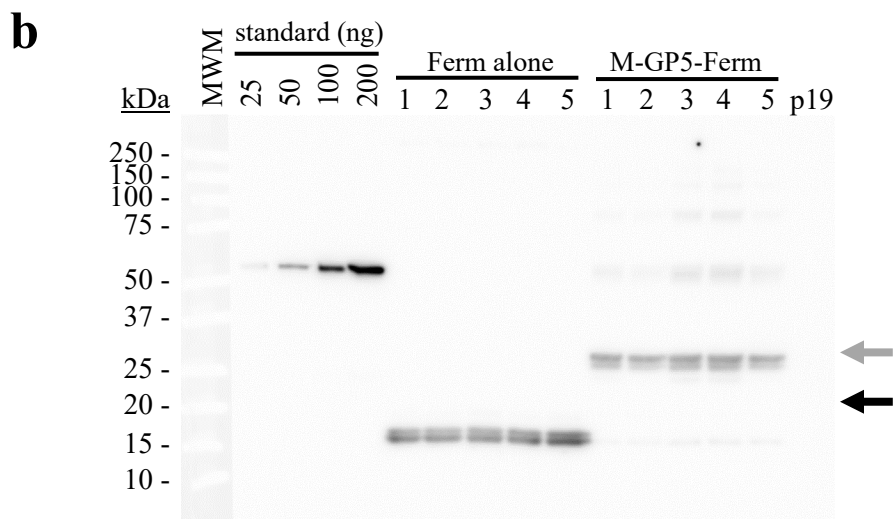
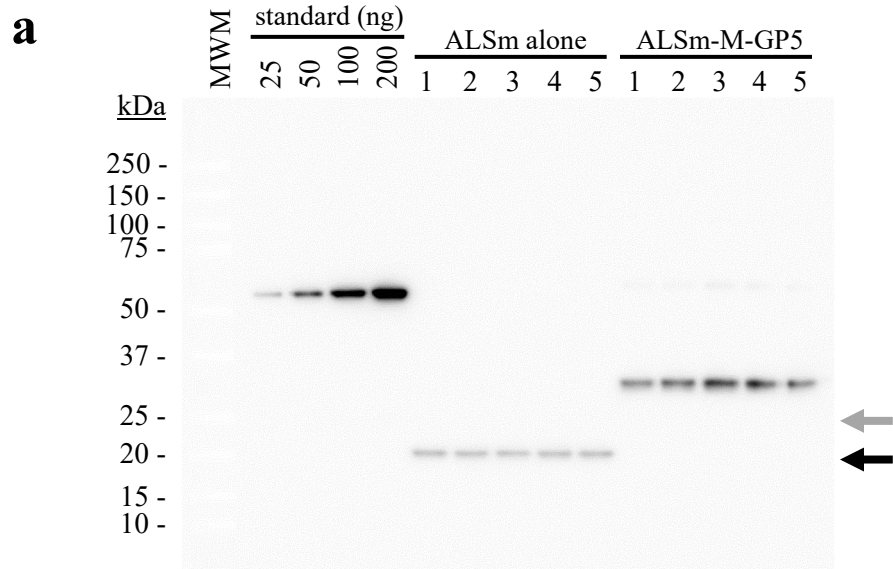
### 3.5 Expression analysis of fusion proteins

From the infiltrations examining the soluble accumulation of the self-assembling proteins on their own it was clear that all three proteins can be expressed in plants (Figure 12). Sampling for that experiment was performed by pooling the leaf discs from all five plants for each protein (Figure 6a) to observe general trends of soluble recombinant protein accumulation. For the protein fusions, sampling was done separately for each plant resulting in five replicates per protein for each experiment (Figure 6b), and this was repeated three times resulting in a total of 15 biological replicates per treatment. This increase in sample size allowed for better statistical analysis of the resulting soluble protein accumulation levels.

All three transient expression experiment replicates for each self-assembling protein and corresponding M-GP5 fusion produced similar results, and only one representative blot is shown in Figure 14 for each combination. All expected sizes are in the figure legend. The modified ALSm appears as a single band, confirming that the additional band observed in Figure 12 was indeed the result of glycosylation. Two bands were expected for both ALSm-M-GP5 and M-GP5-Ferm due to N44 glycosylation of GP5. For ALSm-M-GP5, only one band was observed approximately 33 kDa in size (Figure 14a). The presence of a single band is consistent with efficient glycosylation, resulting in no detectible band for the unglycosylated monomer. This is supported by the size of the band running larger than the expected unglycosylated monomer size of 24.9 kDa. For Ferm, two bands of almost identical size were observed with and without the epitope (Figure 14b). For the unfused Ferm, the size difference is too small for it to be the result of a glycan chain at an undetected glycosylation site. The upper Ferm band is



**Figure 14. Self-assembling proteins and their M-GP5 fusions accumulate well in plants.** All proteins were targeted to the apoplast and sampled on 7 dpi. The numbers across the top (1-5) identify the biological replicates. Black arrows show expected sizes of unglycosylated self-assembling protein monomers, and grey arrows show the expected sizes of the unglycosylated fusion protein monomers. Each N-linked glycan chain adds about 2.5 kDa to the protein size. **a)** ALSm and ALSm-M-GP5. The expected sizes for the unglycosylated ALSm and ALSm-M-GP5 monomers are 18.9 kDa and 24.9 kDa respectively. The larger bands observed for ALSm-M-GP5 may be due to dimeric assembly. One  $\mu$ l of soluble plant protein extract was loaded per ALSm lane, and 0.5  $\mu$ l was loaded per ALSm-M-GP5 lane. **b)** Ferm and M-GP5-Ferm. The expected sizes for the unglycosylated Ferm and M-GP5-Ferm monomers are 21.4 kDa and 27.4 kDa respectively. The multiple banding observed for M-GP5-Ferm may be due to assembly products. Ten  $\mu$ l of soluble plant protein extracts were loaded per lane. The p19 acts as a negative control. **c)** TMVc and TMVc-M-GP5. The expected sizes for the unglycosylated TMVc and TMVc-M-GP5 monomers are 19.8 kDa and 25.8 kDa respectively. The higher band observed for TMVc-M-GP5 is expected to be due to dimerization. Five  $\mu$ l of soluble plant protein extracts were loaded per lane. For all three blots, a protein standard with known quantities was used and detection was performed using anti-HA primary antibodies at a 1:5000 dilution. There was approximately 0.1 mg of plant tissue fresh weight in 1  $\mu$ l of protein extracts, however, tissue weights were recorded and considered for accumulation values later on. Therefore, each lane in these blots has a slightly different amount of plant tissue fresh weight even when the same volume of extract was loaded. Each blot is representative of three transient expression experiments.



likely the monomer, with the lower band possibly being the result of cleavage within the C-terminal c-Myc tag. Further experimentation showed only the upper band is visible when detected with an anti-c-Myc antibody (data not shown), supporting this interpretation of the result. The double bands for M-GP5-Ferm on the other hand is likely due to GP5 N44 glycosylation as it was still detectible with anti-c-Myc tag detection (data not shown).

TMVc was not modified to eliminate glycosylation sites, and therefore two bands were expected without the epitope present as observed previously for apoplast targeting (Figure 12c). With the M-GP5 epitope, a third band was expected to correspond with N44 glycosylation of GP5. TMVc on its own did appear as expected, however, only two bands were present for TMVc-M-GP5. Again, this could be the result of efficient glycosylation, and only the double and triple glycosylated forms of these proteins are detected. The efficient glycosylation of APO-targeted TMVc was previously observed (Figure 12c), supporting this possibility. Also, the TMVc-M-GP5 doublet ran over 5 kDa larger than its expected unglycosylated monomeric size, which is likely due to the addition of 2 and 3 N-linked glycan chains on these proteins. Every TMVc and TMVc-M-GP5 sample contains a lower band approximately 10 kDa in size (Figure 14c), coinciding with the degradation product observed previously for TMVc (Figure 12c).

Typically, the intensity of bands in a Western blot corresponds to the amount of protein present. When examining the ALSm fusion, it is clear that the bands are more faint without the epitope than for ALSm-M-GP5 (Figure 14a). The ALSm lanes of this blot were also loaded with double the volume of soluble plant protein extract than the ALSm-M-GP5 samples, making the greater soluble accumulation of ALSm-M-GP5 compared to ALSm abundantly clear.

For Ferm and TMVc, considering that consistent amounts of protein extracts were loaded into each of their lanes, it appears as though there may be slightly more band intensity and therefore protein accumulation for the self-assembling proteins alone compared to their M-GP5 fusions (Figure 14b and c). These differences in band intensity are not as pronounced as between ALSm and ALSm-M-GP5.

For all of the M-GP5 fusion proteins there are faint bands of larger sizes present (Figure 14), probably due assembly of protein dimers and multimers. For ALSm-M-GP5 and TMVc-M-GP5 only one extra band is present in each lane approximately twice the size of the lower bands. Therefore, these are likely protein dimers. There are several upper bands observed for M-GP5-Ferm, likely corresponding to multimeric assembly of this protein. Therefore, portions of the fusion proteins appeared to maintain some quaternary assembly despite exposure to denaturing and reducing conditions.

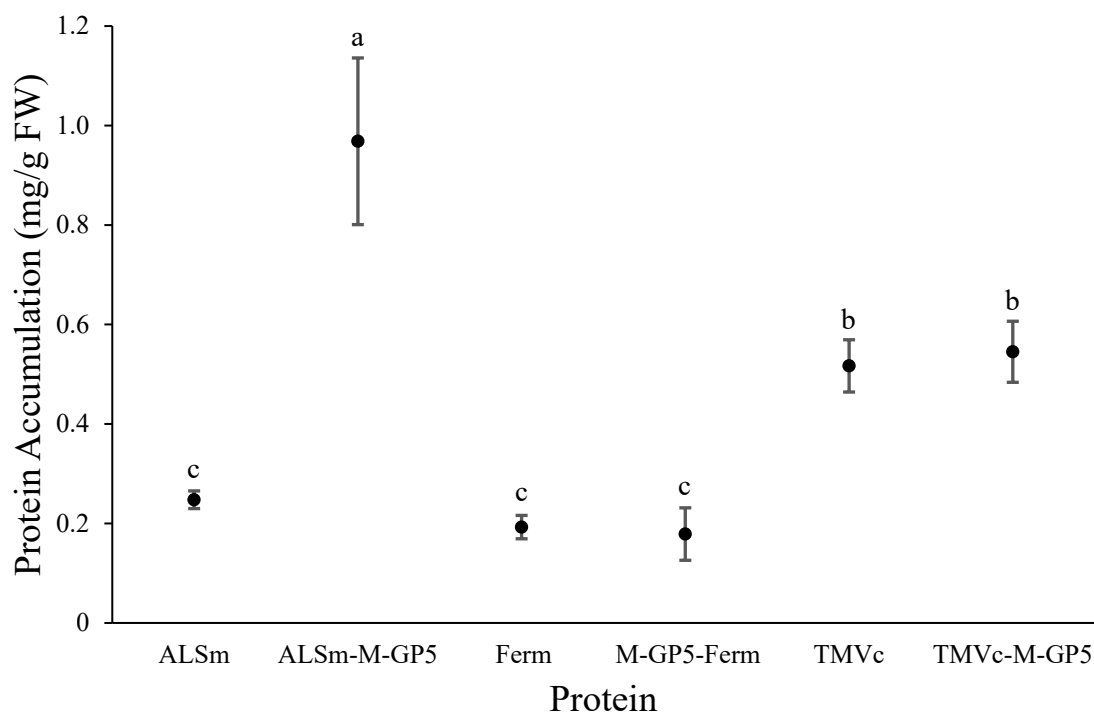
### **3.6 Quantification and statistical analysis of accumulation**

The soluble protein accumulation levels were acquired from quantifying the Western blots for each self-assembling protein and their M-GP5 fusion. The intensity of the bands shown in Figure 14, as well as of the other two replicates were quantified against a standard curve of known concentrations of a synthetic protein. The soluble protein accumulation levels of the 15 replicates were then converted into mg of protein per gram of fresh plant weight by considering the amount of plant protein extraction buffer used, the mass of plant tissue per sample, and dilutions of samples loaded on gels (i.e. volume of extract loaded per lane).

The statistical analysis was conducted using R. No major concerns were found upon initial exploration of the data such as zero inflation, outliers, or collinearity between variables. A GLS model was applied to the data because it does not assume equal variance. This assumption is characteristic of linear models, however, validation steps of the data found there to be unequal variance present between the proteins. The model equation best suited for the data (evaluated by corrected Akaike's Information Criterion values) examined soluble protein accumulation as a function of protein, transient expression experiment number (1-3), and their interaction. A variance of identity residual structure was also used allowing separate variances for each protein. Further validation found no concerns with this model. See Appendix 4 for full post hoc test results.

The statistical analysis of these results (Figure 15) further confirmed the visual observation of the blots. ALSm-M-GP5 had an average soluble protein accumulation of 0.968 mg/g (s.e. 0.077), which was found to be significantly higher than the accumulation for ALSm of 0.248 mg/g (s.e. 0.008) (p-value =  $5.6 \times 10^{-6}$ ). Therefore, the addition of the M-GP5 epitope onto ALSm clearly increases its soluble protein accumulation upon plant-expression.

For Ferm, the quantification and statistical analysis (Figure 15) shows the average soluble protein accumulation was 0.193 mg/g (s.e. 0.011), which was not statistically different from 0.179 mg/g for M-GP5-fer.m (s.e. 0.024) (p-value = 0.9961). For TMVc, the average accumulation was 0.517 mg/g (s.e. 0.024), which was not significantly different than 0.545 mg/g for TMVc-M-GP5 (s.e. 0.028) (p-value = 0.9936) (Figure 15). Therefore, the addition of the M-GP5 epitope onto both Ferm and TMVc does not appear to alter their soluble accumulation upon plant expression.



**Figure 15. Recombinant soluble protein accumulation of ALSm, Ferm, TMVc, and their M-GP5 fusion proteins.** The points denote mean values of 15 biological replicates over three independent experiments (five biological replicate per experiment), with bars above and below representing their 95% confidence intervals. Letters represent significance groupings (p-value < 0.05). All samples were targeted to the apoplast and sampled on 7 dpi. Protein amounts were quantified from blots by GelQuant software using the standard of known concentration as a reference. Protein accumulation levels were calculated based on plant tissue mass, amount of soluble plant protein extraction buffer used, and amount of extracts loaded such that accumulation would be represented as mg of recombinant protein per gram of plant tissue fresh weight (FW). The data from all 15 biological replicates were compiled and examined in R, where statistical analysis was performed using a GLS model that allowed for separate variance per protein. Significance was examined by ANOVA, followed by a Games-Howell post-hoc test.

## 4 DISCUSSION

### 4.1 Protein nanoparticles

Nanoparticle-based vaccines provide safe and effective alternatives to traditional vaccine approaches. Unlike live attenuated vaccines, self-assembling protein nanoparticles are unable to replicate or revert to virulence as they lack genetic material (Francis, 2018; Vetter *et al.*, 2018). Nanoparticle-based vaccines are also efficient at producing strong and long lasting immune responses, a benefit they have over many traditional subunit or killed virus vaccine methods (Francis, 2018; Vetter *et al.*, 2018). Because of these discoveries, more research has been allocated to studying different nanoparticles, like protein nanocages and VLPs, and their use in vaccine development.

Some recombinantly produced VLPs of pathogenic viruses are now commercially available as vaccines. Two vaccines available for HPV (Gardasil<sup>®</sup> and Cervarix<sup>®</sup>) are prime examples of this, where the recombinant expression of a major HPV capsid protein (L1) results in self-assembling VLPs able to induce protection against HPV infection in humans (Wang and Roden, 2013). A multivalent influenza VLP produced in plants by Medicago Inc. (Quebec, QC) has also recently passed phase 3 clinical trials and is awaiting approval by Health Canada (Ward *et al.*, 2020). This vaccine uses co-expression of influenza H5N1 and H1N1 hemagglutinin to produce enveloped VLPs able to induce complete protection from both homologous and heterologous virus challenges (D'Aoust *et al.*, 2008; Ward *et al.*, 2020). For other pathogens, different methods such as attaching epitopes containing neutralizing antibody binding sites from the pathogen onto protein nanoparticles allows for the production of effective vaccines against these diseases (Francis, 2018).

There are different methods of attaching epitopes to protein nanoparticles. One method is genetic fusion, and involves adding the epitope's nucleotide sequence within or at one end of the self-assembling protein's DNA sequence so both proteins are translated into one uninterrupted amino acid sequence (Brune and Howarth, 2018). Chemical conjugation of an epitope to a nanoparticle surface is also possible. This involves production of the nanoparticle and epitope separately then, through the use of chemical reactions, the proteins are connected by a cross-linker (Brune and Howarth, 2018). SpyCatcher-SpyTag is another method of epitope attachment using two domains of a modified *Streptococcus pyogenes* protein, termed SpyCatcher and SpyTag (Hatlem *et al.*, 2019). One domain is genetically fused to the nanoparticle and the other to the epitope. When the two domains interact they form a covalent bond, creating a stable connection between the two proteins (Hatlem *et al.*, 2019). However, chemical conjugation and SpyCatcher-SpyTag technology require production and purification of both the nanoparticle and epitope separately before combining them and allowing fusion to occur (Brune and Howarth, 2018; Hatlem *et al.*, 2019). Keeping costs low is essential for producing veterinary vaccines (Schillberg and Finner, 2021) and genetic fusion, as performed in this study, is the cheapest method for these purposes as it avoids purification and allows direct oral administration using leaves expressing the fused nanoparticles (Chan and Daniell, 2015; Topp *et al.*, 2016).

In this project ferritin from *H. pylori* (Kanekiyo *et al.*, 2015), a bacteria found in the human stomach, lumazine synthase from *A. aeolicus* (Min *et al.*, 2014), an underwater hyperthermophilic bacterium, and the tobacco mosaic virus coat protein (Brown *et al.*, 2013) were selected for study. Recombinant Fer produced in bacteria (Harper, 2019;



Kanekiyo *et al.*, 2015), insect cell cultures (Qu *et al.*, 2020), and mammalian cell cultures (Kanekiyo *et al.*, 2013) accumulated at relatively high levels, assembled into nanoparticle structures, and tolerated the addition of foreign peptides on the particles' outer surface. Recombinant expression of ALS in bacteria resulted in varying levels of solubility depending on the study, with high solubility reported by Min *et al.* (2014) and low solubility reported by Harper (2019). Mammalian expression of ALS found the wild-type protein fused to a small peptide was mainly insoluble (Zhang *et al.*, 2020). The current study found that soluble ALS and Fer accumulate in the plant secretory pathway, although insoluble recombinant protein accumulation was not investigated.

The tobacco mosaic virus has been studied in a variety of expression systems. Expression of wild-type TMVc in bacteria, yeast, and plants did not produce expected nanorod structures when RNA was absent (Kadri *et al.*, 2013; Lomonossoff and Wege, 2018; Thuenemann *et al.*, 2021). This is because the interaction of RNA with TMVc is required for encapsulation and assembly of the TMV nanorod particle. The modified TMVc sequence used in this project contained a two amino acid substitution (E50Q and D77N). Brown *et al.* (2013) found that when produced in *E. coli*, the modified TMVc is soluble and self-assemble into stable nanorod structures in the absence of RNA. Such assembled nanorods of modified TMVc are capable of displaying foreign peptides on their outer surface (Brown *et al.*, 2013). The current study found high soluble accumulation of this modified TMVc in plants. Further, the modified TMVc accumulated to significantly higher levels than ALSm and Fer alone.

In this study, Fer and TMVc showed no significant change in soluble protein accumulation when fused to the M-GP5 epitope, while ALSm-M-GP5 accumulated

significantly higher than ALSm. In contrast, previous work in *E. coli* examining ALS and Fer fused to the GP5 ectodomain from the PRRSV found the addition of this epitope caused the recombinant proteins to become completely insoluble (Harper, 2019). The improved solubility of fusion proteins in this study compared to *E. coli* may be due to the M and GP5 ectodomain fusion, which was shown to increase solubility compared to the GP5 ectodomain alone (Harper, 2019). Another possible cause for the increased solubility is the presence of N-linked glycans introduced in the plant secretory pathway, as glycosylation of certain proteins was shown to increase their stability (Solá and Griebenow, 2009). This explanation is supported by a previous study where the introduction of new glycosylation sites improved the soluble accumulation of ALS and Fer fused to foreign peptides expressed in mammalian cells (Zhang *et al.* 2020). Other factors contributing to the higher fusion protein solubility in *N. benthamiana* compared to *E. coli* could be disulfide bond formation and ER chaperones present in plants, both of which were shown to aid in protein folding and stability (Robinson and Bulleid, 2020; Strasser, 2018).

## **4.2 Livestock vaccines**

Viral diseases are one of the major hurdles facing the livestock industry, especially because viruses are not controlled by antibiotics. Vaccines provide an effective method for controlling the spread of these diseases. While there are many ways to design and produce vaccines, keeping production costs low is essential for veterinary medicine (Schillberg and Finnern, 2021) as livestock farmers are unlikely to purchase expensive medications for their animals (MacDonald *et al.*, 2015). High accumulation levels of recombinant proteins are a major factor in limiting the production costs. For industrial production of pharmaceutical proteins in plants for human use, an unofficial threshold of 1% of total

soluble protein, or approximately 0.1 mg/g of recombinant protein per plant fresh weight, is used to evaluate the economic viability (Rybicki, 2009). Next to accumulation levels, working with insoluble proteins requires more laborious and costly unfolding and refolding steps. Therefore using soluble proteins, which tend to be correctly folded, is preferred.

Plant-produced vaccines have the potential to be administered orally to livestock, which provides several benefits over injection. One benefit is avoiding the need to purify the recombinant proteins because the plant tissue can be lyophilized and fed to the animals (Chan and Daniell, 2015; Topp *et al.*, 2016). This greatly reduces purification costs (Schillberg and Finnern, 2021). A second benefit is the technical ease of oral administration compared to other methods when considering an agricultural setting. The lyophilized plant tissue could be mixed into the livestock's feed, as opposed to injection of the vaccine by a veterinarian (Chan and Daniell, 2015). A third benefit of oral administration is the direct delivery of the vaccine to the mucosal tissue of the intestines, improving the resulting IgA titres (Topp *et al.*, 2016). In contrast, injections tend to produce strong IgG response in the blood and a weaker IgA response. IgA antibodies are present in mucosal tissues such as the respiratory, digestive, and reproductive tracts. Because PRRSV infects at mucosal tissues (Pileri and Mateu, 2016), IgA antibodies are better suited for interrupting initial infection of the pigs (Kolotilin *et al.*, 2014; Topp *et al.*, 2016).

The development of many plant-produced vaccines against livestock diseases have shown promising results in recent years. One example is the production of a multimeric H5 hemagglutinin in *N. benthamiana* to vaccinate chickens against H5N1 avian influenza (Phan *et al.*, 2020). High protective immunity was conferred against H5N1 challenge to chickens injected with crude plant extract containing recombinant proteins mixed with an

adjuvant. Alfano *et al.* (2015) demonstrated the transplastomic production of BLS fused to a bovine rotavirus surface protein in tobacco. This fusion protein was both soluble and stable, and injection of soluble plant extracts with an adjuvant into hens induced high titres of neutralizing antibodies against the rotavirus. Alfano *et al.* (2015) did not examine the assembly of BLS, however, studies have shown BLS to form decameric nanoparticles (Harper, 2019). Therefore, plants have been demonstrated to be an effective platform for the production of recombinant livestock vaccines.

Recent work in the Menassa lab has been investigating transient expression of BLS in plants as a scaffold for the M-GP5 PRRSV epitope. This work has found high soluble accumulation of BLS on its own expressed in *N. benthamiana*, and similar results for a BLS and M-GP5 fusion (Hong Zhu and Rima Menassa, personal communications). The self-assembly of these fusion proteins into stable, decameric nanoparticles similar to BLS alone has also been confirmed using size-exclusion chromatography and transmission electron microscopy (Hong Zhu and Rima Menassa, personal communications). Because of these promising results, I wanted to examine the plant-production of some larger protein nanoparticles previously studied in bacteria.

The current study aimed to investigate the plant-production of a PRRSV M-GP5 epitope genetically fused to three larger protein nanoparticles. I found high soluble accumulation of ALSm-M-GP5, M-GP5-Ferm, and TMVc-M-GP5 upon plant-production. ALSm-M-GP5 had soluble accumulation almost 10x the 0.1 mg/g unofficial industry threshold (Rybicki, 2009), accumulating more than any other protein examined in this study. While M-GP5-Ferm and TMVc-M-GP5 had lower soluble accumulations, they were still over 2x and 5x higher than the industry threshold, respectively. Therefore, all three of

these fusion proteins should be investigated further for production of a vaccine against PRRS, with ALSm-M-GP5 being the best candidate at this time.

### **4.3 Protein mobility and assembly in gels**

Protein migration through gels is impacted by their molecular weight, charge, and shape. The latter two are standardized during an SDS-PAGE through the use of SDS and both denaturing and reducing conditions, resulting in the proteins being separated by only molecular weight. Some of the proteins in this study ran at slightly different sizes than expected, especially ALS targeted to the three subcellular compartments of the secretory pathway. There are several explanations for this, including disproportional SDS binding to the proteins. SDS masks the charge of proteins and is expected to bind proteins uniformly. However, studies have found that some proteins differ in their SDS binding capacities, altering mobility through the gel (Rath *et al.*, 2009; Shi *et al.*, 2012). Another possibility is based on the shape of the proteins. The samples are denatured so proteins do not maintain any of their native conformation. However, some proteins may not fully denature and consequently maintain globular structures, thereby affecting their gel mobility.

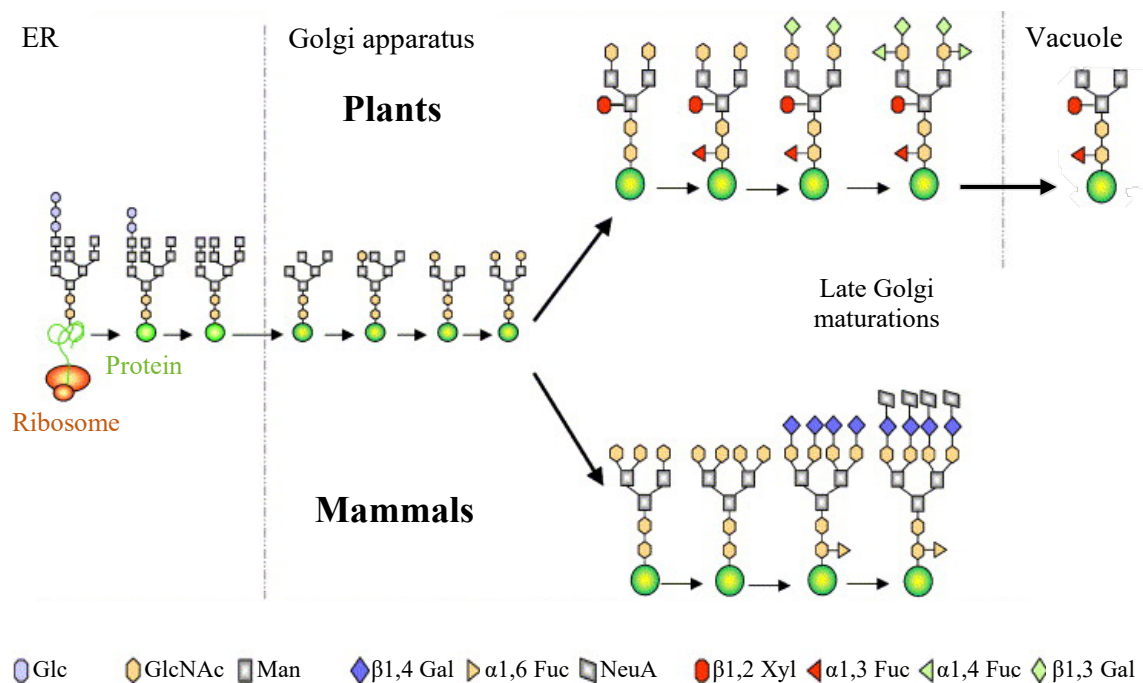
The SDS-PAGE and sample preparation exposed the samples to denaturing and reducing conditions that should have also separated any assembled proteins into monomers. For all of the M-GP5 fusion proteins there are faint bands of larger sizes present, likely due assembly of protein dimers and multimers. One explanation for this could be that not all the proteins are fully denatured, resulting in some remaining joined to one or more other subunits. Another possible cause is disulfide bond formation among the M and GP5 portions of the epitope between subunits. Disulfide bonds are formed in the ER (Robinson and Bulleid, 2020), the first organelle of the secretory pathway that ends

with secretion of proteins to the apoplast. Denaturation alone does not break disulfide bonds, but reducing agents do. While there was a reducing agent (DTT) present in the SDS-PAGE sample loading dye, it is possible some of these bonds may have remained intact. Inter-subunit disulfide bonds would also explain why the assembly products were only observed for the fusion proteins as ALSm, Ferm, and TMVc on their own lack the M-GP5 epitope where this bond would occur.

#### **4.4 N-linked glycosylation**

A major benefit of using a plant expression system is the variety of post-translational modifications they can add to proteins. One modification of particular interest, especially for this project, is N-linked glycosylation. Recombinant proteins targeted to certain plant subcellular compartments will carry different compositions of their N-linked glycan chains (Figure 16). The initial steps of N-linked glycosylation occur in the ER, where glycan chains are co-translationally joined to proteins at asparagine residues within NXS and NXT amino acid motifs (where  $X \neq P$ ) (Gomord *et al.*, 2010; Strasser *et al.*, 2021). After enzymatic modification in the ER, the N-linked glycan chains are mainly composed of mannose sugar residues (Figure 16).

Plant proteins that pass from the ER to the Golgi apparatus before secretion to the apoplast receive further N-linked glycan chain modifications. Most of the terminal mannose residues are replaced with more complex sugars (Figure 16) (Gomord *et al.*, 2010; Margolin *et al.*, 2020). Instead of being secreted, certain proteins that pass through the ER and Golgi enter the vacuole. Passage into the vacuole results in the enzymatic removal of many peripheral sugar residues, producing smaller glycan chains with some complex sugar residues from the late Golgi (Strasser *et al.*, 2021) (Figure 16). Plant and mammalian cells



**Figure 16. N-linked glycan chains from plants and mammals.** An initial N-linked glycan chain is enzymatically added co-translationally to proteins in the ER and is composed of N-acetylglucosamine (GlcNAc: yellow hexagon), mannose (Man: grey square), and glucose residues (Glc: violet circle). Passage through the ER results in removal of the terminal Glc residues. Subsequent passage of the glycoprotein into the early Golgi apparatus results in further modification at the periphery of the glycan chain. Modification of the N-linked glycan chain then deviates between plants and mammals in the late Golgi. In mammals sialic acid (NeuAc: grey rhombus),  $\alpha$ 1,6 fucose (Fuc: yellow triangle), and  $\beta$ 1,4 galactose (Gal: blue rhombus) residues are added. In plants  $\beta$ 1,2 xylose (Xyl: red circle),  $\alpha$ 1,3 fucose (red triangle),  $\alpha$ 1,4 fucose (green triangle), and  $\beta$ 1,3 galactose (green rhombus) residues are added. Plant proteins targeted to the vacuole receive further processing of their N-linked glycan chains, where the terminal Gal, Fuc, and GlcNAc residues are removed. Figure modified from Faye *et al.* (2005).

both allow for the complex processing of proteins, which includes their ability to add N-linked glycans to proteins. The initial steps of adding and modifying these N-linked glycan chains in the ER are identical (Gomord *et al.*, 2010; Margolin *et al.*, 2020). These modifications diverge, however, as the proteins pass through the late Golgi (Figure 16). In plants, these modifications involve a  $\beta$ 1,2 xylose and both  $\alpha$ 1,3 and  $\alpha$ 1,4 fucose residues (Strasser *et al.*, 2021). In mammals, the larger glycan chains include an  $\alpha$ 1,6 fucose and sialic acid residues (Margolin *et al.*, 2020).

The PRRSV infects and replicates within pig cells, therefore its proteins are subject to mammalian post-translational modifications. During replication, the viral envelope proteins are co-translationally inserted into the ER membrane where they acquire the initial stages of N-linked glycosylation. The immature PRRSV particles pass through the ER and Golgi to exit the cell, resulting in the mature PRRSV having the complex N-linked glycan chains characteristic of the mammalian late Golgi (Figure 16) (Li *et al.*, 2015).

Recent advances have been made to knock out plant specific glycosylation enzymes such as beta1,2 xylosyl transferase and alpha 1,3 fucosyl transferase, and to introduce mammalian-specific glycans such as terminal sialic acid, allowing for the production of recombinant proteins in plants with the N-linked glycan patterns characteristic of mammals (Margolin *et al.*, 2020). In wild-type plants, the complex N-linked glycan chains on apoplast-targeted proteins most closely match those from the mammalian late Golgi (Margolin *et al.*, 2020; Strasser *et al.*, 2021). In this study, the plant-produced recombinant proteins appear to be receiving glycosylation. Therefore, the apoplast targeted fusion proteins are likely acquiring complex N-linked glycans on the M-GP5 epitope most similar to those of the native PRRSV.



## 5 CONCLUSIONS AND FUTURE STUDIES

This study was the first to demonstrate that recombinant ALS, Fer, and the modified TMVc can be expressed in plants. Addition of the M-GP5 epitope had little effect on the soluble accumulation of both TMVc and Ferm, resulting in insignificant changes to their soluble protein accumulation levels. In contrast, the ALSm-M-GP5 fusion had soluble accumulation levels significantly higher than any other protein analyzed in this study. Because of this, I propose that ALSm-M-GP5 is the best candidate to advance to future research.

Further study is required to investigate whether plant-produced ALSm, Ferm, and the modified TMVc assemble into their native quaternary structures with and without the M-GP5 epitope fused. This can be examined using methods such as transmission electron microscopy, immunogold localization, and sedimentation velocity studies (Boigard *et al.*, 2017; Harper, 2019). Correct assembly of protein nanoparticles and surface display of M-GP5 epitopes is essential for multivalent presentation to the immune system. As such, determining the quaternary structure of these proteins produced in plants for the first time will be an essential step in advancing this PRRS vaccine candidate prior to animal testing. If producing these proteins in certain plant subcellular compartments inhibits their assembly or epitope display, they lose the major benefits of nanoparticle-based vaccines regardless of their accumulation levels.

Following confirmation of nanoparticle assembly and epitope display, animal trials must be conducted to examine the immunogenicity of the vaccine candidate(s). Initially investigating the immune response in smaller animal models, such as mice, provides information on the resulting antibody response (Boigard *et al.*, 2017), especially IgG and

IgA levels. It would also be important to examine possible routes of administration, such as oral and parenteral, and whether adjuvants are required. Adjuvants are often administered in conjunction with vaccines to improve their recognition and interaction with the immune system (Lemoine *et al.*, 2020). Nanoparticle-based vaccines show improved interaction with the immune system due to their size and display of many epitope copies, therefore adjuvants may not be required (López-Sagaseta *et al.*, 2016; Zhao *et al.*, 2014). For injected vaccine trials, serum samples can be used to examine whether the immune response is against the M-GP5 epitope and/or the nanoparticle scaffold. For oral administration trials, antibodies from the animals' feces and intestines can be examined. The resulting antibodies must recognise the M-GP5 epitope for any immunity to be conferred against the PRRSV. The production of neutralizing antibodies can be explored in virus neutralization assays, and protection from infection can be explored through challenge experiments with the PRRSV in pigs following treatment with the vaccine candidate(s), providing information on whether they provide protective immunity (Li *et al.*, 2021; Zuckermann *et al.*, 2007).

While there is much work yet to be completed before any of the vaccine candidates produced here can be used in the fight against PRRS, this work provides a promising foundation for their further study and commercialization. It also adds to the ever-growing evidence that producing protein pharmaceuticals in plants is an innovative and viable alternative to more traditional approaches.

## 6 REFERENCES

- Alfano, E.F., Lentz, E.M., Bellido, D., Dus Santos, M.J., Goldbaum, F.A., Wigdorovitz, A. and Bravo-Almonacid, F.F. (2015) Expression of the multimeric and highly immunogenic *Brucella* spp. lumazine synthase fused to bovine rotavirus VP8d as a scaffold for antigen production in tobacco chloroplasts. *Front Plant Sci* **6**, 1170-1170.
- Ansari, I.H., Kwon, B., Osorio, F.A. and Pattnaik, A.K. (2006) Influence of N-linked glycosylation of porcine reproductive and respiratory syndrome virus GP5 on virus infectivity, antigenicity, and ability to induce neutralizing antibodies. *J Virol* **80**, 3994-4004.
- Bacher, A. and Mailänder, B. (1978) Biosynthesis of riboflavin in *Bacillus subtilis*: function and genetic control of the riboflavin synthase complex. *J Bacteriol* **134**, 476-482.
- Bachmann, M.F. and Jennings, G.T. (2010) Vaccine delivery: a matter of size, geometry, kinetics and molecular patterns. *Nat Rev Immunol* **10**, 787-796.
- Barbosa Viana, A.A., Pelegrini, P.B. and Grossi-de-Sá, M.F. (2012) Plant biofarming: novel insights for peptide expression in heterologous systems. *Biopolymers* **98**, 416-427.
- Bartlett, B.L., Pellicane, A.J. and Tyring, S.K. (2009) Vaccine immunology. *Dermatol Ther* **22**, 104-109.
- Bevan, M., Barnes, W.M. and Chilton, M.-D. (1983) Structure and transcription of the nopaline synthase gene region of T-DNA. *Nucleic Acids Res* **11**, 369-385.
- Bhaskar, S. and Lim, S. (2017) Engineering protein nanocages as carriers for biomedical applications. *NPG Asia Mater* **9**, e371.
- Boigard, H., Alimova, A., Martin, G.R., Katz, A., Gottlieb, P. and Galarza, J.M. (2017) Zika virus-like particle (VLP) based vaccine. *PLoS Negl Trop Dis* **11**, e0005608.
- Brisse, M., Vrba, S.M., Kirk, N., Liang, Y. and Ly, H. (2020) Emerging concepts and technologies in vaccine development. *Front Immunol* **11**, 583077.
- Brown, A.D., Naves, L., Wang, X., Ghodssi, R. and Culver, J.N. (2013) Carboxylate-directed *in vivo* assembly of virus-like nanorods and tubes for the display of functional peptides and residues. *Biomacromolecules* **14**, 3123-3129.
- Brune, K.D. and Howarth, M. (2018) New routes and opportunities for modular construction of particulate vaccines: stick, click, and glue. *Front Immunol* **9**, 1432.

- Chan, H.-T. and Daniell, H. (2015) Plant-made oral vaccines against human infectious diseases - are we there yet? *Plant Biotechnol J* **13**, 1056-1070.
- Charerntantanakul, W. (2012) Porcine reproductive and respiratory syndrome virus vaccines: immunogenicity, efficacy and safety aspects. *World J Virol* **1**, 23-30.
- Cho, K.J., Shin, H.J., Lee, J.H., Kim, K.J., Park, S.S., Lee, Y., Lee, C., Park, S.S. and Kim, K.H. (2009) The crystal structure of ferritin from *Helicobacter pylori* reveals unusual conformational changes for iron uptake. *J Mol Biol* **390**, 83-98.
- Conley, A.J., Joensuu, J.J., Jevnikar, A.M., Menassa, R. and Brandle, J.E. (2009) Optimization of elastin-like polypeptide fusions for expression and purification of recombinant proteins in plants. *Biotechnol Bioeng* **103**, 562-573.
- Cutt, J.R., Dixon, D.C., Carr, J.P. and Klessig, D.F. (1988) Isolation and nucleotide sequence of cDNA clones for the pathogenesis-related proteins PR1a, PR1b and PR1c of *Nicotiana tabacum* cv. Xanthi nc induced by TMV infection. *Nucleic Acids Res* **16**, 9861-9861.
- D'Aoust, M.A., Lavoie, P.O., Couture, M.M., Trepanier, S., Guay, J.M., Dargis, M., Mongrand, S., Landry, N., Ward, B.J. and Vezina, L.P. (2008) Influenza virus-like particles produced by transient expression in *Nicotiana benthamiana* induce a protective immune response against a lethal viral challenge in mice. *Plant Biotechnol J* **6**, 930-940.
- Das, P.B., Dinh, P.X., Ansari, I.H., de Lima, M., Osorio, F.A. and 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.
- Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huberk, R., Feldman, R.A., Short, J.M., Olsen, G.J. and Swanson, R.V. (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**, 352-358.
- Diaz, D., Care, A. and Sunna, A. (2018) Bioengineering strategies for protein-based nanoparticles. *Genes (Basel)* **9**, 370.
- Duan, X., Nauwynck, H.J. and Pensaert, M.B. (1997) Effects of origin and state of differentiation and activation of monocytes/macrophages on their susceptibility to porcine reproductive and respiratory syndrome virus (PRRSV). *Arch Virol* **142**, 2483-2497.
- Durham, A.C.H., Finch, J.T. and Klug, A. (1971) States of aggregation of tobacco mosaic virus protein. *Nat New Biol* **229**, 37-42.

- Eclercy, J., Renson, P., Lebret, A., Hirchaud, E., Normand, V., Andraud, M., Paboeuf, F., Blanchard, Y., Rose, N. and Bourry, O. (2019) A field recombinant strain derived from two type 1 porcine reproductive and respiratory syndrome virus (PRRSV-1) modified live vaccines shows increased viremia and transmission in SPF pigs. *Viruses* **11**, 296.
- Faye, L., Boulaflous, A., Benchabane, M., Gomord, V. and Michaud, D. (2005) Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. *Vaccine* **23**, 1770-1778.
- Francis, M.J. (2018) Recent advances in vaccine technologies. *Vet Clin North Am Small Anim Pract* **48**, 231-241.
- Gnanasambandam, A., Polkinghorne, I.G. and Birch, R.G. (2007) Heterologous signals allow efficient targeting of a nuclear-encoded fusion protein to plastids and endoplasmic reticulum in diverse plant species. *Plant Biotechnol J* **5**, 290-296.
- Gomord, V., Fitchette, A.C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D. and Faye, L. (2010) Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol J* **8**, 564-587.
- Gonin, P., Pirzadeh, B., Gagnon, C.A. and Dea, S. (1999) Seroneutralization of porcine reproductive and respiratory syndrome virus correlates with antibody response to the GP5 major envelope glycoprotein. *J Vet Diagn Invest* **11**, 20-26.
- Guo, J., Zhou, A., Sun, X., Sha, W., Ai, K., Pan, G., Zhou, C., Zhou, H., Cong, H. and He, S. (2019) Immunogenicity of a virus-like-particle vaccine containing multiple antigenic epitopes of *Toxoplasma gondii* against acute and chronic toxoplasmosis in mice. *Front Immunol* **10**, 592.
- Harper, O.H. (2019) *Engineering self-assembling proteins to produce a safe and effective vaccine for porcine reproductive and respiratory syndrome*. M.Sc. Thesis, Department of Biology, University of Western Ontario, London, Ontario, Canada.
- Hatlem, D., Trunk, T., Linke, D. and Leo, J.C. (2019) Catching a SPY: Using the SpyCatcher-SpyTag and related systems for labeling and localizing bacterial proteins. *Int J Mol Sci* **20**, 2129.
- Haynes, J.R., Cunningham, J., von Seefried, A., Lennick, M., Garvin, R.T. and Shen, S.H. (1986) Development of a genetically-engineered, candidate polio vaccine employing the self-assembling properties of the tobacco mosaic virus coat protein. *Biotechnology (N Y)* **4**, 637-641.

- Holtkamp, D.J., Kliebenstein, J.B., Neumann, E.J., Zimmerman, J.J., Rotto, H.F., Yoder, T.K., Wang, C., Yeske, P.E., Mowrer, C.L. and Haley, C.A. (2013) Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers. *J Swine Health Prod* **21**, 72-84.
- Hood, E.E., Gelvin, S.B., Melchers, L.S. and Hoekema, A. (1993) New *Agrobacterium* helper plasmids for gene transfer to plants. *Transgenic Res* **2**, 208-218.
- Jiang, W., Jiang, P., Li, Y., Tang, J., Wang, X. and Ma, S. (2006) Recombinant adenovirus expressing GP5 and M fusion proteins of porcine reproductive and respiratory syndrome virus induce both humoral and cell-mediated immune responses in mice. *Vet Immunol Immunopathol* **113**, 169-180.
- Kadri, A., Wege, C. and Jeske, H. (2013) *In vivo* self-assembly of TMV-like particles in yeast and bacteria for nanotechnological applications. *J Virol Methods* **189**, 328-340.
- Kanekiyo, M., Bu, W., Joyce, M.G., Meng, G., Whittle, J.R., Baxa, U., Yamamoto, T., Narpala, S., Todd, J.P., Rao, S.S., McDermott, A.B., Koup, R.A., Rossmann, M.G., Mascola, J.R., Graham, B.S., Cohen, J.I. and Nabel, G.J. (2015) Rational design of an Epstein-Barr virus vaccine targeting the receptor-binding site. *Cell* **162**, 1090-1100.
- Kanekiyo, M., Wei, C.-J., Yassine, H.M., McTamney, P.M., Boyington, J.C., Whittle, J.R.R., Rao, S.S., Kong, W.-P., Wang, L. and Nabel, G.J. (2013) Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies. *Nature* **499**, 102-106.
- Karch, C.P. and Burkhard, P. (2016) Vaccine technologies: from whole organisms to rationally designed protein assemblies. *Biochem Pharmacol* **120**, 1-14.
- Kay, R., Chan, A., Daly, M. and McPherson, J. (1987) Duplication of CaMV 35S promoter sequences creates a strong enhancer for plant genes. *Science* **236**, 1299-1302.
- Kim, W.I., Kim, J.J., Cha, S.H., Wu, W.H., Cooper, V., Evans, R., Choi, E.J. and Yoon, K.J. (2013) Significance of genetic variation of PRRSV ORF5 in virus neutralization and molecular determinants corresponding to cross neutralization among PRRS viruses. *Vet Microbiol* **162**, 10-22.
- Kimman, T.G., Cornelissen, L.A., Moormann, R.J., Rebel, J.M.J. and Stockhofe-Zurwieden, N. (2009) Challenges for porcine reproductive and respiratory syndrome virus (PRRSV) vaccinology. *Vaccine* **27**, 3704-3718.
- Klug, A. (1999) The tobacco mosaic virus particle: structure and assembly. *Philos Trans R Soc Lond B Biol Sci* **354**, 531-535.

- Kolotilin, I., Topp, E., Cox, E., Devriendt, B., Conrad, U., Joensuu, J., Stöger, E., Warzecha, H., McAllister, T., Potter, A., McLean, M.D., Hall, J.C. and Menassa, R. (2014) Plant-based solutions for veterinary immunotherapeutics and prophylactics. *Vet Res* **45**, 117-117.
- Lai, H. and Chen, Q. (2012) Bioprocessing of plant-derived virus-like particles of Norwalk virus capsid protein under current Good Manufacture Practice regulations. *Plant Cell Rep* **31**, 573-584.
- Latham, T. and Galarza, J.M. (2001) Formation of wild-type and chimeric influenza virus-like particles following simultaneous expression of only four structural proteins. *J Virol* **75**, 6154-6165.
- Lemoine, C., Thakur, A., Krajišnik, D., Guyon, R., Longet, S., Razim, A., Górska, S., Pantelić, I., Ilić, T., Nikolić, I., Lavelle, E.C., Gamian, A., Savić, S. and Milicic, A. (2020) Technological approaches for improving vaccination compliance and coverage. *Vaccines (Basel)* **8**, 304.
- Li, J., Tao, S., Orlando, R. and Murtaugh, M.P. (2015) N-glycosylation profiling of porcine reproductive and respiratory syndrome virus envelope glycoprotein 5. *Virology* **478**, 86-98.
- Li, Y., Zhang, Y., Liao, Y., Sun, Y., Ruan, Y., Liu, C., Zhang, M., Li, F., Li, X., Fan, S., Yi, L., Ding, H., Zhao, M., Fan, J. and Chen, J. (2021) Preliminary evaluation of protective efficacy of inactivated Senecavirus A on Pigs. *Life (Basel)* **11**, 157.
- Lico, C., Santi, L., Twyman, R.M., Pezzotti, M. and Avesani, L. (2012) The use of plants for the production of therapeutic human peptides. *Plant Cell Rep* **31**, 439-451.
- Lomonosoff, G.P. and Wege, C. (2018) TMV particles: the journey from fundamental studies to bionanotechnology applications. *Adv Virus Res* **102**, 149-176.
- López-Sagasetta, J., Malito, E., Rappuoli, R. and Bottomley, M.J. (2016) Self-assembling protein nanoparticles in the design of vaccines. *Comput Struct Biotechnol J* **14**, 58-68.
- Loving, C.L., Brockmeier, S.L. and Sacco, R.E. (2007) Differential type I interferon activation and susceptibility of dendritic cell populations to porcine arterivirus. *Immunology* **120**, 217-229.
- Lu, B., Stubbs, G. and Culver, J.N. (1996) Carboxylate interactions involved in the disassembly of tobacco mosaic tobamovirus. *Virology* **225**, 11-20.
- Lund, P. and Dunsmuir, P. (1992) A plant signal sequence enhances the secretion of bacterial ChiA in transgenic tobacco. *Plant Mol Biol* **18**, 47-53.

- Lunney, J.K., Fang, Y., Ladinig, A., Chen, N., Li, Y., Rowland, B. and Renukaradhya, G.J. (2016) Porcine reproductive and respiratory syndrome virus (PRRSV): pathogenesis and interaction with the immune system. *Annu Rev Anim Biosci* **4**, 129-154.
- MacDonald, J., Doshi, K., Dussault, M., Hall, J.C., Holbrook, L., Jones, G., Kaldis, A., Klima, C.L., Macdonald, P., McAllister, T., McLean, M.D., Potter, A., Richman, A., Shearer, H., Yarosh, O., Yoo, H.S., Topp, E. and Menassa, R. (2015) Bringing plant-based veterinary vaccines to market: managing regulatory and commercial hurdles. *Biotechnol Adv* **33**, 1572-1581.
- Mardassi, H., Massie, B. and Dea, S. (1996) Intracellular synthesis, processing, and transport of proteins encoded by ORFs 5 to 7 of porcine reproductive and respiratory syndrome virus. *Virology* **221**, 98-112.
- Margolin, E.A., Strasser, R., Chapman, R., Williamson, A.-L., Rybicki, E.P. and Meyers, A.E. (2020) Engineering the plant secretory pathway for the production of next-generation pharmaceuticals. *Trends Biotechnol* **38**, 1034-1044.
- Marillonnet, S. and Werner, S. (2015) Assembly of multigene constructs using Golden Gate cloning. *Methods Mol Biol* **1321**, 269-284.
- Marsian, J. and Lomonosoff, G.P. (2016) Molecular pharming - VLPs made in plants. *Curr Opin Biotechnol* **37**, 201-206.
- Meining, W., Mortl, S., Fischer, M., Cushman, M., Bacher, A. and Ladenstein, R. (2000) The atomic structure of pentameric lumazine synthase from *Saccharomyces cerevisiae* at 1.85 Å resolution reveals the binding mode of a phosphonate intermediate analogue. *J Mol Biol* **299**, 181-197.
- Meo, S.A., Bukhari, I.A., Arkram, J., Meo, A.S. and Klonoff, D.C. (2021) COVID-19 vaccines: comparison of biological, pharmacological characteristics and adverse effects of Pfizer/BioNTech and Moderna Vaccines. *Eur Rev Med Pharmacol Sci* **25**, 1663-1669.
- Miletic, S., Simpson, D.J., Szymanski, C.M., Deyholos, M.K. and Menassa, R. (2016) A plant-produced bacteriophage tailspike protein for the control of *Salmonella*. *Front Plant Sci* **6**, 1221.
- Min, J., Kim, S., Lee, J. and Kang, S. (2014) Lumazine synthase protein cage nanoparticles as modular delivery platforms for targeted drug delivery. *RSC Adv* **4**, 48596-48600.
- Mohsen, M.O., Zha, L., Cabral-Miranda, G. and Bachmann, M.F. (2017) Major findings and recent advances in virus-like particle (VLP)-based vaccines. *Semin Immunol* **34**, 123-132.



- Mor, T.S. (2015) Molecular pharming's foot in the FDA's door: Protalix's trailblazing story. *Biotechnol Lett* **37**, 2147-2150.
- Nielsen, H.S., Oleksiewicz, M.B., Forsberg, R., Stadejek, T., Bøtner, A. and Storgaard, T. (2001) Reversion of a live porcine reproductive and respiratory syndrome virus vaccine investigated by parallel mutations. *J Gen Virol* **82**, 1263-1272.
- Noor, R. (2021) Developmental status of the potential vaccines for the mitigation of the COVID-19 pandemic and a focus on the effectiveness of the Pfizer-BioNTech and Moderna mRNA vaccines. *Curr Clin Microbiol Rep*, 1-8. Advance online publication.
- Offit, P.A. (2005) The Cutter incident, 50 years later. *N Engl J Med* **352**, 1411-1412.
- Okuda, Y., Kuroda, M., Ono, M., Chikata, S. and Shibata, I. (2008) Efficacy of vaccination with porcine reproductive and respiratory syndrome virus following challenges with field isolates in Japan. *J Vet Med Sci* **70**, 1017-1025.
- Pasternak, A.O., Spaan, W.J.M. and Snijder, E.J. (2006) Nidovirus transcription: how to make sense...? *J Gen Virol* **87**, 1403-1421.
- Pelham, H.R.B. (1990) The retention signal for soluble proteins of the endoplasmic reticulum. *Trends Biochem Sci* **15**, 483-486.
- Phan, H.T., Pham, V.T., Ho, T.T., Pham, N.B., Chu, H.H., Vu, T.H., Abdelwhab, E.M., Scheibner, D., Mettenleiter, T.C., Hanh, T.X., Meister, A., Gresch, U. and Conrad, U. (2020) Immunization with plant-derived multimeric H5 hemagglutinins protect chicken against highly pathogenic avian influenza virus H5N1. *Vaccines (Basel)* **8**, 593.
- Pileri, E. and Mateu, E. (2016) Review on the transmission porcine reproductive and respiratory syndrome virus between pigs and farms and impact on vaccination. *Vet Res* **47**, 108.
- Pirofski, L.A. and Casadevall, A. (1998) Use of licensed vaccines for active immunization of the immunocompromised host. *Clin Microbiol Rev* **11**, 1-26.
- Pogue, G.P., Lindbo, J.A., McCulloch, M.J., Lawrence, J.E., Gross, C.S. and Garger, S.J., inventors; Large Scale Biology Corporation, assignee. Production of a parvovirus vaccine in plants as viral coat protein fusions. United States patent US 7,270,825 B2. 2007 Sep 18.
- Popescu, L.N., Tribble, B.R., Chen, N. and Rowland, R.R.R. (2017) GP5 of porcine reproductive and respiratory syndrome virus (PRRSV) as a target for homologous and broadly neutralizing antibodies. *Vet Microbiol* **209**, 90-96.

- Prather, R.S., Rowland, R.R., Ewen, C., Tribble, B., Kerrigan, M., Bawa, B., Teson, J.M., Mao, J., Lee, K., Samuel, M.S., Whitworth, K.M., Murphy, C.N., Egen, T. and Green, J.A. (2013) An intact sialoadhesin (Sn/SIGLEC1/CD169) is not required for attachment/internalization of the porcine reproductive and respiratory syndrome virus. *J Virol* **87**, 9538-9546.
- Qi, M., Zhang, X.-E., Sun, X., Zhang, X., Yao, Y., Liu, S., Chen, Z., Li, W., Zhang, Z., Chen, J. and Cui, Z. (2018) Intranasal nanovaccine confers homo- and hetero-subtypic influenza protection. *Small* **14**, 1703207.
- Qu, Z., Li, M., Guo, Y., Liu, Y., Wang, J. and Gao, M. (2020) Expression, purification, and characterisation of recombinant ferritin in insect cells using the baculovirus expression system. *Biotechnol Lett* **42**, 57-65.
- Rath, A., Glibowicka, M., Nadeau, V.G., Chen, G. and Deber, C.M. (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc Natl Acad Sci U S A* **106**, 1760-1765.
- Robinson, P.J. and Bulleid, N.J. (2020) Mechanisms of disulfide bond formation in nascent polypeptides entering the secretory pathway. *Cells* **9**, 1994.
- Rowland, R.R., Lawson, S., Rossow, K. and Benfield, D.A. (2003) Lymphoid tissue tropism of porcine reproductive and respiratory syndrome virus replication during persistent infection of pigs originally exposed to virus in utero. *Vet Microbiol* **96**, 219-235.
- Rybicki, E.P. (2009) Plant-produced vaccines: promise and reality. *Drug Discov Today* **14**, 6-24.
- Rychlowska, M., Gromadzka, B., Bieńkowska-Szewczyk, K. and Szewczyk, B. (2011) Application of baculovirus-insect cell expression system for human therapy. *Curr Pharm Biotechnol* **12**, 1840-1849.
- Schillberg, S. and Finnern, R. (2021) Plant molecular farming for the production of valuable proteins - critical evaluation of achievements and future challenges. *J Plant Physiol* **258-259**, 153359.
- Shaaltiel, Y., Bartfeld, D., Hashmueli, S., Baum, G., Brill-Almon, E., Galili, G., Dym, O., Boldin-Adamsky, S.A., Silman, I., Sussman, J.L., Futerman, A.H. and Aviezer, D. (2007) Production of glucocerebrosidase with terminal mannose glycans for enzyme replacement therapy of Gaucher's disease using a plant cell system. *Plant Biotechnol J* **5**, 579-590.

- Shi, Y., Mowery, R.A., Ashley, J., Hentz, M., Ramirez, A.J., Bilgicer, B., Slunt-Brown, H., Borchelt, D.R. and Shaw, B.F. (2012) Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. *Protein Sci* **21**, 1197-1209.
- Silhavy, D., Molnár, A., Lucioli, A., Szittyá, G., Hornyik, C., Tavazza, M. and Burgyán, J. (2002) A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. *EMBO J* **21**, 3070-3080.
- Snijder, E.J., Kikkert, M. and Fang, Y. (2013) Arterivirus molecular biology and pathogenesis. *J Gen Virol* **94**, 2141-2163.
- Solá, R.J. and Griebenow, K. (2009) Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci* **98**, 1223-1245.
- Stoian, A.M.M. and Rowland, R.R.R. (2019) Challenges for porcine reproductive and respiratory syndrome (PRRS) vaccine design: reviewing virus glycoprotein interactions with CD163 and targets of virus neutralization. *Vet Sci* **6**, 9.
- Strasser, R. (2018) Protein quality control in the endoplasmic reticulum of plants. *Annu Rev Plant Biol* **69**, 147-172.
- Strasser, R., Seifert, G., Doblin, M.S., Johnson, K.L., Ruprecht, C., Pfrengle, F., Bacic, A. and Estevez, J.M. (2021) Cracking the “sugar code”: a snapshot of N- and O-glycosylation pathways and functions in plants cells. *Front Plant Sci* **12**, 640919.
- Suerbaum, S. and Michetti, P. (2002) *Helicobacter pylori* infection. *N Engl J Med* **347**, 1175-1186.
- Thuenemann, E.C., Byrne, M.J., Peyret, H., Saunders, K., Castells-Graells, R., Ferriol, I., Santoni, M., Steele, J.F.C., Ranson, N.A., Avesani, L., Lopez-Moya, J.J. and Lomonosoff, G.P. (2021) A replicating viral vector greatly enhances accumulation of helical virus-like particles in plants. *Viruses* **13**, 885.
- Topp, E., Irwin, R., McAllister, T., Lessard, M., Joensuu, J.J., Kolotilin, I., Conrad, U., Stöger, E., Mor, T., Warzecha, H., Hall, J.C., McLean, M.D., Cox, E., Devriendt, B., Potter, A., Depicker, A., Viridi, V., Holbrook, L., Doshi, K., Dussault, M., Friendship, R., Yarosh, O., Yoo, H.S., MacDonald, J. and Menassa, R. (2016) The case for plant-made veterinary immunotherapeutics. *Biotechnol Adv* **34**, 597-604.
- Van Breedam, W., Van Gorp, H., Zhang, J.Q., Crocker, P.R., Delputte, P.L. and Nauwynck, H.J. (2010) The M/GP(5) glycoprotein complex of porcine reproductive and respiratory syndrome virus binds the sialoadhesin receptor in a sialic acid-dependent manner. *PLoS Pathog* **6**, e1000730.

- VanderWaal, K. and Deen, J. (2018) Global trends in infectious diseases of swine. *Proc Natl Acad Sci USA* **115**, 11495-11500.
- Vetter, V., Denizer, G., Friedland, L.R., Krishnan, J. and Shapiro, M. (2018) Understanding modern-day vaccines: what you need to know. *Ann Med* **50**, 110-120.
- Vitale, A. and Raikhel, N.V. (1999) What do proteins need to reach different vacuoles? *Trends Plant Sci* **4**, 149-155.
- Waidner, B., Greiner, S., Odenbreit, S., Kavermann, H., Velayudhan, J., Stahler, F., Guhl, J., Bisse, E., van Vliet, A.H., Andrews, S.C., Kusters, J.G., Kelly, D.J., Haas, R., Kist, M. and Bereswill, S. (2002) Essential role of ferritin Pfr in *Helicobacter pylori* iron metabolism and gastric colonization. *Infect Immun* **70**, 3923-3929.
- Wang, A., Chen, Q., Wang, L., Madson, D., Harmon, K., Gauger, P., Zhang, J. and Li, G. (2019) Recombination between vaccine and field strains of porcine reproductive and respiratory syndrome virus. *Emerg Infect Dis* **25**, 2335-2337.
- Wang, J.W. and Roden, R.B.S. (2013) Virus-like particles for the prevention of human papillomavirus-associated malignancies. *Expert Rev Vaccines* **12**, 129-141.
- Ward, B.J., Gobeil, P., Séguin, A., Atkins, J., Boulay, I., Charbonneau, P.-Y., Couture, M., D'Aoust, M.-A., Dhaliwall, J., Finkle, C., Hager, K., Mahmood, A., Makarkov, A., Cheng, M.P., Pillet, S., Schimke, P., St-Martin, S., Trépanier, S. and Landry, N. (2021) Phase 1 randomized trial of a plant-derived virus-like particle vaccine for COVID-19. *Nat Med* **27**, 1071-1078.
- Ward, B.J., Makarkov, A., Séguin, A., Pillet, S., Trépanier, S., Dhaliwall, J., Libman, M.D., Vesikari, T. and Landry, N. (2020) Efficacy, immunogenicity, and safety of a plant-derived, quadrivalent, virus-like particle influenza vaccine in adults (18–64 years) and older adults ( $\geq 65$  years): two multicentre, randomised phase 3 trials. *Lancet* **396**, 1491-1503.
- Wei, Z., Lin, T., Sun, L., Li, Y., Wang, X., Gao, F., Liu, R., Chen, C., Tong, G. and Yuan, S. (2012) N-linked glycosylation of GP5 of porcine reproductive and respiratory syndrome virus is critically important for virus replication *in vivo*. *J Virol* **86**, 9941-9951.
- Wills, R.W., Doster, A.R., Galeota, J.A., Sur, J.H. and Osorio, F.A. (2003) Duration of infection and proportion of pigs persistently infected with porcine reproductive and respiratory syndrome virus. *J Clin Microbiol* **41**, 58-62.
- Wise, A.A., Liu, Z. and Binns, A.N. (2006) Three methods for the introduction of foreign DNA into *Agrobacterium*. *Methods Mol Biol* **343**, 43-53.

- Wissink, E.H., Kroese, M.V., van Wijk, H.A., Rijsewijk, F.A., Meulenbergh, J.J. and Rottier, P.J. (2005) Envelope protein requirements for the assembly of infectious virions of porcine reproductive and respiratory syndrome virus. *J Virol* **79**, 12495-12506.
- Wu, K., Malik, K., Tian, L., Hu, M., Martin, T., Foster, E., Brown, D. and Miki, B. (2001) Enhancers and core promoter elements are essential for the activity of a cryptic gene activation sequence from tobacco, tCUP. *Mol Genet Genomics* **265**, 763-770.
- Zhang, B., Chao, C.W., Tsybovsky, Y., Abiona, O.M., Hutchinson, G.B., Moliva, J.I., Olin, A.S., Pegu, A., Phung, E., Stewart-Jones, G.B.E., Verardi, R., Wang, L., Wang, S., Werner, A., Yang, E.S., Yap, C., Zhou, T., Mascola, J.R., Sullivan, N.J., Graham, B.S., Corbett, K.S. and Kwong, P.D. (2020) A platform incorporating trimeric antigens into self-assembling nanoparticles reveals SARS-CoV-2-spike nanoparticles to elicit substantially higher neutralizing responses than spike alone. *Sci Rep* **10**, 18149.
- Zhang, M., Han, X., Osterrieder, K. and Veit, M. (2021) Palmitoylation of the envelope membrane proteins GP5 and M of porcine reproductive and respiratory syndrome virus is essential for virus growth. *PLoS Pathog* **17**, e1009554.
- Zhang, X., Konarev, P.V., Petoukhov, M.V., Svergun, D.I., Xing, L., Cheng, R.H., Haase, I., Fischer, M., Bacher, A., Ladenstein, R. and Meining, W. (2006) Multiple assembly states of lumazine synthase: a model relating catalytic function and molecular assembly. *J Mol Biol* **362**, 753-770.
- Zhang, X., Meining, W., Fischer, M., Bacher, A. and Ladenstein, R. (2001) X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile *Aquifex aeolicus* at 1.6 Å resolution: determinants of thermostability revealed from structural comparisons. *J Mol Biol* **306**, 1099-1114.
- Zhang, X., Xu, K., Ou, Y., Xu, X. and Chen, H. (2018) Development of a baculovirus vector carrying a small hairpin RNA for suppression of sf-caspase-1 expression and improvement of recombinant protein production. *BMC Biotechnol* **18**, 24.
- Zhao, L., Seth, A., Wibowo, N., Zhao, C.X., Mitter, N., Yu, C. and Middelberg, A.P. (2014) Nanoparticle vaccines. *Vaccine* **32**, 327-337.
- Ziebuhr, J., Snijder, E.J. and Gorbalenya, A.E. (2000) Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. *J Gen Virol* **81**, 852-879.
- Zuckermann, F.A., Garcia, E.A., Luque, I.D., Christopher-Hennings, J., Doster, A., Brito, M. and Osorio, F. (2007) Assessment of the efficacy of commercial porcine reproductive and respiratory syndrome virus (PRRSV) vaccines based on measurement of serologic response, frequency of gamma-IFN-producing cells and virological parameters of protection upon challenge. *Vet Microbiol* **123**, 69-85.

Zylberman, V., Craig, P.O., Klinke, S., Braden, B.C., Cauerhff, A. and Goldbaum, F.A. (2004) High order quaternary arrangement confers increased structural stability to *Brucella* sp. lumazine synthase. *J Biol Chem* **279**, 8093-8101.

## 7 APPENDICES

**Appendix 1.** Nucleotide sequences of codon-optimized genes for plant expression. Codons modified for amino acid substitutions are shown in red.

<b>ALS</b>	<p>CAGATCTATGAAGGTAAATTGACTGCTGAAGGTCTTAGATTTGGT  ATTGTTGCTTCTCGGTTTAATCATGCTCTTGTTGATAGACTTGTGG  AAGGTGCTATTGATTGTATTGTTAGGCACGGCGGTCGTGAGGAA  GATATAACACTAGTTCGGGTCCCTGGTTCTTGGGAGATTCCTGTT  GCTGCTGGTGAAGTTGCTAGAAAGGAAGATATTGATGCTGTTAT  AGCTATTGGTGTTCTTATTAGAGGTGCTACTCCTCATTTTGATTAT  ATTGCTAGTGAAGTTTCTAAGGGTCTTGCTAATCTTTCTCTTGAA  CTTAGAAAGCCTATTACTTTTGGTGTATTACTGCTGATACTCTTG  AACAAAGCTATTGAAAGAGCTGGTACTAAGCATGGTAATAAGGGT  TGGGAAGCTGCTTTGTCTGCTATTGAAATGGCTAATCTTTTAAAG  TCTCTTAGA</p>
<b>ALSm</b>	<p>CAGATCTATGAAGGTAAATTGACTGCTGAAGGTCTTAGATTTGGT  ATTGTTGCTTCTCGGTTTAATCATGCTCTTGTTGATAGACTTGTGG  AAGGTGCTATTGATTGTATTGTTAGGCACGGCGGTCGTGAGGAA  GATATAACACTAGTTCGGGTCCCTGGTTCTTGGGAGATTCCTGTT  GCTGCTGGTGAAGTTGCTAGAAAGGAAGATATTGATGCTGTTAT  AGCTATTGGTGTTCTTATTAGAGGTGCTACTCCTCATTTTGATTAT  ATTGCTAGTGAAGTTTCTAAGGGTCTTGCTCAACTTTCTCTTGAA  CTTAGAAAGCCTATTACTTTTGGTGTATTACTGCTGATACTCTTG  AACAAAGCTATTGAAAGAGCTGGTACTAAGCATGGTAATAAGGGT  TGGGAAGCTGCTTTGTCTGCTATTGAAATGGCTAATCTTTTAAAG  TCTCTTAGA</p>
<b>TMVc</b>	<p>TCTTACTCTATTACTACTCCTTCTCAATTTGTTTTTCTTTCTTCTGC  TTGGGCTGATCCTATTGAATTGATTAATCTTTGTACTAATGCTCTT  GGTAATCAGTTTCAAACCTCAACAAGCTAGGACTGTTGTTCAAAG  ACAATTTTCTCAAGTTTGGGAAGCCTTCTCCACAAGTTACTGTTAG  ATTCCTGATTCTGATTTTAAGGTTTACAGATATAATGCTGTTTTG  AATCCTCTTGTTACTGCTTTGCTTGGAGCTTTTGATACTAGAAAT  AGGATTATTGAAGTTGAAAATCAAGCTAATCCAACACTACTGCTGA  AACTCTTGATGCTACAAGAAGAGTTGATGATGCTACTGTTGCTAT  TAGGTCTGCTATTAATAACCTTATTGTTGAACTTATTAGGGGTAC  TGGTTCTTATAATAGGTCATCTTTTGAATCTTCTTCTGGTCTTGTT  TGGACTTCTGGTCCTGCTACT</p>

<b>Fer</b>	<p>CTATCAAAAGATATTATTA AATTGTTAAATGAGCAAGTTAATAA  GGAGATGAATAGCTCTAATCTTTATATGTCTATGAGTAGTTGGTG  TTATACACATTCTCTTGATGGTGCAGGATTATTTCTTTTTGATCAT  GCAGCTGAAGAATATGAACATGCTAAGAAATTAATAGTTTTTCTT  AATGAGAACAATGTACCTGTTCAATTGACTTCAATTTCTGCTCCC  GAGCATAAGTTTGAAGGATTGACTCAGATTTTCCAAAAGGCGTA  CGGTCATGAACAGCATATAAGTGAATCTATTAATAACATAGTTG  ATCACGCTATTAAGAGTAAAGACCACGCAACGTTTAACTTCCTA  CAATGGTATGTAGCTGAGCAGCATGAGGAAGAGGGTCTTGTTTAA  AGATATTCTTGATAAAATTGAATTGATTGGTAATGAGAATCATG  GTCTTTATCTTGCTGATCAATATGTTAAAGGTATAGCTAAGTCTA  GAAAGTCT</p>
<b>Ferm</b>	<p>CTATCAAAAGATATTATTA AATTGTTAAATGAGCAAGTTAATAA  GGAGATG<b>CAG</b>AGCTCTAATCTTTATATGTCTATGAGTAGTTGGTG  TTATACACATTCTCTTGATGGTGCAGGATTATTTCTTTTTGATCAT  GCAGCTGAAGAATATGAACATGCTAAGAAATTAATAGTTTTTCTT  AATGAGAACAATGTACCTGTTCAATTGACTTCAATTTCTGCTCCC  GAGCATAAGTTTGAAGGATTGACTCAGATTTTCCAAAAGGCGTA  CGGTCATGAACAGCATATAAGTGAATCTATTAATAACATAGTTG  ATCACGCTATTAAGAGTAAAGACCACGCAACGTTTAACTTCCTA  CAATGGTATGTAGCTGAGCAGCATGAGGAAGAGGGTCTTGTTTAA  AGATATTCTTGATAAAATTGAATTGATTGGTAATGAGAATCATG  GTCTTTATCTTGCTGATCAATATGTTAAAGGTATAGCTAAGTCTA  GAAAGTCT</p>
<b>M</b>	<p>GGTTCCTCTCTTGATGATTTTTGCCATGATTCTACCGCTCCTCAA  AGGTT</p>
<b>GP5</b>	<p>AATGCTAGT<b>GCT</b>GACTCTAGTTCACACCTCCAGCTCATCTATAAT  TTGACTCTTTGTGAACTT<b>GCT</b>GGTACTGAT</p>
<b>Flexible linkers</b>	<p>GGTGGTTCAGGTGGTTCTGGTGGTAGTGGAGGTAGT</p>



**Appendix 2.** Amino acid sequences of all proteins produced recombinantly. Amino acids substituted from their wild type sequences are shown in red, and GP5's glycosylation site is shown in blue.

<b>APO-ALS</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYAQIYEGK LTAEGLRFGIVASRFNHALVDRLVEGAIDCIVRHGGREEDITLVRV PGSWEIPVAAGELARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGL ANLSLELRKPITFGVITADTLEQAIERAGTKHGNKGWEAALSAIEM ANLFKSLREQKLISEEDL
<b>VAC-ALS</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYAQIYEGK LTAEGLRFGIVASRFNHALVDRLVEGAIDCIVRHGGREEDITLVRV PGSWEIPVAAGELARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGL ANLSLELRKPITFGVITADTLEQAIERAGTKHGNKGWEAALSAIEM ANLFKSLREQKLISEEDLNGLLVDTM
<b>ER-ALS</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYAQIYEGK LTAEGLRFGIVASRFNHALVDRLVEGAIDCIVRHGGREEDITLVRV PGSWEIPVAAGELARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGL ANLSLELRKPITFGVITADTLEQAIERAGTKHGNKGWEAALSAIEM ANLFKSLREQKLISEEDLKDEL
<b>APO-ALSm</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYAQIYEGK LTAEGLRFGIVASRFNHALVDRLVEGAIDCIVRHGGREEDITLVRV PGSWEIPVAAGELARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGL AQLSLELRKPITFGVITADTLEQAIERAGTKHGNKGWEAALSAIEM ANLFKSLREQKLISEEDL
<b>VAC-ALSm</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYAQIYEGK LTAEGLRFGIVASRFNHALVDRLVEGAIDCIVRHGGREEDITLVRV PGSWEIPVAAGELARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGL AQLSLELRKPITFGVITADTLEQAIERAGTKHGNKGWEAALSAIEM ANLFKSLREQKLISEEDLNGLLVDTM
<b>ER-ALSm</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYAQIYEGK LTAEGLRFGIVASRFNHALVDRLVEGAIDCIVRHGGREEDITLVRV PGSWEIPVAAGELARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGL AQLSLELRKPITFGVITADTLEQAIERAGTKHGNKGWEAALSAIEM ANLFKSLREQKLISEEDLKDEL
<b>APO-Fer</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYALSKDII KLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEE YEHAKKLIVFLNENNPVQLTSISAPEHKFEGLTQIFQKAYGHEQ HISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKI ELIGNENHGLYLADQYVKGIAKSRKSEQKLISEEDL

<b>VAC-Fer</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYALSKDII KLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEE YEHAKKLIVFLNENNVVQLTSISAPEHKFEGLTQIFQKAYGHEQ HISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKI ELIGNENHGLYLADQYVKGIAKSRKSEQKLISEEDLNGLLVDTM
<b>ER-Fer</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYALSKDII KLLNEQVNKEMNSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEE YEHAKKLIVFLNENNVVQLTSISAPEHKFEGLTQIFQKAYGHEQ HISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKI ELIGNENHGLYLADQYVKGIAKSRKSEQKLISEEDLKDEL
<b>APO-Ferm</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYALSKDII KLLNEQVNKEMQSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEE YEHAKKLIVFLNENNVVQLTSISAPEHKFEGLTQIFQKAYGHEQ HISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKI ELIGNENHGLYLADQYVKGIAKSRKSEQKLISEEDL
<b>VAC-Ferm</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYALSKDII KLLNEQVNKEMQSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEE YEHAKKLIVFLNENNVVQLTSISAPEHKFEGLTQIFQKAYGHEQ HISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKI ELIGNENHGLYLADQYVKGIAKSRKSEQKLISEEDLNGLLVDTM
<b>ER-Ferm</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYALSKDII KLLNEQVNKEMQSSNLYMSMSSWCYTHSLDGAGLFLFDHAAEE YEHAKKLIVFLNENNVVQLTSISAPEHKFEGLTQIFQKAYGHEQ HISESINNIVDHAIKSKDHATFNFLQWYVAEQHEEEVLFKDILDKI ELIGNENHGLYLADQYVKGIAKSRKSEQKLISEEDLKDEL
<b>APO-TM<sub>v</sub>c</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYASYSITT PSQFVFLSSAWADPIELINLCTNALGNQFQTQQARTVVQRQFSQV WKPSPQVTVRFPDSDFKVYRYNAVLNPLVTALLGAFDTRNRRIEV ENQANPTTAETLDATRVRVDDATVAIRSAINNLIVELIRGTGSYNRS SFESSGLVWTSGPATEQKLISEEDL
<b>VAC-TM<sub>v</sub>c</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYASYSITT PSQFVFLSSAWADPIELINLCTNALGNQFQTQQARTVVQRQFSQV WKPSPQVTVRFPDSDFKVYRYNAVLNPLVTALLGAFDTRNRRIEV ENQANPTTAETLDATRVRVDDATVAIRSAINNLIVELIRGTGSYNRS SFESSGLVWTSGPATEQKLISEEDLNGLLVDTM
<b>ER-TM<sub>v</sub>c</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYASYSITT PSQFVFLSSAWADPIELINLCTNALGNQFQTQQARTVVQRQFSQV WKPSPQVTVRFPDSDFKVYRYNAVLNPLVTALLGAFDTRNRRIEV ENQANPTTAETLDATRVRVDDATVAIRSAINNLIVELIRGTGSYNRS SFESSGLVWTSGPATEQKLISEEDLKDEL

<b>APO-ALSm-M-GP5</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYAQIYEGK LTAEGLRFGIVASRFNHALVDRLVEGAIDCIVRHGGREEDITLVRV PGSWEIPVAAGELARKEDIDAVIAIGVLIRGATPHFDYIASEVSKGL AQLSLELRKPITFGVITADTLEQAIERAGTKHGNKGWEAALSAIEM ANLFKSLRGGSGGSGGSGGSSLDLDFCHDSTAPQKVGGSGGSGG SGGSNASADSSSHLQLIYNLTLCEL <del>AGT</del> DEQKLISEEDL
<b>APO-M-GP5-Ferm</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYAGSSLDLDF CHDSTAPQKVGGSGGSGGSGGSNASADSSSHLQLIYNLTLCEL <del>AGT</del> DGGSGGSGGSGGSLSKDIIKLLNEQVNKEMQSSNLYMSMSSWCYT HSLDGAGLFLFDHAAEEYEHA <del>K</del> LIVFLNENNVQVQLTSISAPEHKF EGLTQIFQKAYGHEQHISESINNIVDHAIKSKDHATFNFLQWYVAEQ HEEEVLFKDILDKIELIGNENHGLYLADQYVKGIAKSRKSEQKLISEE DL
<b>APO-TM<sub>v</sub>c-M-GP5</b>	MGFFLFSQMPSFFLVSTLLLFLIISHSSHAGGYPYDVPDYASYSITTP SQFVFLSSAWADPIELINLCTNALGNQFQTQQARTVVQRQFSQVW KPSPQVTVRFPDSDFKVYRYNAVLNPLVTALLGAFDTRNRIIEVEN QANPTTAETLDATTRVDDATVAIRSAINNLIVELIRGTGSYNRSSFE SSSGLVWTSGPATGGSGGSGGSGGSSLDLDFCHDSTAPQKVGGSGG GGSGGSGGSNASADSSSHLQLIYNLTLCEL <del>AGT</del> DEQKLISEEDL

**Appendix 3.** Nucleotide sequences for primers (5' - 3').

<b>35S Forward</b>	CCTTCGCAAGACCCTTCCTCTAT
<b>Nos Reverse</b>	CCGGCAACAGGATTCAATCTTAA
<b>Xpress Forward</b>	TCTCTATGATGACGATGACAAGGG

**Appendix 4.** Results of Games-Howell post-hoc test from R (p-values).

	ALSm	ALSm-M-GP5	Ferm	M-GP5-Ferm	TM <sub>v</sub> c
ALSm-M-GP5	$5.6 \times 10^{-6}$	-	-	-	-
Ferm	0.4588	$3.1 \times 10^{-6}$	-	-	-
M-GP5-Ferm	0.3760	$1.8 \times 10^{-6}$	0.9961	-	-
TM <sub>v</sub> c	$1.3 \times 10^{-6}$	0.0013	$1.2 \times 10^{-8}$	$9.8 \times 10^{-9}$	-
TM <sub>v</sub> c-M-GP5	$8.3 \times 10^{-5}$	0.0031	$8.6 \times 10^{-6}$	$3.8 \times 10^{-6}$	0.9936

## 8 CURRICULUM VITAE

**Name:** Jordan VanderBurgt

### **Post-secondary Education and Degrees:**

#### **Master of Science in Biology**

**2019 - present**

Western University, London, Ontario

*Thesis title:* “Production of self-assembling protein nanocages and virus-like particles displaying porcine reproductive and respiratory syndrome virus epitopes in *Nicotiana benthamiana*”

#### **Bachelor of Science Honors (4 years)**

**2015 - 2019**

Western University, London, Ontario

#### **Honors Specialization in Genetics**

*Honors thesis title:* “Identifying the cinnamate 4-hydroxylase enzyme family in soybean and determining their subcellular localization”

### **Honors and Awards:**

Award for best oral presentation, Canadian Society of Plant Biologists conference, 2021

Ontario Graduate Scholarship, Western University, 2020

Gold Medal Honors Specialization in Genetics, Western University, 2019

Dean’s Honor List, Western University, 2015-2019

Albert O. Jeffery Scholarship in Genetics, Western University, 2017

Western Scholarship of Excellence, Western University, 2015

### **Related Work Experience:**

Teaching Assistant, Western University 2019-2020

Molecular Genetic Laboratory (Biol 2583F), Advanced Genetics (Biol 3595B)

### **Presentations:**

- VanderBurgt, J., Menassa, R. (2021, June) Production of self-assembling virus-like particles displaying PRRSV epitopes in *Nicotiana benthamiana*. Oral presentation at the Canadian Society of Plant Biologists 2021 conference, Online.
- VanderBurgt, J., Menassa, R. (2021, May) Production of self-assembling virus-like particles displaying PRRSV epitopes in *Nicotiana benthamiana*. Oral presentation at the 47<sup>th</sup> Annual AGSBS Biology Symposium, Online.
- VanderBurgt, J., Menassa, R. (2019, September) Production of self-assembling protein nanocages and virus-like particles displaying PRRSV epitopes in *Nicotiana benthamiana*. Poster presented at the Biology Graduate Research Forum, London, Ontario.