Engineering the *Nicotiana benthamiana* secretory pathway for the production of Lujo virus vaccines and diagnostic tools

Abdul Rahman Isaacs

Dissertation presented for the Degree of

MASTER OF SCIENCE

In the Department of Molecular and Cell Biology

Faculty of Science

University of Cape Town

July 2022

Supervised by Dr. Emmanuel Margolin, Dr. Ann Meyers and Prof. Ed Rybicki

Table of Contents

Plagiarism declaration

The work in this thesis was conducted in the Biopharming Research Unit (Department of Molecular and Cell Biology, University of Cape Town) and the Vaccinology research group (Institute of infectious Diseases and Molecular Medicine, University of Cape Town). Research was performed under the guidance of Dr Emmanuel Margolin, Professor Ed Rybicki and Dr Ann Meyers.

I, Abdul Rahman Isaacs, hereby declare that the work in this thesis is my work (except where acknowledgements and references indicate otherwise) and that neither the whole work nor any part of it has been or is being submitted for another degree in this or any other university. Furthermore, I give permission to the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner.

Signed by candidate

Abdul Rahman Isaacs

Acknowledgements

I would like to thank the following individuals who have helped me on my journey thus far:

All my family especially my wonderful parents, Moutiel and Sajeda Isaacs for their never ending kindness, support and belief in me through my hardest days. I am truly lucky to have them.

My incredible supervisors: Emmanuel Margolin, Ann Meyers and Ed Rybicki for their great effort, support, patience and helpful advice throughout my Msc. A special thanks to Anna-Lise Williamson for extending her lab group, an office and many other resources to me during my Msc.

My colleagues at the Biopharming Research Unit, Vaccinology research group and Medical Virology division for their assistance in saving my experiments when they seemed to be failing like the Hindenburg.

The University of Cape Town, National research foundation (NRF) and Poliomyelitis research foundation (PRF) for providing financial assistance and making this project possible.

My exceptional friends: Jessica Croudace, for restoring my belief in the world and making it a better place, Thaabiet Parker and Eesaa Harris for the Vida meetups and helpful life advice, Dr. Leah Whittle for all the lab guidance, tea (and coffee), Abdullah Akhalwaya for always being there and keeping me focused. The Tai Chi club and skate group which prevented me from falling into insanity.

Dr Ghadija Hayat for providing great healthcare and Dr Sherin Bickrum for teaching me how to eat an elephant.

And anyone that helped to make my life and journey better in any way, I sincerely thank you from the bottom of my heart.

Abstract

Sub-Saharan Africa is severely deficient in vaccine manufacturing facilities that can keep up with the rate of emergence of viral pathogens. As seen with the Covid-19 pandemic, outbreaks of disease can be extremely detrimental to economies and put severe strain on the public health sector. Vaccination offers a solution to reduce the difficulties that accompany viral outbreaks. Lujo virus, an emerging arenavirus responsible for causing a devastating haemorrhagic fever with an 80% mortality rate, currently has no vaccines or diagnostic tools available. In this study we developed a production pipeline in HEK293T cells and *N. benthamiana* for the protein LUJV GP-CΔTM (Lujo virus glycoprotein precursor without the transmembrane region). The LUJV GP-CΔTM was constructed from gene sequences of the envelope glycoprotein of Lujo virus and then adapted to production in HEK293T cells using the DNA expression vector pTHpCapR. An *N. benthamiana* plant protein expression system was developed in parallel to compare the production utility of both systems with an emphasis being placed on the plant system. Plant expression systems are arguably cheaper and more easily automatable than traditional mammalian expression technologies. This makes them suitable for vaccine protein production in lower socio-economic countries that have an overwhelming burden of disease and poor health care systems. LUJV GP-CΔTM was successfully expressed in both HEK293T cells and *N. benthamiana*. It was confirmed that the LUJV GP-CΔTM undergoes the post translational modifications glycosylation and proteolytic cleavage in recombinant HEK293T cells and was produced in a conformation that allowed successful purification via a His tag sequence that was inserted into the LUJV GP-CΔTM protein gene sequence. On the other hand, *N. benthamiana* does not endogenously express the Site-1 protease needed for cleavage of the Lujo glycoprotein, and thus this needed to be co-expressed. Attempts were made to detect the Site-1 protease including extraction into buffers with different pHs and ammonium sulfate precipitation to concentrate the protein. However, I was not able elicit proteolytic cleavage of the LUJV GP-CΔTM nor detect the Site-1 protease in *N. benthamiana.* This is probably attributable to the innate responses against the hostile nature of proteases to non-native expression hosts. Proteins were expressed in *N. benthamiana* through the use of previously established protocols utilizing recombinant *Agrobacterium tumefaciens* to deliver synthesized genes to plant cells and induce their expression. It was determined that co-expression of the molecular folding chaperone calreticulin is necessary for LUJV GP-CΔTM to accumulate at detectable levels. Co-expression of an oligossacharyl transferase LmSTT3D, isolated from *Leishmania major,* may increase the glycan occupancy of the LUJV GP-CΔTM protein, indicated by a molecular mass shift. However, further experimental lines of evidence are needed –such as glycosylation mapping to determine if this is the case. Other parameters such as the optimal day of protein harvest

and optical density of recombinant *Agrobacterium tumefaciens* strains were recorded. Collectively these findings serve as a prototype pipeline for the production of commercially relevant immunogens, diagnostic tools and virus-like particles. This study narrows the focus for bottlenecks in plant protein production, not just for Lujo virus proteins, but for arenaviruses generally and potentially also other haemorrhagic fever-causing viruses. Future efforts should be directed towards addressing the barriers to plant production of complex viral antigens, and to further investigate the utility of mammalian cell-produced LUJV GP-CΔTM in animal studies.

Chapter 1: Literature review

1.1. Introduction

The increasing incidence of infectious disease outbreaks worldwide highlights the urgent need for the development of new technologies for the rapid large-scale production of vaccines and biopharmaceuticals [1]. These infectious disease outbreaks are largely driven by zoonoses from animal hosts [1] and therefore the development of surveillance is also important to pre-empt potential zoonosis outbreaks. Historically, many of these outbreaks have occurred in developing countries where the healthcare infrastructure is poorly equipped to respond, with examples including the 2014-2016 Ebola virus outbreak in West Africa [2], the increasing spread of Lassa virus in Nigeria since the 2018 outbreak [3, 4], as well as severe acute respiratory virus (SARS-CoV-2) from 2019 until the present [4]. The majority of the population in these poor countries live under highly concentrated and unsanitary conditions and – when coupled with poor healthcare infrastructure and limited funds for vaccine manufacturing – this results in unacceptable mortality rates [5, 6].

In addition to the direct burden of morbidity and mortality arising from disease outbreaks, they also have severe economic implications [1]. This was demonstrated recently during the SARS-CoV-2 pandemic with the need to impose nationwide lockdowns, the resulting loss of jobs and economic burden, and the initial lack of availability of vaccines [7]. Emergence of these diseases is often driven by environmental and ecological factors but may also be closely tied in with socio-economics and politics [1]. While developed countries are better equipped to respond to infectious disease outbreaks, developing countries often have limited infrastructure, ineffective healthcare systems and sub-optimal living conditions, which may exacerbate the impact of the disease. Even the promising rVSV-ZEBOV-GP candidate vaccine for Ebola could not be produced in the vaccine facilities in the DRC, probably due to limited infrastructure [8]. Irrespective of the origin of an infectious disease, the abovementioned conditions will cause an increase in the infection rate in low socio-economic regions compared to more affluent countries. Disastrous outbreaks of disease may be pre-empted by political decisions that result in economic recessions - where the people barely surviving between pay-cheques are the most affected [9]. This scenario was demonstrated in the 2019 COVID-19 outbreak where national lockdowns were implemented around the world for long periods of time, resulting in companies and employees losing large amounts of money and business [9, 10].

The specific knock-on effects of infectious disease outbreaks are difficult to prepare for, making it imperative to constantly find ways to combat infectious diseases and pre-empt the negative effects

of outbreaks. Furthermore, new, more virulent viruses can emerge from cross-species transmission and other cryptic sources, causing recurring outbreaks or even establishing a natural reservoir in humans [11, 12]. Since there is a lack of information regarding the total distribution and diversity of viruses as a whole - which is further complicated by fast mutation rates of some viruses – it is difficult to predict when, where and with what severity a novel virus will emerge. This further highlights the need to find cost-effective and efficient technologies, such as broad spectrum vaccines, that can combat a variety of viruses in a timeous manner. To tackle issues contextual to developing countries, there needs to be emphasis on these vaccines being rapidly scalable and easily produced - circumventing the need for complex infrastructure and funds. Here I will discuss the example of arenaviruses as emerging pathogens, and explore their infectious dynamics as well as past outbreaks to demonstrate the need and direction for future vaccine design.

1.2. Arenaviruses as emerging human pathogens

Arenaviruses are diverse and widely spread, with multiple species known to cause human disease [13]. However the most severe consequence of arenavirus infection is the development of a haemorrhagic fever and shock syndrome which can be associated with a high mortlity rate even between 80-100% for certain arenaviruses [14, 15]. Since the discovery of lymphocytic choriomeningitis virus in 1933 [13], subsequent species of arenaviruses have been identified as responsible for outbreaks around the globe, some of which carry a high fatality rate and continue to spread to new patients today [3, 13]. There are no licensed vaccines or diagnostic tools to combat arenaviruses save a candidate vaccine for Junin virus [16] and an antiviral drug that is somewhat effective for Lassa virus [17]. This presents a dire case for the healthcare system in developing countries as they will not be sufficiently equipped to prepare for and contain these outbreaks, resulting in a higher fatality rate. Metagenomic analyses have been performed in an attempt to classify arenaviruses and try to predict emergence of new virulent strains [18]. However, there does not seem to be a link between virulence of arenaviruses and a shared a common ancestor, which highlights the ineffective epidemiological foresight surrounding arenaviruses [19]. More worringly, not all the natural host reservoirs are known for haemorrhagic fever-causing arenaviruses [13] and attempts at discovering the natural hosts have sometimes led to discovery of even more novel arenaviruses [20]. This review explores some of these concepts further and illustrates the currently ineffective tools available, as well as the general attitude of neglect of arenavirus research as a whole despite the viruses being responsible for a large number of casualties globally.

1.3. Arenavirus classification

Lymphocytic choriomeningitis virus was the frst arenavirus identified in 1933 [13], which afterwards served as a model arenavirus for further discoveries. The International Committee on Taxonomy of Viruses documents the genera in the *Arenaviridae* family: there are over 33 mammal-infecting arenaviruses (mammarenaviruses) and three reptile–infecting (reptarenaviruses) [13]. Inclusion of genera including snake-borne hartmaniviruses and antennaviruses that infect frogfish are reported [18, 21]; however, they remain a topic of divergence in the literature and will thus not be discussed at length. Figure 1 shows a phylogenetic tree depicting phylogenetic relationships between some of the main arenaviruses, especially with regards to mammalian pathogenicity (mammarenaviruses). The sequence data used for classification was based on the envelope glycoprotein precursor (GP-C) of each arenavirus (see 1.4 structural organization of arenaviruses). Mammarenaviruses are subgrouped into Old World (OW) and New world (NW) arenaviruses, originating in the Americas or Southeast Asia and Sub-saharan Africa respectively [13]. Old world (OW) arenaviruses are grouped together in one clade while New World arenaviruses are divided into clades A,B,C, and the recombinant A/rec clade.

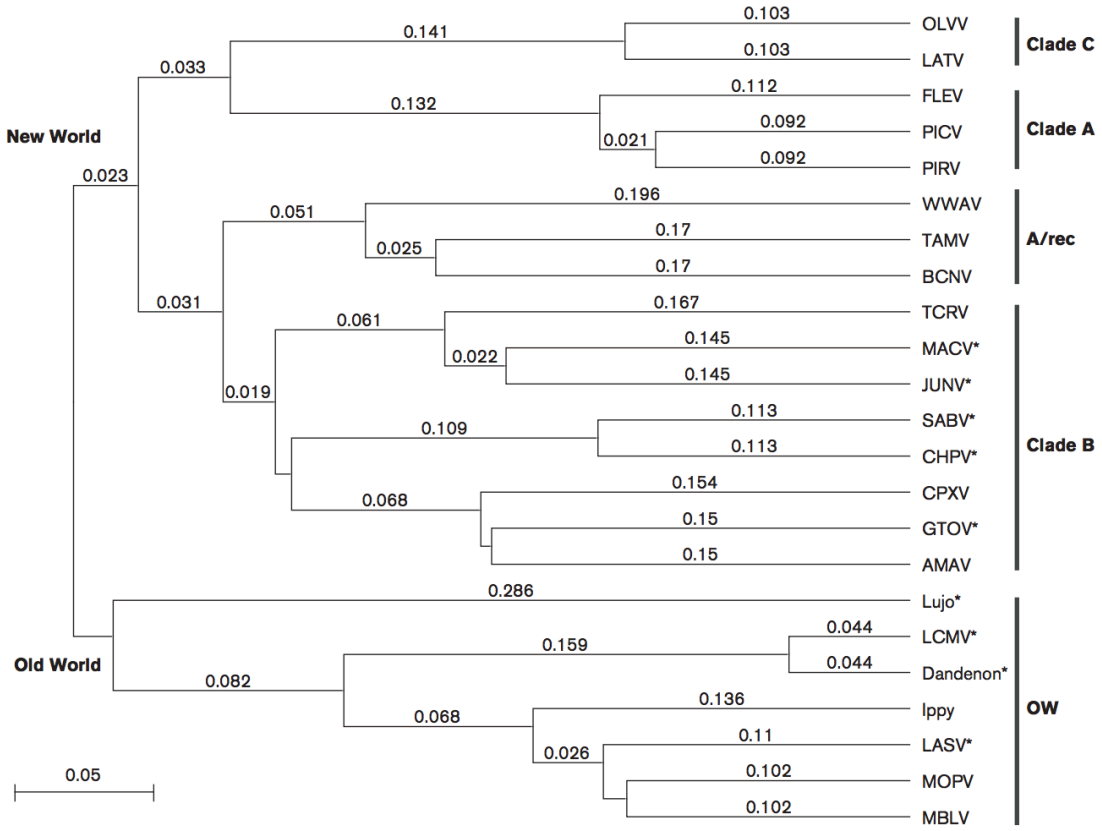


Figure 1: Phylogenetic tree of multiple arenavirus GP-C. Phylogenetic tree showing sequence alignments of the GP-C from different arenaviruses by clustal W analysis performed and published by [19]. The method used to generate the tree was the unweighted pair group with arithmetic mean and was performed in MacVector 12.6.0. The horizontal bar at the bottom left represents the protein differences in amino acid changes with 0.05 amino acid changes per the distance indicated for arenavirus glycoproteins. Arenavirus GP-Cs aligned include: Oliveros virus (OLVV), Latino virus (LATV), Flexal virus (FLEV), Pichinde vrus (PICV), Pirital virus (PIRV), White water Arroyo virus (WWAW), Tamiami virus (TAMV), Bear Canyon virus (BCNV), Tacaribe virus (TCRV), Machupo virus (MACV), Junin virus (JUNV), Sabia virus (SABV), Chapare virus (CHPV), Capixi virus (CPXV), Guanarito virus (GTOV), Amapari virus (AMAV). Lujo virus, Lymphocytic choriomeningitis virus (LCMV), Dandenon virus, Ippy virus, Lassa virus (LASV), Mopeia virus (MOPV) and Mobala virus (MBLV). Disease causing viruses are denoted by an asterisk.

The majority of haemorrhagic fever-inducing NW arenaviruses are in Clade B. In the phylogenetic tree shown in Figure 1 it is seen that viruses that share a common evolutionary ancestor do not necessarily

resemble each other with respect to pathogenicity [19]. For example, the haemorrhagic fever inducing LASV shares a common ancestor with MOPV and MBLV, which are not linked to human disease. The same is seen for the pathogenic GTOV which shares close evolutionary relations with the rather tame AMAV and CPXV. This presents a case that genomic comparison techniques need to evolve to be more useful in helping to predict emergence of highly destructive strains.

While there are many similarities between the OW and NW viruses, each group diverges with regards to receptor preference for cellular entry as well as immune elicitation, which may influence the resulting clinical symptoms in patients [19, 21]. OW arenaviruses LASV, LCMV and Clade C NW viruses have a glycoprotein that preferentially attaches to the cellular α-dystroglycan receptor as a means of host cell entry with clade B viruses making use of transferrin 1 and Lujo virus the neurophin -2 receptor [21, 22]. With regards to clinical symptoms, there are differences: for example, Junin virus (NW) which causes a cytokine storm in human pathogenesis, whereas Lassa virus (OW) elicits a more overall immune suppression [19].

1.4. Structural organization of arenaviruses

Arenaviruses are enveloped viruses whose genomes are made up of two ambisense negative-sense RNA strands. The short (S) segment encodes for the envelope glycoprotein precursor (GP-C) and nucleoprotein (N), whereas the matrix zinc binding protein (Z) as well as an RNA dependent RNA polymerase (L) are encoded by the long (L) segment [23-25]. The GP-C requires co-translational cleavage from a signal peptidase as well as cleavage by a site-1 protease (S1P) to produce the mature glycoproteins GP1 and GP2 to form the GP complex, allowing recognition of host receptors and subsequent endocytosis into cells [23, 26, 27]. From studies on Lassa virus it was found that after S1P cleavage the GP-C is presented as part of a trimer of heterotrimers: GP1 is a host cell attachment protein while GP2 facilitates membrane fusion during entry [28, 29]. A stable signal peptide (SSP) forms the third part of the heterotrimer and needs to be myristoylated to perform an integral role in facilitating pH–induced fusion into the host cell [30]. The GP1-GP2-SSP protomer complex has 11 *N*-glycosylation sites resulting in 25% of the total GP-C complex being made up of glycans [29]. The general genome organization of arenaviruses are depicted in Figure 2 as described. GP-1 interacts with host cell receptors while GP-2 acts to stabilise the binding via ionic interactions. The Z protein is seen to play a structural role within the virion, which aids in assembly and release of mature virions [23, 24]. Gene mapping has revealed a region of amino acids within the nucleoprotein of lymphocytic choriomeningitis virus (LCMV) that diminishes the type I interferon response (IFN) of the host—which is of pivotal importance to progression of infection [31, 32]. Inspection of the crystal structure of the

LASV nucleoprotein has resulted in documentation of a N- terminal m7GpppN mRNA cap binding site that allows the use of host cellular mRNA to bind and be used as a primer for the transcription of viral mRNA [32]. Thus, the nucleoprotein may be a desirable target for disruption of the arenavirus life cycle, as it contributes in multiple ways to virulence.

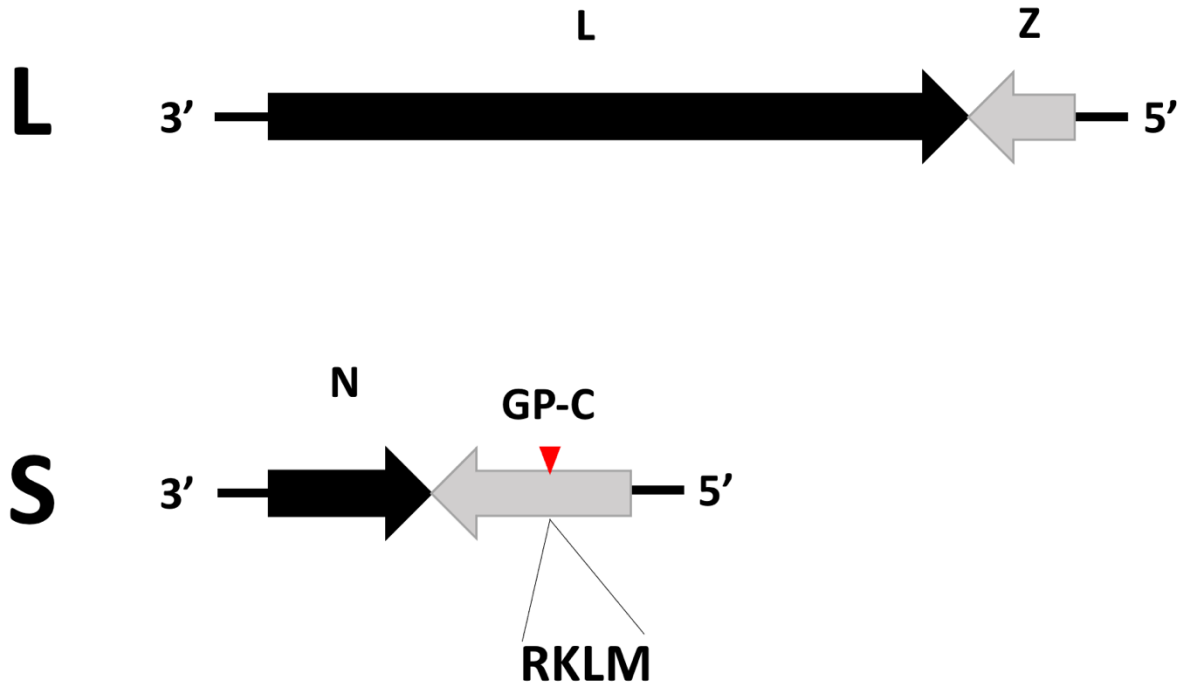


Figure 2: General genome organization of arenaviruses. Genome organization of arenaviruses adapted from Ferron et al. [23]. The Figure depicts gene sequences encoding the L and Z matrix proteins in the long segment (L) as well as the nucleoprotein (N) and Glycoprotein precursor (GP-C) in the short (S) segment. Sequences containing an open reading frame from viral RNA is shown in black while sequences whose mRNA is transcribed from copy RNA is shown in grey. A red arrow indicates the Site 1- protease RKLM cleavage motif that separates the glycoprotein precursor (GP-C) into GP1 and GP2. However the RKLM motif is only present in new world arenaviruses [33]. (note diagram not drawn to scale)

Cryo-electron microscopy (Cryo-EM) has been applied to characterising the structure of the two arenaviruses LASV and MACV. The results revealed that the proteome had much in common with influenza virus, with the L protein resembling other (-) sense viral polymerases as they possessed a catalytic centre with six motifs to enable polymerase activity [34].

1.5. General arenavirus reproductive cycle

Arenaviruses are not significantly distinct from each other with regards to cellular entry, protein expression, virion construction and release – however, there are some differences with regards to protein interactions that will not be detailed in this review. The general arenavirus lifecycle is detailed in Figure 3.

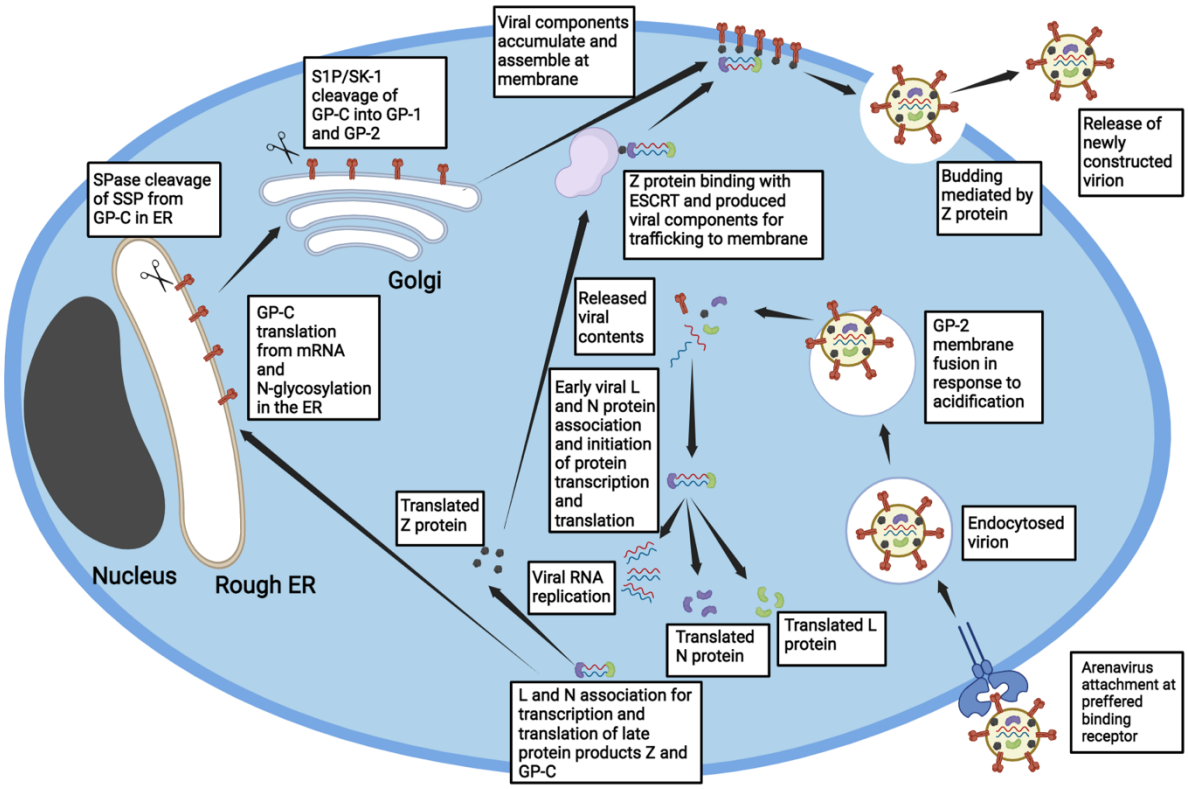


Figure 3: General arenavirus lifecycle. Most arenaviruses have a preference for either transferrin-1, neurophin-2 or α- dystroglycan receptors on the outside of host cells. The image shows the general pathway of viral attachment, fusion, processing, budding and release of new viral particles. Early protein products L and N tend to be expressed earlier while Z proteins and GP-C get expressed in the later stages of the cell takeover [21]. Details with regards to the post translational modifications (N-glycosylation and proteolytic cleavage) and cleavage products/associations such as the GP2/GP1/SSP complex are not shown. The diagram was inspired by [21] and was made in Biorender.com. It is not drawn to scale.

During cellular entry, the GP1 of arenaviruses bind to their preferred cellular receptor, either α-dystroglycan, transferrin 1 or neurophin-2, after which the viral particle is endocytosed [19, 21, 22]. In response to acidification, GP2, a class I fusion protein, undergoes conformational changes to form a six helix conformation, which results in membrane fusion of the viral envelope with the endosome, and release of the viral nucleic acids into the cytoplasm of the host cell [35]. All three Arenavirus Z, N and L proteins are subsequently transcribed and translated in the cytoplasm of the host cell [21]. The N protein and L protein (encoding an RNA–dependent RNA polymerase) released from the virion forms a complex that initiates viral gene transcription of the L and N proteins as well as mediates modifications for stability of the transcript through the insertion of a 5' mRNA cap [25, 32]. The N protein and L protein form ribonucleoprotein complexes with viral RNA [36]. The GP-C is targeted to the endoplasmic reticulum, whereafter it undergoes N-glycosylation and SSP cleavage by a signal peptidase enzyme leaving only the combined GP1 and GP2 [37]. Further proteolytic cleavage of the GP1/GP2 peptide is performed by the subtilisin kexin isozyme 1/ site 1 protease (SK1/S1P) in the Golgi network, releasing GP1 from GP2 [38]. This process is quite important as they need to be cleaved to form multimers which are crucial for their encorporation into the final synthesized virion particles [29]. In the late stages of the viral life cycle the translated viral proteins SSP, GP1 and GP2 associate to form a trimer of heterotrimers [29]. The Z matrix protein plays a crucial role in the final steps of virion construction as it interacts with the GP-SSP complex, N protein as well as the ribonucleoprotein complex [36]. The Z matrix protein also interacts with the endosomal sorting complex required for transport (ESCRT) which comprises a host of proteins involved in endosomal transportation and protein interactions [39]. Collectively, these protein-protein interactions allows translocation of all the viral components to the host cell membrane where the components assemble and bud into daughter virions that are released from the cell [21, 39].

1.6. Arenavirus clinical signs and diagnosis

Arenavirus infection manifests a spectrum of severity from mild infection to a fatal inflammatory fever, although asymptomatic infection has been documented as well. Humans become infected through inhalation of viral particles through freshly urine-contaminated soil [40], animal host bites or coming into contact with excretion / other body fluids from rodents that live around farm areas and find food and shelter in human dwellings [18, 41, 42]. Subsequent human to human transmission is then likely via aerosolized viral particles being taken into lungs where they can then bind macrophages

and other cells to start the infection [21]. Infection generally results in baseline symptoms of fever and different types of haemorrhaging, which can lead to shock and organ failure [15]. Other general reported symptoms are weakness, malaise, coughing, headaches; sore throats were seen in LASV infections and aesthenia, dizziness and mucosal rashes were observed in the South American arenavirus infections [43]. For Lujo virus, commonalities across all patients included fever, headaches, myalgia, sore throats and often a rash [14, 44]. The most severe pathogenesis of arenavirus infection is the development of haemorrhagic fever. In the case of Lassa virus, it is responsible for around 500 000 infections with a 1% mortality rate yearly in West Africa alone [43, 45]. In Lassa virus infections viraemia levels can escalate, resulting in a patient's innate and adaptive immune response being ineffective, which can be lethal [15]. JUNV, MACV, Sabia virus, CHPV and GTOV induce a similar haemorrhagic fever in certain regions in South America [45].

Due to the generalized spectrum of symptoms, early diagnosis of arenavirus infection is very difficult and is usually achieved through viral isolation, followed by reverse transcription polymerase chain reaction (RT-PCR) and enzyme–linked immunosorbent assays (ELISA) [41, 44, 46]. Immunofluorescence assays that detect serum antibodies binding to fixed virus infected cells may be useful in detecting antibodies against haemorrhagic fever-causing arenaviruses as each virus presents a unique fluorescence signature [46]. Additionally nucleoproteins of various arenaviruses have been produced in a recombinant baculovirus expression system including Luna virus, LCMV, LASV, LUJV, JUNV, GTOV, MACV and CHPV [46]. These can be used for IgG antibody detection assays [46]. Despite the availability of these assays, they may not be as useful in a large scale outbreak as positive results are only indicative of retrospective infection.

On a molecular level, viral propagation through cells has been shown to induce stress signals in cells causing the production of reactive oxygen species (ROS) as well as disruptions in physiological calcium homeostasis that can cross-activate apoptotic pathways resulting in cell death [47, 48]. LCMV has also been shown to inactivate host defences by disrupting the activity of type I interferon (IFN-I) which has important functions in host immunity [32, 47]. Numerous studies conclude that the host cell apoptosis and cellular damage is the primary cause of immunopathology seen with haemorrhagic fevers [45]. Interferons as well as other chemokines are crucial in the activation and maturation of lymphocytes such as macrophages, T- cells and dendritic cells, which all compose the first line of defence to dampen the impact of viral infection [49-51]. While these chemokines are highly useful and up-regulated in response to initial cell death, they can also be utilized by arenaviruses to sustain replication, causing further tissue damage resulting in lesions especially within the liver, lungs, spleen, kidneys, adrenal glands, heart and ovaries in JUNV, MACV and LCMV infected patients [45].

1.7. Arenavirus distribution and ecology

The *Arenaviridae* family comprises many viruses, not all of which cause human disease [15, 41, 52]. Although the general focus of this review is arenaviruses that cause haemorrhagic fevers as they present an outbreak threat, discussion of other arenaviruses is included briefly.

Arenaviruses often occupy animal hosts as their natural reservoir, each virus with some level of specificity for a chosen host [41]. Expectedly, the incidence of an arenavirus is often in line with the distribution and ecology of its animal host [13, 41]. Mammarenaviruses are found mosty in rodent hosts including mice, rats and even pets such as Syrian hamsters and guinea pigs have reportedly transmitted the virus [13, 41]. These rodents, once infected, often have lifelong persistence of infection and may be consistently contagious [53], even without evident symptoms of disease [19]. Lassa virus is found to be transmitted through a rodent *Mastomys natalensis*, a rodent found widely distributed throughout Africa, which is the predominant host reservoir [52]. LASV has a high prevalence in West African countries including Nigeria, Sierra Leone, Guinea and Liberia as well as surrounding areas [41, 52]. Other Arenaviruses including SABV, MACV, JUNV, GTOV and CHPV are found in Brazil, Bolivia, Argentina, Venezula and Bolivia respectively [19]. GTOV, MACV and JUNV were found to infect mice such as *Calomys musculinus,* which lives in a wide variety of habitats distributed across the Argentina — whereas some viruses such as SABV do not have an agreed-upon host reservoir [42].

Known viruses in the genus *Reptarenavirus* including *Alethinophid 1, 2* and *3* are known to infect snakes including tree boas and boa constrictors but it is not clear if they are necessarily the natural host reservoir. *Alethinophid reptarenavirus 1* and *2* are found in the United States of America while *Alethinophid 3* reptarenavirus is prevalent in Germany and the Netherlands.

1.8. Arenavirus outbreaks and incidence of disease

Of the broad array of arenaviruses classified, only a select few cause human disease and have outbreak potential. LASV, first characterized in 1969 [3], has been responsible for outbreaks in Sierra Leone and Liberia, most recently in Nigeria in 2018 and was at a time endemic between Senegal and Cameroon [15]. Lassa virus infects tens of thousands of people/year, remains endemic in West Africa, and has around a 30% mortality rate [3, 54]. GTOV was responsible for an outbreak in Venezuela in 1989 with repeat cases still ocurring up until 2006 [42]. In total there were 618 cases, 23% of which were fatal.

In 2004 Chapare virus was discovered in Bolivia, however, it was responsible for only one reported case that was fatal [19]. In the 1960s MACV was identified to be responsible for around 1000 cases and further down the line for 200 cases (several small outbreaks) of Bolivian haemorrhagic fever between 2007 and 2008 with a 20% mortality rate [19]. Around this time, in 2008, LUJV was also discovered independently and recognised as being responsible for an outbreak of 5 cases with a devastating 80% mortality rate in the Lusaka-Johannesburg region in Southern Africa [55]. JUNV, first isolated in 1958, [40] is responsible for causing Argentinian haemorrhagic fever in Pampa, Argentina and mortality rates have been shown to go as high as 30% [42, 56]. However, after implementation of a vaccine numbers were reduced from 300-1000 cases/year to 30-50 cases/year [19, 56]. All the arenavirus outbreaks seem to present themselves in countries which are still considered developing. These countries struggled to battle viral outbreaks in the past and will likely continue to do so due to their impoverished health systems. Poor health practices in these countries may also facilitate the increase of nosocomial outbreaks that may be more devastating, thus it is pertinent to develop affordable and effective methods to combat arenaviruses. They are also geographically seperated with regard to origin of OW and NW viruses, which may suggest some historical migration. Studies of the incidence and distribution of arenaviruses are severely neglected and as a result it is hard to predict where and to what extent a previously known arenavirus will cause an outbreak – or where a new one will emerge. This presents a case for the need for vaccines that will combat a broad spectrum of arenaviruses to reduce the possibilites of an outbreak.

1.9. Immunization as a strategy to protect against infectious diseases

Vaccines are the most cost effective way to combat infectious diseases. Their usage prevents the need for post –infection therapeutics and hospital overloading in the event of an outbreak where large scale hospitalization is needed. Beyond vaccines, outbreak control measures are limited to supportive and anti-viral therapy [57] and quarantine control measures. While these measures are useful, they are most effective if used in conjunction with vaccines, while antiviral drugs only become crucial treatments for viruses such as HIV and human cytomegalovirus where there is no licensed vaccine available [57]. In the case of viruses with great variation and high antigenic drift such as HIV and influenza there is a great possibility for antivirals becoming ineffective due to evolution of resistance, highlighting their inadequacies, and consequently a great dependence on vaccines to combat diseases [57, 58]. While this is a similar challenge for vaccines, these are more viable for a rapid response compared to the more gradual prolonged response effected by antiviral drugs. [59].

Over the last few decades great effort has been expended in an attempt to combat infectious diseases. Vaccines have shown great promise as a medical intervention to combat infectious diseases and the prevention of over 2.5 million deaths each year makes it arguably one of the best public health inventions in human history [60]. Many vaccines have been licensed including the recent rVSV-ZEBOV-GP vaccine for Ebola [8], multiple vaccines for hepatitis B [61] and human papillomavirus [62] among others.

Unfortunately in Africa, vaccine manufacturing capacity is limited and largely dependent on other countries, given that only 1% of vaccines used in Africa are manufactured locally in Morocco, Senegal, Tunisia and South Africa [63]. Such a strong reliance on other countries for vaccines results in their increased cost, a large delay in receiving doses even when time is critical during a viral outbreak and the potential of receiving vaccine doses close to expiry, as seen during the SARS-CoV-2 pandemic [63, 64]. These factors contribute to a poor response to disease outbreaks and an increased fatality rate in developing countries where the spread of disease tends to be the highest due to unsanitary conditions and inadequate healthcare systems.

1.10. Vaccines against arenaviruses

As mentioned previously, there are no licensed vaccines to combat arenaviruses. However, there has been some research carried out on their development by different strategies.

1.10.1. Live attenuated vaccines

Live attenuated vaccines (LAV) have been documented as an effective vaccine platform claiming, credit for the eradication of small pox, and the control of poliovirus and measles through the use of vaccination programmes [65]. LAV approaches involve taking an existing virus and modifying genes or other components to reduce the virulence factors before being administered to patients. The live virus will then elicit an immune response in the vaccinee with little to no pathology compared to natural infection.

A live attenuated vaccine for JUNV Candid #1 was developed through serial passaging to introduce mutations in the virus [66]. The vaccine was highly effective in conferring protection against JUNV infection in humans and had passed all the clinical trials to be licensed for human use [67]. Concerns

of safety with regards to reversion to virulence were expressed for Candid #1 as it was seen that the attenuation of JUNV relied solely on a F427I mutation in the GP2 subunit [66]. This affected the rate of adoption of Candid #1. However, there seemed to be genetic interactions between the mutated residue and a stable signal peptide that served as protective mechanism against reversion to virulence [66]. Attenuation of JUNV has also been documented after inducing a V64G mutation in the RING domain of the Z protein demonstrating multiple targets for attenuation – and therefore suggests multiple ways to increase the safety of the vaccine [68].

Furthermore, there are developments for a LASV live attenuated vaccine, where LASV was attenuated through the codon de-optimization of the GP-C [69]. The resultant candidate LAV, rLASV-GPC/CD, conferred complete protection against LASV infection, indicating the utility of this vaccine technology for use against arenaviruses [69].

1.10.2. Viral vector-based vaccines

Viral vector-based vaccines are a popular approach that utilizes attenuated viruses to deliver heterologous antigens to host cells [58, 70]. In the conventional serial passaging method of attenuation, the prospective vector virus is used to infect host cells and continuously passaged until the infectious genes of the virus are mutated making it relatively safe for use [58]. Once attenuated, other viral genes of interest can be inserted into the vectors and propagated to high viral titre levels [58]. The use of viral vaccine vectors has proved efficient in inducing an effective immune response, one that occurs during natural infection, has potential for great scalability to produce high viral titre levels and is generally safe for use [58]. The viral vector of the rVSV-ZEBOV-GP vaccine has recently been licensed for use against Ebola as well as ChimeriVax-JE against Japanese encephalitis and the Dengue vaccine Dengvaxia®- which comprises a yellow fever virus modified to encode dengue proteins, demonstrating the utility of this type of vaccine[8, 58, 71]. There are limited viral vector vaccine developments for arenaviruses; however, an attenuated version of the non-pathogenic MOPV, $MOPV_{ExoN6b}$, was modified to present proteins of some common pathogenic arenaviruses including SABV, GTOV, LUJV, CHPV, MACV and LASV [72]. The Lassa protein modified version of MOPV was shown to be promising as a vaccine for non-human primates and serves as a proof of concept for the technology to be applied more extensively to the other arenaviruses [72]. A known disadvantage is the possibility of diminished efficacy of vector vaccines in individuals that require multiple vaccinations due to host immunity to the vector. In the case of such patients, either a different vaccine approach would need to be used or the viral vector would need to be modified.

1.10.3. Virus-like particle vaccines

Virus-like particles (VLPs) refer to vaccines that resemble natural virions in the sense that they comprise the external structural components of a viruses assembled so they mimic the virion, but do not contain infectious genetic material [73, 74]. These particles can be extremely diverse, allowing the customizability of the surface with respect to charges, general shape and chemical modifications — which have been arguably revolutionary in the field of vaccines as they can be used to elicit a variety of immune responses [74, 75]. [76]. VLPs have an organised structure allowing repetitive epitopes for presentation to the immune cells making them effective in combating diseases [73]. Success in multivalency has been shown when a VLP was produced that presented 3 different HA subtypes and conferred protection in animal models against influenza [77]. VLPs are safe to use, providing a useful alternative to other vaccine platforms such as attenuated and whole killed vaccines that sometimes have undesirable effects and the potential for reversion to virulence [73, 78]. VLPs are usually stable in a cellular environment and due to their complex particle structure, have the potential to elicit a robust immune response, in contrast to soluble protein vaccines that can be more prone to degradation and can be mistaken for host proteins due to their lack of complexity [75]. The VLP platform has documented success and some vaccines have been licensed for commercialization including the SciBVac™ vaccine for Hepatitis B [79] and Gardasil9® for the prevention of human papillomavirus infections (HPV) [74, 76]. SciBVac™ has shown to be safe for use even in immunocompromised individuals and was shown to induce high antibody titre levels in the host even with low inoculation doses [79]. In the context of arenaviruses, a candidate VLP vaccine GEO-LM01 has showed utility in pre-clinical trials conferring 100% protection against lethal challenge with Lassa virus as well as induction of Lassa virus specific $CD8^+$ and $CD4^+$ cell responses [80]. GEO-LM01 was developed by inserting genes of LASV into attenuated modified vaccina Ankara virus (MVA) to express the LASV GP-C and Z proteins [80]. Expressed Z and GP-C proteins, assembled to form VLPs in the vaccinee [80]. Other attempts at incorporation of the Lassa glycoprotein complex into VLPs also elicited have shown to have immunogenic potential [81] including neutralizing antibody responses in rabbits [82].

1.10.4. DNA vaccines

DNA vaccines comprise naked DNA plasmids that harness host cell machinery following immunization to express a target antigen which induces an immune response [83]. Once injected into a host cell, DNA and in some cases immune factors encoded by the DNA expression system are targeted to the nucleus for subsequent transcription, and then cytoplasmic translation *in vivo* [84]. Once released into the cytoplasm, the antigenic proteins are then presented by professional antigen presenting cells to induce humoral and cell-mediated immune responses against the antigen [83]. DNA vaccines are safe as there is no risk of reversion to virulence as is the case with live-attenuated vaccines and there have been few side effects associated with their use in humans [83]. Additionally, DNA expression systems are easy to modify, cheap to synthesize and are relatively heat stable making them suited for use in poorer countries and for rapidly responding to disease outbreaks [83]. Despite the advantages of DNA vaccines, concerns have been raised that there may be risk of integration with the host genome or continuous unrestricted expression of undesirable proteins; previously, transient proteins have been shown to be expressed for up to a year post transfection [85].

DNA vaccines have shown to be effective against a variety of pathogens and have shown efficacy in various animal models including against West Nile virus in horses and infectious necrosis virus in salmon, although none have received licensure for human use [86, 87]. However, DNA vaccines have great relevance for human use and in the context of this review, have shown to be useful in inducing neutralizing antibodies in rabbit models to combat arenaviruses. Other studies involving mouse models supported DNA vaccines inducing protective cellular immunity against LCMV and PICV in mouse models [88].

1.10.5. Subunit vaccines

Subunit vaccines consist of viral components that are targets of the host immune response during natural infection. Unlike whole killed and attenuated virus vaccines, there is no risk of reversion to virulence or pathology anticipated for the use of subunit vaccines as there is no viral genetic material [78]. The main pathogenic effects of a virus are often due to changes in the cell from the viral lifecycle and immune evasion tactics in totality which is absent when using isolated parts [74, 89]. While traditional vaccine systems causes a more potent immune response, issues regarding their safety can be overcome using subunit vaccines and success has been shown in enhancing the immunogenicity of

subunit vaccines using immune-stimulatory adjuvants [90]. Here I explore the advantages and potential for virus like particles and soluble protein vaccines as an alternative vaccine technology.

1.11. Molecular pharming of pharmaceuticals

In Section 1.10 I discussed different common vaccine production approaches. For commercial production, these approaches are generally performed using mammalian cell expression systems that require expensive equipment and reagents, and are sensitive to contamination especially when using multiple viral vectors or LAV in the same production facility.

Molecular pharming of recombinant proteins is an increasingly popular platform to produce biopharmaceuticals. Plant-based expression requires a less expensive infrastructure than mammalian cells to establish a cGMP-compliant manufacturing facility, and the production costs are similarly lower. In addition, plant-based manufacturing is also suitable for rapid large scale production of vaccine antigens [91-95] — which is critical to respond to disease outbreaks. This is especially appealing for resource-limited regions which lack the infrastructure to manufacture pharmaceuticals and bear a large burden of infectious disease [92]. In contrast, mammalian expression systems, a staple approach of previous vaccine research, requires a huge capital investment attributed to the need for a cell fermentation facility, a continuous need for very specialized skills (compared to genetically modified plants that can be farmed with less technical knowledge after initial setup), as well as a need for expensive highly sterile enclosed conditions [92]. In contrast plant systems utilize a cheaper non-sterile, open environment Lastly, plant-based pharmaceuticals do not risk accidental contamination with live mammalian viruses as these viruses require mammalian hosts.

Nicotiana benthamiana is becoming the staple host for plant-based protein expression due to its high biomass, ease of cultivation, inefficient gene silencing and compatibility with CRISPR/Cas 9 systems [96, 97]. Extensive research has been conducted using *N. benthamiana* as a model system to characterize different aspects of protein expression. This has resulted in improved yields and the production of correctly folded proteins—making it a good choice for an expression host [98].

Research attempts have showed some promising plant-based vaccine technology including a quadrivalent virus-like particle vaccine for influenza that passed the clinical trials up till phase three, as well as a triple monoclonal antibody mix (ZMapp) for the treatment of Ebola virus infections [99-101]. Other cases of use of plant-based technology include the SARS-CoV-2 vaccine candidate RBD121-

NP [102] and Medicago Inc.'s plant based SARS-CoV-2 vaccine Covifenz which was licensed for use in Canada but unfortunately had adoption hesitancy elsewhere due to political reasons [103]. These served as a proof of concepts for the utility in plant based protein production.

Although many proteins can be expressed at useful levels in plants by stable and transient transfection — even up to 0.5g/kg of fresh leaf weight - certain viral glycoproteins that are potential vaccine antigens are expressed at low levels and may also not undergo crucial post translational modifications (PTMs) that occur in mammalian hosts and which are necessary for an appropriate immune response [104, 105]. Viruses utilize host enzymatic machinery to introduce PTMs that are crucial for viral protein recognition, replication, protein synthesis and elicitation of an immune response [106]. Arenaviruses are devoid of PTM enzymes encoded in their RNA and are thus reliant on the host cells for PTMs such as glycosylation, phosphorylation and methylation, proteolytic cleavage and chaperone–mediated folding all of which can affect their pathogenesis [106]. This dependence placed on the host environment makes it necessary to address PTM irregularities when attempting to produce viral proteins in their non-native host, as discrepancies may result in dysfunctional proteins.

1.11.1 Molecular engineering strategies to support the production of complex biomolecules in plants

Since vaccines operate on the premise of antigenic proteins inducing immunity in a host, inappropriate PTMs remain a potential obstacle to the viability and adoption of plant expression systems for routine vaccine production [107]. There is, however, an increasing understanding of the challenges of producing commercially relevant proteins along the plant secretory pathway [92], and several of these bottlenecks to recombinant protein production have been addressed, increasing the complexity of proteins that can be made in the plant expression system [104]. Figure 4 shows an overview of general protein processing in the ER that mirrors the path that transiently produced mammalian proteins follow. For native mammalian proteins to be produced correctly in plants, differences between the plant and mammalian environment need to be addressed so that appropriately humanised vaccine antigens can be made.

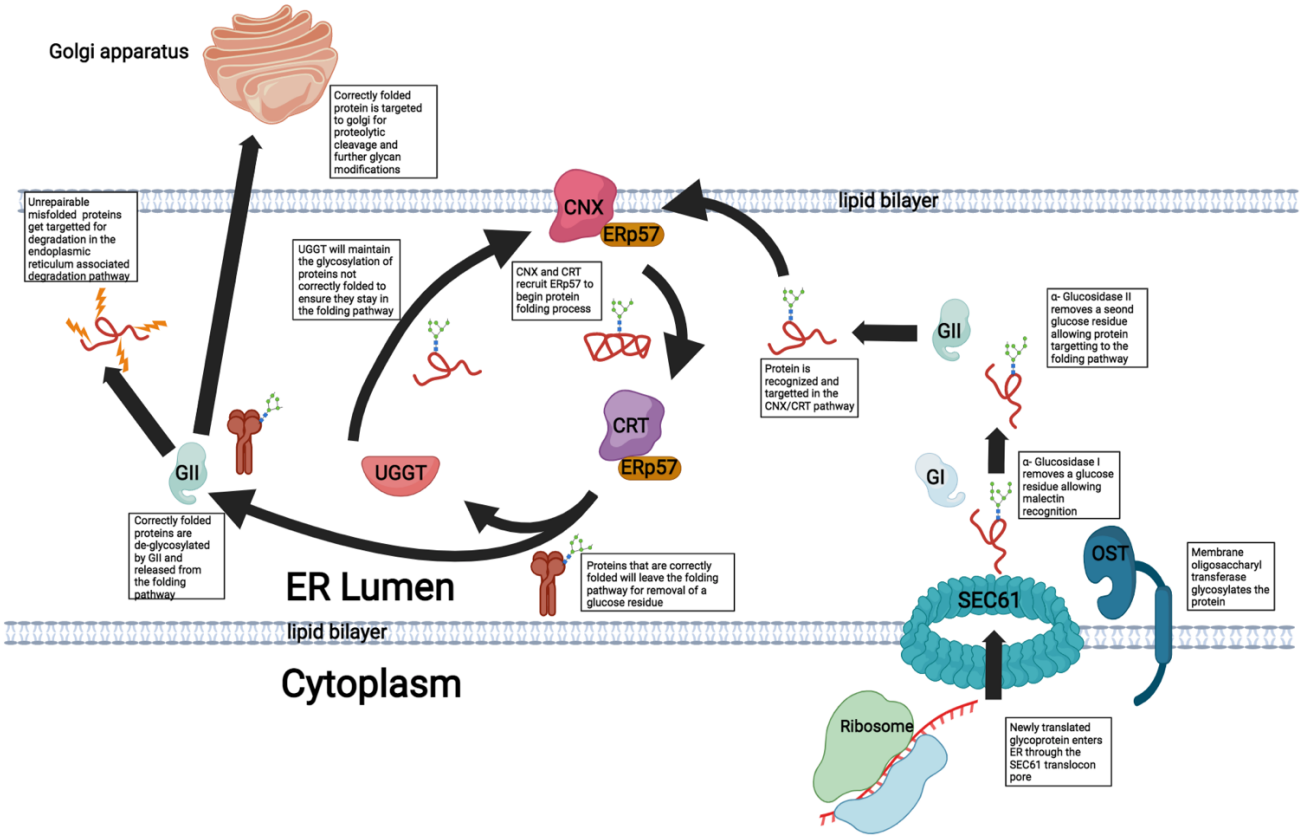


Figure 4: Overview of protein processing along the secretory pathway. The Figure shows the basic protein modification pathway as expected in a mammalian environment. The main aspects of the secretory pathway include protein glycosylation, protein folding, and targeting to the Golgi apparatus and proteolytic cleavage. Glycosylation tends to be one of the more important modifications as it occurs immediately after translation and affects all downstream protein modifications. This also seems to be a point of divergence between plant and mammalian cells and therefore should be carefully considered when attempting to produce mammalian proteins in a plant environment. The general process it as follows: At the ER surface, newly translated proteins enter the SEC61 translocon pore and are glycosylated by an oligosaccharyltransferase (OST). The protein undergoes removal of 2 glucose residues in succession by α-glucosidase I (G-I) and α-glucosidase II (G-II), respectively. This leaves the protein for recognition by lectin chaperones calnexin (CNX) and calreticulin (CRT) that engage the oxidoreductase ERp57 and collectively begin the protein folding process. Proteins correctly folded are further de-glycosylated by G-II making them unrecognizable by the molecular chaperones and targeted along the pathway to the Golgi for proteolytic cleavage and glycan modification. Proteins that fold incorrectly are re-cycled in the pathway by UDP: glucose glycoprotein glycosyltransferase (UGGT) that ensures that they remain glycosylated

until they are folded. However, should the protein become misfolded and unsalvageable it is then degraded in the ER-associated degradation pathway. Figure adapted from [108] and created in biorender.com. This diagram is not drawn to scale.

1.11.2. Importance and challenges of glycosylation for plant produced viral proteins

Glycosylation occurs in both plant and mammalian cells. The two major types of glycosylation, *N*-glycosylation and *O*-glycosylation, function in many common cellular processes including cell-cell signalling and protein folding [109-111]. *N*-linked glycosylation refers to the attachment of sugar groups (glycans) to asparagine residues of proteins [110]. *N*-glycans play a crucial role especially in the context of viral glycoproteins as they direct processes such as fusion, recognition, transport of proteins, binding to cellular receptors and mitigation of antibody responses [110, 112]. Mutations in some glycosylation sites of MACV resulted in an attenuation of the virus, showing the importance of glycosylation in viral pathogenesis [113]. For plant produced recombinant proteins to be immunogenic against human viruses it is necessary that they are modified with human-like glycan arrangements, as these viruses have evolved to propagate and navigate in a mammalian cell environment [97, 114].

N-glycans are important for recognition by molecular chaperones [110]. Without correct glycosylation, proteins may be misfolded which can result in protein aggregates and ER stress/ necrosis in plants [32, 110].

While the early stages in the glycosylation process remain similar in the ER of both plant and mammalian cells, the site of complex glycan formation at the Golgi apparatus has diverging glycan formation under normal circumstances [115]. The inability of plant cells to generate mammalian-like complex glycans can be attributed to a lack of glycosyltransferases similar to those seen in mammalian cells. Furthermore, certain plant specific glycan arrangements not seen in mammalian cells may be adversely immunogenic for proteins that are intended for use human use [116]. Figure 4 shows that glycosylation by the OST complex is the first modification after entry into the ER. Should the glycosylation not occur in a mammalian-like manner, the subsequent modifications that depend on glycan recognition may be aberrant and in some cases even detrimental to the plant tissue.

In some cases, such as with human IgG Fc protein, glycan sites can be underglycosylated *in planta* compared to expression in their native hosts [116]. Differences in glycosylation efficiency between organisms is likely due to different recognition preferences of oligosaccharyltransferase (OST)

complexes that catalyse the early stages of N-glycosylation [115, 117]. Recently, a case success has been shown in increasing the glycan occupancy of recombinant proteins through co-expression of an OST, LmSTT3D from *Leishmania major* —serving as a proof of concept for future co-expression experiments [116, 118]. However co-expression with OSTs needs to be implemented with caution as it's possible they could disrupt other pathways and have unintended effects that may reduce the protein yield obtained [104].

Another crucial type of glycosylation is *O*-glycosylation by N-acetylgalactosaminyltransferases. In contrast to N-glycosylation, *O*-glycosylation of recombinant protein is much harder to achieve in plant systems as the entire pathway is absent in plants and needs to be introduced. *O*-glycans however were not explored in great depth in this review.

1.11.3. Engineering the proteolytic machinery

As outlined in section 1.5 describing the arenavirus lifecycle, proteolytic cleavage is needed for viral glycoprotein cleavage, protein maturation and incorporation into viral particles. To date, the proteases furin as well as site-1 protease (S1P) are not known to be present in the *N. benthamiana* secretory pathway and require artificial co-expression if needed [119, 120]. Co-expression of furin has shown to be effective in the production of mature *Bacillus anthracis* protective antigen 83 (PA83), transforming growth factor beta (TGF-B), blood clotting factor IX [119, 121] as well as the HIV gp140 antigen [122]. It is not known if other proteases that are required for glycoprotein processing are present in plants. The site-1 protease/ SK-1, for example, is integral to cleavage of arenavirus glycoproteins yet it is unclear if the enzyme is present along the plant secretory pathway [38]. Without proteolytic cleavage, the SSP-GP1-GP2 complex cannot form and will thus not be produced as soluble proteins or virus like-particles in plants, making this a necessary factor for consideration when expressing mammalian viral glycoproteins in plants [29].

While not always the case for all recombinant proteins, a potential factor limiting high protein yields in plants is the presence of endogenous plant proteases and protein degradation pathways such as endoplasmic reticulum associated degradation (ERAD) that destroy non-native proteins [123, 124]. As shown in Figure 4 this occurs only to terminally misfolded proteins, and as mentioned, the incidence of this increases with underglycosylation/incorrect glycan groups re-emphasizing glycosylation as a pivotal step.

However, commercially-relevant increases in recombinant accumulation may be achieved by targeting aspects of the plant degradation environment, including targeting protease cleavage active sites/ silencing proteins, modifying cleavage sites/ stabilizing target proteins, and the use of protease inhibitors (PIs).

Attempts to express monoclonal antibodies such as anti-tumour mAB H10 and anti-HIV mAB 2G12, showed that the protein degradation in plants was limited to certain antibody regions [125]. These findings elucidate potential target cleavage sequences that can be modified to avoid recognition by endogenous proteases. Moreover the susceptibility of recombinant proteins for degradation in plants can be reduced by the co-expression of protease inhibitors (PIs) such as HsTIMP, cystatin SICY58, NbPot1 and NbPR4, all of which have shown to reduce protein degradation [126].

An additional strategy to mitigate unwanted degradation in plants is to silence relevant protease genes through gene knockdown of protease genes or by co-expression of protease inhibitors, however, this may have unintended consequences on the host plant [123]. This has been proposed and implemented using RNA-interference approaches to silence C1A proteases [127] as well as the CysP6 protease encoding gene [128] — both of which resulted in an increase in recombinant protein accumulation (albeit marginal) relative to unmodified plants.

Removal of proteases may compromise the structural integrity of proteins as proteolysis is an important factor in the prevention of pathologies related to incorrect protein folding [129]. Efforts should be directed to characterization of disruptive proteases and their protein targets so that these approaches can be optimized for efficiency as there is great mystery surrounding protease targets, especially given the large protease repertoire in plants (1243 confirmed in *N. benthamiana*) [130]. Until this is better classified, novel recombinant protein expression studies will need preliminary testing to mitigate protease interference — which is arguably a factor prolonging the full embracing of plant based expression systems. However co-expressing PIs may undermine crucial viral maturation proteases such as furin/S1P and this may disrupt viral protein maturation.

1.11.4. Improving glycoprotein folding in plants

As shown in Figure 4, protein folding of arenavirus glycoproteins in the ER is guided by chaperones such as Binding immunoglobulin protein (BiP) and calnexin/calreticulin (CNX/CRT) [21, 129]. This is to ensure that the proteins are properly folded and to remove incorrectly folded protein [129]. Failure

to do so can cause ER stress and, in the context of plant produced proteins, can cause tissue necrosis, reducing the both yield and structural integrity of proteins harvested [131], which is of utmost importance in the context of commercial relevant protein production.

To ensure smooth execution of these processes, proteins need to be correctly folded into a native conformation and with the relevant PTMs (especially glycosylation). The secretory pathway spatially separates modifications in a checkpoint-like manner meaning glycan arrangements need to be relevant for progression of proteins [108]. To support protein folding, cells have evolved to make use of molecular chaperones such as BiP, calnexin (CNX) and calreticulin (CRT) to aid in the folding of newly transcribed proteins. CNX binds membrane proteins within the ER membrane and guides the maturation of most glycoproteins while CRT, a soluble chaperone, is based within the ER lumen and binds soluble proteins [132, 133]. Binding of both CNX and CRT to proteins is dependent on their specific glycan arrangement — suggesting that glycosylation needs to occur correctly and prior to recognition for folding. CNX and CRT can interact with other chaperones allowing the catalysis of disulfide bonds, thus tertiary and quaternary protein structures, with the aid of the enzyme protein disulfide isomerase (PDI) [133]. Should protein refolding fail, it is unsurprising that the CNX/CRT pathway is also involved in targeting permanently misfolded proteins for ER–associated degradation [132, 134]. Collectively this highlights the relationship between glycosylation, misfolded proteins and the effects on a plant/ mammalian host [135]. Due to differences in molecular chaperone availability between plant and mammalian cells, co-expression of CRT/CNX has been proposed for transient protein production, and has shown an improvement in glycoprotein accumulation, and a reduction of ER stress relative to control samples. This held true for the HIV gp140 antigen as well as glycoproteins from other viruses including chikungunya virus, Rift Valley fever, SARS-CoV-2 as well as Epstein-Barr virus [104, 136].When aiming to produce glycoproteins, it is reasonable to co-express CRT/CNX in plants despite these enzymes being natively present as they may accumulate at lower levels than what will be needed for folding.

1.11.5. Introducing additional auxiliary cellular machinery and plant modifications to support more mammalian-like post-translational processing

There are a complex array of factors and many differences between mammalian and plant environments. In the context of heterologous protein expression, PTMs including glycosylation and proteolytic processing needs to be prioritized to ensure that the proteins are modified in a way that

elicit the correct response based on the context of their intended use [137]. In cases where certain PTMs do not occur in plants or at insufficient levels, the requisite cellular machinery can be transiently co-expressed to support protein production and proper folding. It is relevant that there are auxiliary aspects of transient protein production that can also be taken into account. For example complex N-glycans tend to be undesirably cleaved by endogenous β-*N*-acetyl-hexosaminadases [138]. This can be prevented by co-expressing sequences that will elicit RNA interference (RNAi) and degradation of these enzymes leaving more complex glycans on transient protein [138]. Alternatively, proteins can be expressed in plant tissue lines that are modified for certain aspects such as ΔXT/FT plants that are deficient for plant specific α-1,3-linked fructose & β-1,2 linked xylose.

Lastly, PTMs such as sulfation and γ-carboxylation have been shown to be relevant in certain cases for biopharmaceutical production and are useful to consider [137]. Addressing the protein maturation constraints along the plant secretory pathway will allow for the production of complex proteins making it easier to produce vaccines that are more effective.

1.12. Project rationale

LUJV is an arenavirus and a potentially emerging threat that is known for causing a singular but severe haemorrhagic fever outbreak in the sub-Saharan regions of Lusaka and Johannesburg. Due to only one known outbreak of LUJV, there is little patient data available aside from that which was reported by the National Institute of Communicable Diseases, Johannesburg, South Africa [139]. Five people were infected, and four died, for a mortality rate of 80%. The onset of the fever is characterized by non-specific symptoms, organ failure and bleeding, making it difficult to diagnose and promptly apply quarantine measures. In a study attempting to find the reservoir for LUJV, two novel arenaviruses were discovered in the Lusaka region in Zambia [14]. The exact source from which the primary patient contracted the virus remains unclear. Further investigation revealed the patient may have been in contact with dogs, horses, cats or rodents, and a tick bite may have been implicated as well [14, 44].

LUJV is not always easily classified in the new or old world groups of *Arenaviridae* [140] despite the classification into the Old World group by [19] in Figure 1. This is attributed to LUJV being unique in using a neurophin 2 (NRP2) receptor as a means of entry into host cells as opposed to the α-dystroglycan and transferrin-1 receptors recognized by the old and new world arenaviruses, respectively [22, 140]. A highly virulent virus that evolved a new entry route for host infection is a

great cause for concern and brings up questions into the driving forces that result in these evolutionary differences.

Additionally, there is no diagnostic test currently available to identify LUJV, and the virus was characterised during the 2008 outbreak by means of full genome pyrosequencing [14, 55]. This is a time consuming and impractical response. With no current licensed vaccine for LUJV, there is an urgent need to produce effective immunogenic antigens and explore possible types of treatment against it and other arenaviruses.

The LUJV symptoms are non-specific and accompany numerous pathological conditions, making the diagnosis of this viral haemorrhagic fever difficult— re-enforcing the need for the development of relevant diagnostic tools and full characterization of the symptoms caused by LUJV infection.

The current conclusions from the treatments used and outcomes of the patients is highly fragmentary and may prove to give minimal insight into preparation for another potential outbreak, especially since the majority of the treatment used was focused on symptomatic treatment and general supportive care interventions. Attempts have been made to develop diagnostic tools for arenavirus infection such as ELISAs and the use of monoclonal antibodies to detect LASV and JUNV NPs. The diagnostics are very specific, and lack broad spectrum detectability amongst arenaviruses [46]. There is of course utility in targeted detection, as it allows discernibility between other arenaviruses.

The outbreak revealed that general health facilities are not equipped to respond quickly enough to dampen the mortality rate. A diagnostic tool and vaccine are viable solutions to respond to outbreaks ,however, there is limited manufacturing capacity and funding for this in Africa.

Plant molecular pharming of recombinant proteins is a sensible solution to develop vaccines at low costs for developing countries. Research efforts are increasingly directed to optimizing the production and quality of proteins produced in *N. benthamiana* and many previous constraints related to protein folding, post translational modification and prevention of proteolysis have been addressed. These approaches provide an alternative to mammalian cells, which are more expensive and require a sophisticated infrastructure [104, 141].

Arenaviruses naturally infect mammalian hosts [14, 22, 140]. In the context of this project It is useful to express recombinant proteins in mammalian cells in parallel with plant cultures to assess protein yield, compare glycosylation profiles and consider the cost differences between the two systems. Mammalian cells are the accepted host platform for this type of research and thus will concomitantly

allow insight into the level of effectiveness of different co-expressed proteins. Systematically addressing folding constraints in plants will allow the production of appropriately folded antigens, and when applied to the production of Lujo virus, will serve as a proof of concept for other fatal emerging arenaviruses in Africa and elsewhere.

1.13. Aims and objectives

The main aim of this project was to develop a cost-effective and scalable protein expression system for the production of recombinant LUJV glycoprotein antigens in plants. The secondary aim of this project was to clone and express these recombinant genes into a mammalian expression system for a comparison of feasibility with the plant expression system. The project was designed to apply novel molecular engineering approaches to support the production of the viral glycoprotein in plants. The plant-produced proteins were compared to the equivalent mammalian cell-produced proteins to identify differences arising from production in plants.

Objectives

- Design synthetic constructs encoding glycoprotein genes from LUJV for expression in plants and mammalian cells.
- Construct recombinant expression vectors for the heterologous expression of LUJV viral antigens in plants and mammalian cells.
- Generate recombinant *A. tumefaciens* strains for transient expression of heterologous antigens in plants.
- Verify transient expression of heterologous proteins in plants and mammalian cells.
- Systematically improve production of LUJV antigens in plants by co-expression of human chaperones to support folding, site 1 protease to support processing and heterologous oligosaccaryltransferase (OST) enzymes; using mammalian cell-produced antigens for comparison.
- Develop and implement a purification strategy to recover the soluble LUJV glycoprotein from plant and mammalian cells.
- Determine the site-specific glycan occupancy and complexity of plant-produced and mammalian cell-derived glycoproteins.

Chapter 2: Development and optimization of a mammalian protein expression system for the Lujo virus glycoprotein

2.1 Introduction

The main aim of the research described in this chapter was to develop a mammalian protein expression system for the production of the Lujo virus precursor glycoprotein (LUJV GP-CΔTM) antigen as a vaccine candidate or a diagnostic tool for Lujo virus. As detailed in chapter one, the GP-CΔTM glycoprotein precursor protein should be cleaved by a signal sequence cleaving Spase I and Site-1 protease (S1P) to produce GP1 (a receptor binding protein) , GP2 (a membrane fusion protein) and a stable signal peptide (SSP) all three of which combine to form a heterotrimer [21]. This study tested the transient expression of the S1P in a mammalian system despite this protein being endogenously expressed. This was tested because it was not clear if the natural levels of S1P expression was sufficient for proteolytic cleavage of proteins or if overexpression may improve yields in some way.

Mammalian expression systems are the more conventional systems used for recombinant protein expression which may have higher yields and lower production issues relative to plant systems but are expectedly more expensive. In this chapter, we tested the expression of LUJV GP-CΔTM and compared some protein dynamics with expression using plants (described in chapter 3).

2.1.1 HEK293T cells as a baseline protein production host cell line

Human embryonic kidney cells (HEK293T) are suitable protein expression hosts for mammalian viral proteins and are widely accepted as a research tool due to their efficacy in recombinant protein expression, less demanding growth conditions [142], the large body of research surrounding mammalian protein expression [143] and compatibility with the novel pTHpCapR expression vector (further discussed in 2.1.2). Viral proteins have evolved to replicate and propagate in those mammalian cellular conditions naturally, making protein folding, cleavage and post translational modifications more realistic in terms of the viral lifecycle. In the case of prophylactics and therapeutics, correct protein conformation and modification is crucial in eliciting meaningful immunogenicity in a vaccinated host, which is not always as easy to achieve in plant cells.

2.1.2 pTHpCapR transfection of HEK293T cells

pTHpCapR is recently developed DNA vaccine vector showing great utility in preclinical vaccine experiments and protein production that was modified by Tanzer et al. [144] and further described by R. Chapman and E.P Rybicki [145]. This expression vector utilizes a porcine circovirus type 1 (PCV-1) enhancer element as well as a PCV-1 capsid protein promoter (pCap) that have shown great increases in recombinant antigen production and immunogenicity in the context of HIV-1 polyprotein (*grttnC*) production in macaques, rabbits and mice at a 4-10-fold lower DNA dose than is generally used. The backbone vector pTH was chosen by the developers of pTHpCapR due to its potency as an antigen expressing agent in previous studies and trials.

2.1.3 Cost of mammalian protein production

An important aspect of this study a is the cost implications of vaccine production since the focus is on mass production in lower socio-economic regions. It is difficult to accurately estimate the costs of vaccine production due to the variable costs in a vaccine production plant, including the cost of equipment, running costs and electricity in the region, cost of consumables and delivery to the area, environmental conditions that may influence the experiments, wages in the region of interest and policies regarding ethics and licensing. Manufacturing condition adjustments based on these factors and more can result in drastic changes in vaccine production costs, however, an estimation can be provided based on the particular costs in this study at the time of research.

Figure 2.1 summarizes a general workflow in mammalian-based antigen expression and gives general estimates related to the costs involved in each step (with the estimations being closely representative of this particular research project).

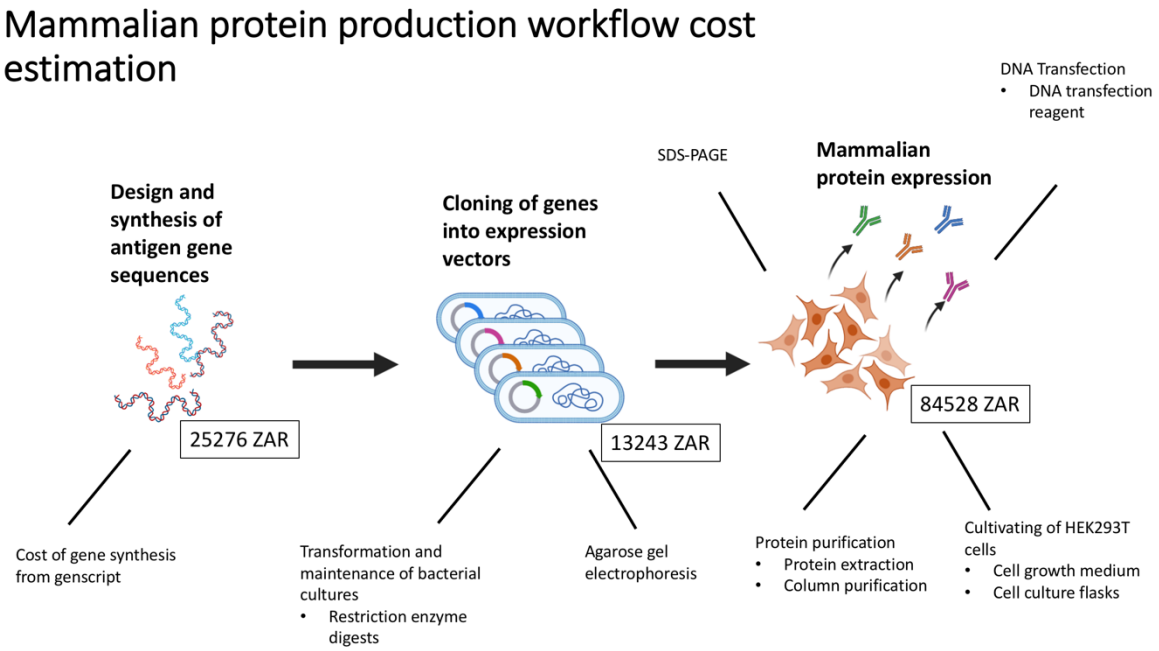


Figure 2.1: Cost estimation of consumables and variable factors in mammalian protein production using the pTHpCapR-HEK293T protein expression system. Above outlines general cost estimation over the main steps used in this study to produce LUJV GP-CΔTM antigen from HEK293T cells. The variables represented above are estimates based on the particular methods used in and at the time of this study. They are not necessarily representative of a mass production attempt although they may provide an indication for it. The cost estimations above also focus on project specific sequences and consumables and do not take into account general routine equipment available for scientific research.

2.2 Materials and methods

2.2.1 Design and synthesis of LUJV GP-C and S1P gene sequences

For mammalian protein expression in HEK293T cells, the pTHpCapR vector was chosen to express the LUJV glycoprotein precursor as well as the Site-1 protease., The S1P (Accession: Q14703) and LUJV GP-C (Accession: C5ILC1) sequences were obtained from UniProt. The transmembrane (TM) region in the LUJV GP-C was not clearly defined in UniProt or GenBank and was therefore inferred by performing a sequence alignment with the GP-C of other arenaviruses (**Figure 2.2**) for which the TM region was annotated. The inferred transmembrane region of LUJV GP-C was then removed to allow it to be secreted and thus easily harvested from the HEK293T growth medium.

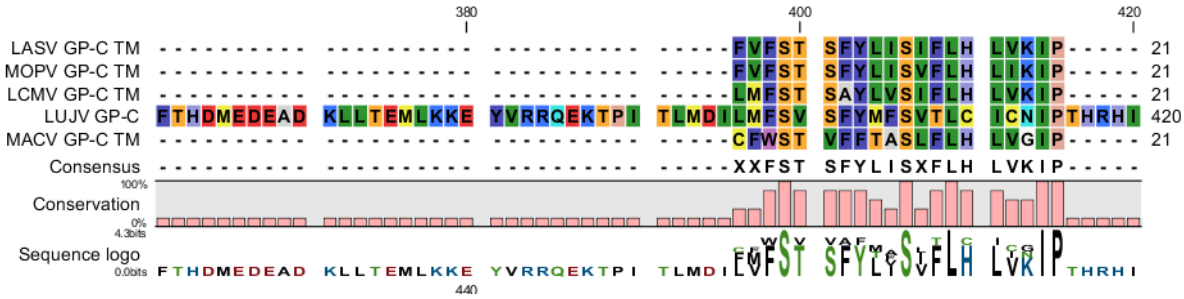


Figure 2.2. Sequence alignment of LUJV GP-C with transmembrane regions of related arenaviruses. A sequence alignment-based approach to find the transmembrane region of LUJV GP-C. LASV- Lassa virus, MOPV- Mopeia virus, LCMV- Lymphocytic choriomeningitis virus, LUJV-Lujo virus, MACV-Machupo virus. Amino acid region 396-415 was removed from the LUJV-GP-C to prevent it from being membrane bound when expressed in host cells.

Both the LUJV GP-CΔTM and S1P gene sequences were modified to include a Kozak sequence to serve as a protein translation initiation site. Additionally, restriction sites were added to the 5' and 3' end of each sequence to allow for convenient restriction endonuclease digests in the subsequent cloning steps. *BshT*1 and *Hind*III were added to the proximal region of both sequences while *Eco*RI and *Xho*I were inserted at the distal end of the sequence. A His tag motif preceded by a flexible linker sequence was added to the 3' end of the LUJV GP-CΔTM gene sequence to allow for protein purification using a cobalt affinity column that binds the His tag. The LUJV GP-CΔTM sequence has a stable signal peptide at the 5' end that may end after residue 59 [55]. The S1P was designed to include its native signal

peptide sequence with the same RE sites on the 5' and 3' terminus as LUJV GP-CΔTM. **Figure 2.3** below illustrates the LUJV GP-CΔTM and SIP sequences that were synthesized by GenScript. The LUJV GP-CΔTM and S1P sequences were received having been inserted into pUC57 vectors.

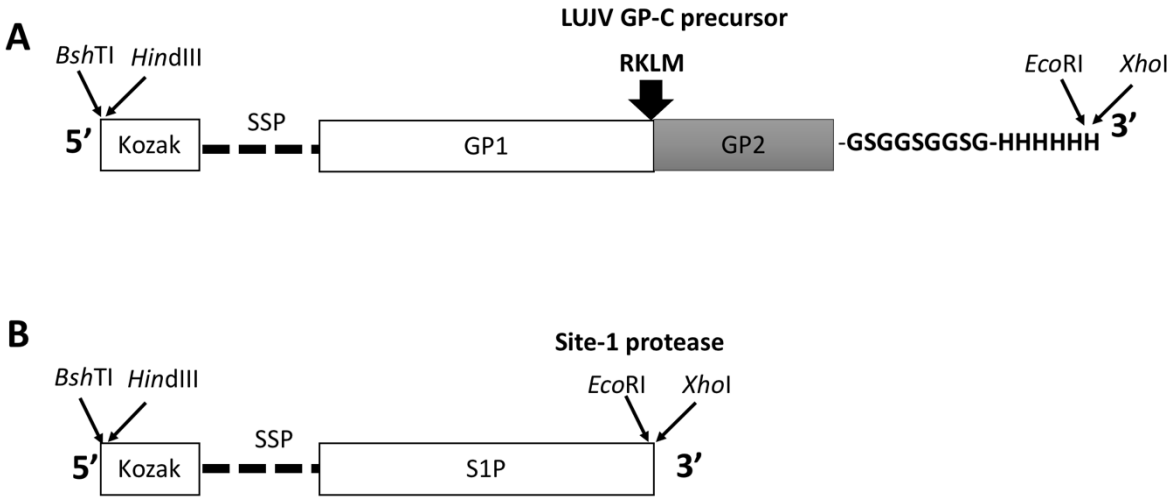


Figure 2.3: A Gene design of LUJV GP-CΔTM modifications before cloning into pTHpCapR. Representation of the modified LUJV glycoprotein precursor gene sequence**.** The gene sequence excluded the transmembrane region found in the LUJV GP-C and was modified to include a stable signal peptide (SSP), a Kozak sequence, restriction sites: *Bsh*TI, *Hind*III (5' end) and *Eco*RI, *Xho*I (3' end), a linker sequence and a His tag. GP1 and GP2 are the two cleavage products that are produced when the S1P recognizes and cleaves the RKLM motif. **B: Gene design of S1P modifications before cloning into pTHpCapR.** Representation of the Site 1-protease gene sequence ordered from GenScript. The S1P sequence was modified to include a signal peptide, restriction sites*: Bsh*TI, *Hind*III (5' end) and *Eco*RI, *Xho*I (3' end),

2.2.2 DNA restriction digests

To make final expression vectors pTHpCapR: LUJV GP-CΔTM and pTHpCapR: S1P, the constructs Puc57: LUJV GP-CΔTM, Puc57:S1P as well as pTHpCapR were digested using Fast digest restriction enzymes (ThermoFisher, USA) *Hind*III and *Eco*RI. The basic restriction digest components and concentrations for all 3 plasmids were the same: 1 μg of DNA was digested for each construct and reactions were performed in 50 μl at 37°C for 1 hour. De-glycosylation digests were performed using the protein deglycosylation mix (New England Biolabs, UK) and PNGase- F (New England Biolabs, UK). 100 μl of

protein was used and processed according to the denaturing method in the manufacturer's recommendations.

2.2.3 Agarose gel excision of DNA fragments for cloning

To assess the success of the cloning of LUJV GP-CΔTM and S1P restriction enzyme-digested DNA was electrophoresed on agarose gels. The fragments of the appropriate sizes were visualized using long wave ultra-violet light (365nm) to reduce as much DNA damage as possible. The DNA fragments were then excised using a scalpel and purified using a QIAquick® gel extraction kit (Qiagen, Germany) according to the manufacturers recommendations before being subject to ligation reactions.

2.2.4 Agarose gel electrophoresis

All agarose was dissolved in 1X TAE. The gels were then visualized on a Molecular Imager® Gel Doc™ XR+ imaging system (Bio-Rad, USA). The ThermoFisher scientific 1KB gene ruler (REF: SM0311, ThermoFisher, USA) was used as a DNA molecular weight marker on the agarose gels.

2.2.5 Ligation reactions to construct the pTHpCapR: LUJV GP-CΔTM and pTHpCapR: S1P vectors

Ligation of compatible DNA fragments into pTHpCapR was set up as follows: All reactions utilized an insert:vector backbone ratio of 3:1. The components were mixed with T4 DNA ligase and buffer (ThermoFisher, USA) before incubation for 1 hour at room temperature. The ligation reaction was performed at room temperature for 1 h before the DNA was used to transform *Escherichia coli* (*E. coli*) (see section on bacterial transformation)

2.2.6 Sequencing of pTHpCapR: LUJV GP-CΔTM and pTHpCapR: S1P

The cloned plasmids pTHpCapR: LUJV GP-CΔTM and pTHpCapR: S1P were sequenced by Macrogen to confirm correct cloning and integrity of the plasmids.

2.2.7 Bacterial transformation and growth conditions

2.2.7.1 *Escherichia coli DH5α* DNA transformation and culturing conditions

Escherichia coli DH5α (*E. coli*) was stored at -80°C and transformed according to the following guidelines: 1-10 ng of plasmid DNA was added to 15µl of slow–thawed *E. coli* DH5α cells and mixed via gentle pipetting. After being left to cool on ice for 15 minutes the cells were then heat shocked at 37°C for 45 seconds or 42°C for 30 seconds. The cells were placed back on ice for 2 minutes before being supplemented with 400 µl of LB and gentle agitation at 37°C for one hour. The transformed *E. coli* was then plated onto Luria agar (LA) plates and incubated at 37°C overnight before being stored at 4°C. In the case of liquid cultures, *E. coli* was cultivated in Luria Bertani (LB) broth with gentle agitation at 37°C overnight. Agar and liquid broth were supplemented with 50mg/ml of ampicillin for all plasmids used in Chapter two.

2.2.7.2 DNA propagation and extraction

Following transformation of *E. coli* with the relevant DNA vectors, the DNA was propagated by inoculating cells into 100ml of LB broth with gentle agitation overnight at 37°C. The following morning the cells were subjected to plasmid DNA extraction using the QIAGEN® Plasmid Plus Maxi kit (Cat: 12963) and the protocol was followed according to the manufacturer's recommendations. Extracted DNA was assessed for purity and yield using a nanodrop ND-1000 (Inqaba biotech™, South Africa) before being stored at 4°C for short term storage or -20°C for long term storage.

2.2.8. Growth conditions and maintenance of human embryonic kidney (HEK293T) cells.

Immortalized HEK293T cells (Cat: CRL-3216™, ATCC®, USA) were used in this project due to their compatibility with expressing genes from the pTHpCapR vector. Growth medium comprised of Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (FBS, v/v) and 1% penicillin/streptomycin mix (v/v). HEK293T cells were cultivated from quick thawed freezer stocks (cell density at approximately 1.7 million cells per tube) that were added dropwise to 5ml of growth media and after gentle mixing, incubated at 37°C with 95% humidity and 5% CO_2 in a T25 flask overnight. The T25 flask was used initially to ensure close contact of the cells with the adherent plastic. The next day

the cells were checked under the microscope for confluence and then seeded into a bigger T75 flask before being used in experiments. Cells were passaged as follows: Growth media was aspirated from the flask followed by a gentle PBS wash. After aspiration of the PBS (Phosphate buffer saline), one ml of trypsin was added, and the flask was incubated at 37°C for 30 seconds allowing the cells to dislodge. Dislodged cells were then resuspended in 4 ml of added media to make up a total of 5 ml. Cells grown in increasing flask sizes were passaged in the same manner except with greater volumes. General volumes of each component are summarized in **Table 2.1**.

Table 2.1: Summary of tissue culture flask components for maintenance and passaging of HEK293T cells. The cells were retrieved from -80°C freezer stocks and cultivated in T-25 flasks until growth and morphology was as intended. The cells were then passaged and split into T-75 flasks for passaging three times a week whilst on standby for experiments.

Flask size	Surface area (mm^2)	0.05% Trypsin (ml)	Growth medium (ml)
T-25	2500	1	3-5
T-75	7500	2	8-15
T-175	16000	4	15-30

2.2.8.1 HEK293T cell counting procedure and confluency determination

General percentage confluency of HEK293T cells was estimated visually using a microscope and cells diluted 1/10 in trypan blue were counted on a haemocytometer to distinguish between live and dead cells.

2.2.8.2 Transfection and protein expression of LUJV GP-CΔTM and S1P in HEK293T cells

Small scale HEK293T cell transfections were performed using 6 well culture plates with a seeding density of 1 million cells per well supplemented in 2ml of growth medium (as described in 2.2.8). HEK293T cells were incubated for 24 hours at 37°C until 60-80% confluency was reached prior to

transfection. One μg of recombinant plasmid DNA was added to 100μl of Gibco DMEM in a sterile 1.5 ml Eppendorf tube (Sigma-Aldrich, South Africa). Three μl of X-tremeGENE DNA Transfection reagent (Roche, South Africa) was added to the DNA mix and incubated at room temperature for 30 mins to allow the transfection reagent to bind to the DNA and form complexes to facilitate transfection. The transfection mix was then added in a dropwise manner to wells whilst the media underwent gentle manual agitation.

2.2.9 Protein SDS gels and western blotting of the HEK293T expressed LUJV GP-CΔTM and S1P proteins.

2.2.9.1 Preparation of Sodium Dodecyl sulfate (SDS) gels for protein separation based on size

SDS-PAGE (Sodium dodecyl sulfate polyacrylamide agarose gel electrophoresis) analysis was performed on the HEK293T-expressed LUJV GP-CΔTM and S1P proteins to separate them according to size and prepare for a western blot to assess purity and expression levels in the experiments. Thirty microlitres of protein samples were prepared with 10μl of 4X β-mercaptoethanol dye. Cell media containing secreted protein was added directly to protein dye. Protein from cell lysate was prepared by the addition of Glo™ Lysis buffer (Promega, Australia) to cell substrate according to the manufacturers recommendations before adding to protein dye. The samples were then incubated at 95°C for 5 minutes to denature the protein. These were then electrophoresed through a polyacrylamide gel (Percentage specified on gel images) at 150V until the proteins passed through the stacking gel whereafter the voltage was increased to 200V until completion (approximately 100 minutes). The gel setup utilized was the Mini-PROTEAN® Tetra SDS-PAGE system (Bio-Rad, USA). For all protein gels the Color prestained Protein Standard, Broad Range molecular weight marker (New England Biolabs, USA) was used.

2.2.9.2 Western blotting and detection of LUJV GP-CΔTM and S1P expressed proteins from HEK293T cell lysate.

To evaluate the utility of the mammalian expression systems in producing the desired recombinant protein, proteins separated on SDS polyacrylamide gels were subject to western blots using appropriate alkaline phosphatase conjugated antibodies for detection. This included a primary mouse anti-6X Histidine antibody (Dilution 1:2000, Cat: MCA1396, Bio-Rad, USA) and secondary goat anti-mouse IgG conjugated to AP (Dilution 1:5000, Cat: A3562, Sigma, USA) for LUJV GP-CΔTM. An anti-S1P

rabbit polyclonal antibody was used for site -1 protease detection (Dilution: 1:5000, Cat: Ab140592, Abcam, UK) and an anti-rabbit IgG (whole molecule)- alkaline phosphatase antibody (goat) was used as a secondary antibody (Dilution: 1:5000, Cat: A3687-1M1, Sigma, USA). Proteins separated on SDS polyacrylamide gels were soaked in transfer buffer along with a strip of Immun-Blot PVDF nitrocellulose Membrane (Bio-Rad, USA) and extra thick western blotting filter paper (7 cm by 8.4 cm, ThermoFisher, USA) before being subject to a Trans-blot® SD semi-dry transfer cell (Bio-Rad, USA). Blotting paper was placed down followed by the SDS gel, nitrocellulose paper and another piece of blotting paper before covering and being subjected to the PowerPac ™ Universal (Bio-Rad, USA) at 25V for one hour. After transfer, the nitrocellulose membrane was incubated in blocking buffer for 1 hour, primary antibody with block buffer overnight and then washed 4 times in PBS for 15 mins before adding secondary antibody diluted in block buffer at the dilution ratios mentioned above. After four more PBS washes, the nitrocellulose membrane was subjected to the KPL BCIP/NBT Membrane Phosphatase Substrate System (Cat:5420-0030, SeraCare, USA) for visualization of protein bands. As soon as the expected bands were visualized, the reaction was stopped using de-ionized water and left to dry before imaging on a Molecular Imager® Gel Doc™ XR+ imaging system (Bio-Rad, USA).

2.2.9.3 Purification and concentration of LUJV GP-CΔTM

Large batches of His tagged expressed proteins (which includes secreted LUJV GP-CΔTM from the growth media) were purified using the HisPur™ Cobalt resin (ThermoFisher, USA) according to the manufacturer's recommendations for "purification using a gravity flow column". The cobalt resin was placed over some glass wool in a 10 mL syringe and a needle was placed into pump tubing connected to the Gilson Minipuls 3 peristaltic pump. The pump would then pump liquid into the airtight syringe. Protein lysate from HEK293T cells was harvested and centrifuged at 2500 RPM for 15 minutes to remove cell debris. After the addition of binding buffer, the supernatant of the cell lysate was pumped through the column followed by wash buffer to remove any unwanted contaminants and elution buffer to collect all the remaining protein bound to the column.

Eluted protein was concentrated using a 10kDa Vivaspin protein concentrator spin column (Cytiva, USA). All column purification and concentration steps were performed at 4°C.

2.3 Results and brief discussion

2.3.1 Restriction digests of pTHpCapR: LUJV GP-CΔTM and pTHpCapR: S1P.

To express the LUJV GP-CΔTM and S1P proteins in HEK293T cells, it was necessary to clone the synthesized regions (obtained from GenScript) into the mammalian expression vector pTHpCapR. The overview of the cloning strategy is shown below for LUJV GP-CΔTM and S1P in **Figures 2.4 and 2.5**, respectively.

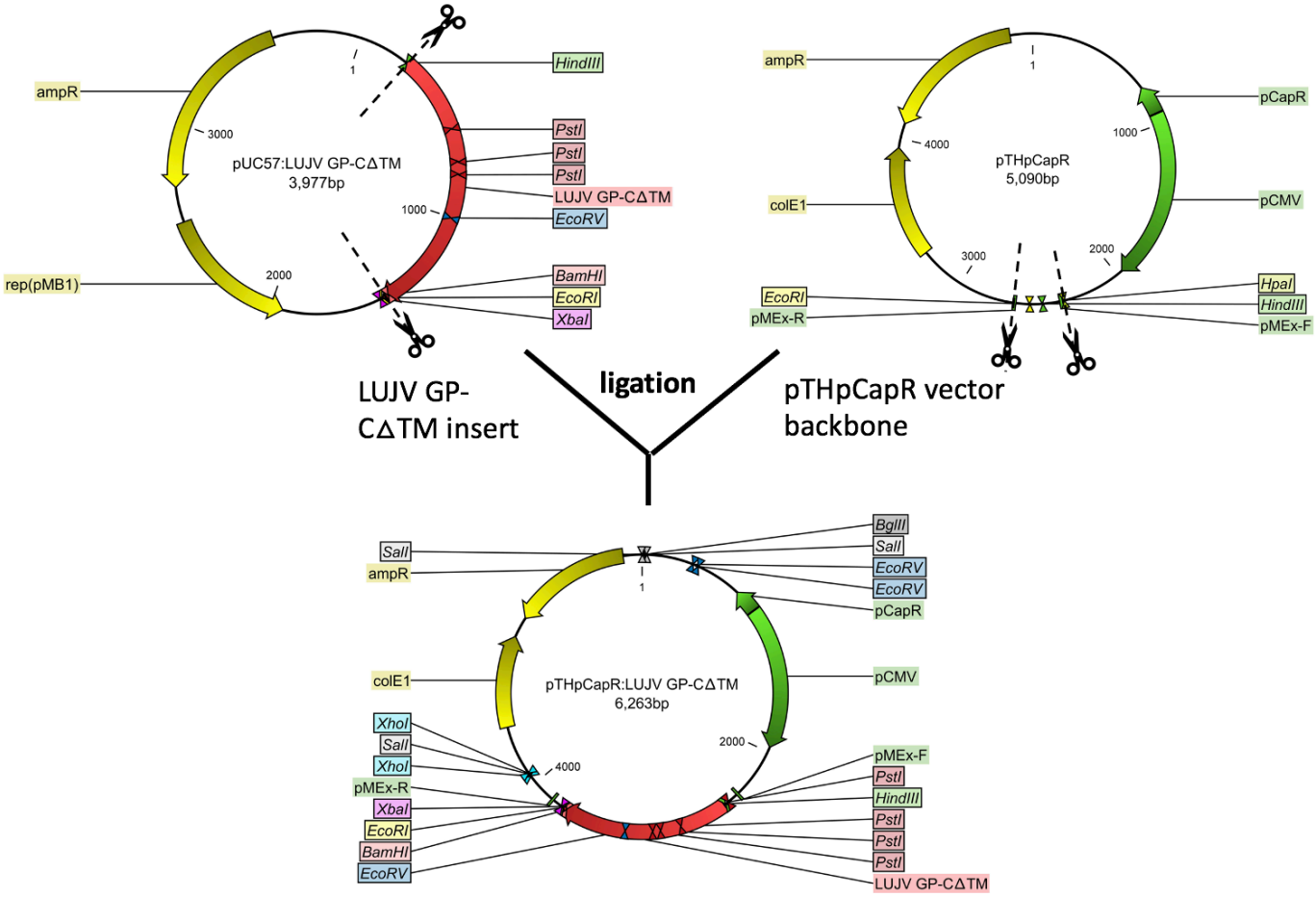


Figure 2.4: cloning strategy to construct pTHpCapR: LUJV GP-CΔTM. The LUJV GP-CΔTM gene construct was synthesized by GenScript and inserted in the pUC57 plasmid. The pTHpCapR vector was used as a backbone into which to clone LUJV GP-CΔTM. pTHpCapR has elements: pCMV-cytomegalovirus promoter, pCapR- a PCV-1 capsid protein promoter, ColE1-origin of replication for *E. coli*, pMEx-F- Forward primer binding site, pMEx-R- reverse primer binding site, ampR- Ampicillin resistance. pUC57 has elements: ampR- Ampicillin/Carbenicillin resistance, rep(pMB1)- *E. coli* replication site. Both constructs were digested with *Eco*RI and *Hin*dIII and cloned into pTHpCapR resulting in pTHpCapR: LUJV GP-CΔTM. Points of cleavage are indicated by scissors and dotted lines.

LUJV GP-CΔTM and S1P were successfully digested from pUC57 using *Eco*RI and *Hin*dIII to yield fragments of 1261 and 3187 bp, respectively (results not shown). The inserts were then subcloned into *Eco*RI/*Hin*dIII digested pTHpCapR (5017 bp fragment after digest) to produce pTHpCapR: LUJV GP-CΔTM and pTHpCapR: S1P, respectively.

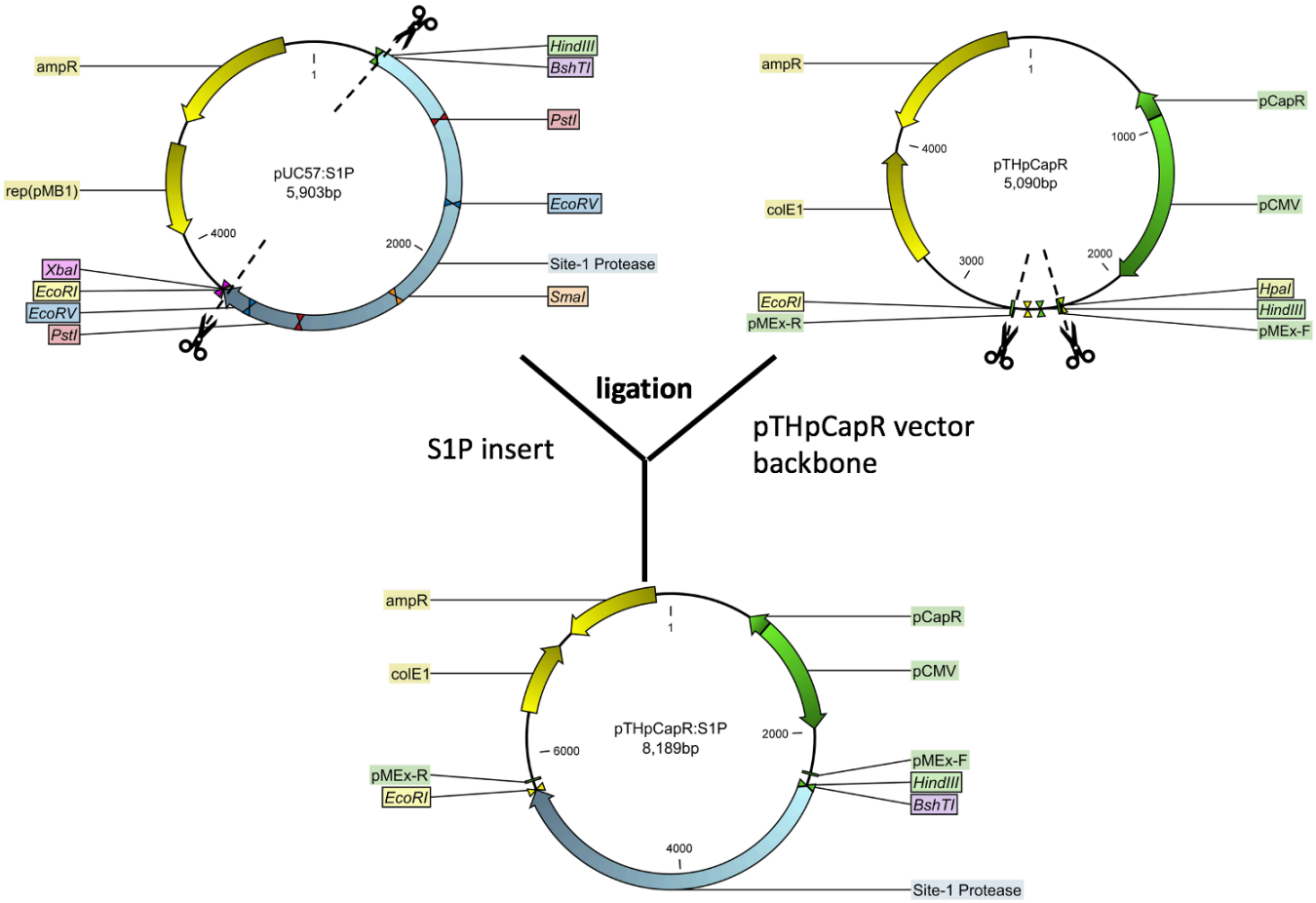


Figure 2.5: Cloning strategy to construct pTHpCapR: S1P. The S1P gene construct was synthesized by GenScript and inserted in the pUC57 plasmid. The pTHpCapR vector was used as a backbone for which to clone the S1P insert into. pTHpCapR has elements: pCMV-cytomegalovirus promoter, pCapR- a PCV-1 capsid protein promoter, ColE1-origin of

replication for *E. coli*, pMEx-F- Forward primer binding site, pMEx-R- reverse primer binding site, ampR- Ampicillin resistance. pUC57 has elements: ampR- Ampicillin/Carbenicillin resistance, rep(pMB1)- *E. coli* replication site. Both constructs were subjected to EcoRI and HindIII restriction endonuclease digestion to make the DNA compatible for ligation to produce the final mammalian expression vector pTHpCapR: S1P. Plasmid DNA was propagated in *E. coli* before being extracted and subject to enzyme restriction digest activity. Points of cleavage are indicated by scissors and dotted lines.

These newly modified expression vectors were subsequently transformed into *E. coli* Dh5α cells for glycerol stock storage. A single colony was chosen and cultivated in liquid cultures to allow for a large-scale DNA isolation of which DNA was used in subsequent experiments. The pTHpCapR: S1P and pTHpCapR: LUJV GP-CΔTM vectors were subject to *Eco*RI/*Hin*dIII restriction digests and then visualized on agarose gels (**Figures 2.6 and 2.7**).

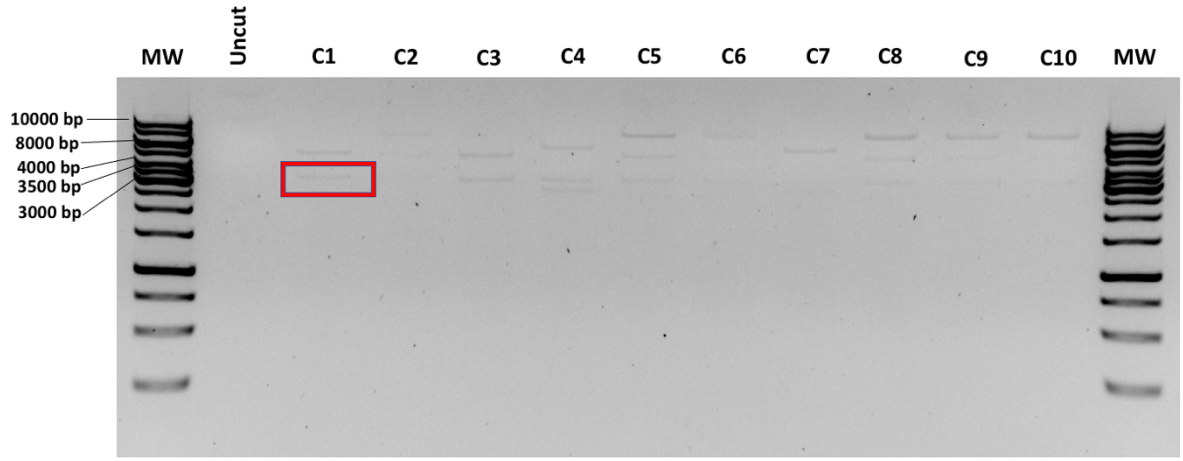


Figure 2.6: 1% Agarose gel of pTHpCapR: S1P digested products to confirm successful cloning. Confirmation restriction digests of pTHpCapR: S1P using *Eco*RI and *Hin*dIII from miniprep of transformed *E. coli* colonies post antibiotic screening. Molecular weight marker- 1KB Gene ruler DNA ladder (ThermoFisher, USA), uncut pTHpCapR: S1P vector, screened colonies 1-10, molecular weight marker- 1KB DNA ladder. S1P is expected at 3187 bp. S1P digested fragment from colony 1 is indicated by a red box and was used for subsequent experiments.

DNA extracted from colonies putatively harboring pTHpCapR: S1P were screened for successful transformation by digestion with *Eco*RI and *Hin*dIII. DNA from colonies 1 and 3 showed the expected banding pattern of 3187 bp (insert) and 5011 bp (vector). Colony 1 was selected for further work. The same insert cleaving digest was performed on the pTHpCapR: LUJV GP-CΔTM vector (**Figure 2.7**).

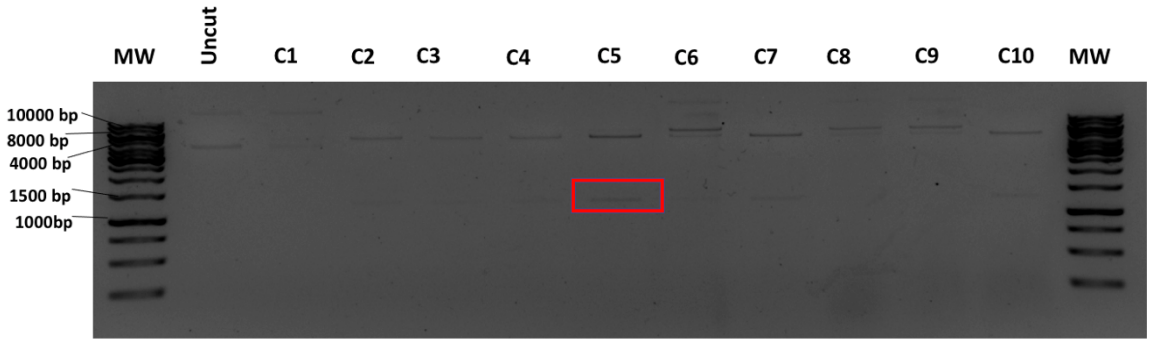


Figure 2.7: 1 % Agarose gel of pTHpCapR: LUJV GP-CΔTM digested products to confirm successful cloning. Restriction digests of pTHpCapR:LUJV GP-CΔTM DNA extracted from transformed antibiotic screened *E. coli* colonies using *Eco*RI and *Hin*dIII to confirm successful cloning. Molecular weight marker- 1KB DNA ladder, uncut pTHpCapR: LUJV GP-CΔTM, transformed *E. coli* colonies 1-10, Molecular weight marker-1KB DNA ladder. The cleaved LUJV GP-CΔTM insert from colony 5 is indicated by a red box and is expected at 1261bp. This colony was cultured and used for subsequent experiments.

The pTHpCapR: LUJV GP-CΔTM is 6263 bp in size and was seen in the uncut plasmid (**Figure 2.7**), however, a larger band above 10000 bp was seen in the uncut, C1, C6, C8 and C9 lanes. This band may be an indication of a mixed culture or amplification of an unknown bacterial borne segment. C5 was chosen as it reflected the expected digest result of 1261 bp (insert) and 5002 bp (vector) which was used in subsequent experiments.

Table 2.2: Table of plasmid vectors, enzymes and expected fragment sizes from restriction digests. Expected plasmid DNA sizes after restriction digests are to be used in conjunction with cloning strategy (**Figures 2.4 and 2.5**).

Plasmid	Enzyme digest	Expected fragment sizes (bp)

pUC57:S1P (manufacturer vector with S1P insert, 5903 bp)	*Eco*RI+HindIII	3187 (S1P insert) +2716 (pUC57 vector backbone)
pUC57: LUJV GP-CΔTM (manufacturer vector with LUJV GP-CΔTM insert, 3977 bp)	*Eco*RI+*Hin*dIII	1261 (LUJV GP-CΔTM insert) +2716 (pUC57 vector backbone)
pTHpCapR: LUJV GP-CΔTM (mammalian expression vector with LUJV GP-CΔTM insert, 6263 bp)	*Eco*RI+*Hin*dIII *Bsh*TI/*Kpn*I *Hpa*I/*Bam*HI *Kpn*I	1261 (LUJV GP-CΔTM insert) +5002 (pTHpCapR vector backbone) 1009+5262 1353+4914 6263
pTHpCapR: S1P (Mammalian expression vector with S1P insert, 8198 bp)	*Eco*RI+*Hin*dIII *Hpa*I/*Sma*I *Sal*I *Pst*I	3187 (S1P insert) +5011 (pTHpCapR vector backbone) 2063+6126 5975+2188+38 5555+2014+632
pTHpCapR (mammalian expression vector backbone, 5090 bp)	*Eco*RI+*Hin*dIII	81+5017 (pTHpCapR vector backbone)

Both newly generated expression vectors pTHpCapR: LUJV GP-CΔTM and pTHpCapR: S1P were subjected to multiple restriction digests (as summarized in **Table 2.2**) to further confirm their genetic integrity. The digest fragments were as expected with smaller fragments <100 bp usually not visible due to low concentrations or migrating off the gel. Cloning was a success and the pTHpCapR: LUJV GP-CΔTM and pTHpCapR: S1P plasmids were transformed into *E. coli* for a large-scale isolation of DNA to be used in subsequent experiments.

2.3.2 Mammalian protein expression

Once the integrity of the pTHpCapR: LUJV GP-CΔTM and pTHpCapR: S1P plasmids was confirmed through the digests, the next step was to assess their ability to express the recombinant proteins in HEK293T cells. The plasmids were transfected into HEK293T cells using the X-tremeGENE HP DNA transfection reagent. The cells were then lysed and the cell contents (HL) along with the media (HM) were visualized on a western blot using an antibody specific to the C-terminal His tag (**Figure 2.8**).

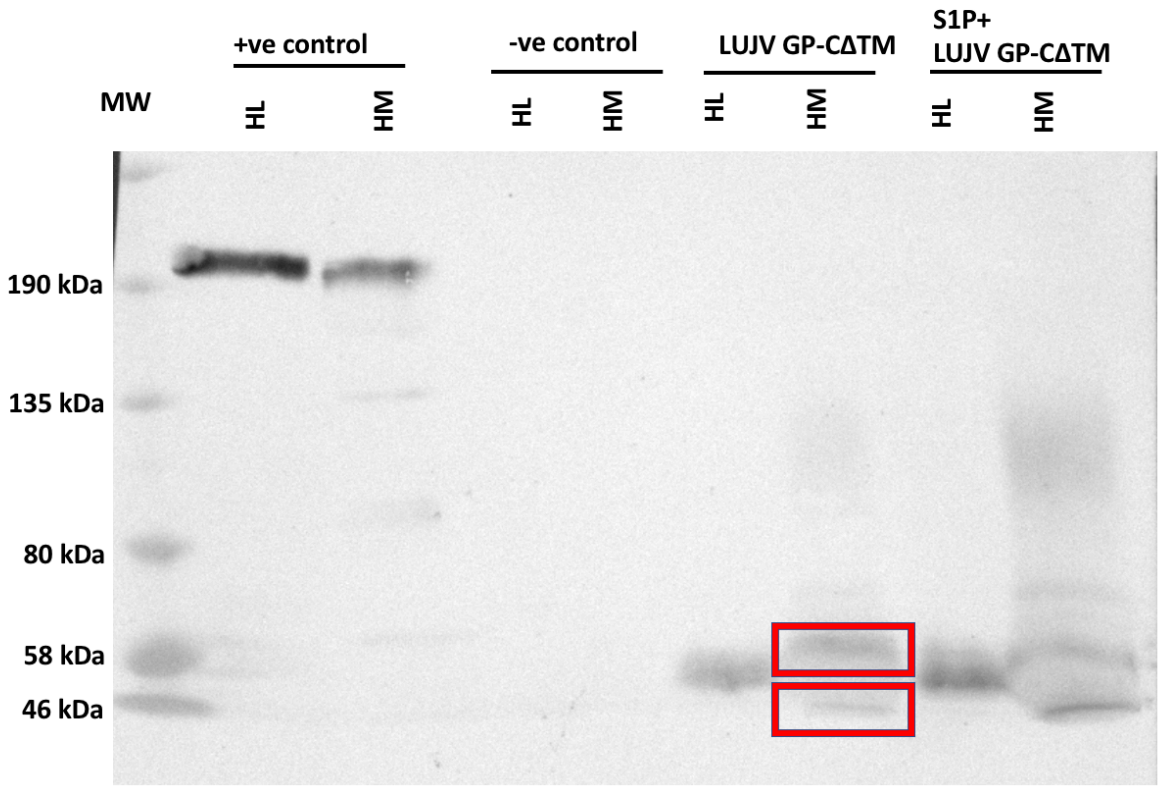


Figure 2.8: 12% western blot probing for His tag of LUJV GP-CΔTM expressed in HEK293T cells. Protein expression products from HEK293T cells after DNA transfection with the pTHpCapR: LUJV GP-CΔTM and pTHpCapR: S1P vectors. Primary antibody was 1:2000 mouse anti-6X Histidine antibody (Cat: MCA1396, Bio-Rad,USA) and secondary was 1:5000 goat anti-mouse IgG conjugated to AP (Cat: A3562, Sigma, USA). HEK293T lysate and media extracted protein are denoted by HL and HM respectively for each sample. 245 kDa molecular weight marker, positive control- Covid spike protein expected at ~ 200 kDa, negative control- transfection reagent only, LUJV GP-CΔTM protein products, LUJV GP-CΔTM and pTHpCapR:

S1P protein products. Suspected glycoforms in the media of the experimental transfections are indicated by red boxes.

A positive control was included that comprised an expression vector expressing the SARS-CoV-2 spike protein (200 kDa) that had previously been shown to interact with the His antibody. The negative control comprising untransfected cells, showed no bands for His antibody associating proteins in either HL or HM as expected. The HL and HM had similar banding patterns to the samples not co-transfected with S1P which was expected as S1P is naturally present in mammalian cells [33]. The LUJV GP-CΔTM, approximately 47 kDa in size was seen in HL at around 58 kDa with two unknown proteins detected in the media (HM) (red boxes). At this point it was not clear what the two proteins were but since the LUJV GP-C protein sequence excluded the transgenic region it was theorized that this may be a cleavage product or 2 different glyco-forms of the LUJV GP-CΔTM. Co-expression with S1P was also explored to assess if it would alter the cleavage of the LUJV GP-CΔTM in case S1P was not expressed at sufficient levels endogenously. Co-expression with S1P did not seem to cause a difference in the expression of the LUJV GP-CΔTM.

S1P was detected in the S1P transfected HL at the expected size of 117 kDa (red box) using a polyclonal antibody with specific affinity for S1P (**Figure 2.9**). Being a protein with a transmembrane region, S1P is expected to be found more abundantly in the lysate rather than the media.

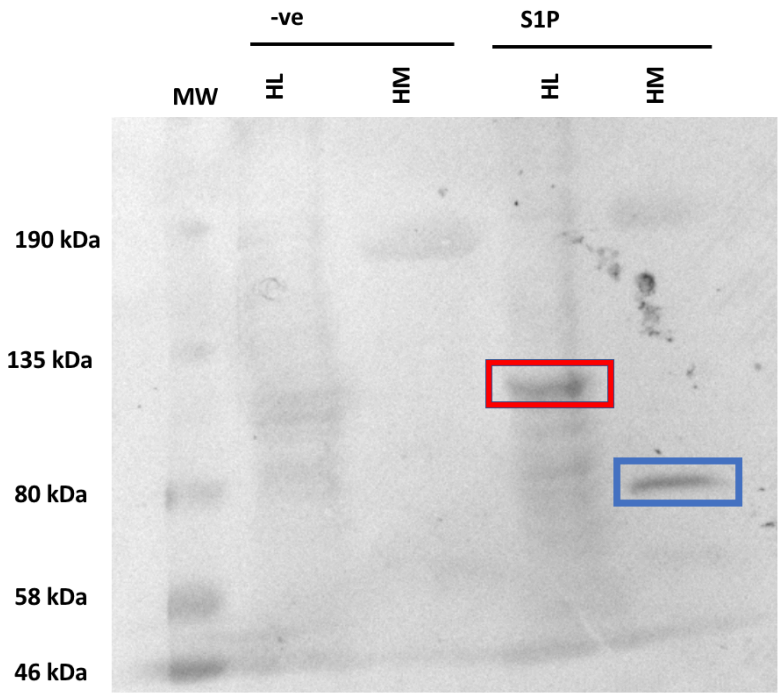


Figure 2.9: 10% Western blot of HEK293T expressed protein probing for S1P. An anti-S1P rabbit polyclonal antibody was used for S1P (Dilution: 1:5000, Cat: Ab140592, Abcam, UK) and an anti-rabbit IgG (whole molecule)- alkaline phosphatase antibody (goat) was used as a secondary antibody (Dilution: 1:5000, Cat: A3687-1M1, Sigma, USA). Proteins from cell lysate and media was denoted by HL and HM respectively for each sample.245 kDa molecular weight marker, negative control- transfection reagent only, S1P binding products from pTHpCapR: S1P transfected cells. S1P expected at 117kDa (red box) and potential S1P cleavage product is seen in the media (blue box).

An 80 kDa product seen in the media was possibly a secreted cleavage product of S1P (blue box). Very faint bands were seen at 117 kDa in the lysate of the negative control; this is likely endogenous S1P being detected, although not quantitative, the intensity of the band may suggest qualitatively that endogenous S1P levels are low.

2.3.3 Mammalian protein purification

After a large-scale expression in HEK293T cells, the LUJV GP-CΔTM protein band around 58 kDa was confirmed by Coomassie-staining after the protein was purified from transfected mammalian cells using the Hispur™ Cobalt Resin (ThermoFisher, USA) (**Figure 2.10A**). In addition, a second band of approximately 70-75kDa was observed.

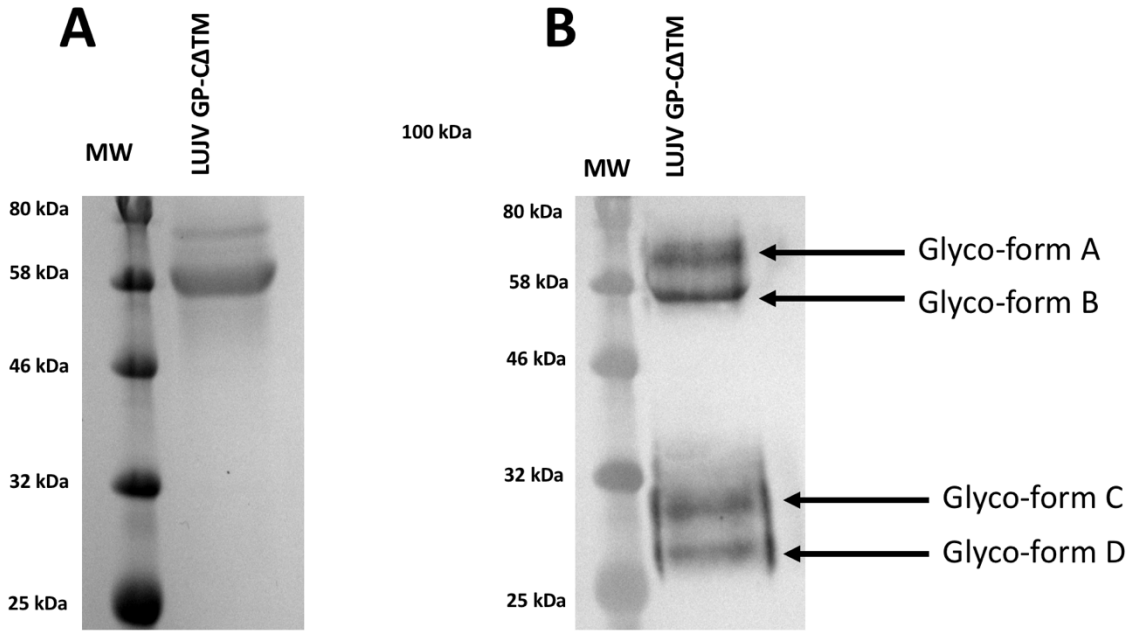


Figure 2.10: A: 12% Coomassie gel of purified LUJV GP-CΔTM Purified LUJV GP-CΔTM expressed in HEK293T cells and His tag purified from the cell media subjected to a Coomassie stain. 245 kDa molecular weight marker, purified LUJV GP-CΔTM. **B: 12% Western blot showing purified LUJV GP-CΔTM** Western blot of LUJV GP-CΔTM purified from HEK293T cells. The protein was probed with a primary mouse anti-6X Histidine antibody (Cat: MCA1396, Bio-Rad, USA) and secondary was 1:5000 goat anti-mouse IgG conjugated to AP (Cat: A3562, Sigma, USA). 245 kDa molecular weight marker, purified LUJV GP-CΔTM. Glyco-form variants A-D are shown in the LUJV GP-CΔTM purified protein indicating different levels of glycosylation. Glyco-forms A and B are likely un-cleaved LUJV GP-CΔTM while glyco-forms C and D are cleavage products (GP 2).

Western blotting of a cognate gel with anti-His antibodies confirmed the presence of both bands as LUJV GP specific. It is possible that these represent 2 glyco-forms (glyco-forms A and B) of LUJV GP-CΔTM. To confirm this, the protein was subjected to 2 different kinds of de-glycosylation treatments: de-glycosylation mix (Deg) which cleaves N and O-linked glycans, and PNGase F which cleaves N glycans only. Treated proteins were separated by SDS-PAGE and probed on a western blot to see if the protein seen was indeed the LUJV GP-CΔTM. Surprisingly, another band was seen just above the 58 kDa expected product along with two other species of protein (glyco-form C and D) between 25 and 32 kDa (Figure 2.10B). One cleavage product GP2 is expected around 29 kDa when cleaved from

GP1, however, cannot be visualized on its own in this manner. The extra band seen above the LUJV GP-CΔTM and its cleavage products were likely glyco-forms as each additional glycan adds approximately 2-3 kDa. The de-glycosylation results are shown in **Figure 2.11**.

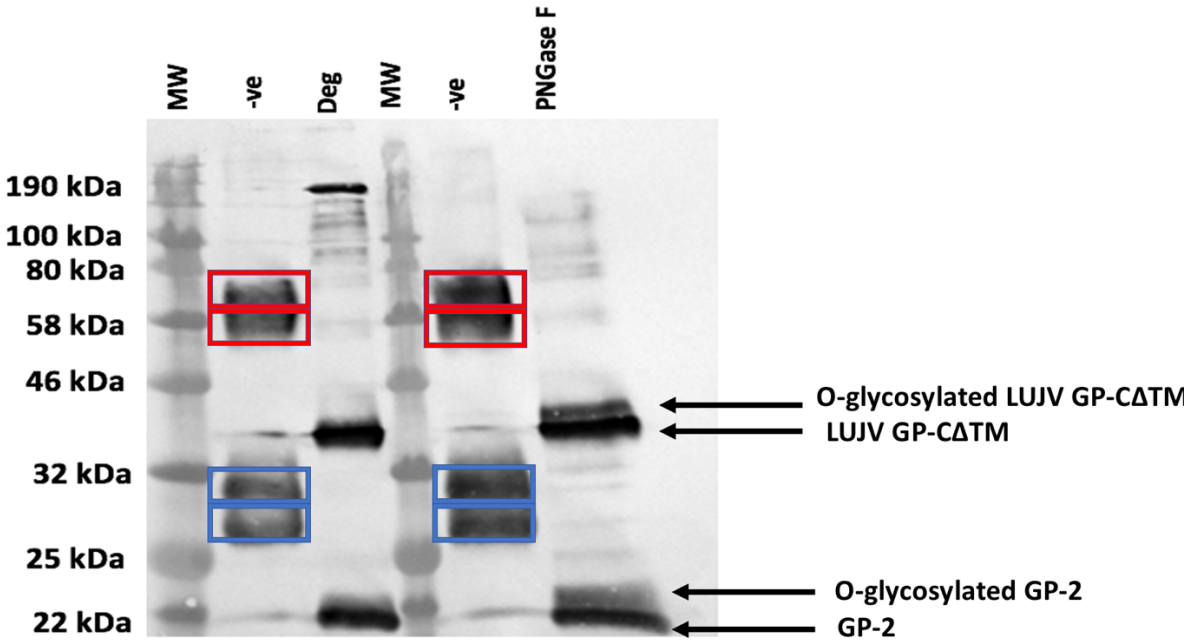


Figure 2.11: 12% Western blot of de-glycosylation reactions to assess glycosylation profile of LUJV GP-CΔTM LUJV GP-CΔTM purified from HEK293T cells subjected to de-glycosylation digests before probing with primary mouse anti-6X Histidine antibody (Cat: MCA1396, Bio-Rad, USA) (1:2000) and 1:5000 goat anti-mouse IgG conjugated to AP (Cat: A3562, Sigma, USA). 245 kDa Molecular weight marker, purified LUJV GP-CΔTM untreated, LUJV GP-CΔTM treated with de-glycosylation mix (Deg) to remove both N and O-linked glycans, 245 kDa molecular weight marker, purified LUJV GP-CΔTM, LUJV GP-CΔTM treated with PNGase – F to remove N-linked glycans.

The negative control (-ve, **Figure 2.11**) was untreated LUJV GP-CΔTM extracted from mammalian cells as shown in **Figure 2.12B**. The banding pattern with the potential glycosylation species was confirmed in this blot. When subject to the de-glycosylation mixture (Deg), the two upper bands (red boxes)

consolidated into a single band of approximately 40 kDa while the two lower bands (blue boxes) conform to a single band at 22 kDa This was confirmation of the theory that two out of the four bands were glycosylation products as de-glycosylation treatment was effective. A band was seen at 190 kDa which may have been degraded protein aggregates or may be attributed to contamination of other protein within the de-glycosylation mix. PNGase F treatment showed a similar consolidation of bands as the de-glycosylation mix result except both bands had a second band very near to but slightly above them. PNGaseF does not cleave O-glycans like the de-glycoslyation treatment does so it is likely that the upper band of each band seen was an O-glycosylated glycoform.

2.3.3 Large scale purification of mammalian cells

Having established that the LUJV GP-CΔTM is expressed in HEK293T cells, the system was tested for compatibility with large scale purification procedures. The LUJV GP-CΔTM was extracted and purified as done previously except this time using substantially more transfected cells (27 T175 flasks at a total of 810 ml). Samples from the purification process were loaded onto a SDS polyacrylamide gel, proteins separated by electrophoresis and either stained with Coomassie blue (**Figure 2.12A**) or transferred to nitrocellulose for western blotting probing with anti-His antibodies (**Figure 2.12B**).

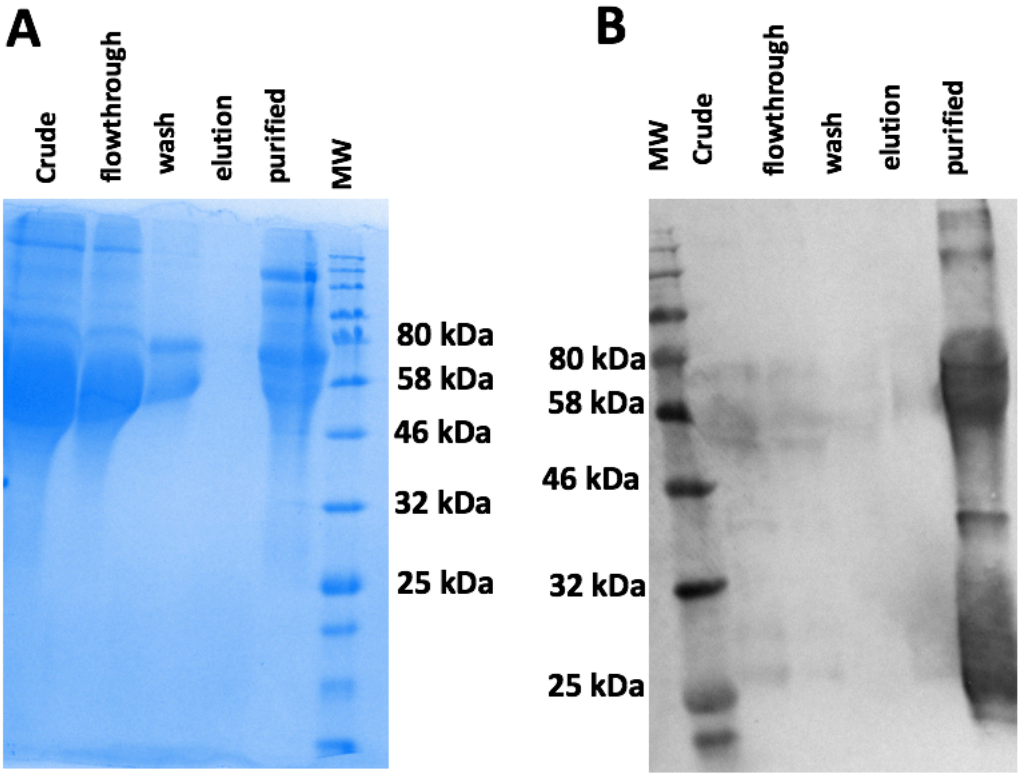


Figure 2.12: A: 12% Coomassie-stained gel of large-scale purification and concentration of LUJV GP-CΔTM from HEK293T cells. LUJV GP-CΔTM His tag purified from HEK293T cells (27 T175 flasks) using previously optimized expression conditions. For this experiment PEI (polyethylenimine) was used as a transfection reagent as opposed to the X-tremeGENE DNA HP transfection reagent (Cat: 06366236001, Roche, Germany) as this was the cheaper option. Crude HEK293T lysate, flowthrough fraction, wash buffer, elution buffer, concentrated eluted protein **B: 12% Western blot showing large scale purification and concentration of LUJV GP-CΔTM from HEK293T cells.** A western blot probing for LUJV GP-CΔTM from the Coomassie shown in Figure 2.10 A. His tag purified fractions of LUJV GP-CΔTM were probed with with primary mouse anti-6X Histidine antibody (Cat: MCA1396, Bio-Rad,USA) (1:2000) and 1:5000 goat anti-mouse IgG conjugated to AP (Cat: A3562, Sigma, USA). Crude HEK293T lysate, flowthrough fraction, wash buffer, elution buffer, concentrated eluted protein.

The extraction resulted in very large amounts of protein (2.16 mg/L cell culture) to the point where there were multiple bands and non-specific binding to the His antibody (**Figure 2.12**). Large bands were seen around 58 kDa as expected from previous extractions for LUJV GP-CΔTM which decreased in size from the crude lysate to the flowthrough to the wash steps. Elution showed very little protein as this is diluted in 100 ml of buffer. The LUJV GP-CΔTM was observed once purified along with a few other prominent protein bands that are likely contaminating cellular proteins. The GP2 cleavage products seen previously at 28 kDa (**Figure 2.10B**) could not be seen in **Figure 2.12A** which is likely due to its low concentration as expected. When subjected to a western blot, the large-scale purified sample of expressed LUJV GP-CΔTM was barely detected from crude to elution fractions (**Figure 2.12B**) but could be visualized after purification in the purified fraction. This is likely due to the large amounts of cellular protein that blocked the LUJV GP-CΔTM from being detected sufficiently. After purification, LUJV GP-CΔTM was very concentrated so showed a smear around the expected size of 58 kDa as opposed to the 2 bands expected from **Figure 2.10B**. The same was seen for GP2 and its glyco-forms at 28 kDa. In the purified sample the bands seen towards the top of the gel are likely His tag binding to cellular proteins. These bands were only seen in the large-scale purification due to the excessive levels of protein.

2.4 Detailed discussion

To produce a mammalian expression system for the production of LUJV GP-CΔTM antigens, the vector pTHpCapR: LUJV GP-CΔTM was constructed along with pTHpCapR: S1P to facilitate maturation of LUJV GP-CΔTM. The protein expression methods used in this study offer favourable tools for research and protein production purposes as viral subunits can be used under Biosafety level 2 conditions. Live Lujo virus needs a Biosafety level 4 facility and is comparatively more expensive due to extra safety procedures and protocols. The pTHpCapR vector was developed and shown to have previous success in producing recombinant proteins in mammalian cells including the production of HIV-1 subtype C mosaic Gag VLPs, which was found to be highly immunogenic [145, 146]. pTHpCapR was also utilized for recombinant mammalian protein expression for developments in a lumpy skin disease virus vaccine (LSDV) [147], the HIV envelope protein [148], as well as the *Theileria parva* surface protein p67 to combat East coast fever [149]. A combination of the cytomegalovirus promoter and the porcine circovirus enhancer element (not shown in **Figures 2.4 and 2.5**) in pTHpcapR increases recombinant protein expression substantially [148]. Following endonuclease verification of successful cloning, the DNA was subject to HEK293T transfection in order to assess the suitability of the expression system in

producing LUJV GP-CΔTM antigens. The LUJV GP-CΔTM was successfully expressed in HEK293T cells when using either PEI or the X-tremeGENE HP DNA transfection reagent. The X-tremeGENE HP DNA transfection reagent seemed to have a high utility for recombinant protein experiments as it seemed to increase the transfection rate of the vector. For western blot analysis this was important as there were many cellular protein contaminants in the samples of which visualization could be avoided due to fast detection of the target proteins. Transfection with PEI seemed to result in a lower transfection rate, which caused the LUJV GP-CΔTM to be visualized along with the contaminants in the NBT/BCIP reaction. Due to the high cost of the X-tremeGENE HP DNA transfection reagent it is sensible to use it only for small scale research experiments while the cheaper PEI can be used for the large-scale mammalian infiltrations as was done in this study. The LUJV GP-CΔTM expressed product seemed to be cleaved into GP1 and GP2 with or without co-expression of the pTHpCapR: S1P construct. This was expected as it is documented that S1P is present in mammalian cells [150] and it was now confirmed that the S1P expresses at sufficient levels in HEK293T cells for proteolytic cleavage of the LUJV GP-CΔTM. The transfection of pTHpCapR: S1P seemed to increase the expression of S1P relative to the endogenous levels, however, this was determined visually on a western blot and is not reliably quantitative – but may be worth exploring should this be a desire in future research.

The LUJV GP-CΔTM was visualized at ~ 58 kDa which is slightly higher than the protopram predicted size of 47 kDa and the expected size from previous studies (52.3 kDa) [55]. It is notable that the protopram prediction software does not account for increases in protein size due to glycosylation so the perceived size would be heavier due to the addition of glycan groups. Surprisingly there seemed to be 2 protein species around each expected LUJV GP-CΔTM fragment i.e the 58 kDa uncleaved fragment and the GP2 cleavage product which was seen at 26.8 kDa and is consistent with the previous genetic characterization of LUJV proteins [55]. GP1 is documented to be 18.9 kDa in size but could not be visualized in this study on its own due to it depending upon the His tag attached to GP2 for visualization [55]. Further de-glycosylation enzymatic reactions to investigate these phenomena confirmed these to be different glyco-forms. This is consistent with reports of the LUJV GP-C having 12 putative glycosylation sites (6 sites in GP1 and 6 sites in GP2) [55]. Our results showed the majority of the molecular mass seemed to be N-linked glycans with a small amount (3-5 kDa) being from O-linked glycosylation. When de-glycosylated completely the LUJV GP-CΔTM showed a band around 40 kDa which is expectedly smaller than the documented size of the LUJV GP-C (52.3 kDa) as the suspected transmembrane region (amino acid 396-415) had been removed during the gene design. Mammalian cells are the natural habitat for LUJV so it is expected that the LUJV GP-CΔTM would be produced in an immunogenic confirmation similar to that of natural infection. In depth structural analysis of proteins were, however, not explored in this study.

Large scale transfection of HEK293T cells yielded highly concentrated protein that seemed to blur the distinction between LUJV GP-CΔTM and other cellular proteins being detected, however, contaminants seemed to visually decrease in each subsequent fraction in the purification steps. The results at the very least determined that the LUJV GP-CΔTM was in the correct conformation to allow the His tag to bind to the purification column, classifying this method as reliable. It is hard to discern if the excess contamination of unwanted cellular proteins would cause adverse effects in the context of a vaccine.

Chapter 3: Development and optimization of a plant-based protein expression system for the Lujo virus glycoprotein

3.1 Introduction

Chapter 2 detailed the production of a mammalian expression system for the LUJV GP-CΔTM protein. The study revealed useful insight into the expected sizes, types of glycosylation and capacity for purification of LUJV GP-CΔTM to serve as a baseline for which to begin plant production experiments as an alternative technology. In Chapter 3, previously developed methods for producing transgenic *N.* benthamiana plants, including *Agrobacterium*-mediated transformation were explored. These methods were applied to the S1P and LUJV GP-CΔTM protein sequences synthesized allowing the production and optimization of LUJV GP-CΔTM, S1P and well as a variety of auxiliary maturation proteins.

The pEAQ-*HT* plasmid has shown to be an effective protein expression vector for *N. benthamiana* having shown previous successful production of VLPs and plant made proteins [151]. LUJV GP-CΔTM and S1P protein sequences were spliced into pEAQ-*HT* and delivered to *N. benthamiana* for transient protein expression through the use of plasmid transformed *Agrobacterium*.

3.1.1 Agrobacterium-mediated transformation of *N. benthamiana*

Agrobacterium tumefaciens – now more correctly known as *Rhizobium radiobacter* - is a naturally occurring soil inhabiting bacterium that delivers genes to plants, allowing production of nutrients for its own survival [152]. Expression vectors such as pEAQ-*HT* with spliced target proteins can be transformed into attenuated *A. tumefaciens* [151, 152]. The transformed bacterial cultures are then forced into the abaxial airspaces of *N. benthamiana* via vacuum infiltration where they can begin their gene delivery process. Multiple expression vectors can be constructed that can be used to deliver genes with this method and this forms the basis of setting up a plant environment that is conducive to the production of a given protein. The overall *Agrobacterium*-mediated transformation process is depicted in **Figure 3.1**.

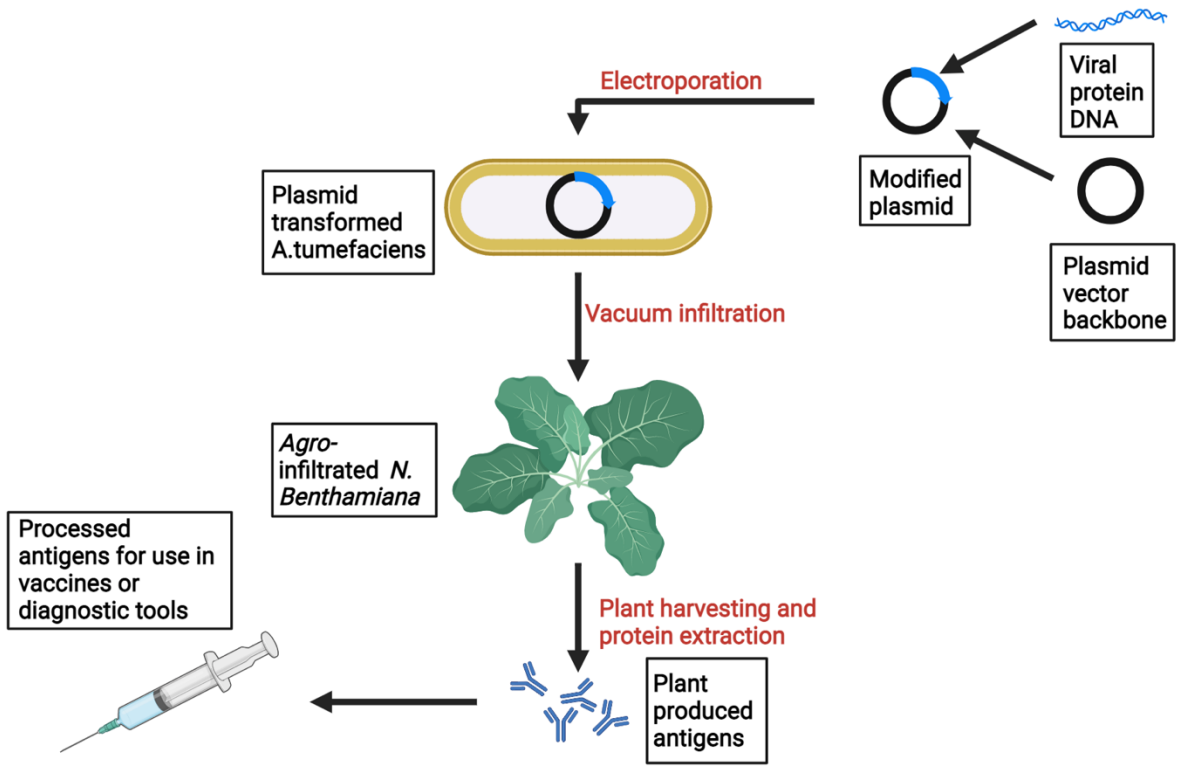


Figure 3.1: Overview of producing recombinant proteins in plants using *Agrobacterium*-mediated transformation. General workflow of producing proteins in plants. Target protein sequences from viruses or other organisms are cloned into plant expression vectors (in this study pEAQ-*HT*). *A. tumefaciens* is transformed with the modified plasmid via electroporation. Recombinant *A tumefaciens* is then cultured before being vacuum infiltrated into the abaxial air spaces of *N. benthamiana* where it can elicit expression of target proteins. After growth, the leaves are harvested, and protein is purified and concentrated before being ready for use in vaccines and diagnostic tools. A single plasmid is depicted for simplicity; however, auxiliary proteins were produced in the same manner.

N. benthamiana is well documented as a host expression plant making it appealing for this type of biopharmaceutical production. *N. benthamiana* plants are grown in optimal conditions beforehand until such time as they are ready for *Agro*-infiltration. Expression vector-containing A*. tumefaciens* are cultured to an optical density (OD) at which they will be suitable for plant infiltration i.e where there will be a balance between optimal protein production and necrosis of the plant. Necrosis occurs when the plant rejects tissue with high concentrations of non-native proteins and if they are in an

unfavorable conformation. It was thus important to also consider the optimal day post infiltration (DPI) as well as the helper proteins with which to co-infiltrate to prevent misfolding and degradation of proteins.

3.1.2 Cost of plant-based protein production

As mentioned in Chapter 2 the cost of production for a given biopharmaceutical is difficult to estimate due to the variables involved including the changing prices of reagents, current production capacity available, ethical and economic factors, and other research costs. Here we give a general comparison between mammalian and plant-based protein production costs. Figure 3.2 depicts the overview of the main financial milestones in plant-based protein production.

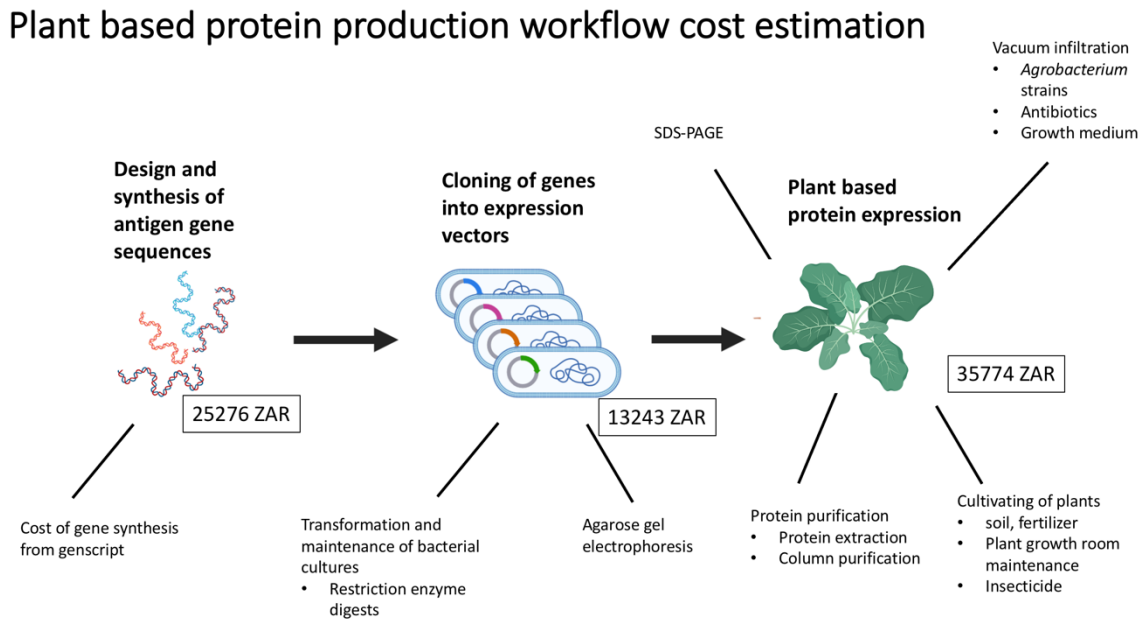


Figure 3.2: Overview of costs involved in plant-based protein production. Synthesis and cloning of genes are generally the same between plant and mammalian system except with a difference in the usage of the vector backbone. The depiction does not take into account the cost of equipment, which is important but varies, and is re-usable/sharable in a laboratory environment.

While the initial vector generation is similar, the main divergence occurs with regards to protein expression in the host cells and the variables and running costs associated therewith. Plant systems require additional expression vectors in the cloning steps as auxiliary proteins are needed to foster a mammalian-like cellular environment. However, plant biomass is cheap to cultivate relative to

mammalian cells as it does not require expensive growth medium/transfection reagent and is less sensitive for day-to-day maintenance relative to mammalian cultures. It is easier to train nonspecialized personnel in plant biopharmaceutical production than it is for mammalian cell culturing which requires BSL2 conditions.

3.2 Materials and methods

3.2.1 Design and synthesis of LUJV GP-CΔTM and S1P gene sequences

Design and synthesis of LUJV GP-CΔTM and S1P sequences was detailed in Chapter 2 (2.2.1). The sequences synthesized were cloned this time into pEAQ-*HT* as opposed to pTHpCapR. The gene arrangement in this Chapter is depicted in **Figures 3.3** and **3.4** (see Results).

3.2.2 Restriction endonuclease digests to produce plant expression vectors pEAQ-HT: LUJV GP-CΔTM and pEAQ-HT: S1P

To create plant expression vectors pEAQ-*HT*: LUJV GP-CΔTM and pEAQ-*HT*: S1P, different restriction enzymes (refer to Figures 3.3 and 3.4) were used to cut out LUJV GP-CΔTM and S1P from the pUC57 vectors in the same manner as described in Chapter 2 (2.2.2). All three plasmids were subject to a restriction digest with 1 μg of. DNA Reaction digest components were added to an Eppendorf tube for each plasmid and incubated at 37°C for 1 hour. Restriction enzymes used were obtained from ThermoFisher, USA.

3.3.1 Restriction digests of modified pEAQ-*HT* vectors

Construction of the pEAQ-*HT*: LUJV GP-CΔTM and pEAQ-*HT*: S1P vectors were performed as follows. The pEAQ-*HT* vector backbone was excised and purified as the base plant expression vector. The LUJV GP-CΔTM insert (1261 bp) was cut out from the pUC57 plasmid (2716 bp) and using *Bsh*tI and *Xho*I and then electrophoresed on an agarose gel before being excised for purification. After gel excision and purification, the DNA was subject to ligation reactions where S1P and LUJV GP-CΔTM was cloned into pEAQ-*HT*. The pEAQ-*HT*: LUJV GP-CΔTM and pEAQ-*HT*: S1P vectors were then tested for structural

integrity using multiple restriction enzyme combinations. **Table 3.4** shows the expected sizes of pEAQ-*HT*: LUJV GP-CΔTM and pEAQ-HT: S1P when subject to an array of restriction digests.

Table 3.1: Expected fragment sizes of plasmids after being subject to restriction digests. Fragment sizes after confirmation and cloning digests. To be considered in conjunction with cloning strategy (**Figures 3.3 and 3.4**)

Plasmid	**Restriction enzyme digests**	**Expected fragment sizes (bp)**
pUC57:S1P (manufacturer vector with S1P insert, 5909 bp)	*Bsh*TI+XhoI	3180 (S1P insert) +2737 (pUC57 vector backbone)
pUC57: LUJV GP-CΔTM (manufacturer vector with LUJV GP-CΔTM insert, 3983 bp)	*Bsh*TI+XhoI	1256 (LUJV GP-CΔTM insert) +2735 (pUC57 vector backbone)
pEAQ-*HT*: LUJV GP-CΔTM (Plant expression vector backbone with LUJV GP-CΔTM insert, 11201 bp)	*Bsh*TI+XhoI *Kpn*I *Hpa*I *Bam*HI *Xba*I	1256 (LUJV GP-CΔTM insert) +9953 (pEAQ-*HT* vector backbone) 9871+825+517 11201 1039+10170 10714+495
pEAQ-*HT*: S1P (vector backbone with S1P insert: 13125 bp)	*Bsh*TI+*Xho*I *Sma*I+*Bam*HI *Xba*I *Hpa*I+*Not*I *Nco*I	3180 (S1p insert) +9953 (pEAQ-*HT* vector backbone) 2249+10880 10880+1752+495+10 3730+9399

		8836+3277+1024
pEAQ-*HT* (Plant expression vector backbone, 10003 bp)	*Bsh*TI+*Xho*I	9953 (pEAQ-*HT* vector backbone) +58

3.2.3 Agarose gel excision of plasmid DNA fragments for cloning

Gel excision of DNA fragments were performed in the same manner as described in Chapter 2 (2.2.3) according to the manufacturer's recommendations with no deviations from the protocol.

3.2.4 Agarose gel electrophoresis of plasmid DNA

All plasmid DNA fragments in Chapter three were analyzed using agarose gel electrophoresis in the same manner as described in Chapter 2 (2.2.4).

3.2.5 Ligation reactions to construct the pEAQ: LUJV GP-CΔTM and pEAQ: S1P vectors

Ligation reactions for pEAQ: LUJV GP-CΔTM and pEAQ: S1P vectors were performed in the same manner as detailed in Chapter 2 (2.2.5) with an insert:vector ratio of 3:1. The reaction components are detailed in **Table 2.2**. Digest components were incubated at room temperature for 1 hour. Plasmid DNA was then transformed into E.*coli* before screening for antibiotic selection and storage.

3.2.6 Sequencing of DNA vectors

The constructed pEAQ: LUJV GP-CΔTM and pEAQ: S1P vectors were sequenced by Macrogen to confirm their integrity.

3.2.7 Bacterial transformation

The bacterial transformation of recombinant bacteria is the same as the methods described in Chapter 2 (2.2.7.1). However, in addition to transforming *Escherichia coli* DH5α cells with the constructed

plasmids for storage and propagation, the DNA was transformed into *A. tumefaciens* which was used as a plant gene delivery system. Recombinant *A. tumefaciens* strains were made by using electroporation transformation according to the following method: 100 ml LB was inoculated with a target A*. tumefaciens* strain and incubated overnight at 27°C with gentle agitation. Overnight cultures were centrifugated at 4000 RPM for 10 min to concentrate the cells and wash them. Supernatant was discarded from the centrifugation tubes and the bacterial cells were resuspended in 5 ml sterile water. The centrifugation step was repeated whereafter the bacterial cells were resuspended in 5 ml 10% glycerol and centrifuged once more as described above. Bacterial pellets were resuspended in 10% glycerol to a final volume of 5 ml before being aliquoted and stored at -80°C in preparation for the transformation steps.

In preparation for electroporation, electrocompetent *A. tumefaciens* cells were thawed on ice and 100 µl of thawed cells aliquoted into a chilled electroporation cuvette (0.1 cm). Three hundred ng of recombinant plasmid DNA was added and pipetted slowly up and down in the cuvette. After incubating the cuvette on ice for 5 minutes, electroporation was performed with settings: 1.8kV, 25 µF and 200 Ω on the electroporator. Nine hundred µl of LB was added to the cuvette and the contents were transferred to a clean Eppendorf tube for incubation at 27°C for 2 hours. The cells were then plated onto antibiotic supplemented agar and grown at 27°C for 3 days before making glycerol stocks from recombinant colonies. Glycerol stocks for *A. tumefaciens* were made in 25% glycerol and stored at -80°C.

3.2.8 Bacterial growth conditions and antibiotic selection

E. coli was cultured in Luria Bertani broth (LB) at 37°C and supplemented with antibiotics according to the recombinant vector contained. *A.tumefaciens* ,however, was grown in LB broth base at 27°C-29°C and has intrinsic antibiotic resistance (depending on the strain) in addition to the plasmid contained. **Table 3.2** below summarizes the strains and plasmids used as well as the antibiotics used for their selection.

Table 3.2: Bacterial strains and antibiotic resistances used for plant-based protein expression. The Table shows the bacterial strains used and the antibiotics that cultures were supplemented with to facilitate selection during incubation steps. Note the strains and vector plasmids have different antibiotic resistances and thus a mix of antibiotics are used for the most accurate selection process.

Certain insertions also have their own antibiotics resistances for example pMP90:LmsTT3D has Spectinomycin resistance (50 mg/ml) and $Hexo_3$ RNAI has resistance to Kanamycin (100 mg/ml).

Bacterial strain	**Antibiotic resistance**
E. coli	None
A.tumefaciens AGL1 : pEAQ-*HT*	Carbenicillin (100 mg/ml, AGL1 specific) Kanamycin (100mg/ml, pEAQ-*HT* specific)
A.tumefaciens GV3101:pMP90	Gentamycin (10mg/ml, pMP90 specific) Rifampicin (20mg/ml, GV3101 specific)

3.2.9 DNA propagation and extraction

DNA propagation and extraction from bacteria was performed in the same manner as described in Chapter 2 (2.2.7.2).

3.2.10 Plant protein expression and cultivation

In Chapter 3, protein was expressed in either wild type *N. benthamiana* or ΔXT/FT plants – glyco-engineered mutants that are deficient for producing plant-specific xylose and fucose glycans allowing a more mammalian-type cellular glyco-profile [153]. Plants were grown up from seed and *A. tumefaciens* infiltrated when the plants were around 3-4 weeks old. After infiltration, plants were returned to the growth room until ready for harvest. *N. benthamiana* was cultivated from seed that

was soaked in 0.05 % gibberellic acid overnight before planting. Seeds were planted in fine soil and watered three times a week. Seedlings were transferred to bigger pots and grown for about three weeks before being infiltrated. Plants were grown in containing a 2:1 ratio of peat to vermiculite and a controlled growth room with 16h artificial light and 8h darkness cycles with 45% humidity and at a constant temperature of 25°C.

3.2.10.1 *A. tumefaciens* mediated transformation

In preparation for plant infiltration, *A.tumefaciens* cultures were inoculated into 500 ml of LBB (Luria base broth) supplemented with antibiotics (see Table 2.3) and grown over 3 days at 27°C with gentle agitation. The day before infiltration, 100 μl of acetosyringone was added to a final concentration of 20 μM. On infiltration day the cultures were sampled and analyzed for optical density (OD_{600}) before being diluted in resuspension solution to get their desired OD_{600}. Acetosyringone was added once more to a final concentration of 200 μM and the cultures left for 1-4 hours. The plant leaves were then submerged into the bacterial cultures and subject to vacuum infiltration (-90 kPa for approximately 5 seconds before rapid release of pressure) before being returned to the plant room for protein expression.

3.2.11 Plant protein extraction

After protein expression, plant leaves were harvested on day 3,4 or 5 post infiltration (DPI) and stored at -80°C. If the protein extraction process begun the same day, then freezing was not necessary. To begin protein extraction, the leaves were flash frozen in liquid nitrogen, ground up using a pestle and then mixed with triple the leaf weight in extraction buffer (1X PBS pH 7.8) as well as protease inhibitor. For large scale experiments, two leaf volumes of buffer were used to reduce the consumption of protease inhibitor and flash freezing was not performed. The mixture was then liquidized in a blender and agitated gently at 4°C for 1 hour. The crude lysate was then filtered through 22-24 μM pore Miracloth™ (Millipore, USA). Afterwards, the mixture was centrifugated at 10000 RPM for 15 mins and the supernatant retained for quantification and purification. The supernatant was aliquoted and stored at - 20°C.

3.2.12 Plant produced protein quantification

Protein quantification was performed using the DC Protein Assay kit (Bio-Rad, USA) according to the manufacturer's recommendations.

3.2.13 Plant protein purification

Plant extracted proteins were filter sterilized using the low protein binding Stericup® Quick release durapore® 0.22 μm filtration system (Millipore, USA). The filtrate was then purified through the Hispur™ Cobalt Resin (Thermofisher, USA) in the same manner as described in Chapter 2 (2.2.9.3)

3.2.14 Sodium Dodecyl Sulfate (SDS) polyacrylamide gels to separate and analyze purified protein based on size.

Plant protein was subject to SDS gels in the same manner as described in Chapter 2 (2.2.9.1)

3.2.15 Western blotting to analyze purified protein

Western blotting was performed in the same manner as described in Chapter 2 (2.2.9.2)

3.3 Results and brief discussion

The HEK293T mammalian expression system detailed in Chapter 2 enabled production of high yields of LUJV GP-CΔTM for use as a positive control. Mammalian cell expression is generally the accepted standard for glycoprotein production and thus served as a suitable comparison for plant expression. The vectors pEAQ-*HT*: LUJV GP-CΔTM and pEAQ-HT: S1P were successfully generated in a manner similar to the mammalian expression vectors described in Chapter 2. **Figures 3.3** and **3.4** show the cloning strategy to produce these vectors.

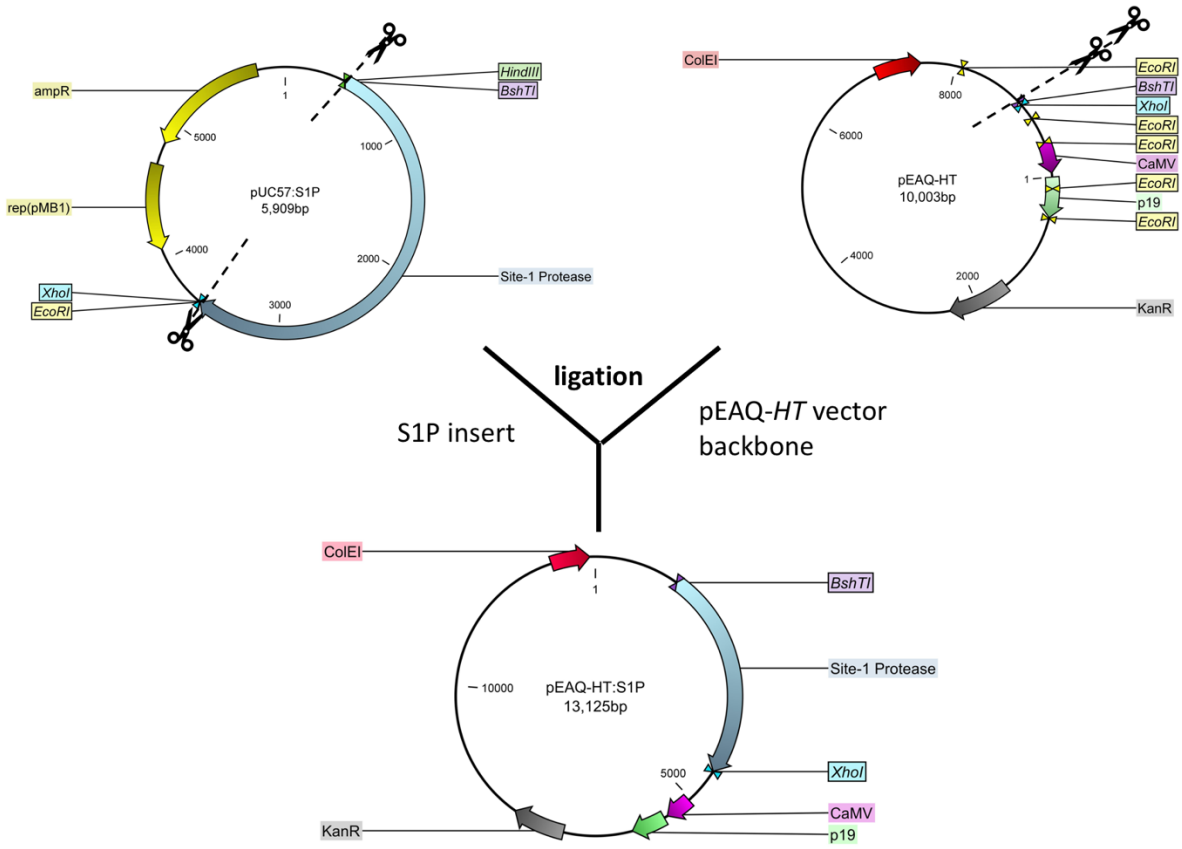


Figure 3.3: Cloning strategy to produce pEAQ-*HT:* S1P from pUC57:S1P and pEAQ-*HT*.S1P was obtained from GenScript in the pUC57 plasmid. pEAQ-*HT* was used as a vector backbone for which to clone S1P in for plant expression. pUC57 contains elements: ampR- Ampicillin/ Carbenicillin resistance, rep(pMB1)- *E. coli* replication site. pEAQ-*HT* contains elements: KanR- Kanamycin resistance, CaMV-35S Cauliflower mosaic virus promoter, ColE1- Origin of replication for *E. coli,* p19- tomato bushy stunt virus RNA silencing suppressor. *Both* plasmids were made to be compatible using *Bsh*TI and *Xho*I endonuclease digests. The fragments were then ligated together to produce the vector pEAQ-*HT*: S1P

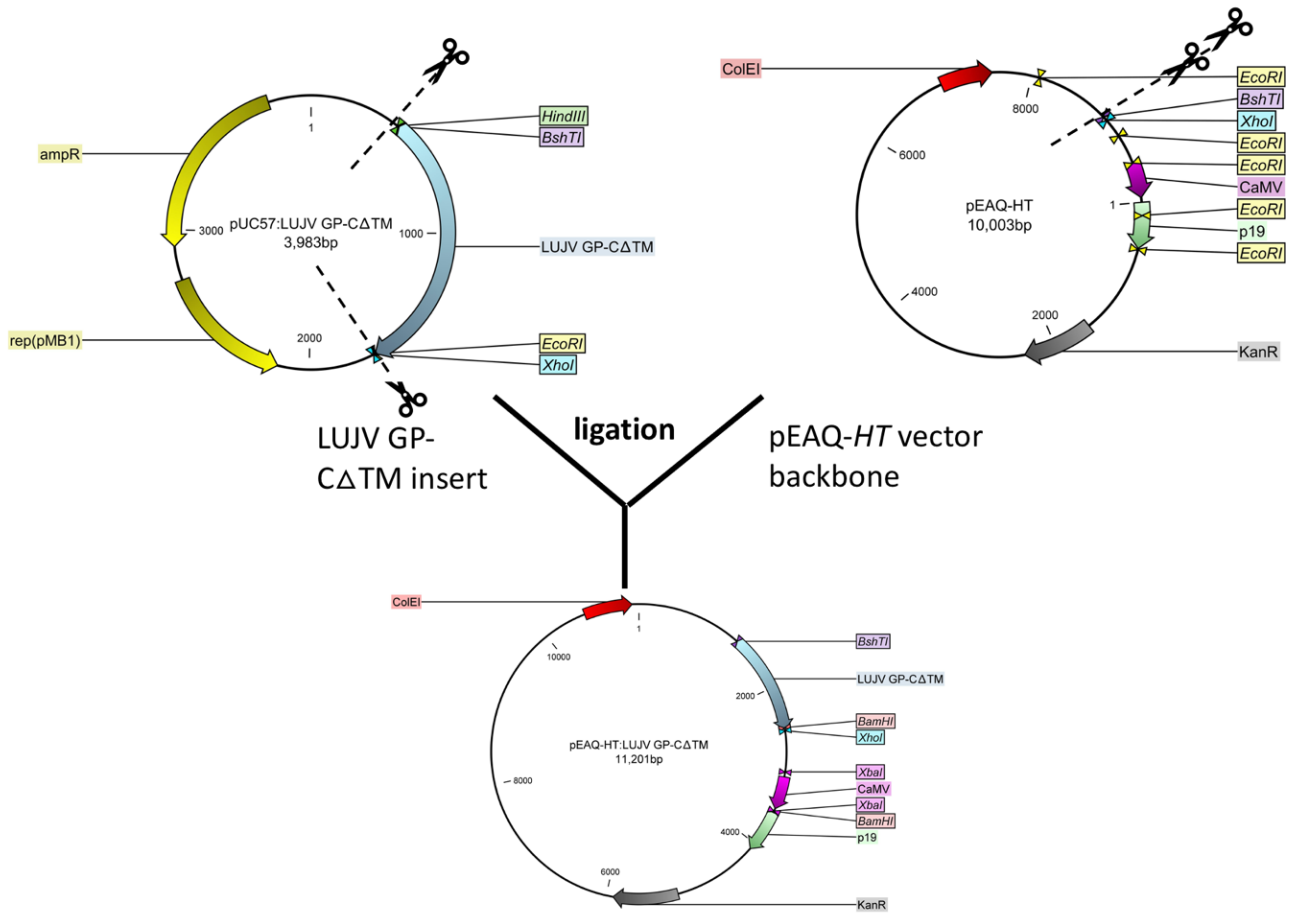


Figure 3.4: Cloning strategy to product pEAQ-*HT:* LUJV GP-CΔTM from pUC57: LUJV GP-CΔTM and pEAQ-*HT.* The LUJV GP-CΔTM sequence obtained from GenScript inside the pUC57 plasmid described in Chapter 2 was reused. pUC57 contains elements: ampR- Ampicillin/ Carbenicillin resistance, rep(pMB1)- *E. coli* replication site. pEAQ-*HT* contains elements: KanR- Kanamycin resistance, CaMV-35S Cauliflower mosaic virus promoter, ColE1- Origin of replication for *E. coli,* p19- tomato bushy stunt virus RNA silencing suppressor. LUJV GP-CΔTM was cut out from the pUC57 plasmid and inserted into the plant expression vector backbone pEAQ-*HT. Bsh*TI and *Xho*I were used for endonuclease digests to make compatible fragments for cloning. Scissors indicate cut points. The fragments were then ligated together using T4 DNA ligase and then transformed into *E. coli* for propagation and storage in glycerol stocks.

3.3.1 Transformation and restriction digest confirmation of recombinant *A.tumefaciens* strains

Following confirmation of successful plasmid cloning, the next step was to transform the plant expression vectors pEAQ-*HT*: LUJV GP-CΔTM and pEAQ-*HT*: S1P into *A. tumefaciens* in preparation for *A. tumefaciens* mediated transformation of *N. benthamiana.* Unlike the pTHpCapR vector, pEAQ-*HT* cannot be used to transfect plant cells directly and thus the new modified pEAQ vectors were transformed into *A. tumefaciens* (AGL1 strain) to mediate transfection of the intended DNA to live *N. benthamiana* plants. Electrocompetent *A. tumefaciens* AGL1 cells were transformed with both pEAQ-*HT* constructs via electroporation. To confirm AGL1 transformation, colony PCR using primers that bind just upstream and downstream of the insert region were used to confirm amplification of S1P and LUJV GP-CΔTM respectively. The PCR results are shown in **Figures 3.5** and **3.6**.

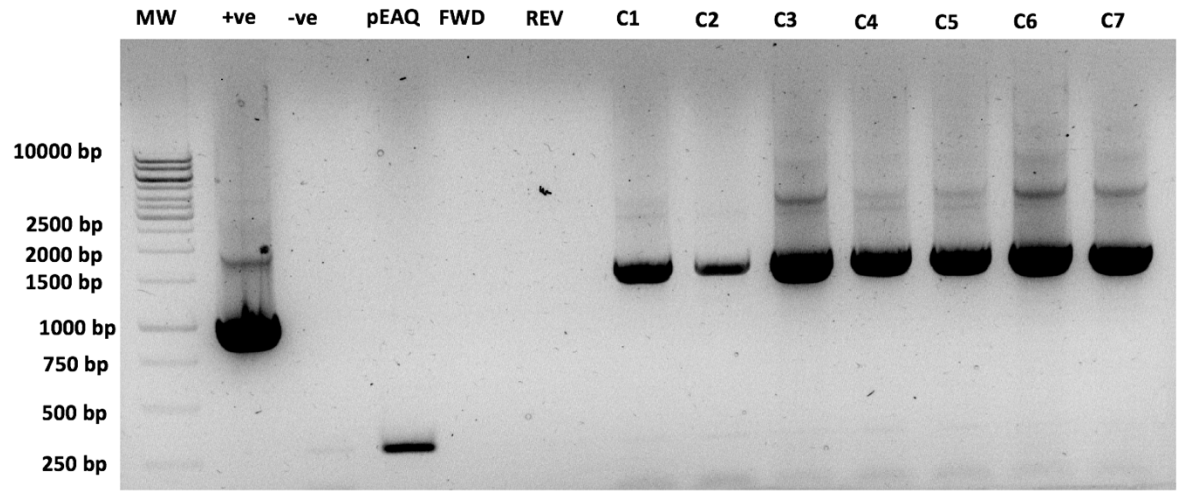


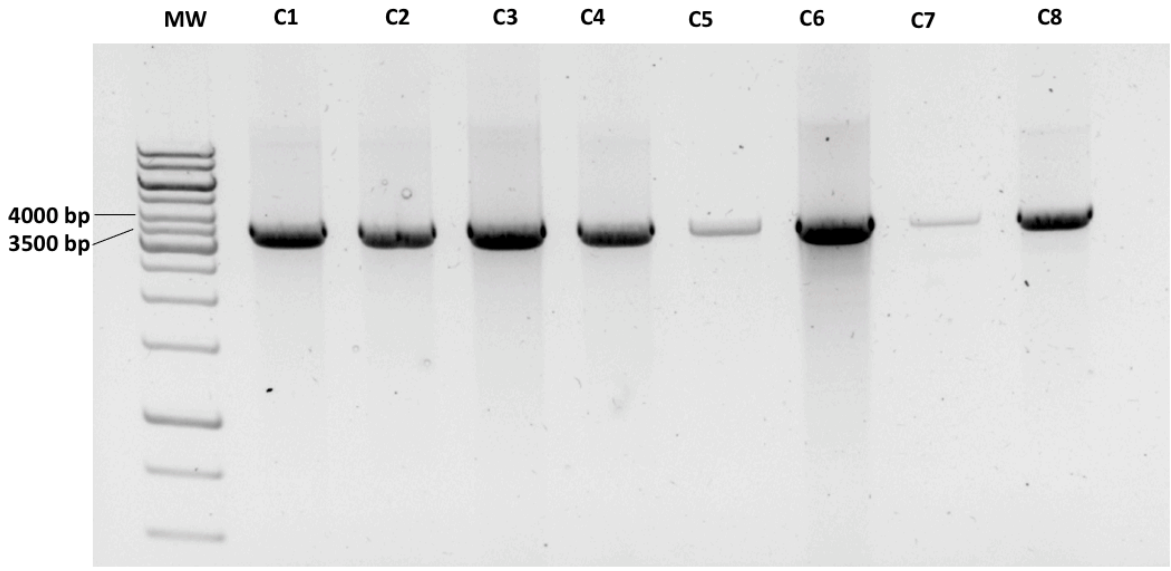
Figure 3.5: PCR on *A.tumefaciens* AGL1 to confirm successful transformation with plant expression vector pEAQ-*HT*: LUJV GP-CΔTM. 1% Agarose gel with PCR products from a pEAQ-*HT* primer that amplifies over the insert region (i.e LUJV GP-CΔTM). Electrophoresis was performed directly on a toothpick prick of bacterial colonies. PCRs were conducted using ImmoMix™ (Bioline, UK) and in 25 µl volumes. 1KB DNA ladder, +ve - pEAQ:Shuni N, -ve- no DNA template control, pEAQ only vector, Forward primer only mix, Reverse primer only mix, AGL1: pEAQ-*HT:* LUJV GP-CΔTM colonies C1-C7. The LUJV GP-CΔTM insert expected at 1261 bp. FWD primer sequence: 5'-TTC TTC TTC TTG CTG ATT GG-3'. REV primer: 5'- CAC AGA AAA CCG CTC ACC-3'.

The positive control in **Figure 3.5** contained a PCR reaction amplifying a Shuni N protein cloned into the pEAQ-*HT* vector and was expected at 1000 bp (+ve). Nothing was seen for the negative control (-

ve, no DNA template) and forward/reverse only primer mix (FWD and REV). The pEAQ only DNA template reaction (lane 4) showed a 250 bp product as expected, when no insert is present in the vector. The PCR on putatively transformed AGL1 colonies all yielded a product consistent with the expected size of LUJV GP-CΔTM around 1500 bp (1261 bp for LUJV GP-CΔTM as well as the 250 bp segment from the pEAQ-*HT* vector as seen in the +ve control) confirming successful transformation (C1-C7). An extra band of approximately 2500 bp was also visible and was possibly due to non-specific binding of primers. Cells from colony 2 were chosen for subsequent experiments. **Figure 3.6** shows the PCR confirmation of successful cloning of pEAQ-*HT:* S1P into *A. tumefaciens* AGL1.

Figure 3.6: PCR on *A.tumefaciens* AGL1 to confirm successful transformation with plant expression vector pEAQ-*HT: S1P*. 1% Agarose gel with PCR products from a pEAQ-*HT* primer that amplifies over the insert region (i.e amplifies S1P). 1KB DNA ladder, AGL1: pEAQ-*HT*: S1P colonies C1-C8. S1p insert expected at 3187 bp.

The PCR on AGL1: pEAQ-*HT*: S1P (**Figure 3.6**) was performed and the resulting colonies showed the expected band size of 3187 bp. Colony 3 was chosen for subsequent experiments. The controls used in **Figure 3.6** were performed in parallel and under the same reaction conditions as **Figure 3.5** so were not included in **Figure 3.6** as well.



3.3.2 Molecular chaperone co-expression

One of the challenges in producing LUJV GP-CΔTM was to determine what molecular chaperones were needed to facilitate its protein folding and production. In an attempt to overcome this, a recombinant *A. tumefaciens* strain harbouring the molecular chaperone calreticulin (CRT) was co-infiltrated with pEAQ- *HT*: LUJV GP-CΔTM as it was previously shown to improve the production of HIV gp140 vaccine candidates in plants [122]. A strain harbouring calnexin (CNX) was also included in the co-expression experiments as an additional molecular chaperone which is involved closely with CRT [134]. CNX and CRT gene sequences were previously constructed by the research group and cloned into the pEAQ-*HT* vector to create pEAQ-*HT*:CRT and pEAQ-*HT*: CNX. These constructs were donated for use in this study and not re-constructed. To date, there is no published account describing the production of LUJV proteins in plant cells. I initially investigated a few useful variables associated with this type of production including optimal DPI, molecular chaperone requirements and necrotic effects of the protein on the plant cells, all of which were previously unknown. In a pilot experiment, *N. benthamiana* was transfected with LUJV GP-CΔTM and plant material was harvested on days 3 and 5 post infiltration (**Figure 3.7**).

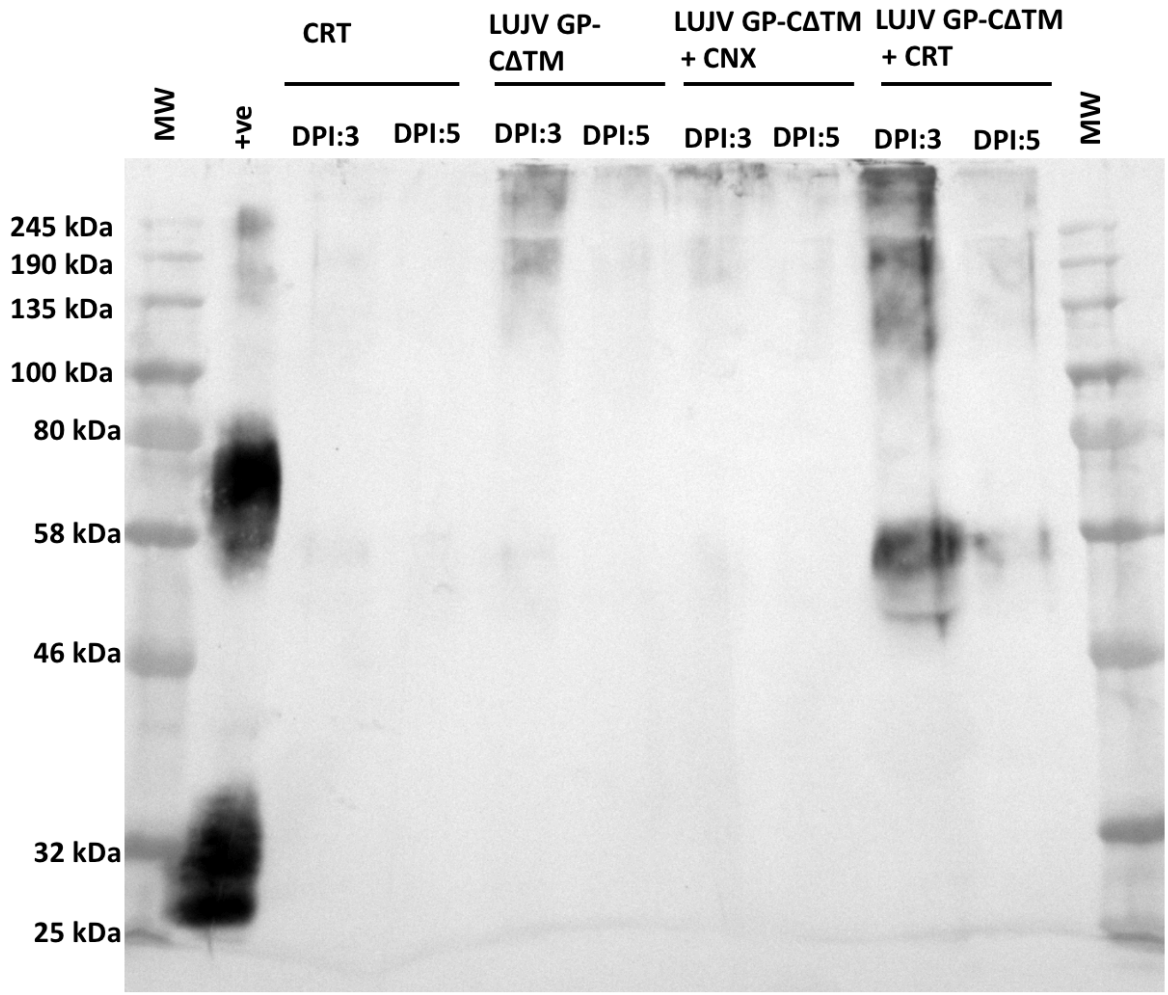


Figure 3.7: Western blot of LUJV GP-CΔTM expression in *N. benthamiana* and the effects of molecular chaperone co-expression on protein production. 12% western blot showing pilot experiments of LUJV GP-CΔTM protein production in plants. Antibody: primary mouse anti-6X Histidine (Dilution 1:2000, Cat: MCA1396, Bio-Rad, USA) and secondary goat anti-mouse IgG conjugated to AP (Dilution 1:5000, Cat: A3562, Sigma, USA) for LUJV GP-CΔTM. LUJV GP-CΔTM was co-expressed with molecular chaperones CNX and CRT, all three proteins of which were delivered through *A.tumefaciens*-mediated transformation and from pEAQ-*HT* vectors.Each expression construct combination was harvested on days 3 and 5 post infiltration (DPI). 245 kDa molecular weight marker, positive control- LUJV GP-CΔTM produced in HEK293T cells (see Chapter 2), CRT only control DPI 3, CRT only control DPI 5, LUJV GP-CΔTM DPI 3, LUJV GP-CΔTM DPI 5, LUJV GP-CΔTM+ CNX DPI 3, LUJV GP-CΔTM+ CNX DPI 5, LUJV GP-CΔTM+CRT DPI 3, LUJV GP-CΔTM+ CRT DPI 5, 245 kDa molecular weight marker .LUJV GP-CΔTM is expected at 47 kDa but found to present slightly higher, likely due to glycan groups increasing the protein weight.

When expressed alone, the LUJV GP-CΔTM protein was not seen near its expected size of 47 kDa (**Figure 3.7**). Instead, a smear of protein was seen towards the top of the blot (>245 kDa) on DPI 3. Previously this smear was seen associated with protein aggregates when producing plant made proteins such as HIV gp140 which may be due to misfolding and clumping of proteins that affects the ability of proteins to migrate in the agarose gel [136]. LUJV GP-CΔTM was co-expressed with the molecular chaperone's CRT and CNX to see if this would improve protein production (**Figure 3.7)**. The positive control seemed to be too concentrated (+ve, LUJV GP-CΔTM extracted from HEK293T cells) to clearly define LUJV GP-CΔTM and the GP2 cleavage product as described in Chapter 2, but served as a general size guide for the plant produced LUJV GP-CΔTM. The GP2 cleavage product around 28 kDa was not seen in any plant produced LUJV GP-CΔTM. Nothing was seen when CRT was expressed alone in *N. benthamiana* as a transfection control. This is expected as the CRT construct does not have a His tag for detection. A molecular chaperone may improve this occurrence if this is the case. When CNX was co-expressed (LUJV GP-CΔTM+ CNX), there did not seem to be an improvement for the LUJV GP-CΔTM protein production, however, the protein aggregates towards the top of the blot (DPI 3) seemed to have decreased, which suggests CNX does reduce misfolding. It was determined in Chapter 2 that glycosylation of LUJV GP-CΔTM occurs in HEK293T cells, so it was possible that this occurs in plant cells as well to a lesser extent. Finally, when co-expressed with CRT (LUJV GP-CΔTM+CRT), the LUJV GP-CΔTM protein seemed to appear around 58 kDa (expected size at 47 kDa). It is likely that the protein is seen slightly higher than the expected size due to extra weight from glycan groups. This was the case for mammalian produced LUJV GP-CΔTM Chapter 2 (**Figure 2.10**), however, it seems that it occurs to a lower extent in *N. benthamiana* which is inferred by the LUJV GP-CΔTM protein being slightly lighter than the mammalian produced one. Additionally, the LUJV GP-CΔTM protein was only detected when the plants were harvested on day 3 (LUJV GP-CΔTM+ CRT, DPI:3). The results showed that the LUJV GP-CΔTM required co-expression with CRT to be produced at a detectable yield that could be observed by western blot.

Images of *N. benthamiana* were taken after recombinant protein expression to give an indication of necrosis and general health of the plant. Visible plant pathology was used in conjunction with the protein blots (3.3.4) to infer concepts about the protein expression. In the investigation into the effects of LUJV GP-CΔTM on the pathology of plants, plants were monitored for 5 days and harvested on DPI 3 and 5. The OD_{600} (optical density) at which vector transformed *A. tumefaciens* was used was also noted. **Figure 3.8** shows pictures of *N. benthamiana* leaves after recombinant protein expression (western blot results seen in Figure 3.7)

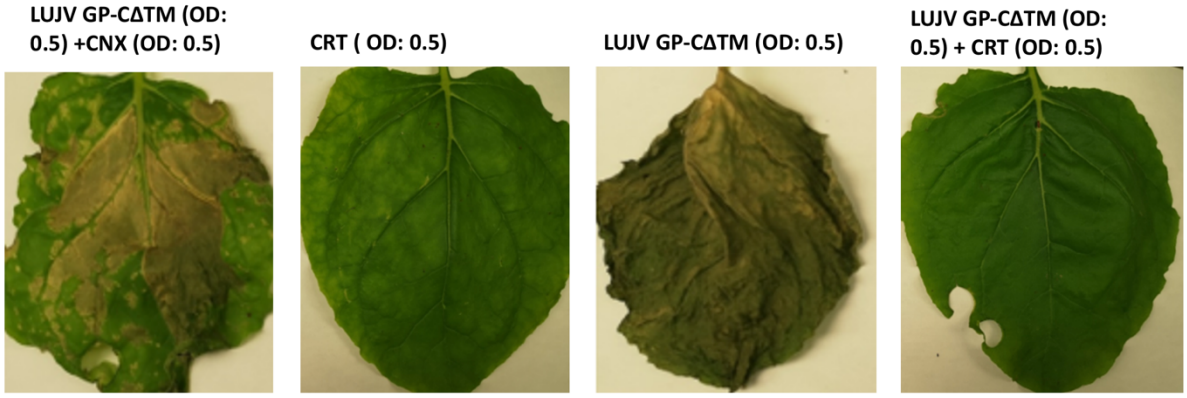


Figure 3.8: *N. benthamiana* pathology after recombinant protein expression with LUJV GP-CΔTM, CRT and CNX. Pictures were taken on day 3 post infiltration. All plants were grown in the same conditions and planted at the same time. Necrosis is characterized by discoloration and shriveling of leaves. For small scale plant experiments, clippings of leaves were taken.

A.tumefaciens containing the LUJV GP-CΔTM expression vector, and infiltrated at an OD_{600} of 0.5, caused necrosis of infiltrated leaves (**Figure 3.8**). This effect was to some extent diminished by CNX co-expression and reduced considerably more when co-expression with CRT. These findings substantiate the results seen in **Figure 3.10** where LUJV GP-CΔTM was not detected without the co-expression of CRT. The extent of necrosis seemed to correlate with poor accumulation of protein. Previously, co-expression of levels of BiP, PDI and bZIP60 (ER stress response markers) were tested in plants co-expressing CRT and the markers decreased by around 2 fold when supplemented with CRT relative to controls [122]. This corroborates with the data observed in this work and signifies the utility of using CRT to reduce ER stress and potentially misfolding. The experiment was repeated with LUJV GP-CΔTM transformed *A.tumefaciens* infiltrated with a bacterial culture at a lower OD to see if this would reduce plant toxicity (**Figure 3.12**).

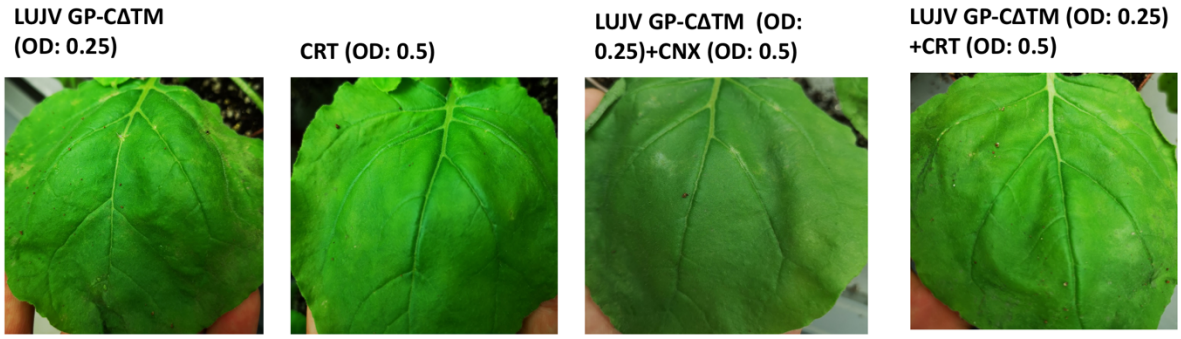


Figure 3.9: *N. benthamiana* pathology after *A.tumefaciens*-mediated recombinant protein expression with LUJV GP-CΔTM, CRT and CNX. Pictures were taken on day 3 post infiltration. This experiment was performed in the same manner as in Figure 3.9 except *A.tumefaciens,* containing the LUJV GP-CΔTM expression vector, was infiltrated at a lower OD_{600} (0.25 compared to 0.5 previously) in an attempt to offset the shriveling and discoloration of leaves associated with necrosis.

A reduction in OD from 0.5 to 0.25 seemed to considerably reduce the necrotic effect and discoloration of leaves seen in **Figure 3.8**. In **Figure 3.9** the plants retained a healthy green physiology, which implied the plant cells underwent reduced ER stress. These parameters were recorded and maintained for subsequent experiments i.e *A.tumefaciens* containing the CRT expression vector (Infiltrated at an OD_{600} of 0.5) as a molecular chaperone, *A.*tumefaciens containing the LUJV GP-CΔTM expression vector and infiltrated at an OD_{600} of 0.25 and harvesting of LUJV GP-CΔTM on day three as the most conclusive protein accumulation was on this day.

As detailed in Chapter 1 and 2, It is of great importance that arenavirus glycoproteins undergo proteolytic cleavage for maturation. To encourage in vivo proteolytic cleavage of LUJV GP, pEAQ-*HT*: LUJV GP was co-infiltrated with pEAQ-*HT*: S1P. Initially, pEAQ-*HT:* S1P was co-infiltrated at an OD_{600} of 0.25 along with LUJV GP-CΔTM and CRT to determine whether LUJV GP-CΔTM could be cleaved into GP1 and GP2 in vivo.

3.3.3 Proteolytic cleavage attempts of LUJV GP-CΔTM through S1P co-expression

After assessing the potential plant pathology cause by S1P recombinant expression, proteins were extracted and subjected to a western blot (**Figure 3.14)**.

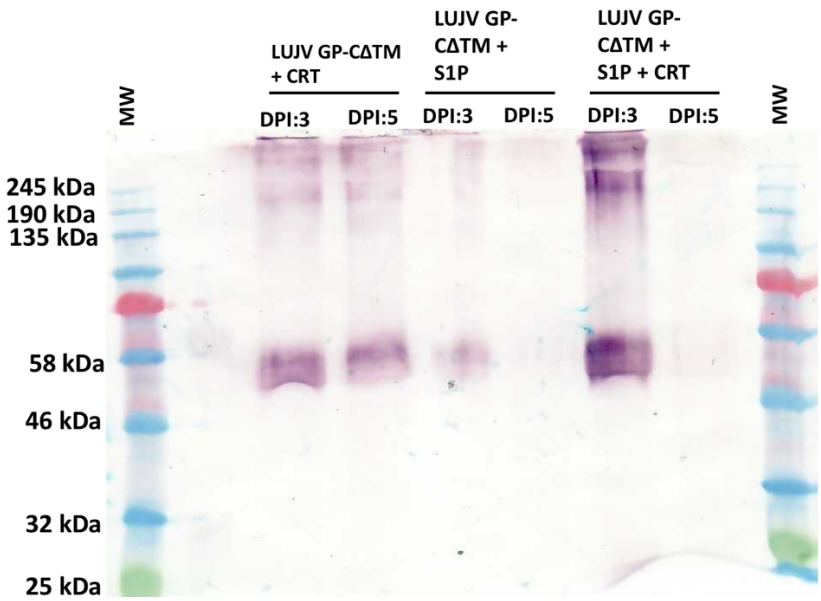


Figure 3.10: Western blot of LUJV GP-CΔTM cleavage attempt following S1P co-expression. 12% Western blot of proteins extracted from *N. Benthamiana* following recombinant protein expression experiments. Antibody: primary mouse anti-6X Histidine (Dilution 1:2000, Cat: MCA1396, Bio-Rad, USA) and secondary goat anti-mouse IgG conjugated to AP (Dilution 1:5000, Cat: A3562, Sigma, USA) for LUJV GP-CΔTM. Plant protein was harvested on days 3 and day 5 post infiltration. 245 kDa broad range molecular weight marker, LUJV GP-CΔTM+ CRT DPI 3, LUJV GP-CΔTM+ CRT DPI 5, LUJV GP-CΔTM+ S1P DPI: 3, LUJV GP-CΔTM+ S1P DPI: 5, LUJV GP-CΔTM + CRT + S1P DPI: 3, LUJV GP-CΔTM + CRT + S1P DPI:5. 245 kDa broad range molecular weight marker.

Co-expression with S1P increased the band intensity of LUJV GP-CΔTM as well as the protein aggregates > 245 kDa on DPI 3 when co-expressed with CRT (**Figure 3.10**, LUJV GP-CΔTM + CRT+S1P, DPI:3). This may suggest that S1P co-expression increases the accumulation of LUJV GP-CΔTM when in the presence of CRT. LUJV GP-CΔTM + CRT expression was used as a control and the results remained as detected previously (Figure 3.7) with a LUJV GP-CΔTM product at 58 kDa (**Figure 3.10**, LUJV GP-CΔTM+ CRT, DPI: 3 and 5). However, when expressed with S1P alone (LUJV GP-CΔTM + S1P, DPI:3 and 5), the LUJV GP-CΔTM seemed to decrease in detection relative to the LUJV GP-CΔTM+CRT

samples. Western blots are not quantitative but equal amounts of total soluble protein were added to each well allowing a qualitative indication of protein expression. Additionally, S1P seemed to not contribute to the cleaving the LUJV GP-CΔTM as no cleavage product was seen around the expected size for GP2 at 32kDa (**Figure 3.10**, LUJV GP-CΔTM + CRT + S1P). In the LUJV GP-CΔTM + CRT + S1P samples all three proteins were expressed but with no significant changes to cleavage of LUJV GP-CΔTM. An intermediary experiment probing for S1P with a polyclonal antibody, from the samples seen in **Figure 3.10**, yielded a blank western blot with no bands (result not shown). The visual plant physiology on day three for the S1P co-expression experiment (**Figure 3.10**) is shown in **Figure 3.11** below. Co-infiltration with S1P (**Figure 3.11**) did not seem to cause a distinctive pathology relative to the LUJV GP-CΔTM+ CRT control and on a visual level the plants seemed to be in a suitable condition for protein harvesting (i.e., not necrotic).

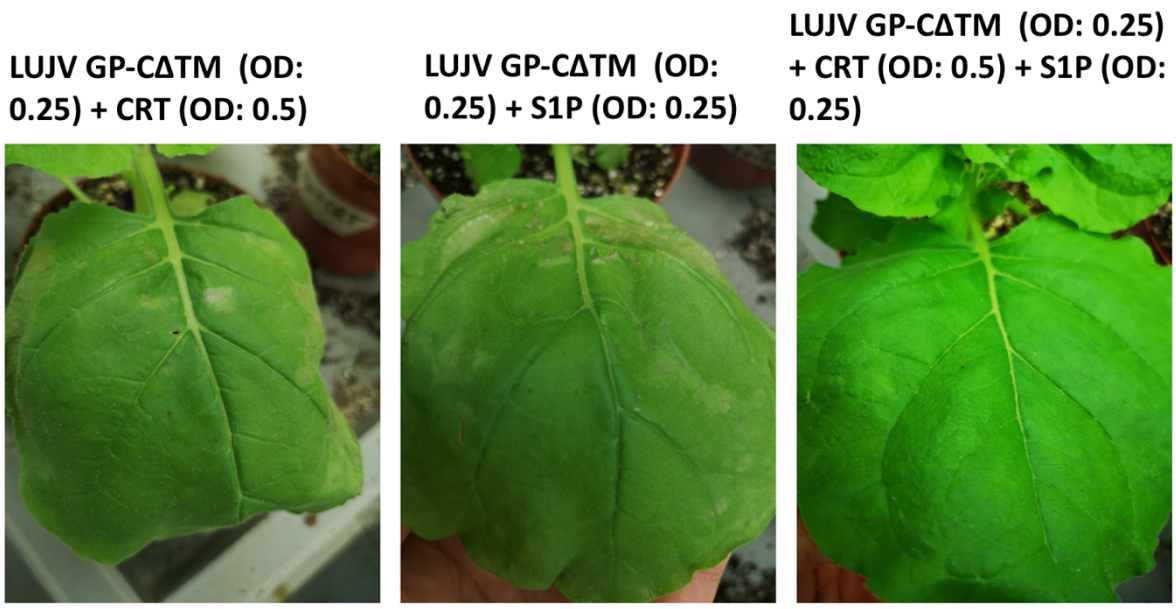


Figure 3.11: *N. benthamiana* pathology after *A. tumefaciens* mediated recombinant protein expression with LUJV GP-CΔTM, CRT and S1P. Protein co-expression experiments were repeated as previously with the addition of S1P as an expression construct. Pictures were taken on day three post infiltration.

At this time, it seemed as though S1P may have been accumulating at low levels. S1P has a known transmembrane region making it somewhat hydrophobic and thus it was possible it could not be

extracted sufficiently at this pH. It was not clear at what pH S1P would extract optimally thus S1P and LUJV GP-CΔTM was expressed in *N. benthamiana* as performed previously and then extracted into buffers of varying pH in an attempt to visualize any S1P present. S1P has a theoretical iso-electric point of 8.91 (protopram analysis) and was expected to be visible in a buffer at pH 8.5 if it was being expressed. Unfortunately, despite being extracted in 5 different buffer pHs (pH 6.5-10.5), S1P was not visualized on a western blot in any sample when co-expressed with LUJV GP-CΔTM or CRT (results not shown).

At this time, it was suggested that S1P was being expressed at levels too low for detection. To try and increase yields, an ammonium sulfate precipitation of crude leaf extract from plants infiltrated with S1P was carried out. This involves slowly adding small amounts ammonium sulfate that binds to water molecules and outcompetes the protein, making it more likely that proteins with hydrophobic regions such as S1P may be more stable at these solubility levels and thus be visualized. It is also a means of concentrating protein to increase the chances of visualizing S1P on a western blot. In parallel to this, varying OD_{600}'s of S1P containing *A.tumefaciens* cultures were tested in an attempt to find an optimal balance of increased protein expression and reduced necrosis.

N. benthamiana plants were agroinfiltrated with S1P at an OD_{600} of 0.5 and 1 and crude leaf extracts subjected to an ammonium sulfate precipitation in increasing increments of 20% in an attempt to concentrate S1P and visualize it on a western blot. The expected band size of 117 kDa was still not present on the western blot suggesting that the S1P was not being expressed in *N. benthamiana*. The lack of observation of proteolytically cleaved LUJV GP-CΔTM when co-expressed with S1P (**Figure 3.10**) supports this conclusion.

Upon further research it was discovered that specific glycans were needed for LUJV GP-CΔTM to be in the correct conformation for recognition of S1P and subsequent cleavage [154]. S1P may not have recognized LUJV GP-CΔTM for cleavage due to under-glycosylation in plants. It was not clear that increasing the glycan occupancy on LUJV GP-CΔTM would necessarily result in cleavage by S1P given poor detection levels of the protease previously (results not shown). However, it was plausible enough to continue experimentation in an attempt to subject LUJV GP-CΔTM to proteolytic cleavage. Previously, under-glycosylation was remedied by using an oligosaccharyl transferase LmSTT3D isolated from *Leishamania major* [116]. Co-expression of this enzyme previously showed increased glycosylation of transiently-expressed human protein [116]. In an attempt to conduct proteolytic cleavage and remedy the suspected under-glycosylation in plants, LmSTT3D was co-expressed with LUJV GP-CΔTM. Co-expression of LmSTT3D with LUJV GP-CΔTM would also allow an investigation into

the glycosylation ability of LmSTT3D. Additionally, a different plant cell line ΔXT/FT was incorporated into the study as having a deficiency for producing plant-like xylose and fucose glycans. This was chosen with the intention of facilitating a more mammalian-like glycosylation profile thus improving the prospects for cleavage of LUJV GP-CΔTM by S1P. The results are shown in **Figure 3.12**.

3.3.4 Glyco-optmization of transiently expressed LUJV GP-CΔTM in *N. benthamiana*

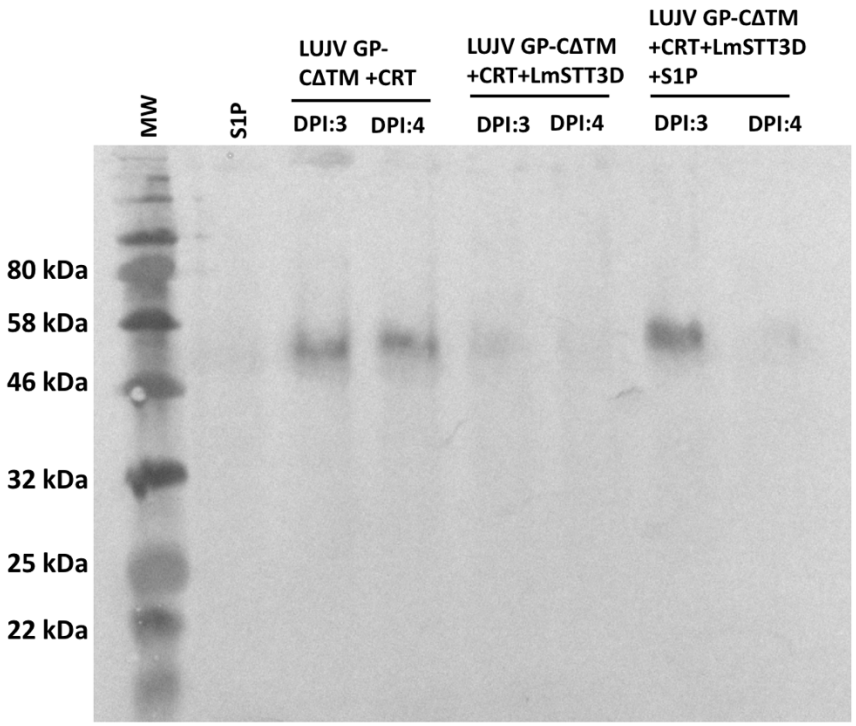


Figure 3.12: Western blot probing for LUJV GP-CΔTM to assess the glycosylation utility of lmSTT3D in an attempt to increase glycan occupancy of LUJV GP-CΔTM for recognition and cleavage by S1P. 12% western blot showing the LUJV GP-CΔTM after co-expression with the oligosaccharyl transferase lmsTT3D (OD_{600}: 0.25) isolated from *Leishmania major* and previously optimized auxiliary co-expression proteins. 245 kDa broad range molecular weight marker, S1P negative control, LUJV GP-CΔTM + CRT, LUJV GP-CΔTM +CRT+ LmSTT3D DPI 3 and 4, LUJV GP-CΔTM+CRT+S1P+LmSTT3D experiment DPI 3 and 4. Antibody: primary mouse anti-6X Histidine (Dilution 1:2000, Cat: MCA1396, Bio-Rad, USA) and secondary goat anti-mouse IgG conjugated to AP (Dilution 1:5000, Cat: A3562, Sigma, USA) for LUJV GP-CΔTM

The S1P negative control showed no bands as expected (**Figure 3.12**, S1P). Co-expression of S1P and LUJV GP-CΔTM with LmSTT3D did not seem to affect the cleavage status of LUJV GP-CΔTM as only the un-cleaved LUJV GP-CΔTM was seen faintly just below 58 kDa in all LUJV GP-CΔTM samples. LmSTT3D cannot be detected on the western blot, however, functionally the addition of glycans causes an increase in approximately 3 kDa per glycan added resulting in the increased size shift seen in the LmSTT3D co-expressed samples. This was difficult to observe in **Figure 3.12.** There were no cleavage products of LUJV GP-CΔTM and no further attempts at proteolytic processing were made. In the final optimization experiments the focus was shifted to increasing glycan occupancy on LUJV GP-CΔTM. The experiment in **Figure 3.12** underwent a technical repetition to better visualize the LUJV GP-CΔTM better after being co-expressed with LmSTT3D as the protein may have degraded during the extraction process in **Figure 3.12.**

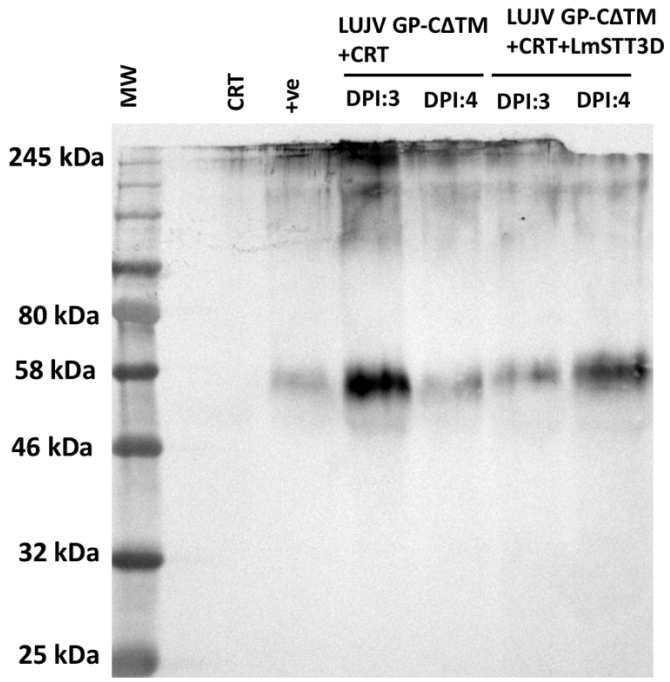


Figure 3.13: Western blot probing for LUJV GP-CΔTM to assess the glycosylation utility of lmSTT3D in an attempt to increase glycan occupancy of LUJV GP-CΔTM 12% western blot showing the LUJV GP-CΔTM after co-expression with the oligosaccharyl transferase lmsTT3D (OD_{600}: 0.25) isolated from *Leishmania major* and previously optimized auxiliary co-expression proteins. 245 kDa broad range molecular weight marker, empty lane, CRT negative control, LUJV GP-CΔTM + CRT protein expressed form a previous infiltration (positive control), LUJV GP-CΔTM +CRT day 3 and day 4, LUJV GP-CΔTM

+CRT+ LmSTT3D DPI 3 and 4. Antibody: primary mouse anti-6X Histidine (Dilution 1:2000, Cat: MCA1396, Bio-Rad, USA) and secondary goat anti-mouse IgG conjugated to AP (Dilution 1:5000, Cat: A3562, Sigma, USA) for LUJV GP-CΔTM.

The results in **Figure 3.13** show an expected empty lane for the negative control with CRT only. LUJV GP-CΔTM +CRT was expressed as a positive control on days 3 and day 5 as has been repeated many times in this study (**Figure 3.13**, +ve). When co-expressed with LmSTT3D, LUJV GP-CΔTM seems to increase very slightly in size (LUJV GP-CΔTM +CRT + LmSTT3D, DPI:3 and 4). This size shift could be attributed to addition of glycan groups onto LUJV GP-CΔTM through the action of LmSTT3D. **Figure 3.14** is a technical repeat of **Figure 3.13** except the SDS gel was run for longer to make the size difference clearer. LUJV GP-CΔTM seemed to be optimal to harvest on day 3 but when co-expressed with LmSTT3D, optimal harvest seems to be on day 4 based on this day being optimal for LmSTT3D glycosylation of proteins (determined previously, unpublished data). Subsequent experiments were carried out taking into account those parameters.

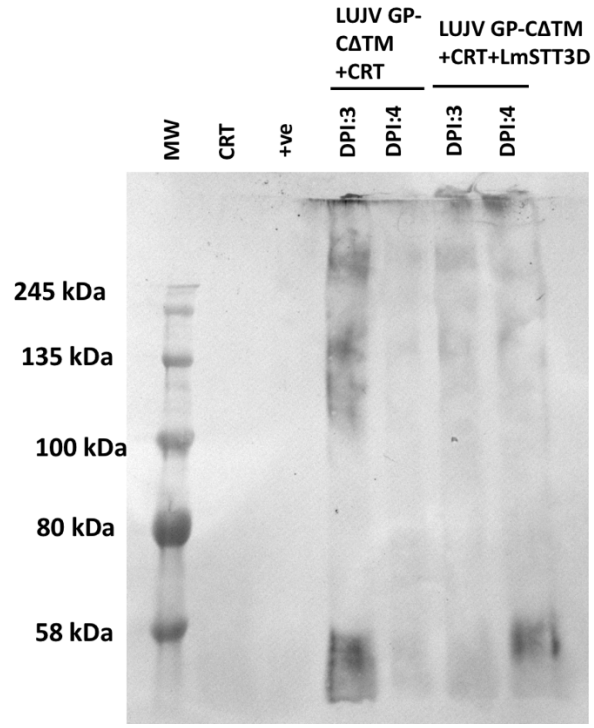


Figure 3.14: Western blot probing for LUJV GP-CΔTM to assess the glycosylation utility of lmSTT3D in an attempt to increase glycan occupancy of LUJV GP-CΔTM 10% western blot showing the LUJV GP-CΔTM after co-expression with the oligosaccharyl transferase lmsTT3D (OD_{600}: 0.25) isolated from

Leishmania major with previously optimized expression conditions. 245 kDa broad range molecular weight marker, CRT negative control, LUJV GP-CΔTM + CRT protein expressed form a previous infiltration (positive control), LUJV GP-CΔTM +CRT day 3 and day 4, LUJV GP-CΔTM +CRT+ LmSTT3D DPI 3 and 4. Antibody: primary mouse anti-6X Histidine (Dilution 1:2000, Cat: MCA1396, Bio-Rad, USA) and secondary goat anti-mouse IgG conjugated to AP (Dilution 1:5000, Cat: A3562, Sigma, USA) for LUJV GP-CΔTM.

Figure 3.14 shows a more distinctive size shift between the LUJV GP-CΔTM +CRT and the LUJV GP-CΔTM +CRT + LmSTT3D samples increasing the evidence of LmSTT3D being a reliable glycosylation tool in plants.

3.3.5 Large scale plant protein expression and purification

Using all the optimized parameters for producing glyco-optimized LUJV GP-CΔTM in plants, a large-scale experiment with infiltration of 60 plants was performed to assess the viability of large-scale production (**Figure 3.15**) and purification using HisPur™ Cobalt Resin.

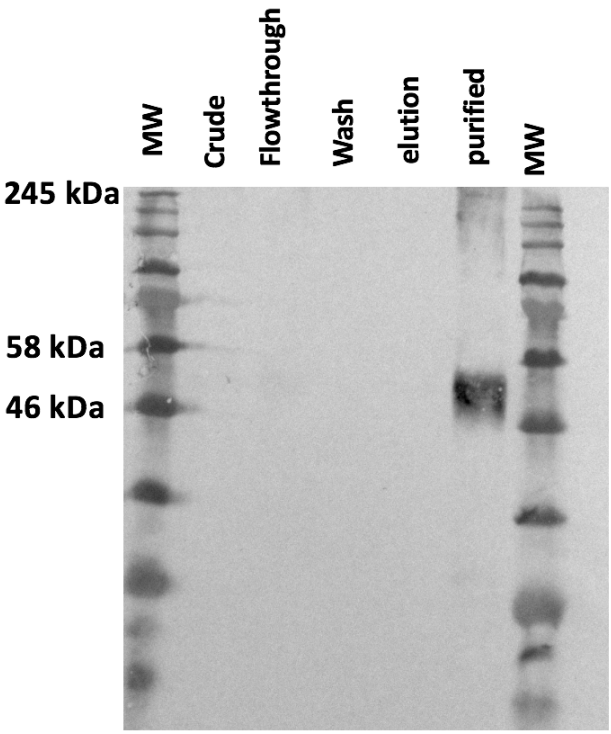


Figure 3.15: Western blot of large-scale purification of *N. benthamiana* produced LUJV GP-CΔTM through cobalt column purification. 12% western blot probing for his tag to visualize LUJV GP-CΔTM

produced in wild type *N. benthamiana*. Antibody: primary mouse anti-6X Histidine (Dilution 1:2000, Cat: MCA1396, Bio-Rad, USA) and secondary goat anti-mouse IgG conjugated to AP (Dilution 1:5000, Cat: A3562, Sigma, USA) for LUJV GP-CΔTM. Protein was purified using a 10 kDa spin column. 245 kDa broad range molecular weight marker, crude plant extract, flowthrough fraction, wash fraction, elution fraction, purified protein, 245 kDa molecular weight marker.

Figure 3.15 shows the results of the purified protein (quantified at 0.85 mg/kg plant biomass) with LUJV GP-CΔTM seen just above 46 kDa confirming the method of capture as viable for future harvesting of this protein using the optimized expression conditions (purified). The samples in **Figure 3.15** were subject to a Coomassie stain but no bands were visualized, likely due to the concentration of the LUJV GP-CΔTM being too low. Protein was not seen in any other lane possibly due to excess cellular proteins co-purified that cause a reduction in the signal of LUJV GP-CΔTM. The purification procedure in **Figure 3.15** was repeated using *N. benthamiana* ΔXT/FT plants and included co-expression of pEAQ- *HT:* $Hexo_3$RNAi – a construct that prevents unwanted glycan truncation by β-*N*-acetyl-hexosaminidases, thus facilitating a more mammalian-like glyco-environment [138]. (**Figure 3.16**).

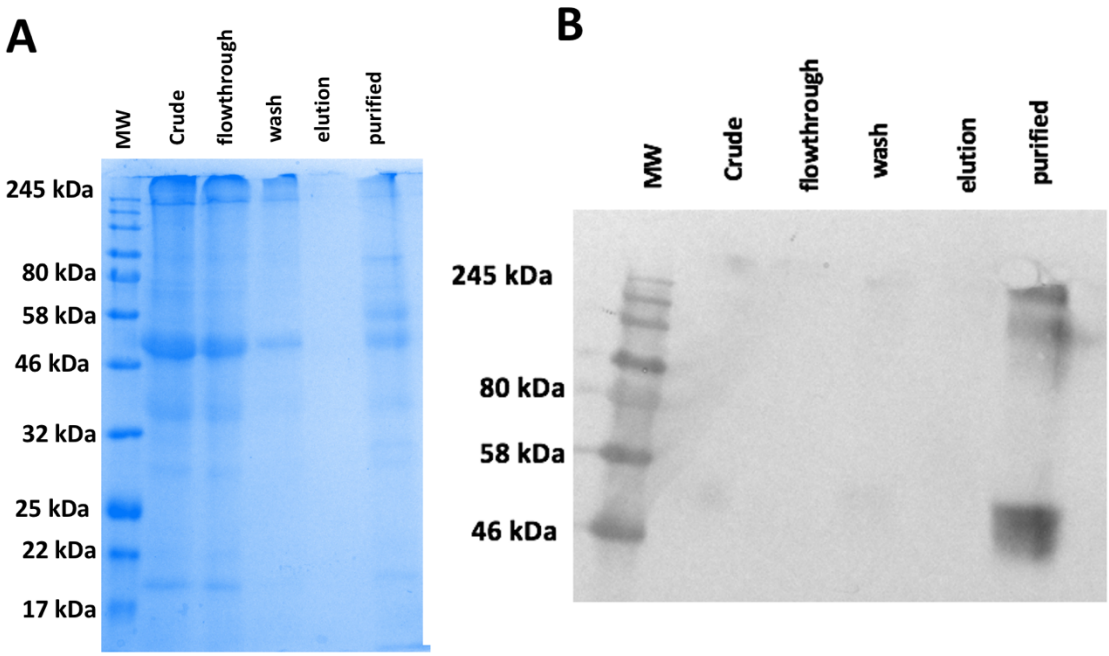


Figure 3.16: A Coomassie stained gel of large-scale cobalt column purified glycol-optimized LUJV GP-CΔTM produced in *N. benthamiana* ΔXT/FT plants. Coomassie SDS gel showing LUJV GP-CΔTM

purified after a large-scale protein infiltration (60 plants). LUJV GP-CΔTM transformed *A.tumefaciens* was co-expressed with CRT and lmSTT3D transformed *A.tumefaciens* as previously optimized. *A.tumefaciens* containing the *Hexo$_3$RNAi* expression vector, was infiltrated at an OD_{600} of 0.25. 245 kDa broad range molecular weight marker, crude plant extract, flowthrough fraction, wash fraction, elution fraction, purified protein. Protein was purified using a 10 kDa spin column. **B western blot of LUJV GP-CΔTM large scale purified from Coomassie produced in ΔXT/FT plants**. 12: western blot probing for his tag to visualize LUJV GP-CΔTM. Antibody: primary mouse anti-6X Histidine (Dilution 1:2000, Cat: MCA1396, Bio-Rad, USA) and secondary goat anti-mouse IgG conjugated to AP (Dilution 1:5000, Cat: A3562, Sigma, USA) for LUJV GP-CΔTM. 245 kDa broad range molecular weight marker, crude plant extract, flowthrough fraction, wash fraction, elution fraction, purified protein.

Many background bands were detected in the crude, wash and flowthrough fractions of the large scale LUJV GP-CΔTM purification from ΔXT/FT plants (**Figure 3.16 A**, crude, flowthrough, wash). These were likely excess plant proteins that were co-purified, however, it is expected based on the sheer volume of the protein extracted (3.1 mg/kg plant biomass). The LUJV GP-CΔTM protein was seen just above 46 kDa in these fractions. No bands were seen in the elution fraction as they are diluted in elution buffer but then concentrated once again in the purified fraction. Background detection above 245 kDa coincides with the size at which protein aggregates tend to accumulate at the top of the gel as mentioned previously. When subject to a western bot, the LUJV GP-CΔTM protein was faintly seen in the crude and wash fractions (**Figure 3.16 B**, crude, wash) The faint detection is likely due to an abundance of excess cellular protein co-purified as similarly seen in **Figure 3.15**. No bands were seen in the elution fraction as expected due to dilution. LUJV GP-CΔTM was visualized in the purified fraction around 46 kDa and protein aggregates >245 kDA were seen as in **Figure 3.15.**

3.4 Detailed discussion

I used the gene sequences described in Chapter 2 produce the plant expression vectors pEAQ-*HT*: LUJV GP-CΔTM and pEAQ-*HT*: S1P in an attempt to express recombinant LUJV GP-CΔTM and S1P in *N. benthamiana*. Using *A.tumefaciens* as plant infiltration tool, we set out to do a series of protein expression experiments in *N. benthamiana.* Parameters for optimal protein expression were recorded including the optimal OD of recombinant *A.tumefaciens* at the time of infiltration, protein folding machinery needed for optimal folding and expression, optimal day for harvesting of plants, type of purification column needed, prospects for proteolytic cleavage, potential plant lines for protein expression, and enzymes that enhance glycan occupancy of proteins. These aspects were investigated

to assess the differences in producing viral proteins in plant vs mammalian expression systems and note any commercially relevant differences between the two.

Initial experiments showed success in producing LUJV GP-CΔTM in *N. benthamiana.* LUJV GP-CΔTM was somewhat toxic to *N. benthamiana* as it induced necrosis and increased the accumulation of misfolded aggregates [152] as was seen in previous attempts with the HIV gp140 protein [136]. This is expected as LUJV GP-CΔTM is a non-native protein and likely contributes to ER stress and outwards pathology. However, co-expression with CRT offset significant levels of plant pathology and misfolded aggregates, making this a necessary auxiliary protein for LUJV GP-CΔTM expression in plants. Aggregated proteins have immunogenic effects depending on the composition and amount present [155]. Aggregations co-eluted down the line into vaccines may have an impact on the effectiveness of the vaccine which warrants investigation. CRT also increased the detection levels of LUJV GP-CΔTM in a similar manner to other proteins such as the HIV envelope protein as well as Hepatitis C, Influenza and Ebola proteins [122]. It is notable that *A. tumefaciens* also illicits a portion of the necrotic immune response seen in plants, [152], however, CRT, being a native protein in *N. benthamiana* was included as a control to correct for this. pEAQ-*HT:* CRT is a more reliable control for assessing the effects of transient protein expression than the empty pEAQ-*HT* vector which is sometimes used. While the blank vector will give an indication of the necrotic effects of *A. tumefaciens*, it does not give an indication of the stress due to expression of a recombinant protein. Visual assessment of the transfected plants showed that the optimal OD_{600} for LUJV GP-CΔTM was 0.25 while for CRT it was 0.5. The vacuum method used in this study offers a superior method of *A. tumefaciens* infiltration as it distributes the bacteria across all the leaves of the entire plant compared to syringe infiltration that has local infiltration of *A. tumefaciens* and uneven protein expression, making different samples difficult to compare.

Proteolytic cleavage of LUJV GP-CΔTM was attempted by co-transfecting the plants with S1P; however, this proved to be unsuccessful. Plant-produced S1P could not be detected on western blots using our extraction methods and even when using different extraction buffer pH's and OD_{600}'s it could not be detected. It is possible that S1P produced in plants was processed in a way that makes it undetectable by the S1P antibody. Attempts at increasing the glycan occupancy of LUJV GP-CΔTM to make it better recognized by S1P were also inconclusive. In all the experiments attempted here, S1P could not be made to elicit proteolytic cleavage of LUJV GP-CΔTM nor was it detected. This was evident by the absence of the cleavage product GP2 in western blots. Proteolytic cleavage experiments were aborted, and efforts were channeled into increasing the glycan occupancy of LUJV GP-CΔTM. Previously, *A.tumefaciens* contraining the *LmSTT3D* expression vector performed glycosyltransferase

activity in *N. benthamiana* when infiltrated at an OD_{600} of 0.1 [118]. My results suggest LmSTT3D may be effective in increasing glycan occupancy of LUJV GP-CΔTM when *A.tumefaciens* containing the LUJV GP-CΔTM expression vector are infiltrated at an OD_{600} of 0.25, however, a lower OD_{600} was not considered as outwards pathology of the plants seemed to be minimal even at this OD_{600}. This was deduced based on the size shift seen before and after co-expression with lmSTT3D, however, it is not clear which sites were glycosylated (out of 12 putative sites [55]) and to what extent these were plant-like glycans. The LUJV GP-CΔTM was detectable on DPI: 3 but seemed to drop significantly on DPI:5. Previously, LmSTT3D modified IgA antibodies expressed in *N. benthamiana* were harvested on DPI:4 [118]. Similarly, DPI:4 was the last day LUJV GP-CΔTM was detectable and was thus the optimal day for harvesting of protein when used in conjunction with LmSTT3D. $Hexo_3$RNAi was expressed to facilitate a more mammalian-like glycol-forms on LUJV GP-CΔTM. However, I had no method to investigate the extent to which this exercise had any effect.

LUJV GP-CΔTM was successfully scaled up and purified under commercially-relevant bulk production conditions (60 plants). LUJV GP-CΔTM was successfully captured using the HisPur™ Cobalt Resin described in Chapter 2. This purification protocol was utilized for both WT *N. benthamiana* transfections and glyco-optimized ΔXT/FT plants. Some cellular proteins were co-purified in the large-scale purification; however, further investigation is needed to determine whether these or the protein aggregates seen would have an effect on the immunogenicity of plant produced antigens.

This chapter documented a scalable method for producing LUJV GP-CΔTM in *N. benthamiana* which may be adaptable to other arenaviruses.

Chapter 4: Conclusion

Lujo virus causes a severe haemorrhagic fever with an 80% mortality rate and to date does not have any licensed vaccine or diagnostic tool available. Affordable and scalable vaccine production facilities are in short supply in Africa making alternative, cheaper options more viable. In this study I designed and documented a production pipeline for a potential immunogenic antigen LUJV GP-CΔTM that may be used as in a vaccine or diagnostic tool. The LUJV GP-CΔTM antigen was transfected into HEK293T cells and *N. benthamiana* in parallel to assess production bottlenecks and general cost differences between the two protein expression systems. Here I summarize the main elements of the study and any shortcomings that should be investigated in future work.

Parameters such as plant cell line, method of transfection, optimal harvesting time, method of purification and accessory co-proteins needed for producing LUJV GP-CΔTM were recorded. To produce LUJV GP-CΔTM in *N. benthamiana*, plants need to be grown for three weeks followed by recombinant *A. tumefaciens* infiltration with the desired protein. LUJV GP-CΔTM requires co-expression with CRT and lmSTT3D possibly increases glycan occupancy of LUJV GP-CΔTM. This was noted by a size shift of LUJV GP-CΔTM when co-expressed by lmSTT3D which suggests addition of glycans, however, glycosylation mapping should be attempted to more accurately assess the types and extent to which LUJV GP-CΔTM is glycosylated. This has downstream effects in terms of protein folding and generation of protein aggregates.

Neutralization assays and animal challenge with LUJV GP-CΔTM produced through this study's methods may be a prospect for future research. This study did not explore the structural integrity of expressed LUJV GP-CΔTM, which leaves questions to be answered about the conformation and multimerization of the LUJV GP-CΔTM protein. This should be investigated by subjecting the LUJV GP-CΔTM to a blue native PAGE which will give an indication of protein interactions with the LUJV GP-CΔTM. Attempts at eliciting proteolytic cleavage of LUJV GP-CΔTM were, however, abortive. The S1P could not be detected in *N. benthamiana* and failed to cleave LUJV GP-CΔTM. The study suggested that S1P was not expressing at low levels nor was the anti S1P antibody faulty since it detected HEK293T produced S1P. It is possible that S1P was produced in an abstract conformation that prevented it from being detected, however, failure to cleave LUJV GP-CΔTM suggested additional potential issues. It is not unexpected that it would be difficult to produce a protease in a non-native system as it may be highly destructive to a host environment –so will be rejected by the host quickly. Despite glycosylation attempts of LUJV GP-CΔTM to help S1P better recognize and cleave it, these were unsuccessful. It may be the case that S1P has additional folding or maturation requirements to

function such as expression of CNX which was never co-expressed with S1P. While proteolytic cleavage is likely necessary for the incorporation of LUJV GP-CΔTM into VLP's, it is possible that the LUJV GP-CΔTM can be used as a diagnostic tool or immunogenic antigen. For future plant work, it may be possible to present the Lujo spike complex on VLPs if the gene sequences are redesigned to include linker sequences separating GP1 from GP2 on the LUJV GP-CΔTM protein or cleaved through a means other than S1P. The LUJV Z protein can then be produced and expressed in conjunction with LUJV GP-CΔTM to try and form multimeric protein complexes and subsequently VLPs. The LUJV GP-CΔTM produced here needs to be assessed for structural integrity through methods such as cryo-electron microscopy.

The basic cost of recombinant protein production was documented with the mammalian expression system costing approximately R123047 and the plant expression system being significantly cheaper around R74293. While the initial gene design, synthesis and cloning steps are similar, the main difference in cost between the two expression systems lies in the running costs of plants vs mammalian cells. Plant biomass and *A. tumefaciens* cultures are cheaper whereas mammalian cultures need expensive growth medium and complex incubators. While there are many factors that contribute to the cost of research (not all of which was explored here) this gave a general indication of the affordability of plant expression systems making them appealing for use in impoverished countries where funds are limited.

For future experiments neutralization assays can be performed with infected sera to assess the immunogenicity of LUJV GP-CΔTM and animal studies may help to further strengthen the results. Collectively the study yielded 3 glyco-variations of the LUJV GP-CΔTM protein: HEK293T (2.16 mg/L cell culture), WT *N. benthamiana* (0.85 mg/kg plant biomass) and *N. Benthamiana* ΔXT/FT (3.1 mg/ kg plant biomass) LUJV GP-CΔTM forms. These proteins serve as a starting point that needs to be investigated for immunogenicity as a vaccine component or diagnostic tool for Lujo virus. The data generated here gives crucial insight into overcoming bottlenecks for producing Lujo viral proteins in *N. benthamiana* and HEK293T cells making the realization of a Lujo virus vaccine a much closer reality. Future studies can build on the data generated here and apply these to other arenaviruses which may lead to the development of a broad-spectrum diagnostic tool that will be of extreme benefit in the face of emerging arenaviruses.

Appendix: Reagent compositions

Elution buffer (for protein purification):

10.21g Imidazole

1.461g NaCl (sodium chloride)

25mL Tris-HCL (Trizma base hydrochloric acid)

0.018g EDTA (Ethylenediaminetetracetic acid)

Add water up to 500ml pH 8

Running buffer (For protein analysis) (1X):

6.04g Trizma Base

28.8g Glycine

Distilled water to 2L

2g Sodium dodecyl sulfate (SDS)

Transfer buffer (For protein analysis) (1X):

6.04g Trizma Base

28.8g Glycine

200 mL Methanol

Distilled water to 1L

Blocking buffer: (For protein analysis)

10 mL 10X PBS

5g Skim milk powder

1 mL 10% Tween 20

Distilled water to 100 mL

Wash buffer (For protein analysis):

10 mL 10X PBS

1 mL 10% Tween 20

Distilled water to 100 mL

Protein loading dye (4X):

4 mL 100% Glycerol

1.6 mL 1.5 M Tris/HCL pH 6.8

3.9 mL distilled water

0.8g Sodium Dodecyl sulfate

4 mg Bromophenol blue

0.5 mL β- mercaptoethanol

Supplemented DMEM (For HEK293T cell growth):

450 mL Dulbecco's Modified Eagle Medium (DMEM)

5 mL Penicillin/streptomycin (10 mg/mL)

50 mL Fetal Bovine serum (FBS)

LB Broth Base (*A. tumefaciens* growth medium*)*:

2.5 g Tryptone

12.5 g Yeast extract

5g NaCl

10 mL 1M 2-(N morpholino) ethanesulfonic acid MES

pH to 5.6

Induction Medium (1X, for *A. tumefaciens* mediated transformation):

10mM 2-(N morpholino) ethanesulfonic acid MES

20μM Acetosyringone

pH to 5.6

Add LB up to 1L

Resuspension medium (1X, For *A. tumefaciens* mediated transformation):

10mM 2-(N morpholino) ethanesulfonic acid MES

pH to 5.6

200μM Acetosyringone

10mM $MgCl_2$

Distilled water to 1L

References

1. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451:990-3; doi: 10.1038/nature06536.

2. Kamorudeen RT, Adedokun KA, Olarinmoye AO. Ebola outbreak in West Africa, 2014–2016: Epidemic timeline, differential diagnoses, determining factors, and lessons for future response. Journal of Infection and Public Health. 2020;13:956-62.

3. Akpede GO, Asogun DA, Okogbenin SA, Okokhere PO. Lassa fever outbreaks in Nigeria. Expert Review of Anti-infective Therapy. 2018;16:663-6; doi: 10.1080/14787210.2018.1512856.

4. Musa SS, Zhao S, Abdullahi ZU, Habib AG, He D. COVID-19 and Lassa fever in Nigeria: A deadly alliance? International Journal of Infectious Diseases. 2022;117:45-7.

5. Eisenstein M. Disease: Poverty and pathogens. Nature. 2016;531:S61-S3; doi: 10.1038/531s61a.

6. Kapiriri L, Ross A. The Politics of Disease Epidemics: a Comparative Analysis of the SARS, Zika, and Ebola Outbreaks. Global Social Welfare. 2020;7:33-45; doi: 10.1007/s40609-018-0123-y.

7. Deb P, Furceri D, Ostry JD, Tawk N. The economic effects of Covid-19 containment measures. Open Economies Review. 2022;33:1-32.

8. Burki T. Ebola virus vaccine receives prequalification. The Lancet. 2019;394(10212):1893; doi: 10.1016/s0140-6736)32905-8.

9. McKibbin, W. and Fernando, R., 2021. The global macroeconomic impacts of COVID-19: Seven scenarios. *Asian Economic Papers*, *20*, pp.1-30.

10. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International Journal of Surgery. 2020;76:71-6; doi: 10.1016/j.ijsu.2020.02.034.

11. Sharp PM. Origins of Human Virus Diversity. 2002;108(3):305-12; doi: 10.1016/s0092-8674(02)00639-6.

12. Afrough B, Dowall S, Hewson R. Emerging viruses and current strategies for vaccine intervention. Clinical & Experimental Immunology. 2019;196(2):157-66; doi: 10.1111/cei.13295.

13. Moolla N, Weyer J. The Arenaviridae. In: Emerging and Reemerging Viral Pathogens. Elsevier; Casablanca, Morroco.2020. p. 69-100.

14. Sewlall N, Paweska J. Lujo virus: current concepts. Virus Adaptation and Treatment. 2017;Volume 9:41-7; doi: 10.2147/vaat.S113593.

15. Moraz ML, Kunz S. Pathogenesis of arenavirus hemorrhagic fevers. Expert Review of Anti Infective Therapy. 2011;9:49-59; doi: 10.1586/eri.10.142.

16. Barrera Oro JG, McKee KT. Toward a vaccine against Argentine hemorrhagic fever. Bulletin of the Pan American Health Organization (PAHO); 25 , 1991. 1991.

17. McCormink J, King Y, Webb P. Lassa fever: effective therapy with rivabirin. New Englland Journal of Medicine. 1986;314:2-26.

18. Pontremoli C, Forni D, Sironi M. Arenavirus genomics: novel insights into viral diversity, origin, and evolution. Current Opinion in Virology. 2019;34:18-28; doi: 10.1016/j.coviro.2018.11.001.

19. McLay L, Liang Y, Ly H. Comparative analysis of disease pathogenesis and molecular mechanisms of New World and Old World arenavirus infections. Journal of General Virology. 2014;95(Pt 1):1-15; doi: 10.1099/vir.0.057000-0.

20. Ishii A, Thomas Y, Moonga L, Nakamura I, Ohnuma A, Hang'ombe B, et al. Novel arenavirus, zambia. Emerging Infectious Diseases. 2011;17:1921.

21. Hallam SJ, Koma T, Maruyama J, Paessler S. Review of Mammarenavirus Biology and Replication. Frontiers in Microbiology. 2018;9:1751; doi: 10.3389/fmicb.2018.01751.

22. Matthijs Raaben, Lucas T. Jae, Andrew S. Herbert, John M. Dye, Thijn R. Brummelkamp, Whelan SP. NRP2 and CD63 Are Host Factors for Lujo Virus Cell Entry. Cell Host and Microbe. 2017;22:688-96; doi: 10.1016/j.chom.2017.10.002.

23. Ferron F, Weber F, De La Torre JC, Reguera J. Transcription and replication mechanisms of Bunyaviridae and Arenaviridae L proteins. Virus Research. 2017;234:118-34; doi: 10.1016/j.virusres.2017.01.018.

24. Strecker T, Eichler R, Meulen J, Weissenhorn W, Dieter Klenk H, Garten W, et al. Lassa virus Z protein is a matrix protein and sufficient for the release of virus-like particles [corrected]. J Virol. 2003;77(19):10700-5; doi: 10.1128/jvi.77.19.10700-10705.2003.

25. Lennartz F, Hoenen T, Lehmann M, Groseth A, Garten W. The role of oligomerization for the biological functions of the arenavirus nucleoprotein. Arch Virol. 2013;158(9):1895-905; doi: 10.1007/s00705-013-1684-9.

26. Beyer WR, Popplau D, Garten W, von Laer D, Lenz O. Endoproteolytic processing of the lymphocytic choriomeningitis virus glycoprotein by the subtilase SKI-1/S1P. Journal of Virology. 2003;77:2866-72; doi: 10.1128/jvi.77.5.2866-2872.2003.

27. Pinschewer DD, Perez M, de la Torre JC. Role of the virus nucleoprotein in the regulation of lymphocytic choriomeningitis virus transcription and RNA replication. Journal of Virology. 2003;77:3882-7; doi: 10.1128/jvi.77.6.3882-3887.2003.

28. Li S, Sun Z, Pryce R, Parsy M-L, Fehling SK, Schlie K, et al. Acidic pH-induced conformations and LAMP1 binding of the Lassa virus glycoprotein spike. PLoS Pathogens. 2016;12:e1005418.

29. Hastie KM, Zandonatti MA, Kleinfelter LM, Heinrich ML, Rowland MM, Chandran K, et al. Structural basis for antibody-mediated neutralization of Lassa virus. Science. 2017;356:923-8.

30. York J, Nunberg JH. Myristoylation of the arenavirus envelope glycoprotein stable signal peptide is critical for membrane fusion but dispensable for virion morphogenesis. Journal of Virology. 2016;90:8341-50.

31. Martinez-Sobrido L, Emonet S, Giannakas P, Cubitt B, Garcia-Sastre A, de la Torre JC. Identification of amino acid residues critical for the anti-interferon activity of the nucleoprotein of the prototypic arenavirus lymphocytic choriomeningitis virus. Journal of Virology. 2009;83:11330-40; doi: 10.1128/JVI.00763-09.

32. Qi X, Lan S, Wang W, Schelde LM, Dong H, Wallat GD, et al. Cap binding and immune evasion revealed by Lassa nucleoprotein structure. Nature. 2010;468:779-83; doi: 10.1038/nature09605.

33. Rojek JM, Lee AM, Nguyen N, Spiropoulou CF, Kunz S. Site 1 protease is required for proteolytic processing of the glycoproteins of the South American hemorrhagic fever viruses Junin, Machupo, and Guanarito. Journal of Virology. 2008;82:6045-51; doi: 10.1128/JVI.02392-07.

34. Peng R, Xu X, Jing J, Wang M, Peng Q, Liu S, et al. Structural insight into arenavirus replication machinery. Nature. 2020;579:615-9; doi: 10.1038/s41586-020-2114-2.

35. Eschli B, Quirin K, Wepf A, Weber J, Zinkernagel R, Hengartner H. Identification of an N-terminal trimeric coiled-coil core within arenavirus glycoprotein 2 permits assignment to class I viral fusion proteins. Journal of Virology. 2006;80:5897-907; doi: 10.1128/JVI.00008-06.

36. Fehling SK, Lennartz F, Strecker T. Multifunctional nature of the arenavirus RING finger protein Z. Viruses. 2012;4:2973-3011; doi: 10.3390/v4112973.

37. Eichler R, Lenz O, Strecker T, Garten W. Signal peptide of Lassa virus glycoprotein GP-C exhibits an unusual length. FEBS Letters. 2003;538:203-6.

38. Burri DJ, Pasqual G, Rochat C, Seidah NG, Pasquato A, Kunz S. Molecular characterization of the processing of arenavirus envelope glycoprotein precursors by subtilisin kexin isozyme-1/site-1 protease. Journal of Virology. 2012;86:4935-46; doi: 10.1128/JVI.00024-12.

39. Wolff S, Ebihara H, Groseth A. Arenavirus budding: a common pathway with mechanistic differences. Viruses. 2013;5:528-49; doi: 10.3390/v5020528.

40. Weissenbacher MC, Sabattini MS, Avila MM, Sangiorgio PM, De Sensi MR, Contigiani MS, et al. Junin virus activity in two rural populations of the Argentine hemorrhagic fever (AHF) endemic area. Journal of Medical Virology. 1983;12:273-80.

41. Charrel RN, de Lamballerie X. Zoonotic aspects of arenavirus infections. Vetirinary Microbiology. 2010;140:213-20; doi: 10.1016/j.vetmic.2009.08.027.

42. Sarute N, Ross SR. New World Arenavirus Biology. Annual Review of Virology. 2017;4:141-58; doi: 10.1146/annurev-virology-101416-042001.

43. Moraz MLK, S., . Pathogenesis of arenavirus hemorrhagic fevers. Expert Review of Anti-Infective Therapy. 2011;9:49-59.

44. Sewlall NH, Richards G, Duse A, Swanepoel R, Paweska J, Blumberg L, et al. Clinical Features and Patient Management of Lujo Hemorrhagic Fever. PLoS Neglected Tropical Diseases. 2014;8:e3233; doi: 10.1371/journal.pntd.0003233.

45. Meyer B, Groseth A. Apoptosis during arenavirus infection: mechanisms and evasion strategies. Microbes and Infection. 2018;20:65-80; doi: 10.1016/j.micinf.2017.10.002.

46. Fukushi S, Tani H, Yoshikawa T, Saijo M, Morikawa S. Serological assays based on recombinant viral proteins for the diagnosis of arenavirus hemorrhagic fevers. Viruses. 2012;4:2097-114; doi: 10.3390/v4102097.

47. Rodrigo WW, Ortiz-Riano E, Pythoud C, Kunz S, de la Torre JC, Martinez-Sobrido L. Arenavirus nucleoproteins prevent activation of nuclear factor kappa B. J Virology. 2012;86:8185-97; doi: 10.1128/JVI.07240-11.

48. Xing J, Ly H, Liang Y. The Z proteins of pathogenic but not nonpathogenic arenaviruses inhibit RIG-I-like receptor-dependent interferon production. Journal of Virology. 2015;89:2944-55; doi: 10.1128/JVI.03349-14.

49. Carnec X, Baize S, Reynard S, Diancourt L, Caro V, Tordo N, et al. Lassa virus nucleoprotein mutants generated by reverse genetics induce a robust type I interferon response in human dendritic cells and macrophages. Journal of Virology. 2011;85:12093-7; doi: 10.1128/JVI.00429-11.

50. Meyer B, Ly H. Inhibition of Innate Immune Responses Is Key to Pathogenesis by Arenaviruses. Journal of Virology. 2016;90:3810-8; doi: 10.1128/JVI.03049-15.

51. Montoya M. Type I interferons produced by dendritic cells promote their phenotypic and functional activation. Journal of the American Society of Hematology. 2002;99:3263-71; doi: 10.1182/blood.v99.9.3263.

52. Smither AR, Bell-Kareem AR. Ecology of Lassa Virus. Current Topics in Microbiology and Immunology. 2021; doi: 10.1007/82_2020_231.

53. Gryseels S, Rieger T, Oestereich L, Cuypers B, Borremans B, Makundi R, et al. Gairo virus, a novel arenavirus of the widespread Mastomys natalensis: Genetically divergent, but ecologically similar to Lassa and Morogoro viruses. Virology. 2015;476:249-56.

54. Roberts L: Nigeria hit by unprecedented Lassa fever outbreak. .American Association for the Advancement of Science. University of Cape Town. 2018.

55. Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, et al. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. PLoS Pathogens. 2009;5:e1000455; doi: 10.1371/journal.ppat.1000455.

56. Gómez RM, de Giusti CJ, Vallduvi MMS, Frik J, Ferrer MF, Schattner M. Junín virus. A XXI century update. Microbes and Infection. 2011;13:303-11.

57. De Clercq E, Li G. Approved Antiviral Drugs over the Past 50 Years. Clinical Microbiology Review. 2016;29(3):695-747; doi: 10.1128/CMR.00102-15.

58. De Vries RD, Rimmelzwaan GF. Viral vector-based influenza vaccines. Human Vaccines & Immunotherapeutics. 2016;12(11):2881-901; doi: 10.1080/21645515.2016.1210729.

59. Bobrowski T, Melo-Filho CC, Korn D, Alves VM, Popov KI, Auerbach S, et al. Learning from history: do not flatten the curve of antiviral research! Drug Discovery Today. 2020;25:1604-13; doi: 10.1016/j.drudis.2020.07.008.

60. Oñate AA, Wan Y, Moreno A. Editorial: Molecular Vaccines Against Pathogens in the Post-genomic Era. Frontiers in Cellular Infection Microbiology. 2020;9; doi: 10.3389/fcimb.2019.00443/full.

61. Haber P, Moro PL, Ng C, Lewis PW, Hibbs B, Schillie SF, et al. Safety of currently licensed hepatitis B surface antigen vaccines in the United States, Vaccine Adverse Event Reporting System (VAERS), 2005–2015. Vaccine. 2018;36:559-64; doi: 10.1016/j.vaccine.2017.11.079.

62. Pinto LA, Dillner J, Beddows S, Unger ER. Immunogenicity of HPV prophylactic vaccines: Serology assays and their use in HPV vaccine evaluation and development. Vaccine. 2018;36:4792-9; doi: 10.1016/j.vaccine.2017.11.089.

63. Otu A, Osifo-Dawodu E, Atuhebwe P, Agogo E, Ebenso B. Beyond vaccine hesitancy: time for Africa to expand vaccine manufacturing capacity amidst growing COVID-19 vaccine nationalism. The Lancet Microbe. 2021;2:e347-e8.

64. Expiring vaccines doomed South Africa's rollout from the start: report. In. businesstech.co.za: businesstech; 2021.

65. Minor PD. Live attenuated vaccines: Historical successes and current challenges. Virology. 2015;479-480:379-92; doi: 10.1016/j.virol.2015.03.032.

66. York J, Nunberg JH. Epistastic Interactions within the Junín Virus Envelope Glycoprotein Complex Provide an Evolutionary Barrier to Reversion in the Live-Attenuated Candid#1 Vaccine. Journal of Virology. 2017;92; doi: 10.1128/jvi.01682-17.

67. McKee Jr KT, Oro JGB, Kuehne AI, Spisso JA, Mahlandt B. Candid No. 1 Argentine hemorrhagic fever vaccine protects against lethal Junin virus challenge in rhesus macaques. Intervirology. 1992;34:154-63.

68. Hallam SJ, Manning JT, Maruyama J, Seregin A, Huang C, Walker DH, et al. A single mutation (V64G) within the RING Domain of Z attenuates Junin virus. PLoS Neglected Tropical Diseases. 2020;14:e0008555; doi: 10.1371/journal.pntd.0008555.

69. Cai Y, Ye C, Cheng B, Nogales A, Iwasaki M, Yu S, et al. A Lassa fever live-attenuated vaccine based on codon deoptimization of the viral glycoprotein Gene. mBio. 2020;11; doi: 10.1128/mBio.00039-20.

70. Lauer KB, Borrow R, Blanchard TJ. Multivalent and multipathogen viral vector vaccines. Clinical Vaccine Immunology. 2017;24; doi: 10.1128/CVI.00298-16.

71. Ewer KJ, Lambe T, Rollier CS, Spencer AJ, Hill AV, Dorrell L. Viral vectors as vaccine platforms: from immunogenicity to impact. Current Opinions in Immunology. 2016;41:47-54; doi: 10.1016/j.coi.2016.05.014.

72. Carnec X, Mateo M, Page A, Reynard S, Hortion J, Picard C, et al. A vaccine platform against arenaviruses based on a recombinant hyperattenuated Mopeia virus expressing heterologous glycoproteins. Journal of Virology. 2018;92; doi: 10.1128/JVI.02230-17.

73. Lua LHL, Connors NK, Sainsbury F, Chuan YP, Wibowo N, Middelberg APJ. Bioengineering virus-like particles as vaccines. Biotechnology and Bioengineering. 2014;111:425-40; doi: 10.1002/bit.25159.

74. Mohsen MO, Zha L, Cabral-Miranda G, Bachmann MF. Major findings and recent advances in virus-like particle (VLP)-based vaccines. Seminars in Immunology. 2017;34:123-32; doi: 10.1016/j.smim.2017.08.014.

75. Plummer EM, Manchester M. Viral nanoparticles and virus-like particles: platforms for contemporary vaccine design. Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology. 2011;3:174-96; doi: 10.1002/wnan.119.

76. Harper DM, Demars LR. HPV vaccines – A review of the first decade. Gynecologic Oncology. 2017;146:196-204; doi: 10.1016/j.ygyno.2017.04.004.

77. Pushko P, Pearce MB, Ahmad A, Tretyakova I, Smith G, Belser JA, et al. Influenza virus-like particle can accommodate multiple subtypes of hemagglutinin and protect from multiple influenza types and subtypes. Vaccine. 2011;29:5911-8; doi: 10.1016/j.vaccine.2011.06.068.

78. Hajj Hussein I, Chams N, Chams S, El Sayegh S, Badran R, Raad M, et al. Vaccines through centuries: major cornerstones of global health. Frontiers in Public Health. 2015;3; doi: 10.3389/fpubh.2015.00269.

79. Zuckerman J. Evaluation of a new hepatitis B triple-antigen vaccine in inadequate responders to current vaccines.Hepatology, 2001;34:798-802; doi: 10.1053/jhep.2001.27564.

80. Salvato MS, Domi A, Guzmán-Cardozo C, Medina-Moreno S, Zapata JC, Hsu H, et al. A single dose of modified vaccinia Ankara expressing Lassa virus-like particles protects mice from lethal intra-cerebral virus challenge. Pathogens. 2019;8:133.

81. Branco LM, Grove JN, Geske FJ, Boisen ML, Muncy IJ, Magliato SA, et al. Lassa virus-like particles displaying all major immunological determinants as a vaccine candidate for Lassa hemorrhagic fever. Virology Journal. 2010;7:1-19.

82. Muller H, Fehling SK, Dorna J, Urbanowicz RA, Oestereich L, Krebs Y, et al. Adjuvant formulated virus-like particles expressing native-like forms of the Lassa virus envelope surface glycoprotein are immunogenic and induce antibodies with broadly neutralizing activity. NPJ Vaccines. 2020;5:71; doi: 10.1038/s41541-020-00219-x.

83. Pereira VB, Zurita-Turk M, Saraiva TDL, De Castro CP, Souza BM, Mancha Agresti P, et al. DNA vaccines approach: from concepts to applications. 2014;04(02):50-71; doi: 10.4236/wjv.2014.42008.

84. Gary EN, Weiner DB. DNA vaccines: prime time is now. Current Opinion in Immunology. 2020;65:21-7; doi: 10.1016/j.coi.2020.01.006.

85. Li S, Maclaughin FC, Fewell JG, Gondo M, Wang J, Nicol F, et al. Muscle-specific enhancement ofgene expression by incorporation ofSV40 enhancer in the expression plasmid. Nature. 2001;8:494-7.

86. Alonso M, Leong JA. Licensed DNA vaccines against infectious hematopoietic necrosis virus (IHNV). Recent Patents on DNA and Gene Sequences. 2013;7:62-5; doi: 10.2174/1872215611307010009.

87. Ulmer JB, Wahren B, Liu MA. Gene-based vaccines: recent technical and clinical advances. Trends in Molecular Medicine. 2006;12:216-22; doi: 10.1016/j.molmed.2006.03.007.

88. Rodriguez-Carreno MP, Nelson MS, Botten J, Smith-Nixon K, Buchmeier MJ, Whitton JL. Evaluating the immunogenicity and protective efficacy of a DNA vaccine encoding Lassa virus nucleoprotein.Virology. 2005;335:87-98; doi: 10.1016/j.virol.2005.01.019.

89. Moyle PM, Toth I. Modern subunit vaccines: development, components, and research opportunities. ChemMedChem. 2013;8:360-76; doi: 10.1002/cmdc.201200487.

90. Pulendran B, Ahmed R. Immunological mechanisms of vaccination. Nature Immunology. 2011;12:509-17; doi: 10.1038/ni.2039.

91. Rybicki EP. Plant-produced vaccines: promise and reality. Drug Discovery Today. 2009;14:16-24; doi: 10.1016/j.drudis.2008.10.002.

92. Ma JKC, Christou P, Chikwamba R, Haydon H, Paul M, Ferrer MP, et al. Realising the value of plant molecular pharming to benefit the poor in developing countries and emerging economies. Plant Biotechnology Journal. 2013;11:1029-33; doi: 10.1111/pbi.12127.

93. Gengenbach BB, Opdensteinen P, Buyel JF. Robot Cookies - Plant cell packs as an automated high-throughput screening platform based on transient expression. Frontiers in Bioengineering and Biotechnology. 2020;8:393; doi: 10.3389/fbioe.2020.00393.

94. Chung SH, Bigham M, Lappe RR, Chan B, Nagalakshmi U, Whitham SA, et al. A sugarcane mosaic virus vector for rapid in planta screening of proteins that inhibit the growth of insect herbivores. Plant Biotechnology Journal. 2021;19:1713-24; doi: 10.1111/pbi.13585.

95. Venkataraman S, Hefferon K, Makhzoum A, Abouhaidar M. Combating human viral diseases: will plant-based vaccines be the answer? Vaccines (Basel). 2021;9; doi: 10.3390/vaccines9070761.

96. Bally J, Nakasugi K, Jia F, Jung H, Ho SY, Wong M, et al. The extremophile *Nicotiana benthamiana* has traded viral defence for early vigour. Nature Plants. 2015;1:15165; doi: 10.1038/nplants.2015.165.

97. Montero-Morales L, Steinkellner H. Advanced plant-based glycan engineering. Frontiers in Bioengeering and Biotechnology. 2018;6:81; doi: 10.3389/fbioe.2018.00081.

98. Bally J, Jung H, Mortimer C, Naim F, Philips JG, Hellens R, et al. The rise and rise of *Nicotiana benthamiana*: A plant for all reasons. Annual Review of Phytopathology. 2018;56:405-26; doi: 10.1146/annurev-phyto-080417-050141.

99. Pillet S, Couillard J, Trépanier S, Poulin J-F, Yassine-Diab B, Guy B, et al. Immunogenicity and safety of a quadrivalent plant-derived virus like particle influenza vaccine candidate—Two randomized Phase II clinical trials in 18 to 49 and ≥50 years old adults. PLOS ONE. 2019;14:e0216533; doi: 10.1371/journal.pone.0216533.

100. Richard T. Davey J et. al. A randomized, controlled trial of ZMapp for Ebola virus infection. New England Journal of Medicine. 2016;375:1448-56; doi: 10.1056/nejmoa1604330.

101. Ward BJ, Seguin A, Couillard J, Trepanier S, Landry N. Phase III: Randomized observer-blind trial to evaluate lot-to-lot consistency of a new plant-derived quadrivalent virus like particle

influenza vaccine in adults 18-49 years of age. Vaccine. 2021;39:1528-33; doi: 10.1016/j.vaccine.2021.01.004.

102. Royal JM, Simpson CA, McCormick AA, Phillips A, Hume S, Morton J, et al. Development of a SARS-CoV-2 Vaccine Candidate Using Plant-Based Manufacturing and a Tobacco Mosaic Virus-like Nano-Particle. Vaccines (Basel). 2021;9(11); doi: 10.3390/vaccines9111347.

103. Duong D, Vogel L: Why is WHO pushing back on a Health Canada–approved Medicago SARS-CoV-2 vaccine? In.: Canada Medical Association; 2022. CMAJ 2022 April 4;194:E504-5. doi: 10.1503/cmaj.1095992

104. Emmanuel Aubrey Margolin et al. Engineering the Plant Secretory Pathway for the Production of Next-Generation Pharmaceuticals. Trends in Biotechnology 2020.38, 1034-1044

105. Zhang Y, Li D, Jin X, Huang Z. Fighting Ebola with ZMapp: spotlight on plant-made antibody. Science China Life Sciences. 2014;57:987-8; doi: 10.1007/s11427-014-4746-7.

106. Kumar R, Mehta D, Mishra N, Nayak D, Sunil S. Role of host-mediated post-translational modifications (PTMs) in RNA virus pathogenesis. International Journal of Molecular Science. 2020;22; doi: 10.3390/ijms22010323.

107. Mateo M, Reynard S, Carnec X, Journeaux A, Baillet N, Schaeffer J, et al. Vaccines inducing immunity to Lassa virus glycoprotein and nucleoprotein protect macaques after a single shot. Science Translational Medicine. 2019;11.

108. Margolin E, Crispin M, Meyers A, Chapman R, Rybicki EP. A roadmap for the molecular farming of viral glycoprotein vaccines: Engineering Glycosylation and Glycosylation-Directed Folding. Frontiers in Plant Science. 2020;11:609207; doi: 10.3389/fpls.2020.609207.

109. Olszewski NE, West CM, Sassi SO, Hartweck LM. O-GlcNAc protein modification in plants: Evolution and function. Biochimica et Biophysica Acta (BBA) - General Subjects. 2010;1800:49-56; doi: 10.1016/j.bbagen.2009.11.016.

110. Hebert DN, Lamriben L, Powers ET, Kelly JW. The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. Nature Chemical Biology. 2014;10:902-10; doi: 10.1038/nchembio.1651.

111. Kallolimath S, Castilho A, Strasser R, Grünwald-Gruber C, Altmann F, Strubl S, et al. Engineering of complex protein sialylation in plants. Proceedings of the National Academy of Sciences. 2016;113:9498-503; doi: 10.1073/pnas.1604371113.

112. Feng T, Zhang J, Chen Z, Pan W, Chen Z, Yan Y, et al. Glycosylation of viral proteins: Implication in virus-host interaction and virulence. Virulence. 2022;13:670-83; doi: 10.1080/21505594.2022.2060464.

113. Koma T, Huang C, Coscia A, Hallam S, Manning JT, Maruyama J, et al. Glycoprotein N-linked glycans play a critical role in arenavirus pathogenicity. PLoS Pathog. 2021;17:e1009356; doi: 10.1371/journal.ppat.1009356.

114. Bowie AG, Unterholzner L. Viral evasion and subversion of pattern-recognition receptor signalling. Nature Reviews Immunology. 2008;8:911-22; doi: 10.1038/nri2436.

115. Strasser R. Plant protein glycosylation. Glycobiology. 2016;26:926-39; doi: 10.1093/glycob/cww023.

116. Castilho A, Beihammer G, Pfeiffer C, Göritzer K, Montero-Morales L, Vavra U, et al. An oligosaccharyltransferase from *Leishmania major* increases the N-glycan occupancy on recombinant glycoproteins produced in *Nicotiana benthamiana*. Plant Biotechnology Journal. 2018;16:1700-9; doi: 10.1111/pbi.12906.

117. Jeong IS, Lee S, Bonkhofer F, Tolley J, Fukudome A, Nagashima Y, et al. Purification and characterization of *Arabidopsis thaliana* oligosaccharyltransferase complexes from the native host: a protein super-expression system for structural studies. The Plant Journal. 2018;94:131-45; doi: 10.1111/tpj.13847.

118. Goritzer K, Goet I, Duric S, Maresch D, Altmann F, Obinger C, et al. Efficient N-glycosylation of the heavy chain tailpiece promotes the formation of plant-produced dimeric IgA. Frontiers in Chemistry. 2020;8:346; doi: 10.3389/fchem.2020.00346.

119. Wilbers RHP, Westerhof LB, Van Raaij DR, Van Adrichem M, Prakasa AD, Lozano-Torres JL, et al. Co-expression of the protease furin in *Nicotiana benthamiana* leads to efficient processing of latent transforming growth factor-β1 into a biologically active protein. Plant Biotechnology Journal.2016;14 pp. 1695-1704 doi: 10.1111/pbi.12530.

120. Thomas EL, van der Hoorn RAL. Ten prominent host proteases in plant-pathogen interactions. International Journal of Molecular Science. 2018;19; doi: 10.3390/ijms19020639.

121. Mamedov T, Musayeva I, Acsora R, Gun N, Gulec B, Mammadova G, et al. Engineering, and production of functionally active human Furin in *N. benthamiana plant*: In vivo post-translational processing of target proteins by Furin in plants. PLoS One. 2019;14:e0213438; doi: 10.1371/journal.pone.0213438.

122. Margolin E, Oh YJ, Verbeek M, Naude J, Ponndorf D, Meshcheriakova YA, et al. Co-expression of human calreticulin significantly improves the production of HIV gp140 and other viral glycoproteins in plants. Plant Biotechnology Journal. 2020; 18, pp 2109-2117doi: 10.1111/pbi.13369.

123. Mandal MK, Ahvari H, Schillberg S, Schiermeyer A. Tackling unwanted proteolysis in plant production hosts used for molecular farming. Frontiers in Plant Science. 2016;7; doi: 10.3389/fpls.2016.00267.

124. Vembar SS, Brodsky JL. One step at a time: endoplasmic reticulum-associated degradation. Nature Reviews Molecular Cell Biology. 2008;9:944-57; doi: 10.1038/nrm2546.

125. Donini M, Lombardi R, Lonoce C, Di Carli M, Marusic C, Morea V, et al. Antibody proteolysis: a common picture emerging from plants. Bioengineered. 2015;6:299-302; doi: 10.1080/21655979.2015.1067740.

126. Grosse-Holz F, Madeira L, Zahid MA, Songer M, Kourelis J, Fesenko M, et al. Three unrelated protease inhibitors enhance accumulation of pharmaceutical recombinant proteins in Nicotiana benthamiana. Plant Biotechnology Journal. 2018;16:1797-810; doi: 10.1111/pbi.12916.

127. Kim NS, Kim TG, Kim OH, Ko EM, Jang YS, Jung ES, et al. Improvement of recombinant hGM-CSF production by suppression of cysteine proteinase gene expression using RNA interference in a transgenic rice culture. Plant Molecular Biology. 2008;68:263-75; doi: 10.1007/s11103-008-9367-8.

128. Duwadi K, Chen L, Menassa R, Dhaubhadel S. Identification, characterization and down-regulation of cysteine protease genes in tobacco for use in recombinant protein production. PLoS One. 2015;10:e0130556; doi: 10.1371/journal.pone.0130556.

129. Strasser R. Protein quality control in the endoplasmic reticulum of plants. Annual Review of Plant Biology. 2018;69:147-72; doi: 10.1146/annurev-arplant-042817-040331.

130. Jutras PV, Dodds I, Van Der Hoorn RA. Proteases of *Nicotiana benthamiana*: an emerging battle for molecular farming. Current Opinion in Biotechnology. 2020;61:60-5; doi: 10.1016/j.copbio.2019.10.006.

131. Wang W, Vinocur B, Shoseyov O, Altman A. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends in Plant Science. 2004;9:244-52.

132. Molinari M, Eriksson KK, Calanca V, Galli C, Cresswell P, Michalak M, et al. Contrasting Functions of Calreticulin and Calnexin in Glycoprotein Folding and ER Quality Control. Molecular Cell. 2004;13(1):125-35; doi: 10.1016/s1097-2765(03)00494-5.

133. Tannous A, Pisoni GB, Hebert DN, Molinari M. N-linked sugar-regulated protein folding and quality control in the ER. Seminars in Cell & Developmental Biology. 2015;41:79-89; doi: 10.1016/j.semcdb.2014.12.001.

134. Caramelo JJ, Parodi AJ. Getting in and out from calnexin/calreticulin cycles. J Biological Chemistry. 2008;283:10221-5; doi: 10.1074/jbc.R700048200.

135. Hamorsky KT, Kouokam JC, Jurkiewicz JM, Nelson B, Moore LJ, Husk AS, et al. N-Glycosylation of cholera toxin B subunit in Nicotiana benthamiana: impacts on host stress response, production yield and vaccine potential. 2015;5:8003; doi: 10.1038/srep08003.

136. Margolin E, Chapman R, Meyers AE, van Diepen MT, Ximba P, Hermanus T, et al. Production and immunogenicity of soluble plant-produced HIV-1 subtype C envelope gp140 immunogens. Frontiers in Plant Science. 2019;10:1378; doi: 10.3389/fpls.2019.01378.

137. Walsh G, Jefferis R. Post-translational modifications in the context of therapeutic proteins. Nature Biotechnology. 2006;24:1241-52; doi: 10.1038/nbt1252.

138. Alvisi N, van Noort K, Dwiani S, Geschiere N, Sukarta O, Varossieau K, et al. beta-Hexosaminidases along the secretory pathway of *Nicotiana benthamiana* have distinct

specificities toward engineered Helminth N-glycans on recombinant glycoproteins. Frontiers in Plant Science. 2021;12:638454; doi: 10.3389/fpls.2021.638454.

139. Paweska JT, Sewlall NH, Ksiazek TG, Blumberg LH, Hale MJ, Lipkin WI, et al. Nosocomial outbreak of novel arenavirus infection, Southern Africa. Emerging Infectious Diseases. 2009;15:1598-602; doi: 10.3201/eid1510.090211.

140. Cohen-Dvashi H, Kilimnik I, Diskin R. Structural basis for receptor recognition by Lujo virus. Nature Microbiology. 2018;3:1153-60; doi: 10.1038/s41564-018-0224-5.

141. Nandi S, Kwong AT, Holtz BR, Erwin RL, Marcel S, McDonald KA. Techno-economic analysis of a transient plant-based platform for monoclonal antibody production. 2016;8:1456-66; doi: 10.1080/19420862.2016.1227901.

142. Dyson MR. Fundamentals of expression in mammalian cells. Advanced Technologies for Protein Complex Production and Characterization. 2016:217-24.

143. Khan KH. Gene expression in mammalian cells and its applications. Advanced pharmaceutical bulletin. 2013;3:257.

144. Tanzer FL, Shephard EG, Palmer KE, Burger M, Williamson A-L, Rybicki EP. The porcine circovirus type 1 capsid gene promoter improves antigen expression and immunogenicity in a HIV-1 plasmid vaccine. Virology Journal. 2011;8:1-10.

145. Chapman R, Rybicki EP. Use of a novel enhanced DNA vaccine vector for preclinical virus vaccine investigation. Vaccines. 2019;7:50.

146. Chapman R, Jongwe TI, Douglass N, Chege G, Williamson A-L. Heterologous prime-boost vaccination with DNA and MVA vaccines, expressing HIV-1 subtype C mosaic Gag virus-like particles, is highly immunogenic in mice. PloS One. 2017;12:e0173352.

147. van Diepen M, Chapman R, Douglass N, Whittle L, Chineka N, Galant S, et al. Advancements in the growth and construction of recombinant lumpy skin disease virus (LSDV) for use as a vaccine vector. Vaccines (Basel). 2021;9; doi: 10.3390/vaccines9101131.

148. Ximba P, Chapman R, Meyers AE, Margolin E, van Diepen MT, Williamson AL, et al. Characterization and immunogenicity of HIV envelope gp140 Zera((R)) tagged antigens. Frontiers in Bioengineering and Biotechnology. 2020;8:321; doi: 10.3389/fbioe.2020.00321.

149. Whittle L, Chapman R, van Diepen M, Rybicki EP, Williamson AL. Characterization of a novel chimeric theileria parva p67 antigen which incorporates into virus-like particles and Is highly immunogenic in Mice. Vaccines (Basel). 2022;10; doi: 10.3390/vaccines10020210.

150. Seidah NG, Mowla SJ, Hamelin J, Mamarbachi AM, Benjannet S, Touré BB, et al. Mammalian subtilisin/kexin isozyme SKI-1: a widely expressed proprotein convertase with a unique cleavage specificity and cellular localization. Proceedings of the National Academy of Sciences. 1999;96:1321-6.

151. Peyret H, Lomonossoff GP. The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. Plant Molecular Biology. 2013;83:51-8; doi: 10.1007/s11103-013-0036-1.

152. Norkunas K, Harding R, Dale J, Dugdale B. Improving agroinfiltration-based transient gene expression in *Nicotiana benthamiana*. Plant Methods. 2018;14:71; doi: 10.1186/s13007-018-0343-2.

153. Hurtado J, Acharya D, Lai H, Sun H, Kallolimath S, Steinkellner H, et al. In vitro and in vivo efficacy of anti-chikungunya virus monoclonal antibodies produced in wild-type and glycoengineered Nicotiana *benthamiana* plants. Plant Biotechnol Journal. 2020;18:266-73; doi: 10.1111/pbi.13194.

154. Eichler R, Lenz O, Garten W, Strecker T. The role of single N-glycans in proteolytic processing and cell surface transport of the Lassa virus glycoprotein GP-C. Virology Journal. 2006;3:41; doi: 10.1186/1743-422X-3-41.

155. Ratanji KD, Derrick JP, Dearman RJ, Kimber I. Immunogenicity of therapeutic proteins: influence of aggregation. Journal of immunotoxicology. 2014;11:99-109.