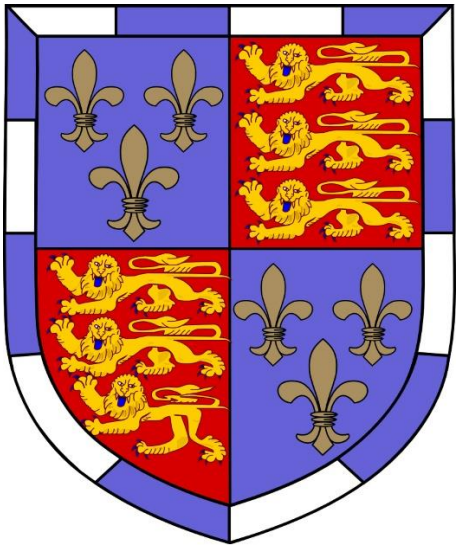
Activation of the TLR4 Signalling Complex by the Ebola Virus Glycoprotein



Michael Jonathan Scherm

Department of Biochemistry
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Christ's College OCT 2021

1 Preface

I hereby declare that the contents of this dissertation are the result of my own work. It, the dissertation, is original and has not been submitted in whole or in part for consideration for any other degree or qualification at the University of Cambridge, or any other University. Unless specifically indicated in the text, this work does not include any results done in collaboration. This dissertation does not exceed the word limit of 60,000 words.

Michael Jonathan Scherm

2 Abstract

Activation of the TLR4 Signalling Complex by the Ebola Virus Glycoprotein

Michael Jonathan Scherm

Infection by the Ebola virus, a member of the Filoviridae family of negative sense unsegmented RNA viruses, leads to acute viral haemorrhagic fever with a fatality rate of greater than 50 % and is currently incurable. As such, understanding the biochemical mechanisms and molecular interactions underpinning Ebola Virus Disease represents a clearly unmet therapeutic need. Initially, the virus replicates to very high levels in macrophages and dendritic cells, both evading recognition and also actively suppressing the innate immune system. The virus load leads to the massive release of cytokines, an unregulated cytokine storm which develops into sepsis-like haemorrhagic fever characterised by tissue damage, loss of vascular integrity and multi-organ failure. The spike glycoprotein (GP1,2), the only surface protein within the Ebola genome, has been identified as the driver for this hyperinflammatory response at the site of infection. Additionally, GP1,2 is shed into the bloodstream following proteolytic cleavage of GP1,2 by TACE protease at the virion membrane. Shed GP migrates through the bloodstream to remote tissues, activating innate immune cells leading to host-wide cytokine storms. The innate immune receptor Toll-like receptor 4, a pattern-recognition receptor (PRR), usually a bacterial LPS sensor with protective pro-inflammatory signalling, has been identified as the source of hyperinflammation during Ebola virus infection. The molecular basis of TLR4 agonism by GP1,2 and also other viral proteins is unknown to date, and this body of work goes some way to explore this interaction.

This thesis explores the Ebola virus glycoprotein and the associated hyperactivation of the TLR4 signalling pathway. Understanding the role of the innate immune system in Ebola virus infection and the central role of the viral glycoprotein in the disease is vital for developing better treatment and finding potential therapeutic targets. The primary focus during this investigation was the involvement of GP glycosylation in TLR4 activation, as glycans contribute to over 50% of the molecular weight of GP and cover most of the protein surface. As TLR4 is not known to be lectin-like, capable of binding glycans, investigations into a glycoprotein of viral origin capable of TLR4 activation, could reveal novel mechanisms that

may lead to therapeutic advances. To experimentally address these aims, I present the first expression of native and stable Zaire EBOV GP in HEK293 under BSL2 conditions with native folding, glycosylation, and full activity. Additionally, to allow the exploration of ideal expression conditions for native GP, multi-step protein purification strategies were developed to ensure high yields while maintaining protein integrity and avoiding contamination through LPS.

The establishment of these novel methodologies laid the foundation to study the interaction between TLR4 and GP. Through establishing TLR4 activity assays I assessed the involvement of GP glycosylation through individual glycosylation mutants and processed glycans. Further, the possible interaction between the TLR4 signalling complex and GP was assessed with biochemical interaction studies. Surprisingly, although the initial hypothesis was built on the contrary, the findings suggest that the glycosylation of ZEBOV GP, does not play a specific role in the activation of TLR4 signalling complex. However, the glycosylation was found to be vital to ensure GP stability and shielding of vulnerable domains from proteases.

Ultimately, I was able to set up novel EBOV GP production methods and established efficient purification strategies which may be utilised at low cost with high yield of glycosylated and stable GP. I uncovered the importance of glycosylation for GP stability, and if absent has an indirect negative impact on protein activity and consequently TLR4 activation. These discoveries and novel methodologies contribute largely towards our understanding of EBOV GP and TLR4 activation, laying the foundation for further investigations into the unusual receptor-ligand interaction and how it may be therapeutically exploited.

3 Dedication

To my parents and grandmother

and

My 9-year-old self who apparently signed with "Dr Mikel"

4 Acknowledgments

Firstly, I would like to thank the Oliver Gatty foundation for their funding, providing me with the stability and opportunity to conduct my PhD. Particularly, I would like to thank my supervisor Professor Nicholas John Anderson Gay for providing me the guidance and freedom to follow my own path in the lab, allowing me to grow as the independent and confident scientist I am today. Thank you for giving me the opportunity to experience Cambridge and Christ's College. I would also like to express my sincere gratitude to Dr. Olaniyi Opaleye for his help on countless occasions, especially during times of illness and injuries when I was not able to work at my normal level and his inextinguishable kindness.

Additionally, I am thankful to my colleagues I have worked with over the years. I thank Dr. Johannes Lauenstein as a scientist and friend who made working till midnight after gradhall to finish that very important purification a little less insufferable. Dr. Atul Udgata, whose never-ending wisdom filled the room on any occasion, with or without request. Dr. Sandro Soares for showing me how easy cloning can be, sometimes. Dr. Lee Hopkins and Dr. Miranda Lewis for teaching me the ropes when it comes to keeping cells alive, particularly stable cell lines. Furthermore, I thank every other member of the NJG lab for making this PhD never boring and creating a space to flourish.

Importantly, I would like to thank Dr Darcie Mulhearn for being the best. Thank you for everything I could ever ask for and so much more. Your compassion and sense of humour made grey times less grey and bright times even brighter. The owner of the most infectious giggle I know, made to make one smile.

Certainly, without a life outside of the lab I would not have gotten this far. Thank you Dr Meltem Gurel for offering home cooked meals away from home, Emile Marin for looking after my physical health, and for the best Mediterranean housemates one could wish for. Thanks to my two Bulgarians, Aleksandar Gyorev for being the cheer-up and oracle when I needed it and Dr Sammy Mahdi for being a sarcastic but supportive German in disguise. Thanks to Max Krause from back home who would always pick me up when visiting, only to kidnap me for a beer before allowing me to see my parents.

Finally, I would like to thank my parents, Hilde and Martin from the bottom of my heart. Without their care, unlimited support, and freedom, I would not be the man I am today. Even if it meant me only being home twice a year, you encouraged and trusted me to follow my own path.

5 Table of Contents

6 List of Figures

7 List of Tables

8 List of Abbreviations

Abbreviation	Full name
°C	Degree Celsius
µg	Microgram
µl	Microliter
µM	Micromolar
Aa	Amino acids
ADAM	A disintegrin and metalloproteinase
AMP	Adenosine mono phosphate
AP-1	Activator protein 1
APS	Ammonium persulfate
ARD5	Monoclonal antibody against tim
Asn	Asparagine
BCAP	B-cell adapter for PI3K
BSA	Bovine serum albumin
BSL	Bio safety level
$CaCl_2$	Calcium chloride
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
CatB	Cathepsin B
CatL	Cathepsin L
CD 14	co-receptor cluster of differentiation 14
cGAMP	cyclic GMP-AMP
cGas	cGAMP synthase
CIP	Calf Intestinal Alkaline Phosphatase
CL	Cell lysis
cm	Centimeter
CO_2	Carbon dioxide
CRL	C-type lectin receptors
Cys	Cysteine
D	Dialysis sample
DAMP	Damage associated molecular patterns
DC	Dendritic cell
ddH_2O	Double distilled water
DENV	Dengue virus
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dsRNA	Double stranded RNA
E or Elu	Elution fraction
EBOV	Ebola Virus
ECD	Ectodomain
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
Endo H	Endoglycosidase H
ER	Endoplasmic reticulum
EtOH	Ethanol
EVD	Ebola virus disease
F	RSV fusion protein
FBS	Fetal bovine serum
FT or Flow Thr	Flow through fraction

Gal	Galactose
Glu	Glutamate
GMP	Guanosine mono phosphate
GnTI-	N-acetylglucosaminyltransferase
GP	Glycoprotein
GPcl	Cleaved Glycoprotein
h	Hour
HA	Human influenza hemagglutinin
HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HMGB1	High mobility group box protein 1
HR	Heptad repeats
HRP	Horseradish peroxidase
IEC	Ion exchange chromatography
IFL	Internal fusion loop
IgG	Immunoglobulin G
IKK	Inhibitor of nuclear factor kappa-B kinase
IL	Interleukin
IMAC	Immobilized metal-affinity chromatography
IRAK	interleukin receptor-associated kinase
IRF	IFN-regulatory factor
kb	Kilobases
KCl	Potassium chloride
kDa	Kilo Daltons
LB	Luria-Bertani) liquid medium
LPS	Lipopolysaccharide
LPSRS	LPS from Rhodobacter sphaeroides
LRR	Leucine-rich repeats
M	Molar
MAL	MyD88-adapter-like
MAPK	Mitogen-activated protein kinases
MBL	Mannose-binding sera lectin
MD2	Myeloid differentiation protein 2
Med	Medium
MgCl	Magnesium chloride
$MgSO_4$	Magnesium sulphate
MHC	Major histocompatibility complex
min	Minute
ml	Milliliter
MLD	Mucin-like domain
MnC_{l2}	Manganese chloride
MoA	Mechanism of action
MyD88	myeloid differentiation factor 88
NaCl	Sodium chloride
Neu	Neuraminic acid
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
ng	Nanogram
Ni2+	Nickel

N-linked	Glycan linked to nitrogen of asparagine
NLR	Nucleotide oligomerisation domain (NOD)-like receptors
nm	Nanometer
NOD	Nucleotide oligomerisation domain
NP	Nucleoprotein
NPC1	Niemann-Pick Disease intracellular Cholesterol transporter 1
NS1	Non-structural protein 1
OAS	Oligoadenylate synthase
OD	Optical density
O-Glyco	O-Glycosidase
O-linked	Glycan linked to oxygen of serine/threonine
ON	Over night
ORF	Open reading frame
PAMP	Pathogen-associated molecular patterns
PBST	phosphate-buffered saline with Tween 20
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEI	Polyethyleneimine
PNGase F	Peptide - N -Glycosidase
PRR	Pattern recognition receptors
PS	Phosphatidylserine
PYHIN	Pyrin and HIN200 domain-containing
RBD	Receptor binding domain
RCF	Relative centrifugal force
REBOV	Reston Ebola virus
RIG-I	Retinoic acid-inducible gene-I
RIP	Receptor-interacting serine/threonine-protein
RLR	Retinoic acid-inducible gene-I (RIG-I)-like receptor
RNA	ribonucleic acid
RNAi	Interfering RNA
RT	Room temperature
s	Second
SARM	Sterile-alpha and Armadillo motif containing protein
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SDS	Sodium dodecyl
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
Ser	Serine
sGP	Secreted glycoprotein
SOC	Super Optimal broth with Catabolite repression
SP	Signalling peptide
ssGP	Small, secreted glycoprotein
ssRNA	Single stranded RNA
T- cell	T lymphocyte
TACE	TNFα-converting enzyme

TAE	Tris-Acetate-EDTA
TAK1	TGF-activated kinase 1
TAK242	Resatorvid
TANK	TRAF family member-associated NF-kappa-B activator
TBK1	TANK-binding kinase 1
TCR	T-cell receptor
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
Thr	Threonine
TIM-1	T cell immunoglobulin and mucin domain 1
TIR	Toll/IL-1R resistance domain
TLR	Toll-like receptor
TMD	Transmembrane domain
TNF	Tumour necrosis factor
TRAF	Tumor necrosis factor receptor-associated facto
TRAM	TRIF-related adaptor molecule
TRIF	TIR-domain-containing adapter-inducing interferon-β
TRIS	tris (hydroxymethyl)aminomethane
Tris HCl	Tris (hydroxymethyl)aminomethane hydrochloride
U	Enzyme Units
V	Volt
Val	Valine
VHD	Viral haemorrhagic disease
VIC	Viral Haemorrhagic Fever Immunotherapeutic Consortium
VP	Virion protein
W	Wash fraction
WT	Wild type
Xaa	Any amino acid
ZEBOV	Zaire Ebola Virus

1 Introduction

This thesis explores the Ebola virus glycoprotein and the associated innate immunity signalling mechanism, central to the severity of Ebola Virus Disease. Within this Chapter a brief overview of the relevant literature is presented. Understanding the role of the innate immune system in Ebola virus infection and the central role of the viral glycoprotein in the disease is vital for identifying potential therapeutic targets.

1.1 The Ebola Virus

1.1.1 Identification of Ebola virus species and their natural reservoirs

The first documented outbreak of the Ebola Virus (EBOV) occurred in 1976 in the Democratic Republic of the Congo (DRC, formerly Zaire), starting with a 44-year-old patient who presented with feverish illness and subsequently claiming 280 fatalities within two months [1]. Initially, the disease was misidentified as severe cases of known viral haemorrhagic diseases (VHD) such as Crimean Congo haemorrhagic fever or Marburg disease, due to identical pathology. Additionally, electron microscopy of human liver samples of patients revealed a morphology indistinguishable from the Marburg Virus (MARV) and a model was created (Figure 1.1)[2]. However, due to the distinct features in serology the VHD was classified as Ebola virus, after the nearby Ebola River. Several months earlier, another VHD outbreak in Sudan was retrospectively identified as a different EBOV subtype with distinct phylogeny to the Zaire EBOV species (ZEBOV). At the time of writing, six species of EBOV have been identified: Zaire, Sudan, Bundibugyo, Tai Forest, Reston, and Bombali [3].

The EBOV species of Zaire, Sudan, and Bundibugyo cause Ebola Virus Disease (EVD) with an average case fatality of 50%, varied from 25% to 90% subject to the subtype and individual outbreak (Zaire being the most pathogenic). The other EBOV subtypes are less pathogenic as exemplified by the Tai Forest species which causes a milder form of VHD and is associated with non-fatal infections in humans, whereas the Reston species (native to the Philippines) and the Bombali species (discovered in Sierra Leone, 2018) are considered non-pathogenic to humans [4]. Understanding the fundamental biochemical differences between these species, and how this correlates with pathogenicity and their ability to modulate the immune system is key to effectively curing EVD.

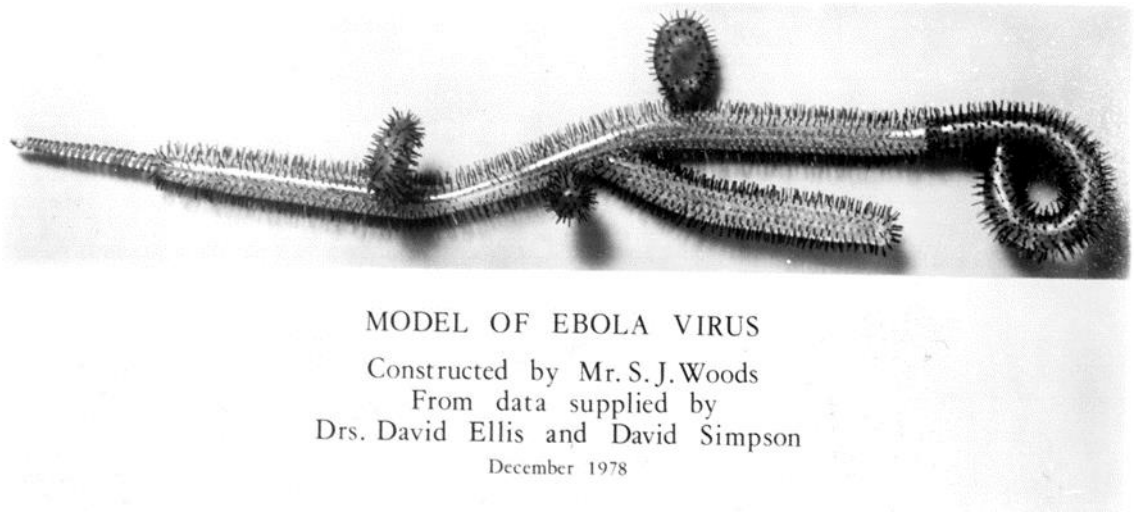


Figure 1.1: First model of the Ebola Virus. The model was constructed based on the first microscopy images of the virus [2].

Despite the ability to be infected by EBOV and the subsequent severity of the EVD, humans are considered non-native hosts to all EBOV species. In nature, human-primates such as chimpanzees, gorillas, and forest duikers were the first identified EBOV carriers. However, these carriers are referred to as "dead-end" hosts, like humans, with the lethality of infection preventing further host-to-host viral transmission [5]. It has been argued that the most likely natural reservoirs for Ebola viruses are bat species and a few non-human primates, in which the virus causes little to no pathogenicity, thus facilitating transmission [6]. It is plausible that the virus is transmitted to humans during bat hunting, with other forest animals as intermediate hosts, and through bodily fluids in bat droppings [7]. Supporting this theory of EBOV mainly existing in bat reservoirs, sequencing of free-tail fruit bats identified them as a direct source of the Sierra Leone outbreak [8]. The human-non-pathogenic Reston virus presents an exception whose natural hosts are native pigs of the Philippines[9]

Despite humans being natural dead-end hosts, the occurrence of a spill-over event from animals to humans facilitates the incidence of human-to-human transmission via direct contact with bodily fluids of an infected person. Additionally, following acute infections, EBOV can persist in immune-privileged sites in a dormant state being shielded by the survivor's immune system, i.e. the central nervous system, interior of the eyes, heart, and genitalia[9]. This persistence can last over a year with a high risk of leading to new infections. One of the often-observed locations associated with the persistence of EBOV and high viral load is semen. Having recovered from EVD and being symptom-free for many months, any individual, particularly men may transmit EBOV through unprotected oral, vaginal, or anal sex, a common scenario in parts of the world such as the war-ridden DRC with little to no access to contraception to prevent sexually transmitted diseases[10].

1.1.2 Pathogenicity and clinical symptoms of EBOV infection

Once infected by the EBOV the subsequent clinical symptoms can be grouped by the initial 'dry' symptoms occurring up to 7 days post-infection and 'wet' symptoms typically appearing 14 days after infection, classified as mild to severe and possible lethality. The first symptoms are fever, fatigue or asthenia, myalgia, and arthralgia[11]. As many other, less-severe, viral infections share similar symptoms a non-neglectable percentage of EVDs are overlooked and misdiagnosed. Approximately two weeks post-infection the typical and characteristic symptoms of EVD are observed. These range from the worsening of the 'dry' symptoms together with anorexia, nausea, vomiting, and abdominal pain in mild cases. During severe disease progression a system-wide cytokine storm, the molecular mechanism discussed in detail in Section 1.2.5, leads to septic shock and sudden haemorrhages at mucosal sites internally and externally. The excessive bleeding from eyes, mouth, skin lesions, and bloody diarrhoea rapidly deplete the patients of necessary fluids, followed by multi-system organ failure and fatal outcomes in 50% of all infections [12].

On a cellular level, an infection by EBOV first takes place at mucosal membrane sites. Although endothelial cells are the first site of viral infection, EBOV subsequently infects and primarily replicates in monocytes and dendritic cells, eliciting its associated pathogenesis via deregulation of both the innate and adaptive immune response[13]. The innate immune system is first affected by this virus, for example by inhibition of Type-I interferon response and the functional impairment of dendritic cells and natural killer cells [14]. The viral envelope is covered by the spike glycoprotein (GP), a transmembrane protein with extensive and complex glycosylation capable of shielding from and misdirecting immune cells responsible for antibody production [15]. As the virus evades the immune system it can freely replicate, leading to high viral load and inflammation signalling throughout every tissue [16]. Additionally, in all infected cell types, albeit through unknown mechanisms, the EBOV impairs expression and transport of many cell surface proteins including integrins and some immune receptors such as MHC I & II. This downregulation does not only inhibit the recognition of viral epitopes and the adaptive response but also destabilises cell-to-cell adhesion [17]. During the Ebola infection the constant inflammatory signalling impacts the surrounding tissue to such an extent that unregulated cell death occurs, resulting in collateral damage and vascular leakage. Here, the tight junctions between endothelial cells surrounding the blood vessels are disrupted, allowing viral particles and viral proteins to enter the blood stream, infect and activate remote cells in completely different tissues from the initial site of infection Figure 1.2.

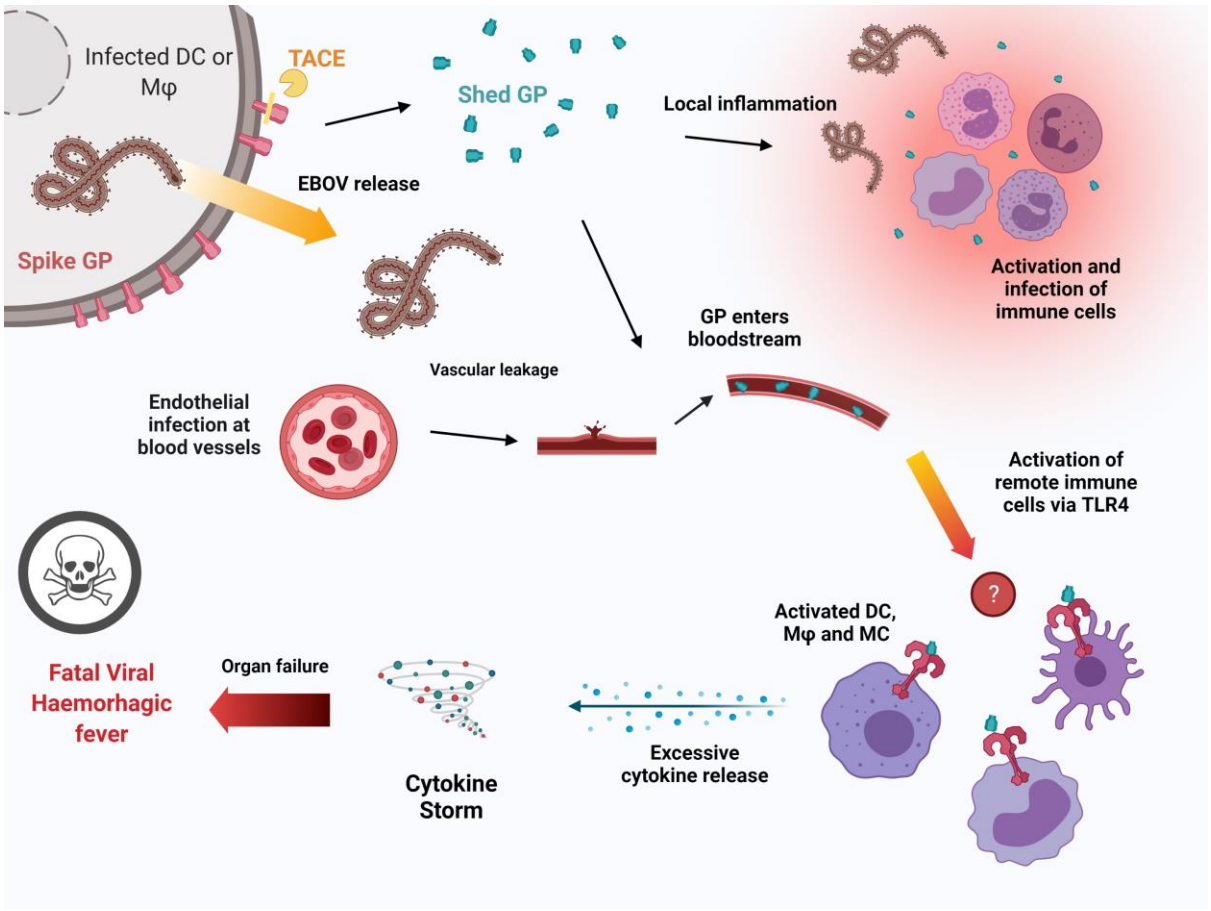


Figure 1.2. Shedding of GP leads to a cascade of inflammatory reactions. The sequential effects of EBOV infection and the role of GP on the host's body and immune system are depicted in this schematic. After TACE cleave, shed GP (blue) activates TLR4 locally leading to local inflammation. Following, the vascular leakage induced by endothelial cell infection, GP diffuses into the bloodstream, migrates and activates cells remote from the site of infection. This system-wide activation leads to the escalating activation of cytokines, formation of a cytokine storm and finally to fatality in the majority of patients.

The EBOV infection affects not only the innate and adaptive immune response but also disrupts the highly complex interplay and communication between the two immune systems. The interaction between dendritic cells and T-cells is severely disrupted due to downregulated surface receptors required for communication but also the hyperinflammatory microenvironment caused by the cytokine storm and permanent activation of MHC-peptide/TCR interaction (despite downregulation). This disruption affects the activation of the adaptive immune response especially[18]. The disordered DC/T synapse, a communication complex of immune and signalling receptors, fails to induce CD4 T cell clonal expansion and consequently prevents the activation of naïve CD8 T and B cells leading to T cell apoptosis and inhibited antibody production [19]. However, the exact mechanism remains to be identified, particularly due to the unspecific downregulation of cell surface proteins.

The pathogenicity and cytotoxicity of an EBOV infection is not solely caused by the vast levels of viral titres in the patient's blood but can mostly be attributed to the effects of GP, which act far beyond the requirement for membrane fusion [20]. The primary effects caused by GP are immuno-evasion, cytokine dysregulation, and the glycoprotein-mediated cytopathology associated with the disruption of host surface proteins and cell rounding[21]. Survivors of EVD have proven to show an extraordinarily robust adaptive immune systems capable of resisting the viral inhibitions, attacks and restricting the otherwise unhindered replication through mounting a response and production of EBOV-specific IgG antibodies. In over 50% of infections, Ebola's pathogenesis impairs the innate and adaptive immune response to such an extent that the virus can replicate almost freely in the host, due to the mechanisms outlined above [22]. If this propagation is left unchecked the viral titres in the blood can exceed 1 million infectious viral particles/mL (10×10^5 TCID 50)and overburden the body within days. Fatalities correlate with high concentrations of virus in the blood and survivors consequently have the least amount of virus present and highest number of neutralising antibodies. Understanding the biochemical differences between survival and lethality on both a cellular and biochemical basis is key to treatment and prevention of this disease.

1.1.3 Treatment and Prevention

To date there is no cure for EBOV infection. Although vaccines, neutralising antibodies, and anti-viral drugs for EVD exist, and many therapeutic options have been fast-tracked through clinical trials, their mode of action is symptomatic treatment i.e., easing severe symptoms and reducing mortality, rather than preventative measures and protection from novel infections. Additionally, a full cure has yet to be identified. Examples of these current available treatments are outlined below.

Much work has been devoted to the development of antibodies that can neutralise EBOV. The monoclonal antibody KZ52, isolated originally from a human survivor of the 1995 Kiewit outbreak, is the first and most extensively researched EBOV neutralising antibody, targeting a glycan-free domain of GP1,2 as revealed via X-ray crystallography [23]. While follow-up studies in rodent models revealed that the antibody reduced ZEBOV lethality, similar immunity was not evident in non-human primates and clinical trials were discontinued. Since then, the Viral Haemorrhagic Fever Immunotherapeutic Consortium (VIC) was founded to identify further anti-EBOV therapeutic antibodies. This global academic initiative used the combined efforts of over 100 research groups to create a database of all antibodies and their corresponding targets against EBOV GP. During the 2014-16 ZEBOV epidemic in Western Africa the efforts were intensified and gave rise to ZMapp (Mapp Biopharmaceutical, Inc., San Diego, CA), a trimeric mAb cocktail effectively used to treat subsequent outbreaks [24].

Independently of academic efforts, pharmaceutical companies produced vaccines and small anti-viral molecules treatments against EBOV resulting in two licensed Ebola vaccines. MERCK (Germany) utilised the vaccine Ervebo, a live vesicular stomatitis virus carrying the EBOV GP gene with an efficacy of 97.5% at preventing infection, while Janssen developed a two-dose vaccine with the vaccine Zabdeno (Ad26.ZEBOV) as the first dose, based on their AdVac viral vector technology and Mvabea (MVA-BN-Filo) as the second dose, based on Bavarian Nordic's MVA-BN® technology with no official efficacy data but immunogenicity data suggests a predicted survival probability of 53.4% and protection of up to 80%[252627]. More recently in 2019, Regeneron announced two successful monoclonal antibody-based treatment approaches: Antibody cocktail, REGN-EB3 and the single antibody mAb114 capable of reducing mortality rates by 33% and 35% respectively[26]. The exact binding has not been published. The Regeneron antibodies showed such promise in clinical trials that they were approved for field utilisation, especially considering the benefits in early-stage patients[28].

In recent years, the race to develop new anti-EBOV vaccines has decelerated, influenced by the success of previous treatments, particularly transfer of neutralising ntibodies such as ZMapp and REGN-EB3,however new vaccines are still being developed and anti-viral compound treatments investigated. Noteworthy is the novel vaccine technology developed at Scripps Institute by Jiang Zhu's research group. By designing a DNA strand containing the genetic information for the optimised EBOV glycoprotein, the nanoparticle-forming unit, a locking domain, and a T-cell epitope, Zhu et al. were able to create self-assembling nanoparticles that form EBOV-mimicking virus-like particles in cells [29]. In addition to an EBOV vaccine, this technology is currently used in accelerated clinical trials against SARS-CoV-2, the cause for the 2019 pandemic lasting to date.

Along with vaccines and antibodies, antiviral compounds such as Remdesivir (GS-5734), a broad-spectrum EBOV antiviral developed by Gilead bioscience, were established as an additional treatment to reduce disease severity, and increase survival of patients [30]. Thus, vaccines and anti-virals have provided tremendous resources in treating the severity of EBOV disease, yet there remains much to be done towards improving vaccines and finding a potential cure for people who have been infected. Solving the mystery of the infection and inflammation mechanism of EBOV would certainly give rise to novel therapeutic targets or therapies with different and perhaps more effective mechanism of actions (MoA).

1.1.4 Genomic organization and processing of EBOV GP

In order to explore therapeutic intervention of EBOV infection, it is first important to understand the viral organisation and significance of the glycoprotein. The Ebola virus belongs to the family of

filoviridae, with negative-sense single-stranded RNA genomes. The Ebola genome encodes seven genes: the nucleoprotein (NP), virion protein 24 (VP24), VP30, VP35, VP40, polymerase (L), and glycoprotein (GP) (Figure 1.3)[31]. Taken together in a structural context, NP, the minor nucleoprotein (VP30), VP35 and the RNA polymerase (L) build the nucleocapsid containing the genomic RNA. VP24 and VP40 form a matrix and connect to the nucleocapsid with the viral envelope which is comprised of a lipid bilayer densely covered by the glycoprotein [32]. The membrane bound trimeric GP is critical in the EBOV life cycle, as it is solely responsible for attachment, membrane fusion and infection of host cells. Moreover, full length GP is responsible for critical pathogenicity, including immune evasion and cytokine storm[33]. As the primary driver in EBOV infection, understanding GP and its role in EVD is vital to enable efficient targeting for therapeutic use.

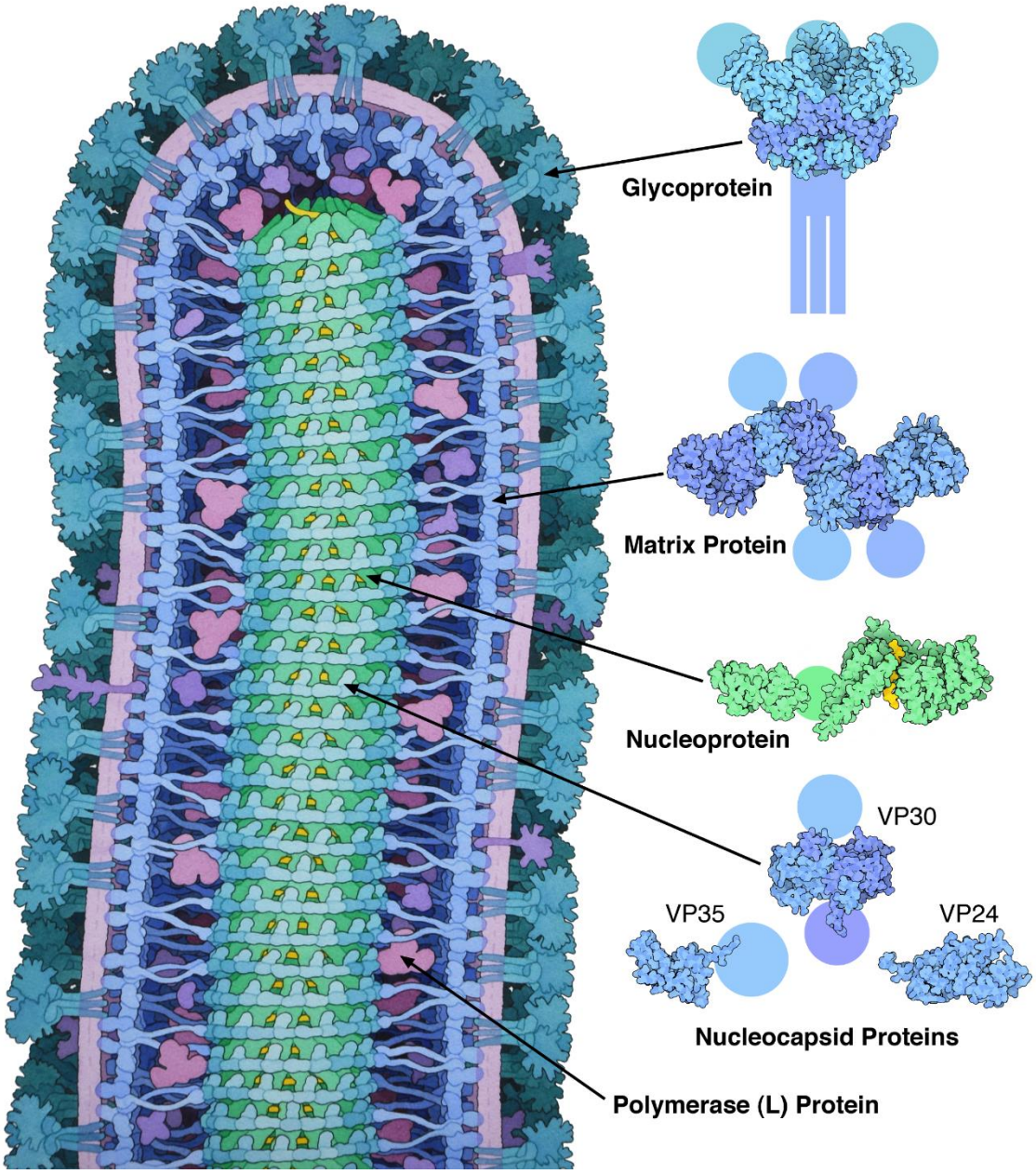


Figure 1.3: The proteins of the Ebola Virus. Schematic drawing of a cross section through the Ebola virus shows the RNA genome (yellow), the viral proteins (blue, green, purple) and the virion membrane (magenta). The atomic structures of each protein are depicted on the right, with unresolved parts of the structure indicated as a circle. Adapted from Goodsell D, 2014.[34]

Due to reading errors of the GP gene during transcription and multiple post-translational processing of precursors, three GP protein variants are expressed by the virus. For example, a unique sequence of seven uracil residues withing the GP gene causes transcriptional stuttering and subsequent frameshift mutation in approximately 20% GP transcription [35].

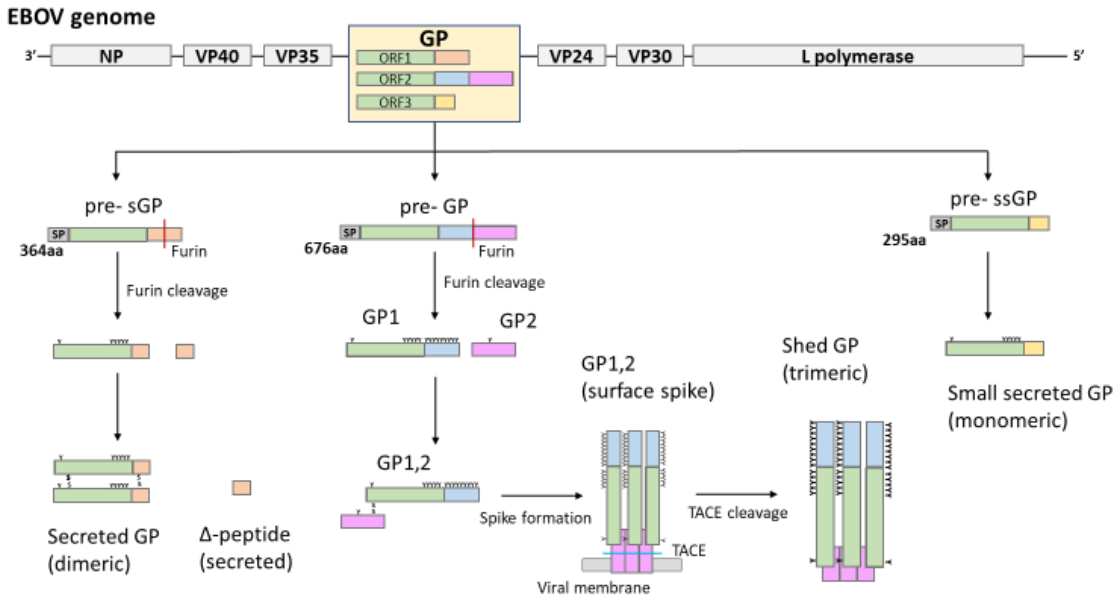


Figure 1.4: The three protein variants of the EBOV GP gene. Due to polymerase stuttering the GP gene is transcribed into three proteins of different lengths. About 80 % of the gene gets transcribed in ORF1 leading to the formation of pre-sGP, 20 % are transcribed to pre-GP in ORF2 and a miniscule amount of <1% are transcribed at ORF3 in pre-ssGP. The sequence in common among all proteins is green, while the different sequences after the stuttering site are orange for sGP, blue and pink for GP and yellow for ssGP. The signalling peptide is indicated as SP (grey). Glycans are depicted as Y-shapes (black). Length of each protein is indicated by XXX aa the number of residues in the peptide chain.

The majority of the GP gene expression occurs without transcriptional shifting and gives rise to the pre-sGP. Post-translational furin-cleavage processes pre-sGP into the 364-residue-long secreted glycoprotein (sGP), a C-terminal fragment and a short peptide, called Δ peptide [36]. After post-translational modifications, such as cleavage by furin in the Golgi, two sGP monomers form a homodimer held together by disulphide bonds and are subsequently secreted [37]. Even though sGP has no known function in infection, it has been suggested that the secretion of sGP plays an important role during the virus' immune system evasion after discovering that survivors' antibodies favour sGP

over the full-length GP and bind with a stronger affinity [38]. However, the full-length form of EBOV GP, GP1,2 is central for the pathogenicity and severe immune reaction (Figure 1.4).

During transcriptional stuttering at the repetitive uracil region, a -1 frameshift causes skipping the stop codon and the synthesis of the 676-amino acid long, full length type I transmembrane GP (pre-GP1,2). As a result of this frameshift, GP and sGP have distinct C-termini following the identical 295 amino acid N-terminus [36]. The pre-GP protein is further cleaved by furin in the Golgi, into two monomers GP1 and GP2 which finally form the 'mature' heterodimer, GP1,2, through a disulphide linkage between Cys53 of GP1 and Cys609 of GP2 [39]. While sGP forms dimers, the final GP1,2 assembles into a 450-kDa spike-trimer at the surface due to the unique structure of the C-terminus. Cleavage by the metalloprotease TNFα-converting enzyme (TACE), a member of the ADAM (a disintegrin and metalloproteinase) proteinase family, also termed ADAM-17, at the transmembrane domain leads to the shedding of the GP1,2 form which is termed shed GP [40].

Another creation of transcriptional stuttering of the GP gene is the small, secreted glycoprotein (ssGP) resulting from a frame shift to -2 ORF in less than one percent of transcription. However, due to the low quantity, lack of any immunogenic activity or impact on both host and virus, this form of altered GP has been deemed a by-product of the stuttering required to produce sGP and GP1,2 [36].

As the name suggests, GP1,2 is heavily glycosylated, especially the 150-residue mucin-like domain (MLD) with up to 80 sites of O- and N-linked glycosylation [39]. The MLD has been shown to be essential for the activity of the glycoprotein in causing dysregulated inflammation and cytokine storms, all contributing to the viral pathogenicity [41]. In 2014, the activity of GP1,2 was first linked to Toll-like receptor 4 (TLR4), when NFκB and inflammatory cytokine production was observed via the TLR4 signalling pathway. Pre-treatment of cells with anti-TLR4 antibodies or deglycosylase was shown to block or decrease the immune cell activation following an incubation with EBOV GP1,2 [20].

1.1.5 Glycosylation and structural features of EBOV GP

The two monomers GP1 and GP2 that form the heterodimeric GP1,2, that assembles into the trimeric spike protein at the virion-membrane, have different functions vital to the viral survival and thus, represent distinct therapeutic opportunities that remain to be explored.

The GP1 monomer has two primary functions in 1) shielding the virus from the immune system with the glycan cap and the heavily glycosylated MLD and 2) attachment to cells by interaction with the cell surface receptors [42]. Structurally, GP1 can be further divided into 3 subdomains: the base that interacts with GP2; the head, which is the central core of the protein; and the glycan cap, which is located

proximally and is heavily glycosylated. The GP1 base interacts with GP2 through the disulphide bridge Cys53 (GP1) and Cys609 (GP2) and a discontinuous section forming a hydrophobic interaction surface with the internal fusion loop (IFL) and the heptad repeat region of GP2. While the head of GP1 is also mostly composed of discontinuous sections between the cap and the base, the glycan cap is comprised of a continuous but highly flexible chain with extensive glycosylation. This glycosylation includes the mucin-like C-terminal domain with more than 20 experimentally proven oligosaccharides and many more predicted [32].

On the other hand, the GP2 monomer is primarily responsible for membrane fusion and can be divided into several domains: the hydrophobic internal fusion loop (IFL); the two heptad repeats (HR1, HR2); the disulphide bond connecting it to GP1; the membrane adjacent stalk domain, and the transmembrane anchoring helices for stabilising trimer formation[32].

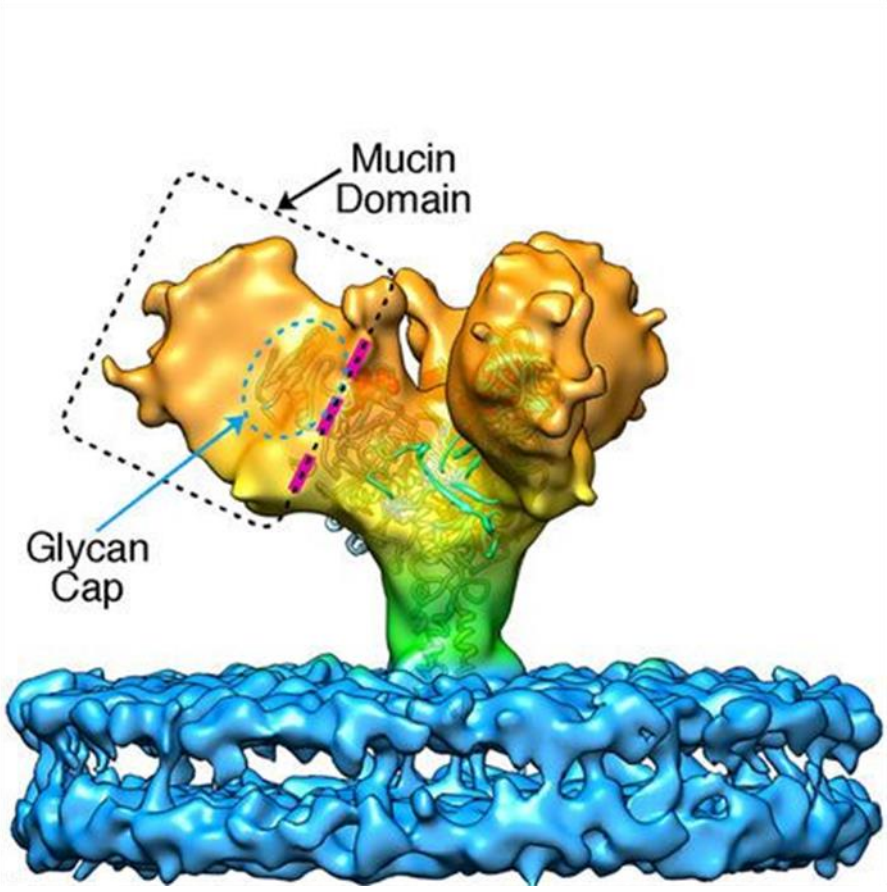


Figure 1.5: Cryo-Electron microscopy structure of EBOV GP with intact MLD glycosylation. Structure of the EBOV GP within the virion envelope at 11 Å resolution, *in situ* from a purified virus particle. The cryo-EM structure was docked to the crystal structure 5JQ7 to indicate the newly resolved electron density of the mucin-like domain. The location of GP1 is indicated in orange and yellow, while GP2 is coloured in green and the virion membrane coloured in blue. Adapted from Beniac et al. 2017 [43].

The IFL is arguably the most essential domain of GP2 due to its ability to initialise the membrane fusion of EBOV in the endosome and presents a key EBOV therapeutic target. Prior to the proteolytic cleavage

in the endosome, the hydrophobic part of the IFL, consisting mostly of tryptophan, leucine, isoleucine, proline, and phenylalanine, is held in place by a hydrophobic pocket in the base of GP1 and shielded from solvent [44].

The EBOV GP, particularly the glycan cap, is heavily glycosylated. With over 50 different N-glycans structures including high mannose, hybrid, and tetra-antennary complex glycans, such as fucose and sialic acids the vast surface of the spike trimer is covered by a glycan shield[22]. In addition to the 17 N-linked glycans, approximately 80 O-linked glycans have been identified and/or predicted, doubling the molecular weight of the GP1,2 from the initial 75 kDa in amino acids to 150 kDa. A 150-residue-long sequence, MLD, located at the C-terminus of the glycan cap has the highest concentration of glycosylation sites within GP. The glycosylation density in combination with its flexibility and lack of secondary structure enables this domain to shield immunogenic protein epitopes and hinder antibody binding. In fact, most accessible epitopes for effective antibody neutralisation are at the base of GP1, the GP2 fusion loop, parts of GP1 head-domain while antibodies against the heavily glycosylated glycan cap or MLD have not shown effective neutralisation [22].

Although EBOV and the glycoprotein have been the centre of many investigations and initiatives since its discovery, the first high resolution structure was only solved in 2008. The crystal structure of the ZEBOV GP trimer (PDB code 3CSY) was solved through crystallisation of the complex of EBOV GP and KZ52 antibody Fabs derived from a human survivor [23]. A decade earlier the crystal structure of GP2 was determined at a low-resolution, suggesting a trimeric protein shape [35]. However, only a partial crystal structure lacking the MLD was solved as the heavy and heterogenous glycosylation prevented crystallisation. Due to the advancement in electron microscopy technology and its increased use in recent years, the complete structure of GP1,2 including the MLD has now been determined [45]. More recently, a cryo-EM structure of GP1,2 of 11 Å resolution bound to the endogenous virion envelope was determined (Figure 1.5). Even though the resolution was lower in comparison to other EM structures of the protein, for the first time the transmembrane domain of GP2 and the orientation of the MLD and the glycan cap could be shown [43]. Considering the importance of the GP for immune evasion and TLR4 engagement, these previous investigations into the protein and glycan structure underline the importance for further research to understand how GP can be targeted.

1.1.6 Mechanism of EBOV infection and endosomal escape

Although the cellular mechanism of action through which EBOV elicits infection has been investigated well, understanding the molecular mechanism has proven more challenging and remains unknown. In the last decade, the virion-bound spike glycoprotein GP was identified as the main driver of viral

membrane fusion and infection with host cells [33] . More recently, the host-associated proteins crucial for EBOV infection and endosomal escape, T cell immunoglobulin and mucin domain 1 (TIM-1) and Niemann-Pick Disease intracellular Cholesterol transporter 1 (NPC1) respectively were discovered (Figure 1.6) [46 47]. Understanding the precise biochemistry of EBOV infection will pave the way for novel and effective targets, such as these key proteins: GP, TIM-1, and NPC1.

Further knowledge of the structure and interaction surfaces of such key viral proteins allows for an understanding of how the virus evades the immune system and thus helps combat this immune evasion. For example, the N-terminal region (54-201aa), termed receptor binding domain (RBD), is located at the top of GP, slightly recessed in the cup-like head domain. There it is masked from the immune system by the unstructured, flexible, and highly glycosylated mucin-like domain and the adjacent GP1 glycan cap. This RBD is believed to be required for or at least supporting the interaction of EBOV and the receptor, required for viral uptake [47].

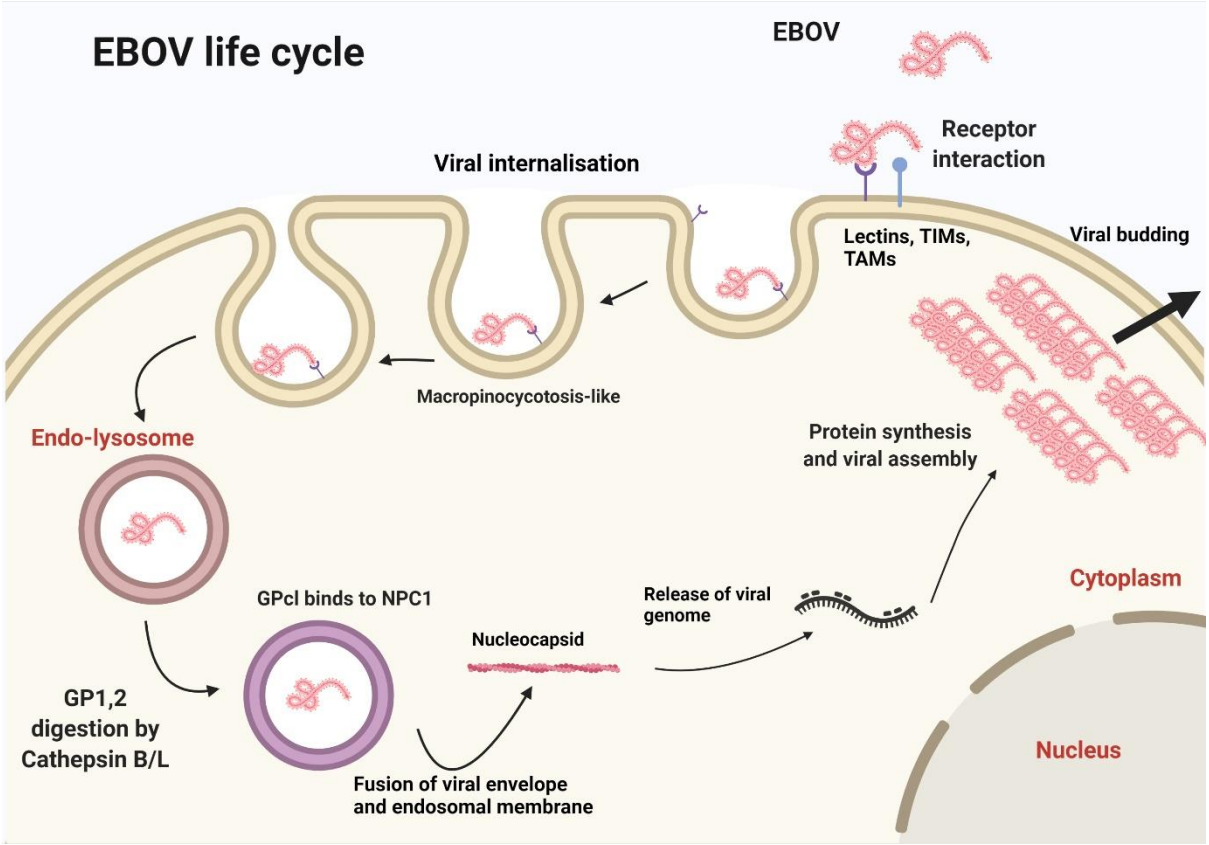


Figure 1.6: Ebola virus life cycle in an infected cell. The infection of a cell (dendritic) by EBOV (pink) starts by interaction with extracellular membrane receptors leading to the internalisation of the virus by a micropinocytosis-like process. This results in the formation of a vesicle which merges with endo-lysosomes. Within the acidic environment GP is cleaved by cathepsins, forming GPcl which interacts with Niemann-Pick disease type c cholesterol transporter (NPC1) leading to membrane fusion of the virus and the endosome. This membrane fusion releases the viral proteins and genome into the cytosol where protein synthesis and viral assembly occurs. The newly formed EBOV viral particles bud at the cell membrane and escape.

Another therapeutic option is targeting the host proteins which the EBOV receptor interacts with in order to infect cells. In the case of ZEBOV infection, the virus binds TIM-1, a phosphatidylserine (PS) receptor that usually binds PS on cell surfaces to initiate cell clearance of apoptotic cells. Viruses such as EBOV are capable of mimicking these apoptotic bodies by accumulating PS on their virion membrane. It has been suggested that TIM-1 could promote viral infections through interaction with virion-associated PS [48]. Reflecting the importance of this for disease severity, TIM-1-deficient mice had significantly better survival following ZEBOV infection than their wild-type counterparts. Furthermore, ectopic TIM-1 expression in cells increases EBOV infection 20-fold [49]. Equally, RNAi treatment to reduce TIM-1 surface expression or exposure with ARD5, a TIM-1 specific monoclonal antibody, decreases the infection severity and subsequent cytokine storm. As such, TIM-1 is a potential target for antiviral therapy, and further understanding of the molecular mechanism of this and other proteins will help identify novel EBOV treatments.

Once the EBOV has entered the host cell via phagocytosis, the virus is engulfed by a vesicle which matures into an endosome[50]. Usually, when endogenous or foreign bodies are taken up by endosomes, they are degraded by the combination of an acidic environment and the presence of proteases. However, EBOV has evolved to utilise that environment to its advantage to escape the endosome and deliver its genetic material into the cytoplasm where it is translated and assembled into new viral particles [32]. Again, investigating the molecular basis of this would allow to understand the precise protein residues involved and would highlight ways to exploit them for novel therapies. Here, the acidic endosomal environment protonates key histidine residues of GP, initiating a conformational change that exposes the b13-b14 loop of GP1. This loop is then targeted by the protease cathepsin L (CatL) and cathepsin B (CatB). In this process GP1 is trimmed from 130 kDa to 19 kDa, while still being bound to GP2 through disulphide linkages, generating the primed cleaved GP (GPcl). During this important cathepsin cleavage, the glycan cap and MLD of GP are removed, revealing an initially shielded hydrophobic cavity. The new accessible hydrophobic cavity of GP interacts with the Niemann-Pick C1 endosome receptor (NPC-1), thus revealing another potential therapeutic target [51]. Upon engagement of GPcl with NPC-1, GPcl undergoes further conformational changes, with the b7-b8 internal fusion loop (IFL) moving downwards while the α1 helix moves closer to NPC1 (Figure 1.7). Through this interaction the IFL is positioned at the top of GPcl in a pre-hairpin intermediate and following a 'backfold' distorts the viral and endosomal membrane, facilitating the formation of a hemi-fusion stalk. After juxtaposition of IFL and the transmembrane domains the fusion pore is established. This triggers membrane fusion with the endosomal membrane, thus enabling EBOV to escape the endosome and replicate in the cytosol [52]. To date, it is unknown whether GPcl may have more

functions in addition to endosomal escape such the activation of immune receptors leading to signal transductions (Figure 1.6).

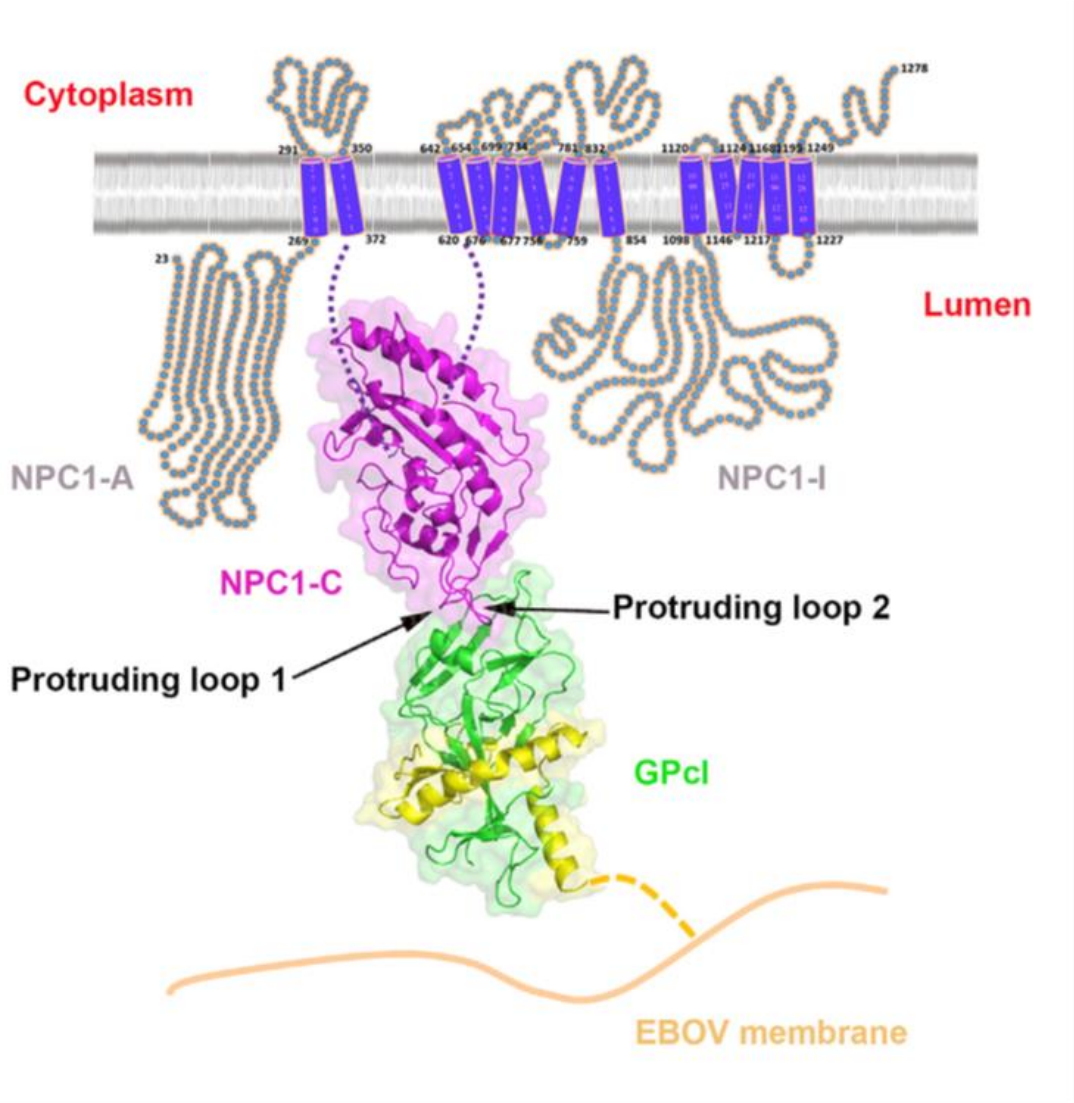


Figure 1.7: Complex of GPcl with NPC1-C. Schematic depicting of the crystal structure of prime cleaved GPcl in complex with domain C of Niemann-Pick C1. Following cleavage of GP by CatB/CatL the newly formed GPcl interacts with NPC1. Two protruding loops of NPC1 engage with a hydrophobic cavity at the head of GPcl. This interaction leads to further hydrophobic changes in GPcl releasing the additional N-terminal end of IFL and finally triggering membrane fusion and core pore formation, allowing the virus to escape the endosome. Adapted from [46].

EBOV infections dysregulate immune responses to such a degree that protective mechanisms cause severe damage to the host, in addition to supressing antiviral activity and maintaining uncontrolled viral replication, usually leading to death. Mechanistic insights how EBOV modulates the innate immune response and also affects the innate-adaptive interface may help to develop a better understanding of EBOV infection.

1.2 Innate immunity associated with EBOV infection

The EBOV is capable of modulating and altering the immune system, particularly dysregulating the innate immune system which plays a central role in EVD as a consequence [13]. To understand the effects that EBOV infection has on the body it is important to revisit the mechanisms of innate immunity and connection to disease.

1.2.1 Pattern recognition receptors in innate immunity

All multicellular organisms use an innate immune system as the first line of defence against infections and cellular damage. Innate immune responses are mediated by cell surface, cytosolic, and endosomal pattern recognition receptors (PRRs)[53]. PRRs are capable of recognising pathogen-associated molecular patterns (PAMPs), which are conserved microbial ligands such as bacterial polysaccharides and lipids; glycol- and lipoproteins; and foreign, predominantly viral nucleic acids [53]. In addition to PAMPs, PRRs respond to damage associated molecular patterns (DAMPs), also known as alarmins, first proposed by P Matzinger as the 'danger theory' which primarily consist of proteins released upon cell damage or death that activate the immune system [54].

PRRs can be classed into several major groups including the Toll-like receptors (TLRs); the nucleotide oligomerisation domain (NOD)-like receptors (NLRs); the Retinoic acid-inducible gene-I (RIG-I)-like receptor (RLR) family; the PYHIN (pyrin and HIN200 domain-containing) family; the C-type lectin receptors (CRLs) and oligoadenylate synthase (OAS) proteins, and the related protein cyclic GMP-AMP (cGAMP) synthase (cGAS), each with their own unique signalling pathway and recognised ligands [55]. The TLR family is the main scope of this work, and as such will be described in more detail due to their important role in EVD.

1.2.2 Toll-like receptors signalling pathways

TLRs are a family of class I transmembrane pattern recognition receptors. As class I receptors, TLRs contain a leucine-rich repeat (LRR) ectodomain for recognising molecular ligands and the intracellular Toll/IL-1R (TIR) domain important in receptor dimerisation and signal transduction [56]. TLR activation by PAMPs or DAMPs mediates inflammatory responses to cell damage and protective innate immune reactions against infections [57]. Overall, since the initial discovery, retrospectively named human TLRs4 in 1997 by Medzhitov, Janeway *et al.,* based on work by Hoffmann and Beutler *et al.,* ten human TLR receptors have been characterised including their activation mechanisms and the majority of their ligands (Figure 1.8) [59] [60] [61].

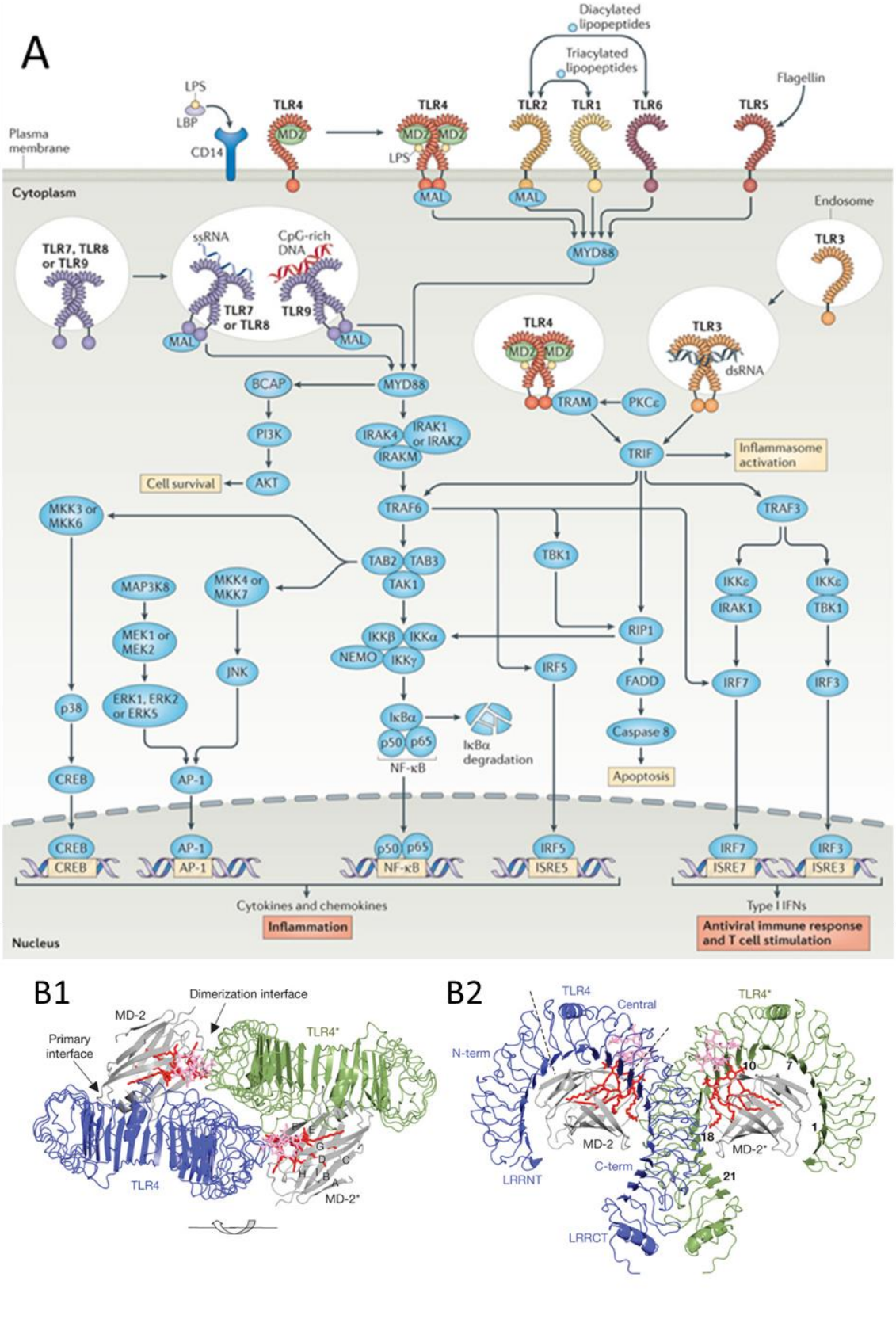


Figure 1.8: Overview of TLR signalling pathways and TLR structure. A) Toll-like receptors (TLRs) are located at the cell surface and in endo-lysosomes, where they recognise pathogen-associated molecular patterns or danger-associated molecular patterns. TLR stimulation leads to the pathway activation involving myeloid differentiation primary response protein 88 (MYD88) and or TIR domain-containing adaptor protein inducing IFNβ (TRIF). Engagement with other pathways ensures regulation of the signalling cascade leading either to cell survival or apoptosis, and the transcription of pro-inflammatory

cytokines and chemokines, and type I interferons (IFNs). AP-1, activator protein 1; CREB, cAMP-responsive element-binding protein; dsDNA, double-stranded DNA; dsRNA, double-stranded RNA; ERK, extracellular signal-regulated kinase; FADD, FAS-associated death domain; IκBα, inhibitor of NF-κBα; IKK, inhibitor of NF-κB kinase; IRAK, interleukin-1 receptor-associated kinase; IRF, IFN-regulatory factor; ISRE, IFN-stimulated response element; JNK, JUN N-terminal kinase; LBP, LPS-binding protein; LPS, lipopolysaccharide; MAL, MYD88 adaptor-like protein; MAP3K, mitogen-activated protein kinase kinase kinase 8; MD2, myeloid differentiation factor 2; MEK, mitogen-activated protein kinase/ERK kinase; MKK, mitogen-activated protein kinase kinase; NEMO, NF-κB essential modulator; NF-κB, nuclear factor-κB; PI3K, phosphoinositide 3-kinase; PKCε, protein kinase Cε; RIP1, receptor-interacting protein 1; ssRNA, single-stranded RNA; TAB, TAK1-binding protein; TAK1, TGFβ-activated kinase 1 (also known as MAP3K7); TBK1, TANK-binding kinase 1; TRAF, tumour necrosis factor receptor-associated factor; TRAM, TRIF-related adaptor molecule. B) Crystal structure of TLR4-MD2-LPS complex. B1) Top view of symmetrical dimer of TLR4-MD2-LPS complex. The primary interface between TLR4 and MD2 is formed prior to LPS binding, while the dimerisation interface is induced following the binding of LPS. B2) Side view of the complex. The interacting module numbers of LRRs of TLR4 and the β strands of MD-2 are given in black. TLR4 is categorised in central and N- and C-terminal domains. LRRNT and LRRCT indicate the termini of LRR modules. TLR4 is coloured in green and purple, MD-2 in grey, the lipid A component of LPS is red and the carbohydrates components in pink. Adapted from [62] [63].

Based in their subcellular localisation, TLRs either belong to the group located in the endosome, such as TLR3, TLR7, TLR8, TLR9, or the group that directly recognises ligands at the cell surface and signal including TLR1, TLR2, TLR3, TLR4, TLR5, TLR6, TLR10 (although TLR4 is also capable of being activated in the early endosomal compartment)[64].

With the exception of TLR4, which requires another protein, TLRs form homo or heterodimers and this interaction alone is sufficient for recognising foreign molecules and initiating signal transduction. Activation of TLR4 on the other hand requires the co-receptor myeloid differentiation factor 2 (MD2) that binds the lipid A acyl chain of bacterial lipopolysaccharide ligand (LPS) in a hydrophobic pocket, leading to receptor dimerisation and signalling (Figure 1.8) [65]. To bind and transfer LPS to TLR4 MD2 requires CD14 to collect LPS and shuttle it to MD2. The lack of co-factor requirement for activation of the other TLRs reflects their different ligands. For example, TLR2 recognises triacylated or diacetylated lipoproteins and induces signalling via heterodimerisation with TLR1 or TLR6, respectively[66]. The latest discovered TLR10 is known to form heterodimers with TLR1 and TLR2, but is considered an orphan receptor with unknown ligands, whereas the interaction between the TLR5 and bacterial flagellin at the cell surface is well studied[67]. Furthermore, endo-lysosomal TLR3, TLR7, TLR8 and TLR9 form homodimers and recognise viral nucleic acids, specifically dsRNA (TLR3), ssRNA (TLR7 and TLR8), and unmethylated CpG nucleotides (TLR9)[68]. Thus, although TLRs are a diverse family of receptors with different ligands and binding partners, they are unified by a common theme of dimerization induced by ligand recognition at the TLR LRR ectodomains.

TLR dimerisation moves TIR domains close to each other to create a scaffold for the recruitment of six TIR- domain containing adaptor proteins MyD88, MAL, TRIF, TRAM, SARM and BCAP [69] [70] [71] [72] [73] [74]. With the exception of TLR3, the recruitment of MAL to the TIR domain of TLRs initialises a myeloid differentiation factor 88 (MyD88)-dependent signalling pathway. In the case of TLR3, signalling transduction occurs through a MyD88-independent pathway that requires TRAM to signal through TRIF adaptor protein. TLR4 is uniquely capable of utilising both MyD88-dependent and -independent

pathways. Initially, TLR4 is activated at the cell surface by extracellular LPS triggering the MyD88-dependent pathway but also, after TLR4 receptor internalisation to the endosome, TRAM and TRIF activate the MyD88-independent pathway without the need for additional proteins besides MD2 [72]. The EBOV GP is capable of activating TLR4 before and after receptor endocytosis, giving it a unique ligand activity.

Of the six TIR-domain containing adaptor proteins, only SARM and BCAP are negative regulators of TLR activation [73] [74] [75]. On a molecular level, TLR activation leads to a kinase signalling cascade, which increases the expression and activity of immune system specific transcription factors. Generally, all TLR signalling leads to the activation of Nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) and activator protein-1 (AP-1) although, depending on the subcellular localisation of TLRs and numerous signalling adaptor proteins, other transcription factors such as interferon regulatory factors (IRFs) can also be activated [76] which also occurs in EVD. Despite the general signalling pathway similarities, the MyD88-dependent and MyD88-independent pathways recruit different adaptor and cascade associated proteins and thus activate a different suite of transcription factors. Further differences between the two pathways will be outlined below.

Upstream of the MyD88-dependent pathway, MAL associates with TIR domain of activated TLRs and recruits MyD88. Subsequently, the Myddosome is formed, a self-assembling helical complex comprised of the oligomeric IL-1R-associated kinase (IRAK) family proteins IRAK1, IRAK2, and IRAK4 and the initial seed protein, MyD88[77]. Following the Myddosome assembly a series of adaptor proteins are activated in a signalling cascade. Firstly, phosphorylation of IRAKs activates the E3 ubiquitin ligase tumour necrosis factor receptor-associated factor-6 (TRAF6). Activation of TRAF6 leads to the poly-ubiquitination of the Myddosome, further stabilising the complex but also self-ubiquitination[78]. This in turn initiates the inhibitor of nuclear factor kappa-B kinase (IKK) complex formation and the MAPKKK7, also known as TAK1, complex formation. While the TAK1 complex formation activates the mitogen-activated protein kinase (MAPK) cascade which finally promotes AP-1 family transcription factors activation, TAK1 also phosphorylates the IKK complex which leads to the assembly of the NF-κB complex and downstream cytokine expression[79] which is extensive during EVD.

The MyD88 independent pathway leads to activation of the transcription factors NF-κB, AP-1 and IRF family including IRF3, IRF5, IRF7 through different protein complexes and enzymes starting with the TRAM and TRIF complex formation. In this pathway, NF-κB complex activation is a result of the activation of two merging pathways activated by TRIF, the TRAF6 pathways or Receptor-interacting serine/threonine-protein I (RIPs) pathway[80]. IRF family transcription factors activation, leading to the production of type I interferons including IFNβ, occurs when TRAF3 is activated, mediating the

complex formation and activation of TANK-binding kinase 1 (TBK1) and IKKε which induces IRF activation[81].

These molecular changes resulting from TLR activation elicit a cellular effect, activating innate immune cells. For example, macrophages and dendritic cells, exhibit enhanced phagocytosis of bacteria and infected cells and secretion of pro-inflammatory cytokines and chemokines. While cytokine secretion has an effect on local cells, the secretion of chemokines recruits and activates additional immune cells, of both the innate and the adaptive immune system[82]. Particularly, promoting the proliferation and activation of adaptive immune cells lacking PRRs or specifically TLRs such as naïve T-cells is an important step in the defence against infections. In addition to T-cells there are other cell types of the adaptive immune system, B-cells, which express PRRs including TLRs. These cells do not require external activation and upon ligand recognition they undergo differentiation into mature, antibody producing cells while secreting co-stimulatory molecules[83]. Therefore, TLR activation is a key step in engaging the innate and adaptive immune response to PRRs, and as such a key player in inflammation and disease.

1.2.3 TLR4 and inflammatory disease signalling

The signalling processes mediated by the TLRs are implicated in the initiation and progression of many human diseases[84]. This includes incurable syndromes such as gram-negative septic shock, severe sepsis and viral haemorrhagic fever following an infection but also sterile diseases such as allergic airway diseases, rheumatoid arthritis, atherosclerosis, and the neuroinflammation associated with neurodegenerative diseases such as Alzheimer's and Parkinson's. The infection-associated diseases such as EVD have a common rapid onset of intra-vascular coagulation due to disruption of the vascular endothelium, which if untreated results in multi-organ failure and death[18]. One such cause of infection associated inflammation is via LPS associated TLR4-signalling, which can lead to severe inflammation and subsequent sepsis[85]. Following the heterodimer TLR4-MD2 formation, a second TLR4-MD2 complex binds, leading to a tetramer formation. Only this resulting complex is able to induce signal transduction and cytokine upregulation and release[86]. It is believed that the co-receptor cluster of differentiation 14 (CD14) presents LPS to MD2 and enhances the binding of TLR4/MD2 complex and ligands resulting in complex activation and downstream signalling (Figure 1.8), although signalling without CD14 can also be observed [87]. The mechanism of TLR4 activation has been thought to be binary with either an 'on' or 'off' state. However, despite this all-or-nothing response, the levels of intracellular downstream signalling activity can range from mild to severe. This phenomenon is caused

by MyDDosomes, intricate complexes of different proteins, downstream of TLR4, which are formed upon receptor activation (see section 1.2.1). Here TLR4 initiates nucleating signalling, initiating MyDDosomes assembly, which is independent of receptor activation after nucleation. The pace of this MyDDosome scaffold assembly, the number and the complex size may vary and has been suggested to regulate the extend of the immune response to microbial molecules e.g., LPS which triggers strong immune reactions and indicates bacterial infection[88].

The activation of TLR4 by LPS is designed to be almost instantaneous in order to counteract the fast replication and spread of bacterial infections. At a normal level of TLR4 activation, the release of pro-inflammatory cytokines is a beneficial protection mechanism. However, overstimulation of TLR4, in diseases such as EVD, Alzheimer's, cancer and arthritis leads to incorrect response levels and a damaging excessive inflammatory response[89]. This "cytokine storm" damages the surrounding tissue and can be detrimental to the host if left unchecked. Although TLR4 has been reported to induce anti-viral immunity via the TRIF-IRF pathway, it may still occur that the pathogens are not specifically targeted and can replicate unhindered and infect further cells[90]. As the infection reaches an acute level, the rate of TLR4 activation is rapidly increased and the dysregulated pro-inflammatory cytokine and other damage associated protein activation is capable of causing more damage to the host than the original infection.

1.2.4 TLR4 in infectious diseases

In addition to TLR4's ability to recognise bacterial LPS, the receptor is also activated either directly or indirectly by non-LPS ligands, with distinct structures such as DAMPs and other fungal, parasitic and viral PAMPs such as the EBOV GP (Figure 1.9). These protein ligands (either endogenous or pathogen derived) are capable of activating TLR4 and are a common feature in disease pathogenesis and more often a main contributor, however none have been proven to bind TLR4 directly to date. One example of a protein DAMP that activates TLR4, is high mobility group box 1 (HMGB1) protein. This architectural transcription factor and alarmin is passively released upon cell injury or death and actively released following excessive cell stress[91]. It has been implicated in chronic sterile inflammation and diseases such as arthritis, where its release causes constant activation of TLR4 and pro-inflammatory cytokine release[92]. An example of a parasitic PAMP capable of activating TLR4 is glycoinositolphospholipid (GIPL) from *Trypanosoma cruzi*, the tropical parasite causing Chagas disease or American

trypanosomiasis which most likely interacts with the hydrophobic MD2 barrel [93]. In regard to fungal activation of TLR4, *Candida albicans* is perhaps most noteworthy with its small-secreted cysteine-rich protein (Sel1). This yeast is a common member of the human gut flora but also the most prevalent cause of fungal infections in humans due to its opportunistic pathogenic nature when it invades the gut epithelium barrier and enters the bloodstream. Here the pathogen encounters nitrogen limitation and serum, a hallmark of human bloodstreams which lead to the upregulation of Sel1[94].

While the list of viral proteins that induce an inflammatory response through TLR4 is progressively expanding, the most studied include the dengue virus (DENV) non-structural protein 1 (NS1); RSV fusion protein (F); the vesicular stomatitis virus glycoprotein (VSV G) and the EBOV glycoprotein GP, albeit for the latter, more research is required due to the highly complex and irregular nature of this activation [90 95 96 20]. Thus, it is clear that TLR4 can be activated by many ligands from a variety of pathogens, and while it can provide a protective mechanism, subversion of this can cause disease. One of the clearest instances of this is the interaction of TLR4 and EBOV PAMPs, where an evolutionary protective mechanism against LPS can be lethal if falsely activated, as is the case for EVD.

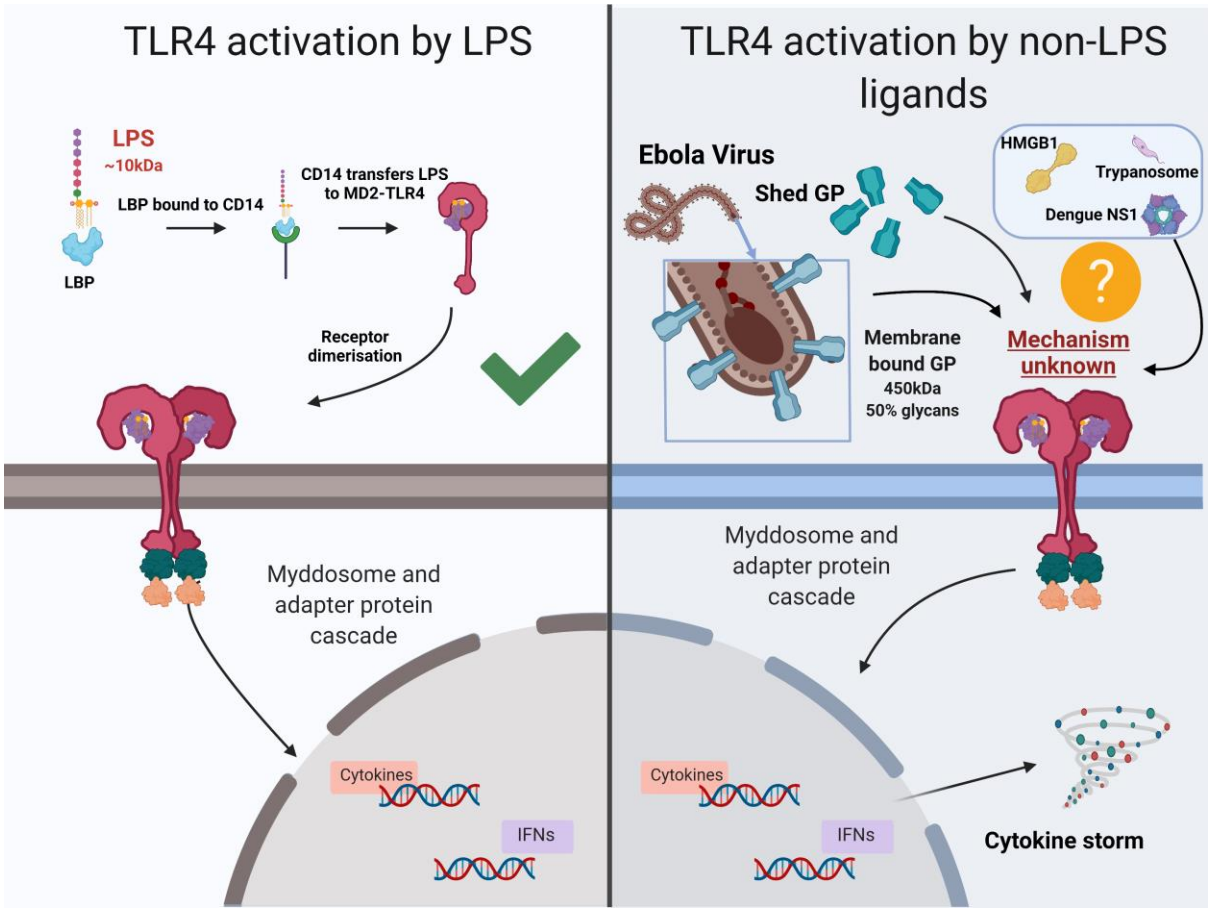


Figure 1.: Mechanism of TLR4 activation by LPS and non-LPS ligands. The mechanism of LPS recognition upon a bacterial infection starts with capturing of LPS by LBP. LBP interacts with CD14 leading to the transfer of LPS to the MD2-TLR4 complex. Undergoing a conformational change upon binding the complex dimerises and the signalling cascade is initiated (left). The dynamics of non-LPS TLR4 ligands capturing, including PAMPs such as EBOV GP and DAMPs such as HMGB1 as well as the mechanism of activation and the related cytokine storm is unknown to date. CD14, Cluster of differentiation 14; DAMPs, Danger associated molecular patterns; GP, Glycoprotein of the Ebola virus; HMGB1, High-mobility group box protein 1; IFNs, interferons; LBP, Lipid binding protein; LPS, Lipopolysaccharide; MD2, myeloid differentiation protein 2; NS1, non-structural protein 1; TLR, Toll-like receptor.

1.2.5 Ebola virus glycoprotein and Toll-like receptor 4 activation

Following investigations of history's largest EBOV epidemic in 2014-2016 (Western Africa), TLR4 is proposed to be one of the pathogen recognition receptors that is activated and could potentially bind the secretory viral protein, EBOV-shed GP [97] [98]. In turn, TLR4 was thought to be a key driver of severe inflammation in viral haemorrhagic fever due to sepsis-like disease symptoms and excessive activation of cytokines and signalling factors [20]. The contribution of TLR4 to EVD fatality was subsequently supported by mouse studies, where pre-treatment or daily dosing with TLR4 antagonists including eritoran and TAK242 or anti-TLR4 antibodies increased EVD survival of up to 70% [7]. Although initially the mechanism of EBOV activation of TLR4 was unclear, the viral glycoprotein GP1,2, a heavily glycosylated transmembrane protein, has been identified as a driving force[21]. To date, GP is the only Ebola virus protein found to facilitate TLR4 signalling, and although much remains to be uncovered regarding the specifics of this mechanism, the current understanding will be next discussed.

GP can engage with the cell surface pattern-recognition receptor TLR4, which usually responds to bacterial LPS. This recognition is extremely sensitive as a bacterial infection needs to be counteracted imminently due to their rapid exponential growth and adaptability[85]. In the case of an EBOV infection a TLR4 immune response exists through the TRIF-IRF pathway, however due to the viral immune evasion, such as endosomal escape, glycan shielding and anti-viral downregulation, EBOV can mostly propagate and replicate unhindered, thus further activating TLR4, leading to hyperactivation and the cytokine storm[99] [7]. If this TLR4 hyperactivation would be restricted to infected cell types and their neighbouring cells, then perhaps EBOV infection could be less severe and treatment more effective. However, shed GP, in addition to being an immune decoy, diffuses in the surrounding tissues and blood vessels, thus initiating widespread TLR4 activation far from the site of infection[21]. This systemic TLR4-induced inflammatory response, causing activation of non-infected dendritic cell and macrophages, is the main reason for disease severity and the high mortality rate[21].

In addition to the deregulation of the immune system, the EBOV GP expression leads to the disruption of host's surface proteins and cell rounding[100]. This is supported by both *in vivo* and cell overexpression studies, where overexpression of GP has been shown to affect the expression of host surface proteins involved in cell adhesion, including α and β integrins, and immune surveillance such as MHC class I

and II[101]. Moreover, one of the main factors contributing to the pathology of EBOV infection besides expression in immune cells, is the expression of GP in endothelial cells[18]. Here the disruption of surface adhesion proteins leads to the opening of tight junctions between the individual endothelial cells. As the wall of blood vessel is mostly comprised of these cells, vascular leakage occurs which evolves further to haemorrhaging and internal bleeding[13]. This leads to the typical symptoms of a severe EBOV infection and EVD progression such as bleeding from the mucosal membranes including eyes and mouth and organ failure [102].

Deletion of the heavily glycosylated mucin-like domain or enzymatic de-glycosylation abolishes most of the cytotoxic and cytopathological effects of GP, highlighting the impact this protein has on disease and severity, and consequently the opportunity it provides to both validate the biochemistry of pathology and uncover novel drug targets[103]. This activation can be reduced by pre-treating the cells with anti-TLR4 antibodies, supporting the initial mouse studies, or deglycosylation of the shed GP1,2. Additionally, treatment with mannose-binding sera lectin (MBL), which sequesters GP, also prevents TLR4 activation[20]. Taken together, these studies support the theory that EBOV GP activates TLR4 signalling. However, the mechanism is more complex than a simple GP-TLR4 complex as the GP of Reston Ebola virus (RESTB), an Ebola virus non-pathogenic to humans does not elicit a strong immune response in animals or in macrophage cell cultures [104]. The exact mechanism by which glycosylation of GP has such an effect on TLR4 activation and the host's body in general, despite its vastly different structure to the native ligand LPS remains to investigated and is the major topic of this body of work.

In this section I have provided an overview of the relevant literature and highlighted that EBOV GP is the key protein to combat EVD. Although several therapeutic strategies are available, the interaction with the TLR4 signalling complex, particularly the role of glycosylation has not been explored and represents a potential avenue for future therapeutic mechanisms of activation (MoAs).

1.3 Aims and rationale

The Glycoprotein of the Ebola Virus is one of the main contributors to the severe symptoms of the Ebola Virus Disease[21]. During the infection, GP immobilised at the virion membrane and also shed GP activates TLR4 at the site of infection and throughout the body. Although the importance of TLR4 in EVD is clear, the precise biochemical mechanisms by which EBOV GP produces the inflammatory signalling and the fatal "cytokine storm" are unknown and present an unmet need in the field[41 99]. Understanding this activation mechanism would allow identification of vulnerabilities in viral infection and also expand our available strategies of treating it. The introduction presents a literature review on EBOV GP and TLR4, followed by an overview of the known strategies to express and purify proteins and how they may apply to GP. The methodologies and experiments are presented in Chapter 3 to 5 and the final discussion in Chapter 6.

This dissertation investigates the activation of the TLR4 signalling complex by the Ebola Virus Glycoprotein. The study was primarily focused on the glycosylation of GP and its role in the activation of TLR4, based on previous findings suggesting that glycan removal from GP can significantly reduce activation of TLR4[20]. The final aim of this investigation is to identify specific targets for therapeutic possibilities and to generally broaden the knowledge of the role of GP. Additionally, discoveries and protocols developed in this research may be utilised to gain further insight in other diseases such as Dengue and the involvement of TLR4.

2 Experimental strategies to express and purify EBOV GP

The production of native and fully functional EBOV GP with high yields in academic settings has proven unsuccessful for many years. In the study by Bradfute et al., production and purification of EBOV GP in insect and mammalian expression systems was specifically and unsuccessfully addressed[105]. Here, they were unable to express unmodified full-length GP from the Zaire EBOV serotype in sufficient quantities, the most pathogenic serotype responsible for the epidemics in the past decades [105].

This Chapter explores the existing strategies to optimise protein expression and possible approaches towards purifying a complex protein such as EBOV GP are summarised. The applicability of each method for EBOV GP expression and purification is reviewed and the technique chosen for future experiments towards expressing and purifying native, glycosylated GP in large quantities.

2.1 Cytotoxicity of EBOV GP

The expression of ZEBOV GP induces dramatic morphological changes including cell rounding and detachment in cells over-expressing GP. This phenomenon is called GP-mediated cytopathology or cytotoxicity and presents a challenge to overcome when studying this protein[100]. By nature, this protein is cytotoxic to the cells expressing it, a feature contributing to the fatality of disease *in vivo* and complicating expressing large quantities of the GP for experimental use. The ability to cause adhesion loss and membrane protrusions has been linked to the extensively glycosylated mucin-like domain in GP1 and the GXXXA motif within the transmembrane domain of GP2 to date but other undiscovered motifs are most likely involved[106 107].

A modification or deletion of these domains could circumvent the problem, however if not executed correctly this alteration could impact other GP features such as stability and glycosylation that are essential for function. The ability to impact the host and immune receptors in the context of human infection, the downregulation of membrane receptors associated with the cytopathology and cytotoxicity of GP does not only enable the virus to evade the immune system but also also the disruption of tight junctions and spreading through the blood[106]. Generally, when expressing EBOV GP in mammalian over-expression systems, special safety precautions must be taken to avoid contact with the protein as much as possible due to the cytotoxic and highly immunogenic nature mentioned above, especially in the large quantities resulting from over-expression.

2.2 Choice of expression system

To study the interaction of the Ebola Virus Glycoprotein with the immune receptor TLR4, the expression of both proteins and associated co-receptors is required. The choice of the optimal expression system varies from protein to protein and can influence subsequent experiments. From the available protein expression cell systems: bacterial, insect or mammalian, the latter was chosen due to several reasons. Firstly, mammalian cells offer a complex arsenal of chaperones, glycosylation machinery, stabilising proteins, and disulphide isomerases for production of secreted viral proteins which are not necessarily present to the same extent in other expression systems and are required for expression of 'native' and functional EBOV GP[108]. Secondly, given that the glycosylation of EBOV GP may be crucial for activation of the TLR4 signalling complex, an expression system that fulfils that requirement is fundamental [20]. The expression system influences the complexity and composition of glycans of GP, with baculovirus infected insect cells yielding simple and less branched sugars, while mammalian expression retains full complexity and sialyation of N- and O-linked glycosylation. Thirdly, bacterial membranes contain LPS, a ligand of the TLR4 complex, capable of activating the innate immune system in picogram quantities[63]. As such, bacterial cell expression systems would present a source of LPS that would contaminate and confound studies which use mammalian signalling assays to assess EBOV GP activity. Lastly, the expression system is recommended be the same as the signalling system to avoid background signalling to foreign proteins and possible activation or steric inhibition of receptors with molecules including lipids[109]. As such it was imperative that a mammalian cell expression system be optimised to express GP and for subsequent experiments within this thesis.

The fully processed spike trimeric GP has a molecular weight of 450 kDa of which half are glycans. As glycans are vital for active glycoproteins these differences may have a substantial impact on the activity, stability, and function of the glycoprotein. The key differences lie in the complexity of N-linked glycans produced by the mammalian system and the often decreased or even missing O-linked glycosylation in insect cell produced recombinant proteins. In the case of EBOV GP, the mammalian derived proteins should be able to demonstrate high complexity glycans with noticeable sialyation over the simple and high mannose glycans from insect cell derived protein[110].

Although generally the recombinant protein yield of mammalian system expression is significantly lower than bacterial or insect cell expression, when the protein under investigation is as large and complex as the secreted glycoprotein of the Ebola virus, the yield of functional and correctly folded/processed protein is relatively higher. Due to the arguments presented here the mammalian expression system was selected. Although, there has been success of using a yeast-based expression system to produce mammalian proteins for structural biology, this method was not taken into

consideration due to the available facilities and complications associated with issues in protein folding, glycosylation and other translational modifications of complex mammalian or viral glycoproteins, which includes EBOV GP [111]. In order to address the low yield proteins such as EBOV GP expression in mammalian cells, several approaches are commonly used, which will be discussed below.

2.3 Codon optimisation and mRNA secondary structure

The sequence encompassing the start codon in the mRNA influences the recognition by the ribosome. Choosing the optimal sequence ensures the correct codon to anti-codon recognition and enhances the rate of translation. Additionally, strong secondary mRNA structures act sterically as obstacles and reduce the efficiency of translation. It is advantageous to reduce if not fully remove secondary structures, especially close to the 5' UTR of the mRNA strand, to exploit the full translation potential and increase protein yield. Additionally, commonly used codons are capable of accelerating elongation, while rare codons impede translation. This codon bias can be utilised to improve the protein expression significantly. However, precautions must be taken as translation elongation rates have co-evolved to assist co-translational folding of proteins and carefree changes can affect functional protein structure or may add undesirable post-mutations in the resulting peptide sequence due to frame shifts[112].

2.4 Choosing the correct vector for protein expression

The choice of expression vector also influences the efficiency of recombinant protein translation and production. Depending on the host expression system, specific sequences within the vector and/or vector copy number can increase or stabilise protein production and influence the plasmid stability within host cells. The correct vector must be chosen to fit the host and regulated so as to not excessively burden the host with a high number of copies, leading to cell stress responses or loss of the plasmid during cell division, thus decreasing the protein production. Secondly, selecting the correct promotor for the recombinant protein is important. Usually, a strong promotor such as CMV, leading to recombinant protein over-expression is preferred. However, in some instances, such as the expression of proteins involved in the regulation of the immune system e.g., TLR4, over-expression can be detrimental to the host cell, leading to cell stress and potential cell death. As such a choice of a 'weak' promotor is recommended[113].

An example of a plasmid with a high copy number and a strong promotor is pcDNA 3.1, one of the most common plasmids for protein expression in mammalian cells. When working on recombinant

proteins with cytotoxic or cytopathologic properties such as EBOV GP, the over-expression requires special fine-tuning to ensure optimal balance between protein yield and cell death[114]. In addition to the effect of promotor choice on protein expression, in recent years the plasmid backbone has been shown to affect the expression levels of the target protein. Bradfute et al. discovered that while keeping the recombinant protein, the tags, promotors, and antibiotic selections markers identical, and only changing the vector backbone, the exact same proteins, in this case filovirus glycoproteins can have vastly varying levels of expression[115].

2.5 Secretory leader peptide

When choosing mammalian or insect cell over bacterial expression systems, applicable to the research on EBOV GP, harvesting the recombinant protein from the cell medium instead of cell pellets is recommended as these expression systems require a significantly lower cell density to grow and express efficiently. To ensure secretion of the target protein, a signalling peptide, an N-terminal amino acid sequence (5-30 residues) is added to the recombinant protein. This sequence is recognised by the signal recognition particle causing ER trafficking and subsequent secretion. Selecting the correct signalling peptide can increase protein expression yield several-fold, varying from protein-to-protein depending on size and complexity.

While some signalling peptides e.g., from immunoglobulins, are broadly effective, others are only narrowly applicable, and some proteins require specific sequences. An engineered or foreign SP can illustrate superior features over the native signalling peptide. Finding the correct candidate may require empirical trials if the target protein is known to be difficult to express due to its size or complexity, a challenge which EBOV GP presents to the highest degree[116]. To further guarantee accumulation of recombinant protein in the cell medium, anchoring domains such as transmembrane domains, where applicable and non-essential for stability or protein function, should be deleted to avoid capturing the protein in the membrane.

2.6 Fusion tags

Besides expressing the protein of interest directly there is the well-established technique of generating recombinant fusion proteins. Here the target protein is fused with another protein domain or full-length protein at the C- or N-terminus[117]. These joined molecules are known as ‘tags’ and commonly linked to the target protein via additional amino acid linker to reduce steric interference and improve accessibility. There are two main purposes to design recombinant fusion proteins. Firstly,

vastly improved recovery and purification and secondly, to improve protein expression, stability, and solubility[118]. For the former, the most popular fusion tags are the poly-histidine tag (His-tag) and the Streptavidin specific Strep-tag due to cost, efficiency and specificity, while the Maltose-binding protein (MBP) tag, human serum albumin or constant antibody domains are used to improve yield and protein stability[119].

Occasionally, the C- or N-terminal location of a fusion tag, may differentially change the effect on solubility and purification[120]. The glycoprotein of EBOV does requires fusion tags for the purification and protein detection but not in order to enhance solubility or stability as the protein was evolutionary designed to be shed from the virion surface, diffuse to the bloodstream, and reach remote targets. In regard to the correct choice of fusion tags for the optimal expression and purification of GP, tag compatibility with secreted proteins and a small tag size to reduce steric hindrance during GP processing and trimeric spike formation are the most essential criteria.

2.7 Mutations to increase protein yield

Deletion of specific residues has also been shown to impact stability and the yield or recombinant proteins[121]. Post-translational modifications such as glycosylation and phosphorylation introduce further steps required for a fully processed and active protein. If the over-expressed protein has many modifications, such as Ebola viral glycoproteins that are extensively glycosylated, the expression system can be burdened with the enhanced high expression and post-translation processing leading to a reduction in the final yield. To circumvent this, specific motifs in the gene of interest required for the recognition by processing enzymes can be modified; the culture medium can be supplemented with enzyme-specific inhibitory compounds to stop the addition of post-translational modifications or using a cell line deficient in processing enzyme activity. While targeting and mutating motifs within the gene under investigation has the advantage of avoiding off-target effects which might influence the expression system, such mutations may affect protein stability and function.

Secreted glycoproteins and transmembrane receptors in particular rely heavily on glycans for activity and mutating these carelessly can inhibit protein-protein interactions and/or cause a loss of function[122]. During an investigation to produce stable and homogeneous Zaire EBOV GP for structural analysis Eos et al. mutated individual glycosylation sites and deleted the mucin like domain (MLD) to reduce the heterogenicity and noise introduced by the extensive and variable glycans. Whereas the deletion of the MLD did not alter the overall protein stability, mutating individual glycosylation sites noticeably impacted and impaired degradation and expression, indicating that some of these are

essential for correct folding and processing of the mature GP[121]. The details of the findings will be discussed in Chapter 4 when illustrating the glycosylation of GP.

In addition to single amino acid changes, one can delete entire domains or structures to increase protein yield. This approach is commonly used in structural biology as large quantities of soluble, stable, and homogeneous protein are necessary to elucidate high-resolution structures. The main methods include the deletion of transmembrane or other hydrophobic domains to improve solubility and increase secreted protein amounts; deletion of unstructured domains and non-essential loops and the removal of larger processing or enzymatic cleavage sites such as glycan caps to establish independence from rate limitations during expression by reducing complexity and heterogenicity[121].

Such approaches have been used to stably produce and optimises yields of recombinant EBOV GP: mutating the editing site of GP genes to ensure exclusive expression of full length GP1,2 and avoid sGP and ssGP; deleting the transmembrane domain in GP2 to increase secretion of the glycoprotein and removing the Furin cleavage site removal connecting GP1 and GP2 for eased purification. However, when blocking Furin processing of GP, the structure of the heterodimer GP1,2 and the formation of the trimeric spike protein are altered and no longer relate to the native state[123]. This is something that needs to be avoided or rectified to truly understand GP interactions.

2.8 Media formulations

For protein expression in mammalian cells such as human embryonic kidney (HEK) 293 cells, routinely, Dulbecco's modified Eagle's media is chosen with additional supplements such as > 4.5 g/L glycose, sodium pyruvate and 5-10% fetal bovine serum (FBS). To further enhance protein expression by up to 50 %, the media can be supplemented with non-essential small molecules such as sodium butyrate [124]. When expressing secreted proteins which may interact with molecules within the serum or for cell lines prone to intracellular protein aggregation, special formulated serum-free media can maximise protein recovery and cell viability. Additionally, cell lines can be supplemented with an array of antibiotics to negate possible bacterial contaminations, commonly penicillin and streptomycin or to select a stable cell line for cells carrying the expression plasmid with an incorporated, antibiotic-specific resistance marker[125]. Modification of the media may be beneficial for the expression and stability of GP influencing cytotoxicity and the ability to efficiently purify shed GP.

2.9 Comparing Stable, transient expression and inducible protein expression

In transient expression, a vector containing the gene of interest is introduced to the cells chemically or via electroporation for a limited period of time and not integrated into the genome, thus circumventing the time-consuming selection of a clonal cell line. The main advantage of this method is the short time required from cloning to protein expression. However, transient expression has its limitations pertaining to the underlying biology of mammalian tissue culture, such as inter-cell line differences in transfection efficiency and cytotoxic effects of transfection reagents, both of which limit recombinant protein yield[126]. Additionally, while the cells divide and grow the plasmid number stays constant, and eventually the cells expressing the recombinant protein are outcompeted by 'non-transfected' cells for nutrients and growth, leading to a dilution effect. Therefore, the protein collection is significantly time limited due to cell stress, overgrowth and this dilution effect. Further, the costs of consumables e.g., transfection reagents and DNA vectors accumulate and can burden a laboratory[127]. Considering these limitations, when large quantities of one homogeneous protein are required for investigations such as crystallisation and binding studies, the generation of a stable cell line is superior.

Stable mammalian cell expression can be driven by integrating the gene of interest into the host genome. As integration is random, the position within the host's genome can influence the expression levels, not to mention possibly disrupt endogenous gene transcription. To ensure the highest yield and protein quality, several selected clones need to be screened. Alternatively, the stable cell lines can be generated by autonomous replication and nuclear retention through viral elements provided by viruses, often adenoviruses such as Epstein-Barr virus. Each virus has different copy numbers which influence the transgene amplification and can lead to an under-production of the target protein or overloading of the cells, depending on infection efficiency[128]. Thus, expression systems need to be optimised carefully, especially when attempting to express a cytotoxic protein such as GP. Furthermore, when investigating cytotoxic proteins, cell stress and death limits the efficiency of stable cells growth and recombinant protein yield. This can be avoided by using an inducible promotor. One option to overcome this problem would be expressing the EBOV GP via an inducible stable cell line; however, this was not feasible within the facilities available.

2.10 Co-expression with processing enzymes

Another approach used to increase protein yield is to over-express proteins related to the enzymatic processing of proteins during production and maturation[129]. The spike glycoprotein of the Ebola virus is processed and undergoes several changes to mature from the monomeric pre-GP to the trimeric spike protein GP (as discussed in the introduction, section 1.1.4). Co-expression of Furin convertase to ensure efficient processing of pre-GP and the metalloprotease TACE to cleave membrane-associated GP to increase the secretion of shed GP would overall increase the yield of functional GP in the culture medium[123].

2.11 Choice of purification method

To study proteins after successful expression, the protein of interest must be accumulated and isolated via purification techniques. If the expressed protein of interest accumulates in the cytosol, as is the case for most proteins, or in the cell membrane when dealing with lipophilic transmembrane proteins, then the cells must be lysed through the use of detergents, enzymes, or physical methods such as sonication to receive a crude cell extract. This extract contains cell debris such as organelles or membrane residues that are commonly removed from the cell extract via centrifugation, leaving a supernatant containing mostly proteins. When collecting secreted proteins instead of using chemicals to lyse cells, only centrifugation is required to remove cells as the target protein accumulated in the cell media.

When purifying proteins, the final use of the target protein (e.g., use in therapy) will dictate the order and type of purification steps affecting yield and purity. As each purification step contributes to loss of final protein yield, the purification strategy must be carefully chosen to fit the target protein's properties including size, charge, and solubility. Conceptually, each purification step decreases the yield by around 20% but is capable of raising the purity up to >99% [130]. Two-step purification methods are usually utilised as they offer an optimal compromise between yield and purity. Commonly, affinity chromatography is employed as the first step which captures tags fused to the protein of interest, followed by a second purification step, often size-exclusion chromatography (SEC). SEC separates smaller from larger molecules through pores in a polymer matrix column or Ion-exchange chromatography (IEC) which which protein separation by charge using either positive (anion exchange) or negative (cation exchange) beads to attract functional groups of the protein with opposite charges.

Affinity chromatography allows purification of most proteins without prior knowledge of composition or properties, allowing cost effective and high throughput purification in native conditions. The first commercially available tags were inspired by naturally occurring epitopes, a short antigenic amino acid sequence which is specifically recognised by antibodies[130]. Conventionally, the earliest epitope tags including human c-myc proto-oncogene product (EQKLISEEDL), the human influenza virus hemagglutinin protein (YPYDVPDYA), and the hydrophilic FLAG epitope (DYKDDDDK) have been primarily used in protein visualisation due to the high specificity and virtually non-existent cross reactivity[131 132 133]. With the antibody production and purification methods steadily improving these naturally occurring epitope tags have become a viable option for protein purification due to reduction in cost and yield improvements.

Purification of the tagged recombinant protein can be achieved by using immobilised monoclonal antibody beads e.g., Sepharose beads, under non-denaturing and native conditions, and straightforward elution by change in pH or chelating agents to reduce binding between antibody and epitope antigen. However, antibody-based purification matrices require special storage conditions and a protease free environment due to their considerably lower stability compared to metal ion associated columns or matrices[134].

Over the past decade artificially engineered tags have been developed and optimised to allow improved purification under sub-optimal conditions e.g., inclusion bodies or protein membrane localisation. The first tag of this type was developed by Sassenfeld and Brewer, using six consecutive arginines and captured using a cationic resin. Succeeding the polyArg tag, as the most popular poly-amino acid tag is the polyhistidine (polyHis) tag due to its advantages and uses[135]. PolyHis tags can be fused on both the C and N-terminus of the target protein and purification can be conducted independently on denaturing or native conditions. In addition, due to the small size, low structural interference, and high purification yield, the His-tag is vastly popular in biophysical and structural studies. To purify polyHis fusion proteins immobilized metal-affinity chromatography (IMAC) is performed to capture the metal binding histidine peptides using positively charged metal ions predominantly Nickel (Ni2+), but also Cobalt (Co2+), Copper (Cu2+), or Zinc (Zn2+)[136]. Due to these features, the polyHis-tag dominates the single-step purification approaches and the field of biophysics with over 60% of all submitted structures containing a poly-His tag [137]. The only disadvantages are possible contamination of the purified protein with metal ions, which may trigger cell signalling via metal sensing receptors or the possibility to co-purify cellular proteins which naturally contain high amounts of histidine or other negatively charged groups, which may be solved with metal-chelating agents and buffer optimisation.

If the use of metal ions during the purification process must be avoided if the protein is to be used for therapeutic use or itself has metal ion binding capabilities, then the use of the Strep-tag (WRHPQFGG), a peptide that was initially selected due to its affinity to the streptavidin core, a bacterial protein and then later improved to Strep-tag II (WSHPQFEK) is a recommended the choice [138]. If the target protein requires additional stabilisation and improved solubility larger fusion tags may be used such as Maltose-binding protein (MBP), a protein originating from the malE gene of Escherichia coli K12 [139]. The disadvantage of an MBP-fusion protein and other expression- and solubility- enhancing protein tags, is that due to the substantial size of 42kDa of the tag, it may interfere with activity and binding capabilities of the protein of interest and has to be removed after purification which may revert the solubility benefit of the MBP-tag and lead to protein aggregation and precipitation[140].

As ZEBOV GP is natively soluble and stable, an enhancing tag is not required, while a small sized fusion tag would be preferred to avoid any steric interference during the complicated enzymatic processing, protein maturation and complex formation process. Notably, as the expression occurs within the mammalian expression system, the chosen tag must be compatible with that system and the associated complications including possible contamination by mammalian proteins.

2.12 Two-step purification

Unfortunately, single-step purification processes are not compatible with every intended purpose for the target protein and often two or more purification steps are required to assure purity at the cost of potential yield. One strategy would be to employ two affinity tags in tandem utilising tandem affinity purification (TAP) e.g., a poly-His tag and a Strep II tag fusion protein to increase target specificity[141]. Another strategy, and more routinely utilised is the use of IEC, to separate the single step purified protein by charge or gel filtration technique SEC, where proteins are separated by their size and conformation[142].

One of the advantages of using IEC or SEC to further purify proteins, in addition to increasing purity and removal of contaminants, is the removal of degradation products of the protein of interest resulting in one homogenous protein sample in a buffer of choice. These techniques work well due to fact that proteins absorb UV light at 280nm and complete purification systems like the AKTA FPLC combine all tools and measurements to accurately detect protein migration and elution to estimate purity and sample stability. IEC utilises the fact that each protein has an individual isoelectric point, a particular pH at which the protein holds a neutral net charge, where the number of positive and negative charges is equal. By altering buffer conditions during the purification process, to increase or reduce the positive or negative charges on the protein depending on pH, a homogenous final sample

can be achieved provided a prior primary purification with a protein solution of satisfactory purify, usually above 80%[143].

Gel filtration on the other hand is mostly independent of pH and is able to separate proteins by their size and shape. The pore sizes in the column matrix influences the purification where smaller proteins pass through the beads' pores delaying their migration through the column, while proteins larger than the pores bypass the beads and migrate unhindered until eluted. The resolution of the column, its ability to separate proteins accurately by their size, is not only affected by the pore sizes but also by the overall column volume and choice of beads, commonly used Sepharose beads with low protein binding capabilities. In addition to purifying proteins by their size, SEC can be used to confirm protein interaction and oligomeric state, e.g., if a sample of two interacting proteins is injected, then the protein complex will migrate and elute together visible by a single peak in the 280nm UV spectrum, while monomeric individual proteins would show separated elution points[142]. Occasionally, it is necessary to remove fusion tags after purification as they may interfere with target protein activity or affect the structure. A solution to this problem is the incorporation of short sequences between fusion tag and protein that can be recognised by site specific proteases, the most commonly used protease include tobacco etch virus (TEV) protease, SUMO protease and thrombin [144] [145] [146].

Each purification strategy has its specific range of suitable buffers and conditions that need to be optimised on a per protein basis. When using the most popular single-step purification methods of IMAC to capture polyHis-tagged proteins two main problems arise, firstly the fact that chelating agents such as EDTA often present in cell media may strip the column from capturing metal ions and that bovine serotransferase, a major component in fetal bovine serum binds to positively charged molecules[147]. In order to guarantee artifact-free results during this dissertation using most native state of EBOV GP as possible is a requirement, therefore addressing and optimising problems when purifying mammalian-derived proteins was a fundamental process of this dissertation.

Having reviewed and assessed the strategies to express and purify native GP in this Chapter, the expression trials and extensive optimisation steps required to fulfil that goal are discussed and presented in chapter 3.

3 Expressing and purifying the Ebola Virus glycoprotein in large quantities while maintaining glycosylation, stability, and functionality

3.1 Introduction

In this chapter the approaches taken during my PhD towards expressing native, active, and stable Ebola virus glycoprotein in large quantities are presented. Although soluble and heavily modified ectodomain of EBOV GP is commercially available, the alterations, particularly including the removal of glycosylation sites could possibly impact the endogenous behaviour of GP, which is of vital importance when trying to study the mechanism under physiological conditions [121].

For this, a comprehensive review of the possible strategies was undertaken, including the tailoring for EBOV GP and reasoning for choosing a particular methodology in the chapter 2. A schematic summary is presented in Figure 3.1. This process required extensive optimisation due to the novel approach of expressing active and unaltered EBOV GP. As outlined above this is a critical and unmet need as the activity of GP and effect it has on the immune system can only be studied with physiologically relevant protein. To guarantee artifact-free results during this dissertation using most native state of EBOV GP as possible is a requirement, therefore addressing and optimising problems when purifying mammalian-derived proteins was a fundamental process of this dissertation.

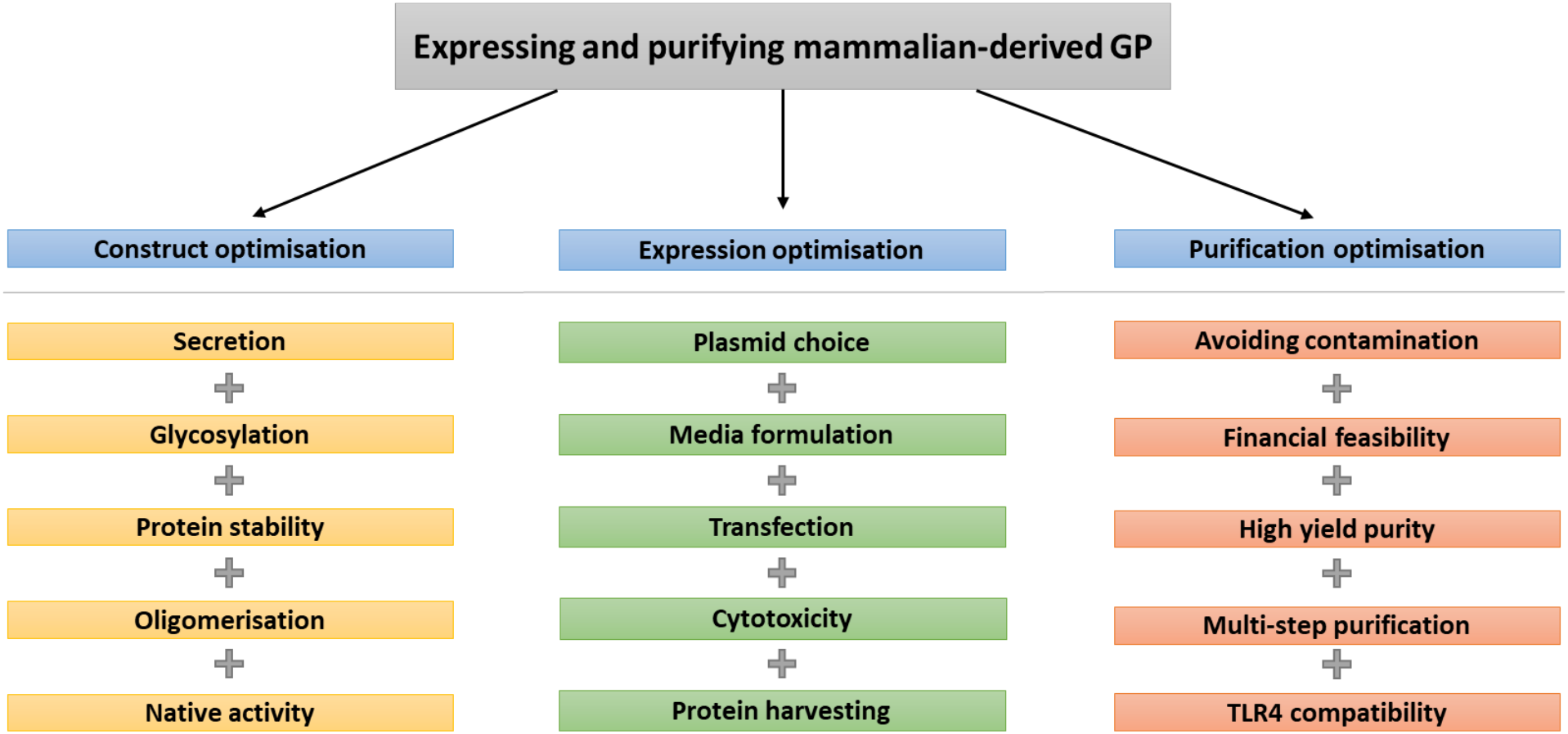


Figure 3.1. Optimisation steps to express and purify EBOV GP. The optimisation steps to express GP in mammalian cells and purify the protein with high yields were separated into three groups, constructs, expression and purification.

3.2 Results

3.2.1 Expressing EBOV GP in HEK293T cells

To investigate ZEBOV GP, the primary sequence was obtained (GenBank: AAB81004.1), and a synthetic gene for wild type full length GP1,2 Zaire Makona was designed and requested (synthesis performed externally; see methods for details). Next, to express the glycoprotein of EBOV the mammalian cell line human embryonic kidney (HEK) 293 T was chosen to ensure physiologically relevant expression conditions with maximal post translational processing including glycosylation and minimal LPS contaminations. The first full length wild-type GP construct (GPori) with a c-terminal HA-tag was expressed in HEK 293 T cells via transient transfection using jetPEI (according to manufacturers' instructions) to investigate the expression efficiency of GPori in pMA-RQ, with the signalling peptide native to GP and the cleavage at the ectodomain through natively expressed metalloprotease TACE.

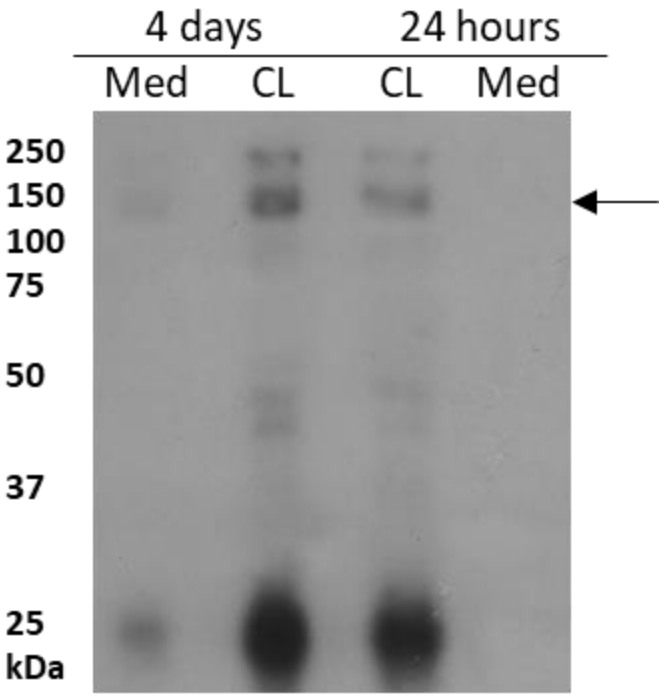


Figure 3.2: Expression of WT GPori in HEK293T cells with low secretion. HEK293T cells were transfected with GPori plasmid and both culture medium and cell lysate were collected 4 days post-transfection. The samples were run on a non-gradient, reducing 12% SDS-PAGE gel, followed by western blotting with anti-HA (mouse). Band at 25 kDa indicates reduced GP2 with C-terminal HA-tag. Arrow at 150 kDa indicates full-length GP1,2. Med: culture medium containing GP; CL: cell lysate containing GP.

The culture medium was harvested and expression of GPori was assessed via western blotting. From the initial expression trial, it became clear that full length GPori was successfully expressed, however insufficient quantities were detected in the cell medium, and the majority was located in the cell lysis

fraction (Figure 3.2). Glycosylation was assumed at this step due to the visible band at ~150 kDa indicating fully glycosylated GP, whereas GP of ~75 kDa would have indicated a lack of glycosylation. It was hypothesised that expression levels of native TACE are insufficient to allow adequate processing of overexpressed full length GP1,2 into shed GP.

3.2.2 Optimising GP secretion

Following this result, a new construct of GP (GP_ED) was designed lacking the transmembrane domain which ought to lead to artificial expression of shed GP independently of endogenous TACE activity. The deletion of GP TMD sequence (WRQWIPAGIGVTGVIIAVIALFCI) which included the hydrophobic helical domain of GP2, (Cholesterol binding motif underscored) lead to unexpected complications during the expression. Not only was the rate of expression of GP_ED significantly lower compared to full length GP but also an unexplained lack of glycosylation was evident (Figure 3.3). The combination of low expression and reduced glycosylation deemed this particular GP construct unsuitable for the investigation of the mechanism of interaction with the TLR4 signalling complex.

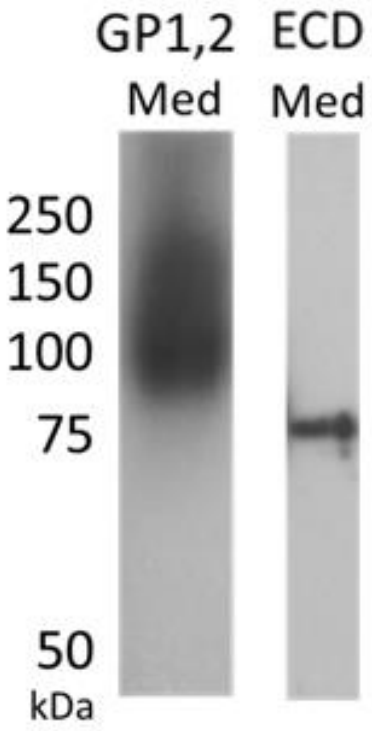


Figure 3.3: GP_ED expressed in HEK293T lacks glycosylation. HEK293T cells were transfected with GPori and GP_ED plasmids and the culture medium was collected 4 days post-transfection. The culture medium was cleared and 10x concentrated prior to running the samples on a non-reducing, gradient-less12% SDS-PAGE gel, followed by western blotting with anti-HA antibodies.. Med: Culture medium containing GP; ECD: ectodomain.

The deletion of GP TMD seemed to have a fundamental impact on the stability of GP during post-translational processing and possibly affecting glycosylation and maturation[148]. This investigation into

the role of GP TMD in the protein maturation process was suspended and will be discussed in section 3.2.3. As the full-length construct GPori yielded sufficient protein expression, two strategies were employed in parallel to address the low rate of TACE processing observed.

The first strategy was to co-express the target protein GPori and processing enzyme TACE with the intention to increase the amount of shed GP. Alternately to co-expression, the TACE construct was designed with an affinity-fusion tag to enable the separate protein expression, purification, and addition of TACE to culture medium of GP expressing cells. By adding TACE to the medium, rather relying on intracellular production, additional cell stress associated with multi-protein over-expression may be avoided if necessary (Figure 3.4 B).

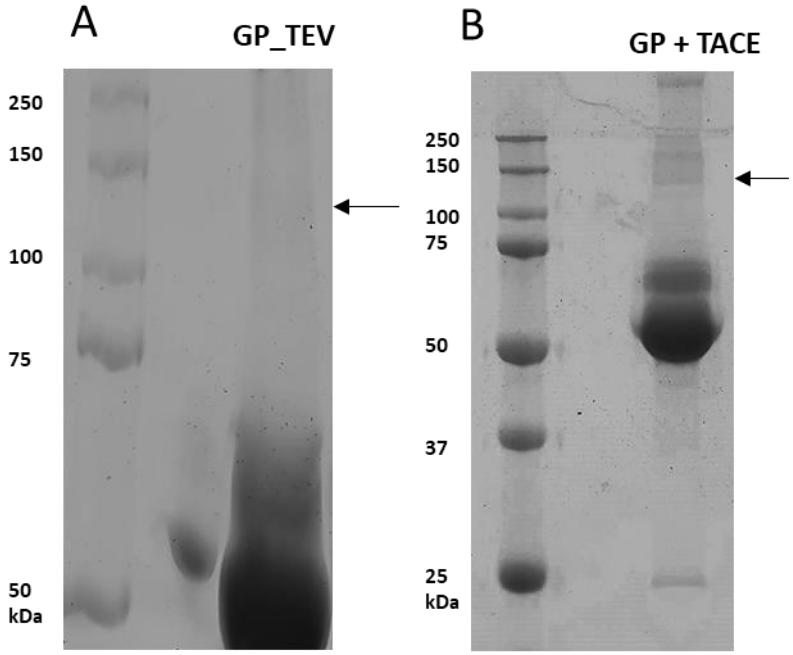


Figure 3.4: Low expression of GP_TEV and improved secretion by co-expressing of GPori with TACE. HEK293T cells were transfected with GP_TEV, GPori and TACE plasmids. For A.) Following the medium change 6 hours- post transfection, 20µg of purified TEV protease were added to the GP_TEV expressing culture at 24-hour intervals. For A.) and B.) The culture medium was collected 4 days post-transfection and cleared. The cleared culture supernatant containing the GP proteins was run on gradient 6-16% SDS-PAGE gels under non-reducing conditions and stained with Coomassie. Detailed methods are found in section 7. Arrow indicates the GP protein. Protein at 50-70 kDa is serum albumin.

The second approach was altering the TACE recognition sequence of GP. The sequence could either be mutated to guarantee the most optimal cutting site for the processing enzyme, or the sequence could be replaced by another targeting site for more efficient proteases. The latter strategy was chosen, as the peptide sequence of GP recognised by TACE is in a flexible and unstructured loop which has not been associated with any additional function to date. The protease recognition sequence of TEV was chosen to replace TACE due to its superior activity (due to optimal sequence choice), commercial availability, and familiarity within the research group. The effect of both approaches on

the expression of full-length GP and amount of GP present in the culture medium was assessed and compared (Figure 3.4). The co-expression of TACE together with GPori as well as the mutation and replacement of the TACE processing site by a TEV target sequence, lead to an increase in shed GP in the cell culture and available for purification (Figure 3.4). However, the GP construct containing a TEV sequence instead of TACE (GP_TEV) affected protein stability with increased protein degradation and an overall reduction in expression. It is possible that, despite the unstructured nature of the TACE loop, a fundamental mutation in that area of the GP may impact protein folding or interfere during processing in the cell.

To recapitulate, expressing full-length EBOV GP and relies on endogenous TACE for shedding by cleavage at the membrane. TACE is expressed in HEK293T cells but at insufficient levels to yield effective GP shedding. Co-transfection with TACE or addition to the cell medium post-transfection increases GP secretion in the medium, while replacing the TACE sequence with another protease recognition sequence destabilises expression and is ineffective. Next it was investigated how the TMD of GP can be mutated to increase GP secretion without affecting glycosylation or protein stability.

Table 3.1: List of GP mutants to investigate transmembrane domain effect on glycosylation and secretion. Taken the construct GP_ECD as a template the residues of the transmembrane domain were re-introduced. The sequence of each modification and the effect on glycosylation and secretion are listed sequentially. The construct GP_TM3 (grey), yielded both, glycosylation, and secretion, unlike established constructs (EOS_ECD_GP) and newly generated. ECD, ectodomain; EOS, construct used by [32] for the crystal structure generation; TM, transmembrane domain.

Construct Name	Modification	Glycosylation	Secretion
EOS_ECD_GP	Deletion of MLD, TMD, mutation of T42V, T230V glycosylation sites	No	Yes
GP_ECD	1-650 (WRQ last residues)	No	Yes
GP_TM3 (later GPΔTM)	1-653 (WIP)	Yes	Yes
GP_TM6	1-656 (WIPAGI)	Yes	No
GP_TM9	1-659 (WIPAGIGVT)	Yes	No
GP_TM15	1-665 (WIPAGIGVTGVIIA)	Yes	No

3.2.3 Optimising glycosylation of secreted GP

Viral surface proteins often have an oligomerisation domain close to or within the transmembrane. It is possible that complete removal of the TMD of GP may interfere with this process. Although successful expression of the ectodomain of GP, capable of forming the trimeric spike protein have been reported [32], the designed recombinant proteins were severely altered including the removal of unstructured domains including MLD, glycosylation sites, mutation of the Furin cleavage sequence and introduction of stabilising residues for the trimerization.

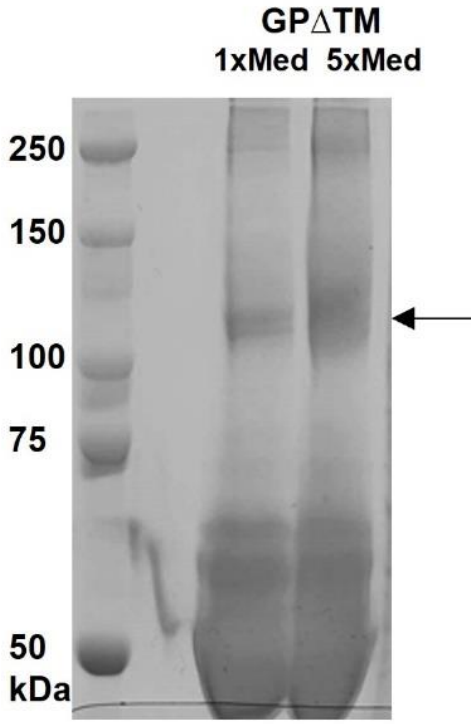


Figure 3.5: Introduction of first TM helix of GP reinstates glycosylation. HEK293T cells were transfected with GP_TM3 plasmid and the medium was collected, cleared and concentrated. A gradient 6-16% SDS-PAGE gel was run under non-reducing conditions and stained with Coomassie.. The complete protocol is listed in section 7. Arrow indicates presence of GP. 1xMed, cleared cultured medium containing GPΔTM and the level of concentration. GPΔTM, Original construct named GP_TM3 and selected over other transmembrane constructs.

New GP constructs were designed to address this phenomenon with step-wise re-introduction at the C-terminus of the TMD sequence three residues at a time. Particular attention was paid to the residues of the α-helix protruding from the membrane, WRQWIP, and a cholesterol-binding motif GXXXA [149] in the middle of the TMD (Table 3.1). All recombinant protein constructs containing this helix were successfully expressed and showed glycosylation and stability. To assess the successful secretion of GP, the medium was collected, and the protein purified as previously described, and glycosylation was assessed with deglycosylation enzymes. Additionally, the cells were lysed and WB performed to assess

the rate of secretion vs cytosolic protein. However, only the construct containing solely the first helix WRQWIP, referred to as GPΔTM henceforth, was secreted without the dependence of TACE, mimicking native shed GP, most likely due to balance between stability and hydrophobicity (Figure 3.5). Glycosylation of GPΔTM was confirmed by incubation with the deglycosylating enzyme, O-glycosidase and the SDS-PAGE migration was compared to untreated GP (Figure 3.6).

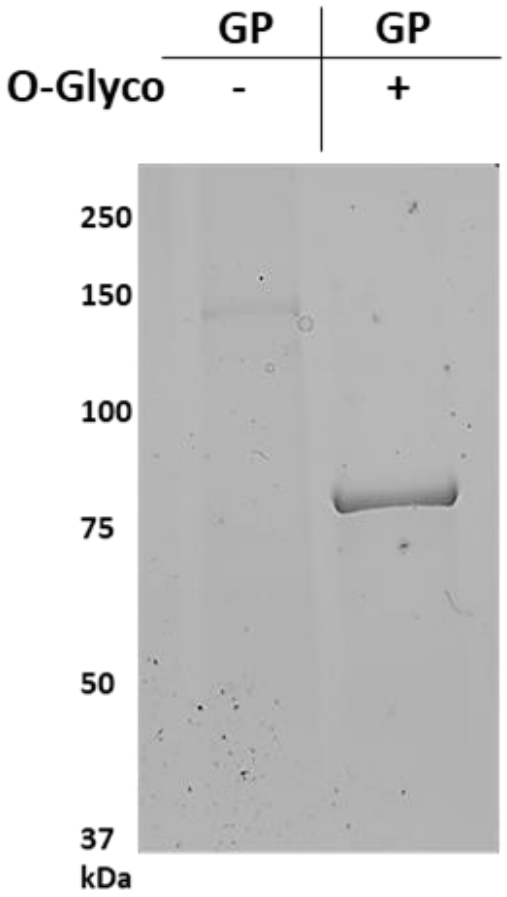


Figure 3.6: Glycosylation of GP confirmed by gel shift with O-glycosidase. GPΔTM was expressed in HEK293T cells and purified as described in section 7. The purified GP was treated with O-glycosidase to confirm accurate glycosylation. O-deglycosylation was performed as described in. The samples were run on a gradient 6-16% SDS-PAGE gel under non-reducing conditions and stained with Coomassie. The '+' and '-' indicate treated and untreated GP. Band migration visible from 145 kDa to 80 kDa. This gel is an example of the routine glycosylation and enzyme activity. In this particular case, the sample was analysed chronologically after optimisation of the purification method.

As the expression yield and stability of recombinant proteins can be influenced by the choice of plasmid, the compatibility of plasmids optimised for cellular secretion regarding the expression and secretion of EBOV GP was next addressed. Due to familiarity and availability within the department the secretory plasmids pHL-sec and pDisplay were chosen and their suitability compared to the construct GPΔTM in pcDNA 3.1. The rate of expression and secretion of GP was vastly different for

each plasmid, with GPΔTM in pDisplay displaying the highest rate of secretion; pcDNA 3.1 displaying the highest expression although not the highest secretion (see Figure 3.2), and lastly pHL-sec indicating no detectable expression. Following this result, pDisplay was chosen as the plasmid to conduct future expression studies. Although a recombinant EBOV GP with an HA-tag was expressed and purified for structural studies by EOS in the past, the construct GPΔTM in pDisplay with a C-terminal HA-tag was only suitable for immuno-detection but not large-scale purification in the laboratory setting available to me; the purification optimisation and methodology is presented and discussed in Chapter 2. Therefore, for large scale purification, a new GP construct was designed to be expressed in pDisplay. The construct consisted of endogenous signalling peptide of GP, followed by a Strep-II tag, and poly His-tag at the N-terminus, and lastly a TEV cleavage sequence (Figure 3.7).

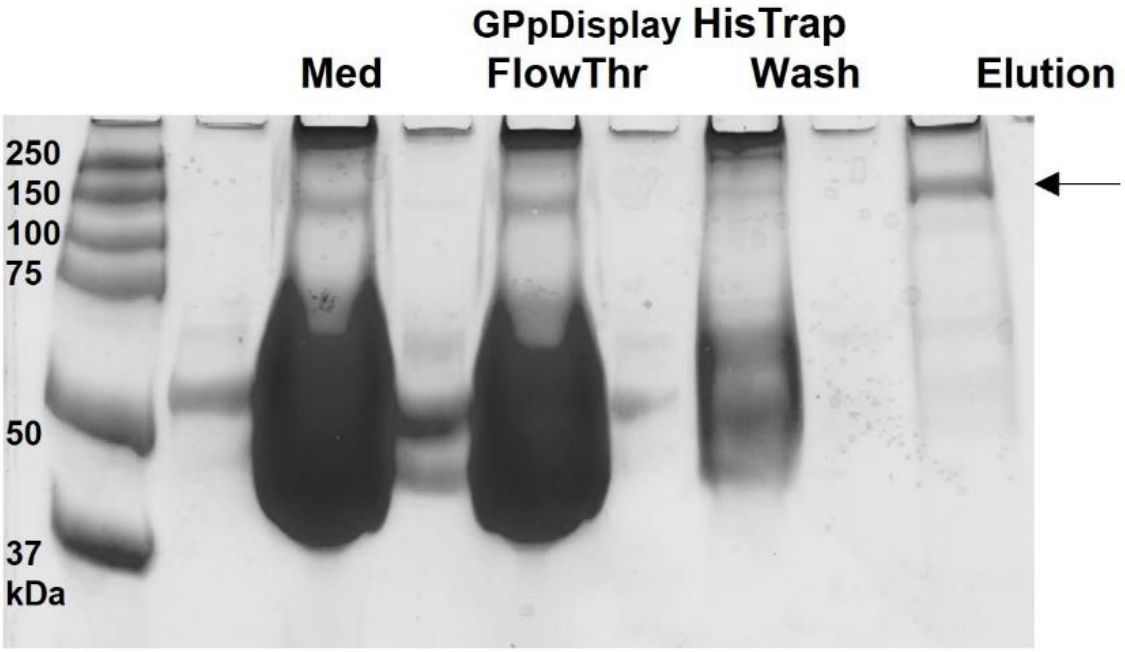


Figure 3.7: Expression of GPpDisplay and high purification with HisTrap. GPΔTM in pDisplay termed GPpDisplay was expressed in HEK293T cells, and the secreted protein was harvested. IMAC purification was performed with a HisTrap column. A gradient SDS-PAGE 6-16% was run under non-reducing conditions and stained with Coomassie. Detailed protocols are listed in section 7. Arrow indicates GP. Med, Culture medium sample containing GP; FlowThr, Flow through of HisTrap column from culture medium; Wash, washing fractions of HisTrap column.

This optimised GP construct lacking all of the transmembrane domain except the first helix WRQWIP, GPΔTM, produced the highest yield of fully glycosylated and stable protein capable of activating the TLR4 signalling complex. Thus, having optimised the expression and successfully purified GP with IMAC, the effects of GP's cytotoxicity and cytopathology were investigated to increase protein yield and improve protein homogeneity.

3.2.4 Addressing cytotoxicity and cytopathology

As shown by G. Simmons et al. the expression of full-length EBOV GP or only the MLD leads to cell rounding and detachment[100]. In the interest of obtaining the greatest quantity of GP the protein expression must be optimised and the balance between cytotoxicity and overexpression achieved. Expression of GP alone is capable of disrupting cell surface receptor expression and consequently cell adhesion within 24 hours. However, the underlying limitations of the mammalian expression system of choice do not favour harvesting within 24 hours, if substantial yields are to be obtained.

To investigate the balance between expression yield and cytotoxicity, next the optimal expression period of ZEBOV GP was explored in transiently transfected HEK293T cells to yield pure and stable protein. Theoretically, a short expression period usually correlates with low yield but homogenous protein while late harvesting leads to higher protein yield but increases the change of partial degradation or heterogenicity due to cell passage effects such as: overgrowth, change in pH, reduced expression, and cell stress. To address this challenge several rounds of optimisation including cell density, media composition, surface to medium ratio, and expression duration were conducted.

When over-expressing intra-cellular recombinant proteins in adherent mammalian cells, commonly cells are transfected at 70-80% confluency, following a media exchange 6 hours and harvesting 48 hours post-transfection. However, the conditions for optimal expression guaranteeing high yields and protein integrity are different when expressing secreted proteins due to the extended duration between transfection and sample harvesting.

The accumulation of the target protein in the cell culture medium requires longer periods of expression due to dependency on the secretory pathway. The extended culture period would lead to culture overgrowth associated with cell stress and protein degradation if the initial cell density remains the same as for intra-cellular proteins. To prevent this, the cells were seeded and transfected at less than half of the manufacturers' recommended density i.e., 2E6 cells seeded and 8E6 cells transfected for a culture on a 15 cm diameter petri-dish. Additionally, to counteract the relatively lower initial cell density at the time of transfection, the transfection media incubation was extended from 6 hours to 16 hours in addition to increasing the recommended PEI transfection agent to DNA ratio from 2:1 to 5:1 to ensure transfection during the initial cell growth period. Although longer incubation periods may result in higher transfection efficiency, the cytotoxic properties of PEI when used in larger quantities limit this use.

To further increase the potential yield of secreted protein, the total expression media volume was increased. To accommodate the maximum amount of medium per cm^2 of adherent cell culture, tissue culture dishes with larger volumetric capacity due to the taller walls and overall geometry were chosen instead of vented flasks. With this setup, it was possible to increase the available volume for protein secretion from the original 2ml to over 10 ml per million cells at the time of transfection. Attempts at using multilayer flask e.g. 5-layer 175cm^2 flask to increase the available media volume over conventional single layer flasks resulted in increased protein yield, however the associated cost-to-benefit ratio did not justify this method.

After the optimal expression conditions for GP were established, a temporal expression analysis was conducted to investigate the timepoint with the highest protein yield, without noticeable GP degradation or cell death induced by cell stress, released proteases and the cytotoxic effect of GP on human cells. For this, a daily time course up to 7 days was performed in HEK293T cells transfected with GPΔTM in pDisplay, referred to as GP, to elucidate optimal time of harvesting. As a result, it became evident that the protein yield increases over time from day 1, with the most noticeable change from day 1 to day 2, and significant degradation occurring after day 4/5 gradually, which was consequently deemed the optimal day of sample acquisition (Figure 3.8).

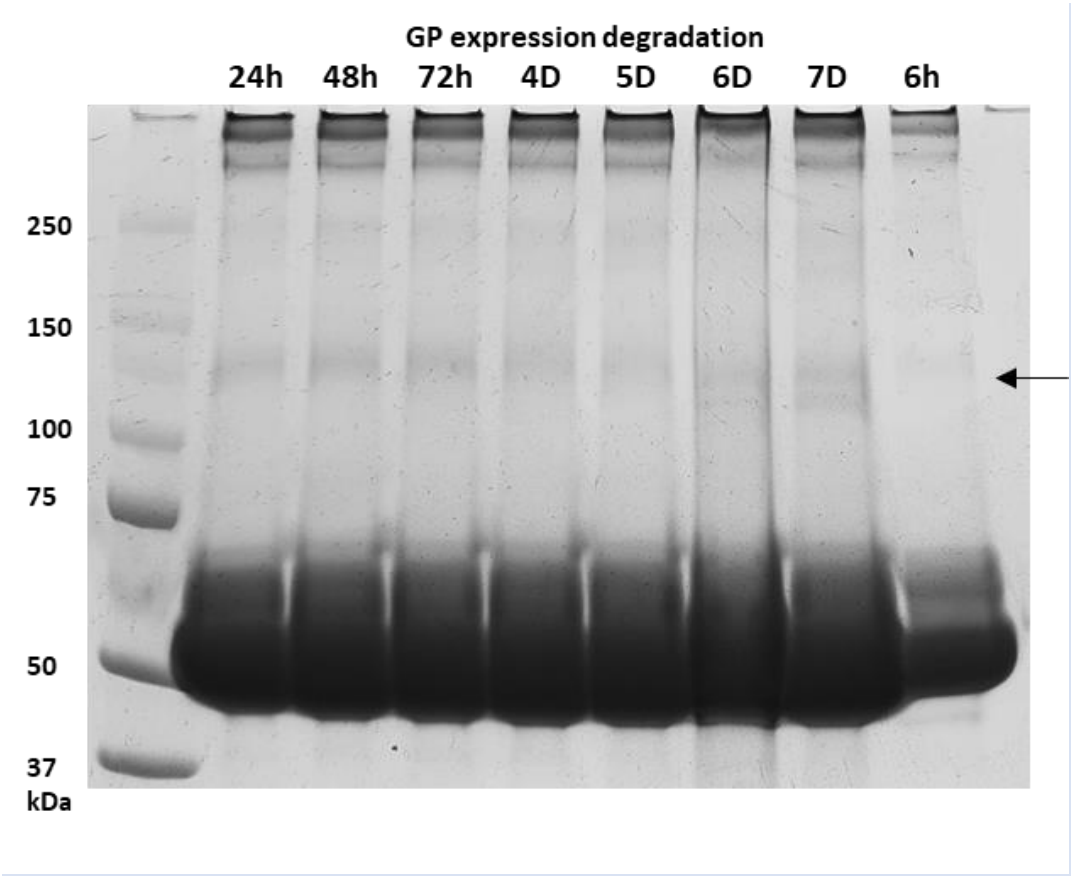


Figure 3.8: Secreted GP starts degrading 4 days post-transfection. GPΔTM was expressed in HEK293T plated in a 12 well plate. The culture medium was sampled six hours post-transfection before changing the medium, collected every 24 hours from a well. The medium was cleared, and protease activity stopped by adding protease inhibitor cocktail for an accurate time point. A gradient SDS-PAGE 6-16% gel was run under non-reducing conditions and stained with Coomassie. Degradation was detected post-day 4. The arrow indicates secreted GP. D, days post transfection.

When replacing the transfection medium mixture with fresh medium, a new batch of DMEM lacking serum of bovine or calf origin was prepared, to eliminate possible interactions and aggregation between ZEBOV GP and proteins in the serum, in addition to non-specific binding and possible contamination during protein purification (see Figure 3.7).

To summarise, the optimal expression of ZEBOV GP in the available facilities was achieved by growing HEK293T cells in 15cm TC dishes, lowering the initial cell density by more than 50%, while using linear PEI at a 5:1 DNA ratio during a 16-hour transfection incubation period, followed by a media change and up to 5-fold upscale in FBS-lacking DMEM, and harvesting after a 4-day expression period.

3.2.5 Purification of EBOV GP

3.2.5.1 *Initial IMAC purification of GP*

The authors of the first study resulting in a high-resolution structure of trimeric spike GP, which required large amounts of stable protein, utilised anti-HA agarose immunoaffinity columns to purify the protein[150]. Replicating the purification process was considered, however not pursued as the required techniques and instruments were not available in the accessible facilities. Additionally, the cost of over £5000 per single expression-purification run were a limiting factor.

As the next possible purification method, IMAC was selected due to the advantages discussed in Chapter 2. However, expression and purification of poly-His tagged GP in mammalian cells presented unique challenges. Previously, purification of GP using IMAC was deemed incompatible with the mammalian expression system due to non-specific binding of contaminants (including biotin and bovine serotransferase) and the chelating agents such as EDTA present in FBS and complete TC medium [105]. The authors present that despite extensive optimisation including numerous wash steps; over 100kDa dialysis; incubation with 8M urea; size exclusion chromatography and utilisation of different metal ions such as cobalt, the contaminants could not be removed and a final sample of GP with high purity was unachievable. As IMAC using a Nickel column is one of the most efficient purification techniques, this challenging purification and the challenges associated with it were addressed and troubleshooted during this dissertation. A trial purification of GP using a Cobalt column was performed, however the yield was significantly lower compared to Nickel column and additional

loss of material associated with subsequent purification steps would have resulted in diminishing returns (Figure 3.9).

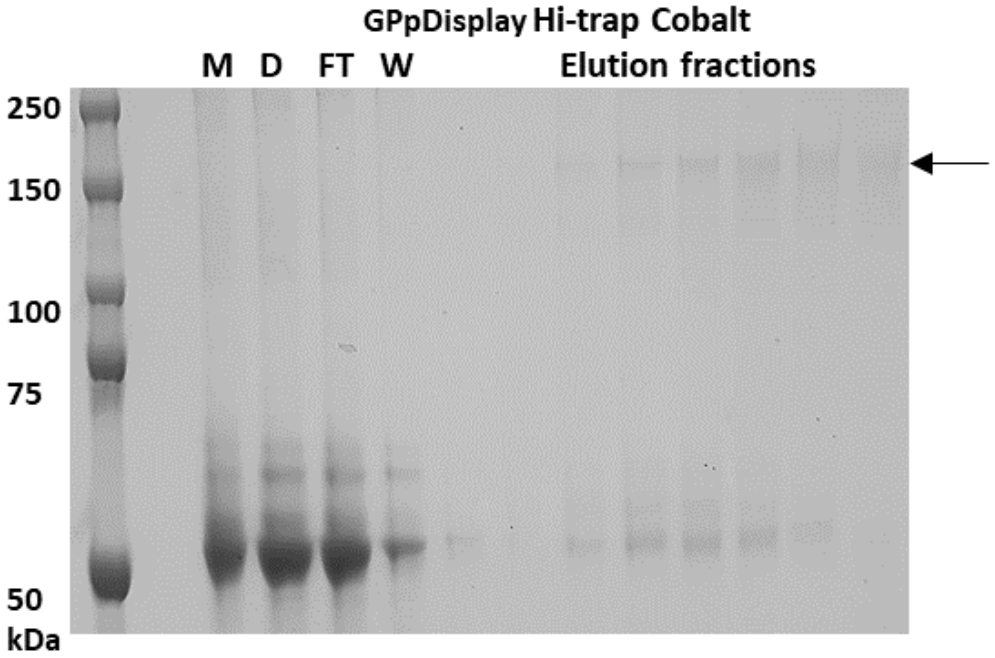


Figure 3.9: Purification of GP using Cobalt IMAC results in low yield. GPΔTM was expressed in HEK293T cells, the medium was harvested, concentrated, and dialysed prior Cobalt IMAC. All fractions were run on a gradient SDS-PAGE 6-12% under non-reducing conditions and stained with Coomassie. The elution fractions are one column volume at a time. Contamination by unknown protein at 55 kDa in elution fraction. See section 7 for more details. Arrow indicates GP. M, Culture medium sample containing GP; D, GP medium sample post-dialysis; FT, Flow through fraction of Cobalt column; W, wash fraction of Cobalt column.

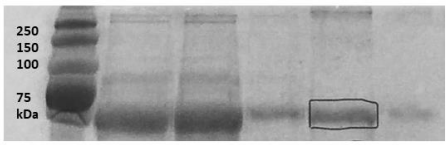
3.2.5.2 *Addressing protein contamination of GP during purification*

IMAC purification of proteins secreted from mammalian cells may lead to column stripping i.e. the detachment of metal ions from the resin and severe reduction in column capacity due to residual chelating agents present in the supernatant. Despite multiple rounds of sample dialysis, the resulting protein solution may still contain residual trace amounts of molecules capable of stripping Ni2+ columns. This may result in sample contamination by Nickel ions capable of activating TLR4, leading to artefacts and false positive activity results. Additionally, Nickel ions are capable of triggering an inflammatory response by directly activating TLR4. Therefore, contamination of GP protein or any other sample by Nickel or Cobalt ions lead to a background activity of TLR4 even with de-activated samples[151].

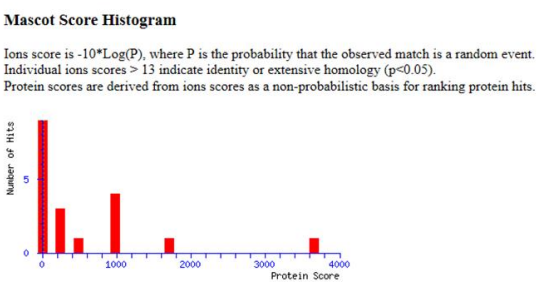
To solve this problem HisTrap excel columns, prepacked with Ni-Sepharose excel affinity resin resistant to stripping were employed here for the purification of EBOV GP, as they are specifically designed for mammalian culture samples containing secreted poly-His tag proteins.

The application of cell cleared culture medium from GP expressing HEK293T directly onto the HisTrap excel column without prior buffer exchange resulted in a mostly pure sample with a clear band around 150kDa. However, the flow-through sample still contained a significant amount of GP which did not bind to the column and was not purified in the process (Figure 3.7). To improve column binding two approaches were investigated. Firstly, the majority of the protein within the cell culture medium is of less than 75 kDa, therefore not GP but rather another protein which may block or sterically hinder the capture of poly-His tag by the column resin. Therefore, SDS-PAGE bands located between 75 and 50 kDa originating from mammalian cell culture medium were identified by mass spectrometric analysis with the main hit as serum albumin (*Bos Taurus*), a main component in FBS (Figure 3.10).

A



B



C

1. 1::sp|cRAP087|P02769|ALBU_BOVIN Mass: 71244 Score: 3658 Matches: 694(694) Sequences: 70(70) emPAI: 626.46
Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
☐ Check to include this hit in error tolerant search or archive report

Query	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Score	Expect	Rank	Unique	Peptide
☑ 141	395.2388	788.4631	788.4644	-1.60	0	(17)	0.02	1		K.LVTDLTK.V
☑ 142	395.2389	788.4633	788.4644	-1.30	0	(25)	0.0031	1		K.LVTDLTK.V
☑ 144	395.2393	788.4640	788.4644	-0.44	0	(18)	0.015	1		K.LVTDLTK.V
☑ 145	395.2394	788.4642	788.4644	-0.21	0	(17)	0.021	1		K.LVTDLTK.V
☑ 146	395.2396	788.4646	788.4644	0.33	0	(18)	0.015	1		K.LVTDLTK.V
☑ 147	395.2398	788.4651	788.4644	0.88	0	(15)	0.028	1		K.LVTDLTK.V
☑ 148	395.2400	788.4655	788.4644	1.42	0	42	7e-05	1		K.LVTDLTK.V
☑ 149	395.2400	788.4655	788.4644	1.49	0	(39)	0.00012	1		K.LVTDLTK.V
☑ 150	395.2403	788.4660	788.4644	2.03	0	(17)	0.018	1		K.LVTDLTK.V
☑ 322	409.7162	817.4178	817.4181	-0.43	0	(15)	0.034	1	U	K.ATEEQLK.T
☑ 323	409.7162	817.4178	817.4181	-0.43	0	(31)	0.00083	1	U	K.ATEEQLK.T
☑ 324	409.7162	817.4178	817.4181	-0.36	0	40	9.9e-05	1	U	K.ATEEQLK.T
☑ 336	410.2083	818.4021	818.4021	-0.08	0	(35)	0.00031	1	U	K.ATEEQLK.T + Deamidated (NQ)
☑ 337	410.2084	818.4022	818.4021	0.07	0	(35)	0.00032	1	U	K.ATEEQLK.T + Deamidated (NQ)

Figure 3.10: Mass-spectrometric analysis identifies serum albumin as the main contaminant in HisTrap elution. A.) SDS-PAGE of fractions from HisTrap purification of GP. The highlighted band below 75 kDa was submitted to the mass-spectrometry analysis. B.) Mascot Score Histogram of the hits. C.) The main hit identified the contaminant as serum albumin, *bos taurus*, a main component in Fetal bovine serum.

To counter this, expression trials in HEK293T cells were conducted with varying concentrations of FBS within the cell culture medium post transfection incubation (Figure 3.11). It became evident that a reduced concentration of FBS between 2% and 4% and an expression period of no more than 4 days

lead to the lowest contamination of serum albumin while maintaining a manageable level of cytotoxicity. Secondly, even though HisTrap excel columns are designed to be used with clarified medium directly it may be advantageous to collect, concentrate and buffer exchange via dialysis prior to IMAC purification to eliminate any possible nickel ion contamination.

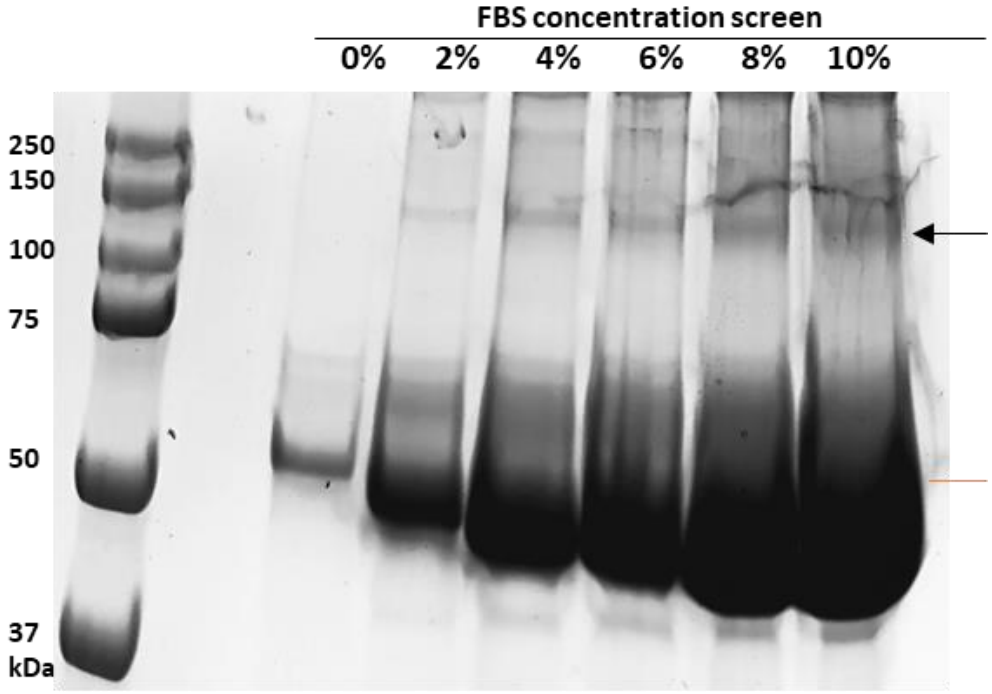


Figure 3.11: Identification of the minimum required FBS to express GP. HEK293T cells were cultured in 6-well plates and transfected with GPΔTM . Six hours post-transfection the culture medium was changed and replaced with medium containing FBS concentrations from 0-10%. The GP-containing medium was harvested and cleared 4 days post-transfection and run on a gradient SDS-PAGE 6-16% gel under non-reducing conditions and stained with Coomassie. Detailed protocol listed in section 7. The arrow indicates GP in the culture medium sample.

3.2.5.3 *Improving GP purification*

Prior to IMAC purification, clarified cell culture supernatant of GP-expressing HEK293T cells was concentrated using a 100kDa (Sartorios) cut off, and transferred to dialysis membrane standard tubing. Initial dialysis of the concentrated GP media in HisTrap calibrating buffer (300 mM NaCl, 20 mM Tris HCl, 20 mM Imidazole) at 4°C resulted in protein precipitation, aggregation, and immense material loss. To prevent protein aggregation and possible denaturation the subsequent dialysis runs were conducted at room temperature to increase molecular movement and an alternative HisTrap buffer composition was designed. The NaCl concentration was reduced from 300 mM to 100 mM, Tris

HCl was replaced by 20 mM HEPES to reduce possible aggregation and the amount of Imidazole was doubled to prevent further unspecific binding of aggregates to columns downstream of the buffer exchange. In anticipation of anion exchange chromatography as a potential secondary purification method for GP with a PI of 6.16, a pH of 8 was chosen to ensure sufficient surface charges for anion exchange. This strategy abolished protein precipitation and set up the GP-containing supernatant for improved HisTrap excel column binding and purification. However, the purified sample contained significant contamination by serum albumin as identified by SEC (Figure 3.12A). Therefore, a second purification step was necessary to achieve the desired GP purity. Following Nickel IMAC purification the eluted GP solution was subjected to further purification steps. Firstly, anion exchange chromatography using a HiTrap Q column via a benchtop peristatic pump was performed and the elution fractions were pooled and applied to gel filtration column to assess the purity and stability of the sample (Figure 3.12B).

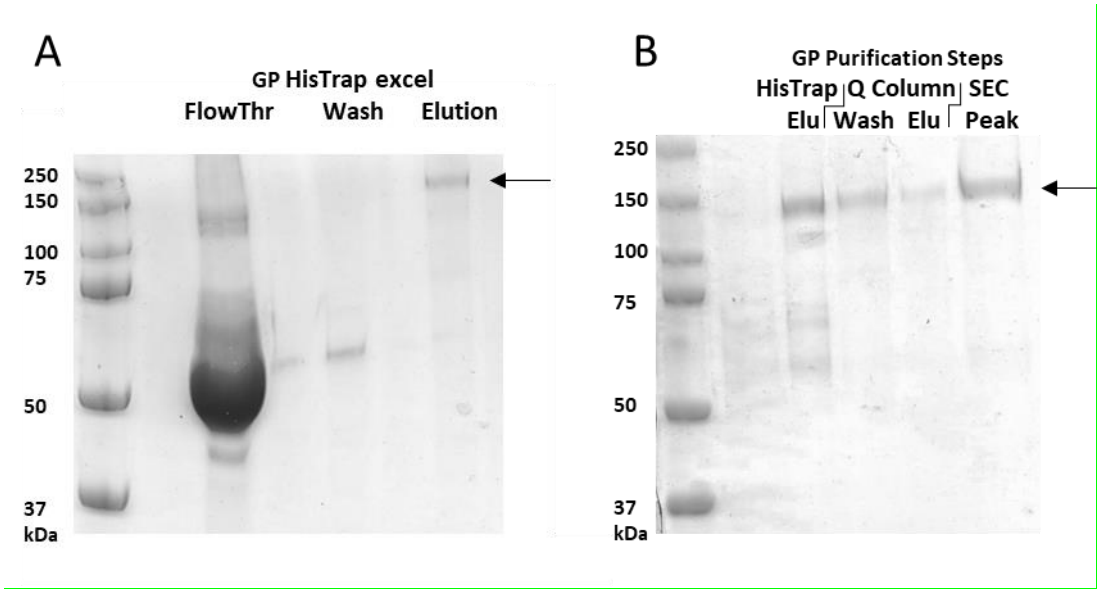


Figure 3.12: The optimised purification for EBOV GP. GPΔTM was expressed by HEK293T cells and the culture medium was harvested 4 days post-transfection. The cleared culture medium was cleared and concentrated followed by dialysis to prepare for protein purification. A.) GP was purified using a HisTrap excel column and the fractions were pooled and run on a gradient 6-16% SDS-PAGE gel under non-reducing conditions and stained with Coomassie. B.) After HisTrap purification of GP, the elution fraction was concentrated, and anion-exchange (q-column) was performed followed by size exclusion chromatography (SEC) to increase and assess sample purity. The fractions were pooled and run on a gradient 6-16% SDS-PAGE gel under non-reducing conditions and stained with Coomassie. The detailed protocol is noted in section 7. The arrow indicates GP bands. FlowThr: flow through of sample application on column; Elu: elution fraction.

Furthermore, SEC was not only utilised as a secondary purification step but also to visualise the oligomeric state of GP. Potentially, certain purification conditions such as the salt concentration in the

buffer may change the equilibrium between GP monomer and GP trimer and, as shown by Rutten et al. 2020, even single point mutations can influence the spike trimerization significantly[152].

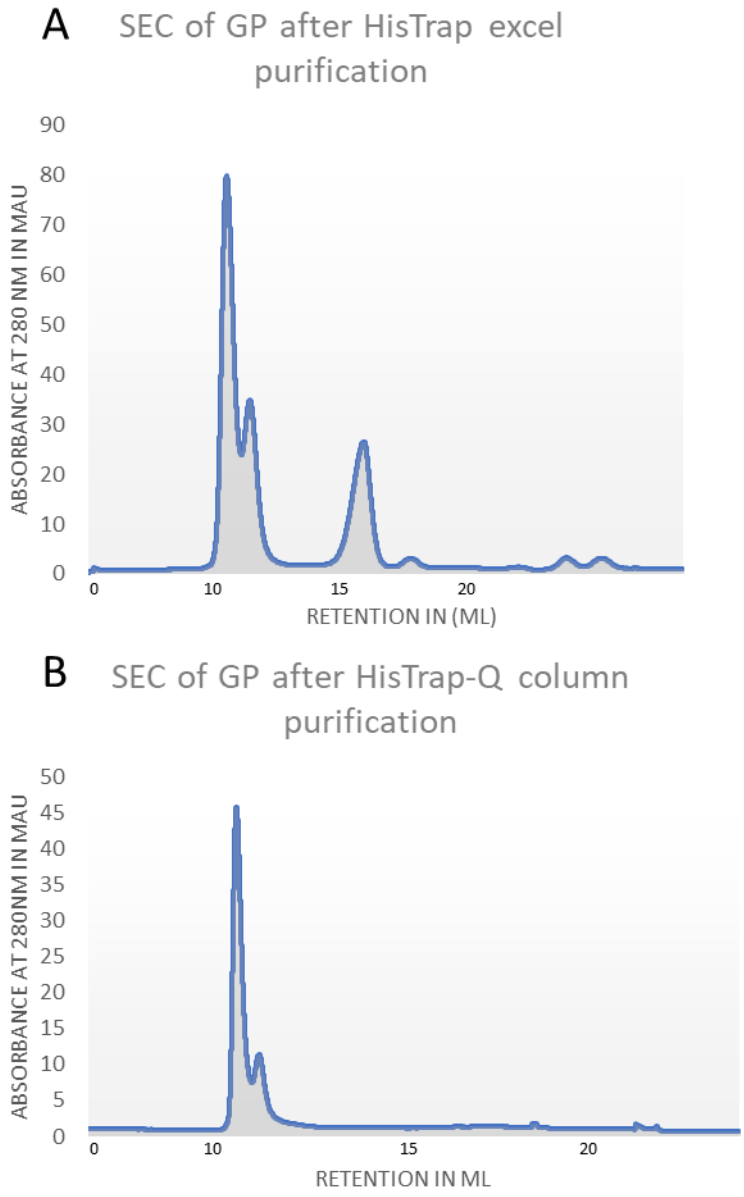


Figure 3.13: SEC elution diagram of EBOV GP on AKTA after single and multi-step purification. GP was purified using HisTrap alone (A) and HisTrap followed by anion exchange (Q-column) (B). The purifity as well as the homogeneity of the GP sample was assessed with size exclusion chromatography on an AKTA FLPC. A.) Three peaks were identified as GP trimer, GP monomer and serum albumin indicating contamination. B.) Two peaks were identified as GP trimer, GP monomer indicating improved purity after multi-step purification. A detailed protocol is noted in Chapter 7.

However, SEC revealed that the majority of purified GP is present in its trimeric state with a portion in monomeric form after two-step purification (Figure 3.13 B). Thus, I have successfully created an easily

accessible method of producing native, oligomeric EBOV GP followed by purification while maintaining stability and LPS-free conditions. Utilising this methodology GP can be produced in basic laboratory conditions without the need for advanced instruments and financial strains.

Table 3.2: Summary of the main optimisation steps to purify EBOV GP.

Step	Initial	Issue	Solution
GP first purification	Cobalt IMAC	Low yield	Only Nickel IMAC
GP first purification	Nickel IMAC	Nickel contamination and column stripping	Use HisTrap excel
GP HEK medium	GP secretion with 10% FBS	Serum albumin contamination	Lower FBS content in medium
GP HEK medium	Direct loading of HEK medium	Serum albumin contamination	Concentrate medium with 100 kDa cut-off for buffer dialysis
GP dialysis	Dialysis in HisTrap calibration buffer at 4°C	Protein precipitation, material loss	Move to RT, lower salt content in buffer
GP dialysis	HisTrap calibration buffer: 300 mM NaCl, 20 mM Tris HCl, 20 mM Imidazole	Protein precipitation, aggregation	HisTrap calibration buffer: 100 mM NaCl, 20 mM HEPES, 40 mM Imidazole
GP purified by IMAC	Single step purification	Contamination by serum albumin	IMAC buffer at pH 8.0 to make anion-exchange compatible

3.2.6 Analysing activity and physiological relevance of purified GP

Following the multi-step purification of EBOV GP and the confirmation of purity via SEC, the protein samples were evaluated by their ability to activate the TLR4 signalling complex. For this NFκB reporter signalling assays that were established in the research group were adapted to GP activity analysis. HEK293 cells carrying plasmids for the expression of TLR4, MD2 and CD14 were selected using antibiotic resistance markers to generate a stable cell line (HEK293 TMC). The stable HEK293 TMC cells were transiently transfected with a plasmid carrying the genes for luciferase under an NFκB promoter. This enabled an indirect readout of NFκB activation via luciferase expression and measurable activity upon substrate stimulation. A luminescence plate reader was utilised for luciferase activity readout

and the luminescence levels were translated into relative NFκB activity and showed that the purified GP was capable of activating TLR4 (Figure 3.14).

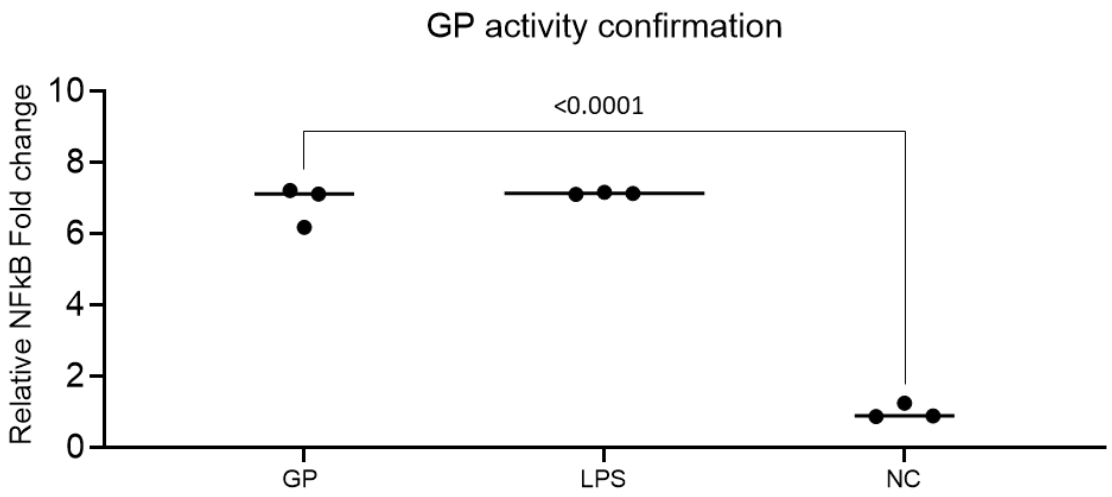


Figure 3.14: Ability of purified GP to activate TLR4 confirmed through NFκB luciferase assay. Purified GP was tested for ability to activate TLR4. Stable TLR4/MD2/CD14 HEK293 cells were used for NFκB signalling assays. The relative NFκB activity as measured by luciferase luminescence after 6-hour induction with 100 ng/µl LPS and 2 µg/µl GP was normalised to the negative control. The values are presented as means of three biological experiments with five technical replicates. The negative control in this case was a random poly-His-tagged non-TLR4- ligand protein purified with the same protocol as GP in order to rule out possible Ni^{2+} and LPS contamination. The negative control was set to 1-fold change and the values normalised to it. Data represent means of three replicates. Statistical significance was assessed using a one-way ANOVA test.

3.3 Discussion

Here I showed that the deletion of the TMD of GP led to the complete loss of glycosylation without affecting secretion. Reintroducing the first few residues of the transmembrane helix led to reinstatement of the glycosylation while keeping the secretion efficient independently of TACE cleavage. It may be possible that the TMD within the c-terminus of GP2 consists of an oligomerisation region necessary for the trimerisation and stability of GP during the post-translational modification process and the removal of it affects glycosylation. The reason why glycosylation but not secretion was affected by the truncation remains unknown, as destabilisation of GP during the post-translational processing should affect the secretion pathways and protein localisation as well. Perhaps the transmembrane domain of GP could be used as a novel therapeutic target to disrupt the GP stability and glycosylation during the post-translational modification process. As the extensive glycosylation acts as a shield protecting the protein from 54mmune-detection this may be an interesting target.

Although the expression and purification strategies presented in this Chapter were able to routinely yield 5-10 mg of GP per litre of culture medium used the process may be improved further. Even though expression systems utilising a cell culture in suspension such as Expi293F allow higher cell

density per volume and consequently result in higher yields, particularly for secreted proteins, adherent HEK293T cells were chosen as the expression system for EBOV GP. Unfortunately, suitable facilities supporting the expression in mammalian suspension cells in a BSL2 environment were not available throughout the research project and expression of GP in BSL1 was not approved within the accessible facilities. Additionally, using a stable cell line with an inducible promoter may be more efficient over the long term due to transfection reagent and DNA extraction kit usage and associated costs. However, using adenoviruses or other means to integrate a gene of the Ebola Virus into the genome of mammalian cells would require more advanced facilities and most likely a BSL3 environment.

The purification could be improved by finding or developing serum-free culture media to eliminate the contamination by serum albumin when using IMAC. Although this process was improved by using HisTrap excel columns, which also avoid any contamination by Nickel ions that could falsely activate TLR4, a different affinity purification may be used for more specific purification such as a strep column. The disadvantage with the strep column would be the contamination with biotin within the serum, leading to inefficient column binding. All in all, improving the purification process of native, and fully glycosylated GP secreted by mammalian cells requires much more extensive optimisation and care compared to the improvements that could be made at the level of expression.

To conclude, in this Chapter I presented the techniques and strategies required to optimally express and purify the complex and cytotoxic EBOV GP trimer, in large quantities while maintaining the native state, glycosylation and activity as assayed by NFκB activation, a world's first. As these approaches were conducted in restrictive circumstances, this suggested methodology can be utilised by researchers constrained by instruments or finances.

4 The role of GP glycosylation in TLR4 activation

4.1 Introduction

In the previous Chapter I presented the strategies developed to express and purify stable and fully glycosylated GP in large quantities. This formed the foundation required to investigate the importance of glycosylation in TLR4 signalling complex activation. The 75 kDa EBOV GP is heavily glycosylated, which can be confirmed via SDS-PAGE with an observed molecular weight of ~150kDa[153]. Prediction software of N- and O-glycosylation on GP reveals 17 possible N-glycan sites, enriched in the N-terminal glycan cap, and over 80 O-glycan sites most of which are located within the unstructured mucin-like domain (MLD) visualised as the purple propeller blade-like structure in Figure 4.1.

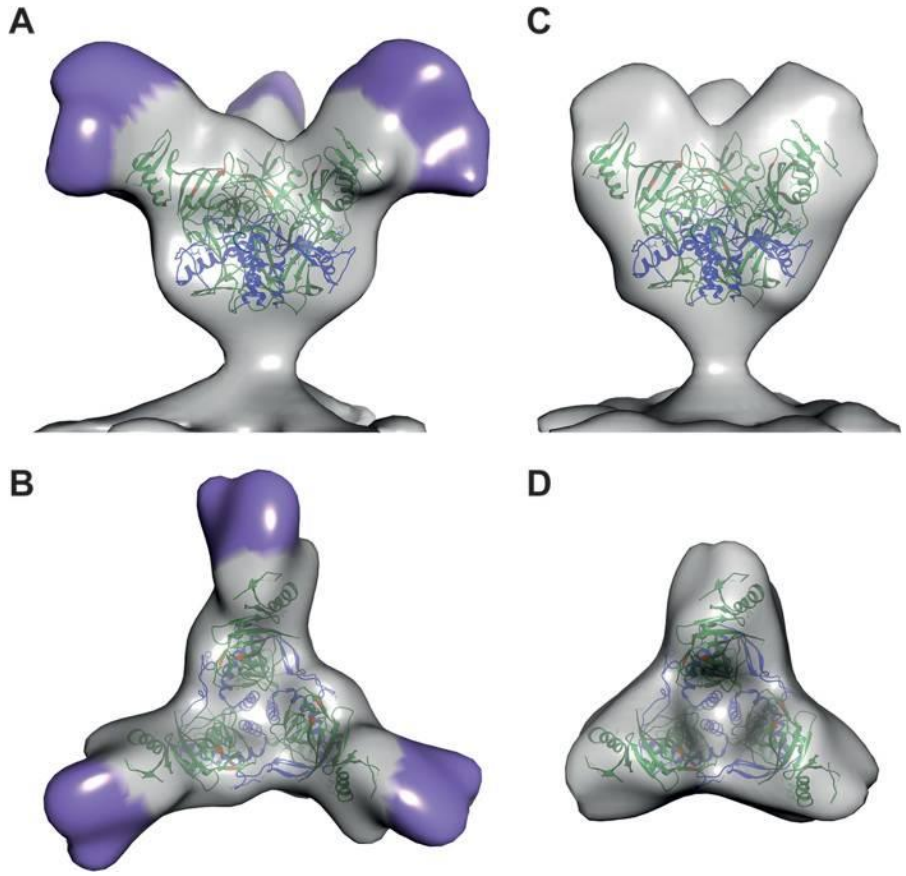


Figure 4.1: Spatial localisation of MLD on EBOV GP trimer determined by cryo-ET. The electron density of trimeric GP immobilised attached to the membrane of virus-like particles. Obtained from cryo-electron tomography was superimposed with the crystal structure 3CSY, [PDB]. A.) Side view of GP trimer electron density with MLD. B.) Top view of GP trimer electron density with MLD. C.) Side view of GP trimer electron density without MLD. D.) Top view of GP trimer electron density without MLD. Purple colouring indicated the unstructured and glycosylated Mucin-like domain (MLD). Adapted from [154]

Previous mass spectrometry studies of modified GP1,2 revealed high mannose, hybrid, bi-, tri-, tetra-antennary complex glycans with ~50 different species suggesting differentially glycosylated sites

varying from protein to protein [153]. About 40% of O-linked glycans were found to contain mono or double sialic acid glycans.

This extensive glycosylation forms a shield around the protein protecting it from antibody recognition and has also been implicated in decreasing surface expression of immune receptors such as MHCI and integrins leading to cell dissociation and death. This cytotoxicity is thought to be dependent on the MLD and the glycan cap of GP1.

Although the glycosylation of GP, particularly the mucin-like domain, has been implicated in GP-mediated cell cytotoxicity, the role in TLR4 activation remains unknown and a potential mechanism is only speculative. In this Chapter I present the approaches taken to identify the role of individual domains of GP and how glycosylation may play in the activation of TLR4.

4.2 Results

4.2.1 Elucidating the Minimal GP domain requirement for activation of TLR4

Firstly, the GP domain requirements for TLR4 activation were investigated by designing and cloning truncation mutants of GP. The minimal domain screen was purposely limited to the glycosylated and MLD carrying GP1, due to the focus on the role of glycosylation which is predominantly at GP1. Overall, five GP truncation constructs were created with pDisplay as a backbone (see Figure 4.2) and their ability to activate TLR4 was compared to that of the optimised GP construct GP1,2 (C-terminal helix of TMD). The first construct was created by removing the intracellular and transmembrane domain of GP (including the helix discussed in Chapter 3), named ECD (GP). The second construct, named GP1, was created by prematurely terminating GP1,2 at the Furin cleavage site after amino acid residue 501. The next construct shortened GP1 to only include all residues upstream and including the MLD, named GP-485. The last two constructs were created by cloning the MLD downstream of the signalling peptide, either in isolation (referred to as MLD henceforth) or also including the final few residues of GP1 in the construct MLD-501. A schematic depiction of the five constructs for minimal domain sreening is shown in Figure 4.2.

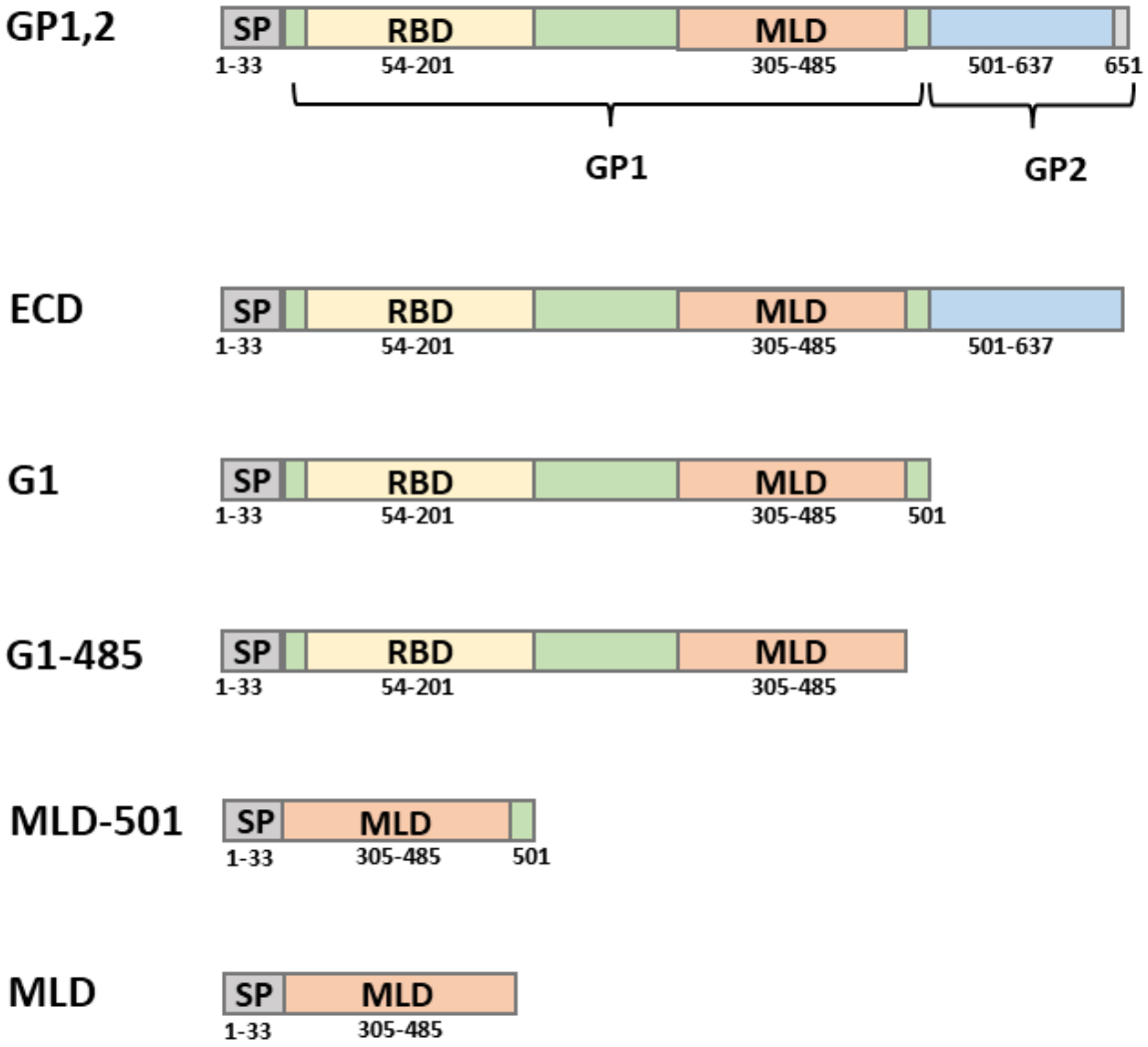


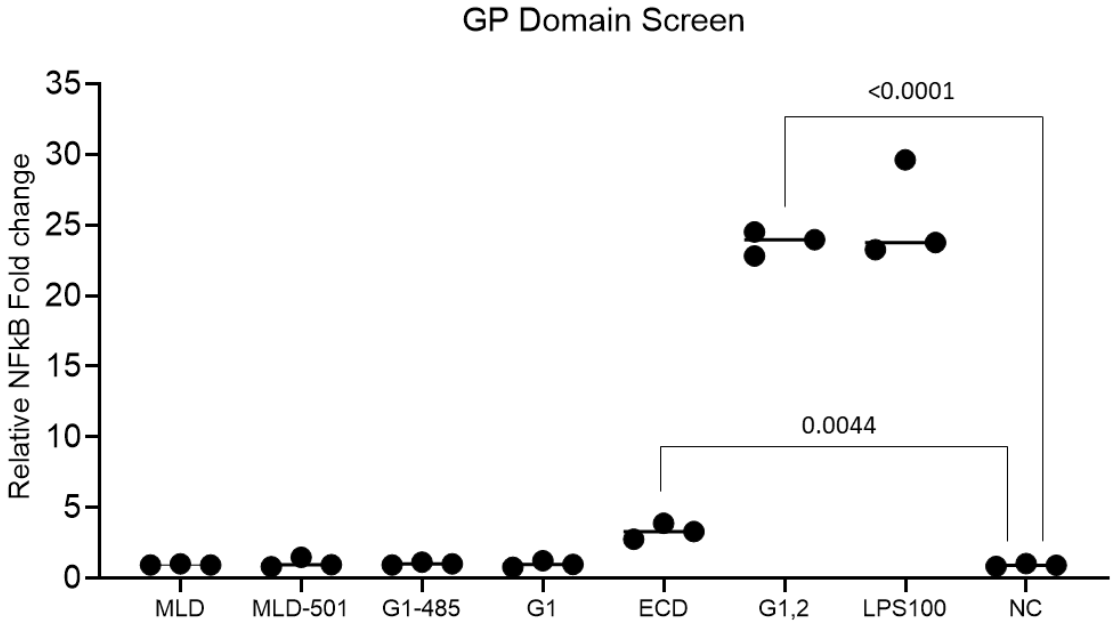
Figure 4.2: Truncation constructs of EBOV GP used for limited domain screen. The domain boundaries are indicated by the residue position number within the polypeptide chain. The individual domains are indicated through colour separation GP1 (green), GP2 (blue), GP: Glycoprotein, ECD: Ectodomain, SP: Signalling peptide (grey), RBD: Receptor binding domain (yellow), MLD: Mucin-like domain (orange) and the C-terminal transmembrane domain starting from residue 638.

To study the activation of TLR4, the GP truncation constructs were expressed in HEK293T cells and proteins were harvested as previously described (see Chapter 7). Due to concerns that the truncation of GP may affect secretion and/or protein stability, a protease inhibitor cocktail was added to the expression medium (see Chapter 7 for details). To increase the amount of protein available for the assay, the culture medium was concentrated and merged with cleared cell lysate, followed by multiple rounds of dialysis in 50-times the sample volume sterile filtered PBS to remove detergents.

IMAC purification of the GP truncation constructs was attempted according to the previously established protocols, however the constructs were incompatible with purification. While GP1,2 and GP ECD were able to be expressed and purified as shown in chapter 3, the remaining constructs G1,

G1-485, MLD-501 and MLD resulted in HisTrap excel column blockage caused by protein precipitation. The common structural feature of these constructs is the lack of GP2 which seemingly affected protein stability and solubility. Due to this complication, for the purpose of analysing TLR4 activation, the concentrated and dialysed samples were used for NFκB luciferase signalling assays instead of the purified protein samples. The presence of each protein species in the crude cell extracts was assessed with western blotting and the samples were diluted according to the associated signal strength to achieve approximate equal amounts of protein present in each sample.

As depicted in Figure 4.3, the sample containing the construct GP1,2 led to a NFκB induction of similar levels as the positive control. All other truncation constructs were not able to activate TLR4, although the ECD construct did show a minor activation of TLR4, however after statistical analysis it was deemed insignificant.



GP mut	MLD	MLD-501	G1-485	G1	ECD	G1,2
WB						
Dilution	1:2	1:1	1:2	1:2	1:1	1:2

Figure 4.3: Limited domain screen of GP required to activate TLR4. The GP truncation constructs were expressed in HEK293T cells. Four days post-transfection the medium was harvested, and the cells were lysed. Both medium and cell lysate were cleared, pooled, concentrated, and dialysed against PBS. A NFκB luciferase activity screen was conducted with stable TMC HEK293 cells and 5 µg/µl of the crude isolates of the GP truncation constructs, previously diluted post-WB. W was carried out with 1:10,000 anti-His and 1:5000 anti-mouse. The negative control consisted of medium and cell lysate from untransfected cells. The negative control was set to 1-fold change and the values normalised to it. Data points represent means of triplicates. Statistical significance was assessed using a one-way ANOVA test.

Based on the outcome of these initial investigations, it was hypothesised that to activate TLR4, the EBOV GP must be as native as possible, while truncations or significant domain mutations may lead to

instability, solubility issues and other complications, both experimentally and *in vivo*. Furthermore, in light of the insight gained during the expression and purification optimisation of GP constructs lacking the TMD in Chapter 3, the ECD construct lacking glycosylation and further modifications and truncations seem to have a similar negative effect on stability and/or solubility. Based on these findings, it seems that truncation of EBOV GP, negatively impacts stability or solubility of the protein, and thus the entire protein is either directly or indirectly required for the activation of TLR4. Next, the role of EBOV GP glycosylation in the activation of the TLR4 signalling complex was studied, by using the full-length GP protein, mutated at various glycosylation sites.

4.2.2 Effect of GP glycosylation to activate TLR4

4.2.2.1 *Investigating the effect of individual glycans in the activation of TLR4*

To assess the impact of GP glycosylation on the activation of TLR4, the protein sequence was analysed with the glycosylation prediction software NetNGlyc and NetOGlyc, which indicates the probability of a glycosylation occurring *in vivo*, with a value of about 0.5 indicating a high likelihood[155]. The software predicted seventeen possible N-linked glycosylation sites, fifteen for GP1, and two for GP2. Ten of these showed a glycosylation potential of > 0.5 and were enriched in the glycan cap (residues 200-300) at position 40, 204, 228, 238, 257, 268, while the others were predicted to be at position 317, 386, 436 and 563 located within or close to the MLD (Figure 4.4). Of over 80 predicted O-linked glycosylation sites, five were predicted to have a high glycosylation likelihood (>0.5), of which >80% were located at the MLD (not shown). Given the average molecular weight of N- and O-linked glycans the theoretical molecular weight of all glycans on the 97 glycosylation sites was calculated as 75 kDa. This theoretical molecular weight would add to the molecular weight of GP as estimated by SDS-PAGE (from 75 kDa to 150 kDa) which is similar to the observed molecular weight of GP of 140-150 kDa.

MGVTGILQLPRDRFKRTSFFLWVIILFQRTFSIPLGVIHNSTLQVSDVDKLVCRDKLSSTNQLRSVGLNLEGNGVATDVP 80
SATKRWGFRSGVPPKVVNYEAGEWAENCYNLEIKKPDGSECLPAAPDGIRGFPRCRYVHKVSGTGPCAGDFAFHKEGAFF 160
LYDRLASTVIYRGTTFAEGVVAFLILPQAKKDFFSSHPLREPVNATEDPSSGYYSTTIRYQATGFGTNETEYLFEVDNLT 240
YVQLESRFTPQFLLQLNETIYTSGKRSNTTGKLIWKVNPEIDTTIGEWAFWETKKNLTRKIRSEELSFTVVSNGAKNISG 320
QSPARTSSDPGTNTTTEDHKIMASENSSAMVQVHSQGREAAVSHLTTLATISTSPQSLTTKPGPDNSTHNTPVYKLDISE 400
ATQVEQHHRRTDNDSTASDTPSATTAAGPPKAENTNTSKSTDFLDPATTTSPQNHSETAGNNNTHHQDTGEESASSGKLG 480
LITNTIAGVAGLITGGRRTRREAIVNAQPKCNPNLHYWTTQDEGAAIGLAWIPYFGPAAEGIYIEGLMHNQDGLICGLRQ 560
LANETTQALQLFLRATTELRTFSILNRKAIDFLLQRWGGTCHILGPDCCIEPHDWTKNITDKIDQIIHDFVDKTLPDQGD 640
NDNWWTGWRQWIPAGIGVTGVIIAVIALFCICKFVF

Figure 4.4. Predicted N-linked glycosylation sites of EBOV GP. RED: Predicted possible sites for glycosylation, Blue: Most likely to be glycosylated, Underscored: Mucin-like domain.

The exact positions of predicted glycosylation sites and their potential (i.e. how likely those sites are occupied by N-linked glycans) are shown in Table 4.1. Although, the threshold was set to 0.5 by the prediction software, due to the proximity to 0.5 the possible glycan occupied sites at residues 296 and 618 were also included in the next investigation steps.

Table 4.1: Summary of all predicted N-glycosylation sites of GP and their potential of glycosylation. The position and sequence of each N-linked glycosylation site of GP is presented together with the potential score indicating the likelihood of glycosylation if above 0.5. The shaded rows indicate the mucin-like domain and the glycosylation sites within it.

Number	Position	Sequence	Potential (0.5 threshold)
1	40	NSTL	0.7499
2	204	NATE	0.6833
3	228	NETE	0.6051
4	238	NLTY	0.7198
5	257	NETI	0.7112
6	268	NTTG	0.6196
7	296	NLTR	0.4817
8	317	NISG	0.6616
9	333	NTTT	0.4946
10	346	NSSA	0.4328
11	386	NSTH	0.6740
12	413	NDST	0.4086
13	436	NTSK	0.6012
14	454	NHSE	0.5020
15	462	NNTH	0.4819
16	563	NETT	0.5234
17	618	NITD	0.4543

Based on the predicted N-glycosylation sites GP mutants were designed with individual altered glycosylation sites. The nine chosen mutations were located outside of the unstructured MLD with seven located within the GP1 N-terminal glycan cap domain (40, 204, 228, 238, 253, 268, 296) and two within the C-terminal region of GP2 (563, 618). Due to the unstructured and non-conserved nature of the MLD, and the extensive O-linked glycosylation located within it, the N-glycosylation sites were deemed to have an unlikely role in the activation of TLR4 and a primary function in immune evasion and epitope shielding and thus discarded from further experimental analysis.

To generate GP constructs with single-amino acid changes to alter glycosylation patters, site directed mutagenesis was used. The N-glycosylation sequons of Asn-Xaa-Ser/Thr were mutated to Glu-Xaa-Ser/Thr or Asn-Xaa-Val to disrupt the glycosylation. The rationale behind conducting a double mutant screen was that Saphire et al deleted all glycosylation sites of GP in order to resolve the crystal structure. They discovered that some glycosylation site mutations depending on to type of mutation, N to Q or S/T to V, may lead to reduced expression and protein instability[32]. To eliminate this as a possibility, individual mutants were created for each glycosylation site.

Table 4.2: Individual glycosylation site mutants of GP. Summary of the N-linked glycosylation sites of GP selected for mutant activity screen. The position and sequence of each site within the peptide chain is noted including the mutations of Asn to Glu and Thr to Val. The grey marked rows contain the mutants which did not express.

Position of Glycan	Sequence	Mutation	Expression
40	NSTL	N40D	No
40	NSTL	T42V	Yes
204	NATE	N204D	No
204	NATE	T206V	Yes
228	NETE	N228D	Yes
228	NETE	T230V	Yes
238	NLTY	N238D	Yes
238	NLTY	T240V	Yes
257	NETI	N257D	Yes
257	NETI	T259V	Yes
268	NTTG	N268D	No
268	NTTG	T270V	Yes
296	NLTR	N296D	No
296	NLTR	T298V	Yes
563	NETT	N563D	No
563	NETT	T565V	Yes
618	NITD	N618D	Yes
618	NITD	T620V	No

Each individual construct with a single glycosylation site mutation of GP and their sequence change is listed in Table 4.2. Of the 18 glycosylation site mutants 12 were able to be expressed in HEK293T cells.

Of the six non-expressed constructs five encompassed a N → D amino acid change, a pattern that was noted but not further investigated due to the common practice of mutating Asn to inhibit glycosylation in biochemistry. A summary of the single glycosylation site mutants of GP which expressed well and were taken forward to be used in TLR4 signalling assays is depicted in Figure 4.5.

The single glycosylation site mutants of GP were investigated for their potential to activate TLR4, to elucidate whether a particular glycosylation site may play a role in the activation. For this NFκB activity was used as a readout for TLR4 activity. As HEK293T cells do not express receptors required to activate the NFκB pathway natively, this readout correlated directly with TLR4 activity. The ability to induce NFκB of each mutant was compared against the positive GP control. The luciferase assay did not show any significant reduction in NFκB activity for any mutant. In fact the difference between wild type GP and each mutant was insignificant with fold change levels around the ten-fold mark compared to the normalised background (Figure 4.6).

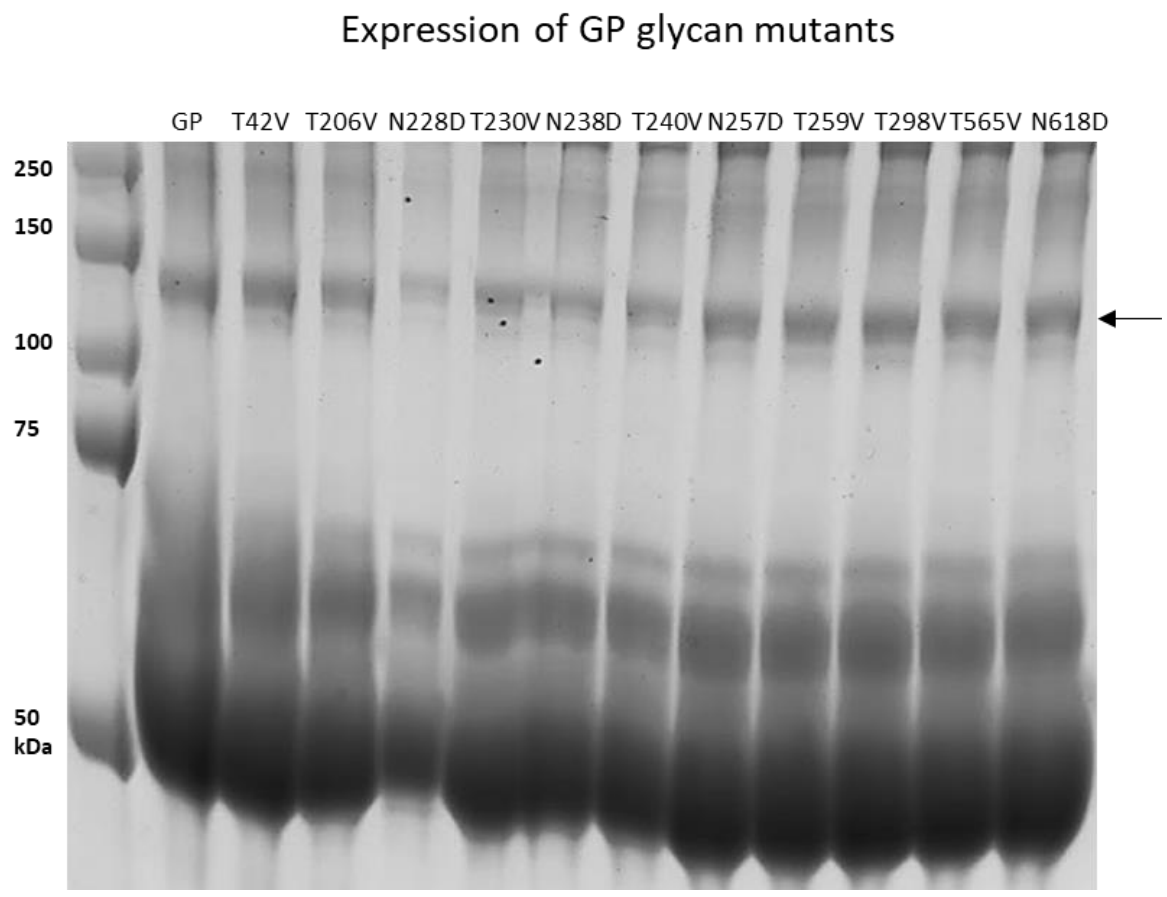


Figure 4.5: Collection of single-glycosylation site GP mutants that express. HEK293T were transfected with plasmids carrying single-glycosylation GP mutants. Four days post-transfection the medium was harvested, cleared, concentrated, and run on a gradient 6-12% SDS-PAGE under non-reducing conditions and stained with Coomassie. The detailed protocol is listed in Chapter 7. The arrow indicates the GP proteins.

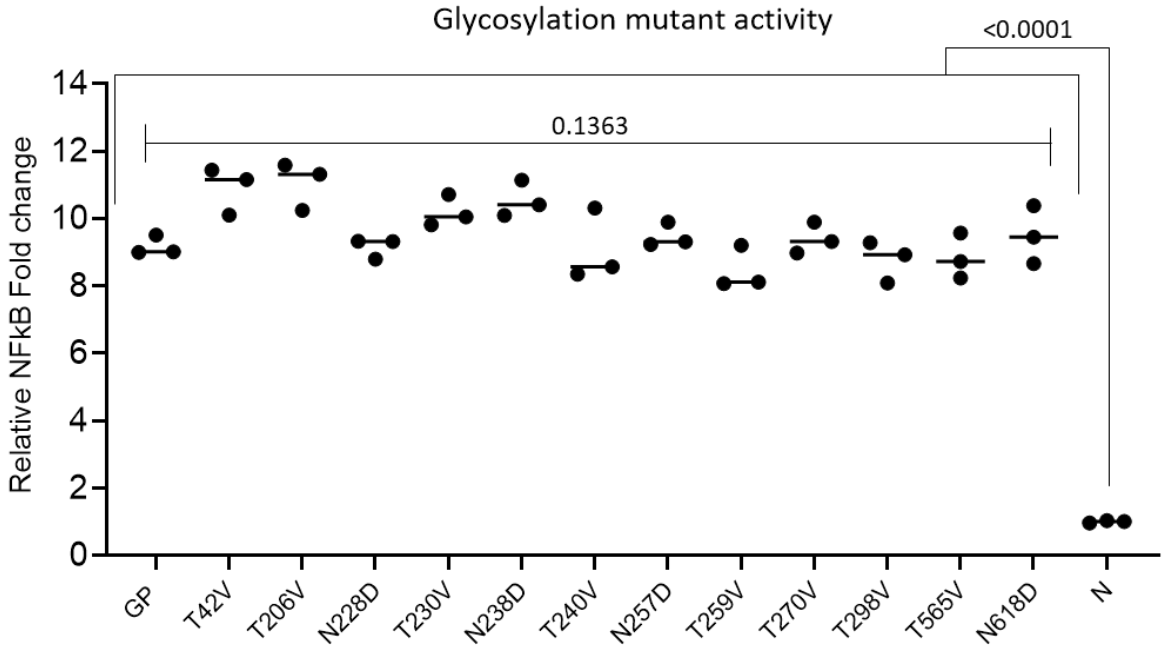


Figure 4.6: The single-glycosylation GP mutants do not show a significant change in TLR4 activation. HEK293T cells were used to express GP constructs with single-glycosylation site mutations. Each of the mutant GP constructs that expressed was purified according to the same protocol as WT GP. A NFκB luciferase activity screen was conducted with stable TMC HEK293 cells and 1 µg/µl of the purified GP mutants. The negative control consisted of purified GP, treated with proteases for 1 hour followed by boiling for 20 minutes. The negative control was set to 1-fold change and the values normalised to it. Data points represent means of triplicates. Statistical significance was assessed using a one-way ANOVA test.

It may be possible that a single glycosylation mutant does not induce enough structural or chemical changes to GP, but multiple mutants may result in significant changes to GP activity. However, as some single glycosylation mutants did not express, it is likely that multiple mutations lead to further destabilisation of GP and expression problems.

4.2.2.2 *Enzymatic removal of glycans of the EBOV GP to study TLR4 activation*

One hypothesis that could explain that mutating individual glycosylation sites did not influence TLR4 activation, is that multiple sites need to be altered. Therefore, to further investigate the role that GP glycosylation may play in the activation of TLR4, a deglycosylation assay was conducted. Here O-glycosidase was utilised to evaluate the overall effect of O-linked glycans of which over 80 are predicted to cover the MLD of GP. To assess which N-linked glycans may be able to affect the activation of TLR4, PNGase F, Endo H, Galactosidase and Neuraminidase were used in the de-glycosylation assay (Figure 4.7).

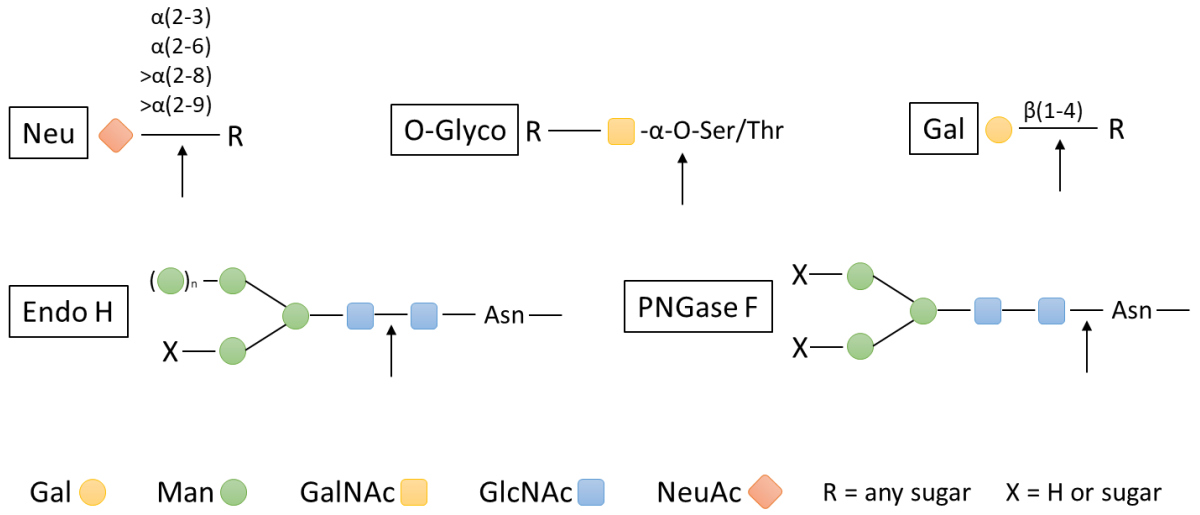


Figure 4.7: Cleavage patterns of de-glycosylation enzymes. Schematic depiction of the cleavage patterns of deglycosidases and substrate specificity. Neu; Neuraminidase; O-Glyco: O-Glycosidase; Gal (enzyme): Galactosidase; Endo H: Endoglycosidase H; PNGase F, Peptide-N-glycosidase F; Gal (sugar), Galactose; Man, Mannose; GalNAc, N-Acetylgalactosamine; GlcNAc: N-Acetylglucosamine; NeuAc: N-Acetyl-Neuraminic Acid; H: Hydrogen; X: extension of mannose; R: any sugar; Asn: Asparagine; Ser: Serine; Thr: Threonine. The arrow indicates the enzymatic cleavage pattern.

The manufacturer's recommended protocol for enzymatic removal of glycans was heavily modified and optimised to be compatible with high levels glycosylation of EBOV GP. A summary of the main challenges and optimisation steps taken is listed in Table 4.3.

Table 4.3: Summary of the optimisations for enzymatic removal of glycans. This table lists the protocols recommended by the manufacturers, the challenge faced when applying them to GP, the optimisation steps taken, and the final effect on the process.

Recommendation	Challenge	Resolution	Effect
Less than 5 µg protein deglycosylated	Insufficient amount for assays	Upscale reaction for at least 1 mg of protein	Sufficient amount for assays
Addition of denaturing buffer and sample boiling	Irreversible denaturation leading to protein in-activity	Instead, add 1.5 M Urea to protein mixture	Partial and reversible unfolding of protein to allow enzyme access
Short 2-hour incubation	Only partial deglycosylation	Increase to 16-hour incubation	Homogenous deglycosylated GP
Incubation at 37 °C	Degradation during 16-hour incubation	Addition of 10% glycerol per volume, and protease inhibitor	Stabilisation of GP protein
Reaction stop at 100 °C	Irreversible denaturation of target protein	Instead, 0.22 µm filtration, dilution in saline solution and concentration at 50 kDa cut-off	Removal of glycans, inactivation of enzymes and buffer exchange for assays

Many of the manufacturer's recommendations include denaturation of the protein of interest to effectively remove glycans. This is incompatible with signalling assays as it might affect the stability and/or correct folding of GP which is required to study the activation of the TLR4 signalling complex. To increase accessibility of the glycans for the enzymes, rather than denaturing GP by boiling and addition of denaturing buffer, 1.5 M urea was added to the de-glycosylation mixture. Due to the extensive glycosylation of GP it was hypothesised that a short incubation of max 2 hours would be insufficient to completely cleave the target glycans, especially in presence of urea. Therefore, the incubation time with the de-glycosylation mixture was increased to 16 hours at 37 °C with addition of a protease inhibitor cocktail. Rather than stopping the reaction by boiling the samples in SDS-PAGE buffer, the deglycosylated protein mixture was 0.22 µm filtered, 10 x diluted in saline solution and concentrated with 50 kDa cut-off Viva Spin centrifugation concentrators. The efficiency of de-glycosylation was assessed via SDS-PAGE.

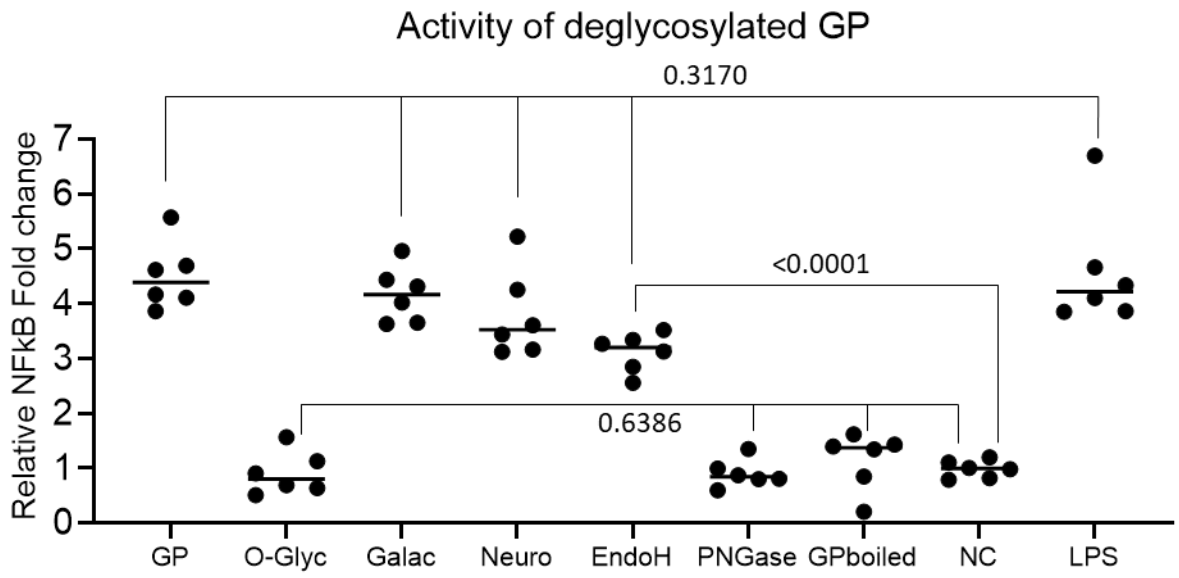


Figure 4.8: Complete removal of O- or N-linked glycans eliminates GP induced TLR4 activation. Purified, fully glycosylated GP was treated with deglycosylating enzymes and prepared for TLR4 signalling assays. A NFκB luciferase activity screen was conducted with stable TMC HEK293 cells and 1 µg/µl of the treated or untreated GP proteins. The negative control consisted of purified GP, buffer exchanged through a polymyxin B column to bind LPS, treated with proteases for 1 hour followed by boiling for 20 minutes. The negative control was set to 1-fold change and the values normalised to it. Data points present individual replicates. Statistical significance was assessed using a one-way ANOVA test. The enzymes' activity was assessed via gel shift SDS-PAGE assays similarly to Figure 3.6.

The effect on TLR4 activation post deglycosylation of GP was analysed with NFκB signalling assays and compared to untreated WT GP (Figure 4.8). The GP samples treated with O-Glycosidase and PNGase F showed a significant reduction in NFκB signalling compared to untreated GP and showed baseline activation at the level of the negative control. The samples treated with Galactosidase and Neuraminidase did not show a significant change compared to untreated GP. Surprisingly, the Endo H

treated sample did not show a significant reduction in activity although the rate of glycan removal is comparable with PNGase F.

Following the observation that the complete removal of both N- and O-linked glycans abolished TLR4-induced NFκB activity, the samples were further analysed to assess why the absence of glycans may influence GP's ability to activate TLR4. One explanation was that complete glycan removal may impact GP stability, in turn preventing the TLR4 activation seen following O-Glycosidase and PNGase F incubation. To investigate this hypothesis a proteolytic screen was next conducted

4.2.3 Partial proteolysis of GP reveals protease vulnerability post-deglycosylation

Removal of glycans of an extensively glycosylated protein such as EBOV GP could affect stability and other functions as the glycans may shield particular epitopes and hydrophobic residues. Such exposure of residues and/or increased instability following glycan removal can increase accessibility to proteolytic digestion. To analyse this effect in the case of GP, the two proteases Trypsin and Pepsin were chosen. Trypsin belongs to the serine protease family and cleaves exclusively c-terminal to arginine and lysine residues between the carboxylic acid group and amino group of the adjacent amino acid at pH 7.0 to 9.0 [156]. As the culture environment of most mammalian cells is at pH 7 or just below, using Trypsin to analyse the stability of GP is feasible, as it would be added into the cell medium. Such a pH reflects TLR4 interactions at the cell surface; on the other hand endo-lysosomes, the other site of TLR4-GP engagement, have a pH between 4.0 to 6.0 depending on the maturation state. Pepsin, an endopeptidase, and the principal protease of the stomach, cleaves broadly but favours phenylalanine and leucine at position P1, and is active at pH 1.5 to 5.0. Therefore, treating GP with Pepsin at acidic pH may be used to simulate the effect on GP stability when located in the endosome[157].

An assay to study the stability of GP post glycan cleavage was established and optimised. Limited proteolysis was conducted with trypsin and pepsin. Rather than of temporal variation the incubation time of 30 min at 37 °C was held constant throughout the assay but a dilution series of protease concentration was prepared ranging from 3 µg/ml to 1 mg/ml. To accommodate the active range of Pepsin the pH of samples treated by it was reduced to pH 4.5 to ensure a balance between protease activity and a non-denatured state of GP. The degradation of untreated GP under these conditions was compared with the cleavage after deglycosylation with O-glycosidase, PNGase F in light of the reduction of TLR4 activation and Endo H due to surprisingly unaffected TLR4 signalling despite similar cleavage patterns (see section 4.2.2.2).

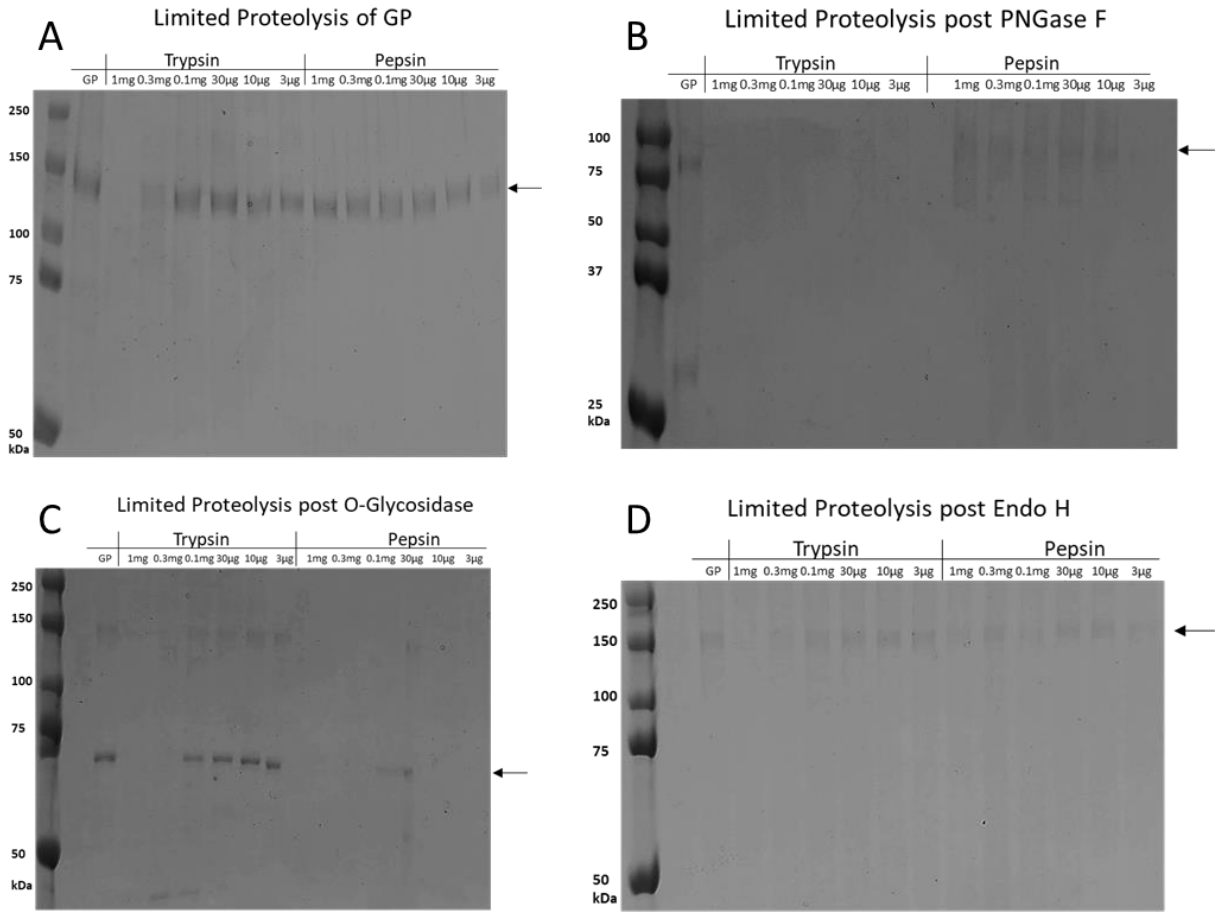


Figure 4.9: Partial proteolysis of GP post-deglycosylation with Trypsin and Pepsin. The impact of deglycosylation by PNGase F, O-glycosidase and Endo H on the stability of GP was analysed by utilising partial proteolysis by the proteases Trypsin and Pepsin, at pH 7.0 and pH 4.5 respectively. The GP samples treated with proteases under different concentrations were run on a gradient 6-16% SDS-PAGE under non-reducing conditions and stained with Coomassie. The concentrations of the proteases used for each sample are indicated by the value and amount/per ml for each lane. The detailed protocol is listed in section 7. The arrow indicates GP.

Limited proteolysis of native and glycosylated GP showed visible degradation of GP at Trypsin concentrations over 0.1 mg/ml while incubation with Pepsin did not yield any noticeable degradation (Figure 4.9A). GP which was previously treated with PNGase F degraded fully at even the lowest concentration of Trypsin while Pepsin degradation was unaffected when compared to untreated GP (Figure 4.9B). The treatment of GP with O-Glycosidase resulted in a heterogenic mixture of glycosylated GP just below 150 kDa and GP lacking O-glycans around 70 kDa. Upon proteolysis of this solution the protein started to degrade at Trypsin levels of 0.1 mg/ml and completely degraded at >0.3/ml mg and above, while any concentration of Pepsin protease led to the complete degradation of GP lacking O-Glycosylation. The sample with a Pepsin concentration of 30 µg/ml was deemed an artifact due to all other concentrations yielding complete GP degradation (Figure 4.9C). When incubating GP with Endo H, the protein band of GP at 150 kDa did not show any significant shift compared to untreated GP suggesting incomplete or inhibited enzymatic activity. Limited proteolysis

of Endo H treated GP showed no visible change to the patterns observed with untreated GP (Figure 4.9D). This particular assay was repeated more often than the usual triplication (n=10) with different batches of Endo H, all yielding the same result of no visible deglycosylation. Endo H removes all mannose-type N-linked carbohydrates and a lack of visible molecular weight band shift may indicate that the N-glycosylation is mostly comprised of complex hybrid glycans.

The observation of an Endo H resilience of GP was also noted by Volchkov et al who studied differential glycosylation of GP derived from infected cultures during the protein maturation and post-translational processing[158]. Here they realised that pre-GP in the ER is susceptible to Endo H evident by a shift from 110 kDa to 75 kDa while GP in the Golgi and after shows no change in molecular weight of 160 kDa. This discovery suggests that production and purification methodology established during this PhD project does not affect the GP glycosylation and mimics native and physiological GP.

To summarise the complete removal of glycans, both N- and O-linked makes GP more susceptible towards proteolytic degradation with the lack of N-glycans making the protein more susceptible to Trypsin while the lack of O-glycosylation significantly increases degradation by Pepsin. In an acidic environment, which may affect the structure and stability of GP generally, the shielding effect of the heavily O-glycosylated MLD seems to be more important than the N-glycans at the glycan cap, possibly due to the unstructured and flexible nature of the MLD domain. The reasons why Trypsin-mediated degradation of GP was not increased to a similar extent as the Pepsin degradation in an acidic environment post O-glycan cleavage are unknown. Perhaps the broad activity of Pepsin compared to the selective residue digestion of Trypsin is response for the phenotype. However, an EBOV GP deglycosylated to such an extent would not occur naturally and may not fully reflect the intricate interaction between GP and the TLR4 signalling complex.

4.2.4 Expression of EBOV GP in GNTI- HEK293 cells

To assess if the reduction in TLR4 signalling post glycan cleavage may be due to an artifact associated with the non-physiological state of GP, an alternative to express GP with reduced glycosylation without the use of enzymatic cleavage was explored. Additionally, as the complete removal of glycans destabilises GP, the effect of expressing GP with reduced glycosylation could contribute to further understanding of the protein.

For this purpose, the Expi293F GnTI- cell line was selected and the effect of expression of GP in these modified cells was explored. The Expi293F GnTI- cell line is derived from engineered Expi293F cells that do not have N-acetylglucosaminyltransferase I (GnTI) activity and therefore lack complex N-

glycans (Figure 4.10). The most common use of this cell line is in structural biology where the complete lack of glycosylation or at least high-yield expression of homogeneously glycosylated recombinant proteins is preferred for high-resolution protein structures.

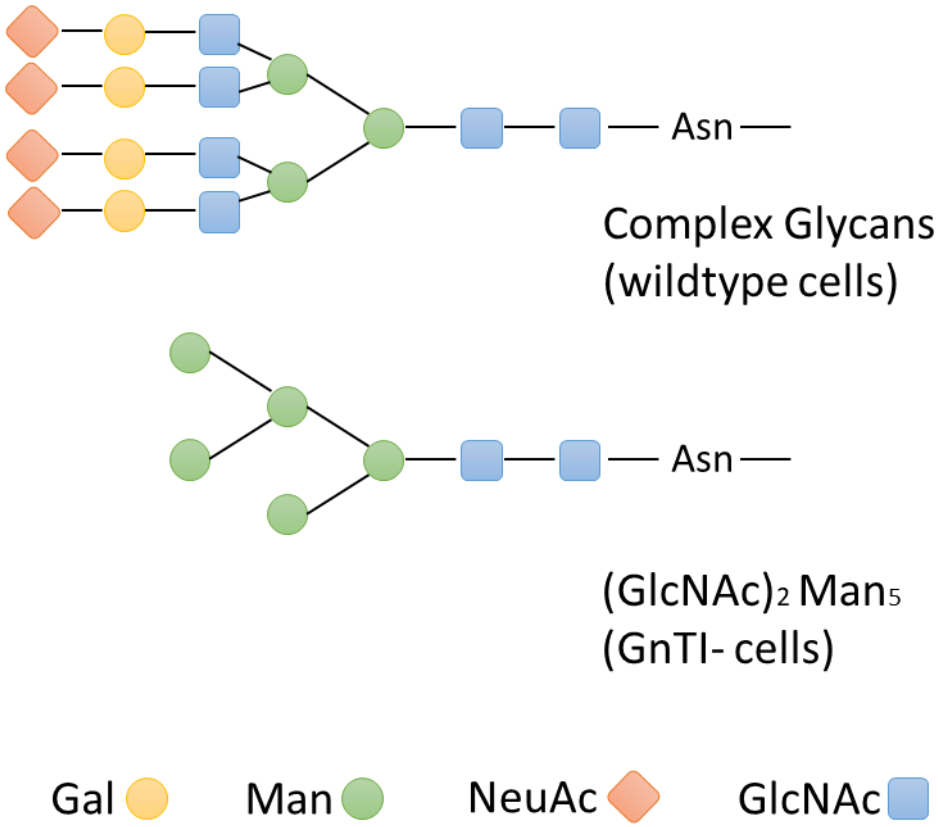


Figure 4.10: Schematic depiction of possible glycosylation in Expi293 GnTI- cells. Comparison of glycosylation in HEK293T cells and GnTI- HEK cells. GnTI: N-acetylglucosaminyltransferase I; Gal: Galactose; Man: Mannose; NeuAc: N-Acetyl-Neuraminic Acid; GlcNAc: N-Acetylglucosamine; Asn: Asparagine.

The maintenance and handling were identical to HEK293T cells apart from the culture media being F12 Ham instead of regular DMEM which is usually required for Expi293F GnTI- cell maintenance in suspension as the cytopathology of GP leads to the detachment of cells. The cell media containing GP was harvested and prepared for HisTrap excel purification as previously described (see Chapter 3). Post-purification a band at around 150 kDa was detectable in the elution fraction, later identified as GP in addition to contamination of serum albumin ~55 kDa (Figure 4.11).

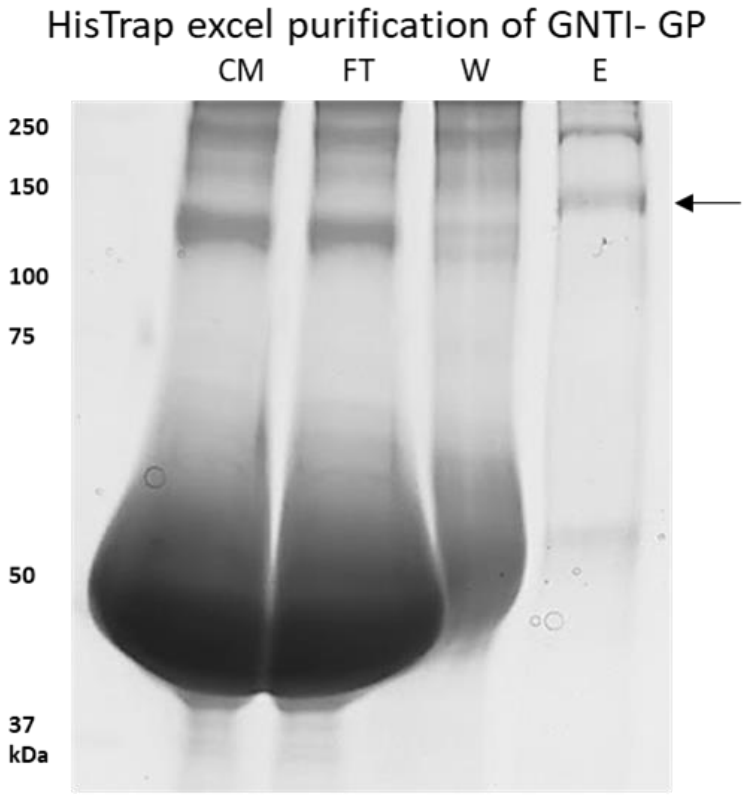


Figure 4.11: HisTrap excel purification of GP expressed in Expi293 GnTI- cells. Expi293 GnTI- cells were used to express GP, the medium was collected 4 days post-transfection and purified according to the same protocol as with HEK293T cell expressed GP. The fractions of the HisTrap excel IMAC purification were run on a gradient 6-16% SDS-PAGE under non-reducing conditions and stained with Coomassie. CM, cell medium containing GP protein; FT, column flow through after sample application; W, column was steps; E, pooled elution fractions. The detailed protocol is listed in Chapter 7. The arrow indicates GP.

Due to the absence of GnTI activity HEK293 GnTI- cells are less resilient and require more care. Additionally, the overall protein yield is lower, and the cells require 10% FBS within the media formulation for maintenance and protein expression which resulted in the increase in serum albumin contamination in the HisTrap excel column IMAC purification of GP (Figure 4.11).

The ability of purified GNTI- GP to activate TLR4 was compared to HEK293T derived GP via a NFκB signalling assay. The difference between the two protein samples was deemed insignificant (Figure 4.12). A change in glycosylation from complex glycans to simple mannose glycosylation did not affect the mechanism by which GP activates TLR4. Seemingly, GP does not possess significant amount of mannose glycans, but rather complex N-linked glycans and even limiting the glycosylation to strictly N-linked high-mannose glycans does not affect protein stability or TLR4 activation.

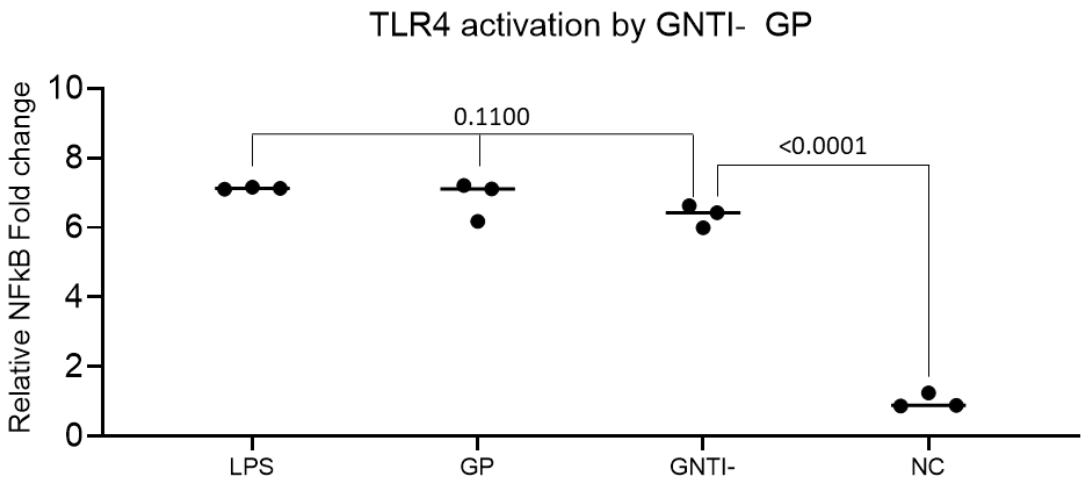


Figure 4.12: GP derived of Expi293GnTI- cells shows no change to wild-type GP TLR4 activity. GP was expressed in HEK293T cells and Expi293GnTI- cells, purified and the ability to activate TLR4 was analysed. A NFκB luciferase activity screen was conducted with stable TMC HEK293 cells and 1 µg/µl of the GP proteins. The negative control consisted of purified GP, treated with proteases for 1 hour followed by boiling for 20 minutes. Data points represent means of triplicates. The negative control was set to 1-fold change and the values normalised to it. Statistical significance was assessed using a one-way ANOVA test.

4.3 Discussion

Following the establishment of expression and purification methodology in Chapter 3, the section presents the approaches taken to study the minimal requirements of GP to activate TLR4 with a focus on the glycosylation patterns and complexities.

The expression and activity analysis of the GP truncation constructs showed that a fundamental sequence alteration or mutation destabilises and affects parts of the post-translational modification process. Individual point mutations of EBOV GP can affect protein expression and stability, an observation which was confirmed through personal correspondence with Erica Ollmann Saphire at The Scripps Research institute. Treatment of GP with deglycosylating enzymes revealed the type of glycans present and that partial removal of glycans does not affect stability or signalling but the complete removal with O-glycosidase or PNGase F leads to a reduction in TLR4 activity and protein stability. It may well be that a complete lack in glycosylation affects protein stability, the protective glycan shield of GP to such an extent that the protein degrades in presence of proteases before TLR4 can be activated. However, it was not possible to study the effect of complete deglycosylation on GP during the expression process and expression was limited to Expi293 GnTI- cells. Investigating how the lack of glycans affects GP protein stability could help further to understand the reason behind reduced TLR4 activation after removal of all O and N- linked glycans on GP. Unfortunately, expressing full length GP in an expression system lacking glycosylation was not possible during this dissertation and would have required further years of optimisation. The research group of Han Xia at the Wuhan Institute of

Virology designed and optimised a way to express EBOV GP in bacterial expression systems and attempts were made to receive samples of the published constructs, however due to the global pandemic these efforts were fruitless. Nevertheless, this remains an avenue to be explored and addressed in the future.

5 Biochemical and functional assays of GP

In the previous Chapter the contribution of GP glycosylation in TLR4 activation was explored and showed that the complete removal of N- or O-glycosylation of GP results in elimination of TLR4 signalling but also significantly destabilises GP and makes it more susceptible to degradation. Next, I investigated the mechanism by which GP activates TLR4. As explained in the introduction (section 1.2.3) the best understood and most studied ligand-receptor interaction of TLR4, is the mechanism by which TLR4 can recognise LPS and how the LPS molecule binds the hydrophobic pocket of the TLR4/MD2 complex. Due to the vast chemical and structural differences between LPS and EBOV GP, a similar activation mechanism is unlikely, but not impossible, whereas a potential novel mechanism of activation may be a possibility. Understanding the interaction in as much detail as LPS activation of TLR4, would widen our understanding of EVD and in turn provide therapeutic insights.

Although both LPS and GP contain saccharides, PAMPs and endogenous glycans in respective order, the structural domain of LPS, which is the main contributor to the interaction of LPS and TLR4/MD2 complex, is the hydrophobic lipid A region, with its acyl chains and not the branched carbohydrate region [63]. On the other hand, the significant glycosylation of EBOV GP makes it likely that the carbohydrates contribute to the mechanism of TLR4 activation. This hypothesis is explored in this Chapter, and with approaches taken to increase the understanding of the activation of TLR4 by GP, with the final aim to characterise the mechanism or interaction.

5.1 Results

5.1.1 Dose response of TLR4 to GP

To investigate whether the activation of TLR4 by GP follows a linear dose-response similarly to LPS[63], stable TMC HEK293 cells were incubated with a varying amounts of GP. In this assay the cells plated in 96-well plates were subjected to GP protein ranging from 0.5 to 2.5 µg and 100 ng/µl LPS for 24 hours. A reason for the extended induction period was to investigate the potency of GP over time within the cell culture. Incubation with the lowest amount of GP lead to a 10-fold increase in NFκB signalling while the highest amount of GP increased signalling 25-fold (Figure 5.1). Although doubling the amount of GP ligand does not yield double the TLR4 activation, the overall trend follows a linear progression. Presence of anti-TLR4 antibody during the incubation results in a significant reduction in activity, confirming that the NFκB activity seen is driven by TLR4 activation.

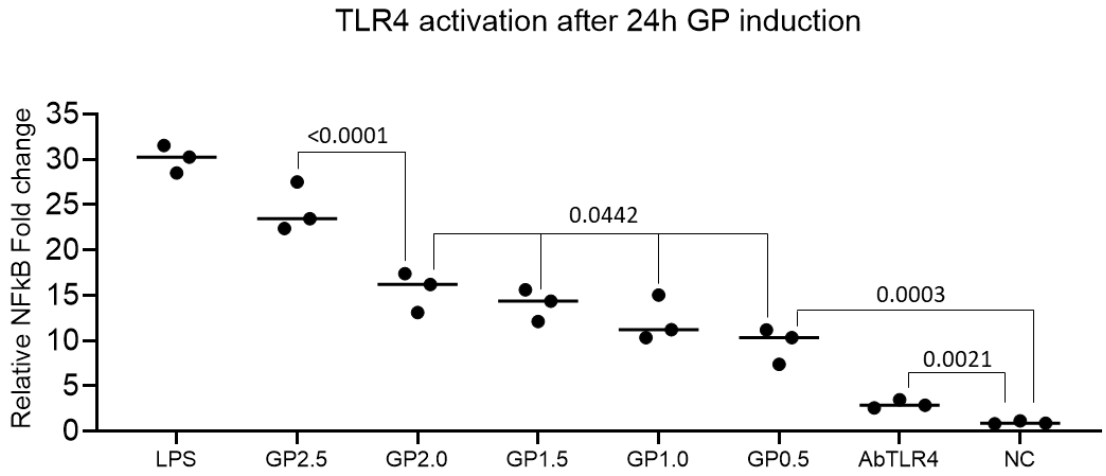


Figure 5.1: Dose response of TLR4 to GP over 24 hours. A NFκB luciferase activity screen was conducted with stable TMC HEK293 cells. A range of amounts of 0.5µg to 2.5µg of purified GP, LPS 100ng/µl was used in the assay with a 24-hour induction time. Additionally, 5µg of anti-TLR4 were added to replicates with 2.5 µg/µl of GP. The negative control consisted of purified GP, treated with proteases for 1 hour followed by boiling for 20 minutes. The negative control was set to 1-fold change and the values normalised to it. Data points represent means of triplicates. Statistical significance was assessed using a one-way ANOVA test.

Additionally, it became evident that induction of cells with native GP over 24 hours induces increased TLR4 activation compared to a shorter six-hour incubation period. Considering the increase in cell stress, cytokine activation and finally a cytokine storm in an infected patient these findings are plausible. Next, the similarity between the activation mechanism of LPS and GP was compared with TLR4 specific inhibitors.

5.1.2 Pre-treatment with LPSRS reduces TLR4 activation by GP

To further investigate the mechanism by which EBOV GP activates TLR4, the possible similarity to the mechanism of activation by LPS was explored, using TLR4 specific inhibitors, targeting the ligand binding domain. Although many TLR4-specific inhibitors exist, the majority target the intracellular TIR domain leading to ligand-independent inhibition. However, to investigate the mechanism of activation (MoA) of GP, a ligand specific antagonist of TLR4 is necessary. To achieve this the TLR4 antagonist LPSRS was used, a LPS isolated from *Rhodobacter sphaeroides* capable of binding the hydrophobic pocket of the TLR4/MD2 complex without receptor activation, hypothesised to be due to the difference in acyl chains to lipid A. Consequently, pre-treatment with LPS-RS inhibits LPS specific activation of TLR4. It was hypothesized that if incubation of the stable TLR4/MD2/CD14 expressing HEK293 cells with LPSRS prior to exposure to GP would reduce the NFκB activation, then the mechanism of activation by GP may be like the LPS-associated mechanism.

To investigate this, a NFκB signalling assay was used using stable TMC HEK cells. The cells were pre-treated with three different concentrations of LPSRS, (100 ng/µl, 300 ng/µl,1000ng/µl), 1 hour prior to a 6-hour TLR4 agonist induction. Subsequently, cells were treated with either LPS and GP at two different concentrations each, 10 ng/µl and 100 ng/µl; 100 ng/µl and 500 ng/µl respectively and their NFκB activity assessed relative to a non-treated control. The rationale of using GP and LPSRS in excess compared to LPS, was to identify competitive binding and possible mechanistic similarities between GP and LPS. Pretreatment with LPSRS, lead to the reduction in NFκB signalling for all samples (Figure 5.2).

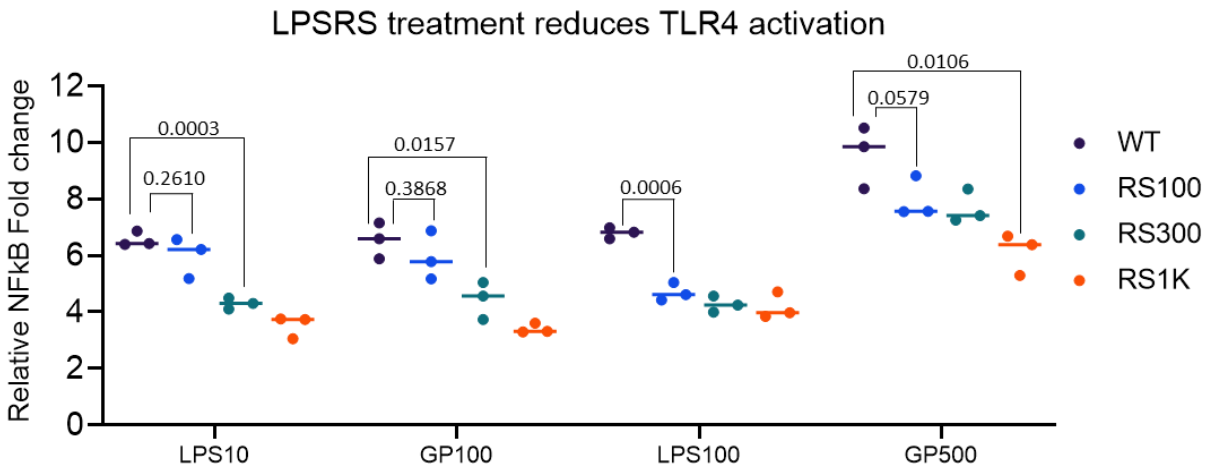


Figure 5.2: Pre-treatment with LPSRS reduces TLR4 activity for LPS and GP. The specific inhibition of LPS and GP induced TLR4 signalling by LPSRS was investigated using the stable HEK293 cells. A NFκB luciferase activity screen was conducted with stable TMC HEK293 cells. The cells were incubated with LPSRS at concentrations of 100 ng/µl, 300 ng/µl and 1 µg/µl one hour pre-induction with the agonists. The cells were stimulated with LPS concentrations of 10 ng/µl and 100 ng/µl, and 100 ng/µl and 500 ng/µl of GP for six hours and NFκB activity was assessed. The measurements were normalised to the the levels of the negative control.The negative control consisted of purified GP, treated with proteases for 1 hour followed by boiling for 20 minutes. The negative control was set to 1-fold change and the values normalised to it. Statistical significance was assessed using a one-way ANOVA test.

The signalling for LPS concentrations was significantly reduced with 300 ng of antagonist, while despite the excess amount of 100 ng/µl of GP, significant reduction was achieved with minimal amount of LPSRS. The evolution of TLR4 to be sensitive to LPS may be the reason for the different rate of signalling reduction. It is noteworthy, that a 10-fold increase in antagonist amount did not yield a 10-fold decrease in signalling activity. It may be possible that treatment of 1 hour only allowed a fraction of TLR4 receptors to be occupied by LPSRS and to achieve higher levels of signalling reduction longer incubation times may be necessary.

Interestingly, while both LPS concentrations showed almost identical signalling patterns despite a 10-fold increase in LPS, the conditions with 500 ng/µl GP showed increased NFκB activation compared to

lower GP concentrations and signalling was only significantly reduced at pre-treatment with an excess of 1µg of LPSRS.

Both, LPS and GP mediated TLR4 signalling could be significantly reduced by pre-treating the cells with LPSRS, albeit only with 1µg of antagonist in some conditions. This similarity suggests that GP may have a TLR4 activation mechanism closely related to LPS, despite the structural and chemical differences. Additionally, as LPSRS is specific to the TLR4/MD2 hydrophobic pocket, the significant reduction of GP-mediated signalling suggests that MD2 is necessary in this mechanism. Therefore, it may be possible that a hydrophobic region of GP interacts with the hydrophobic pocket of the TLR4/MD2 complex.

Table 5.1. Hydrophobic regions of GP analysed by Expasy ProtScale. The approximate residues part of the hydrophobic region are listed alongside the location of known domains.

Hit Number	Residues	Regions/Domains
1	20-29	Signalling Peptide
2	177-187	Receptor binding domain
3	478-493	C-terminal Mucin-like domain
4	527-535	Internal fusion loop
5	654-672	Transmembrane domain

To explore this hypothesis, the sequence of GP was analysed for hydrophobic regions, particularly located in regions that may be accessible through a conformational change. The notable domains and regions are summarized in Table 5.1, as identified via the Expasy Protscale software using Kyte & Doolittle scale for hydrophobicity.

Of the hydrophobic regions, the signalling peptide (Hit 1) and the transmembrane domain (Hit 5) were immediately disregarded due to the signalling peptide being cleaved in the native protein and the TMD being inaccessible due to immobilisation within the cell or virion membrane. The putative receptor binding domain at Hit 2, would be inaccessible too as it is located in the core of the trimeric spike GP. Although the region of hit 3 is located at the peak of the propellor-like mucin like domain, it would be natively shielded by glycans, also making it inaccessible. Lastly, hit number 4 which is part of

the internal fusion loop (IFL) required for membrane fusion and cell infection, is situated roughly in the centre the spike protein at the stalk. Due to the flexible nature of the loop and the accessible location without glycans in immediate proximity or other shielding domains, this may qualify as a hydrophobic and accessible interaction partner of the TLR4/MD2 complex, and thus was explored further to enhance our knowledge of TLR4 activation by GP.

5.1.3 The internal fusion loop of GP as a possible interaction region with MD2 during TLR4 ligand-binding

To further explore this hypothesis a mutant of GP was created lacking the hydrophobic part (red underscored) of the internal fusion loop (deleted hydrophobic region of fusion loop in red) GAAIGLAWIPYFGPAA, termed GPΔFuLo. The GPΔFuLo mutant was successfully expressed and purified following the same procedure and optimisation process as WT GP that was developed in Chapter 3.

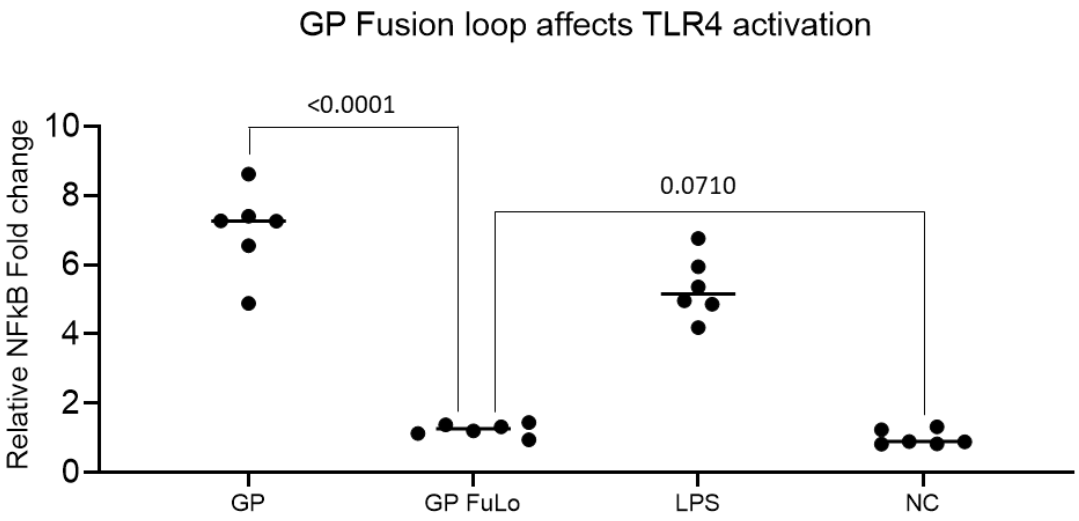


Figure 5.3: Deletion of the GP IFL diminishes TLR4 activity compared to WT GP. The ability of GPΔFuLo to activate TLR4 was assesses and compared to WT GP with stable TMC HEK293 cells. A NFκB luciferase activity screen was conducted with stable TMC HEK293 cells. The cells were incubated with 100 ng/µl LPS and 25 ng/µl of GP for six hours and activity was assessed. The negative control consisted of purified GP, treated with proteases for 1 hour followed by boiling for 20 minutes. Data points present individual replicates. The negative control was set to 1-fold change and the values normalised to it. Statistical significance was assessed using a one-way ANOVA test.

Once purified 200 ng/µl of the GP protein was incubated with stable TMC HEK293 cells and the relative levels of NFκB signalling were compared to WT GP. Compared to WT GP which induced a 7-fold increase in signalling compared to untreated controls on average in the assay, GPΔFuLo showed no significant change to the normalised negative control and was deemed inactive(Figure 5.3). The results

indicate that the IFL of GP affects TLR4 signalling, and a deletion leads to the loss of TLR4 activation. To further investigate the mechanism by which GP activates TLR4, co-immunoprecipitation was utilised to study the effects of glycan alteration and GPΔFuLo on the interaction between the signalling complex and GP.

5.1.4 Co-immunoprecipitation of GP and TLR4/MD2

The complete removal of O- and N-glycans (Chapter 4) or the mutation of the internal fusion loop abolished TLR4 signalling. To further study if these alterations also affect the interaction between EBOV GP and the TLR4/MD2 complex a co-immunoprecipitation protocol was established. For this, HEK293T cells were cultured and seeded onto 6-well plates and co-transfected with GP, TLR4 and MD2. The medium and cells were collected separately 48 hours post-transfection. The medium was cleared and concentrated with a 30 kDa cut-off, while the cells were washed, dissociated, and lysed with HEPES lysis buffer (4 °C, 1 h). As GP is secreted while TLR4 and MD2 are mostly likely found in the cell lysis fraction, the concentrated medium and cell lysate were mixed prior to immunoprecipitation. To confirm expression, and proof of concept of the co-IP protocol, the lysate mix was immunoblotted for anti-FLAG tagged TLR4, anti-myc-tagged MD2 and anti-His tagged GP and normalised to an actin loading control. While TLR4, MD2 and actin were detected by western blotting, GP was not detected in sample (Figure 5.4). Despite multiple replications, it was not possible to detect GP in the lysate-medium mixture. It was hypothesised that the amount of medium collected was insufficient, compared to the overexpression of TLR4 and MD2 in the cell lysis. Therefore, as an alternative to the standard immunoprecipitation protocol using cell lysate and mixing with concentrated culture medium, purified proteins were used to increase the amount of available and detectable protein. Furthermore, as TLR4 is not known to be lectin-like, capable of binding glycans, GP may interact with an unknown protein which forms a complex with TLR4.

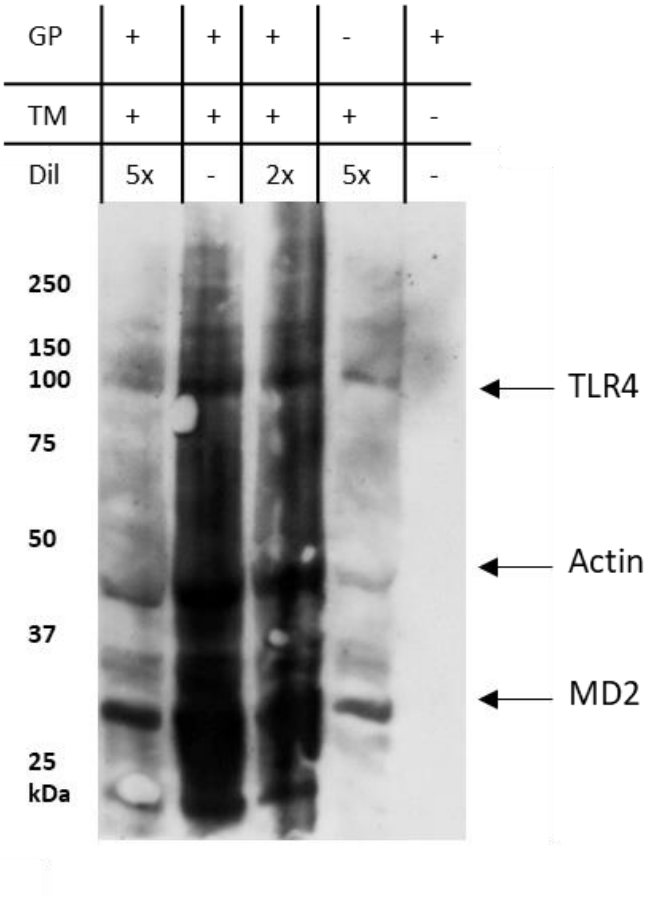


Figure 5.4: TLR4, MD2 and actin but not GP were detected in the cell lysate-mix. Prior to immunoprecipitation to identify the interaction between GP and TLR4/MD2 the cell lysate/input analysed regarding presence of target proteins. Four days post-transfection the medium and the cell lysate were harvested, cleared, mixed, and run on a gradient 6-16% SDS-PAGE under non-reducing conditions. Western blotting was performed with anti-His (GP), anti-TLR4, anti-MD2 and anti-actin. GP was not detected while TLR4, MD2 and actin were. The detailed protocol is listed in section 7. The arrow indicates the proteins targeted by antibodies. GP: fraction containing GP; TM: fraction containing TLR4 and MD2; Dil: dilution of the lysate mix before SDS-PAGE and western blotting.

Consequently, TLR4 can only be activated in presence of the glycan-binding protein. Therefore, using cell lysate may falsely suggest binding of GP and TLR4/MD2, however in fact an unknown protein might be the cause for this 'interaction' but remains undetected by western blotting due to antibodies specific against affinity tags and target proteins. Using purified proteins had the advantage to unambiguously identify protein interaction, not requiring western blotting and antibodies, avoiding artifact formation which may be caused by unknown proteins co-immunoprecipitated with the complex by not detected. Additionally, this method allowed the study of the effect of GP deglycosylation on the interaction with TLR4/MD2 which would otherwise not have been possible within cell lysate. As deglycosylated GP is prone to degradation as indicated by partial proteolysis, addition of deglycosylated GP to cell lysate may lead to the degradation.

To prepare for immunoprecipitation, purified GP constructs, enzymatically treated GP, and commercial TLR4 and MD2 were incubated overnight at 4 °C in HEPES lysis buffer and a sample for IP

input was taken. The following day anti-TLR4 or anti-MD2 antibodies were added at the protein mixture and incubated for 4 hours before adding pre-equilibrated protein A/G beads and incubating for an additional 2 hours. After washing with HEPES buffer, the samples were eluted with an acidic glycine solution and immediately neutralised with NaOH. Elution through boiling in SDS-PAGE sample buffer was avoided due to possible contamination by antibodies and protein A/G.

The five samples, GP treated with ENDO H; treated with PNGase F; treated with O-Glycosidase; GP expressed in HEK293 GnTI- cells and GPΔFuLo were analysed for their interaction with either TLR4 or MD2 separately and compared to WT GP (Figure 5.5). For each IP, the input samples were run on a separate gel prior to antibody incubation and non-specific binding of the GP samples to the beads was assessed by performing IP in absence of TLR4 or MD2. Immunoprecipitation revealed the binding of GP to MD2 and TLR4 as indicated by the visible band of around 150 kDa. No unspecific binding was detected. Despite the difference in TLR4 activation post-deglycosylation, all GP glycosylation samples, including GnTI- derived GP were able to bind both TLR4 and MD2, as indicated by bands at 150 kDa for GTI- GP, and ENDO H treated GP, 75 kDa for PNGase F treated GP and between 50 and 75 kDa for O-Glcyosidase treated GP.

The only exception was GPΔFuLo where a faint band was detected associated with a molecular weight of ~150 kDa in the TLR4 IP and no band in the MD2 IP. This particular result for GPΔFuLo could not be replicated, and all replicates did not show co-immunoprecipitation with TLR4 nor MD2, therefore this particular result was identified as an artifact, most likely resulting from neighbouring lane contamination. The results are summarised in the Table 5.2.

Table 5.2: **Summary of the interaction results from Co-immunoprecipitation.** The construct GPΔFuLo (grey) showed no interaction with TLR4 nor MD2.

Sample	TLR4 interaction	MD2 interaction
GP	Yes	Yes
GP GNTI-	Yes	Yes
GPΔFuLo	No	No
GP PNGase F	Yes	Yes
GP ENDO H	Yes	Yes
GP O-Glyco	Yes	Yes

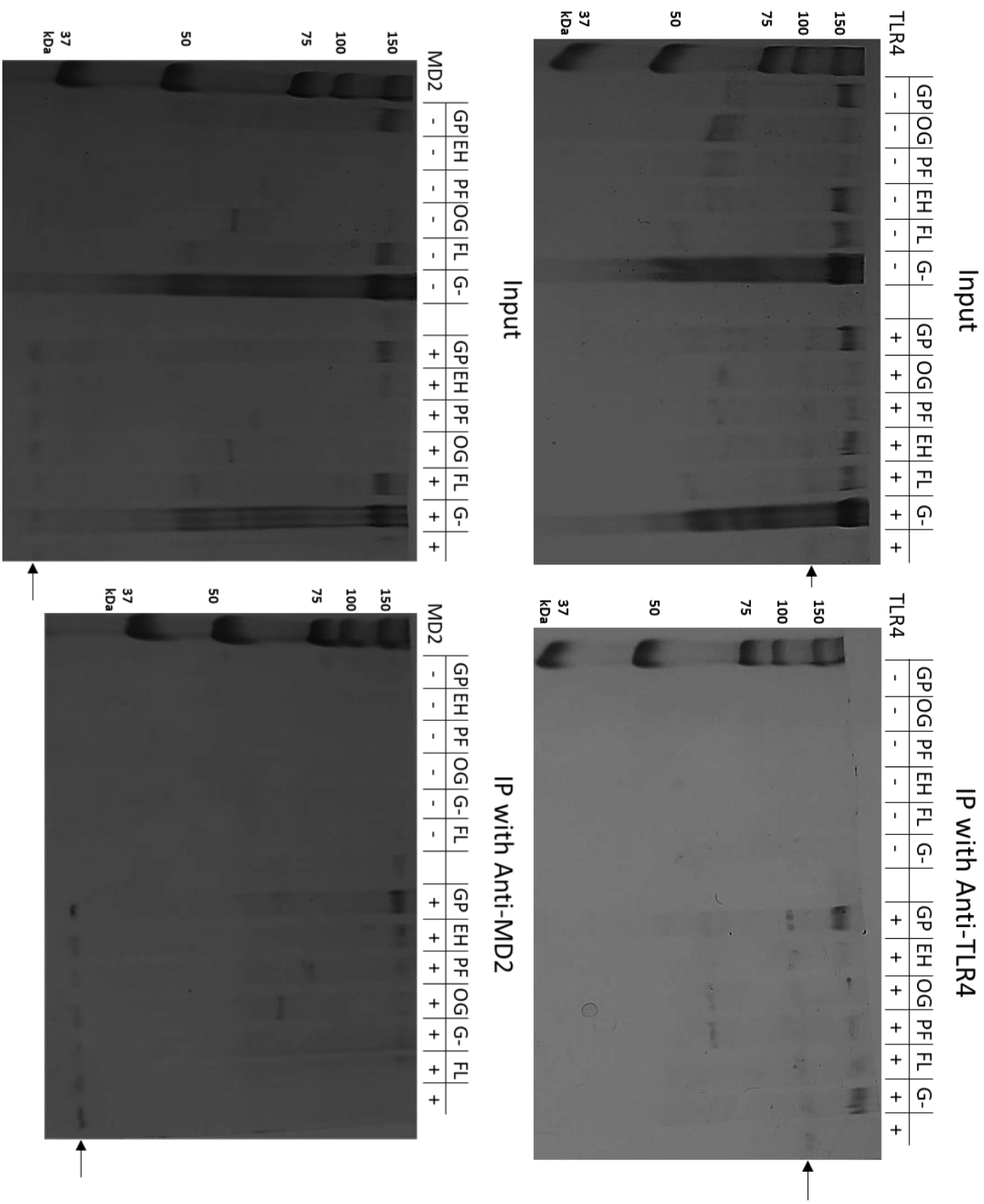


Figure 5.5: Purified GP species were immunoprecipitated with TLR4 and MD2 was the exception of GPΔFuLo. Immunoprecipitation of TLR4 and MD2 was conducted to investigate the interaction with GP and co-immunoprecipitate different GP species. Prior to immunoprecipitation 2µg of TLR4 and MD2, and 5µg of GP were incubated in the IP solution and immunoprecipitated with either anti-TLR4 or anti-MD2 and protein A/G beads. The eluted samples were collected and run on a gradient 6-16% SDS-PAGE under non-reducing conditions and stained with Coomassie. All GP species were co-precipitated with TLR4 or MD2, without any unspecific binding to the beads with the exception of GPΔFuLo which did not bind TLR4 or MD2. The '+' indicates the presence of TLR4 or MD2 whereas the '–' indicates the absence. Black arrow indicates TLR4 and MD2 in respective order. The detailed protocol is listed in section 7. GP, wild-type glycoprotein; OG, GP protein

after treatment with O-glycosidase; PF, GP after treatment with PNGase F; EH, GP after treatment with Endo H; FL, GP with deleted internal-fusion loop GPΔFuLo; G-, GP expressed in Expi293 GnTI- cells.

These results suggest that the IFL of GP plays a vital role in the activation of TLR4 but also the interaction with TLR4 and MD2, unlike the glycosylation of GP whose removal does not impact binding. It may be that the hydrophobic region of IFL mimics the acyl chains of LPS, leading to the activation of the TLR4 signalling complex. Due to the hydrophobic nature of IFL, there may be an unknown further function of lipid binding which is responsible for activating TLR4. Secreted viral proteins such as Dengue NS1 is capable of binding lipids, activating TLR4 and can be blocked with LPSRS [159–161]. Perhaps a similar mechanism akin to GP with hydrophobic domains takes place here. Further, investigation into this phenomenon is required, however, it will be addressed in future research projects.

5.1.5 Reducing TLR4 activation by inhibiting GP shedding with TACE inhibitors

During an active EBOV infection, the virus replicates in cells and GP is subsequently shed from cells and virions alike due to cleavage by endogenous TACE[162]. These shed GP molecules diffuse through the blood stream and can activate TLR4 on cells remote from the infection site. Theoretically, inhibition of TACE could restrict the distribution of GP throughout the host, limit the TLR4 activation to the site of infection and significantly reduce cytokine and chemokine activation. TACE or ADAM17 inhibitors have been used in clinical trials for cancer treatment and are commercially available[163].

To investigate this hypothesis a cell-based co-culturing and inhibition assay was created. HEK293T cells were cultured and transfected with full-length GP plasmid including the transmembrane domain. Additionally, as a negative control to assess TACE expression on its own cells were also transfected only with TACE-containing plasmids. In parallel, stable TMC HEK293 cells were cultured and prepared for NFκB signalling assays in 96 well plates. 24 hours post-transfection the cells were washed, carefully dissociated with TrypLE Express, and the stable cell culture and the GP-expressing culture were merged. The medium of selected wells was supplemented with 10 uM TAPI-1, a TACE specific inhibitor. TrypLE Express was chosen as a dissociation agent as it acts as a Trypsin replacement enzyme while simultaneously not degrading surface proteins, such as GP, which may be vulnerable to proteases as indicated in Chapter 4.

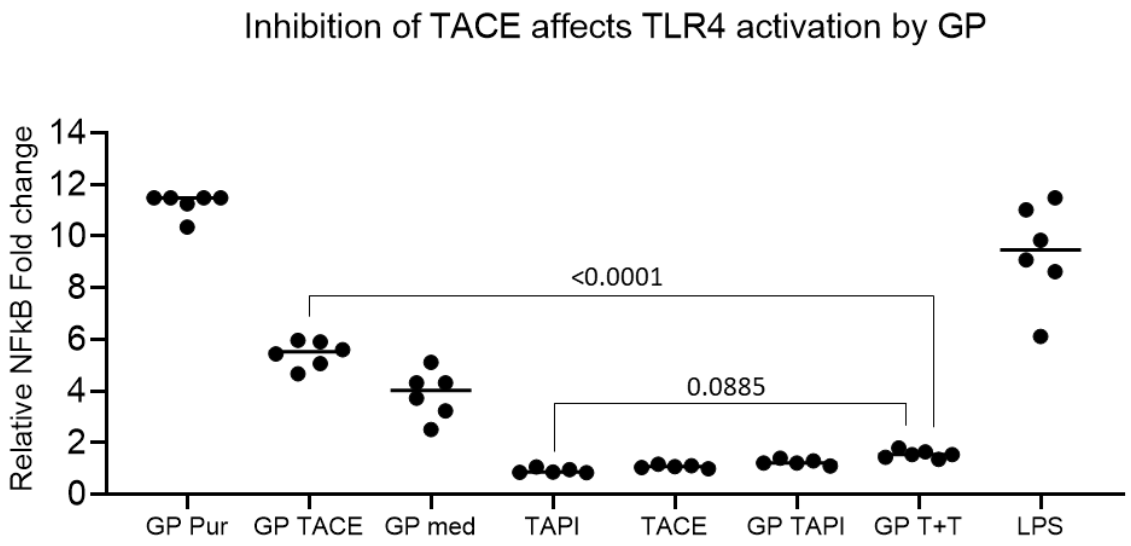


Figure 5.6: Inhibition of TACE eliminated GP-mediated TLR4 activation. The effect of inhibiting TACE activity on the secretion of GP and the activation of TLR4 was investigated. HEK293T cells were transfected with plasmids carrying full-length GP (including TMD) and/or TACE. Post-transfection induction the cells were washed and dissociated with TrypLE Express and merged with cultured Stable TMC HEK293 cells. The TACE inhibitor TAPI-1 was added to selected wells at 10 uM concentrations. The two cell lines were co-cultured for 24 hours and harvested for assessment of NFκB activity. T+T, TACE and TAPI. Expression of TACE alone, and transfection with an empty vector together with TAPI served as the negative control. The negative control was set to 1-fold change and the values normalised to it. Data points present individual replicates. Statistical significance was assessed using a one-way ANOVA test.

The two cell lines were co-cultured for 24 hours, samples were harvested and prepared for NFκB signalling assays as previously described (see methods Chapter 7). As most signalling assays were performed with purified GP instead of direct application of culture medium containing GP, a control with purified GP was included in the co-culture for signal strength comparison. Purified GP showed a similar NFκB induction to LPS of about 10-fold signal increase, levels which were routinely observed in other assays (Figure 5.6). The samples with co-cultured cells expressing GP alone showed up to 5-fold induction of NFκB, while the co-expression with TACE yielded up to 6-fold induction, although no statistical distinction between both. As hypothesised the treatment with TAPI-1 eliminated GP-mediated NFκB activation, even with co-expression of TACE, resulting in no-significant change from the negative control.

5.2 Discussion

In this Chapter I present the approaches taken to investigate the MoA of TLR4 by GP. Firstly, it was confirmed that GP activates TLR4 in a dose-dependent manner, similarly to known agonists such as LPS. The finding that TLR4 activation by GP can be reduced by pre-treatment with the antagonist LPSRS, which blocks the activation mechanism of LPS, specifically suggested a related MoA for GP. Sequence analysis for hydrophobic regions of GP which may mimic the acyl chain of LPS was undertaken and the IFL was marked as a potential target. A deletion of this hydrophobic fusion loop

of GP resulted in a loss of function and no measurable increase in TLR4 activation, suggesting the IFL loop as a key interaction partner in TLR4 activation and thus also a potential therapeutic target for EBOV. Inhibition of TACE resulted in elimination of shed GP, keeping the protein immobilised at the membrane, linked to a fundamental reduction in TLR4 signalling, which would usually be detrimental to the host during EBOV infection. This Chapter therefore presents the identification of three distinct vulnerabilities of GP: LPSRS, GP-IFL and TACE which may be used to target or treat during EVD.

Following this discovery, the interaction between GP and the TLR4/MD2 signalling complex was investigated via immunoprecipitation. The different species of GP with altered glycosylation pattern, discussed in Chapter 4, were also included in this analysis regarding their ability to interact with TLR4 and MD2. Although, previous publications suggest that the glycans of GP are necessary for the activation of TLR4, due to deglycosylation of GP resulting in significant reduction in signalling, this was identified as a possible misinterpretation of results[20]. In fact, the lack of GP glycans does result in reduced TLR4 activation but most likely due to the degradation and instability of GP in culture as the interaction between TLR4 and MD2 remains unaffected. Furthermore, deglycosylated GP would not be found natively in an infected host and does not depict the dynamics of a real infection well. However, the deletion of the IFL of GP, results in both, reduction in TLR4 activation and the elimination of GP interaction with both TLR4 and MD2.

In a final experiment, the effect of TACE inhibitors was investigated, resulting in the inhibition of GP shedding and consequently, activation of remote TLR4 receptors in culture. Further studies are required but these findings indicate possible therapeutic targets in the internal fusion loop of GP, inhibition of TLR4 signalling via LPSRS and using TACE/ADAM17 inhibitors to eliminate shed GP and system-wide activation of TLR4.

6 Overall conclusions and future objectives

6.1 General aim and objectives

In this dissertation I have outlined the approaches taken to investigate the activation of the TLR4 signalling complex by the Ebola Virus Glycoprotein. This is an unmet need in the field, as understanding the precise biochemical mechanisms by which EBOV produces the inflammatory signalling and the fatal "cytokine storm", allows us to identify the vulnerabilities of viral infection and also expand our mechanisms of treating it. More generally, this broadens our understanding of the ways in which TLRs interact with ligands, expanding our knowledge of general themes that can be applied to other contexts. The study was primarily focused on the glycosylation of GP and its role in the activation of TLR4, based on previous findings that treatment of GP with deglycosylation enzymes can significantly reduce activity[20].

Towards this goal strategies to express and purify extensively glycosylated and cytotoxic EBOV GP with high yields, while maintaining the native and active state of the protein were developed. Although established protocols exist, alternative approaches needed to be developed and optimized, as existing methods either express significantly modified versions of EBOV GP or require safety precautions up to BSL4 due to the propagation of live virus in mammalian cells/animals or using virus-like particles[164]. The methodology established during this study to express and purify EBOV GP, has several novel advantages such as the accessibility due to minimal requirements to express and purify; affordability compared to previously established techniques; and lastly the ability to use the protein in assays requiring active GP. Special attention was paid to avoidance of LPS contaminations which would lead to incorrect results when studying the immune signalling induced by GP, as LPS is also able to bind and activate TLR4. Additionally, as the expression method was virion and virus-free it may not require a BSL2 tissue culture room, subject to the regulations of the respective country, again improving the accessibility of a previously restricted protocol. Although this protocol required years of optimization, native and fully glycosylated GP was fully expressed; modifications to this protocol can help similarly difficult to express proteins become more routine and accessible. Following the creation of this protocol, the recombinant GP was successfully expressed and purified to investigate the activation of the TLR4 signalling complex.

6.2 Discussion

6.2.1 Investigation of GP glycosylation and the contribution to TLR4 activation

Previous studies in the field have indicated that glycosylation of GP was thought to be specifically important in the activation of TLR4 as enzymatic glycan removal or mutations of glycosylation sites of GP resulted in reduced TLR4 activation[20]. As TLR4 is not known to have a lectin-like function, i.e. the ability to bind and be activated by carbohydrates, this hypothesis was novel and possibly linked to a new MoA.

During this PhD it was identified that alteration of GP glycosylation does affect TLR4 signalling, however not due to the activation of TLR4 through glycosylation directly. Natively, the glycans form a protective shield around the GP and upon removal or mutation of them, the protein becomes unstable and vulnerable to enzymes, resulting in degradation and loss of activity. This phenotype was falsely identified as TLR4 activation by GP being dependent on glycosylation. This hypothesis was confirmed by immunoprecipitation of TLR4 and GP species with altered or removed glycans showing that glycosylation is not required in the interaction with TLR4. Therefore, it can be concluded that TLR4 is not a lectin and signalling assays alone may lead to wrong presumptions and need to be supplemented with biochemical and biophysical experiments.

6.2.2 Effectiveness of LPSRS on TLR4 activation by GP suggests a similar activation mechanism to LPS

Having shown that GP glycosylation does not directly interact with TLR4, the mechanisms of TLR4 activation by GP were explored. For this, target-specific inhibitors were used to identify or narrow the possibilities. Inhibitors which are TLR4 specific but not target specific such as Tak242, which targets the intracellular TIR domain, leading to inhibition of all TLR4 signalling were considered but would not have contributed to further understanding of the MoA of TLR4 by GP, as they would not allow differentiation between the ligand-receptor binding site [165]. Therefore, LPSRS, which is an antagonist specific to the MoA of the best studied ligand of TLR4, LPS, was used to study the mechanism associated with GP. Using LPSRS to inhibit or reduce TLR4 signalling also served as a further control alongside anti-TLR4 antibodies confirming that GP only activates TLR4 in stable TMC HEK293T cells, and that the NFκB activity observed was not due to an alternative activation mechanism. The reduction of NFκB activity upon pre-treatment with LPSRS for both LPS and GP suggested a similar MoA as although the two ligands are vastly different in structure, size and composition, both agonists

competed with LPSRS for the same active site. This finding serves as the first indication that a hydrophobic region of GP similar to that of the acyl chains of LPS may be involved in the activation of TLR4 in EVD and steps were taken to identify the interacting region.

6.2.3 Internal fusion loop of GP identified as possible interaction site with TLR4

The sequence analysis of GP regarding hydrophobic regions yielded five substantial hits of which only one, the internal fusion loop of GP, was a realistic target due to the location, accessibility, and conservancy. Expression and purification of GPΔFuLo did not pose any further challenges compared to WT GP using the protocol developed in this thesis and the purified protein could be analysed regarding TLR4 activity. The NFκB signalling assays revealed that the deletion of IFL eliminated GP's ability to activate TLR4. Following this discovery, the direct interaction of GP and TLR4 was next explored, including the prime target GPΔFuLo and all glycan alterations.

6.2.4 Approaches to study the interaction of GP and TLR4

Within the field, previous work has shown an interaction between GP and TLR4, using CoIP from lysed cells. To date this has not been investigated further, a need that this dissertation responded to. Additionally, the accumulation of GP in the cell lysate, may indicate inefficient protein secretion and possibly other limitations in protein processing. Although Co-IP from cell lysates is commonly used to confirm protein-protein interaction this method has one distinct flaw: as the standard readout of an IP is western blotting, only the presence of known target proteins may be reported with specific antibodies. The technique does not account for potential protein complexes that the target proteins may be a part of and therefore in close proximity to each other, however with no actual direct interaction between the two. A possible reason for using this preliminary technique in the past may be the lack of techniques and methods to effectively express and purify stable, and natively glycosylated GP. This challenge was addressed in this dissertation and allowed sophisticated analysis of the interaction between TLR4 and GP.

Instead of using crude cell lysates, purified protein samples of GP were used to study the interaction with TLR4 and for first time, also the interaction of GP and MD2. This also allowed studying the effect of enzymatic deglycosylation treatment of GP on the interaction and activity of TLR4, conditions which could not be natively created in the expression culture. This setup made it possible to discover that GP interacts with both TLR4 and MD2 seperately. To interact with TLR4 and MD2, it is likely that GP

binds the hydrophobic pocket of the TLR4/MD2 complex, similarly to LPS, an unexpected and novel finding.

6.2.5 Altering glycosylation does not impact the interaction of GP and TLR4

One interpretation of the finding that glycan removal from GP (via treatment with O-glycosidase and PNGase F) correlates with a reduction with NFκB signalling is that interaction with TLR4 is reduced with these glycan-depleted counterparts. However, all glycan depleted GP species showed interaction with TLR4 and MD2. Therefore, it can be concluded that the glycans of GP do not play an active role in the activation of TLR4 or general function of the protein but rather passively by shielding the vulnerable epitopes, avoiding recognition by enzymes and receptors alike. Furthermore, fully deglycosylated GP would not occur natively and using deglycosylating enzymes in EVD treatment is not advisable due to severe off target effects. Although studying the glycosylation of GP did not lead to further understanding of the TLR4 activation by GP, it did broaden the knowledge about GP and challenges some earlier theories in the field that suggested glycans were a key interacting partner with TLR4.

6.2.6 The IFL of GP is required for the interaction with TLR4 and MD2

The GPΔFuLo species, on the other hand, did not interact with TLR4 and MD2 in the same Co-IP experiment. For the first time, another candidate for TLR4 receptor interaction with EBOV GP was identified, and the results obtained throughout the PhD suggest that the IFL of GP is essential for the activation of TLR4. Although only hypothetical at this point, it may the case that the IFL mimics the acyl chains of LPS and interact with the pocket formed by MD2 and TLR4. Perhaps these findings may contribute to a further understanding of the activation of TLR4 by other molecules such as HMGB1 associated with a broad range of human diseases, whose MoA remains a mystery. Additionally, the protocols developed in this thesis allow to easily produce, purify and analyse cytotoxic and highly glycosylated GP, can also be applied to studying other potential ligands of TLR4, which pose similar obstacles to express native protein for biochemical studies.

6.2.7 Therapeutic potential of findings

In addition to further elucidating the biochemical mechanism by which GP activates TLR4, i.e. competitive binding with LPSRS, the effects of GP glycosylation and the IFL of GP on TLR4 activation, each of these also present an area of therapeutic potential. An additional investigation into the

therapeutic potential of shed GP was also performed. For this, the metalloprotease TACE, the primary facilitator of GP shedding and associated disease complications such as the system-wide activation of TLR4, was investigated to show how a potential inhibition of it may lead to a reduction in TLR4 signalling. The reasons for exploring shed GP as a therapeutic potential are outlined below.

During an infection by EBOV, the levels of GP may reach levels higher than 34 µg/ml in the blood, as estimated from animal studies with infected guinea pigs[166]. During infection the full-length form of GP1,2 including the transmembrane domain is expressed, incorporated into newly formed virions, and presented at the cell membrane of infected cells. The high levels of GP in the blood lead to activation of targets and secretion of cytokines remote of the site of infection are only reached due to the cleavage of GP1,2 into shed GP by the host endogenous TACE/ADAM17. Although inhibition of TACE should lead to the reduction of shed GP, keeping GP immobilized at the membrane, this has never been explored scientifically. In this dissertation the inhibition of TACE/ADAM-17 was identified as a novel target to treat EBOV infection. Upon inhibition of TACE, the amount of shed GP is significantly reduced and the activation of TLR4 is almost eliminated unless the virion or infected cell were to come in direct contact with the receptor. Although all clinical trials studying the use of TACE inhibitors have failed due to side effects and efficacy over existing treatments to date, the identification of new targets in cancer and inflammation are advancing continuously including the discoveries made during this research project[167].

In addition to the newly identified role in TLR4 signalling, the IFL of GP is essential for endosomal escape of the virus. Targeting that region of GP may additionally result in reduced the rate of cell infections and viral propagation as well as reducing the TLR4 associated symptoms associated with EVD, thus providing a multi-faceted treatment. Further, as identified in this thesis LPSRS may be utilised to further reduce TLR4 activation in EVD and the associated cytokine storm, supporting treatment and lessening the burden on the patient.

6.3 Future objectives

The expression and purification of active EBOV GP from HEKL293T cells was successful yielding up to 10 mg/l of pure protein, compared to the previously attempted expression protocols using modified GP which was estimated to be 1 µg/ml prior to purification[20]. Certainly, the methodology to express and purify GP could still be further improved to increase the yield and enable the use of techniques which demand high amounts of protein, such as electron-microscopy.

Subject to available facilities the yield of GP could be improved by using suspension cultures over adherent cultures to increase the medium to cell ratio, thus increasing the available volume to collect expressed GP. Alternatively, to the use of transient transfection, a cell line stably expressing GP with an inducible promotor could be established, streamlining the expression process and eliminating further variables, however the disadvantages would be the time-consuming process and additional safety precautions that come with viral integration of a cytotoxic protein in the mammalian genome.

The purification process could also be improved to be compatible with larger volumes of GP containing culture media while eliminating the need for dialysis. A serum-free media would circumvent the need for dialysis which would avoid further loss of material. It is desirable that the interaction between GP and the TLR4/MD2 signalling complex is further investigated with quantifiable techniques such as SPR to confirm the interaction and deepen the understanding with regards to the kinetics, dynamics and binding mechanisms. Finally, a structure of the possible TLR4/MD2-GP complex would result in fundamental insight of the interaction and deepen the understanding of TLR4 activation in general with potential therapeutic use.

As the effect of glycosylation on the interaction of GP with TLR4 has been shown to be miniscule, it may be possible to limit the glycosylation of GP without affecting the interaction with TLR4/MD2 for structural analysis. As crystal structures of GP lacking the MLD and glycans exist, it may be possible to collaborate with those research groups to obtain a structure of the TLR4/MD2 – GP interaction.

6.4 Concluding remarks

Initially the research focus of the PhD was to study the effect of GP glycosylation on the activation of and potential interaction with the TLR4 signalling complex, based on previous studies suggesting that the removal of glycans or deletion of the mucin-like domain results in reduction of TLR4 activation[20]. During this investigation it was revealed that the mutation of individual glycosylation sites does not affect the activation of TLR4 by GP. Similarly, only the complete removal of O- or N-glycosylation on GP lead to elimination of TLR4 activity, not truncation by enzymes or restricted glycan complexion during expression. However, it was also shown that the removal of glycans results in significant instability of GP and vulnerability to degradation, suggesting that the glycosylation does not only shield epitopes from antibody recognition but also from proteases. Therefore, the removal of glycosylation of GP does lead to reduction or elimination of TLR4 activation, however, not due to previously presumed direct dependence but rather due to protein degradation and instability caused by the lack of glycan shielding.

Two main achievements were accomplished during this project:

Firstly, this dissertation presents the first methodology to express and purify stable and natively glycosylated GP in an accessible manner while maintaining high yields. Lessons may be learned from the established strategies and applied to other heavily glycosylated proteins of cytotoxic or non-cytotoxic nature, requiring particular care when expressing and purifying to maintain activity.

Secondly, three novel vulnerabilities have been identified which may offer targets for future therapeutic used in EVD. Firstly, the internal fusion loop of GP to target the protein directly, leading to loss of TLR4 activation and EBOV ability to infect. Secondly, the use of LPSRS to specifically inhibit TLR4 activation by GP and lastly the use of TACE inhibitors to significantly reduce the amount of shed GP present in the extracellular space. To date the use of LPSRS to inhibit specific TLR4 signalling as well as the use of TACE inhibitors to eliminate GP migration through the patient's body and system-wide sepsis like activation of TLR4 have not been established in the treatment of EVD. Although the vulnerability presented by the IFL requires further confirmation and deeper investigation before presenting a therapeutically translation opportunity, the other two targets are more amendable to being established and trialled in a clinical context, in case of a future EBOV epidemic.

7 Material and Methods

7.1 Genetic materials

7.1.1 Synthetic gene design and acquisition

The amino acid sequence for ZEBOV GP Makona was obtained from GenBank (GenBank: AAB81004.1) and a synthetic gene was ordered from Invitrogen: ID 15ACZN7P; GPEBO in pMA-RQ (ampR); Cloning sites Sfi1; optimised for expression in Homo sapiens cells; referred to as GPori in this report (Figure 7.1).

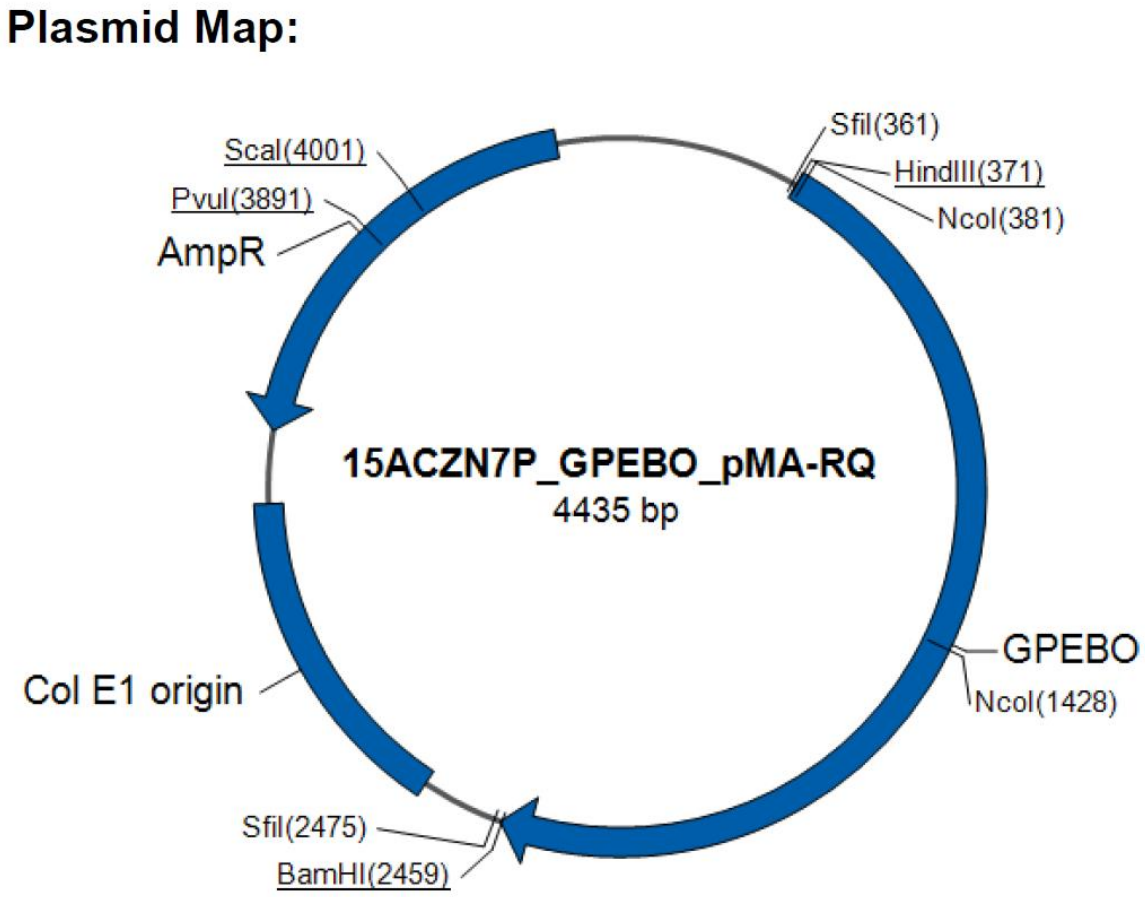


Figure 7.1: The original GP construct obtained from Invitrogen

Following the expression optimisation, a new GP construct was designed and ordered through Synbio Technologies with the endogenous signalling peptide, an N-terminus Strep-II tag, TEV cleavage site and poly-His tag. Further, the gene was optimised to increase yield by avoiding splice cites, RNA secondary structures, optimised GC content and base repeats.

Item	1. Syno® 2.0 Gene Synthesis
Project ID	S011884-01
Gene Name	K-MiniP-KZ-SPSTHFLGP
Gene Length	2199 bp
Vector	Vector name: pcDNA3.1+
	Cloning site: NheI GCTAGC and NotI GCGGCCGC
	Resistance: Ampicillin
	Ship the vector to Synbio Technologies: No
Delivery Form	Standard delivery: 2-5 μg lyophilized plasmid
DNA Sequence	

GCTAGCTagagggtatataatggaagctcgacttccagcttggcaatccggtactgttggtaaagccaccatg
ggtgttacaggaatattgcagttacctcgtgatcgattcaagaggacatcattctttctttgggtaattatcc
ttttccaaagaacattttccatgtcgtactactggagccacccgcagttcgagaaagagaacctgtacttcca
atcccatcaccatcaccatcacggcggcggcggcagcatcccgcttggagttatccacaatagtacattacag
gttagtgatgtcgacaaactagtttgtcgtgacaaactgtcatccacaaatcaattgagatcagttggactga
atctcgaggggaatggagtggcaactgacgtgccatctgcgactaaaagatggggcttcaggtccggtgtccc
accaaaggtggtcaattatgaagctggtgaatgggctgaaaactgctacaatcttgaaatcaaaaaacctgac
gggagtgagtgtctaccagcagcgccagacgggattcggggcttcccccggtgccggtatgtgcacaaagtat
caggaacgggaccatgtgccggagactttgccttccacaaagagggtgctttcttcctgtatgatcgacttgc
ttccacagttatctaccgaggaacgactttcgctgaaggtgtcgttgcatttctgatactgccccaagctaag
aaggacttcttcagctcacacccctgagagagccggtcaatgcaacggaggacccgtcgagtggctattatt
ctaccacaattagatatcaggctaccggttttggaactaatgagacagagtacttgttcgaggttgacaattt
gacctacgtccaacttgaatcaagattcacaccacagtttctgctccagctgaatgagacaatatatgcaagt
gggaagaggagcaacaccacgggaaaactaatttggaaggtcaaccccgaaattgatacaacaatcggggagt
gggccttctgggaaactaaaaaaaacctcactagaaaaattcgcagtgaagagttgtctttcacagctgtatc
aaacggacccaaaaacatcagtggtcagagtccggcgcgaacttcttccgacccagagaccaacacaacaaat
gaagaccacaaaatcatggcttcagaaaattcctctgcaatggttcaagtgcacagtcaaggaaggaaagctg
cagtgtcgcatctgacaacccttgccacaatctccacgagtcctcaacctcccacaaccaaaacaggtccgga
caacagcacccataatacacccgtgtataaacttgacatctctgaggcaactcaagttggacaacatcaccgt
agagcagacaacgacagcacagcctccgacactccccccgccacgaccgcagccggacccttaaaagcagaga
acaccaacacgagtaagagcgctgactccctggacctcgccaccacgacaagcccccaaaactacagcgagac
tgctggcaacaacaacactcatcaccaagataccggagaagagagtgccagcagcgggaagctaggcttaatt
accaatactattgctggagtagcaggactgatcacaggcgggagaaggactcgaagagaagtaattgtcaatg
ctcaacccaaatgcaaccccaatttacattactggactactcaggatgaaggtgctgcaatcggattggcctg
gataccatatttcgggccagcagccgaaggaatttacacagaggggctaatgcacaaccaagatggtttaatc
tgtgggttgaggcagctggccaacgaaacgactcaagctctccaactgttcctgagagccacaactgagctgc
gaacctttttcaatcctcaaccgtaaggcaattgacttcctgctgcagcgatggggtggcacatgccacatttt
gggaccggactgctgtatcgaaccacatgattggaccaagaacataacagacaaaattgatcagattattcat
gattttgttgataaaaccttccggaccaggggacaatgacaattggtggacaggatggagacaatggatac
cggcaggtattggagttacaggtgttataattgcagttatcgctttattctgtatatgcaaatttgtctttTA
GGCGGCCGC

Figure 7.2: Genetic code of the synthetic gene of GP after sequence optimisation for mammalian expression

7.1.2 Plasmids

While the constructs and plasmids containing TLR4, MD2 and CD14 were obtained from the Gay research group stock, the following plasmids and backbones used for cloning and protein overexpression purposes were gifted or obtained from outside of the research group (Table 7.1).

Table 7.1. List of Plasmids

Plasmid name	Producer	Location
pMA-RQ GPEBO	Invitrogen, Thermo Scientific	Merelbeke, Belgium
pcDNA 3.1 + GP	Synbio Technologies	New Jersey, USA
pHLsec from pLexM	Manuela Urbischek	Department of Biochemistry
pDisplay-AP-CFP-TM	Dr Neil Rzechorzek	Department of Biochemistry
pRK5F-TACE	Addgene (31713)	Teddington, UK
Furin-bio-His	Addgene (51755)	Teddington, UK
pBIIX-NF-κB-Luc	Dr Lee Hopkins	Department of Vet. Medicine
hRG-TK (Renilla luciferase)	Dr Lee Hopkins	Department of Vet. Medicine

7.2 Cloning

7.2.1 Primers

The primers for pcDNA 3.1, (Ori)were designed to integrate a N-terminal HindIII and a C-terminal KpnI restriction enzyme cutting site flanking the gene of interest.

Table 7.2 List of primers

Construct name	Forward primer	Reverse primer
GP in pDisplay	ATATATAGGGCCCATCCCGCTTGGAGTTATC	ATTCTGTATATGCAAATTTGTCTTTTAGGGCGCGCCATATATAT
GP in pcDNA	GAATTCCACCACACTGGACTAGTGG	TCACTATAGGGAGACCCAAGCTGG
GP Synbio	GCCGCCAAGCTTGCGATCCCGCTTGGAGTTAT	GGCGGCAAGCTTCGCCTAAAAGACAAATTTGCATATA
GP GP1	TAGGGCGCGCCGATGGTG	TCTTCGAGTCCTTCTCCCGCC
GP GP2	GAAGTAATTGTCAATGCTCAACCCAAATGCAA	GCTGCCGCCGCCGCCGTG

	CCCCAATTTACATTACTGGACTACTCAG	
GPN40D	AGTTATCCACGATAGTACATTACAGGTTAGTGATGTC	CCAAGCGGGATGCTGCCG
GPT42V	CCACAATAGTGTATTACAGGTTAGTGATG	ATAACTCCAAGCGGGATG
GPN204D	AGAGCCGGTCGATGCAACGGAGG	CTCAAGGGGTGTGAGCTGAAG
GPT206V	GGTCAATGCAGTGGAGGACCCGTC	GGCTCTCTCAAGGGGTGT
GPN228D	TTTTGGAACTGATGAGACAGAGTAC	CCGGTAGCCTGATATCTAATTG
GPT230V	AACTAATGAGGTAGAGTACTTGTTCGAGGTTG	CCAAAACCGGTAGCCTGA
GPN238D	CGAGGTTGACGATTTGACCTACG	AACAAGTACTCTGTCTCATTAG
GPT240V	TGACAATTTGGTCTACGTCCAACTTG	ACCTCGAACAAGTACTCTG
GPN257D	GCTCCAGCTGGATGAGACAATAT	AGAAACTGTGGTGTGAATC
GPT259V	GCTGAATGAGGTAATATATGCAAGTGGG	TGGAGCAGAAACTGTGGT
GPN268D	GAAGAGGAGCGACACCACGGGAA	CCACTTGCATATATTGTCTCATTCAGC
GPT270V	GAGCAACACCGTGGGAAAACTAATTTG	CTCTTCCCACTTGCATATATTG
GPN296D	AACTAAAAAAGACCTCACTAGAAAAATTCGCAGTG	TCCCAGAAGGCCCACTCC

GPT298V	AAAAAACCTCGTTAGAAAAATTCGCAGTGAAG	TTAGTTTCCCAGAAGGCC
GPN563D	GCAGCTGGCCGACGAAACGACTC	CTCAACCCACAGATTAAACCATCTTGG
GPT565V	GGCCAACGAAGTGACTCAAGCTCTCCAAC	AGCTGCCTCAACCCACAG
GPN618D	TTGGACCAAGGACATAACAGACA	TCATGTGGTTCGATACAG
GPT620V	CAAGAACATAGTAGACAAAATTGATCAGATTATTC	GTCCAATCATGTGGTTCG
GPΔMLD	ACTCATCACCAAGATACC	TGATACAGCTGTGAAAGAC
GPΔTM	TGGTGGACAGGATGG	CACACTGGACTAGTGCCATCCTGTCCACCA
GP sequencing	ACCATCAGATACCAGGCCAC	TCCTGGTGGTGGGTATTGTT
GP colony PCR	CGCATAAGCTTGCCACCAT	AGGGATCCTCAGAACACGAA
GP_TEV1	CGTAGTCGGGGTAGCCCTGGAAGTACAGATTGTCC	TCCACGACTTCGTGGAGAATCTGTACTTCGAC
GP_TEV2	GGACAATCTGTACTTCCAGGGCTACCCCGACTACG	CGTCGAAGTACAGATTCTCCACGAAGTCGTGGA
GP-ori	GTTTTCCCAGTCACGACGTTGT	CTATGACCATGTTAATGCAGCTGG
GP_MLD	ATGATGACCGGTGAACTGAGCTTCACCGT	CATCATGGTACCTGTGTTGGTGATCAGTCC
GP_485	ATGATGACCGGTATCCCCCTGGGCGTGAT	CATCATGGTACCTGTGTTGGTGATCAGTCCCAGC

GP_MLD_501	ATGATGACCGGTGAACTGAGCTTCACCGTGGTGT	CATCATGGTACCGCGTCTGGTTCTTCTGCCG
GP_ECD	CACGACTTCGTGGACTGAGGATCCCTGGG	GCCCAGGGATCCTCAGTCCACGAAGTCGT

All protein constructs were cloned into pDisplay using the restriction enzymes ApaI and HindIII, incorporated by the primers. Unless stated otherwise, the primers were ordered from Sigma-Aldrich at 100 µM concentration in water and stored at -20 °C.

7.2.2 Polymerase chain reaction (PCR)

7.2.2.1 PCR for restriction enzyme assisted cloning

To create amplicons containing the gene of interest, PCR was carried out using a 3Prime (TECHNE) PCR machine. A PCR reaction mix was prepared, which contained 2x Phusion® High-Fidelity PCR Master Mix (NEB), 2.5 µM of forward and reverse primer, <100 ng template DNA and nuclease free water in a final volume of 50 µl. PCR was performed using the New England Biolabs recommended cycling procedure (Table 3):

Table 7.3: PCR protocol for restriction enzyme cloning

	Temperature in °C	**Duration**	**Number of cycles**
Initial denaturation	98	30 s	1
Denaturation	98	10 s	35
Annealing	(primer dependent)	30 s	
Extension	72	30 s per kb	
Final extension	72	10 min	1
Hold	4	∞	

The annealing temperature varied between 55-72 °C depending on the appropriate primers for different constructs as indicated in Table 7.2. Successful amplification of the gene of interest was assessed using gel electrophoresis and the band of interest extracted before restriction enzyme digestion.

7.2.3 Restriction enzyme cloning

7.2.3.1 Enzymatic digest of vector and insert

Amplicon PCR products and the expression plasmid of choice were digested using restriction enzymes for subsequent ligation. Double digestion with two different enzymes using 5 µg of DNA was performed. The DNA samples were added to 1x CutSmart R Buffer, 10 U of each enzyme with nuclease free water to a final volume of 50 µl. The complete mix was briefly centrifuged using a microfuge and incubated for 1 hours at 37°C or overnight if previous experiments yielded incomplete digestion.

7.2.3.1 Plasmid dephosphorylation

Dephosphorylation of the plasmid was performed prior to the ligation reaction. For the dephosphorylation reaction mixture 10x CutSmart R Buffer was mixed with 1 U Calf Intestinal Alkaline Phosphatase (CIP) and plasmid in a final volume of 50 µl. The solution was incubated for 1 hour at 37 °C, followed by 5 min enzyme heat inactivation at 70 °C. The solution was left to cool down to room temperature (RT) until further processing for ligation.

7.2.4 Plasmid and insert ligation

The digested and dephosphorylated expression plasmid of choice and the digested insert, containing the gene of interest were ligated using T4 DNA ligase(NEB). For optimised ligation the amount of DNA needed for a molar ratio of 1:3 plasmid to insert was calculated using the NEBio Calculator online tool. A 20 µl ligation reaction mix was created, containing 1x T4 DNA ligase buffer, 1 U T4 DNA ligase,100 ng plasmid and the appropriate amount of insert, and Nuclease-free water to a final volume of 20 µl. The mixture was spun down briefly and incubated for 1 hour at RT or overnight at 16 °C. Following incubation, the reaction mixture was heat inactivated at 50 °C for 15 minutes and chilled on ice. 5 µl of the ligation mix was added to 50 µl of chemically competent *E.coli* DH5α cells and bacterial transformation performed as outlines in section 7.2.6. Positive controls containing the uncut plasmid, negative controls containing digested plasmid, and non-dephosphorylated ligations were run alongside the ligation for qualitative comparison.

7.2.5 Agarose gel electrophoresis

DNA fragments from PCR were separated using agarose gel electrophoresis. A 1% agarose gel was prepared with agarose purchased (Merck) and 1xTAE buffer. Electrophoresis was performed at 100V for 1 hour and the DNA samples were subsequently visualised using UV light and SYBRR Safe DNA gel

stain (Invitrogen). Fragment length estimation was achieved by comparing the sample migration with the 2-log DNA standard ladder (NEB).

7.2.5.1 Gel extraction

The DNA fragments with the correct molecular size were extracted and purified using the QIAquick R Gel Extraction Kit after gel electrophoresis (Qiagen). The identified DNA bands were cut from the gel and transferred into a sterile 2ml Eppendorf tube. The manufacturers' instructions were followed according to the protocol except twice the volume of Qiagen QG buffer and PE buffer was used. Elution of the fragments from the column was performed using 50 µl of the QiagenEB buffer.

7.2.5.2 DNA quantification

The purified DNA samples were analysed using NanoDrop Spectrophotometer2000c (Thermo Scientific). A representative amount of 1.5 µl sample was used and the concentration in ng/µl and quality ratio (absorption at 260/280 nm and 230/280 nm) were recorded.

7.2.6 Transformation of bacterial cells

Chemically competent *E.coli* DH5α cells were used in a sterile environment for plasmid quantification and isolation (NEB). After thawing an aliquoted sample on ice, 1 µl of 1-10 ng of plasmid DNA was added and the transformation mix was kept on ice for 25 minutes. Following this step, the mix was heat-shocked in a 42 °C water bath for 45 seconds. The samples were immediately transferred onto ice and incubated for 2 minutes. 350 µl of SOC medium (Invitrogen) was added to the mix and the cells were incubated at 37 °C for 1 hour in a shaking incubator. 100 µl of the incubated transformation solution were plated onto a 10 cm agar plate containing the appropriate antibiotic (37 °C, overnight).

7.2.6.1 Preparation of chemically competent bacterial cells

The chemically competent cells were created using the following protocol. All procedures were performed under sterile conditions or with sterilised equipment. An aliquot of the strain of interest (DH5a) was used to inoculate a 3 ml LB pre-culture which was grown overnight at 37 °C. The following day the pre-culture was added to a 2l shaking flask containing 500 ml SOB^{++} medium (2% Bacto-tryptone, 0.5% yeast extract, 8.5 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$) prepared beforehand and incubated at 16 °C overnight up until an OD_{600} of no more than 0.2. The culture was cooled on ice for at least 10 minutes before centrifugation (15 min, 4000g, 4 °C) to pellet the cells. The supernatant was removed, and the bacterial cell pellet was gently resuspended in 100 ml ice-cold TB buffer (10 mM HEPES, 15 mM $CaCl_2$, 250 mM KCl, 55 mM $MnCl_2$). The cell suspension was incubated for 10 min on ice and centrifuged for (15 min, 4000g, 4°C). After supernatant removal, the pellet was resuspended in 18.6 ml ice-cold TB buffer (10 mM HEPES, 15 mM CaCl2, 250 mM KCl, 55 mM MnCl2)

and 1.4 ml DMSO as a freezing protectant. This final solution was incubated for at least 10 min on ice to ensure cooling during the aliquoting process. Aliquots were snap frozen in liquid nitrogen. The frozen competent bacteria aliquots were stored at -80 °C and kept their high competence for at least 6 months.

7.2.7 Single colony screen

To confirm successful ligation and uptake of the plasmid containing the new insert, the bacterial colonies were subjected to a colony screen. Each selected colony was isolated using an individual sterile loop and transferred to a chessboard-gird-like agar plate for referencing. A colony PCR was performed using a sample of each colony, 2x DreamTag Green PCR Master Mix (Thermo Scientific) with 1.25 units of DreamTag DNA Polymerase and 1 µM of the appropriate forward and reverse primer compatible to the insert/plasmid overlap in a final volume of 25 µl. The PCR was performed according to the protocol in Table 7.4.

Table 7.4: PCR protocol for colony PCR

	Temperature in °C	**Duration**	**Number of cycles**
Initial denaturation	95	2 min	1
Denaturation	95	30 s	30
Annealing	Tm-5	30 s	
Extension	72	2 min	
Final extension	72	10 min	1
Hold	4	∞	

7.2.7.1 Plasmid amplification

Single colonies positive for the plasmid of interest were selected from the E.coli DH5α transformation plates grown overnight and inoculated in 10 ml of Luria broth (LB) medium with the appropriate antibiotic (37 °C, overnight). The cultures were pelleted (10 min, 5000 RCF) and DNA isolation was carried out using the QIAprepR Spin Miniprep Kit according to the manufacturer's protocol. Isolated plasmid DNA used for expression in mammalian cells was sterilised using 0.22 µm Millex GV syringe filters (Merck Millipore) to minimise endotoxin contamination due to DH5α plasmid quantification.

7.2.8 Site-directed mutagenesis

To generate GP constructs with single-amino acid changes to alter glycosylation patters, site directed mutagenesis was used. The N-glycosylation sequons of Asn-Xaa-Ser/Thr were mutated to Glu-Xaa-Ser/Thr or Asn-Xaa-Val to disrupt the glycosylation.

The assure high coverage the Q5® High-Fidelity DNA Polymerase (NEB) was used in site-directed mutagenesis. For this Q5® High-Fidelity DNA Polymerase 2x Master Mix, 1 µM of forward and reverse primer, <10 ng template DNA and nuclease free water were mixed in a final volume of 25 µl. The PCR was performed as summarised in Table 7.5.

Table 7.5: PCR protocol for site directed mutagenesis

	Temperature in °C	**Duration**	**Number of cycles**
Initial denaturation	98	30 s	1
Denaturation	98	10 s	25
Annealing	<72 (primer dependent)	10 s	
Extension	72	20 s per kb	
Final extension	72	2 min	1
Hold	4	∞	

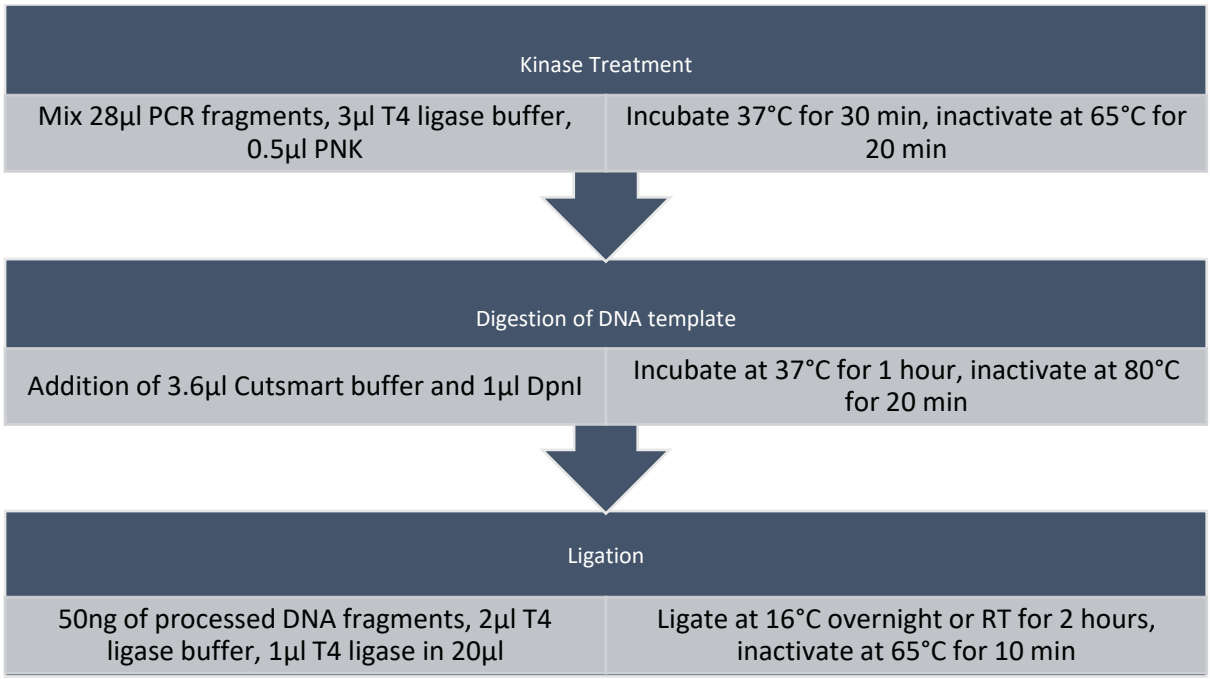


Figure 7.3: Protocol for optimised blunt end ligation following SDM

Following the high-fidelity PCR, the fragments were gel purified, and the concentration was measured. Blunt end ligation of the plasmid containing the gene with the single point mutation was performed according to the protocol summarised in Figure 7.3 and subsequently transformed into chemically competent cells to isolate the DNA as summarised in section 7.2.7.1.

7.2.8.1 DNA sequencing

The constructs obtained from PCR or plasmid amplification were sequenced using the in-house Sanger sequencing facilities to confirm the quality of the constructs. The primers necessary for the sequencing were either present in the facility storage or provided in a concentration of 10 ng/µl. The DNA samples to be sequenced were diluted to a concentration of 100 ng/µl in a total volume of 20 µl using distilled water.

7.3 Cell Biology

7.3.1 Cell handling

For protein expression in mammalian systems HEK293T cells were cultured and maintained per the protocol of the manufacturer (Invitrogen, Thermo Scientific) in a standing incubator at 37 °C and 5% CO2. The cells were maintained in T75 asks (Thermo Scientific) containing DMEM Media GlutaMAX with 10% heat inactivated Fetal Bovine Serum (Gibco, Life Technologies) and 100 U/ml of Penicillin/Streptomycin Solution (Gibco). Cell viability and density was visually determined using a EVOS M5000 microscope (Invitrogen) and quantified using the cell counter Cellometer Mini (Nexcelom) with trypan blue dye (Sigma Aldrich). Cells were passaged every 2-3 days and maintained at a density of 0.5-2E6 cells/ml for a maximin of 20 passages.

If the HEK293 cell line stably expressing hTLR4-MD2-CD14 was to be cultured, the cell handling procedure was adapted. For the first week post-thawing the stable HEK293 cells were handled according to the same protocol as HEK293T cells, to increase cell viability before the selection process. To ensure maintenance of the plasmids coding for the proteins the growth medium was supplemented with Normocin (100 µg/ml), and the selective antibiotics Blasticidin (10 µg/ml) for hTLR4 and Hygromycin B Gold (50 µg/ml) for MD2 and CD14 (InvivoGen). Expi293 GnTI- cells were cultured and maintained according to the same protocol as HEK293T cells with the exception that the medium was DMEM: F-12 Medium (Gibco) and supplemented with 10% PLURONIC F-68.

7.3.2 Thawing cells

The frozen HEK 293T cells were removed from the liquid nitrogen storage and rapidly thawed in a 37 °C water bath for 2 minutes. The cryovials containing 10E6 cells in 1ml of freezing medium were de-contaminated with 70% EtOH and transferred to a 15 ml Flacon tube containing 10 ml pre-warmed DMEM Media GlutaMAX and centrifuged at 1000 RCF for 2 minutes. The supernatant was aspirated, and the cell pellets gently resuspended in DMEM Media GlutaMAX followed by transfer to a T75 ask and incubation overnight. The medium was replaced the following day and the cells passaged 24 hours later. Stable HEK293 expressing TLR4/MD2/CD14 were handled in the same manner.

7.3.3 Freezing cells

HEK 293T or stable HEK293 cells of low passage were collected and frozen for future uses when necessary. For this purpose, the cells were grown to 90% confluency in T75 flasks, washed with pre-warmed 37 °C sterile PBS, followed by gentle dissociation with 0.5 ml/10 cm^2 37 °C Trypsin dissociation agent solution (0.05% trypsin, 0.53 mM EDTA, Merck) for 2 minutes. To inactivate trypsin, the cells detached cells were mixed with 2ml growth medium and viability and concentration was assessed via the cell counter. The cell suspension was transferred to a tube and centrifuged at 1000 RCF/g for 5 minutes. Following the supernatant removal, the cells were resuspended in freezing medium (90% FBS and 10% DMSO) and aliquoted at 10E6 cells per ml in 1.5 ml cryovials. The vials were cooled over 48 hours at -80 °C in a Mr Frosty (Thermo Scientific) cell freezing container at a rate of 1 °C per minute before being transferred to liquid nitrogen storage.

7.3.4 PEI preparation

The transfection reagent polyethylenimine (PEI) was prepared by dissolving linear PEI (Merck) indistilled water while heating the mixture it up to 80 °C. Following its dissolution, the PEI mixture was cooled to RT and pH adjusted to pH 7.4 through 25 mM HEPES. The PEI solution was diluted to 1 mg/ml, filtered through 0.22 µm syringe filters, aliquoting and stored at -20°C. Commercial jetPEI (Polyplus transfection R) was used for initial transfection screens, however discarded due to the same efficiency of linear PEI and the economic advantage of lower cost.

7.3.5 Transfection

Transfections with HEK293T cells were performed after passaging them at least four times. If the recombinant protein of interest was intracellular or membrane associated the following protocol was utilised for transfection.

HEK293T cells were seeded at 5E5 cells/ml 24 h prior to transfection and the cells were checked for a confluency of 70% the day of transfection. A mixture of the plasmid containing the gene of interest and PEI was prepared at 1:2 ratio in 150 mM NaCl of a volume of 1/10 of the final transfection medium mix and incubated for 30 min at RT. The incubated transfection mix was mixed with the remaining 9/10 of culture medium for the respective flask or dish size and added to the cells. About 6 h post-transfection the transfection medium mix was replaced with fresh culture medium and the cells were left to express for 48 h before harvesting.

For secreted recombinant proteins the transfection protocol was altered and optimised accordingly. The cells were seeded at reduced concentration of 2E5/ml the day before transfection onto 15 cm dishes (Thermo Scientific Nunclon Delta Surface). The ratio of DNA to PEI was increased to 1:5 and the transfection culture incubation was also increased from 6 h to 16 h. Following the transfection incubation, the culture medium was replaced with serum-free DMEM, and the cells were left to express and secrete protein for 4 days, followed by culture medium harvesting.

7.3.6 Harvesting of expressed protein

HEK293T cells were then lysed with HEPES lysis buffer (50 mM HEPES, 150 mM NaCl, 2 mM EDTA, 10% glycerol, 0.5% NP-40 pH 7.5) of 1/10th of the recommended culture media volume. The cell solution was transferred to sterile centrifugation tubes and incubated at 4 °C for 30 minutes followed by centrifugation for 10 min (16000 g, 4 °C). The cleared lysate supernatant was collected, filtered through 0.22 µm Millex GV syringe filters (Merck Millipore) and supplemented with 100X protease inhibitor cocktail (Calbiochem) before expression analysis or purification. If the samples were not used immediately then glycerol at 20% final volume was added to the solution and it was stored at -20°C.

For harvesting of secreted proteins such as GP, the cell culture medium was collected 4 days post transfection. The medium was filtered using single use SteriCup vacuum filters to remove cells and other debris and the supernatant was collected in sterile Duran bottles containing a protease inhibitor cocktail (Calbiochem) and stored at 4 °C until protein purification or media concentration.

7.4 Protein analysis methods

7.4.1 SDS polyacrylamide gel electrophoresis (SDS PAGE)

7.4.1.1 *Gel preparation*

SDS Polyacrylamide gels were made using 1.5 mm gel cassettes (Novex Life Technologies). The following substances were mixed in 50 ml Falcon tubes to make up 4 gels comprising of 12% running gel and 5% stacking gel. For the 12% running gels, 10 ml of 1.5 M Tris HCl pH 8.8, 16ml of 30%

acrylamide, 13.2 ml of ddH2O, 400 µl of 10% SDS, 400 µl of 10% APS, and 16 µl of TEMED made up the final gel mix. For the 5% stacking gels, 1.25 ml of 1.5 M Tris HCl pH6.8, 1.7ml of 30% acrylamide, 6.8ml of ddH2O, 100 µl of 10% SDS, 100 µl of 10% APS and 10 µl of TEMED were mixed. TEMED and APS were added and mixed into the solution immediately before pouring the solution into the cassettes. The resolving gel was covered with 70% EtOH to create a level starting point for sample separation. Twenty minutes later the ethanol was discarded, and the stacking gel poured into the cassette and sealed with the combs. The polymerised gels were placed into the mini gel tank (Thermo Scientific), with 1x running buffer. The running buffer was obtained from diluted running buffer (10X) containing 30 g Tris base,144 g glycine and 10 g SDS dissolved in 1 l of distilled water of an approximate pH of 8.3.

7.4.1.2 *Gradient SDS-PAGE gel preparation*

Gradient polyacrylamide gels were utilised to enable improved visualisation of samples with multiple proteins of different molecular weight and improved resolution of high molecular weight samples. To create a gradient gel solution for a 5% stacking gel and a 12% running gel were prepared as previously described. To achieve a gradient, each individual gel was cast separately rather than the staggered approach. Here 2 ml of the 5% gel solution were drawn up in a 10 ml stripette, followed by 7ml of 12% gel solution. After the gel solution for one gel was present within the stripette, air bubbles were cautiously introduced by drawn up air. If done correctly, this would ensure gradual mixing of the 12% solution at the bottom of the stripette with the original stacking gel. After brief mixing, the now gradient gel solution was slowly applied to cassettes and sealed with combs. The sample preparation and SDS PAGE procedure remained the same.

7.4.1.3 *Sample preparation*

The protein samples were prepared by mixing the whole cell lysate with 4x non-reducing loading buffer in a 3:1 ratio. The non-reducing loading buffer contained 0.04% bromophenol blue, 8% SDS, 40% glycerol, and 250 mM Tris HCl pH 6.8 topped up with distilled water to a final volume of 10 ml. The complete mix was heated for 7 minutes at 95 °C and centrifuged for 10 minutes at 13,000 RCF at RT to pellet the cell debris from the lysate. After centrifugation 20 µl of the protein supernatant per sample was applied to each well and gel electrophoresis was performed at 200V for 1 h. To visualise the protein samples and compare them with the molecular weight marker, Precision Plus Protein Dual Colour Standards (BioRad), the gel was stained with Coomassie Brilliant Blue by constant shaking at 12 RCF for 1 h and de-stained using distilled water for 1 h. If the protein samples were to be analysed by western blotting, two polyacrylamide gels were run, one for immediate Coomassie Brilliant Blue staining and one for western blotting.

7.4.1.4 *Native-PAGE*

If interactions or protein homogeneity were to be assessed, then a native-PAGE was performed. For Native-PAGE, 6% acrylamide gel solution was prepared using TRIS CAPS running buffer (30 mM TRIS, 10 mM CAPS pH 9.4). Unlike SDS-PAGE the samples were not heat denatured but mixed with 5x Native-PAGE sample buffer (30 mM TRIS, 10 mM CAPS, 50% glycerol, 0.01% bromophenol blue pH 9.4) less than 10 minutes before gel electrophoresis. If the resolution was inadequate or smears appeared on the gel, then native PAGE detergent buffer was added to the sample before gel electrophoresis. The gels were run in 1x TRIS CAPS (2 h, 150V).

7.4.2 Western blotting

7.4.2.1 *Nitrocellulose membrane transfer*

Western blotting was performed using the Bolt Mini blot Module in the associated Mini Gel Tank (Novex Life Technologies). For the transfer sandwich, assembly filter paper and sponge pad were soaked in 1x transfer buffer containing 20 % MeOH, 10 % 10x transfer buffer (25 mM Tris base, 192 mM Glycine), and 70 % distilled water. Compression, additional transfer buff er, and a blot roller were used to ensure the absence of air bubbles at several steps during the assembly. The assembly performed in the following order from cathode to anode: 4 sponge pads, 1 filter paper, SDS PAGE gel, nitrocellulose membrane, 1 filter paper, 2 sponge pads. Finally, the module was placed into the gel tank and the tank was filled to the indicated line with 1X transfer buffer. Protein transfer from the gel to the membrane was conducted (12V, 1 h). Then, the membrane was removed from the module and prepared for immunoblotting by rinsing it with PBST (1X) containing 0.05% Tween 20, 137 mM of NaCl, 2.7 mM of KCl, 10 mM of Na_2HPO_4, and 1.8 mM KH_2PO_4. Additional to the immunoblotting of the nitrocellulose membrane, the polyacrylamide gel was stained with Coomassie Brilliant Blue by constant shaking at 12 RCF/g; 1 h and de-stained with distilled (1 h) in order to check the quality of the transfer.

7.4.2.2 *Immunoblotting*

Following the protein transfer, the nitrocellulose membrane was blocked with a 50 ml solution of 3% BSA in 1X PBST (1 h, RT, rolling). After blocking of the membrane, 10 ml of the blocking solution was used to dilute the primary antibody to the concentration recommended by the manufacturer, 1:2000 in most cases. The primary antibody solution and the membrane were incubated overnight rolling (4 °C). Afterwards the membrane was washed three times with 1x PBST for 10 minutes each. The

secondary antibody conjugated to horseradish peroxidase (HRP) was diluted to 1:10000 in blocking buffer and incubated with the membrane for 1 h at RT on a tube roller followed by three washes with 1x PBST for 10 minutes each. When using HRP-conjugated primary antibodies the membranes were incubated overnight with a dilution of less than 1:2000. The antibodies used in western blotting or biochemical assays are listed in Table 7.6.

For the signal development of HRP, SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) was used, according to the manufacturer's protocol. For signal detection and visualisation, the membrane was placed in a developing cassette and transferred to the dark room. The Super RX-N medical x-ray film (Fuji) was exposed to the chemiluminescent emitting membrane for different periods of time depending on the signal strength followed by chemical photo developing using Velopex Developer/Fixer (Dental Sky) in a Mini-Medical Series Developer (AFP Imaging Corp).

Table 7.6: List of Antibodies

Antibody	Product Number	Vendor
Anti-TLR4	Ab13867	Abcam
Anti-MD2	Ab24182	Abcam
Anti-CD14	C7673-50TST	SIGMA
Anti-EBOV-GP	40442-MM02	Sino Biological Solution
Anti-beta-actin	Ab8226	Abcam
Anti-Myc-tag	3536055	EMD Millipore Corp
Anti-FLAG-tag	SLBQ7119V	SIGMA
Anti-His-tag	51-9000012	BD Pharmagen
Anti-HA-tag	104K4804	SIGMA
Anti-Strep-tag	StrepMAB-Classic HRP	IBA
Anti-rabbit-HRP	069M4825V	SIGMA
Anti-mouse-HRP	SLBN5472V	SIGMA

7.5 Protein purification

7.5.1 Concentration of cell culture media containing secreted protein

To prepare the harvested solution of secreted protein from mammalian cells for protein purification the culture medium was filtered and clarified using sterile single-use SteriCup vacuum filters to

remove cells and other debris, and 100X protease inhibitor cocktail (Calbiochem) was added to the solution. The cleared cell media was concentrated 20-fold by using Vivaspin 20 centrifugal concentrator PES, 100kDa molecular weight cut-off (Sartorius) in presence of 100X protease inhibitor cocktail (Calbiochem) at 4 °C. The flow-through of the concentration was collected and analysed for cut-off efficiency. The concentrated protein media solution was brought to RT prior to buffer exchange dialysis.

7.5.2 Protein dialysis

Buffer exchange of protein solutions was performed via dialysis to remove small molecules in the concentrated culture media and prepare the protein for affinity chromatography. Following the concentration of the media containing EBOV GP, the concentrated solution was transferred to pre-wetted Spectra/Por 3 dialysis membrane standard RC tubing with a molecular weight cut-off of 6 kDa (Spectrumlabs). The membrane tube was sealed and attached to a Styrofoam flotation gadget, then placed in HisTrap calibrating buffer (100 mM NaCl, 20 mM HEPES, 40 mM Imidazole, pH 8) of a volume 25-fold greater than the sample volume and incubated for 2 hours while stirring at RT. After 2 hours the dialysis buffer was replaced, and the solution was dialysed overnight or until almost colourless.

7.5.3 Nickel affinity purification

Following the harvesting, concentration, and dialysis of the culture medium containing overexpressed protein of interest, the protein was purified using Nickel IMAC with a 5 ml HisTrap excel column (Cytivia) at RT with a peristaltic benchtop P1 pump. The column was equilibrated using HisTrap equilibration buffer prior to protein solution loading at a rate of 5 ml/min. For later analysis a sample of dialysed protein and column flowthrough was collected. The column was washed with 5 column volumes of HisTrap washing buffer and the wash fractions were collected. Elution of the immobilised target protein was performed gradient-free while circulating 1 column volume of HisTrap elution buffer for 10 min with a final elution volume of an addition 1 column volume to ensure complete elution of protein. A collection of all buffers used in the purification process of GP is shown in Table 7.7.

7.5.4 Anion exchange chromatography

The IMAC-purified proteins were further purified with anion exchange HiTrap Q column. The Q column was equilibrated with Q equilibration buffer to maximise protein binding. The Q column was washed for 10 column volumes with HiTrap Q column washing buffer. Initially the protein was eluted on ÄKTA Pure FPLC system (Cytivia) using a 10 column volumes linear gradient of HiTrap elution buffer to analyse the elution fraction and the optimal point of protein recovery. From this point onwards all following elutions were conducted gradient-free while circulating 1 column volume of HiTrap Q elution buffer for 10 mins and 5 final column volumes. A collection of all buffers used in the purification process of GP is shown in Table 7.7.

7.5.5 Size exclusion chromatography

To assess sample purity, condition, and homogeneity of the purified protein solution size exclusion chromatography (SEC) was performed. The protein solution from previous purification steps was concentrated using Vivanspin 20 concentrators (Sartorius) with a 100 kDa cut-off to a final volume of less than 3% of the size exclusion column volume. Unless stated otherwise the samples were loaded onto a Superdex 200 GL 300 column (Cytivia) previously equilibrated with SEC buffer and gel filtration was performed using ÄKTA Pure FPLC system (Cytivia) at a rate of 1 ml/min.

Table 7.7: Buffers used in the purification process of EBOV GP

Buffer Name	Purification step	Content
HisTrap equilibration buffer	Nickel IMAC using HisTrap excel	20 mM HEPES, 100 mM NaCl, 40 mM Imidazole, pH 8.0
HisTrap washing buffer		20 mM HEPES, 75 mM NaCl, 40 mM Imidazole, pH 8.0
HisTrap elution buffer		20 mM HEPES, 75 mM NaCl, 400 mM Imidazole, pH 8.0
HiTrap-Q loading buffer	Anion exchange chromatography using Q HT column	20 mM HEPES, 100 mM NaCl, 100 mM Imidazole, pH 8.0
HiTrap Q washing buffer		20 mM HEPES, 75 mM NaCl, pH 8.0
HiTrap Q elution buffer		20 mM HEPES, 1 M NaCl, pH 8.0
SEC buffer	Size exclusion chromatography using S200 GL10/300	20 mM HEPES, 100 mM Tris, 500 mM NaCl, 1 mM EDTA, pH 8.0

The relevant fractions were collected, analysed using SDS-PAGE, pooled, and prepared for subsequent storage at -80 °C or experiments. A collection of all buffers used in the purification process of GP is shown in Table 7.7.

7.5.6 TEV Protease Cleavage

For constructs containing a TACE mimicking TEV cleavage site the expression medium was supplemented with 5 µg of purified TEV. The shed glycoprotein was harvested as previously described.

Following IMAC purification the poly-His tagged proteins were subjected to TEV protease cleavage to remove the tags. The eluted samples were pooled, and the volume was doubled using HiTrap-Q loading buffer to reduce imidazole concentration. The samples were incubated at 4 °C for 3 h with 2 µg of purified TEV protease. After TEV cleavage, the digested solution was cleared using a 0.2 µm syringe filter and the sample was injected onto a 5 ml HisTrap excel pre-equilibrated with HiTrap-Q loading buffer. The sample flowthrough was collected and subjected to anion exchange chromatography.

7.5.7 Co-Immunoprecipitation

To perform co-immunoprecipitation of overexpressed mammalian proteins HEK293T cells were seeded onto 6-well plates at a density of 2E5 24 hours prior to transfection. On the day of transfection 3 µg of plasmid DNA were mixed with the transfection agent linear PEI at a 1:2 ratio and the HEK293T cells were transfected. Six hours post-transfection the medium was replaced. At 48 hours post-transfection the medium was collected, cell-cleared and concentrated using 30 kDa Vivaspin 2 concentrators while the cells were washed with sterile RT 1x PBS. The cells were dissociated from the plate with cell scrapers, 200 µl of HEPES lysis buffer (50 mM HEPES, 150 mM NaCl, 2 mM EDTA, 10% glycerol, 0.5% NP-40 pH 7.5), supplemented 100X protease inhibitor cocktail (Calbiochem), and the cell lysis solution was transferred to a sterile 1.5 ml centrifugation tubes. The lysis mixture was lysed for 1 h 4 °C and subsequent lysate clearance by centrifugation (16000 g, 10 min, 4 °C). All steps were undertaken in presence of 100 X protease inhibitor cocktail (Calbiochem). In case of immunoprecipitation of purified proteins, 2µg of each target protein mixed with HEPES lysis buffer to a final volume of 20 µl. For IP with purified proteins, the GP proteins produced in house and commercially available MD2 (ab238343, Abcam) and TLR4 (ab233665, Abcam) were used.

Prior to immunoprecipitation the affinity beads with equilibrated with HEPES lysis buffer for 30 minutes at 4 °C. If the target proteins were secreted and intracellular the cleared cell lysate and the concentrated medium were mixed and incubated at 4 °C overnight (for at least 12 hours). The

following day the protein mixture was added to 20 µl of EZview Red FLAG M2 beads (Merck), magnetic anti-Myc beads, magnetic anti-HA beads or Protein A/G beads (Abcam) and the final mixture volume was increased to 500 µl with HEPES lysis buffer and incubated for 2 hours at 4 °C. In case of Protein A/G beads, the samples were incubated with 2 µl of tag-specific affinity antibodies for 4 hours at 4 °C followed by the addition of protein A/G beads and incubation for 2 hours at 4 °C. To remove non-specific proteins not interacting with the protein of interest the bead-protein mixture was washed three times. Each washing step comprised of centrifugation of the samples (400 rcf, 5 min, 4 °C) removal of the supernatant and addition of 500 µl HEPES lysis buffer, followed by rotation incubation for 5 minutes at 4 °C. To elute the immobilised proteins from the beads the samples were either boiled in SDS-PAGE loading buffer for 5 min or incubated with 50 µl of 1 M Glycine pH3 to avoid antibody contamination of the sample. In case of acidic elution, the samples were transferred to a centrifugation tube and the solution was neutralised with 50 µl of 1 M NaOH.

7.5.8 Partial proteolysis

Partial proteolysis was performed with 1mg/ml protease stock solutions prepared by dissolving the proteases Trypsin (Merck) and Pepsin (Merck) in 1x PBS and filtered using 0.22 µm Millex GV syringe filters. A dilution series with protease concentrations of 1 µg/µl, 300 ng/µl, 100 ng/µl, 30 ng/µl, 10 ng/µl and 3ng/µl was prepared. To perform limited proteolysis 5 µl of each protease dilution was mixed with 10 µg of the protein of interest in a final volume of 20 µl with 1x PBS. The mixture was incubated for 30 minutes at RT and the reaction was stopped by adding 6.5 µl of 5X SDS STOP solution (400ul contained: non-reducing SDS-PAGE sample buffer, 10.5 µl protease inhibitor cocktail, 4.16ul 500 mM EDTA), followed by heat inactivation at 95 °C for 5 minutes. Samples were immediately then loaded on a gradient 5-12 % SDS-PAGE for degradation analysis.

7.6 Biochemical assays

7.6.1 Luciferase NFκB reporter assay

7.6.1.1 Luciferase preparation

Luciferase assays were used to analyse the activation and signalling capacity of GP through TLR4 and associated interaction partners. To prepare a luciferase assay HEK293T cells were seeded on 96-well at bottom plates at a concentration of 1.5E5 cells/ ml 48 h prior to transfection. The following DNA

mix protocol for transfection was prepared and is representative for 10 wells for easier calculation and upscaling for individual experiments (Table 7.8).

If the hTLR4/MD2/CD14 stably expressing HEK293 cells were cultured and utilised for the NFκB luciferase reporter assay, the transfection mixture composition was altered. Here only the plasmids containing firefly luciferase and renilla luciferase were added to the solution and supplemented with empty plasmids e.g., pcDNA 3.1 +.

Table 7.8: Luciferase assay transfection. Table and collection of the pipetting scheme used for the luciferase assay, including the plasmid, concentration, and volume necessary for HEK293 cells transfection in 96 well plates. The depicted values are representative for 10 wells.

Plasmid	Concentration (ng/ µl)	Volume (µl)
pCDNA3/TLR4	10	1
pEFIRES/MD2	10	1
pcDNA3/CD14	10	1
pBIIX-luc (NF-κB luciferase)	10	10
pTK hRG (Renilla)	10	5
pcDNA3	100	8.2
150 mM NaCl	N/A	21.18
Total	N/A	50

After the DNA mix was prepared, 2 µl of PEI transfection reagent were pipetted in 48 µl 150 mM NaCl, added to the DNA mix and incubated at RT for 30 minutes. After incubation, the 100 µl transfection mix was diluted with 9 times the volume of advanced Gibco DMEM medium containing 10% Fetal Calf serum, L glutamine and the antibiotic mix of 100 U/ml penicillin/streptavidin to give a final volume of 1 ml. 48 h after plating the medium of the plated HEK293 cells was removed and 100 µl of the diluted DNA/jetPEI mix was added to each well. The cells were then incubated at 37 °C and 5% CO_2 for 48 hours for reporter gene expression. For post transfection incubation, the medium was removed and replaced with advanced Gibco serum-free DMEM medium containing the ligand of interest with 100 µl per well and was left to incubate (6 h, 37 °C). Following this incubation, the supernatant was removed, and the cells were carefully washed with 200 µl 37 °C sterile 1x PBS per well, followed by adding 70 µl 1x passive lysis buffer (Promega) to each per well and the plate was stored at 80 °C.

7.6.1.2 Luciferase reporter assay readout

To assess the expression and activity of NFκB, the cell lysates were assayed for luciferase activity using the Dual-Glo luciferase kit (Promega) with a modified protocol. Of each sample 25 µl were incubated with 50 µl of luciferase substrate buffer II and luminescence was detected using a PHERAstar FS (BMG

Labtech) plate reader. Followed by the NFκB luciferase detection, 50 µl of Stop Glo solution were added to quench the luciferase luminescence. Afterwards the luminescence of the constitutively expressed Renilla luciferase was measured in each sample. Following the normalisation of the luminescence levels of firefly luciferase to Renilla luciferase, and baseline establishment with negative controls, the relative NFκB activity could be assessed and quantified.

7.6.2 Assessment of glycosylation

To assess the glycosylation state of the protein a glycosylation estimation carbohydrate kit (Promega) was utilised, and the assay carried out in 96 well plates according to manufacturer's instructions.

7.6.3 LPSRS competition assay

Ultrapure LPS from photosynthetic bacterium Rhodobacter sphaeroides (LPSRS) was obtained from InvivoGen, dissolved in DMSO at a concentration of 1 mg/ml and stored at -20 °C. Stable HEK293 cells expressing TLR4/MD2/CD14, were cultured and transfected for luciferase signalling as previously described in section 7.3. Prior to the induction with TLR4 agonists such as LPS and GP, the cells were subjected to LPSRS with amounts varying from 100ng to 1 µg, and incubated for 1 hour. The induction, culturing, harvesting, and luciferase readout was carried out as previously described in section 7.6.1.

7.6.4 De-glycosylation assays

The glycans of EBOV GP were cleaved and removed through the enzymatic cleavage with O-Glycosidase, PNGase F, Endo H, Galactosidase, Neuraminidase (NEB) according to manufacturer's instructions. The deglycosylated GP samples were used for further analysis. The protocol was changed to accommodate larger protein quantities as the following:

The reaction volume of each de-glycosylation reaction was increased from 20 µl to 500 µl. The buffers and amount of enzyme was upscaled accordingly. To prepare the removal of O-linked glycans using O-glycosidase and removal of N-Glycans using PNGase F, the protein mixture required to be partially denatured with the supplied denaturing buffer and incubated at 95 °C for 10 min. Although the protocol suggested addition of denaturing buffer and 10 min boiling of the protein sample prior to incubation by O-glycosidase and PNGase F, the protocol was altered to keep GP in its native form without possibly compromising activity while deglycosylating. Instead, Urea was added to the deglycosylation mix at a final concentration of 1.5 M and removed through buffer exchange following the glycan cleavage. To guarantee the complete glycan cleavage and homologous GP species after treatment, the recommended incubation period was significantly extended from 1h to 16 h at 37 °C.

An addition of 10 % glycerol ensured protein stability during the extended incubation period. Following de-glycosylation incubation a sample of the mixtures were taken, heat inactivated at 95 °C for 10 minutes and SDS-PAGE analysis was performed. After confirmation of activity the samples were diluted in PBS and concentrated with 50kDa cut-off.

7.6.5 TACE inhibition assay

For the GP TACE inhibition assay, HEK293T cells and HEK293 cells stably expressing TLR4/MD2/CD14 were cultured as previously described (7.3). The cells were seeded at 1.5E5 cells/ml on 96-well plate format 48 hours before transfection. HEK293T cells were transfected with full length GP in pcDNA and TACE in pRK5F, while stable HEK293 cells were transfected with plasmids carrying NFκB promotor luciferase and renilla luciferase at 1 µg of DNA and 2 µl of PEI per 10 wells as previously described in section 7.6.1. 24 hours post-transfection the culture media was removed, and the cells were washed with 1 x sterile PBS. After removal of the wash solution, 20 µl of TrypLE Express (Gibco) were added per well and incubated for 5 minutes. The detached cell cultures of HEK293T expressing GP and TACE and the stable HEK293 cells were merged and transferred to a new 96-well plate. To enable co-culturing and TACE inhibition, 10 µl of 10 mM (500 µg/100 µl) InSolution TAPI-1-TACE inhibitor (Calbiochem) and 150 µl of growth medium were added to make up a final volume of 200 µl per well. The two cell lines were then co-cultured for 24 hours at 37 °C (Figure 7.4). The established harvesting, storing and read out procedures for luciferase assays were followed from this point onwards.

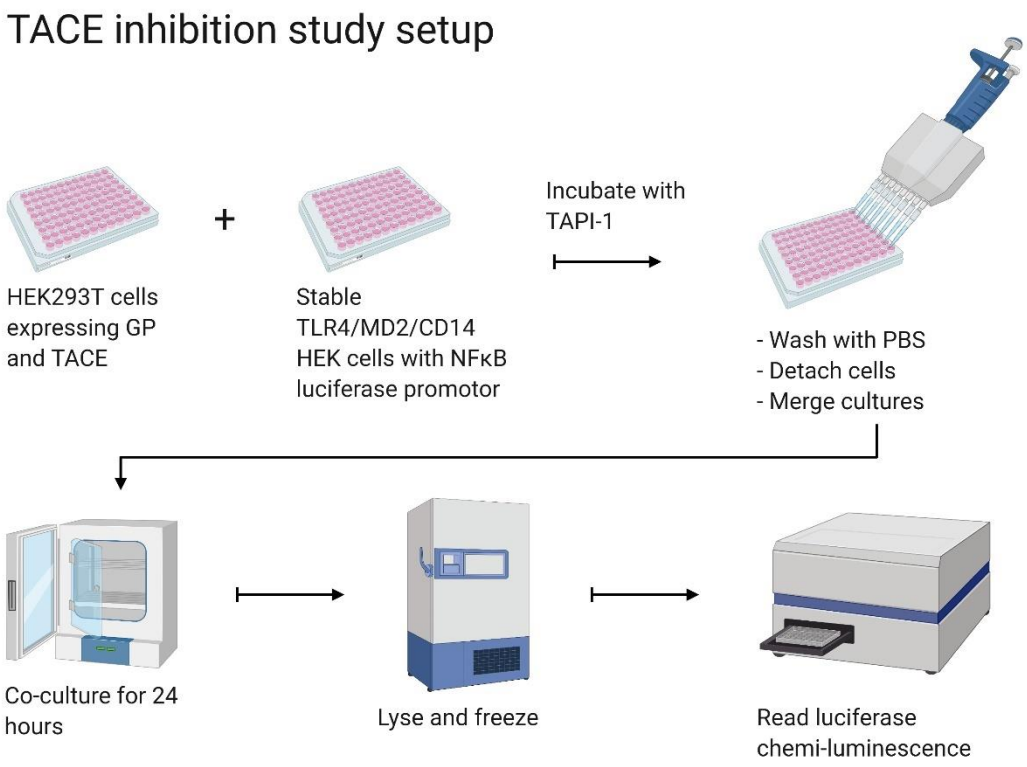


Figure 7.4: Schematic depiction of TACE inhibition TLR4 activation assay

7.7 Statistics

All experiments were conducted in biological triplicates or more and the quantification of luciferase assays was, unless otherwise stated based on technical repeats of 5 or more from each biological replicate. The results are presented as means of each normalised measurement and indicated standard error. Statistical significance was assessed by using a one-way ANOVA test and the p-values presented.

7.8 Safety procedure for EBOV GP

The expression of all Ebola proteins was conducted in the biosafety category lvl 2 (BSL2) tissue culture room of the Department of Biochemistry. All other experiments were conducted in specifically designated zones in the laboratory of Prof Nick Gay whilst wearing appropriate personal safety equipment and with clearly visible signs warning about the protein handled in the environment. All surfaces around the handling were cleaned generously with 1% Virkon and Ethanol, followed by brief washing with water. Disposable equipment such as tubes were collected in a separate container, labelled and double autoclave packed.

8 References

1. Breman JG, Heymann DL, Lloyd G, et al. Discovery and Description of Ebola Zaire Virus in 1976 and Relevance to the West African Epidemic during 2013-2016. *J Infect Dis*. 2016;214:S93-S101. doi:10.1093/INFDIS/JIW207

2. Ellis DS, Simpson IH, Francis DP, et al. Ultrastructure of Ebola virus particles in human liver. *J Clin Pathol*. 1978;31(3):201-208. doi:10.1136/JCP.31.3.201

3. Rojas M, Monsalve DM, Pacheco Y, et al. Ebola virus disease: An emerging and re-emerging viral threat. *J Autoimmun*. 2020;106(November 2019):102375. doi:10.1016/j.jaut.2019.102375

4. Martell HJ, Masterson SG, McGreig JE, Michaelis M, Wass MN. Is the Bombali virus pathogenic in humans? *Bioinformatics*. 2019;35(19):3553-3558. doi:10.1093/BIOINFORMATICS/BTZ267

5. Pourrut X, Kumulungui B, Wittmann T, et al. The natural history of Ebola virus in Africa. *Microbes Infect*. 2005;7(7-8):1005-1014. doi:10.1016/J.MICINF.2005.04.006

6. Leroy EM, Kumulungui B, Pourrut X, et al. Fruit bats as reservoirs of Ebola virus. *Nature*. 2005;438(7068):575-576. doi:10.1038/438575A

7. Baseler L, Chertow DS, Johnson KM, Feldmann H, Morens DM. The Pathogenesis of Ebola Virus Disease∗. *Annu Rev Pathol Mech Dis*. 2017;12:387-418. doi:10.1146/annurev-pathol-052016-100506

8. Goldstein T, Anthony SJ, Gbakima A, et al. The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. *Nat Microbiol*. 2018;3(10):1084-1089. doi:10.1038/S41564-018-0227-2

9. Heeney JL. Ebola: Hidden reservoirs. *Nature*. 2015;527(7579):453-455. doi:10.1038/527453A

10. Purpura LJ, Choi MJ, Rollin PE. Zika virus in semen: lessons from Ebola. *Lancet Infect Dis*. 2016;16(10):1107-1108. doi:10.1016/S1473-3099(16)30330-9

11. Paessler S, Walker DH. Pathogenesis of the Viral Hemorrhagic Fevers. *Annu Rev Pathol Mech Dis*. 2012;8(1):411-440. doi:10.1146/annurev-pathol-020712-164041

12. Leligdowicz A, Fischer WA, Uyeki TM, et al. Ebola virus disease and critical illness. *Crit Care*. 2016;20(1). doi:10.1186/s13054-016-1325-2

13. Paessler S, Walker DH. Pathogenesis of the Viral Hemorrhagic Fevers. *Annu Rev Pathol Mech Dis*. 2013;8(1):411-440. doi:doi:10.1146/annurev-pathol-020712-164041

14. Falasca L, Agrati C, Petrosillo N, et al. Molecular mechanisms of Ebola virus pathogenesis: Focus on cell death. *Cell Death Differ*. 2015;22(8):1250-1259. doi:10.1038/cdd.2015.67

15. Francica JR, Varela-Rohena A, Medvec A, Plesa G, Riley JL, Bates P. Steric shielding of surface epitopes and impaired immune recognition induced by the Ebola virus glycoprotein. *PLoS Pathog*. 2010;6(9). doi:10.1371/journal.ppat.1001098

16. Ning YJ, Deng F, Hu Z, Wang H. The roles of ebolavirus glycoproteins in viral pathogenesis. *Virol Sin*. 2017;32(1):3-15. doi:10.1007/s12250-016-3850-1

17. Francica J. A study of the Ebola virus glycoprotein: Disruption of host surface protein function and evasion of immune responses. *Publicly Access Penn Diss*. Published online 2010. http://repository.upenn.edu/edissertations/198

18. Baseler L, Chertow DS, Johnson KM, Feldmann H, Morens DM. The Pathogenesis of Ebola Virus Disease. *Annu Rev Pathol*. 2017;12:387-418. doi:10.1146/ANNUREV-PATHOL-052016-100506

19. Misasi J, Sullivan NJ. Camouflage and misdirection: The full-on assault of ebola virus disease. *Cell*. 2014;159(3):477-486. doi:10.1016/j.cell.2014.10.006

20. Escudero-Pérez B, Volchkova VA, Dolnik O, Lawrence P, Volchkov VE. Shed GP of Ebola Virus Triggers Immune Activation and Increased Vascular Permeability. *PLoS Pathog*. 2014;10(11). doi:10.1371/journal.ppat.1004509

21. Iampietro M, Younan P, Nishida A, et al. Ebola virus glycoprotein directly triggers T lymphocyte death despite of the lack of infection. *PLoS Pathog*. 2017;13(5):1-27. doi:10.1371/journal.ppat.1006397

22. Iraqi M, Edri A, Greenshpan Y, et al. N-Glycans Mediate the Ebola Virus-GP1 Shielding of Ligands to Immune Receptors and Immune Evasion. *Front Cell Infect Microbiol*. 2020;10(March):1-11. doi:10.3389/fcimb.2020.00048

23. Lee JE, Fusco ML, Hessell AJ, Oswald WB, Burton DR, Saphire EO. Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature*. 2008;454(7201):177-182. doi:10.1038/nature07082

24. Flyak AI, Kuzmina N, Murin CD, et al. Broadly neutralizing antibodies from human survivors target a conserved site in the ebola virus glycoprotein hr2-mper region. *Nat Microbiol*. 2018;3(6):670-677. doi:10.1038/s41564-018-0157-z

25. Woolsey C, Geisbert TW. Current state of Ebola virus vaccines: A snapshot. *PLoS Pathog*. 2021;17(12). doi:10.1371/JOURNAL.PPAT.1010078

26. Choi MJ, Cossaboom CM, Whitesell AN, et al. Use of Ebola Vaccine: Recommendations of the Advisory Committee on Immunization Practices, United States, 2020. *MMWR Recomm Reports*. 2021;70(1):1-12. doi:10.15585/MMWR.RR7001A1

27. Ollmann Saphire E. A Vaccine against Ebola Virus. *Cell*. 2020;181(1):6. doi:10.1016/J.CELL.2020.03.011

28. Markham A. REGN-EB3: First Approval. *Drugs*. 2021;81(1):175-178. doi:10.1007/S40265-020-01452-3

29. He L, Chaudhary A, Lin X, et al. Single-component multilayered self-assembling nanoparticles presenting rationally designed glycoprotein trimers as Ebola virus vaccines. *Nat Commun*. 2021;12(1). doi:10.1038/S41467-021-22867-W

30. Tchesnokov E, Feng J, Porter D, Götte M. Mechanism of Inhibition of Ebola Virus RNA-Dependent RNA Polymerase by Remdesivir. *Viruses*. 2019;11(4). doi:10.3390/V11040326

31. Feldmann H, Klenk H, Sanchez A. Molecular biology and evolution of filoviruses. *Arch Virol Suppl*. 1993;7:81-100. doi:10.1007/978-3-7091-9300-6_8

32. Lee JE, Saphire EO. Ebolavirus glycoprotein structure and mechanism of entry. *Future Virol*. 2009;4(6):621-635. doi:10.2217/fvl.09.56

33. Yang Z, Duckers H, Sullivan N, Sanchez A, Nabel G, Nabel G. Identification of the Ebola virus glycoprotein as the main viral determinant of vascular cell cytotoxicity and injury. *Nat Med*. 2000;6(8):886-889. doi:10.1038/78645

34. Goodsell D. Proteins of the Ebola Virus. *Protein Data Bank*. 2014;Molecule o(10). doi:10.2210/rcsb_pdb/mom_2014_10

35. Weissenhorn W, Carfí A, Lee K-H, Skehel JJ, Wiley DC. *Of Ebola in Zaire in 1995 (Centers., 1996), and Core Prevention*. Vol 2. Hunter and Swanstrom; 1998.

36. Sanchez A, Trappier SG, Mahy BWJ, Peters CJ, Nichol ST. *The Virion Glycoproteins of Ebola Viruses Are Encoded in Two Reading Frames and Are Expressed through Transcriptional Editing (Filovirus/Glycoprotein Gene/Phylogenetic Analysis)*. Vol 93.; 1996.

37. Barrientos LG, Martin AM, Rollin PE, Sanchez A. Disulfide bond assignment of the Ebola virus secreted glycoprotein SGP. *Biochem Biophys Res Commun*. 2004;323(2):696-702. doi:10.1016/j.bbrc.2004.08.148

38. Sanchez A, Yang ZY, Xu L, Nabel GJ, Crews T, Peters CJ. Biochemical analysis of the secreted and virion glycoproteins of Ebola virus. *J Virol*. 1998;72(8):6442-6447. http://www.ncbi.nlm.nih.gov/pubmed/9658086

39. Jeffers SA, Sanders DA, Sanchez A. Covalent Modifications of the Ebola Virus Glycoprotein. *J Virol*. 2002;76(24):12463-12472. doi:10.1128/jvi.76.24.12463-12472.2002

40. Gooz M. ADAM-17: The enzyme that does it all. *Crit Rev Biochem Mol Biol*. 2010;45(2):146-169. doi:10.3109/10409231003628015

41. Okumura A, Pitha PM, Yoshimura A, Harty RN. Interaction between Ebola virus glycoprotein and host toll-like receptor 4 leads to induction of proinflammatory cytokines and SOCS1. *J Virol*. 2010;84(1):27-33. doi:10.1128/JVI.01462-09

42. Zhu W, Banadyga L, Emeterio K, Wong G, Qiu X. The roles of ebola virus soluble glycoprotein in replication, pathogenesis, and countermeasure development. *Viruses*. 2019;11(11). doi:10.3390/v11111002

43. Beniac DR, Timothy BF. Structure of the Ebola virus glycoprotein spike within the

virion envelope at 11 Å resolution. *Sci Rep*. 2017;7. doi:10.1038/srep46374

44. Hoffmann M, Hernández MG, Berger E, Marzi A, Pöhlmann S. The glycoproteins of all filovirus species use the same host factors for entry into bat and human cells but entry efficiency is species dependent. *PLoS One*. 2016;11(2). doi:10.1371/journal.pone.0149651

45. Tran EEH, Simmons JA, Bartesaghi A, et al. Spatial localization of the Ebola virus glycoprotein mucin-like domain determined by cryo-electron tomography. *J Virol*. 2014;88(18):10958-10962. doi:10.1128/JVI.00870-14

46. Wang H, Shi Y, Song J, et al. Ebola Viral Glycoprotein Bound to Its Endosomal Receptor Niemann-Pick C1. *Cell*. 2016;164(1-2):258-268. doi:10.1016/j.cell.2015.12.044

47. Kondratowicz AS, Lennemann NJ, Sinn PL, et al. T-cell immunoglobulin and mucin domain 1 (TIM-1) is a receptor for Zaire ebolavirus and Lake Victoria marburgvirus. *Proc Natl Acad Sci U S A*. 2011;108(20):8426-8431. doi:10.1073/pnas.1019030108

48. Jemielity S, Wang JJ, Chan YK, Ahmed AA, Li W. TIM-family Proteins Promote Infection of Multiple Enveloped Viruses through Virion-associated Phosphatidylserine) TIM-family Proteins Promote Infection of Multiple Enveloped Viruses through Virion-associated Phosphatidylserine. Published online 2013. doi:10.1371/journal.ppat.1003232

49. Brunton B, Rogers K, Phillips EK, et al. TIM-1 serves as a receptor for ebola virus in vivo, enhancing viremia and pathogenesis. *PLoS Negl Trop Dis*. 2019;13(6):1-20. doi:10.1371/journal.pntd.0006983

50. Hoenen T, Watt A, Mora A, Feldmann H. Modeling The Lifecycle Of Ebola Virus Under Biosafety Level 2 Conditions With Virus-like Particles Containing Tetracistronic Minigenomes. *J Vis Exp*. 2014;(91). doi:10.3791/52381

51. Spence JS, Krause TB, Mittler E, Jangra RK, Chandran K. Direct Visualization of Ebola Virus Fusion Triggering in the Endocytic Pathway. *MBio*. 2016;7(1). doi:10.1128/mbio.01857-15

52. Gregory SM, Harada E, Liang B, Delos SE, White JM, Tamm LK. Structure and

function of the complete internal fusion loop from Ebolavirus glycoprotein 2. *Proc Natl Acad Sci*. 2011;108(27):11211-11216. doi:10.1073/pnas.1104760108

53. Takeuchi JS, Akira S. Pattern recognition receptors and inflammation. *Cell*. 2010;140(6):805-820. doi:10.1016/J.CELL.2010.01.022

54. Matzinger P. Tolerance, danger, and the extended family. *Annu Rev Immunol*. 1994;12:991-1045. doi:10.1146/ANNUREV.IY.12.040194.005015

55. Vajjhala PR, Ve T, Bentham A, Stacey KJ, Kobe B. The molecular mechanisms of signaling by cooperative assembly formation in innate immunity pathways. *Mol Immunol*. 2017;86:23-37. doi:10.1016/J.MOLIMM.2017.02.012

56. Gay NJ, Gangloff M. Structure and Function of Toll Receptors and Their Ligands. *Annu Rev Biochem*. 2007;76(1):141-165. doi:10.1146/annurev.biochem.76.060305.151318

57. Akira S, Takeda K. Toll-like receptor signalling. *Nat Rev Immunol*. 2004;4(7):499-511. doi:10.1038/NRI1391

58. Medzhitov R. Toll-like receptors and innate immunity. *Nat Rev Immunol*. 2001;1(2):135-145. doi:10.1038/35100529

59. Medzhitov R, Preston-Hurlburt P, Janeway CA. *A Human Homologue of the Drosophila Toll Protein Signals Activation of Adaptive Immunity*.; 1997.

60. Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in Drosophila adults. *Cell*. 1996;86(6):973-983. doi:10.1016/S0092-8674(00)80172-5

61. Beutler B, Greenwald D, Hulmes JD, et al. Identity of tumour necrosis factor and the macrophage-secreted factor cachectin. *Nat 1985 3166028*. 1985;316(6028):552-554. doi:10.1038/316552a0

62. Gay NJ, Symmons MF, Gangloff M, Bryant CE. Assembly and localization of Toll-like receptor signalling complexes. *Nat Rev Immunol*. 2014;14(8):546-558. doi:10.1038/nri3713

63. Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O. The structural basis of

lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature*. 2009;458(7242):1191-1195. doi:10.1038/nature07830

64. Kagan JC, Su T, Horng T, Chow A, Akira S, Medzhitov R. TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta. *Nat Immunol*. 2008;9(4):361-368. doi:10.1038/NI1569

65. Ohto U, Fukase K, Miyake K, Shimizu T. Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. *Proc Natl Acad Sci U S A*. 2012;109(19):7421-7426. doi:10.1073/pnas.1201193109

66. Schenk M, Belisle JT, Modlin RL. TLR2 Looks at Lipoproteins. *Immunity*. 2009;31(6):847-849. doi:10.1016/J.IMMUNI.2009.11.008

67. Yoon S Il, Kurnasov O, Natarajan V, et al. Structural basis of TLR5-flagellin recognition and signaling. *Science (80-)*. 2012;335(6070):859-864. doi:10.1126/SCIENCE.1215584

68. Barton G, Medzhitov R. Toll-like receptors and their ligands. *Curr Top Microbiol Immunol*. 2002;270:81-92. doi:10.1007/978-3-642-59430-4_5

69. Akira S, Hoshino K, Kaisho T. The role of Toll-like receptors and MyD88 in innate immune responses. *J Endotoxin Res*. 2000;6(5):383-387. Accessed October 4, 2021. https://pubmed.ncbi.nlm.nih.gov/11521059/

70. Fitzgerald KA, Palsson-Mcdermott EM, Bowie AG, et al. Mal (MyD88-adapter-like) is required for toll-like recepfor-4 signal transduction. *Nature*. 2001;413(6851):78-83. doi:10.1038/35092578

71. Yamamoto M, Sato S, Hemmi H, et al. Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. *Science*. 2003;301(5633):640-643. doi:10.1126/SCIENCE.1087262

72. Fitzgerald KA, Rowe DC, Barnes B, et al. LPS-TLR4 signaling to IRF-3/7 and NF-kappaB involves the toll adapters TRAM and TRIF. *J Exp Med*. 2003;198(7):1043-1055. doi:10.1084/JEM.20031023

73. Couillault C, Pujol N, Reboul J, et al. TLR-independent control of innate immunity in Caenorhabditis elegans by the TIR domain adaptor protein TIR-1, an ortholog of

human SARM. *Nat Immunol*. 2004;5(5):488-494. doi:10.1038/NI1060

74. Troutman TD, Hu W, Fulenchek S, et al. Role for B-cell adapter for PI3K (BCAP) as a signaling adapter linking Toll-like receptors (TLRs) to serine/threonine kinases PI3K/Akt. *Proc Natl Acad Sci U S A*. 2012;109(1):273-278. doi:10.1073/PNAS.1118579109

75. Lauenstein J, Scherm M, Udgata A, Moncrieffe MC, Fisher D, Gay NJ. Negative Regulation of TLR Signaling by BCAP Requires Dimerization of Its DBB Domain. *J Immunol*. 2020;204(8):2269-2276. doi:10.4049/JIMMUNOL.1901210

76. Kawai T, Akira S. TLR signaling. *Cell Death Differ*. 2006;13(5):816-825. doi:10.1038/sj.cdd.4401850

77. Gay NJ, Gangloff M, O'Neill LAJ. What the Myddosome structure tells us about the initiation of innate immunity. *Trends Immunol*. 2011;32(3):104-109. doi:10.1016/J.IT.2010.12.005

78. West A, Brodsky I, Rahner C, et al. TLR signalling augments macrophage bactericidal activity through mitochondrial ROS. *Nature*. 2011;472(7344):476-480. doi:10.1038/NATURE09973

79. Symons A, Beinke S, Ley S. MAP kinase kinase kinases and innate immunity. *Trends Immunol*. 2006;27(1):40-48. doi:10.1016/J.IT.2005.11.007

80. Meylan E, Burns K, Hofmann K, et al. RIP1 is an essential mediator of Toll-like receptor 3-induced NF-kappa B activation. *Nat Immunol*. 2004;5(5):503-507. doi:10.1038/NI1061

81. Verstak B, Stack J, Ve T, et al. The TLR signaling adaptor TRAM interacts with TRAF6 to mediate activation of the inflammatory response by TLR4. *J Leukoc Biol*. 2014;96(3):427-436. doi:10.1189/JLB.2A0913-487R

82. Pasare C, Medzhitov R. Toll-like receptors: Linking innate and adaptive immunity. *Adv Exp Med Biol*. 2005;560:11-18. doi:10.1007/0-387-24180-9_2

83. Rawlings D, Schwartz M, Jackson S, Meyer-Bahlburg A. Integration of B cell responses through Toll-like receptors and antigen receptors. *Nat Rev Immunol*. 2012;12(4):282-294. doi:10.1038/NRI3190

84. Corr S, O'Neill LAJ. Genetic variation in Toll-like receptor signalling and the risk of inflammatory and immune diseases. *J Innate Immun*. 2009;1(4):350-357. doi:10.1159/000200774

85. Dauphinee SM, Karsan A. Lipopolysaccharide signaling in endothelial cells. *Lab Invest*. 2006;86(1):9-22. doi:10.1038/labinvest.3700366

86. Miller SI, Ernst RK, Bader MW. LPS, TLR4 and infectious disease diversity. *Nat Rev Microbiol*. 2005;3(1):36-46. doi:10.1038/NRMICRO1068

87. Shuto T, Kato K, Mori Y, et al. Membrane-anchored CD14 is required for LPS-induced TLR4 endocytosis in TLR4/MD-2/CD14 overexpressing CHO cells. *Biochem Biophys Res Commun*. 2005;338(3):1402-1409. doi:10.1016/j.bbrc.2005.10.102

88. Latty S, Sakai J, Hopkins L, et al. Activation of Toll-like receptors nucleates assembly of the MyDDosome signaling hub. *Elife*. 2018;7. doi:10.7554/ELIFE.31377

89. Liu Y, Yin H, Zhao M, Lu Q. TLR2 and TLR4 in autoimmune diseases: a comprehensive review. *Clin Rev Allergy Immunol*. 2014;47(2):136-147. doi:10.1007/S12016-013-8402-Y

90. Modhiran N, Watterson D, Muller DA, et al. Dengue virus NS1 protein activates cells via Toll-like receptor 4 and disrupts endothelial cell monolayer integrity. *Sci Transl Med*. 2015;7(304). doi:10.1126/scitranslmed.aaa3863

91. Štros M. HMGB proteins: Interactions with DNA and chromatin. *Biochim Biophys Acta - Gene Regul Mech*. 2010;1799(1-2):101-113. doi:10.1016/j.bbagrm.2009.09.008

92. Park JS, Svetkauskaite D, He Q, et al. Involvement of Toll-like receptors 2 and 4 in cellular activation by high mobility group box 1 protein. *J Biol Chem*. 2004;279(9):7370-7377. doi:10.1074/jbc.M306793200

93. Medeiros M, Peixoto J, Oliveira A, et al. Toll-like receptor 4 (TLR4)-dependent proinflammatory and immunomodulatory properties of the glycoinositolphospholipid (GIPL) from Trypanosoma cruzi. *J Leukoc Biol*. 2007;82(3):488-496. doi:10.1189/JLB.0706478

94. Wang W, Deng Z, Wu H, et al. A small secreted protein triggers a TLR2/4-dependent inflammatory response during invasive Candida albicans infection. *Nat Commun*.

2019;10(1). doi:10.1038/S41467-019-08950-3

95. Rallabhandi P, Phillips RL, Boukhvalova MS, et al. Respiratory syncytial virus fusion protein-induced toll-like receptor 4 (TLR4) signaling is inhibited by the TLR4 antagonists rhodobacter sphaeroides lipopolysaccharide and eritoran (E5564) and requires direct interaction with MD-2. *MBio*. 2012;3(4). doi:10.1128/MBIO.00218-12

96. Harrison SC. Viral membrane fusion. *Virology*. 2015;479-480:498-507. doi:10.1016/J.VIROL.2015.03.043

97. Ohimain E, Silas-Olu D. The 2013-2016 Ebola virus disease outbreak in West Africa. *Curr Opin Pharmacol*. 2021;60:360-365. doi:10.1016/J.COPH.2021.08.002

98. Saghazadeh A, Rezaei N. Implications of Toll-like receptors in Ebola infection. *Expert Opin Ther Targets*. 2017;21(4):415-425. doi:10.1080/14728222.2017.1299128

99. Saghazadeh A, Rezaei N. Implications of Toll-like receptors in Ebola infection. *Expert Opin Ther Targets*. 2017;21(4):415-425. doi:10.1080/14728222.2017.1299128

100. Simmons G, Wool-Lewis RJ, Baribaud F, Netter RC, Bates P. Ebola virus glycoproteins induce global surface protein down-modulation and loss of cell adherence. *J Virol*. 2002;76(5):2518-2528. http://www.ncbi.nlm.nih.gov/pubmed/11836430

101. Simon EJ, Linstedt AD. Site-specific glycosylation of Ebola virus glycoprotein by human polypeptide GalNAc-transferase 1 induces cell adhesion defects. *J Biol Chem*. 2018;293(51):19866-19873. doi:10.1074/jbc.RA118.005375

102. Wahl-Jensen VM, Afanasieva TA, Seebach J, Ströher U, Feldmann H, Schnittler H-J. Effects of Ebola Virus Glycoproteins on Endothelial Cell Activation and Barrier Function. *J Virol*. 2005;79(16):10442-10450. doi:10.1128/JVI.79.16.10442-10450.2005

103. Okumura A, Pitha PM, Yoshimura A, Harty RN. Interaction between Ebola Virus Glycoprotein and Host Toll-Like Receptor 4 Leads to Induction of Proinflammatory Cytokines and SOCS1. *J Virol*. 2009;84(1):27-33. doi:10.1128/jvi.01462-09

104. Olejnik J, Forero A, Deflubé LR, et al. Ebolaviruses Associated with Differential Pathogenicity Induce Distinct Host Responses in Human Macrophages. *J Virol*.

2017;91(11). doi:10.1128/jvi.00179-17

105. Clarke EC, Collar AL, Ye C, et al. Production and Purification of Filovirus Glycoproteins in Insect and Mammalian Cell Lines. *Sci Rep*. 2017;7(1):1-11. doi:10.1038/s41598-017-15416-3

106. Francica JR, Matukonis MK, Bates P. Requirements for cell rounding and surface protein down-regulation by Ebola virus glycoprotein. *Virology*. 2009;383(2):237-247. doi:10.1016/j.virol.2008.10.029

107. González-Hernández M, Hoffmann M, Brinkmann C, et al. A GXXXA Motif in the Transmembrane Domain of the Ebola Virus Glycoprotein Is Required for Tetherin Antagonism. *J Virol*. 2018;92(13):e00403-18. doi:10.1128/jvi.00403-18

108. Zhu J. Mammalian cell protein expression for biopharmaceutical production. *Biotechnol Adv*. 2012;30(5):1158-1170. doi:10.1016/J.BIOTECHADV.2011.08.022

109. Sauter K, Brcic M, Franchini M, Jungi T. Stable transduction of bovine TLR4 and bovine MD-2 into LPS-nonresponsive cells and soluble CD14 promote the ability to respond to LPS. *Vet Immunol Immunopathol*. 2007;118(1-2):92-104. doi:10.1016/J.VETIMM.2007.04.017

110. Collar AL, Clarke EC, Anaya E, et al. Comparison of N- and O-linked glycosylation patterns of ebolavirus glycoproteins. *Virology*. 2017;502(October 2016):39-47. doi:10.1016/j.virol.2016.12.010

111. Hedfalk K. Further advances in the production of membrane proteins in Pichia pastoris. *Bioengineered*. 2013;4(6):363. doi:10.4161/BIOE.23886

112. Hedfalk K. Codon optimisation for heterologous gene expression in yeast. *Methods Mol Biol*. 2012;866:47-55. doi:10.1007/978-1-61779-770-5_5

113. Wang S, Song X, Zhang K, et al. Overexpression of Toll-Like Receptor 4 Affects Autophagy, Oxidative Stress, and Inflammatory Responses in Monocytes of Transgenic Sheep. *Front Cell Dev Biol*. 2020;0:248. doi:10.3389/FCELL.2020.00248

114. Alazard-Dany N, Volchkova V, Reynard O, et al. Ebola virus glycoprotein GP is not cytotoxic when expressed constitutively at a moderate level. *J Gen Virol*. 2006;87(5):1247-1257. doi:10.1099/vir.0.81361-0

115. Clarke EC, Collar AL, Ye C, et al. Production and Purification of Filovirus Glycoproteins in Insect and Mammalian Cell Lines. doi:10.1038/s41598-017-15416-3

116. Stern B, Olsen LC, Tröße C, Ravneberg H, Pryme IF. Improving mammalian cell factories : The selection of signal peptide has a major impact on recombinant protein synthesis and secretion in mammalian cells. *Trends Cell Mol Biol*. 2007;2:1-17. doi:10.1007/978-1-4020-5476-1_41

117. Uhlén M, Forsberg G, Moks T, Hartmanis M, Nilsson B. Fusion proteins in biotechnology. *Curr Opin Biotechnol*. 1992;3(4):363-369. doi:10.1016/0958-1669(92)90164-E

118. Carter J, Zhang J, Dang T-L, et al. Fusion partners can increase the expression of recombinant interleukins via transient transfection in 2936E cells. *Protein Sci*. 2010;19(2):357-362. doi:10.1002/PRO.307

119. Nallamsetty S, Austin B, Penrose K, Waugh D. Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of Escherichia coli. *Protein Sci*. 2005;14(12):2964-2971. doi:10.1110/PS.051718605

120. Lee J, Kim S. High-throughput T7 LIC vector for introducing C-terminal poly-histidine tags with variable lengths without extra sequences. *Protein Expr Purif*. 2009;63(1):58-61. doi:10.1016/J.PEP.2008.09.005

121. Lee JE, Fusco ML, Abelson DM, Hessell AJ, Burton DR, Saphire EO. Techniques and tactics used in determining the structure of the trimeric ebolavirus glycoprotein. *Acta Crystallogr Sect D Biol Crystallogr*. 2009;65(11):1162-1180. doi:10.1107/S0907444909032314

122. Dalton AC, Barton WA. Over-expression of secreted proteins from mammalian cell lines. *Protein Sci*. 2014;23(5):517-525. doi:10.1002/PRO.2439

123. Volchkov VE, Feldmann H, Volchkova VA, Klenk H-D. Processing of the Ebola virus glycoprotein by the proprotein convertase furin. *Proc Natl Acad Sci*. 2002;95(10):5762-5767. doi:10.1073/pnas.95.10.5762

124. Allen M, Boyce J, Trentalange M, et al. Identification of novel small molecule enhancers of protein production by cultured mammalian cells. *Biotechnol Bioeng*.

2008;100(6):1193-1204. doi:10.1002/BIT.21839

125. Becker B, Cooper M. Aminoglycoside antibiotics in the 21st century. *ACS Chem Biol*. 2013;8(1):105-115. doi:10.1021/CB3005116

126. Jenkins N, Meleady P, Tyther R, Murphy L. Strategies for analysing and improving the expression and quality of recombinant proteins made in mammalian cells. *Biotechnol Appl Biochem*. 2009;53(Pt 2):73. doi:10.1042/BA20080258

127. Frenzel A, Hust M, Schirrmann T. Expression of recombinant antibodies. *Front Immunol*. 2013;4(JUL). doi:10.3389/FIMMU.2013.00217

128. Lufino M, Edser P, Wade-Martins R. Advances in high-capacity extrachromosomal vector technology: episomal maintenance, vector delivery, and transgene expression. *Mol Ther*. 2008;16(9):1525-1538. doi:10.1038/MT.2008.156

129. Sugase K, Landes M, Wright P, Martinez-Yamout M. Overexpression of post-translationally modified peptides in Escherichia coli by co-expression with modifying enzymes. *Protein Expr Purif*. 2008;57(2):108-115. doi:10.1016/J.PEP.2007.10.018

130. Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnol J*. 2012;7(5):620-634. doi:10.1002/biot.201100155

131. Evan G, Lewis G, Ramsay G, Bishop J. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol Cell Biol*. 1985;5(12):3610-3616. doi:10.1128/MCB.5.12.3610-3616.1985

132. Wilson IA, Niman H, Houghten R, Cherenson A, Connolly M, Lerner R. The structure of an antigenic determinant in a protein. *Cell*. 1984;37(3):767-778. doi:10.1016/0092-8674(84)90412-4

133. Hopp TP, Prickett KS, Price VL, et al. A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Bio/Technology 1988 610*. 1988;6(10):1204-1210. doi:10.1038/nbt1088-1204

134. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S. Comparison of affinity tags for protein purification. *Protein Expr Purif*. 2005;41(1):98-105. doi:10.1016/j.pep.2005.01.019

135. Brewer SJ, Sassenfeld HM. The purification of recombinant proteins using C-terminal polyarginine fusions. *Trends Biotechnol*. 1985;3(5):119-122. doi:10.1016/0167-7799(85)90126-X

136. Roth L, Gotsbacher M, Codd R. Immobilized Metal Affinity Chromatography as a Drug Discovery Platform for Metalloenzyme Inhibitors. *J Med Chem*. 2020;63(20):12116-12127. doi:10.1021/ACS.JMEDCHEM.0C01541

137. Carson M, Johnson D, McDonald H, Brouillette C, Delucas L. His-tag impact on structure. *Acta Crystallogr D Biol Crystallogr*. 2007;63(Pt 3):295-301. doi:10.1107/S0907444906052024

138. Korndörfer I, Schlehuber S, Skerra A. Structural mechanism of specific ligand recognition by a lipocalin tailored for the complexation of digoxigenin. *J Mol Biol*. 2003;330(2):385-396. doi:10.1016/S0022-2836(03)00573-4

139. Duplay P, Hofnung M. Two regions of mature periplasmic maltose-binding protein of Escherichia coli involved in secretion. *J Bacteriol*. 1988;170(10):4445-4450. doi:10.1128/JB.170.10.4445-4450.1988

140. Riggs P. Expression and purification of recombinant proteins by fusion to maltose-binding protein. *Mol Biotechnol*. 2000;15(1):51-63. doi:10.1385/MB:15:1:51

141. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol*. 1999;17(10):1030-1032. doi:10.1038/13732

142. Barth H, Boyes B. Size exclusion chromatography. *Anal Chem*. 1992;64(12):428 R-442 R. doi:10.1021/AC00036A023

143. Jungbauer A, Hahn R. Ion-exchange chromatography. *Methods Enzymol*. 2009;463(C):349-371. doi:10.1016/S0076-6879(09)63022-6

144. Parks T, Leuther K, Howard E, Johnston S, Dougherty W. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal Biochem*. 1994;216(2):413-417. doi:10.1006/ABIO.1994.1060

145. Malakhov M, Mattern M, Malakhova O, Drinker M, Weeks S, Butt T. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J*

Struct Funct Genomics. 2004;5(1-2):75-86. doi:10.1023/B:JSFG.0000029237.70316.52

146. Sticha K, Sieg C, Bergstrom C, Hanna P, Wagner C. Overexpression and large-scale purification of recombinant hamster polymorphic arylamine N-acetyltransferase as a dihydrofolate reductase fusion protein. *Protein Expr Purif*. 1997;10(1):141-153. doi:10.1006/PREP.1997.0734

147. Block H, Maertens B, Spriestersbach A, et al. Immobilized-metal affinity chromatography (IMAC): a review. *Methods Enzymol*. 2009;463(C):439-473. doi:10.1016/S0076-6879(09)63027-5

148. Webb SR, Smith SE, Fried MG, Dutch RE. Transmembrane Domains of Highly Pathogenic Viral Fusion Proteins Exhibit Trimeric Association In Vitro. *mSphere*. 2018;3(2):e00047-18. doi:10.1128/msphere.00047-18

149. Hacke M, Björkholm P, Hellwig A, et al. Inhibition of Ebola virus glycoprotein-mediated cytotoxicity by targeting its transmembrane domain and cholesterol. *Nat Commun*. 2015;6(May). doi:10.1038/ncomms8688

150. Lee JE, Fusco ML, Abelson DM, Hessell AJ, Burton DR, Saphire EO. Techniques and tactics used in determining the structure of the trimeric ebolavirus glycoprotein. *Acta Crystallogr Sect D Biol Crystallogr*. 2009;65(11):1162-1180. doi:10.1107/S0907444909032314

151. Schmidt M, Raghavan B, Müller V, et al. Crucial role for human Toll-like receptor 4 in the development of contact allergy to nickel. *Nat Immunol*. 2010;11(9):814-819. doi:10.1038/ni.1919

152. Rutten L, Gilman MSA, Blokland S, Juraszek J, McLellan JS, Langedijk JPM. Structure-Based Design of Prefusion-Stabilized Filovirus Glycoprotein Trimers. *Cell Rep*. 2020;30(13):4540-4550.e3. doi:10.1016/j.celrep.2020.03.025

153. Collar AL, Clarke EC, Anaya E, et al. Comparison of N- and O-linked glycosylation patterns of ebolavirus glycoproteins. *Virology*. 2017;502:39-47. doi:10.1016/j.virol.2016.12.010

154. Tran EEH, Simmons J a, Bartesaghi A, et al. Spatial localization of the Ebola virus

glycoprotein mucin-like domain determined by cryo-electron tomography. *J Virol*. 2014;88(18):10958-10962. doi:10.1128/JVI.00870-14

155. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput*. 2002;310(22).

156. Olsen J, Ong S, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics*. 2004;3(6):608-614. doi:10.1074/MCP.T400003-MCP200

157. Luo Q, Chen D, Boom RM, Janssen AEM. Revisiting the enzymatic kinetics of pepsin using isothermal titration calorimetry. *Food Chem*. 2018;268:94-100. doi:10.1016/J.FOODCHEM.2018.06.042

158. Volchkov VE, Volchkova VA, Slenczka W, Klenk H, Feldmann H. Release of viral glycoproteins during Ebola virus infection. *Virology*. 1998;245(1):110-119. doi:10.1006/VIRO.1998.9143

159. Akey DL, Brown WC, Dutta S, et al. Flavivirus NS1 crystal structures reveal a surface for membrane association and regions of interaction with the immune system. *Science*. 2014;343(6173):881. doi:10.1126/SCIENCE.1247749

160. Gutsche I, Coulibaly F, Voss JE, et al. Secreted dengue virus nonstructural protein NS1 is an atypical barrel-shaped high-density lipoprotein. *Proc Natl Acad Sci U S A*. 2011;108(19):8003-8008. doi:10.1073/PNAS.1017338108/SUPPL_FILE/PNAS.201017338SI.PDF

161. Chao CH, Wu WC, Lai YC, et al. Dengue virus nonstructural protein 1 activates platelets via Toll-like receptor 4, leading to thrombocytopenia and hemorrhage. *PLoS Pathog*. 2019;15(4). doi:10.1371/JOURNAL.PPAT.1007625

162. Dolnik O, Volchkova V, Garten W, et al. Ectodomain shedding of the glycoprotein GP of Ebola virus. *EMBO J*. 2004;23(10):2175-2184. doi:10.1038/sj.emboj.7600219

163. Trad A, Hansen HP, Shomali M, et al. ADAM17-overexpressing breast cancer cells selectively targeted by antibody-toxin conjugates. *Cancer Immunol Immunother*. 2013;62(3):411-421. doi:10.1007/s00262-012-1346-x

164. Warfield K, Bosio C, Welcher B, et al. Ebola virus-like particles protect from lethal

Ebola virus infection. *Proc Natl Acad Sci U S A*. 2003;100(26):15889-15894. doi:10.1073/PNAS.2237038100

165. Matsunaga N, Tsuchimori N, Matsumoto T, Ii M. TAK-242 (resatorvid), a small-molecule inhibitor of Toll-like receptor (TLR) 4 signaling, binds selectively to TLR4 and interferes with interactions between TLR4 and its adaptor molecules. *Mol Pharmacol*. 2011;79(1):34-41. doi:10.1124/MOL.110.068064

166. Wong G, Audet J, Fernando L, et al. Immunization with vesicular stomatitis virus vaccine expressing the Ebola glycoprotein provides sustained long-term protection in rodents. *Vaccine*. 2014;32(43):5722-5729. doi:10.1016/j.vaccine.2014.08.028

167. Calligaris M, Cuffaro D, Bonelli S, et al. Strategies to target ADAM17 in disease: From its discovery to the iRhom revolution. *Molecules*. 2021;26(4). doi:10.3390/MOLECULES26040944