Kent Academic Repository

Generation of SARS-CoV-2 Coronavirus Protein Antigens for Development of Diagnostics and Novel Vaccine Approaches

By: Emily Rose Burrows

MSc-Res in: Biochemistry

Year of Submission: 2022

Signed:

Total word count: 30,542

School of Biosciences, The University of Kent

Declaration

No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent, or any other University or Institution of learning.

Signed:

Date: 8th February 2022

Acknowledgements

I would like to thank my supervisor Professor Mark Smales for giving me the opportunity to carry out this project and as well for all of the support and advice I have received over the last year.

I am profoundly grateful to Jo Roobol, Sarah Martin, Dr Maria Stanley, the COVID Team and the rest of the Smales Laboratory (University of Kent) for answering all of my endless questions, passing on their expertise and wisdom as well providing me with so much support. To have worked on such a project during a pandemic in which there were multiple lockdowns, shift work and laboratory restrictions was a privilege, and I have learnt so much.

I would like to thank my family and friends, particularly my parents and sister, Martin, Jill and Rosie, for their continuous support and help in which this year would not have been possible without them. To Nathan and the 'Guzzle Gang', thank you for keeping me sane during this year and supporting me when times were especially hard.

Finally, a thank you to Auntie Jane and Grandad Ern for always inspiring me to work hard and achieve my goals. Your work ethics are instilled in me and I will carry this throughout life.

Table of Contents

List of s and Tables

List of Tables

List of Figures

Abbreviations

ACE2 – Angiotensin Converting Enzyme 2

Amp – Ampicillin

B.1.1.7 – SARS-CoV-2 variant originated in the UK

B.1.351 – SARS-CoV-2 variant originated in South Africa

bp – Base Pairs

BSA – Bovine Serum Albumin

CD-CHO – Chemically Defined CHO medium

CHO – Chinese Hamster Ovary

CO_2 – Carbon Dioxide

CoV – Coronavirus

COVID-19 – Coronavirus Disease 2019

Da – Daltons

DNA – Deoxyribonucleic Acid

dNTP's – deoxyribonucleotide triphosphate's

E – Envelope

ECL – Enhanced Chemiluminescence

eGFP – Functional Enhanced Green Fluorescent Protein

ELISA – Enzyme Linked Immunosorbent Assay

EOC – End of Culture

ER – Endoplasmic Reticulum

FL – Full Length

GS – Glutamine Synthetase

H_2O – Water

HRP – Horseradish Peroxidase

hygro – Hygromycin

IgG – Immunoglobulin G

Kb – Kilobases

kDa – kilodaltons

LB – Luria Broth

M – Membrane

mAbs – Monoclonal Antibody

MERS – Middle East Respiratory Syndrome

mGS – mouse GS

ml – millilitres

mRNA – messenger RNA

MSX – Methionine Sulfoximine

NaF – Sodium Fluoride

Nat – Native signal peptide not CHO optomised

NaV – Sodium Orthovanadate

NC – Nucleocapsid

ng/µl – nanograms per microlitre

ORF – Open Reading Frame

OSP- Native CHO Optomised signal peptide

OSP- Native signal peptide that is CHO optomised

PBS – Phosphate Buffered Saline

PCR – Polymerase Chain Reaction

PHE – Public Health England

qRT-PCR – Real Time Quantitative Reverse Transcription PCR

RBD – Receptor Binding Domain

RNA – Ribonucleic Acid

RT – Reverse Transcriptase

S – Spike

S1 – Subunit 1

SARS-CoV – Severe Acute Respiratory Syndrome Coronavirus

SDM – Site Directed Mutagenesis

SDS – Sodium Dodecyl Sulfate

SDS-PAGE – SDS – Polyacrylamide Gel Electrophoresis

SP – Supernatant

TAE – Tris-acetate-EDTA

TBST – Tris-Buffered Saline + Tween 20

TE – Tris- EDTA

TMB – Tetramethylbenzidine

TMPRSS2 – Transmembrane Serine Protease 2

VOC – Variant of Concern

µg – micrograms

µl – microlitres

µM – micromolar

Abstract

The SARS-CoV-2 coronavirus pandemic that is reported to have originated in Wuhan in China in late 2019 has since rapidly spread globally resulting in millions of cases and hundreds of thousands of deaths. A rapid global response to this included nationwide lockdowns in countries with massive economic and social implications alongside an escalating case and death rate. In order to identify/diagnose infected individuals and those that have previously been infected to help in the management and treatment of the disease, there has been a concentrated effort to develop diagnostics for SARS-CoV-2 and antibodies raised against this. There has also been a simultaneous drive to develop vaccines to protect against SARS-CoV-2 infection. Although there are multiple approaches to the design of diagnostics and vaccines, each requires the rapid supply of SARS-CoV-2 protein antigens for us in diagnostics or vaccines or to use in assays in their development. The novel SARS-CoV-2 coronavirus genome encodes a number of structural proteins, including the Spike protein that facilitates binding of the virus to the ACE2 receptor during viral cell entry. The aim of this project was to generate coronavirus antigens for the development of novel vaccines and diagnostic tools, particularly of the Spike protein. This was undertaken by investigating recombinant expression of SARS-CoV-2 Spike protein production via the comparison of using different CHO cell line host expression systems, the comparison of various ER signal peptides, the development of stable CHO cell pools expressing the spike protein and measuring patient antibody responses to different mutant spike proteins. SARS-CoV-2 Spike DNA constructs were generated via molecular cloning and various mutants of the Spike generated. Antibody responses were measured using ELISA assays. The various Spike DNA constructs were successfully constructed into commercial vectors as well as the B.1.1.7 and B.1.351 variants as they emerged during the progress of the pandemic. CHO-S cells were found to be more effective than CHO-K1 cells for spike protein production. The native Spike CHO optimised ER signal peptide was the most effective signal peptide of those evaluated giving the highest expression of Spike protein and higher mRNA transcript amounts. Indeed, secreted protein yield, as assessed by western blot, was highest using constructs that gave the highest mRNA transcript amounts. When generating CHO cells stably expressing Spike protein using a glutamine synthetase selection system, 25 μM MSX levels gave rise to higher productivity in the stable cell pools than 50 μM MSX and ELISA with the recombinant Spike proteins showed a concentration-response relationship with the serum of Wuhan infected patients against various spike variants. Overall, this project was able to successfully develop coronavirus Spike protein antigens and improve the Spike protein production process via varying the CHO host cell line, the ER signal peptide selection, development of stably Spike expressing cell pools and modifying the MSX concentration used during stable cell development, collectively resulting in higher Spike protein production.

Chapter 1: Introduction

1.1 SARS-CoV-2: What are coronaviruses?

1.1.1 A background overview on Coronaviruses

Viruses are sub-microscopic infectious agents that consist of an RNA or DNA genome enclosed within a protein capsid coat, with the ability to impact public health all over the world within a matter of months. These agents enter living cells and take advantage of their hosts molecular machinery to replicate their own genetic material(1). They may be enveloped (surrounded by a lipid bilayer) or non-enveloped in nature.

Being a highly diverse family of enveloped positive-sense single-stranded RNA viruses, Coronaviruses (CoVs) are a challenge not only for public health but also of economic concern. These viruses are able to infect humans, mammals and avian species including companion animals and livestock. The classification of coronaviruses lies in the order of the *Nidovirales* and the suborder of *Coronavirinieae* lies within the family classification of *Coronaviridae*. The family is then further defined into the subfamily of *Orthocoronavirinae* which contains four groups of genera: *alphacoronavirus, betacoronavirus, gammacoronavirus* and *deltacoronavirus.* Within these four genera groups, they affect different species. For alphacoronaviruses and betacoronaviruses, they exclusively infect mammalian species. However for gammacoronaviruses and deltacoronaviruses, their host range widens to include avian species (2).

Among the known coronaviruses, seven have been found to be associated with humans (1). The infections these coronaviruses cause result primarily in respiratory and enteric disease (2). Since the twenty-first century, coronaviruses have had a significant impact on public health with three of the seven existing human coronaviruses forming outbreaks causing epidemics/pandemics, all three specifically betacoronaviruses. These include; the severe acute respiratory syndrome coronavirus (SARS-CoV) in 2003, Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012 and the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in 2019, which is still ongoing today. These viruses were able to achieve large-scale transmission during their major outbreak period. MERS-CoV cases still

remain although they are very low with severely reduced transmission. However, the SARS-CoV-2 coronavirus, that was first detected in Wuhan, China, in late 2019 is still a major threat to global public health with case numbers rising and death tolls increasing significantly every day. In terms of transmission, the SARS-CoV-2 virus has been found to be more difficult to control than the previous outbreaks with 207 million cases reported worldwide with more than 4.3 million deaths as a result (as of 16th August, 2021) (3)

Coronaviruses contain a 5' – capped, positive strand RNA molecule that ranges from 26-32 kb containing 6 open reading frames (ORF's). ORF1a/b (the first ORF) makes up two-thirds of the genome and encodes replicase proteins. ORF1a is also where translation begins and continues into ORF1b after a -1 frameshift signal. The rest of the one third of the genome remaining encodes four structural proteins: Spike (S) glycoprotein, Envelope (E) protein, Membrane (M) protein and Nucleocapsid (N) protein (4). Of particular interest to this project is the Spike or S protein that is involved in cellular entry as described in more detail below.

1.1.2 SARS-CoV-2 Structure and Replication

The first steps of coronavirus infection, more specifically SARS-CoV-2 infection, involves the highly specific binding of the Spike glycoprotein to the cellular entry receptor angiotensin-converting enzyme (ACE2) on the surface of target cells. As these entry receptors are expressed and distributed vastly among tissue, this influences pathogenicity and viral tropism where it productively infects the cells. Within the cell, the virus can begin its intracellular life cycle where the coronavirus expresses and replicates its genomic RNA in order to produce full-length copies that are later incorporated into newly created viral particles (2). Figure 1.1.1 outlines the structure of the viral particle and also the intracellular life cycle of the virus.
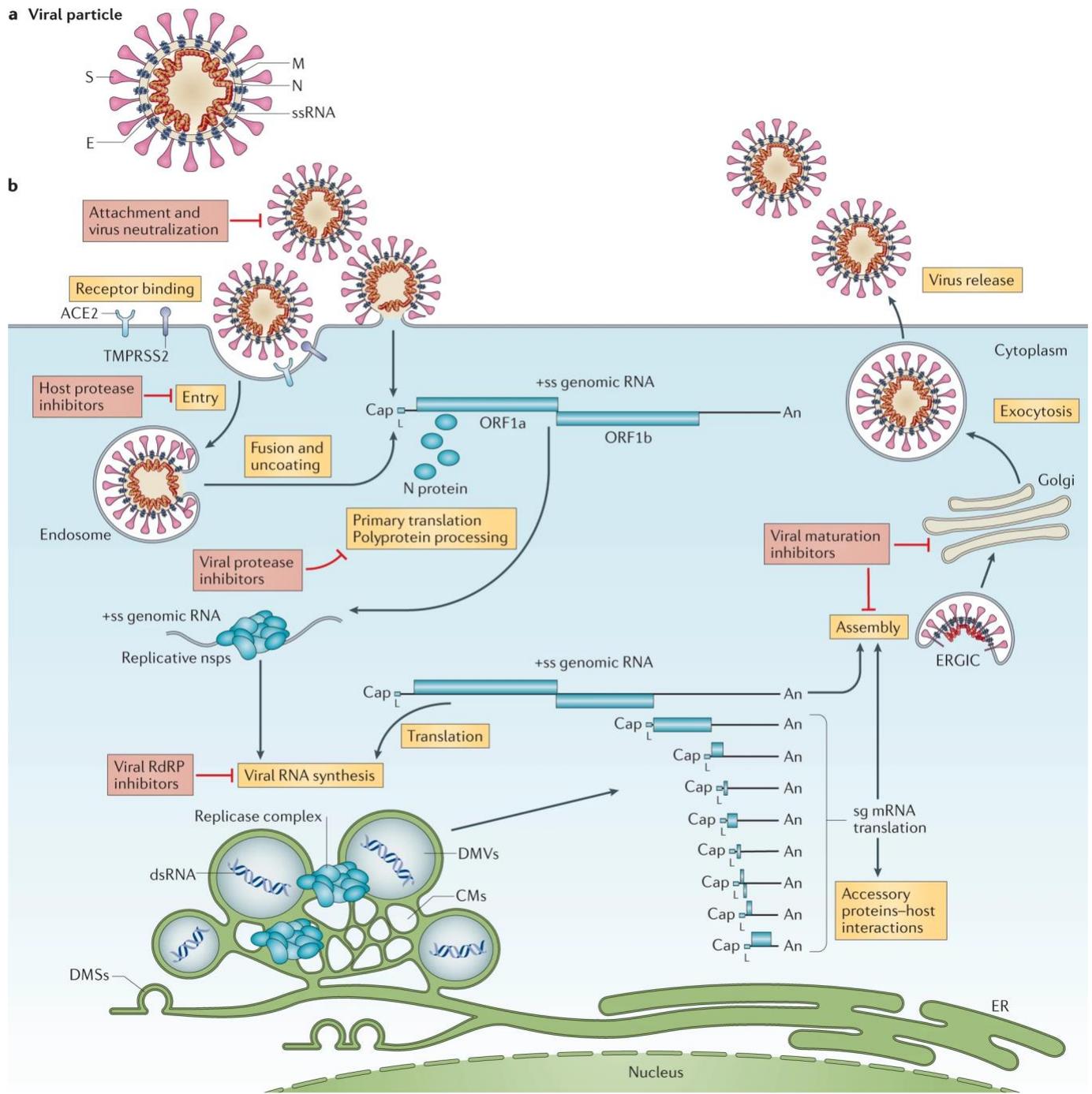
Figure 1.1.1: *Structural components of the coronavirus and also the intracellular life cycle of the coronavirus once infection has occurred. Taken from* (2)

As the infection begins, the SARS-CoV-2 virus targets specific cells notably the bronchial and nasal epithelial and pneumocytes via the S protein binding to the ACE2 receptor on these cells. The TMPRSS2 protein present in the host cell promotes the uptake of the virus by cleaving ACE2 and activating the SARS-CoV-2 S protein mediating entry into the cell (5). There is a high expression of the TMPRSS2 and ACE2 proteins in alveolar epithelial type II cells, essential for SARS-CoV-2 pathogenesis (6). The viral inflammatory response as a result of the infection, consisting of both the adaptive and innate immune response, increases lymphocyte apoptosis and impairs lymphopoiesis (5) hindering the immune response.

As the viral infection spreads, this impacts the integrity of the epithelial-endothelial barrier as well as affecting pulmonary capillary endothelial cells resulting in an increased inflammatory response with an influx of neutrophils and monocytes. The persistent inflammatory response results in increased vascular permeability, oedema and alveolar interstitial thickening. The combination of endothelial barrier disruption, impaired oxygen diffusion capacity and dysfunctional alveolar-capillary oxygen transmission result in the characteristics of COVID-19 disease, with the respiratory tract being severely affected (5).

1.1.3 SARS-CoV-2: The Spike (S) Protein

The Spike protein is a large structural protein with a monomeric size of 180-200 kDa. It is heavily glycosylated and produced as a monomer protein that exists natively as a trimer. The monomer is subjected to proteolytic cleavage processing that is important to its function. The trimer is located on the surface of the viral envelope enabling the S protein to bind to the receptor. The S protein consists of a short intracellular C terminal, an extracellular N-terminus and a transmembrane domain anchored in the viral membrane (Figure 1.1.2A). The spike protein exists in a prefusion conformation as once the virus interacts with the host cell, rearrangement of this protein occurs which consequently allows fusion between the virus and the host cell (7). To evade the immune system surveillance of the host cell during entry, the spike proteins are coated in polysaccharide molecules to act as a camouflage (8).

As shown in Figure 1.1.2B, the spike protein trimer forms a 'bulbous head and stalk' made up of the S1 and S2 subunits. The Receptor Binding Domain (RBD), found in the S1 subunit, binds to the ACE2 receptor in the aminopeptidase N region. The RBD is an important target for

neutralising antibodies. As a result of its structure and role, the S protein has been identified as a target for vaccine development when inducing immunity against viral infection from the SARS-CoV-2 (7) and the majority of vaccines that have been developed, or are in development, to date are based on the S protein.

A

B
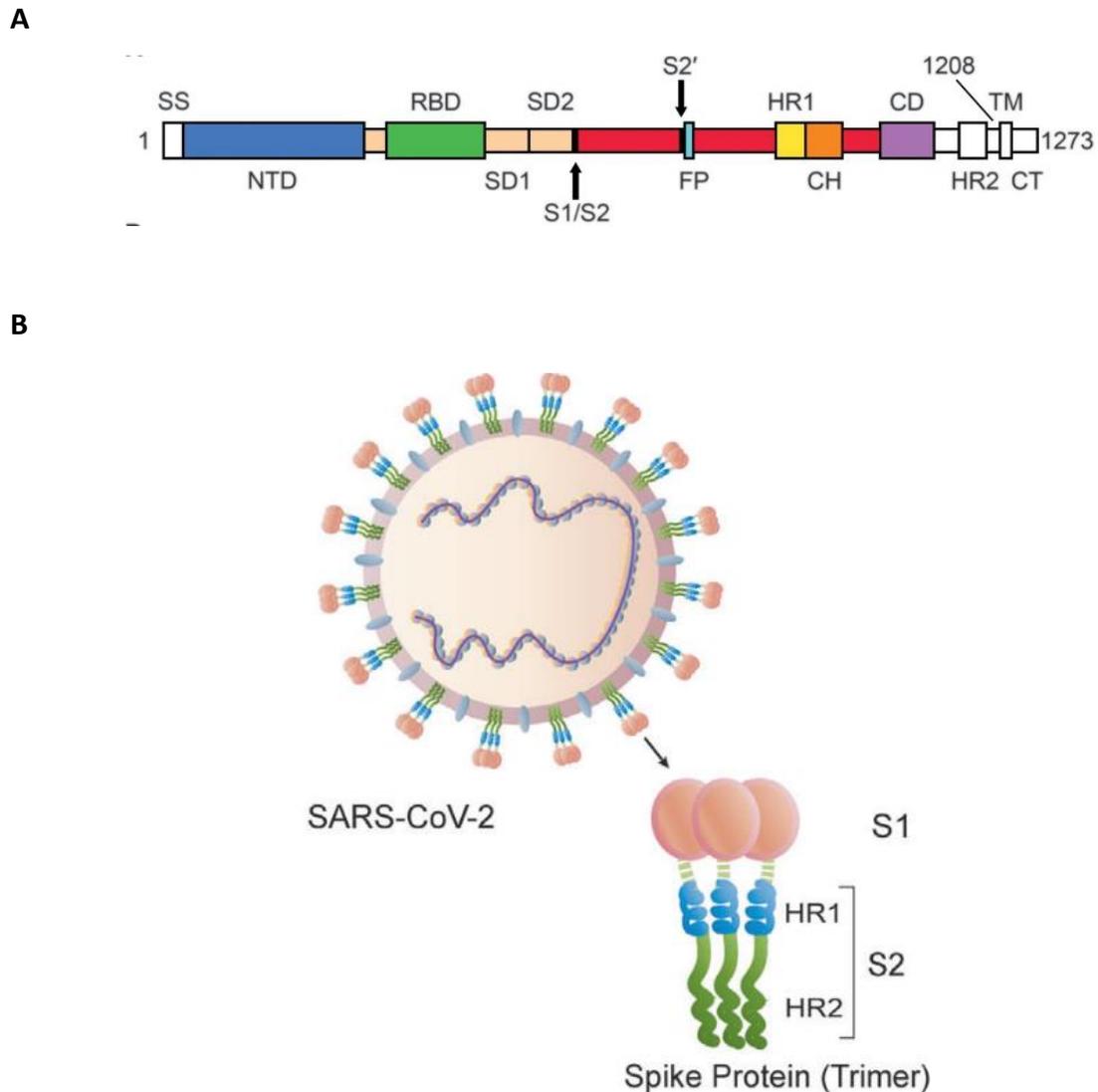
Figure 1.1.2: *(A) Schematic describing the domain makeup of the monomeric S-protein of SARS-CoV-2. SS = ER signal sequence, NTD = N-terminal domain, RBD = receptor binding domain, SD1 = subdomain 1, SD2 = subdomain 2, S1 = S1 protease site, S2 = S2 protease site, FP = fusion protein, HR = heptad repeat, CH = central helix, CD = connector domain, TM = transmembrane domain, CT = cytoplasmic tail. (B) Trimeric structure of the Spike Protein showing the S1 and S2 subunits. Excerpt of figure taken from* (7)

1.1.4 COVID-19: Disease and Transmission

The transmission of the SARS-CoV-2 virus is high due to the virus being able to very successfully enter target cells, mostly thought to be transmitted in droplets expelled during face to face exposure as well as coughing or sneezing being the most effective forms of transmission. Prolonged exposure is associated with higher transmission rates of the disease. Contact surface spread is also another factor with the virus being identified on surfaces for up to 3 to 4 days after inoculation (5). The most common symptoms in hospitalised patients suffering with COVID-19 due to the SARS-CoV-2 virus are fever, with up to 90% of patients experiencing this, dry cough, nausea, vomiting, diarrhoea and myalgia (9).

The long-term effects of COVID-19 are also being recognised. Also known as 'Long Covid', these are symptoms that persist for more than 2 weeks after the onset of the COVID-19 disease and are considered long term effects. A meta-analysis of studies by Lopez-Leon.S *et al* shows 80% of patients with COVID-19 have at least one long-term symptom. The most common were found to be fatigue, headaches, attention disorder and ageusia. (10).

With transmissibility of COVID-19 being high, the disease has become a social phenomenon with the issue of containment of the virus relying heavily on limiting social contact. With responses from public health enforcers emphasising social distancing, stay at home orders and behavioural changes within the individual, not only did this slow the trajectory of viral transmission but also had a serious impact on mental health (11). True quarantine has been shown to have significant effects on mental health including depression, insomnia, post-traumatic stress and generalised anxiety (12). Alongside this, stay at home and social distancing requirements had had enormous economic impacts globally. Therefore, the need for vaccine and diagnostic tools for this pandemic was, and remains, of the highest importance.

1.1.5 The Global Emergence of New Variants

As well as the need to rapidly develop vaccines for the 2019 native (named here as such as the original Wuhan sequenced virus) SARS-CoV-2 virus, the emergence of new variants throughout the pandemic has been of concern. Notably, the B.1.1.7 (from the UK, also referred to as Alpha variant) and B.1.351 (from South Africa, referred to as Beta variant)

variants are/were of notable concern due to their increased transmissibility and various mutations within the spike protein (13). In early 2021 the Alpha variant started to become the dominate variant detected worldwide. However, more recently the now termed Delta variant, that is thought to have originated in India (B.1.617.2) has become the global dominant variant and is the most transmissible and infectious of those variants to have arisen to date. It has rapidly spread globally. This variant arose towards the end of the work undertaken in this thesis.

The B.1.1.7 (Alpha) variant was first detected in September 2020 in Kent, UK. Throughout December 2020, the infections continued to spread throughout the UK with more than 75% of cases in east England and Greater London were estimated to be B.1.1.7. The increasing trend of B.1.1.7 throughout the continents from the first week when the B.1.1.7 variant was reported, the global average days taken to reach a 50% positive rate was 178 days. (14). The reason for the extensive transmissibility of this variant compared to the native Wuhan strain is due to its additional mutations in the spike protein. The mutations are outlined in Table 1.1.1 (15).

Table 1.1.1: *List of spike mutations present in the B.1.1.7 variant. Taken from* (15).

B.1.1.7 Spike Amino Acid Mutation
HV 69-70 deletion
Y144 deletion
N501Y
A570D
P681H
T716I
S982A
D1118H

As shown in Table 1.1.1, there are many mutations that each contribute to the increased transmissibility of this variant including the spike 501 mutation which is a key residues in the RBD, with data suggesting this mutation increases ACE2 receptor affinity (16) thus increasing possible binding to the ACE2 receptor and causing viral replication within cells. The P681H mutation is also of significance as it one of the four residues that makes up the insertion point which creates a furin cleavage site between S1 and S2 in spike. This cleavage site has been shown to facilitate entry into respiratory epithelial cells (17).

The B.1.351 (Beta) variant emerged in South Africa in late 2020, also since becoming dominant locally. 9 spike mutations are present in this variant alongside D614G (18). The B.1.351 variant first emerged in October 2020 and by mid-November, was the dominant variant, superseding the B.1.1.54, B.1.1.56 and C.1 lineages. With over 80,000 excess natural deaths occurring in South Africa by the end of 2020, South Africa is the most largely affected country in Africa (19).

The B.1.617.2 (Delta) variant was first identified in December 2020 in India and within months, was present in 98 countries worldwide. It has been found the Delta variant is between 40-60% more transmissible than the B.1.1.7 variant as well as viral loads of the virus being significantly higher around 1000 higher than other variants. This variant is predominant in countries such as the US and UK with 83% of COVID-19 cases in the US being the B.1.617.2 variant (20). A notable mutation contributing to the efficacy and transmissibility of this variant is P681R. This mutation was found to fuse to the plasma membranes of uninfected cells three times faster than spike proteins that do not possess the P681R mutation, thus speeding up the spread of SARS-CoV-2 from cell to cell (21).

1.1.6 Notable Spike Protein Mutations: D614G and E484K

Of all the SARS-CoV-2 Spike protein mutations that have arisen during the pandemic, the most notable and prevalent is D614G. By early April 2020, it was clear the Spike D614G mutation exhibited behaviours enabling the transmission of the virus such as conferring resistance to neutralising antibodies and altering host species susceptibility. This mutation is found on the surface of the spike protein protomer where contacts are able to be formed with the neighbouring protomer (22). Data from the Korber *et al* 2020 study shows that over the span

of 1 month, the Wuhan variant carrying the D614G spike mutation became the dominant form of SARS-CoV-2 across the globe (22). This mutation has since been found in the B.1.351 variant.

Another notable Spike protein mutation in the B.1.351 and B.1.1.7 variants that enhanced their transmission is the E484K mutation. This mutation impacts the immune system recognition. Whilst the E484K mutation did not initially appear in the B.1.1.7 variant, it appears to have now further mutated with the addition of this variant. Also known as an 'escape' mutation, this mutation helps the virus to evade the immune system (23). With potently neutralising monoclonal antibodies recognising the SARS-CoV-2 RBD being elicited in response to SARS-CoV-2 infection, they are effective in treating and preventing infection (24). The single E4848K mutation in the RBD is able to affect the binding of these neutralising antibodies making the virus more transmissible (25).

1.1.7 The Race for a SARS-CoV-2 Vaccine

One of the most successful and cost-effective public health interventions to prevent disease and the complications this may bring is vaccination. The World Health Organisation (WHO) estimates around 3 million deaths are prevented every year due to vaccination (26). A vaccine is a biological agent that elicits an immune response to a specific antigen which has been derived from an infectious pathogen that causes disease (27). Thanks to advances in molecular biology, immunology and virology, many types of vaccines have been developed including nucleic acid based vaccines (mRNA and DNA), viral vectored vaccines and recombinant vaccines. With 103 million cases of childhood diseases prevented between 1924 and 2010 in the United States through vaccination, this also benefits the economy as children are able to healthily grow up, not need expensive hospitalisation treatment and able to reach adulthood without serious disease as may have happened in the past (28).

In the midst of a once-a-century pandemic, vaccination is important. Even though there are currently anti-viral drugs in clinical trials, these treat infection whereas vaccination prevents infection or at least primes the body to rapidly respond to any infection. Thus, vaccinating as many as possible of the entire population against COVID-19 is/has been considered vital in order to control the pandemic, reduce serious illness and death and allow a return to

normality whereby lockdowns and stay at home legislation are not required to control the virus and the impact on the health of the population. This presents a wide variety of challenges as a vaccine first needs to be developed which is both safe and effective. Then there is the manufacturing and distribution capabilities as well as the administration of the vaccine to consider all within a short timeframe. Some vaccines may also be required to be kept in a 'cold-chain' (e.g. below room temperature and some at temperatures as low as -80°C) which also presents further challenges when it comes to reaching more vulnerable populations (29). Over 100 COVID vaccine candidates are currently under development. As of August 9th, 2021, there were thirteen different types of vaccines currently in use with over 4 billion doses administered globally, a real success story as all of these have been developed since the first outbreak was recognised in early 2020 (3). Traditionally the development and launch of such a vaccine would take many years.

For recombinant protein and vector-based vaccines, the spike protein is the main target. Due to its extracellular location, the S protein can be used as an effective antigen when developing SARS-CoV-2 vaccines, specifically the RBD. Mucosal vaccinations via intranasal or oral routes offers advantages for COVID-19 protection as well as the process of mass immunisation. By formulating vaccines into a dry powder not only reduces the costs of vaccination but also the thermostability of the vaccine making it more accessible to larger numbers of the population as transport and containment does not have to be specific (29). As the Spike protein is the main target for the majority of vaccines now in use and in development, the major focus of this project was in methods to produce recombinant Spike protein that could be made available to help vaccine and diagnostic developments by others.

1.2 Background on Molecular Biology and Cell Culture Systems Used during this Investigation

1.2.1 Molecular cloning to generate plasmids to express recombinant genes of interest

In order to create DNA-based experimental tools which can then be used for expression in mammalian cells, molecular cloning is an essential technique used to produce DNA plasmids containing the exogenous genes of interest to be expressed (e.g. Spike) driven by appropriate promoters for expression in mammalian cells. The most widely used approach involves the preparation of the plasmid vector backbone and target insert DNA, ligation of the insert and

vector, transformation into competent *E. coli* cells and identification of positive clones. Examples of these DNA constructs include a promoter fused to a reporter gene/cDNA sequence which is under the control of a ubiquitous promoter. Vectors used in molecular cloning include various basic features of naturally occurring plasmids such as a replication origin, a drug resistance gene and unique restriction sites allowing the insertion of DNA fragments. An overview of a standard molecular cloning strategy is shown in Figure 1.1.3.
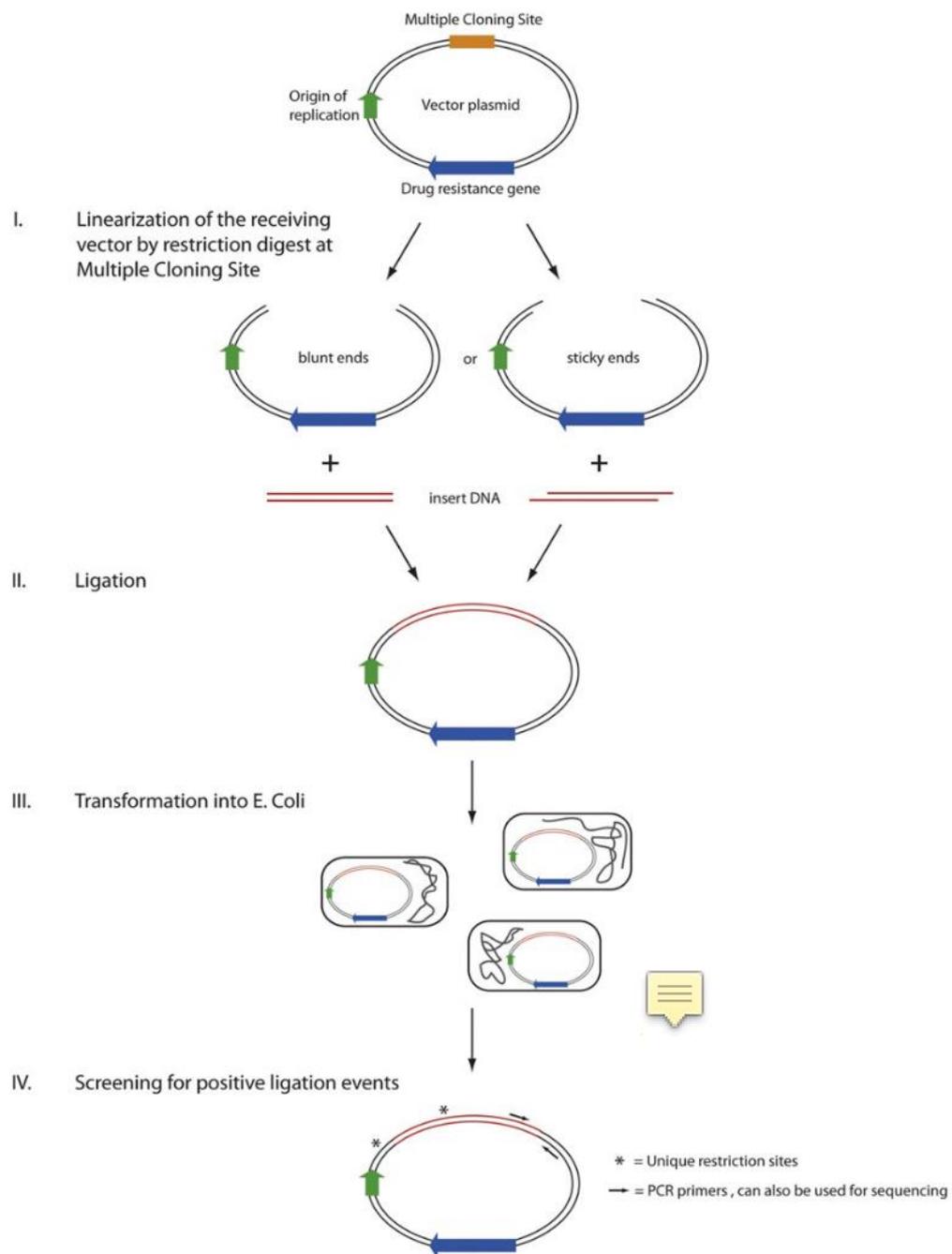
Figure 1.1.3: *Schematic outlining the basic steps involved in generation of a plasmid DNA construct containing a target gene of interest for expression in cell systems*.

Restriction enzyme choice is important when designing a cloning experiment as while some enzymes cut the DNA in one place creating a 'blunt' end, others leave a few bases at the cut site that overhang creating a 'sticky end'. These 'sticky' ends increase the efficiency of ligation thus increasing the likelihood of a successful cloning event (30). Further, restriction enzymes must be selected that do not cut within the sequence of the gene of interest.

1.2.2 Site Directed Mutagenesis

To understand protein function and structure, protein variants sometimes need to be explored such as point mutations, deletions and domain swapping and truncations. Once the DNA is amplified by PCR, these modifications to a gene sequence can be made to replicate such mutations. If internal mutations or deletions are required, different strategies such as site directed mutagenesis need to be applied. This method requires two complementary primers that contain the specific mutation in the middle of the sequence as well as the circular target plasmid DNA. With 10-15 bases flanked either side of the mutation, these anneal to the template DNA. After cycles of DNA annealing and extension the entire plasmid DNA containing the newly added mutations is then amplified. *DpnI* can then be used to reduce background transformation (31). This approach was used in the work described in this thesis to introduced reported mutations to the Spike gene/protein.

1.2.3 Mammalian Cell Culture: CHO Cells in Recombinant Protein Production

When using mammalian cell culture expression systems, Chinese hamster ovary (CHO) cells are the most commonly used mammalian host for commercial production of therapeutic proteins (32). The CHO cell line has many advantages for mammalian protein production due to its ability to synthesise, fold, assemble, secrete, and post-translationally modify proteins with human -like modifications with it being the most commonly used cell line for both stable and transient transfection of such molecules (33). Around 70% of recombinant protein therapeutics are now produced from CHO cells (34). The popularity of the CHO cell line can be attributed to a number of reasons in addition to those listed above. These include it being a 'safe' and well know host to the regulators making approval of biotherapeutic proteins easily attainable as well as the cells ability to be adapted to growth in serum-free suspension conditions, something that is applicable on a large scale when it comes to cultures in

bioreactors (32). The CHO host also exhibits high reproducibility across multiple batches of cultures providing an improved safety profile compared to other types of medium containing animal/human derived proteins (35). Another advantage for the CHO cell line is that while using mammalian cells for protein production can result in a low specific productivity, CHO cells can overcome this disadvantage due to gene amplification and advances in the host cell line and bioprocesses have resulted in product yields being increased from 10's of mg/L 30 or more years ago to > 5g/L.

The selection and gene amplification systems that are most widely used are the glutamine synthetase (GS) – mediated gene amplification and dihydrofolate reductase (DHFR)-mediated are available (32). These metabolic selection systems allow amplification of high gene copy numbers and as a result, high specific protein productivity (36). CHO cells have also demonstrated reduced susceptibility to certain viral infections compared with other mammalian cell lines (37). Out of all the advantages of CHO cells, the one advantage that could be considered the most important is their ability to produce proteins with complex bioactive protein post-translational modifications that are similar to the human produced equivalent (35) enabling therapeutic protein production of molecules such as monoclonal antibodies (mAbs).

Originally, CHO cells were isolated as spontaneously immortalised cells from primary Chinese hamster ovarian cultures. The CHO cell line was later derived into the CHO-K1 and CHO-S cell lines with the CHO-S cell line being adapted to grow in suspension culture (38). CHO cells having the ability to grow in suspension and serum free chemically defined media, enables scalability into large volume bioreactors (up to 20,000 L), with the expression system being highly tolerant to changes in variables such as pH, temperature, pressure, and oxygen levels during the development of recombinant cell lines and subsequent production manufacturing (39).

To select for the successful transfected CHO cells, either of the two metabolic selection systems are routinely used (antibiotic selection systems are generally avoided) with the DHFR system used in DHFR deficient cells whereby the DHFR gene is included on the same plasmid as the gene of interest or methionine sulfoximine (MSX) inhibition of the glutamine

synthetase (GS) gene in CHO cells whereby the gene for GS is included on the same plasmid as the recombinant gene(s) of interest (40). Within the MSX system, the GS gene is used as a dominant selection marker. GS is vital for survival of CHO cells in culture and without glutamine present in the growth medium, only the successful CHO transfectants survive (41). By using the GS system, this also allows for the reduction of by-products, as with the GS gene being reintroduced, ammonia and glutamate is incorporated to form glutamine which is then available for α-ketoglutarate, an example of a TCA cycle intermediate (42). More recently GS knockout CHO cell hosts have been developed whereby there is no endogenous GS expression at all and cells must be provided with L-glutamine in the medium or the gene for GS introduced for cells to survive in the absence of L-glutamine.

1.3 Aims of this project

With the SARS-CoV-2 pandemic gripping the world, the need for vaccines, treatments and diagnostic tools was (and remains) at an all-time high at the start of this project. This project set out to develop vectors and constructs for the production of the SARS-CoV-2 spike protein and variants thereof as they arose and subsequently to express these in CHO cells. The spike protein was the key target due to it being the most widely studied and utilised of the SARS-CoV-2 virus for vaccines and diagnostics, and thus was and remains in high demand in terms of recombinant protein material. Indeed, the demand for spike protein to be produced at higher yields with good quality protein has increased for novel vaccine and diagnostic tool development. The project aimed and described approaches to deliver platforms to generate the spike protein and further enhance the production process via the comparison of signal peptides, CHO cell lines and development of stable cell lines, as well as developing new variants as they emerged.

The overall aims of the project were therefore to:

- Generate various spike gene constructs in commercial vectors suitable for transient and stable expression in mammalian cells, particularly CHO cells
- Generate and clone new variants of the SARS-CoV-2 spike as they emerge including B.1.1.7 and B.1.351
- Perform site directed mutagenesis on spike constructs as new point mutations emerged within the population (D614G and E484K).
- Transiently transfect various spike constructs into cell lines and asses expression
- Compare protein various ER signal peptides on spike constructs and the impact on secreted spike production
- Develop recombinant spike producing CHO stable cell pools for multiple spike protein constructs
- Establish growth curves, western blots and qRT-PCR on stable cell pools to compare effectiveness between constructs
- Measure antibody response to the different spike protein variants between hospital patient SARS-CoV-2 serum samples with either Wuhan or B.1.1.7 infected patients.

Chapter 2

Materials and Methods

2.1 Molecular Biology

2.1.1 Enzymatic restriction digests of plasmid DNA

For each plasmid DNA construct sample, 0.6 μl of plasmid DNA, 1 μl of restriction enzyme, 2 μl of FastDigest™ buffer and 16.4 μl of nuclease free water was added and placed in a water bath to incubate for 1 hour at 37°C. The enzymes used for each plasmid are described in Table 2.1.1.

2.1.2 Plasmid constructs and primers

The plasmid constructs and primers generated and used within the project are outlined in Table 2.1.1 and Table 2.1.2 below.

2.1.3 Agarose gel electrophoresis analysis of DNA and gel extraction

Agarose gels were prepared by dissolving 0.5 g of agarose in 50 ml of 1x TAE buffer (buffer consisting of Tris base, acetic acid and EDTA). 2.5 μl of ethidium bromide was then added, and the mixture poured into a gel mould with a 15 well comb and allowed to set. To run the gel, it was removed from the mould and placed into a tank filled with TAE and run at 120 V for 40 minutes. The gel was loaded with 10 μl of 1 kb DNA ladder (Promega) and 10 μl of the plasmid DNA digest samples. For ligation samples, amplified PCR samples were loaded at 30 μl per well and 15 μl of 1 kb DNA ladder (Promega). Results of the gel were viewed via a standard UV imager. Bands of interest were excised using a sterile razor blade and purified via the Wizard SV Gel and PCR Clean Up System (Promega), following the manufacturers protocol.

Table 2.1.1: *List of DNA constructs that have been generated and/or used throughout the project showing the construct name, gene of interest, size of the plasmid (bp), the restriction sites for insertion of the gene of interest and the backbone*

Construct	Gene of Interest	Size (bp)	Restriction Site	Backbone
Nat-FL-His	Full Length Spike	9186	HindIII/XhoI	pcDNA3.1hygro
Nat-FL-His	Full Length Spike	9298	HindIII/XhoI	pcDNA3.1mGS
Nat-FL-His (D614G)	Full Length Spike	9186	HindIII/XhoI	pcDNA3.1hygro
Nat-S1-His	S1	7605	HindIII/XhoI	pcDNA3.1hygro
Nucleocapsid- CHO	Nucleocapsid	6876	HindIII/XhoI	Industrial Vector
Nat-FL-His - SpyTag	Full Length Spike	9255	HindIII/XhoI	pcDNA3.1hygro
Nat-S1-His-SpyTag	Full Length Spike	7674	HindIII/XhoI	pcDNA3.1hygro
H7-FL-His	Full Length Spike	9198	HindIII/XhoI	pcDNA3.1hygro
H7-FL-His (D614G)	Full Length Spike	9198	HindIII/XhoI	pcDNA3.1hygro
OSP-FL-His-Strep	Full Length Spike	9179	HindIII/XhoI	pcDNA3.1hygro
OSP-FL-His-Strep	Full Length Spike	9322	HindIII/XhoI	pcDNA3.1mGS
OSP-RBD-His-Strep	RBD	6176	HindIII/ApaI	pcDNA3.1hygro
OSP-FL-His-Strep (B.1.1.7)	Full Length Spike	9213	HindIII/XhoI	pcDNA3.1hygro
OSP-FL-His-Strep (B.1.1.7)	Full Length Spike	9325	HindIII/XhoI	pcDNA3.1mGS
OSP-FL-His-Strep (B.1.1.7) (D614G)	Full Length Spike	9213	HindIII/XhoI	pcDNA3.1hygro
OSP-FL-His-Strep (B.1.1.7) (D614G + E484K)	Full Length Spike	9213	HindIII/XhoI	pcDNA3.1hygro
OSP-FL-His-Strep (B.1.351)	Full Length Spike	9213	HindIII/XhoI	pcDNA3.1hygro
OSP-FL-His-Strep (B.1.351)	Full Length Spike	9325	HindIII/XhoI	pcDNA3.1mGS

Table 2.1.2: *List of primers used throughout the project. Columns include the name of the primer, the forward and reverse sequences, the target gene and what the primer was used for within the project*

Primer Name	Forward Sequence	Reverse Sequence	Target Gene	What it is
T7	TAATACGACTCACTATAGGG		Full Length Spike	PCR Colony Screen
BGHR		TAGAAGGCACAGTCGAGG	Full Length Spike	PCR Colony Screen
SDMD614G	GTGCTGTATCAGGGGCGGTGAACTGTACC	GGTACAGTTCACGCCCTGATACAGCAC	Full Length Spike	Adding in point mutation D614G
SDME484K	AACGGCGTGAAAGGCTTCAACTGCTACTTCCCAC	GTGGGAAGTAGCAGTTGAAGCCTTTCACGCCGTT	Full Length Spike	Adding in point mutation E484K
eGFPfwdXbaI	TATTCTAGAGTGAGCAAGGGCGAGGAG		Full Length Spike	Replacing His tag with eGFP
eGFPfwdXhoI	TATCTCGAGGTGAGCAAGGGCGAGGAG		Full Length Spike	Replacing His tag with eGFP
eGFPrvsApaI		ATAGGGCCCTTACTTGTACAGCTCGTCCATGC	Full Length Spike	Replacing His tag with eGFP
NFLrvsXhoI		ATACTCGAGCTGCTCGTATTTGCCGAGCTC	Full Length Spike	Creating new restriction sites for Nat-FL spike
NFLfwdHindIII	TATAAGCTTAAGCTTATGTTCGTGTTCCTGGTGCTCCTG		Full Length Spike	Creating new restriction sites for Nat-FL spike
qRT Spike FL	CTGAACGAGGTGGCCAAGAA	TCAGTGATGATGGTGATGGTGG	Full Length Spike	qRTPCR for FL spike constructs
seqSpike 1	CCTTCCTGCTGAAGTACAAC		Full Length Spike	Sequening midsection of full lengt CHO optomised spike
seqSpike 2	CTTCAACTTCAACGGCCTGAC		Full Length Spike	Sequening midsection of full lengt CHO optomised spike
seqSpike 3	GAGGTGTTCGCCCAAGTGAAG		Full Length Spike	Sequening midsection of full lengt CHO optomised spike
mGSSmalfwd	TATCCCGGGATGGCCACCTCAGCAAG		mouseGS	Addition of restriction sites for mGS
mGSPfolrvs	ATATCCCGGATTAGTTCTTGTATTGGAAGGGTTCG		mouseGS	Addition of restriction sites for mGS

2.1.4 Ligation of targets into the commercial backbone plasmid DNA

The commercial vector used in this project was the pcDNA3.1hygro (ThermoFisher) vector which contains a Hygromycin resistance gene. For each ligation sample, 1 µl of pcDNA3.1hygro, 3 µl of the specific gene insert, 1 µl of 10x ligase buffer, 1 µl of DNA Ligase (Promega T4 DNA ligase M1801) and 5 µl of nuclease free water were added and left for 1 hour to incubate at room temperature.

2.1.5 Transformation of competent DH5alpha *E. coli* cells with plasmid DNA

Competent DH5α *E. coli* cells were supplied by the laboratory. 50 µl of competent *E. coli* cells were added to 10 µl of the ligation samples and left on ice for 20 minutes. The samples were then placed in a water bath for heat shock at 42°C for 1 minute 30 seconds. 450 µl of LB broth (medium consisting of Tryptone, Yeast extract and Sodium Chloride) was then added to each sample and incubated for an hour at 37°C. 100 µl of the samples was then streaked onto LB agar plates (containing a working stock of 100 µg/ml ampicillin) and incubated overnight.

2.1.6 Positive colony PCR screen

When screening for positive colonies following transformation, a PCR colony screen was set up. Individual colonies from the transformation were picked and resuspended into 5 µl of nuclease free water. To each of the colony resuspensions, 15µl of a PCR mastermix was added that included the relevant target forward and reverse primers and the PCR performed. The volumes and reagents added to the PCR mastermix and the GoTaq Thermocycler Programme is outlined below in Tables 2.1.3 and 2.1.4, respectively.

Table 2.1.3: *The reagents and volumes added to the PCR mastermix, per reaction*

Reagent	Volume per reaction (µl)
Nuclease Free Water	7.3
GoTaq 5x Buffer	4
$MgCl_2$	1.6
10mM dNTPs	0.4
Forward Primer (1:10 dilution)	0.8
Reverse Primer (1:10 dilution)	0.8
GoTaq DNA Polymerase (Promega G2 Flexi DNA Polymerase M7801)	0.1

Table 2.1.4: *The steps for the GoTaq Thermocycler Programme for Colony PCR Screening*

Step	Step Description	Temperature °C	Time	
1	Initial Denaturation	94	5 minutes	
2	Denaturation	94	30 seconds	
3	Annealing	50	30 seconds	30 Cycles
4	Extension/Elongation	72	1 min/kbp	
5	Final Extension	72	7 minutes	
6	Storage	4	Hold	

2.1.7 Phusion Cloning

To amplify the number of copies of genes for cloning, Phusion (ThermoFisher Phusion High Fidelity DNA Polymerase 2U/µl F530S) enzyme cloning was used. The following mastermix was prepared and added per reaction as shown in Table 2.1.5 and added to 1 µl of template DNA. Table 2.1.4 outlines the steps of the PCR programme used to amplify the copies of DNA.

Table 2.1.5: *The reagents and volumes used to make up the Phusion PCR mastermix*

Reagent	Volume (μl)
dH$_2$O	33.5
5x HF	10
dNTPs	1
Forward Primer	2
Reverse Primer	2
Phusion Enzyme	0.5

Table 2.1.6: *The PCR Programme steps for Phusion Cloning*

Step	Step Description	Temperature ^{o}C	Time
1	Initial Denaturation	98	30 seconds
2	Denaturation	98	10 seconds
3	Annealing	Primer Temp	30 seconds
4	Extension/Elongation	72	20 secs /kbp
5	Final Extension	72	10 minutes
6	Storage	20	Hold

2.1.8 Overnight culture of *E.coli* colonies

Transformed single colonies were inoculated in overnight cultures at an incubation temperature of 37°C in 5 ml of LB broth complemented with 1 μg/μl ampicillin (Amp).

2.1.9 Isolation of plasmid DNA using commercial mini or maxiprep kits

To isolate plasmid DNA from *E. coli* cells, commercial kits were used. For minipreps the QIAgen Spin Miniprep Kit 250 with a 5 ml culture was used and for Maxipreps, the Invitrogen Thermofisher PureLink HiPure Maxi Kit with a 250 ml culture volume was used following the manufacturers protocols. The purity and concentration of the preps were measured using a Nanodrop 1000 spectrophotometer.

2.1.9.1 Site Directed Mutagenesis of Plasmid DNA

To achieve point mutations within plasmid DNA, site directed mutagenesis was carried out using the commercially available Quikchange Lightning Site Directed Mutagenesis kit (Agilent Technologies) following manufacturer protocols. Primers were designed and ordered from Integrated DNA Technologies. Table 2.1.7 describes which point mutations were carried out in which plasmid during this project.

Table 2.1.7: List of plasmid constructs and the point mutations carried out in each

Plasmid DNA	Point Mutation
Native-Full Length Spike – His	D614G
OSP-FL-Strep-His(pcDNA3.1hygro) (B.1.1.7)	D614G
H7-Full Length-His	D614G
OSP-FL-Strep-His (pcDNA3.1hygro) B.1.1.7	E484K

2.2 Protein Analysis and Characterisation

2.2.1 Transient transfection of CHO cells with plasmid DNA

Transient transfections were carried on CHO-K1 and CHO-S cells as described below.

2.2.2 Electroporation of CHO cells

For electroporation, a Bio-Rad gene pulse electroporator was set up with the following settings; exponential protocol, 300 V, 900 vF, ∞ resistance and 4 mm cuvette. For the CHO-K1 and CHO-S transfections, 4 transfections were carried out per host cell line with 20 μg of DNA being transfected per 20 ml of cells. 700 μl of cell stock (concentrations as in Table 2.1.8) was added to 100 μl of plasmid DNA stock and electroporated before being added to 30 ml of CD-CHO medium. Table 2.1.8 outlines the volumes of cells and cell medium per transfection and the calculations for the DNA stock.

Table 2.1.8: *The volumes of DNA and cell stocks for transient transfections of CHO cells*

Cell Line	Sample	Concentration (µg/µl)	DNA stock (µ)	TE volume (µl)	Viable cells/ml	Cells needed per transfection (ml)	Total CD-CHO medium (ml)
CHO-S	Nat-FL-His	1159	51.9	248.1	3.91×10^6	2.6	30
	Nat-S1-His	467.8	128.4	171.6			
	NC-CHO	1376.5	43.5	256.5			
	OSP-FL	648.2	92.7	207.3			
CHO-K1	Nat-FL-His	1159	51.9	248.1	7.49×10^6	1.3	30
	Nat-S1-His	467.8	128.4	171.6			
	NC-CHO	1376.5	43.5	256.5			
	OSP-FL	648.2	92.7	207.3			

2.2.3 Cell harvesting

Samples for counting of cells using a ViCell instrument were taken on days 4, 5 and 6 of culture when harvesting CHO-S and CHO-K1 cells. To harvest cells, 1 ml samples were taken and spun down in a mini centrifuge for 5 minutes at 1000 rpm. The supernatant was then moved to a fresh centrifuge tube and placed in dry ice. Pellets was resuspended in 500 µl of PBS and recentrifuged at 1000 rpm for 5 minutes. The PBS was then removed from the pellet and the sample placed on dry ice. On day 6, an extra sample was taken of 1×10^7 viable cells. Samples were then stored at -80°C until required for further analysis.

2.2.4 Routine cell passaging

To passage CHO-S and CHO-K1 cells, cells were seeded at 0.2×10^6 viable cells per ml in a 20 ml culture. To measure viable cells, a ViCell analyser was used to determine the number of viable cells/ml. Cells were cultured at 37°C in a 5% CO_2 atmosphere in 125 mL shake flasks.

2.2.5 Determination of protein concentrations in samples by Bradford assay

Protein concentrations from cell pellet and culture supernatant samples were prepped and estimated using a Bradford Assay. 300 μl of lysis buffer (1 protease tablet, 500 μl NaF, 50 μl NaV) was added to pellet samples and placed on ice for 30 minutes. These samples were then centrifuged for 3 minutes at 13000 rpm at 4°C degrees and the lysate removed. Laemmli sample buffer was added to each sample. The Bradford dilutions were then made up with a 1 in 20 dilution. In a 96 well plate, each sample was run in triplicate with 20 μl per well of sample. 200 μl of Bradford reagent was then added to each well and measured on a spectrophotometer measuring the absorbance at 595 nm. A calibration curve with known concentrations of BSA was used to estimate protein concentrations of samples.

2.2.6 Protein analysis by SDS-PAGE

For SDS-PAGE analysis, 10 μl protein samples were prepared in a 5 x Laemmli sample reducing buffer (1 M Tris-HCl pH 6.8, glycerol, SDS, β-mercaptoethanol, Bromophenol Blue, H_2O). The samples were then boiled at 95°C for 10 minutes. The protein samples were loaded into a 12% polyacrylamide gel, components outlined in Table 2.1.9, and run in a 1x SDS running buffer (10x SDS running buffer made up with glycine, SDS, Tris made up to 1 L with ddH_2O at pH 8.8. 100 ml of 10x added to 900 ml of ddH_2O for 1x) at 100 V for 30 minutes and then 150 V for 1.5 hours or until the dye front had reached the bottom of the gel.

Table 2.1.9: *The reagents and volumes required to make a 12% SDS-PAGE resolving gel*

Reagent	Volume
ddH20	2.05 ml
30% Polyacrylamide	0.5 ml
1.5 M Tris pH 8.8	1.6 ml
10% SDS	62.5 μl
10% Ammonium Persulfate	62.5 μl
TEMED	2.5 μl

2.2.7 Transfer of proteins from polyacrylamide to nitrocellulose for western blot analysis

Following SDS-PAGE, the proteins separated were transferred to a nitrocellulose membrane in an electrophoresis gel tank at 0.75 A for 1 hour at 4°C. The gel and membrane were submerged in a 1 x Transfer Buffer prepared from a 10 x stock (144 g Glycine, 30 g Tris, 1 g SDS made up to 1 L with ddH$_2$O water).

2.2.8 Western blot analysis

The primary antibodies used for Spike full length protein were anti-His mouse (Sigma H1029-0.2ML) as the Spike protein contained a C-terminal His-tag. The secondary antibody was anti-mouse (Sigma A4416-1ML). Separated proteins on nitrocellulose were first probed with a primary antibody made up in 5-10 ml 5% milk powder (1:1000 dilution in 3% BSA-TBST). Blots were then washed with 0.1% TBST and shaken overnight at 100 rpm. The secondary antibody was then added (1:1000 dilution in 3% milk-TBST). The loading control used was a β-actin antibody.

2.2.9 Development of western blots

Western blots were incubated in a tray of 1 ml ECL 1 reagent and 1 ml of ECL 2 reagent (ThermoFisher Pierce™ ECL Western Blotting Substrate 32106) for 5 minutes before being exposed and developed on film using an Optimax 2010 film processor.

2.3 Development of Stable CHO Cell Pools

2.3.1 Plasmid linearization

To linearise plasmid DNA in preparation for stable cell line development for CHO-S cells, DNA digests were carried out using the single cut restriction digest enzyme *Fsp1* with 50 μg of plasmid in 100 μl volumes. Minipreps were then made of these linearised plasmids to 'clean' the digests up.

2.3.2 DNA solubilisation

To each linearised plasmid miniprep sample, 10 μl of sodium acetate (3M Sigma S7899) was added as well as 250 μl of ice cold 100% ethanol. The DNA was then pelleted in a chilled

centrifugation at 13000 rpm for 30 minutes at 4°C. The supernatant was removed and 500 μl of ice-cold filter sterilised 70% ethanol was added to the DNA pellets. This was pelleted using a chilled centrifugation at 13000 rpm for 10 minutes at 4°C. The supernatant was then removed, and the DNA pellet allowed to air dry for 1 hour. DNA pellets were then resuspended in 50 μl of filter sterilised TE buffer and vortexed thoroughly. This was left to incubate for 1 hour at room temperature and the concentrations of the DNA were measured on a Nanodrop 1000 Spectrophotometer.

2.3.3 Stable cell line transfection

After linearisation, DNA constructs were transfected via electroporation using a Biorad Genepulser Xcell electroporator. For each of the constructs, 20 μg of the linearised plasmid and $1x10^7$ viable CHO-S cells were electroporated at 300 V and 900 mF with a cuvette diameter of 0.4 mm. Cells were then transferred into a T75 flask with 10 ml of the appropriate medium (CD-CHO 500 ml made up with 250 μl of MSX and 500 μl of anti-clump). T75 flasks were incubated at 37°C in a static incubator with a CO_2 environment of 5% and the cells then left over a 2-week period. After 2 weeks, those cells that had survived the MSX selection and were growing were used to seed shake flasks with either 25 μM MSX present in the medium or 50 μM MSX in the medium. For each shake flask, $0.2x10^6$ cells/ml were seeded in 20 ml of culture.

2.3.4 Batch culture growth curve generation

To determine the growth profile of stable cell pool cultures seeded at $0.2 x 10^6$ cells/ml across a batch culture, ViCell readings were taken each day at the same time measuring the viable cells ($x10^6$) and culture viability (%) of each sample. To take the readings, 500 μl of each culture was taken and added to a cuvette with 500 μl of PBS added to make up the sample to 1 ml.

2.4 RNA Extraction and Purification

2.4.1 Extraction

RNA samples were generated from transfected cell cultures from 1×10^6 viable cell pellets collected at the appropriate time from culture (Day 3).

2.4.2 RNA purification and qRT-PCR analysis

To obtain Day 3 cell lysate RNA samples, the RNA was isolated using the Qiagen RNeasy Kit following the manufacturers protocol. The samples were then quantified using a nanodrop spectrophotometer on the RNA-40 setting. The RNA samples then were subjected to a gDNA treatment in order to remove any unwanted genomic or plasmid DNA. The following protocol was then used to undertake qRT-PCR, as shown in Table 2.2.1. 1 μg dilutions of the RNA were prepared using the nanodrop concentrations. The samples were then plated into a qRT-PCR 96 well plate (with 18 μl of mastermix and 2 μl of the various spike samples added to each well). The mastermix components are outlined in Table 2.2.2 for both ß-actin and Spike samples with respective primers below.

Table 2.2.1: *DNA treatment protocol reagents for removing genomic or plasmid DNA from RNA samples*

DNA Treatment Protocol	Volume (μl)
+ RNA + Water	17
+ DNAse (Promega M6101)	2
+ DNase	1
30 minutes at 37ºC	-
STOP solution	2
10 minutes at 65ºC	-
H_2O	18
-80ºC Storage	-

Table 2.2.2: *Reagents and volumes required for mastermix of RNA samples for both β-actin and spike sample RT-PCR relative mRNA analysis.*

Reagent	Volume (μl)	x No of wells	Total volume (μl)
H_2O	3.8	25	95
Quantifast (Qiagen SYBR Green 400 204154)	12	25	300
Primer	2	25	50
RT	0.2	25	5

2.4.3 Quantitative Real Time PCR (qRT-PCR)

qRTPCR was carried using the iScript ™ One-Step RT-PCR Kit with SYBR Green(BioRAD) and the previously loaded 96 well plate. Primers directed against the heavy and light chains of the Spike mRNA were used (as shown in Primers Table 2.1.2) enabling the detection of the levels of the mRNA in each sample. ß-actin primers were also used as a housekeeping gene, kindly supplied by Joanne Roobol (University of Kent, UK). To programme the Realplex PCR machine the following conditions were set up, as outlined in Table 2.2.3. Once the PCR programme was complete, the relative quantity raw data was collected from each well and the average relative quantity with standard deviation error bars were calculated using β-actin as a normaliser.

Table 2.2.3: qRT-PCR program steps

Step	Temperature (°C)	Time
1.	50	10 minutes
2.	95	5 minutes
3.	95	10 seconds
4.	60	10 seconds
5.	60	20 seconds
6.	Melting Curve	-
7.	20	HOLD

2.5 Enzyme Linked Immunosorbent Assay

2.5.1 ELISA protocol

To perform ELISA, the different spike variant, the original Wuhan variant (here referred to as OSP-FL), B.1.1.7 and B.1.351 spike sample dilutions of 2 µg/mL were made up with coating buffer (0.8 g Na_2Co_3, 1.47 g $NaHCO_3$ made up to 500 ml with dH_2O pH 9.7 (1 M HCl)) and 100 µl of each loaded into the designated wells of a 96 well plate (Greiner Bio-one microplate (Part 655101)) to coat the wells. The dilutions used are outlined in Table 2.2.4.

Table 2.2.4 *Table displaying dilution measurements for spike samples for ELISA*

Sample	Spike Volume (µl) to give 2 µg/ml	Coating Buffer (µl)
OSP-FL (Wuhan)	21.6	3978.4
B.1.1.7	20.9	3979.1
B.1.351	20.6	3987.4

The wells were coated in the layout displayed in Figure 2.1.1. The plate was then covered with foil and left overnight at 4°C. 0.25 g of Caesin was also added to a blocking buffer (50 ml, 1.59 g Na_2CO_3, 2.93 g $NaHCO_3$) and left overnight at 37°C. The wells were then washed 3 x with 100 µl of wash buffer (5.84 g NaCl, 1.15 g Na_2HPO_4, 0.29 g $NaH_2PO_4.2H_2O$, 3.72 g EDTA, 200 µl Tween-20 and 10 ml n-butanol made up to 1 L with ddH_2O) and 200 µl of blocking buffer added to each well. The plate was then left to incubate for 3 hours on an orbital shake. Two virus heat inactivated human serum samples (Liverpool LV094 29.04.20 (Wuhan sample) and Sondus 5730 (B.1.1.7)) dilutions were made up using sample buffer as displayed in Table 2.2.5.
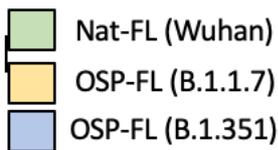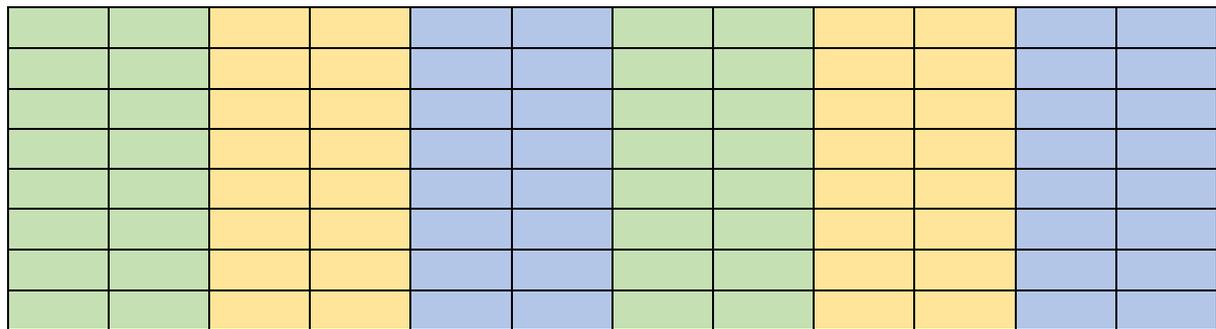
Figure 2.1.1: *Layout of 96 microwell plate with spike protein variants added to designated wells.*

Table 2.2.5 *The dilutions prepared in µl for the virus inactivated human serum samples*

Dilution	Serum Sample (µl)	Sample Buffer (µl)
1:20	32.5 (Serum)	617.5
1:50	23 (Serum)	1127
1:100	500 (1:50 sample)	500
1:250	328 (1:100 sample)	492
1:1000	162.5 (1:250 sample)	487.5

Anti-Spike antibody dilutions were made up using Sino Biologics 40150-R007 (Sars-CoV-2 (2019 nCoV) Spike antibody, Rabbit mAb using sample buffer (12.1 g Tris base, 5.84 g NaCl, 200 µl Tween-20, pH 7.0) as part of the positive control with the dilutions outlined in Table 2.2.6.

Table 2.2.6: *The dilution measurements for the Anti-Spike Ab samples made up with sample buffer.*

Dilution	Anti-Spike Ab Stock	Sample Buffer (ml)
1:1000	2ul	2
1:2500	0.8ml (1:1000 sample)	1.2
1:5000	1ml (1:2500 sample)	1
1:10000	0.5ml (1:500 sample)	0.5

After incubation, the 96 well plate was blotted and 100 µl of the serum dilutions and anti-Spike Ab was added to the designated wells. The plate was left overnight at 4°C then washed 6 x with 100 µl of wash buffer and patted dry on paper towels. 50 µl of secondary antibody (HRP conjugated) was then added to each well. HRP was added to the wells that had been incubated with serum samples and one set of blank wells. Goat anti rabbit antibody – HRP (Sigma A1654-1ML) was added to the wells that had been incubated with anti-SARS-Cov2 spike antibody and the second set of blank wells. The plate was then incubated at room temp for 1 hour with shaking. TMB substrate was prepared using 100 µl aliquots of 0.1 g TMB substrate in 10 mL DMSO and was stored frozen (100 µl TMB substrate, 10 mL substrate buffer and 100 µL H_2O_2 (30%)). The plate was again washed 6 x with 100 µl wash buffer and patted dry on paper towel. 50 µl of TMB substrate was then added to all wells and the plate covered and incubated in the dark for 15 minutes. To stop the reaction, 50 µL of 2.5M H_2SO_4 was added. Absorbance was then read on a plate reader at 450 nm. Absorbances of the mean sample blank was subtracted from the absorbance read for each duplicate serum dilution.

Chapter 3

Results

3.1 Molecular Cloning

3.1.1 Subcloning of the native DNA constructs that encode for the SARS-CoV-2 pike protein

To be able to produce Spike protein for use in novel vaccines and diagnostic tools, DNA constructs were generated to meet the demands of collaborators and for further studies in this project. The Native-FL-His (the coding sequence of the native, full length spike with the trans-membrane domain removed and tagged with a C-terminal His-tag) and Native-S1-His (S1 domain only) were provided in industrial proprietary vectors and therefore the Spike sequences were sub-cloned into a commercial pcDNA3.1hygro vector. To do this, the plasmid DNA provided was digested using the restriction enzymes *HindIII* and *XhoI* (Refer to Section 2.1.1) in order to release the genes of interest and drop out a band of the expected size for each gene sequence. As shown in Figure 3.1.1, agarose gel analysis of the resulting digests should show expected band drop outs for Native-FL-His and Native-S1-His of approximately 3657 bp and 2076 bp respectively, however the products observed were larger than expected when compared to the DNA ladder used. The reason for this was not clear although the 1 kb DNA ladder did not appear to have run correctly compared to other gels. The observed bands were thus assumed to be correct due to the fact there was just one band drop out as expected, and these would be subsequently verified in further digests and sequencing, and so were extracted and used in subsequent ligation experiments into the new pcDNA3.1hygro backbone.
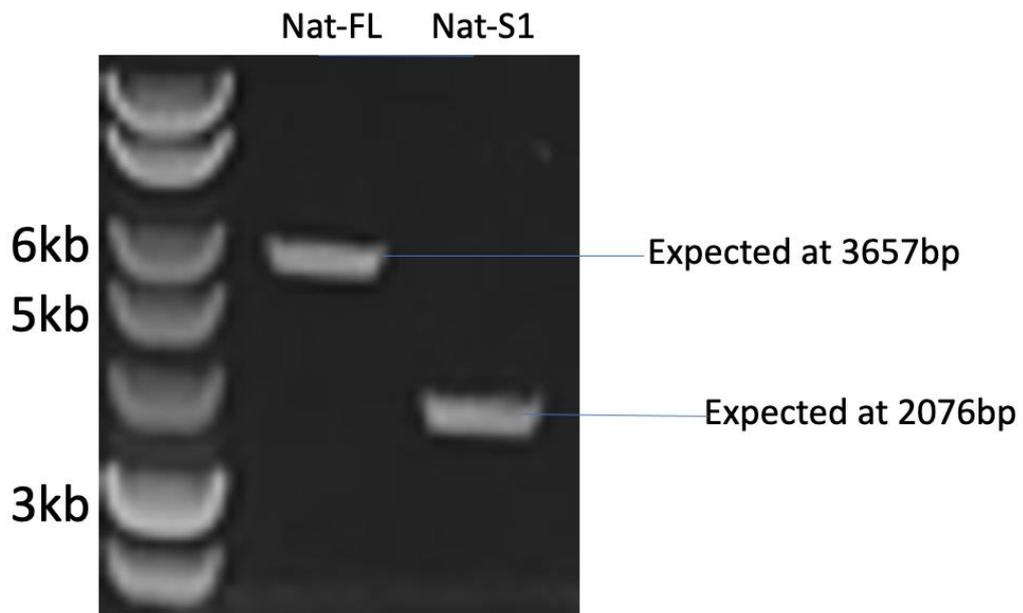
Figure 3.1.1: *Agarose gel analysis showing the result of the restriction enzyme digest of Native-FL-His and Native-S1-His from the original donor plasmid DNA using restriction enzymes HindIII and XhoI to drop out the gene of interest from the donor vector.*

The DNA bands released in the digest shown in Figure 3.1.1 were then extracted and ligated into a HindIII/XhoI pre-cut pcDNA3.1hygro vector. After ligation, the Native-FL-His and Native-S1-His were transformed into DH5α *E. coli* and left overnight. Figure 3.1.2 shows a PCR colony screen on 4 subsequent colonies from each ligation indicating that all four colonies were positive by PCR screening and were therefore taken forward to be grown up and minipreps of the plasmid DNA prepared. The band sizes were also of the expected size compared to the DNA marker ladder and therefore further confirmed that the digest products shown in Figure 3.1.1 were indeed of the correct size.
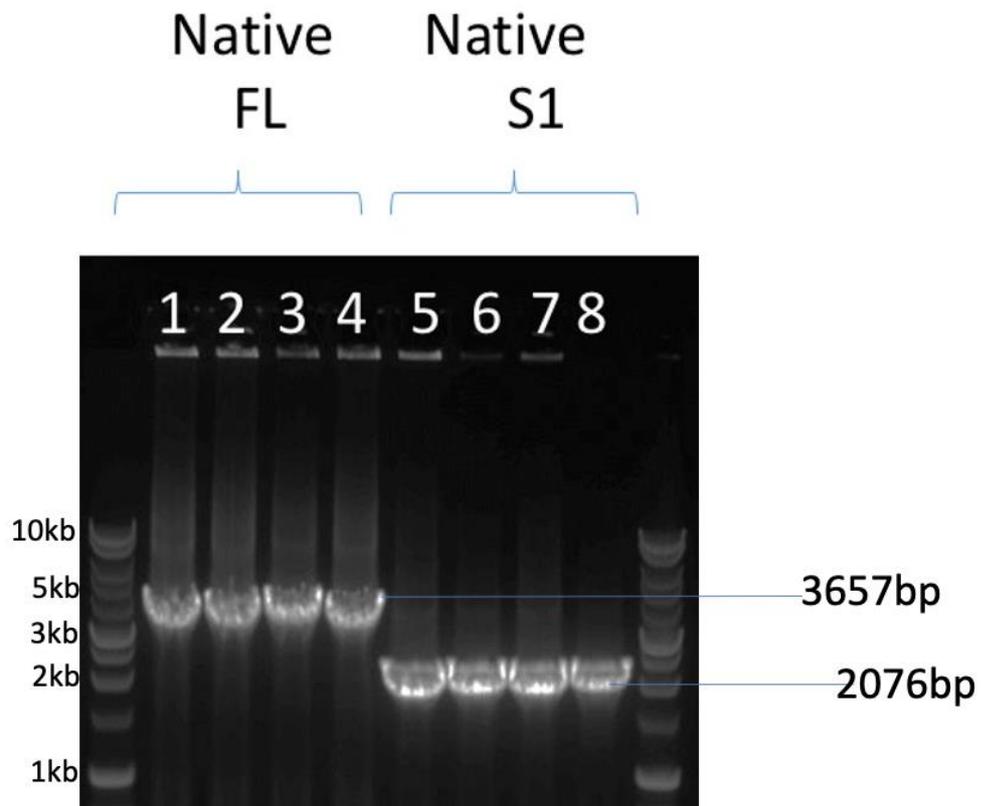
Figure 3.1.2: *PCR colony screen of colonies after ligation of native-FL-His or Native-S1-His band dropouts into the pcDNA3.1Hygro vector showing the positive clones from transformation of DH5alpha E. coli. Agarose gel showing positive clones for 1-4 (Native-FL-His) and 5-8 (Native-S1-His) with band sizes for the full-length products indicated.*

After the positive clones identified in Figure 3.1.2 were grown, minipreps were prepared of the resulting Native-FL-His and Native-S1-His pcDNA3.1hygro vectors (See Appendix 1 and 3) and an analytical restriction enzyme digest and was undertaken with HindIII/XhoI to check the expected band drop outs were present. The results of this digest were analysed by agarose gel electrophoresis and are shown in Figure 3.1.3 below.
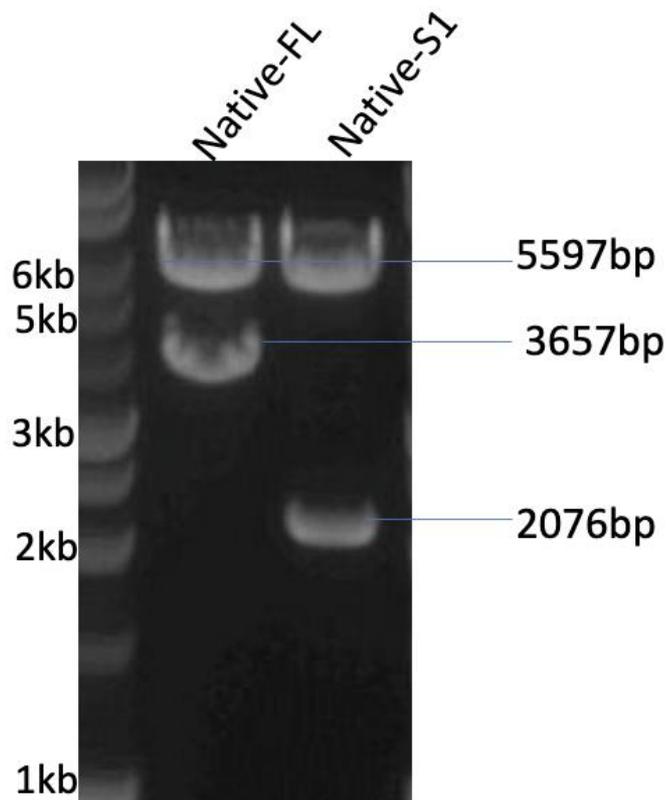
Figure 3.1.3: *Analytical agarose gel analysis of a restriction enzyme digest with HindIII/XhoI confirming the ligation had worked correctly for the Native-FL -His and Native-S1-His genes of interest into pcDNA3.1hygro. A backbone band of approximately 5597 bp was observed for both digests whilst for each target a band drop out of the expected size was observed as highlighted.*

The DNA agarose analysis showed the presence of the expected band drop outs when using HindIII/XhoI were present for the pcDNA3.1hygro backbone at around 5597 bp, as well as the gene inserts, 3657 bp for Native-FL-His and 2076 bp for Native-S1-His. The concentration and quality of the samples were also measured using a nanodrop spectrophotometer and the results are outlined in Table 3.1.1.

Table 3.1.1: *The Native-FL-His(3.1hygro) and Native -S1-His (3.1hygro) purified DNA constructs concentration and quality. The 260/280 measurement indicates the purity of the samples and more specifically, between 1.80-2.00 indicates a high purity.*

Sample	ng/µL	260/280
Nat-FL-His (3.1hygro)	975.4	1.91
Nat-S1-His (3.1hygro)	857.2	1.91

The data in Table 3.1.1 shows that the Nat-FL-His (pcDNA3.1hygro) and Nat-S1-His (pcDNA3.1hygro) construct DNA preparations had high concentrations of 975.4 ng/μl and 857.2 ng/μl respectively, indicating good quantities of the samples. The A260/280 ratio of 1.91 also indicates the purity of the samples to be within the DNA range (1.8-2.0) and not contaminated with RNA.

3.1.2 Generation of His-Spy tag versions of the full-length Spike and S1 domains

Generation of the Native-FL and Native-S1 was then carried out to include a His-SpyTag instead of just a His tag (See Appendix 5 and 6). SpyTags were previously added to the constructs within the spike sequence to aid in the development of a SpyCatcher system in which the spike protein could be attached to a Spy catcher on a different molecule or structure allowing direct linking of the recombinant protein to another structure such as might be appropriate for a vaccine. The SpyTag/SpyCatcher system is a covalent peptide interaction in which the SpyTag, a short peptide, forms an isopeptide bond when engaging with its other half, the SpyCatcher. This system allows for bioconjugation thus an effective mechanism for vaccine development(43). A Native- Nucleocapsid-CHO construct (whereby the codon usage was optimised for Chinese hamster and the sequence made and provided commercially) was also generated by restriction enzyme digest using *HindIII* and *XhoI* restriction enzymes and ligated into the commercial pcDNA3.1hygro backbone (See Appendix 4). Following ligation into the commercial pcDNA3.1hygro vector and transformation into *DH5α E.coli*, a PCR colony screen was undertaken, and the products analysed by agarose gel electrophoresis to identify positive clones. The resulting agarose gel analysis of the PCR screen is displayed in Figure 3.1.4. The expected band drop outs were present in the agarose gel (Figure 3.1.4) with positive clones for each of the samples (Nat-FL-His-SpyTag, Nat-S1-His-SpyTag, Nucleocapsid CHO) with the expected band drop outs of 3726 bp, 2145 bp and 1347 bp, respectively. These samples were then used to prepare minipreps for each of the constructs. The resulting purified DNA samples were then measured and quantified using a nanodrop spectrophotometer and the results are outlined in Table 3.1.2.
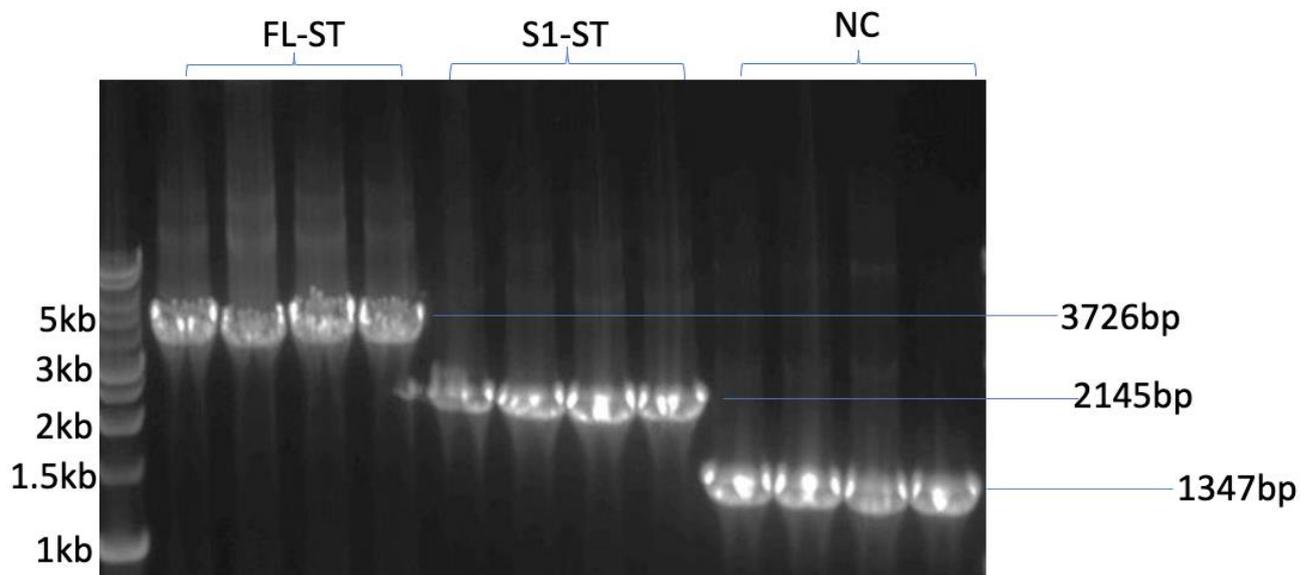
Figure 3.1.4: *Agarose gel analysis of the products of a PCR colony screen highlighting the positive clones of the Nat-FL-His-SpyTag (FL-ST), Nat-S1-His-SpyTag (S1-ST) and Nucleocapsid-CHO (NC) samples with the expected band drop outs at the sizes indicated.*

Table 3.1.2 *The results of using a nanodrop spectrophotometer to quantify the Nat-FL-His-SpyTag, Nat-S1-His-SpyTag and Nucleocapsid-CHO plasmid DNA by measuring the concentration and* purity *of each sample. A 260/280 measurement measures purity and between 1.8-2 indicates it is DNA and not RNA.*

GOI (In pcDNA3.1hygro)	ng/μL	A260/280
Nat-FL-His-SpyTag	574.7	1.88
Nat-S1-His-SpyTag	343.4	1.9
NC-CHO	303.4	1.92

The nanodrop results, as shown in Table 3.1.2, indicate the purified DNA samples were of appropriate quality as the 260/280 results were between 1.80 and 2.0. The concentrations of the samples, however were on the lower end of the scale indicating a low yield with Nat-FL-His-SpyTag having a concentration of 547.7 ng/μl, Nat-S1-His-SpyTag with a concentration of 343.4 ng/μl and Nucleocapsid CHO with a concentration of 303.4 ng/μl. The 260/280 readings of the three samples are all within the 1.8-2.0 bandwidth indicating they are DNA samples. The lower yield could be due to multiple restriction enzyme digests and clean ups resulting in

lower concentrations or that these plasmids were not as well tolerated by the bacteria and thus gave lower yields.

To confirm the ligation between pcDNA3.1hygro vector and the various Spike sequences, an analytical restriction enzyme digest was run using *HindIII/XhoI*. As shown in Figure 3.1.5, the expected band drop outs were present for Nat-FL-His-SpyTag at around 3726 bp, Nat-S1-His-SpyTag around 2145 bp and Nucleocapsid-CHO around 1347 bp. The backbone pcDNA3.1hygro vector expected band drop out at around 5597 bp was also present.
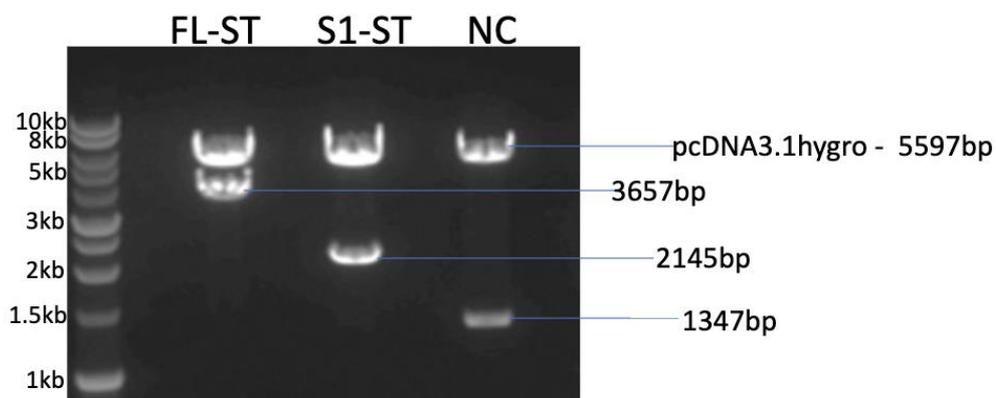
Figure 3.1.5: *Agarose gel analysis of an analytical restriction enzyme digest with HindIII and XhoI showing expected band drop outs for the pcDNA3.1hygro ligation with Nat-FL-His-SpyTag, Nat-S1-His-SpyTag and Nucleocapsid-CHO samples at the expected sizes shown.*

Table 3.1.3 reports the quantification of the purified DNA samples using a nanodrop spectrophotometer. The results, as shown in Table 3.1.3, indicate a high concentration of ng/µl of DNA for Nat-FL-His-SpyTag and Nucleocapsid-CHO at 1159.0 and 1376.5 ng/µl respectively. Nat-S1-His-SpyTag had a lower concentration of 467.8 ng/µl. All three samples were within the 1.8-2.0 260/280 bandwidth thus indicating an appropriate DNA purity.

Table 3.1.3: *The purity (260/280) and concentration (ng/µl) of purified DNA constructs Nat-FL-His-SpyTag, Nat-S1-His-SpyTag and Nucleocapsid-CHO.*

GOI (In pcDNA3.1hygro)	ng/µL	A260/280
Nat-FL-His-SpyTag	1159.0	1.92
Nat-S1-His-SpyTag	467.8	1.88
Nucleocapsid-CHO	1376.5	1.94

3.1.3 Addition of the mouse glutamine synthetase gene into the pcDNA3.1hygro vector as an alternative metabolic selection marker

To make stable recombinant expressing mammalian cell lines a selection marker is usually included in the plasmid DNA to select for cells that have integrated the plasmid into the genome and hence should be expressing the genes of interest in the plasmid. The pcDNA3.1hygro vector has an antibiotic hygromycin resistance marker already in it, however for the production of recombinant proteins that might be used in humans and animals it is desirable to avoid antibiotic resistance markers and hence metabolic selection markers are preferable. A well-known and used metabolic selection marker in mammalian cells, particularly CHO cells, is glutamine synthetase. CHO cells need glutamine synthase (GS) to produce glutamine and survive/grow, and this enzyme can be inhibited by methionine sulfoximine (MSX) and therefore with increasing concentrations of MSX only those cells with sufficient exogenous GS expression can survive. By inserting GS into the plasmids that are then transfected into the cells in the presence of MSX, this enables the cells to be maintained and to select for those that contain exogenous GS and hence the other target genes of interest on the plasmid transfected (in this case the spike). Mouse GS was selected as it is rodent and similar to CHO GS but is different than the endogenous CHO GS so can be differentiated from the endogenous material by PCR.

To add the mouse GS (mGS) into the pcDNA3.1hygro vector (See Appendix 2), Phusion cloning was carried out to amplify copies of mGS using the restriction enzymes *SmaI* and *PfoI* (See section 2.1.7) pcDNA3.1hygro was also digested using *SmaI* and *PfoI*, these digests were then run on a agarose gel and the appropriate bands extracted. The expected band drop outs were around 1100 bp for the mouse GS and 1131 bp for the hygromycin gene of the pcDNA3.1hygro, as shown in Figure 3.1.6. However, the expected band drop outs were not present and the 1 kb DNA ladder had not resolved well making it difficult to distinguish the correct band drop outs. The mGS band drop outs were assumed to be correct around the expected 1100 bp region and were gel extracted and cleaned up for eventual ligation, shown in Figure 3.1.6. The experiment was then also repeated but instead of adding the restriction enzymes *SmaI* and *PfoI* to the pcDNA3.1hygro vector together, the restriction enzymes were added separately, at different temperatures as the manufacturer protocols suggested the restriction enzymes work better at different temperatures. The pcDNA3.1hygro vector was

first digested with *PfoI* at 37°C for an hour and was then cleaned up before re-digesting with *SmaI* at 30°C for an hour. This method was chosen due to the enzymes optimum activity being at different temperatures.
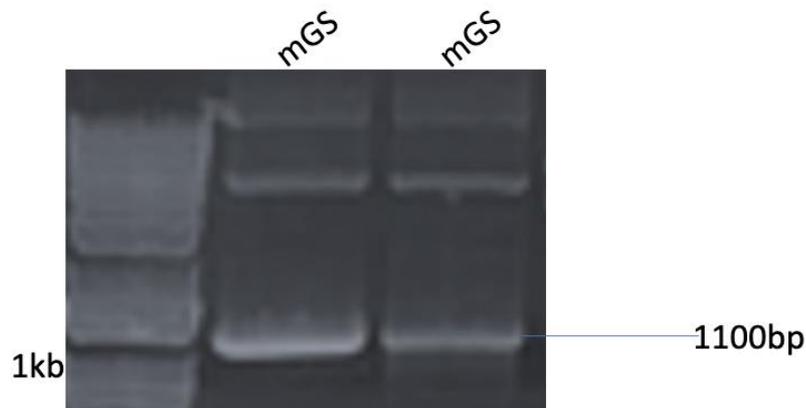
Figure 3.1.6: *Agarose gel analysis showing the band drop outs for the Phusion cloned mGS and at expected band drop out around 1100 bp. The 1 kb DNA ladder was poorly resolved.*

By rerunning the restriction enzyme digest of the pcDNA3.1hygro vector via the new method of two single digests, the agarose gel in Figure 3.1.7 showed the expected band drop outs of 1131 bp for the hygromycin gene for both pcDNA3.1hygro 1 and 2 and 4466 bp for the backbones although very faint bands were visible for the expected band for the hygromycin gene. The faintness of the bands could be explained by the single digests requiring DNA clean up each time resulting in a low yield due to loss during clean up. The bright band at around 5 kb for pcDNA3.1hygro 1 indicates uncut vector which would further explain the faint visibility of the hygromycin gene band due to an inefficient digest. The backbone band of the pcDNA3.1 vector at approximately 4466 bp was gel extracted and ligated with the previous mGS PCR sample.
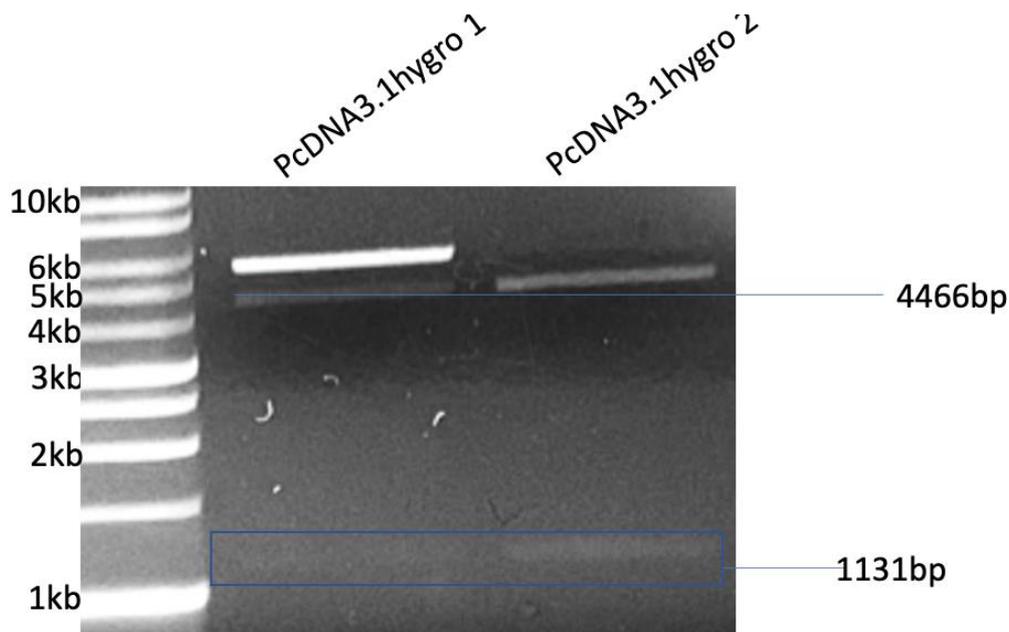
Figure 3.1.7: *Agarose gel outlining results of a Restriction Enzyme Digest of pcDNA3.1hygro using PfoI / SmaI one at a time as a single digest with samples cleaned up in between the two single digests.*
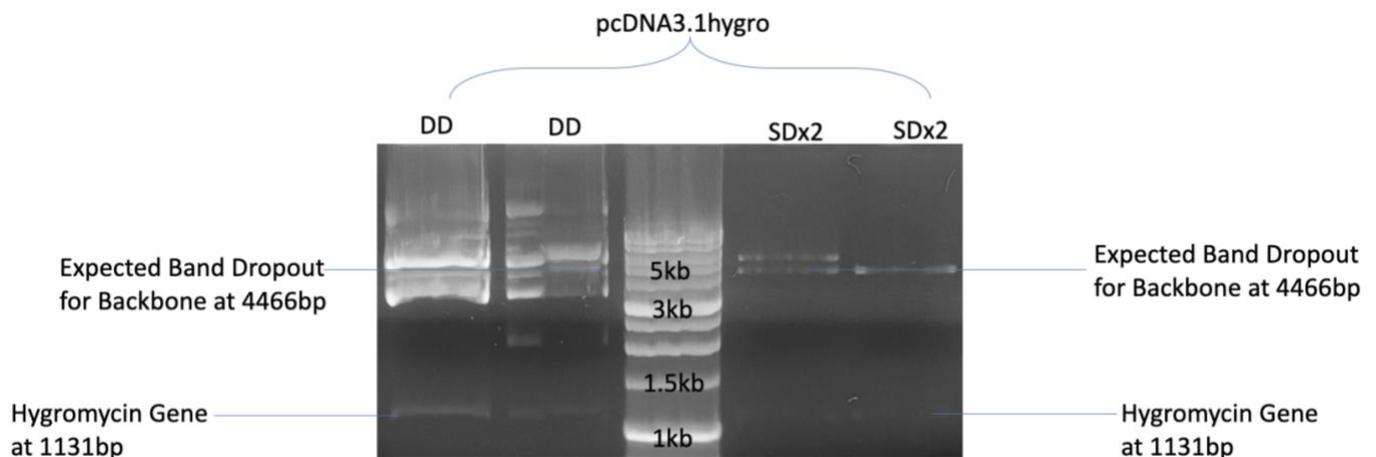
Figure 3.1.8: *Agarose gel analysis comparing the effectiveness of two approaches to digesting the pcDNA3.1hygro vector. Double digest (DD) using PfoI and SmaI at the same time compared to two single digests (SDx2) using PfoI at 37°C and SmaI at 30°C.*

To compare the effectiveness of the simultaneous double digest vs two single digests approach using restriction enzymes *PfoI* and *SmaI*, an agarose gel was run to compare the two approaches. The results of the comparison are outlined in Figure 3.1.8 The image as shown in Figure 3.1.8 indicates that two single digests resulted in a higher loss of yield due to very faint bands of the hygromycin gene at the expected drop out of 1131 bp. The double

digest however had stronger hygromycin band dropout but the backbone dropout at the expected size of 4466 bp was less distinct, indicating uncut vector which could pose an issue. The double digest was run at room temperature and because *PfoI* works more efficiently at 37°C, this could be why the double digest was not as effective. However, running two single digests and cleaning up between them resulted in a high loss of yield so neither approach was completely effective.

The single digest backbone drop-out was extracted and ligated with the previous mGS PCR sample, transformed and then colonies analysed using a PCR colony screen, as shown in Figure 3.1.9, to identify any positive clones. The single digest backbone approach was chosen due to its cleaner band drop out being more precise.
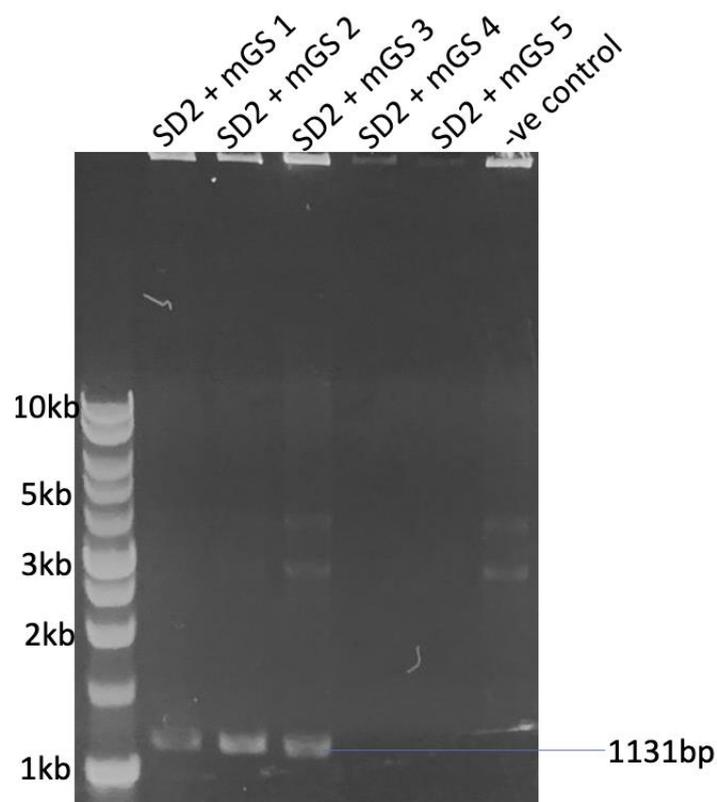
Figure 3.1.9: *Agarose gel analysis of a PCR colony screen indicating the positive clones from the ligation of pcDNA3.1 and mGS PCR product. A negative (-ve) control was also included.*

As shown in Figure 3.1.9 the PCR colony screen indicated that three of the colonies were potentially positive clones due to the expected bands for the mGS at around 1131 bp being present. These clones were then miniprepped and analysed via a restriction enzyme digest to check the ligation was correct and the correct expected band dropout was present. An analytical agarose gel was run to show to the efficiency of the *PfoI* and *SmaI* enzymes. Figure 3.2.1 shows the results of this gel analysis alongside the expected band dropouts shown in a simulated agarose gel, as well as a comparison to *PmlI*, a double cutting enzyme.
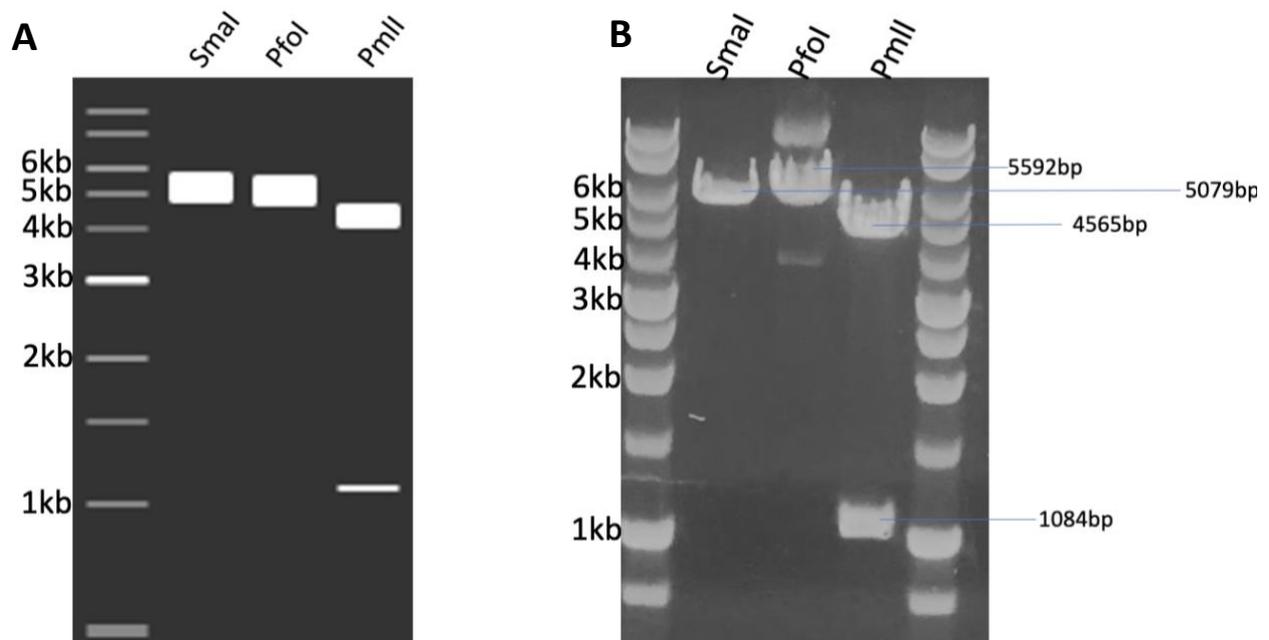
Figure 3.2.1: *Agarose gel analysis showing the simulated expected band dropouts (A) showing the efficiency of the SmaI, PfoI and PmlI enzymes compared to the actual result of the enzyme efficiency gel (B) showing the observed band dropouts when digesting the pcDNA3.1 plasmid.*

As shown in Figure 3.2.1, the expected band drop out for *PfoI* digestion was not correct as where there should be one band drop around 5592 bp due to the nature of being a single cut enzyme, as in Figure 3.2.1 gel (B), there is a second band higher than the expected single band drop out. This indicates an issue with the *PfoI* enzyme and incomplete digestion and could explain why the digest of the pcDNA3.1hygro vector backbone was either incomplete or did not work. There was no such issue with *SmaI* or *PmlI* and the expected band dropouts were present (Figure 3.2.1 (B)) as shown with the simulated gel in Figure 3.2.1 (A) at around 5079 bp for *SmaI* and 4656 bp and 1084 bp for *PmlI*.

To check that the pcDNA3.1 and mGS ligation had worked, test restriction enzyme digests were run to confirm the correct band dropouts were released. *BamHI* was chosen due to the location in the plasmid in which it cuts, directly in mGS but not the backbone, therefore proving the mGS had correctly ligated. The band dropouts were expected to be around 3.8 kb for the pcDNA3.1 backbone and 1.7 kb for the mGS drop out. As shown in Figure 3.2.2, this was not the case. The 1 kb DNA ladder had leaked in the mGS3 sample making it difficult to view the outcome for this sample. For samples mGS1 and mGS2, the 3.8 kb expected band drop out for the pcDNA3.1 backbone was present however the 1.7 kb expected mGS drop out was not, indicating the ligation had not worked.
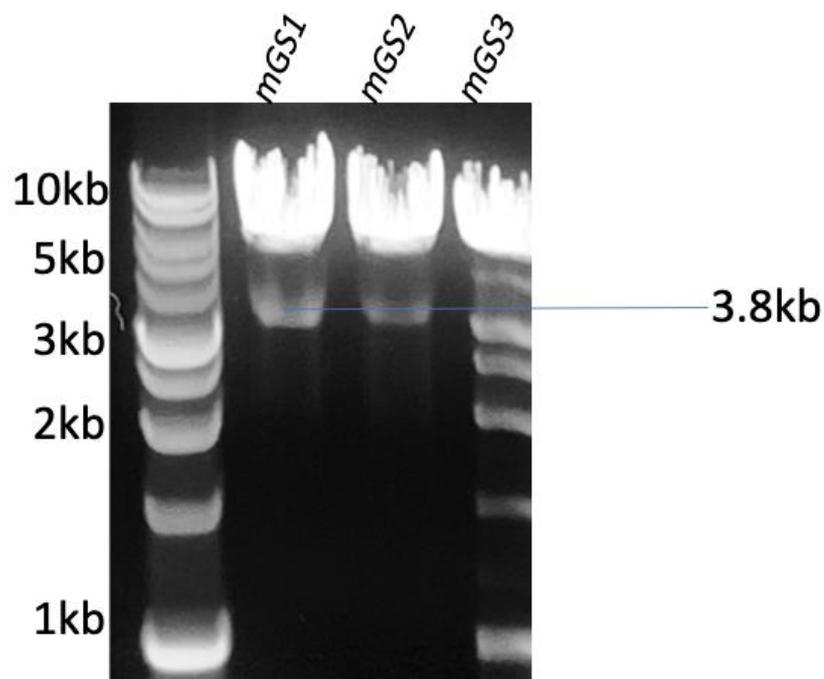
Figure 3.2.2: *Agarose gel analysis showing the result of a test restriction enzyme digest using BamHI to check the ligation of the pcDNA3.1 backbone and the mGS. mGS1, mGS2 and mGS3 are all identical samples of the ligation between pcDNA3.1 and the mGS. Note that the 1 kb DNA ladder had leaked into the mGS3 sample lane*

To further confirm that the ligation had not worked, another test restriction enzyme digest was run with *KpnI* and *PvuI*, these both cut across the mGS or the entire plasmid respectively and would rule out the possibility of the *BamHI* enzyme not working. The results of Figure 3.2.3 indicate again that the ligation had not worked. *KpnI* cuts directly in the mGS and therefore the expected band dropouts should be around 1266 bp for the mGS and 4326 bp for the pcDNA3.1 backbone. As indicated, there was multiple band dropouts for mGS1, mGS2 and mGS3 samples, one of was around the expected 4326 bp dropout. For the *PvuI* digest,

the expected band dropout for the samples should be around 5592 bp but this is not consistent with the results displayed in Figure 3.2.3. Due to the band dropouts for mGS1, mGS2 and mGS3 being the same as the control pcDNA3.1hygro lane, this suggests that the ligation did not work.
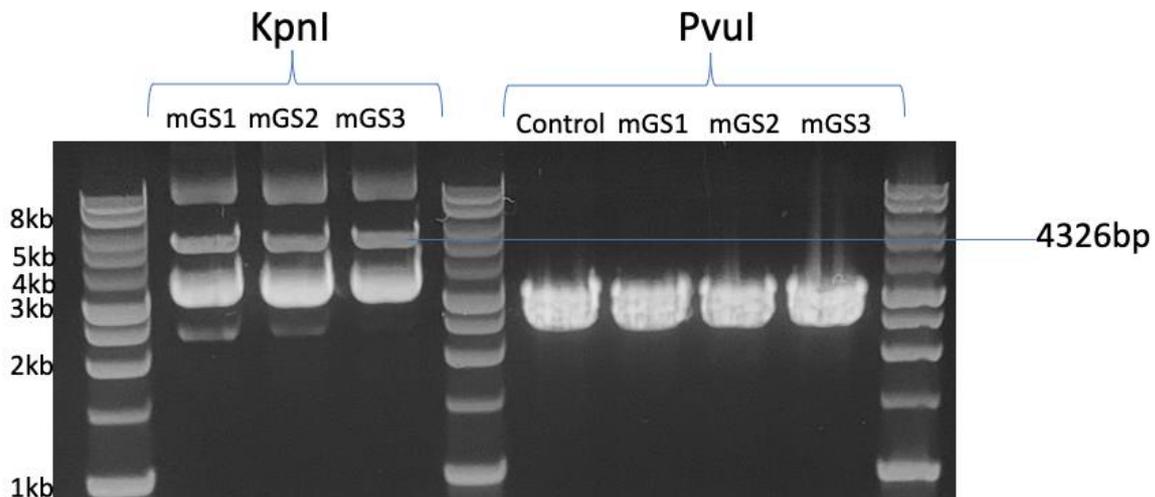
Figure 3.2.3: *Agarose gel analysis showing restriction enzyme digests for the mGS1, mGS2 and mGS3 samples (ligation between pcDNA3.1 +mGS). Digested samples using KpnI and PvuI with a control used in the PvuI restriction analysis.*

Overall, the aim of this experiment was to generate a new construct replacing the hygromycin gene in the commercialised pcDNA3.1hygro vector with a mGS gene, in order to aid stable cell line development and spike protein production. As the construct was not able to be created in a timely manner, it was commercially ordered and made instead.

3.1.4 Generation of eGFP constructs for transient transfection

To be able to visually view the location of the SARS-Cov-2 constructs within the CHO-S cells, and to allow a rapid method to monitor expression, eGFP was tagged to the constructs to be able to determine fluorescence by confocal microscopy, highlighting location of the spike protein in the cell, or by flow cytometry to gauge expression in cells. The restriction sites to use were determined by looking at the sequencing of the SARS-CoV-2 constructs to find where the His-Tag was attached and how to remove this without removing Spike encoding base pairs and for the correct primers to be designed. When designing the Nat-FL-eGFP construct, due to the restriction sites, the Nat-FL had to be re-ligated into a pcDNA3.1hygro vector before adding the eGFP tag. The SARS-CoV-2 constructs were cloned using the Phusion

enzyme to amplify the constructs and the specific restriction sites. These were then run on an agarose gel and gel extracted as shown in Figure 3.2.4 (A). The expected band dropouts were observed as outlined in Figure 3.2.4(B)

A
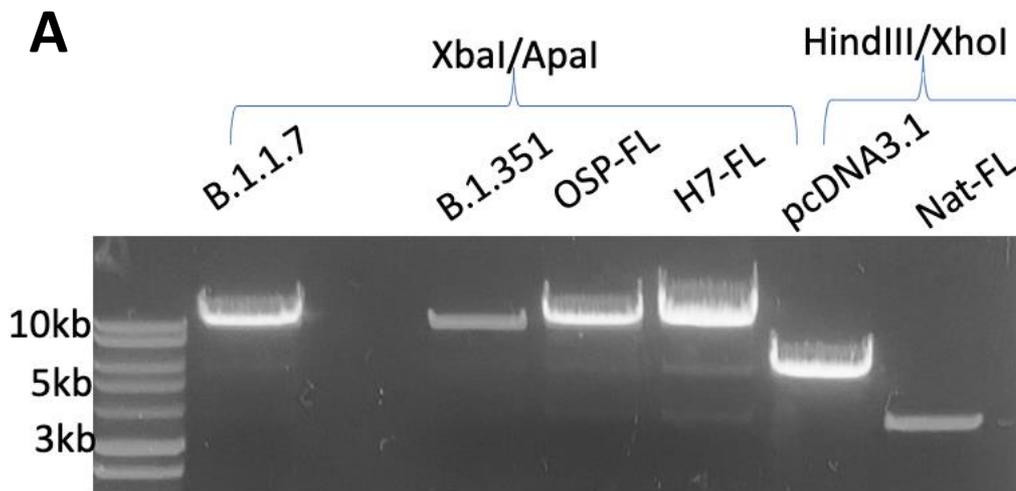
B

Construct	Expected Band Drop (bp)
B.1.1.7	9134
B.1.351	9246
OSP-FL	9142
H7-FL	9149
pcDNA3.1	5523
Nat-FL	5523

Figure 3.2.4: *(A) Agarose gel analysis showing the results of a restriction enzyme digest using either XbaI/ApaI or HindIII/XhoI to digest the B.1.1.7, B.1.351, OSP-FL, H7-FL, pcDNA3.1 and Nat-FL constructs alongside a 10 kb DNA ladder. There was no sample in lane 3. (B) Table showing the expected band dropouts for each of the vector constructs.*

After clean-up of the different SARS-CoV-2 constructs, the eGFP inserts were then amplified using Phusion with primers ate each end designed to incorporate the required restriction sites for the individual SARS-CoV-2 vectors. The PCR products were then run on an agarose gel,

extracted and cleaned up. The result of the agarose gel is reported in Figure 3.2.5(B) with the eGFP inserts and their restriction enzymes and SARS-CoV-2 vectors detailed in Figure 3.2.5(A).

A

eGFP Insert No	SARS-CoV-2 Vector	Restriction enzymes (for primers)
1	Control	HindIII/XhoI
2	B.1.1.7	XbaI/ApaI
3	B.1.351	XbaI/ApaI
4	OSP-FL	XbaI/ApaI
5	H7-FL	XbaI/ApaI
6	Nat-FL	XhoI/ApaI

B
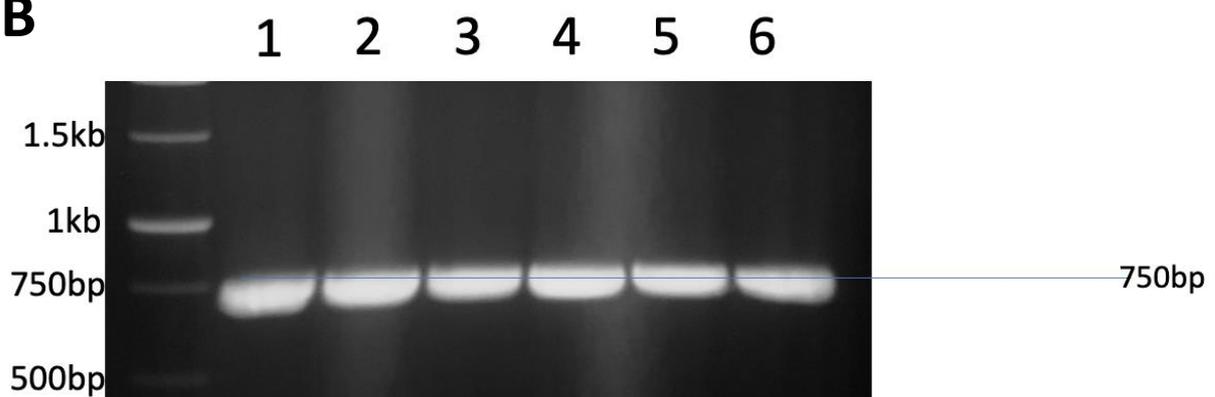
Figure 3.2.5: (*A) Table outlining the labelled 1-6 eGFP inserts, which SARS-CoV-2 vector they are assigned to and the restriction enzyme primers that were used for each. (B) Agarose gel analysis displaying the results of the Phusion PCR products of the eGFP inserts with the expected bands present.*

In order to check the quality of the DNA vectors and inserts before ligation, an analytical agarose gel was run of the clean-up products of the digests. The results of this are outlined in Figure 3.2.6.
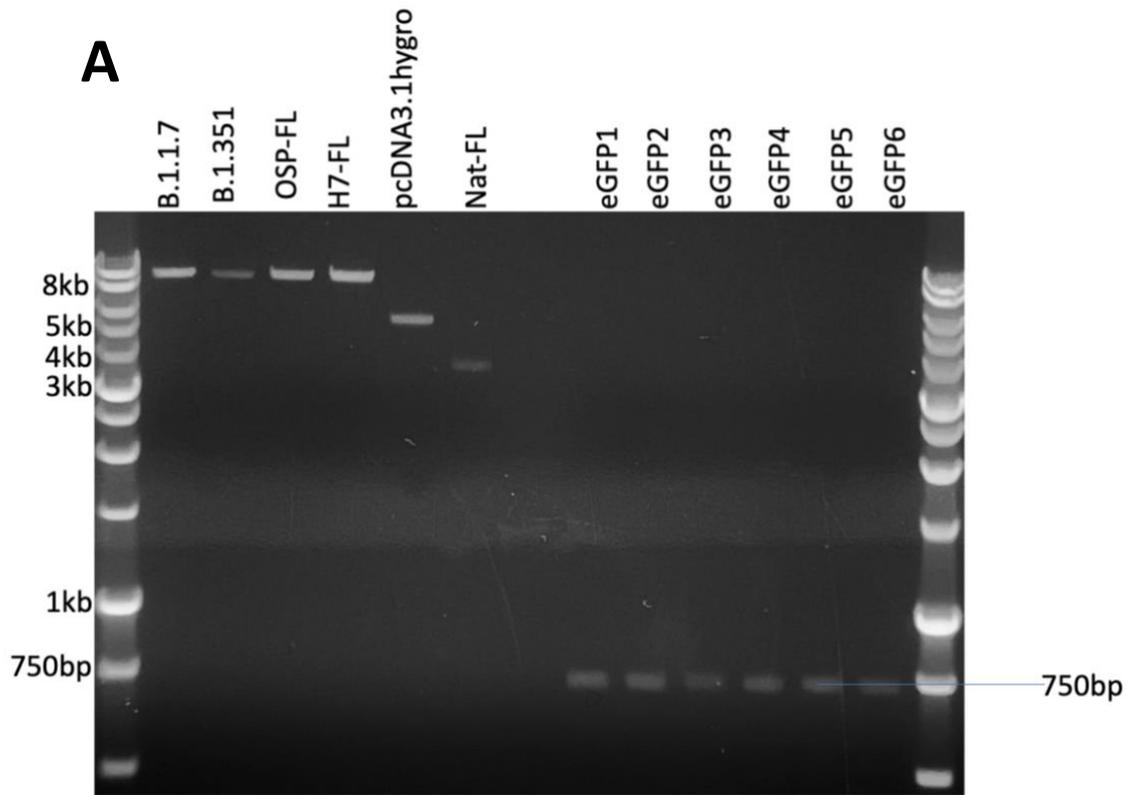
Figure 3.2.6: *(A) Agarose gel analysis detailing the expected band dropouts for the SARS-CoV-2 spike protein construct vectors and their respective eGFP inserts in a side by side comparison. (B) Table showing the expected band dropouts for each vector shown in base pairs (bp).*

The result of the agarose gel, as shown in Figure 3.2.6(A) confirmed the bands to be of the expected size and suitable for subsequent ligation experiments.

3.1.5 Generation of SARS-Cov-2 variant vectors for stable cell line development: B.1.1.7

As the pandemic progressed, there was a need to generate new constructs for variant Spike proteins that had arisen and were soon more prevalent than the original, native SARS-CoV-2 strain. The B.1.1.7 gene sequence was therefore designed and commercially ordered. To replace the variant sequence into the pcDNA3.1hygro vector, a restriction enzyme digest was carried out on the B.1.1.7 spike sequence using the *HindIII* and *XhoI* restriction enzymes. The results of after agarose gel electrophoresis analysis are outlined in Figure 3.2.7 showing the expected band dropout of around 3657 bp. This was then extracted and cleaned up ready for ligation with pcDNA3.1hygro and pcDNA3.1mGS vectors both for transient and stable transfections in future experiments (See Appendix 11 and 12).
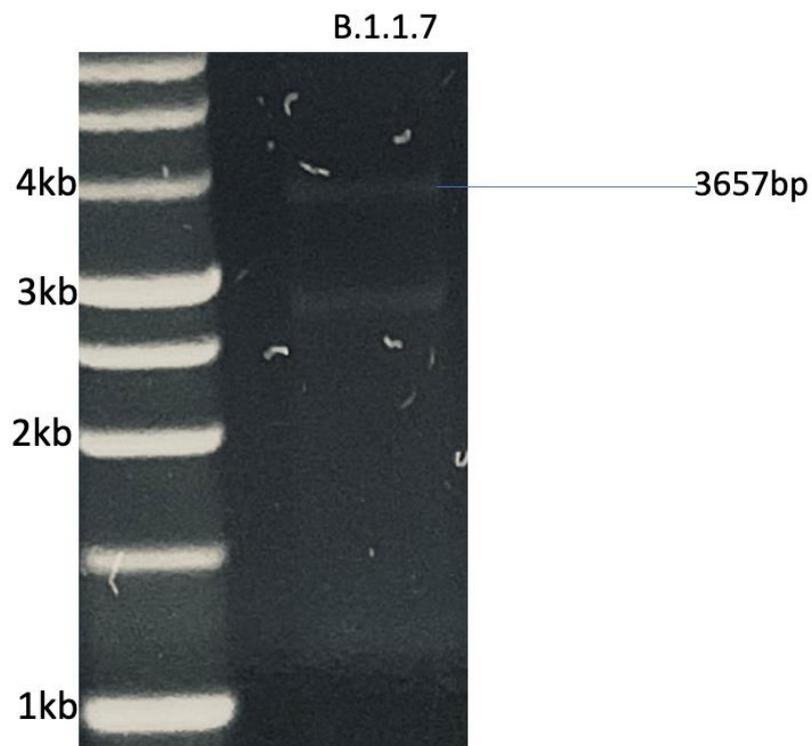
Figure 3.2.7: *Agarose gel analysis showing the result of a restriction enzyme digest using HindIII and XhoI restriction enzymes resulting in the expecting band dropout of around 3657 bp for the B.1.1.7 spike variant gene fragment.*

After ligation of the B.1.1.7 spike gene sequence and the pcDNA3.1hygro and pcDNA3.1mGS vectors, transformations were carried out into *DH5α E. coli* and a PCR colony screen was run to check for positive clones. As outlined in Figure 3.2.8, samples 2, 3 and 4 were positive clones. As such, one of each of the ligations had worked which enabled these to be progressed.
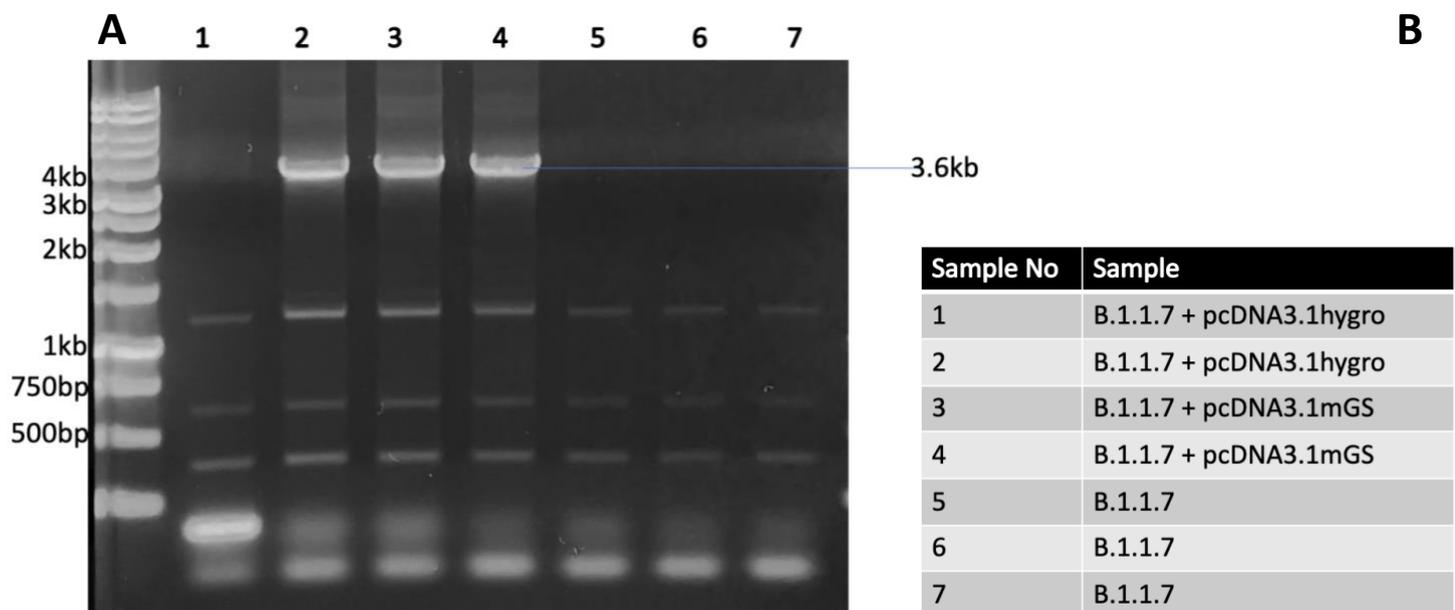
Figure 3.2.8: (*A*) *Agarose gel analysis outlining results of a PCR colony screen to check for positive clones for the ligations of pcDNA3.1hygro and pcDNA3.1mGS with the B.1.1.7 variant spike gene respectively. (B) Table outlining the samples and their corresponding numbers on the lanes of the agarose gel.*

Using the results of the PCR colony screen the B.1.1.7, B.1.1.7 pcDNA3.1hygro and B.1.1.7 pcDNA3.1mGS constructs were then miniprepped and the purified plasmid DNA quantified on a nanodrop spectrophotometer (Table 3.1.4).

Table 3.1.4: *Quantification of the B.1.1.7 constructs with either pcDNA3.1hygro or pcDNA3.1mGS vectors showing the concentration and purity of the purify plasmid DNA (in ng/μl and A260/280 respectively).*

Sample	ng/μl	A260/280
B.1.1.7 + pcDNA3.1hygro	517.6	1.90
B.1.1.7 +pcDNA3.1mGS	601.4	1.90
B.1.1.7	205.8	1.91

Table 3.1.4 shows that while the A260/280 ratio for all three constructs indicated appropriate purity of DNA, the concentrations (ng/μl) were quite low with B.1.1.7 having 205.8 ng/μl. The B.1.1.7 pcDNA3.1mGS had the highest concentration of the three with 601.4 ng/μl, appropriate for easily generating sufficient material for stable cell line development.

To check the ligations were correct, an analytical gel was run using the restrictive enzymes *HindIII* and *XhoI*, as well as *PmlI* to check that the expected band dropouts were present. The digests were run alongside Nat-FL and H7-FL constructs as a point of comparison. The results of these restriction digests are outlined in Figures 3.2.9 and 3.3.1.
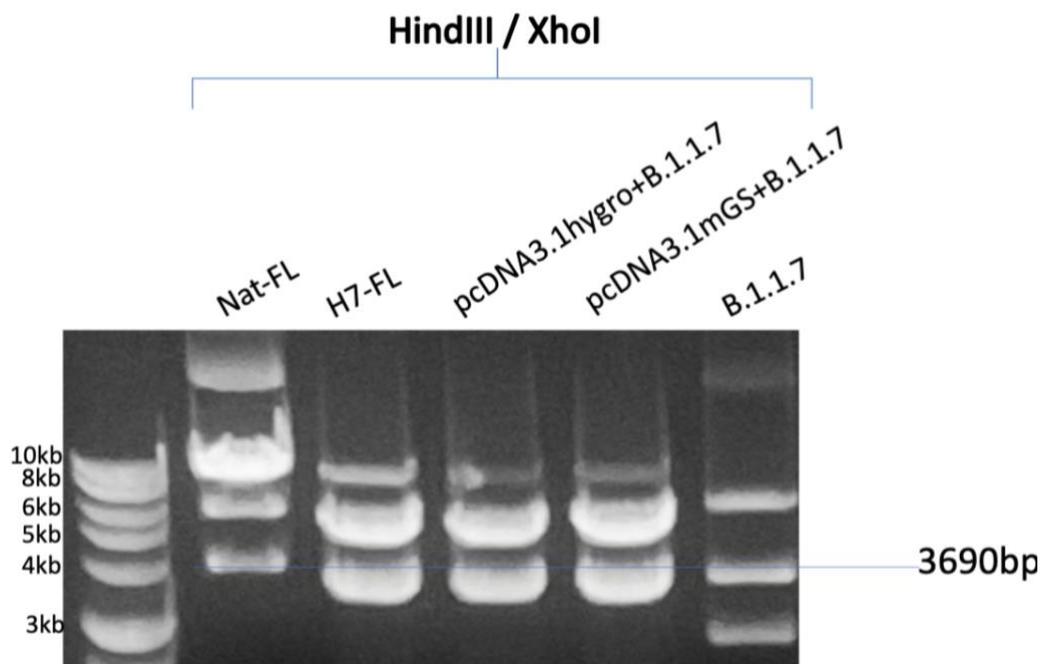
Figure 3.2.9: *Agarose gel analysis showing the results of a HindIII/XhoI restriction of the pcDNa3.1hygro + B.1.1.7 and pcDNA3.1mGS +B.1.1.7 constructs to check for the correct ligation via the expected band dropouts. Compared to Nat-FL and H7-FL constructs. First lane outlining 10 kb DNA band ladder to indicate size of bands.*
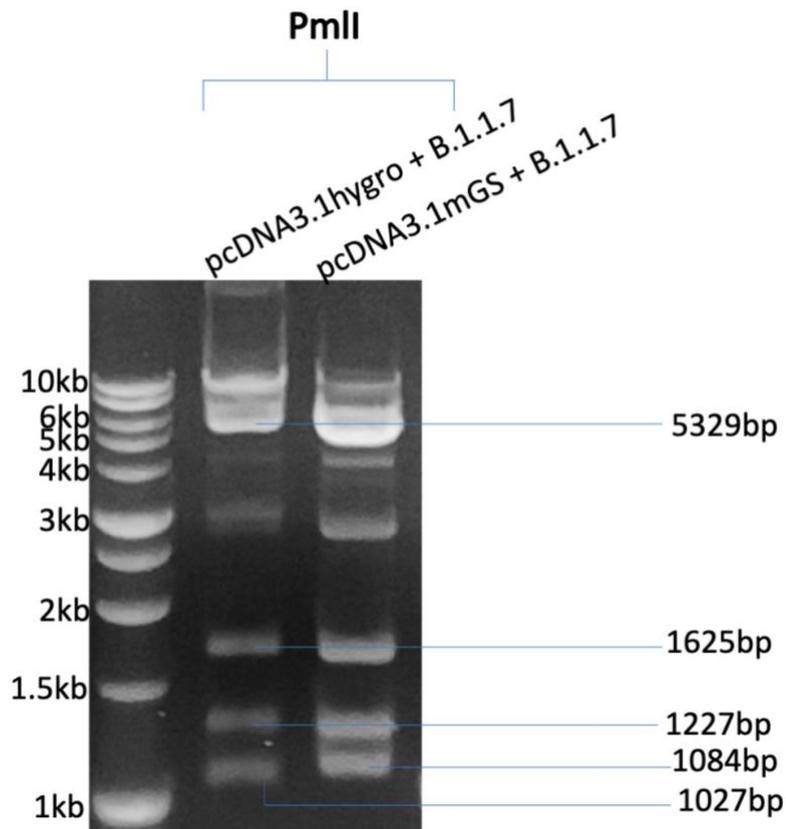
Figure 3.3.1: *Agarose gel analysis showing the results of a PmlI restriction of the pcDNa3.1hygro + B.1.1.7 and pcDNA3.1mGS +B.1.1.7 constructs to check for the correct ligation via the expected band dropouts. First lane outlining 10 kb DNA band ladder to indicate size of bands.*

The data in Figure 3.2.9 shows that the expected band dropouts of 3690 bp for the B.1.1.7 spike gene and 5329 bp were present, however there was a third band above the 5329 bp mark which could imply uncut vector. This could be due to an issue in the enzymes efficiency or too much product being loaded into the digest. To further confirm the correct ligation of the two constructs, Figure 3.3.1 shows the expected band dropouts of 1027 bp, 1084 bp, 1227 bp, 1625 bp and 5329 bp. However, there were also additional band dropouts in the 3 kb and 4 kb regions which again suggests an issue with the enzyme. Maxipreps were then prepared of these constructs and quantified using nanodrop spectrophotometer technology, outlined in Table 3.1.5 below.

Table 3.1.5: *Quantification of the B.1.1.7 in either pcDNA3.1hygro or pcDNA3.1mGS using a nanodrop spectrophotometer to measure the concentration and the purity of the maxiprep samples.*

Construct	ng/µl	A260/280
B.1.1.7 + pcDNA3.1hygro	517.6	1.90
B.1.1.7 + pcDNA3.1mGS	601.4	1.90

3.1.6 Generation of SARS-CoV-2 constructs with the D614G mutation via site directed mutagenesis (SDM)

One of the key mutations that has arisen during the SARS-CoV-2 pandemic to date is the emergence of the D614G mutation. Using the Native-FL-His, H7-FL-His and B.1.1.7(hygro) constructs previously made, the D614G mutation was added via site directed mutagenesis (SDM). During the first round of SDM, the transformation was not successful for any of the constructs apart from Nat-FL-His for which there were 2 colonies on the plate. These were then miniprepped and an analytical restriction enzyme digest using *PmlI* was run on the two samples and analysed by agarose gel electrophoresis to determine if the correct expected band dropouts were present. The results of this gel are outlined in Figure 3.3.2.
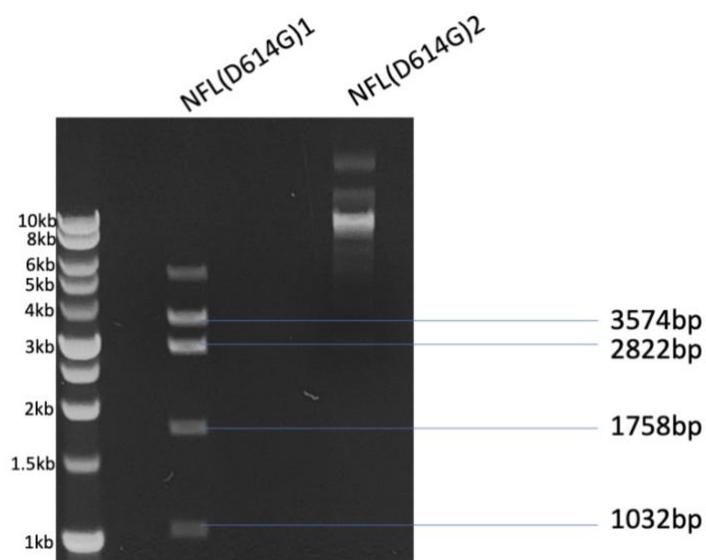
Figure 3.3.2: *Agarose gel analysis showing the results of an analytical restriction enzyme digest using PmlI to check if the colonies for the D614G constructs for the native full length spike had the correct expected band drop outs.*

The results of Figure 3.3.2 suggest that NFL(D614G)1 had the correct expected band drop outs of 1032 bp, 1758 bp, 2822 bp and 3574 bp, however there was an extra band present around the 4 kb region. This could be due to some partially uncut vector. NFL(D614G)2 did not have the correct expected band dropouts. The NFL(D614G)1 plasmid preparation had a concentration of 274.6 ng/μl. The A260/280 purity measurement of this sample was 1.92 indicating an appropriate purity.

The process of site directed mutagenesis was then repeated again for the Nat-FL, H7-FL and B.1.1.7 pcDNA3.1hygro samples but 5 μl of template DNA was added instead of 2μl. This resulted in 4 colonies for each of the constructs which were then miniprepped. All 4 colonies from each construct were miniprepped due to the possibility of the *DpnI* enzyme destroying the original plasmid. These minipreps were then quantified via nanodrop spectrophotometer with the results outlined in Table 3.1.6 below.

Table 3.1.6: *Quantification of the miniprepped D614G samples showing concentration (ng/μl), and purity (A260/280) and which minipreps were chosen to take forward.*

Sample No	Construct Name	Ng/μl	260/280	Good Quality
1	H7-FL(D614G)	12.2	2.48	
2	H7-FL(D614G)	410.5	1.91	✓
3	H7-FL(D614G)	268.3	1.92	✓
4	H7-FL(D614G)	345.6	1.90	
5	B.1.1.7(D614G)	287.5	1.92	✓
6	B.1.1.7(D614G)	441.6	1.90	✓
7	B.1.1.7(D614G)	37.6	2.27	
8	B.1.1.7(D614G)	9.2	2.27	
9	Nat-FL(D614G)	157.9	1.96	✓
10	Nat-FL(D614G)	292.6	1.94	
11	Nat-FL(D614G)	88.9	1.99	
12	Nat-FL(D614G)	7.0	2.29	

The results outlined in Table 3.1.6 indicated that many of the minipreps were not good quality because of their concentration or purity. The samples 2, 3, 5, 6 and 9 were chosen due to

their high concentrations (ng/μl) and purity (A260/280) in the 1.80 – 2.00 range as this indicates a good purity DNA sample. These samples were then investigated using an analytical restriction enzyme digest with *PmlI* to check the expected band dropouts were released. The results of this restriction enzyme digest with *PmlI* is outlined in Figure 3.3.3.
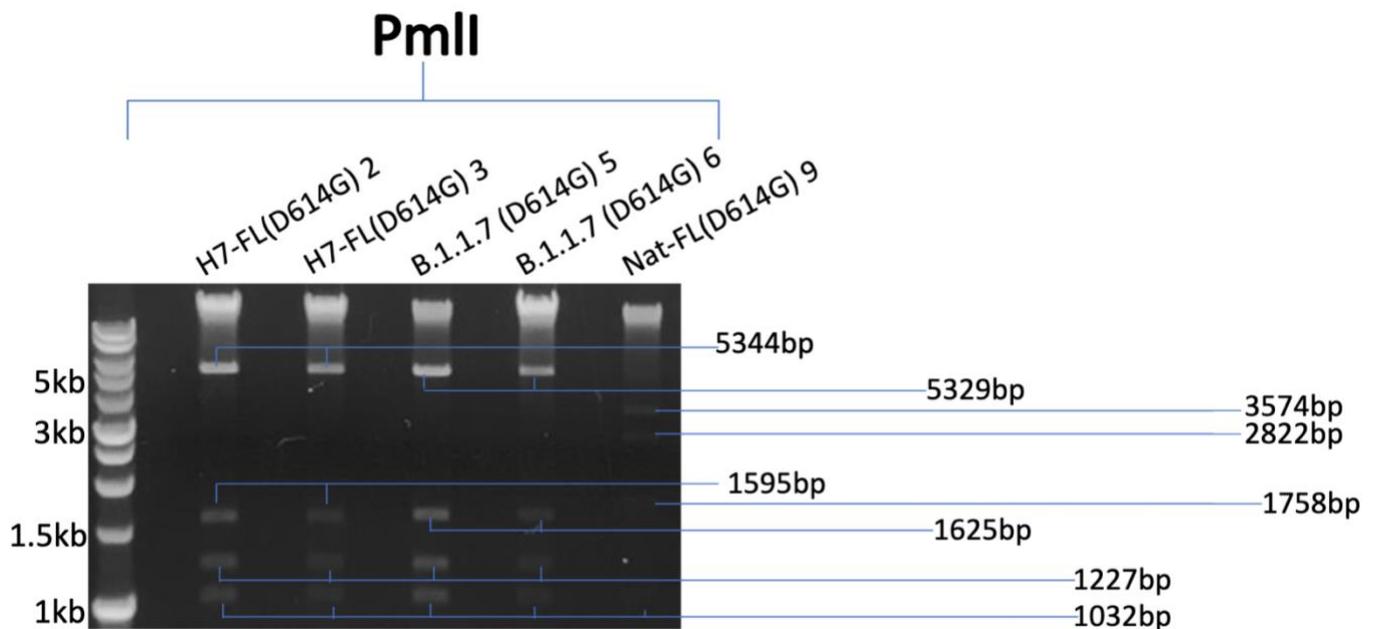
Figure 3.3.3: *Agarose gel analysis showing the results of the analytical restriction enzyme digest using PmlI to check the correct expected band dropouts were present for the three constructs (H7-FL(D614G), B.1.1.7(D614G) and Nat-FL(D614G). Minipreps 2, 3, 5, 6 and 9 were chosen due to their high concentrations and correct purities.*
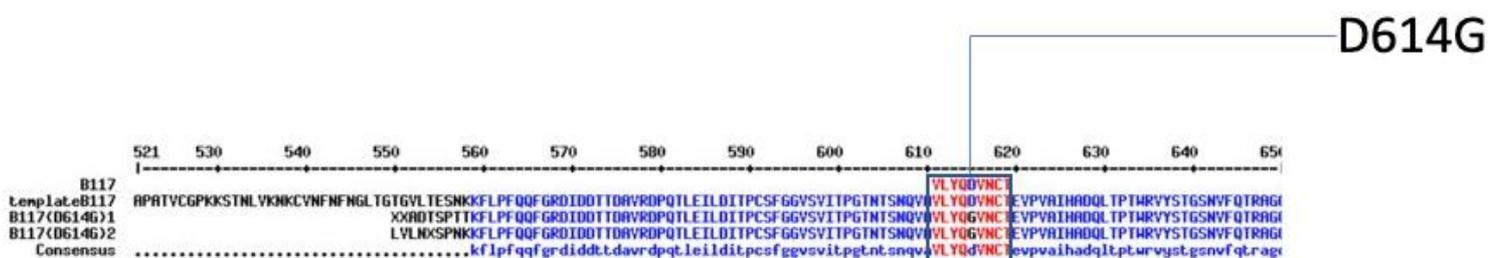
Figure 3.3.4: *Sequence alignments of B.1.1.7 sample showing the D→G mutation at the 614-position comparing the template B.1.1.7 sequence to the B.1.1.7 (D614G) samples*

Figure 3.3.3 indicates that H7-FL(D614G)2, H7-FL(D614G)3, B.1.1.7(D614G)5 and B.1.1.7(D614G)6 all had the correct expected band dropouts. These data provide evidence that the plasmids contained the correct insert that had not been damaged in the SDM process. Nat-FL(D614G)9 also had the correct expected band dropouts however they were faint. All 5 constructs had a lot of uncut vector at the top of the gels which could explain the

reason for the faint band drop outs. In order to check that the site directed mutagenesis was correct and the D614G mutation had been successfully inserted into the sequence, sequencing had to be carried out. As the B.1.1.7 (D614G) was of more relevance, this was sent off for sequencing and then aligned using the tool multialign. The results of the alignment of the sequences are shown in Figure 3.3.4.

Figure 3.3.4 shows the sequence alignment confirming that there was the D→ G change at position 614 in the B.1.1.7(D614) minipreps when compared to the original B.1.1.7 template sequence. The rest of the sequence also confirmed that no further mutations had been introduced during the SDM procedure. Overall, these results indicate the successful generation of B.1.1.7 (D614G) constructs ready for transfection and expression.

3.1.7 Generation of B.1.1.7 SARS-CoV-2 constructs with D614G and E484K mutations by site directed mutagenesis

Another mutation of interest emerged in the B.1.1.7 strain called E484K. This mutation is present in many variants of the SARS-CoV-2 spike protein. To add this mutation to the previously constructed B.1.1.7 (D614G) constructs, site directed mutagenesis was carried out to add the single point mutation E484K. Four subsequent colonies were analysed via an analytical restriction enzyme digest using *HindIII* and *XhoI* to check the correct expected band drop outs were still present. The result of this is shown in Figure 3.3.5.

Figure 3.3.5: *Agarose gel analysis showing the results of an analytical restriction enzyme digest using restriction enzymes HindIII and XhoI to separate the expected band dropouts of the B.1.1.7 spike gene (3690 bp) and the pcDNA3.1hygro vector (5523 bp).*

The result of Figure 3.3.5 shows that the expected band dropouts were present for all four samples. The band dropouts did appear to run slightly higher than expected. Due to the size and drop out of the bands they were expected to be correct. In order to check the single point mutation E484K had been added correctly, the samples were sequenced and then aligned, the results displayed in Figure 3.3.6.

Figure 3.3.6: *The results of sequence alignment comparing the B.1.1.7 (KV) original sequence to the B.1.1.7 D614G E4848K (KVE4848K S1-S4). The sequence alignment shows the E→K change at the 484 position in the sequence as well as the D→G change in 614 position.*

Table 3.1.7: *The quantification, using a nanodrop spectrophotometer, of the concentration (ng/µl) and purity (A260/280) of the B.1.1.7 D614G E484K samples.*

Sample	ng/µl	A260/280
B.1.1.7 D614G E4848K 1	659.0	1.91
B.1.1.7 D614G E4848K 2	610.3	1.91
B.1.1.7 D614G E4848K 3	702.3	1.92
B.1.1.7 D614G E4848K 4	698.6	1.91

The results of the minipreps being quantified (Table 3.1.7) indicated that the B.1.1.7 D614G E484K 3 preparation had the highest concentration of 702.2 ng/µl with a purity (A260/280)

of 1.92 indicating this was a good quality sample. As well as being a good quality sample, the B.1.1.7 D614G E484K 3 sample in Figure 3.3.6 showed no extra mutations compared to the others which had extra mutations in the 630-650 region. In conclusion, B.1.1.7 D614G E484K 3 was found to be the highest quality, reliable construct with no extra mutations and with the correct E484K and D614G mutations.

3.1.8 Determination of Other B.1.351 SARS-CoV-2 Spike Protein Variant Mutations that could be Generated

Through a thorough literature review and continual monitoring of reported variants, the following mutations were selected to be added to the B.1.351 variant construct which would be sent off for synthesis. Sources were limited and not always reputable due to the fast-paced release of information and non-peer reviewed releases, requiring sources to be compared to each other to determine those mutations that we were most confident in, that were reported across various sources and not a single point mutation that only occurred as a one off. The particular identified mutations of interest were as displayed in Table 3.1.8.

Table 3.1.8: *List of mutations gathered to be added to the B.1.351 plasmid construct alongside the sources that state the mutations*

Mutations	Source
L18F	PHE Technical Briefing 6, CoV Lineages.com
D80A	PHE Technical Briefing 6, CoV Lineages.com, GISAID
D215G	PHE Technical Briefing 6, CoV Lineages.com, GISAID
R2461	PHE Technical Briefing 6, CoV Lineages.com
K417N	PHE Technical Briefing 6, CoV Lineages.com, GISAID
E484K	PHE Technical Briefing 6, CoV Lineages.com, GISAID
N501Y	PHE Technical Briefing 6, CoV Lineages.com, GISAID
D614G	Tegally *et al* 2020, GISAID
A701V	PHE Technical Briefing 6, CoV Lineages.com, GISAID
Del L242- 244L	PHE Technical Briefing 6, CoV Lineages.com

The Public Health England VOC Technical briefing (published 13/2/21) outlined a list of mutations that were present in the B.1.351 variant, as shown in Table 3.1.9. As this is a British government document that uses multiple reliable studies to generate the results, and with the high levels of genomic sequencing available in the UK, this was considered a reliable source of such mutations and the choice of the mutations were mainly based from this document.

Table 3.1.9: *List of some of the reported mutations present in the B.1.351 Spike variant. Source taken from 'Public Health England Investigation of SARS-CoV-2 Variants of Concern in England , Technical Briefing 6'. Mutations highlighted in **red** indicates acquisition in subset of isolates within the lineage.*

Gene	amino_acid	actual_nucleotide
S Gene	L18F	21614C>T
	D80A*	21801A>C
	D215G*	22206A>G
	R246I	22299G>T
	K417N*	22813G>T
	E484K*	23012G>A
	N501Y*	23063A>T
	A701V*	23664C>T
	242-244del	

The list of mutations in Table 3.1.9 above was then examined alongside other sources to check the reliability before constructing a plasmid containing these point mutations. Another source used to look at the point mutations present in the B.1.351 was the GISAID database containing the COVID-19 CoV Genetics. The database is consistently updated with the key point mutations found in the various lineages. As outlined in Table 3.2.1, the mutations are displayed in comparison to the reference lineage.

Lineage	Seqs	%Seqs	54	69	80	144	215	417	484	501	570	614	681	688	701	716	769	846	982	1078	1118	1176
Ref			L	HV	D	Y	D	K	E	N	A	D	P	A	A	T	G	A	S	A	D	V
B.1.351	833	25.85	L	HV	A	Y	G	N	K	Y	A	G	P	A	V	T	G	A	S	A	D	V

Table 3.2.1: *The point mutations present in the B.1.351 variant present in comparison to the reference lineage (GISAID). The point mutations in the B.1.351 variant is highlighted in yellow to show the amino acid change between the reference lineage and the B.1.351 variant at the different sequence points*

Table 3.2.1 reports the following mutations to be present in the B.1.351 variant: D80A, D215G, E484K, N501Y, D614G and A701V. In comparison to the PHE VOC briefing, as shown in Table 3.1.9, these mutations are concurrent with each other although not all the mutations in Table 3.1.9 are outlined in Table 3.2.1. This could be due to the limited availability of the data available for this variant at the time of these reports.

Another source 'CoV lineages reports' also had a list of mutations (Figure 3.3.7). The following mutations are concurrent with the PHE VOC Briefing, as shown in Table 3.1.9 as well as the GISAID Database, as shown in Table 3.2.1; D80A, D215G, K417N, A701V, N501Y and E484K.

aa:S:D80A
aa:S:D215G
aa:S:K417N
aa:S:A701V
aa:S:N501Y
aa:S:E484K

Figure 3.3.7 T*aken from 'Cov-lineages.org' detailing the list of Spike amino acid point mutations present in the B.1.351 variant.*

A recent paper by Tegally *et al* 2020 (44) also detailed a list of SARS-CoV-2 B.1.351 variant Spike point mutations. The paper reported the D614G, D80A, D215G, E484K, N501Y and A701V mutations to be present when sampling in October 2020, however by November 2020 three new mutations had been sequenced and added, L18F, R246I and K417N. A deletion was also observed at L242_244L. These results had been gathered from various samples from different laboratories indicating the results to be reliable (44). When comparing Table 3.1.9

to Figure 3.3.7, all the mutations were the same, indicating a high level of reliability in these results which could be used to generate a finalised list of mutations to be added to the B.1.351 variant plasmid being designed.

3.1.9 Subcloning of the B.1.351 SARS-CoV-2 Spike construct into the pcDNA3.1hygro vector

To remove the B.1.351 SARS-CoV-2 spike gene from its original plasmid provided by the gene synthesis company, the plasmid was subjected to a restriction enzyme digest using *HindIII* and *XhoI* with the expected band dropouts being present at approximately 3.6 kb for the B.1.351 gene insert and 2.2 kb for the vector backbone (Figure 3.3.8). The B.1.351 insert was then gel extracted, cleaned up and ligated with the commercial vectors pcDNA3.1hygro and pcDNA3.1mGS (See Appendix 13)
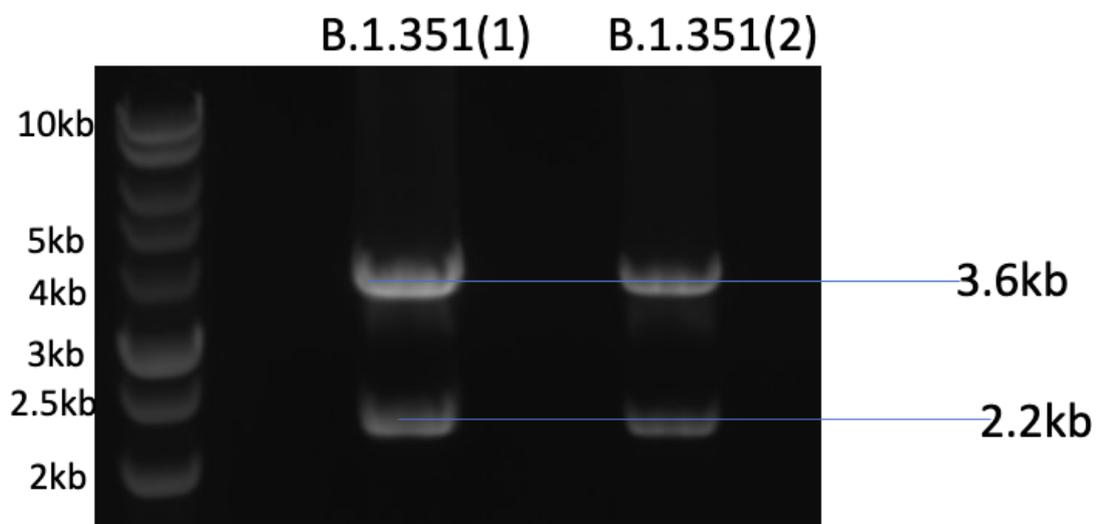
Figure 3.3.8: *Agarose gel displaying the results of a restriction enzyme digest using restrictive enzymes HindIII and XhoI to separate the B.1.351 gene insert (3.6 kb) from its original vector at around the 2.2 kb region.*

The colonies from ligations and transformation of *E. coli* were then subjected to a restriction enzyme digest using *HindIII* and *XhoI* and then analysed on an agarose gel (Figure 3.3.9).

Figure 3.3.9: *Agarose gel analysis of a restriction enzyme digest using HindIII and XhoI to check the ligations between the B.1.351 insert (3690 bp) and the pcDNA3.1hygro and pcDNa3.1mGS, 5523 bp and 5635 bp respectively.*

Figure 3.3.9 data indicates that the ligations were correct for both the constructs B.1.351(pcDNA3.1hygro) and B.1.351(pcDNA3.1mGS), having the expected band dropouts present for the commercial vectors pcDNA3.1hygro (5523 bp) and pcDNA3.1mGS (5635 bp) as well as the B.1.351 insert around 3690 bp. These ligated constructs were then transformed into *DH5alpha E. coli* and a PCR colony screen was run on the colonies to check for positive clones. The results of the PCR colony screen are outlined in Figure 3.4.1.
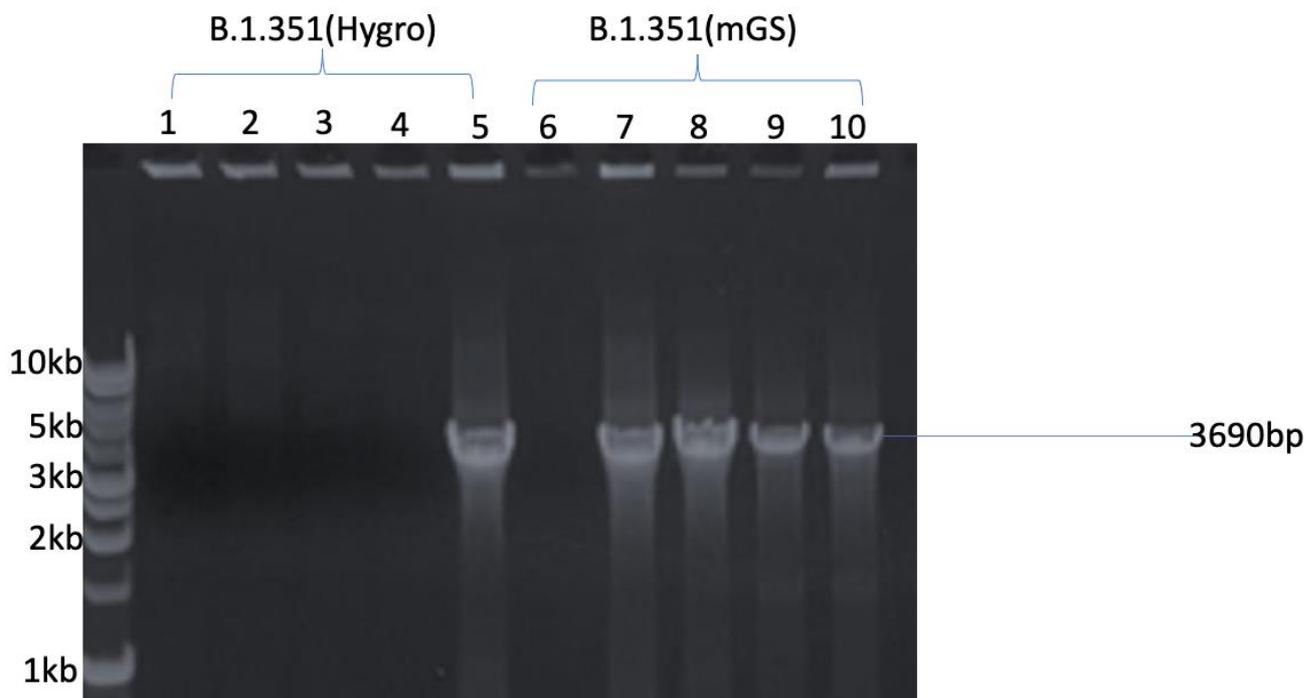
Figure 3.4.1: *Agarose gel analysis of a PCR colony screen of the B.1.351 (pcDNA3.1hygro) and B.1.351(pcDNA3.1mGS) ligations and the result of the transformations in DH5alpha E. coli. The band dropouts indicate a positive clone.*

The result of the PCR colony screen (Figure 3.4.1) indicates 5 of the 10 colonies were positive. Colony 5 of the B.1.351(pcDNA3.1hygro) construct and Colonies 7, 8, 9 and 10 of B.1.351(pcDNA3.1mGS) construct were positive clones. As one of each construct were positive, Colony 5 and Colony 7 were then miniprepped and quantified (Table 3.2.2).

Table 3.2.2: *The quantification of the B.1.351(pcDNA3.1hygro) and B.1.351(pcDNA3.1mGS) constructs using nanodrop spectrophotometry. The concentration (ng/μl) and the purity (A260/280) of the samples are reported.*

Sample	ng/μl	A260/280
B.1.351(pcDNA3.1hygro)	830.4	1.92
B.1.351(pcDNA3.1mGS)	895.0	1.93

Table 3.2.2 indicates that both B.1.351(pcDNA3.1hygro) and B.1.351(pcDNA3.1mGS) plasmid preparations had high concentrations of 830.4 ng/μl and 895.0 ng/μl respectively with good purities (A260/280) of 1.92 and 1.93 respectively. These samples were then checked using an analytical restriction enzyme digest using *PmlI*. The results of this are outlined in Figure 3.4.2
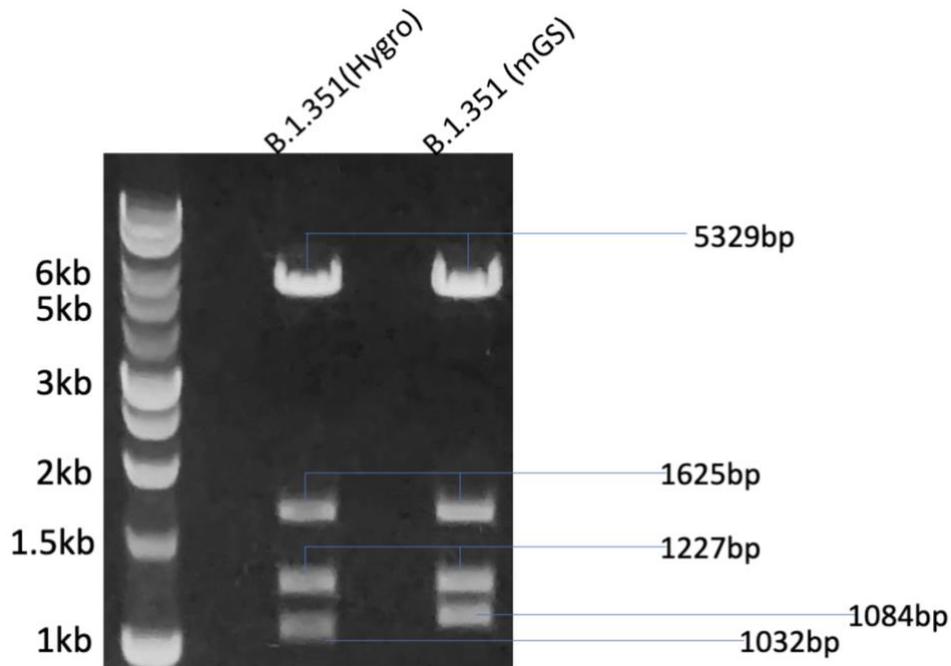
Figure 3.4.2: *Agarose gel analysis of an analytical restriction digest using PmlI showing the expected band dropouts for the B.1.351(pcDNA3.1hygro) and B.1.351(pcDNA3.1mGS) minipreps.*

The results of the analytical restriction digest shown in Figure 3.4.2 confirmed that both the B.1.351(pcDNA3.1hygro) and B.1.351(pcDNA3.1mGS) minipreps had the expected band dropouts as outlined in Figure 3.4.2 with clear bands. The B.1.351(pcDNA3.1hygro) miniprep was then amplified across 5 Maxipreps for transient transfection experiments with an analytical restriction digest run using the *HindIII* and *XhoI* restriction enzymes. The result of this digest is shown in Figure 3.4.3 below.
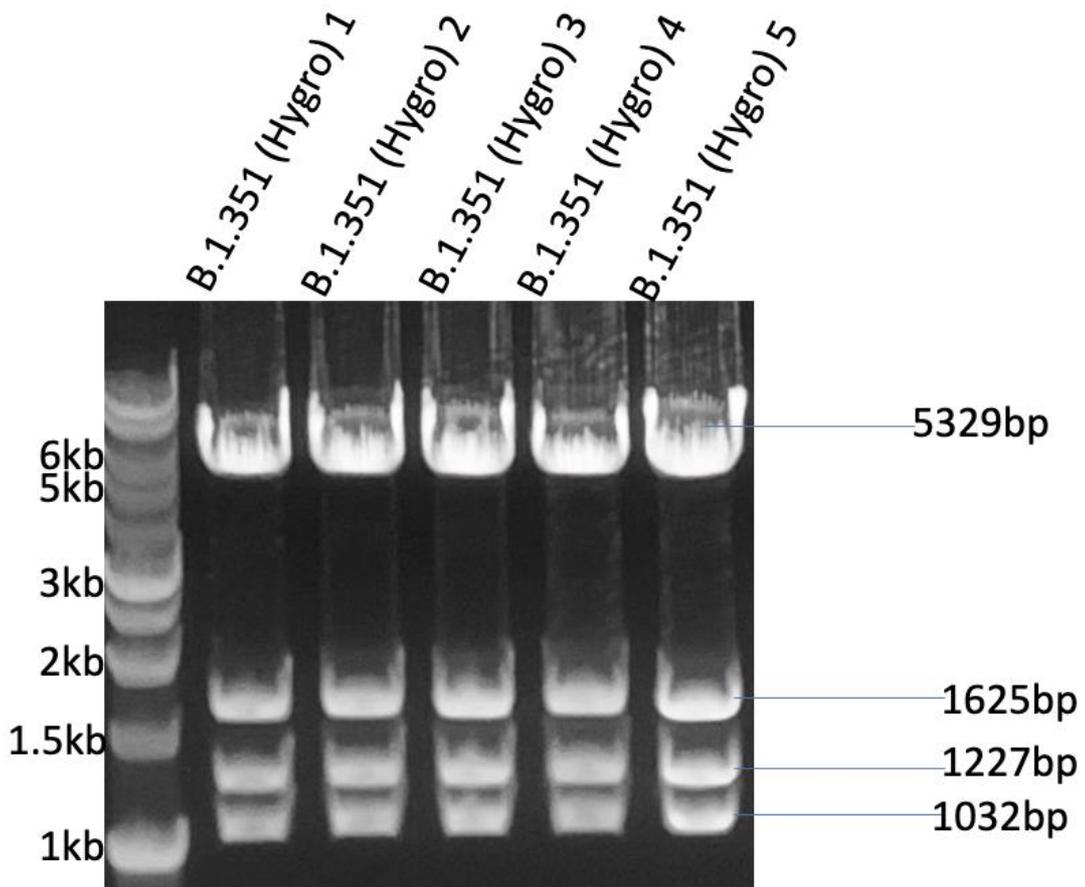
Figure 3.4.3: *Agarose gel analysis of an analytical restriction enzyme digest using restriction enzyme PmlI to check the correct expected band dropouts are present for the B.1.351(pcDNA3.1hygro) samples.*

The results of the restriction enzyme digest using *PmlI* outlined in Figure 3.4.3 suggests that the expected band dropouts are present. As a result of this, the B.1.351 spike gene insert was successfully ligated into the commercial vectors pcDNA3.1hygro and pcDNA3.1mGS ready for future transfections.

3.2 Protein expression

3.2.1 Comparing CHO-K1 vs CHO-S host cell lines via transient transfection of native SARS-CoV-2 constructs

In order to start the optimisation of the quality and quantity of the spike protein being produced transiently from CHO cells, two host cell lines were evaluated. To determine which CHO cell line was most efficient for the expression of the native SARS-CoV-2 constructs, the CHO-S and CHO-K1 cell lines were compared. The same four native SARS-CoV-2 constructs were transiently transfected into both cell lines and protein samples were taken on days 4, 5 and 6 after the transfections. The protein samples were then analysed on western blots to determine the protein expression efficiency of the cell lines. The culture viability and viable cells were also measured of both the CHO-S and CHO-K1 cell lines over day 4-6 post-transfection. The results are displayed in Figure 3.4.4.
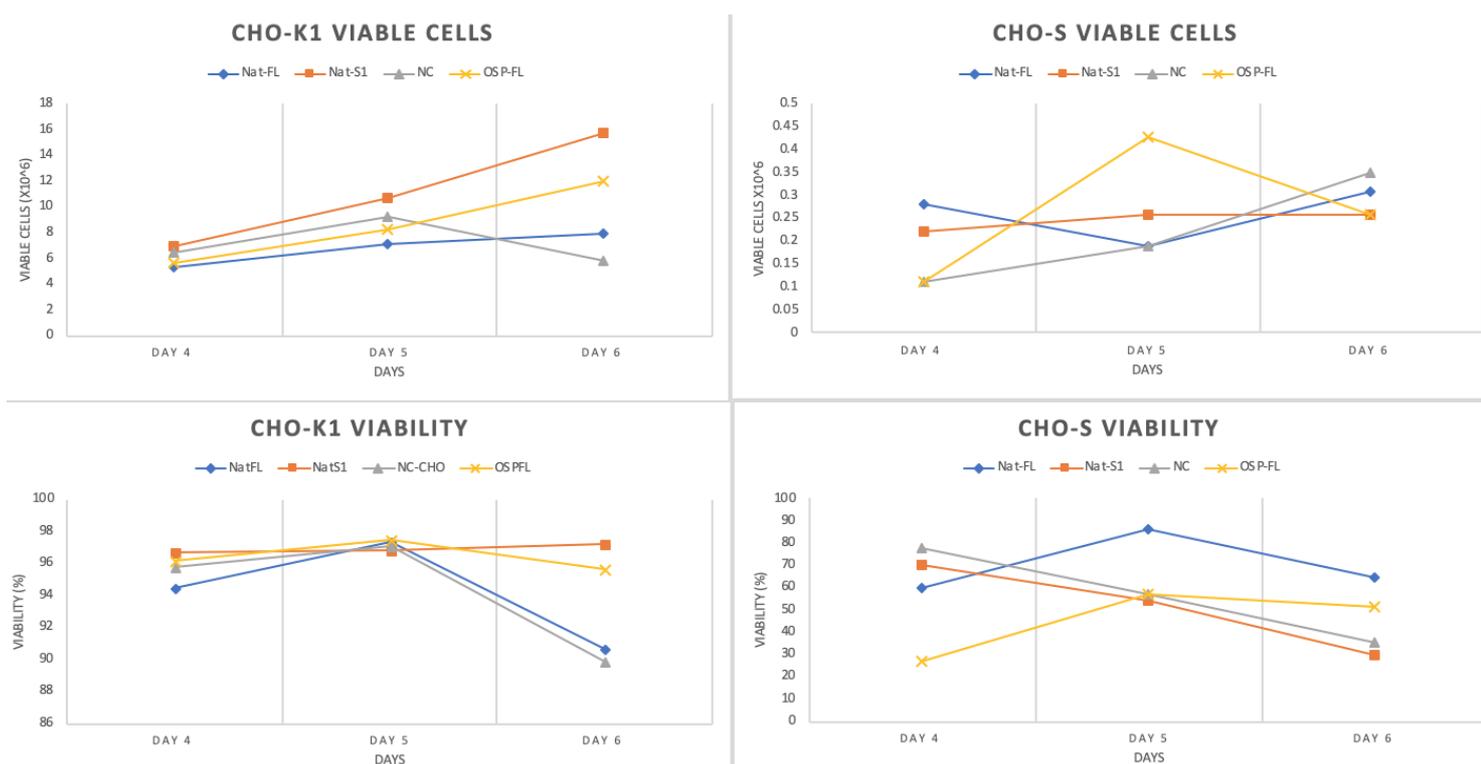
Figure 3.4.4: *The culture viability and viable cell concentration (x 10^6 cells/ml) of the CHO-S and CHO-K1 cell lines over days 4, 5 and 6 using a ViCell analyser.*

Comparing the CHO-S and CHO-K1 cell line data over days 4, 5 and 6 (Figure 3.4.4) revealed that the CHO-K1 cell line showed a gradual increase in viable cell numbers over the three-day period when transfected with the DNA constructs Nat-FL, Nat-S1 and OSP-FL. For Nat-FL, the

viable cells increased from 5.35×10^6 on day 4 to 7.93×10^6 cells/ml on day 6. For Nat-S1, the viable cells had an increase of 8.78×10^6 viable cells/ml from days 4 to 6. For the OSP-FL construct, the viable cells increased from 5.71×10^6 to 11.95×10^6 cells/ml over the three-day period. The NC construct viable cells decreased from days 5 to 6 by around 3.33×10^6 viable cells/ml.

For the CHO-S cells, these were at much lower numbers with the concentration of viable cells being in the $0.1\text{-}0.45 \times 10^6$ cells/ml range, considerably less than the CHO-K1 cells. For the Nat-FL construct the viable cells decreased on days 4 to 5 from 0.28×10^6 to 0.19×10^6 cells/ml but then increased on day 6 to 0.31×10^6. For the Nat-S1 transfected cells, the concentration stayed fairly consistent over the three-day period with only a 0.04×10^6 cells/ml increase. For the OSP-FL transfected cells (See Appendix 8) there was a large difference, on day 5 there was a spike in viable cells from 0.11×10^6 cells on day 4 to 0.43×10^6 cells/ml on day 5. After day 5, however the viable cells decreased to 0.26×10^6 cells on day 6. For the NC construct transfected cells there was a gradual increase in viable cells over days 4, 5 and 6 with an increase of 0.24×10^6 cells over the three-day period. By day 6, all four construct transfected cells were more-or-less the same with only a 0.09×10^6 cells/ml range between them.

Comparing the CHO-S and CHO-K1 culture viability (Figure 3.4.4) the results were quite different. For the CHO-K1 culture viability, the Nat-S1 and OSP-FL viabilities showed consistent levels with the Nat-S1 viability increasing by 0.05% and the OSP-FL decreasing by 0.08% over the three-day period. For the Nat-FL and the NC however, both sets of cells increased in viability on day 5, with 97.4% cells for Nat-FL and 97.1% cells for the NC, however on day 6, both sets of cells showed a decrease in culture viability to 90.6% for Nat-FL and 89.9% for NC.

For the CHO-S cells the four transfected constructs showed decreasing culture viabilities over the three-day period. For the Nat-S1 and NC transfected cells, a steady decline was observed over the three-day period dropping from 70% to 29.6% for the Nat-S1 and 77.8% and 35.4% for NC. However, for Nat-FL and OSP-FL transfected cells there was an increase in culture viability on day 5 at 86.2% for Nat-FL and 57.1% for OSP-FL which then dropped on day 6 to 65% for Nat-FL and 51.9% for OSP-FL.

Initially, by looking at the overall culture viability and viable cell concentrations for both CHO-S and CHO-K1 cell lines transfected with the Spike constructs, the CHO-S cells consistently had lower viable cells and lower culture viability over the days 4, 5 and 6 compared to the CHO-K1 cells whose culture viabilities remained in the 80-99% region and had viable cell concentrations of 5-15x10^6 cells/ml. These data indicate that the CHO-K1 cell host was more robust in term of cell numbers and culture viabilities but does not demonstrate how this reflects recombinant spike protein synthesis.

To examine the transient recombinant spike protein expression from both the cell lines, CHO-S and CHO-K1 cell lysate and supernatant samples were then run on western blots to compare the protein expression over days 4, 5 and 6. The results of the cell lysate blot of the four constructs comparing the CHO-S and CHO-K1 cell lines are outlined in Figure 3.4.5.
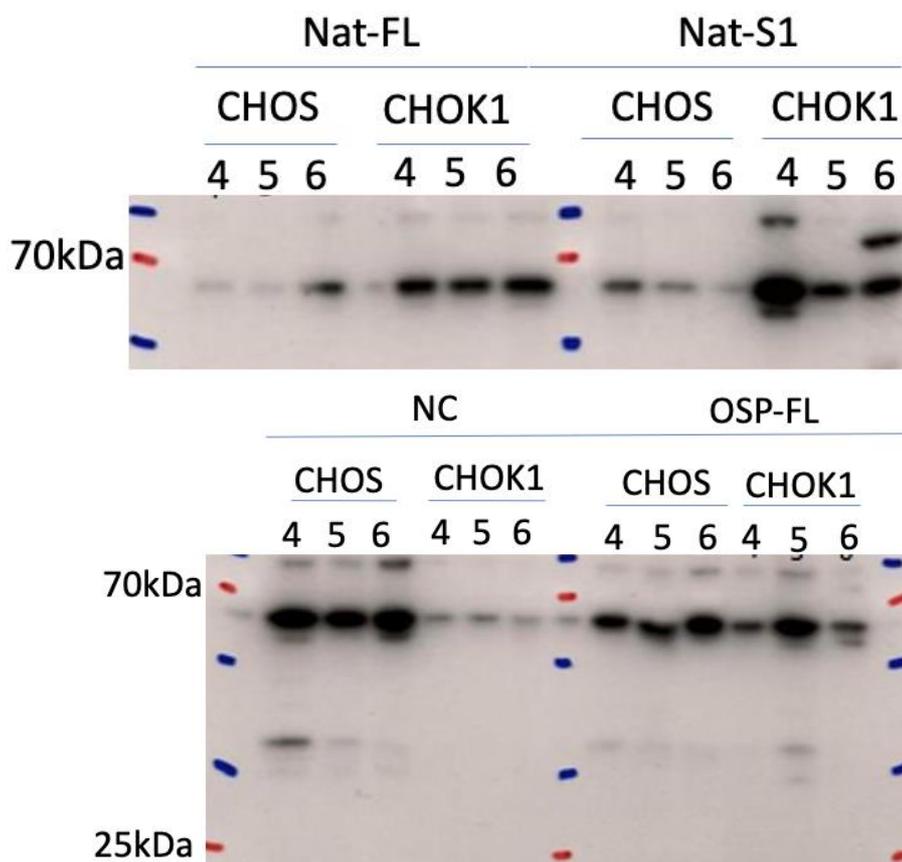
Figure 3.4.5: *Western blots displaying the results of reduced cell lysate protein samples taken from the transient transfection cultures on days 4,5 and 6 of the Nat-FL, Nat-S1, Nucleocapsid (NC) and OSP-FL when blotted with an anti-His Ab and developed for 2 minutes.*

For the CHO-S samples, there were protein bands around the 70 kDa region across days 4, 5 and 6 for all four samples Nat-FL, Nat-S1, NC and OSP-FL. The reason for this is unclear but one possibility is that the anti-His antibody is detecting a cellular protein. There were no band present at the expected size (130 kDa) for the Nat-FL CHO-S days 4, 5 and 6 samples suggesting this is not being expressed, is degraded or rapidly secreted out of the cell. Likewise, for the Nat-S1 CHO-S days 4, 5 and 6 samples there was strong bands in the 70 kDa region. The expected protein band is 75 kDa, and for the CHO-K1 samples, in the day 4 lane, there was a strong band in the 75 kDa region suggesting the expected Nat-S1 protein might be present however on day 5 there was no protein in this region and on day 6 there was yet another protein band above the common 70 kDa band.

In the NC CHO-S samples there were multiple bands present. The first were in the 100 kDa region with increasing density over days 4, 5 and 6. There were the previously described 70 kDa bands that are not NC specific and finally a band in the 40 kDa region on the day 4 sample which is the approximate expected size for the nucleocapsid protein. For the CHO-K1 samples, there were again faint protein bands present in the common 70 kDa region but no bands present in the expected 45 kDa region across days 4, 5 and 6. Finally, in the OSP-FL CHO-S samples, there was also the predominant 70 kDa band but also a band in the 100 kDa region. The expected protein band for this protein is 133 kDa, and therefore there is little evidence of transient expression intracellularly for this protein from these blots.

Overall, there was a common 70 kDa protein band across all four sample blots. The reasoning for this prominent band is unclear but the anti-His Ab is clearly detecting a cellular protein and not the protein(s) of interest as this 70 kDa band was present in all samples. There were no prominent bands in the expected protein band regions, indicating low level of protein expression in the cell lysate and that the protein is either not being expressed at sufficient levels for detection or has been translocated and secreted out of the cell. Blots of the supernatant from the transfections were therefore undertaken as shown in Figure 3.4.6.
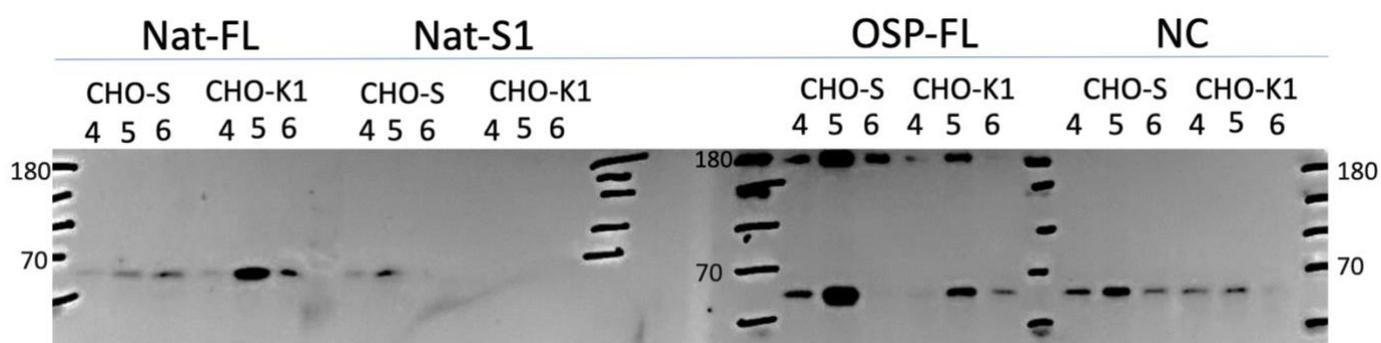
Figure 3.4.6. R*educed supernatant samples of the Nat-FL, Nat-S1, OSP-FL and NC transfected constructs on a western blot displaying protein expression over days 4, 5 and 6 post-transfection. Blots were developed for 5 minutes and probed with an anti-His antibody.*

It was expected that the proteins of interest would be present in the supernatant samples due to the protein being secreted out of the cell. The Nat-FL and OSP-FL was expected to be in the 130 kDa region, the Nat-S1 in 75 kDa and the Nucleocapsid (NC) in the 45 kDa region. For the Nat-FL construct in the CHO-S cell line, no protein was present in the 130 kDa region however there were once again faint protein bands around the 70 kDa region for days 4, 5 and 6 consistent with the intracellular blots. The CHO-K1 cell line showed a similar profile. For the Nat-S1 construct in the CHO-S cell lines, there was again a faint protein band present on days 4 and 5 in the 70 kDa region but there was no protein present in the day 6 lane whilst for the CHO-K1 cell line there were no protein bands present.

In the OSP-FL CHO-S cell line western blot, there were protein bands present in the 180 kDa range which could indicate presence of the OSP-FL spike protein. Others in the laboratory have seen the protein running at this size and this would agree with such results. However, in the CHO-K1 samples, there was only one faint band in the day 5 lane suggesting the CHO-K1 cells were not as effective in synthesising and secreting the protein. Finally, (Figure 3.4.6) the Nucleocapsid (NC) CHO-S western blot showed no protein in the expected region of around 45 kDa however there was the 70 kDa region band present.

Collectively these data show little evidence confirming expression of the target proteins either intra- or extracellularly. The only protein that there was some evidence of expression in the supernatant was for the CHO-S host with the OSP-FL (Full Length spike) in the

supernatant. Sequencing of the constructs showed these to be correct therefore the lack of expression suggests very low expression or rapid degradation of the target proteins. Nevertheless, as the CHO-S cell host did show some evidence of secretory protein production, and in an effort to limit the number of variables in experiments, it was decided to continue using the CHO-S host for further experiments.

3.2.2 Comparing the Effectiveness of the H7 Signal Peptide vs the Spike Native Signal Peptide for Secretory Spike Protein Production

As part of optimising the spike protein production process to be able to meet the demands of collaborators, the effects of different signal peptides were investigated to determine if this impact secretory protein amounts from the CHO-S host. In this experiment the two signal peptides being compared were the native signal peptide from the Spike protein and the reported H7 signal peptide. The H7 signal peptide has previously been used to express heavy and light chains of recombinant monoclonal antibodies and shown to result in high level secretion (45). To compare the effectiveness of the two signal peptides, both H7-FL (See Appendix 7) and Nat-FL constructs were transfected into CHO-S cells and protein samples taken on days 4, 5 and 6 post-transfection. The samples were then analysed by western blots to examine the protein expression from both the constructs. The cell lysate samples western blots are displayed in Figure 3.4.7.

Figure 3.4.7: *Reduced cell lysate samples run on a western blot showing protein expression from transfection of the H7-FL and Nat-FL constructs into CHO-S cells. Both samples probed with anti-His antibody and film developed for 5 minutes.*

Little protein was expected in the reduced cell lysate samples due to the nature of the samples where the majority of the protein should be secreted out of the cell and be present in the supernatant unless there was a block in the secretory pathway. If protein was to get retained in the cell the expected band would be around the 130 kDa area. There was no protein band in the expected 130 kDa region but instead in the 60-70 kDa region for both H7-FL and Nat-FL a prominent band was observed across the three-day period. This is similar to the band observed and described previously in section 3.2.1 and is likely a cellular protein. There was also a protein band in the 40 kDa region in the Day 6 lanes for both Nat-FL and H7-FL constructs. This could be due to a degradation product, but further work would be required to confirm this.

The reduced supernatant samples for the H7-FL and Nat-FL for days 4, 5 and 6 were run on a western blot and the results displayed in Figure 3.4.8. The expected result would be to see protein bands in the 130 kDa region for both the H7-FL and Nat-FL across days 4, 5 and 6. Instead, this was not the case and there was only a band in the 70 kDa region once again for the H7-FL day 4 sample. This is not the spike protein of H7-FL due to the size of the expected band running higher at 130 kDa. There was therefore no evidence of Spike protein expression using either the native or H7 signal peptides in the samples analysed.
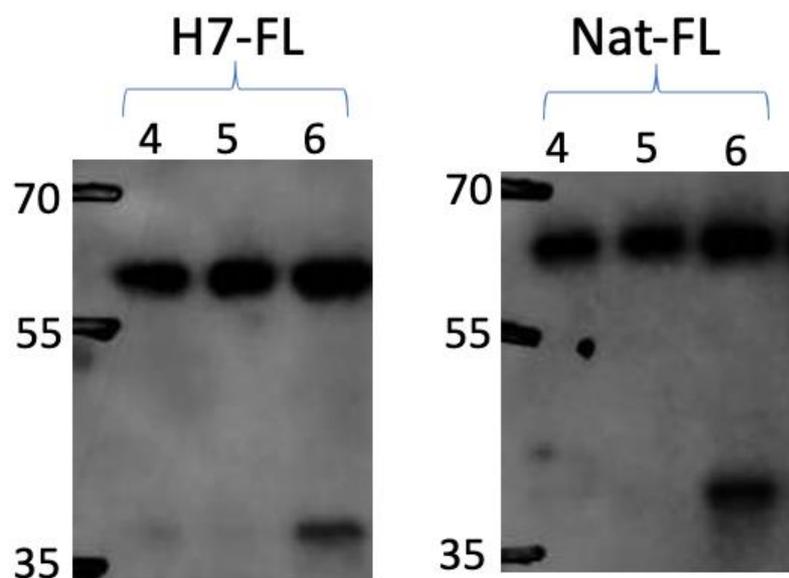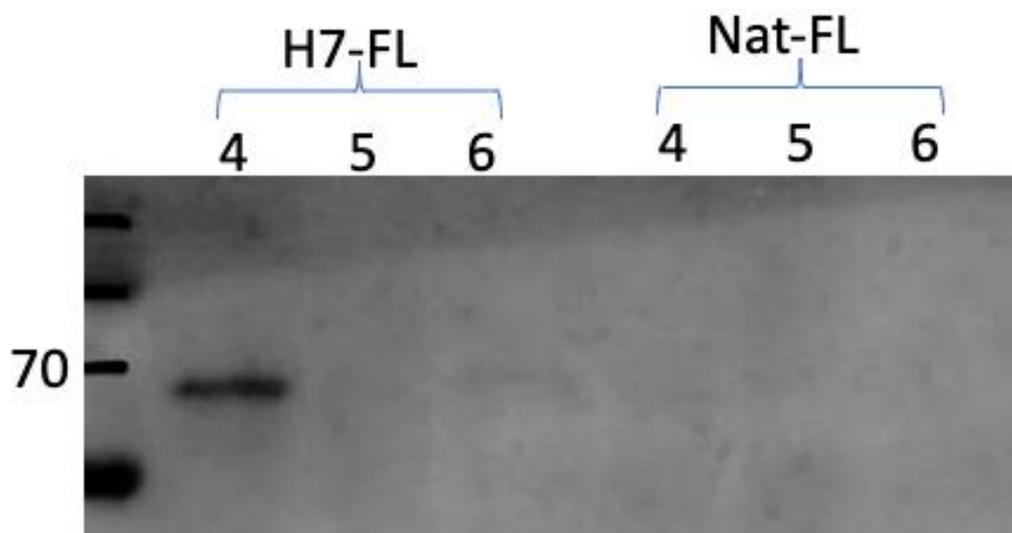

Figure 3.4.8: *Reduced supernatant samples run on a western blot showing protein expression from transfection of the H7-FL and Nat-FL constructs into CHO-S cells. Both samples were probed with anti-His antibody and film developed for 15 minutes.*
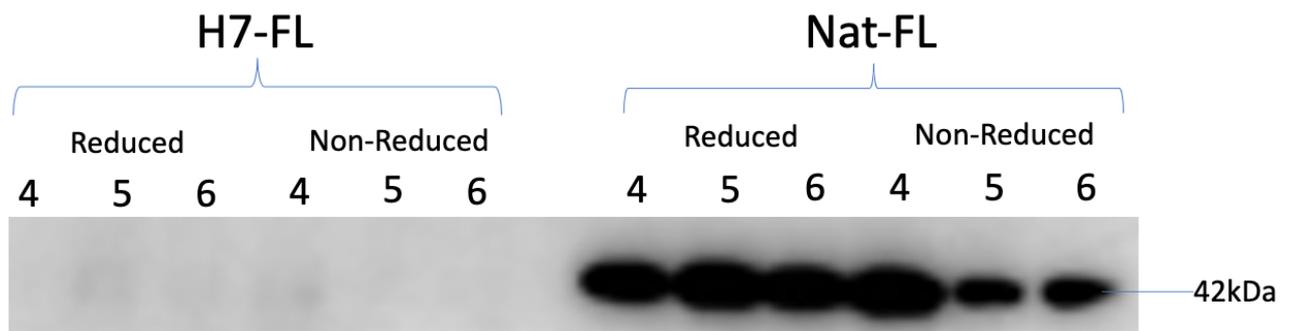
Figure 3.4.9: *Western blot showing the intracellular reduced and non-reduced β-actin loading controls for the H7-FL and Nat-FL constructs transfected into CHO-S cells with samples taken on days 4,5 and 6. Samples probed with Anti-His Ab with film developed for 1 minute.*

A β-actin loading control was also carried out on the intracellular H7-FL and Nat-FL construct transfected CHO-S cells to check the loading of material. As shown in Figure 3.4.9, the western blot shows the Nat-FL cells had a strong β-actin band present in the 42 kDa region for days 4, 5 and 6 however for the H7-FL cells there was a very faint (reduced) or no (non-reduced) β-actin present indicating an issue with the H7-FL transfected cells or samples. Overall, the result of this experiment did not allow a conclusion to be drawn over the comparison of the two signal peptides as secreted full-length spike protein was not detected from either construct.

3.2.3 Stable cell pool development of various SARS-CoV-2 constructs transfected into CHO-S cells

In order to improve the spike protein production process, stable cell pools were constructed to continuously produce the various SARS-CoV-2 Spike proteins instead of transiently transfecting the DNA constructs into the CHO cells and harvesting on day 6. The four DNA constructs that were transfected into CHO-S cells to make stable cell pools were the Nat-FL, H7-FL, OSP-FL and OSP-RBD. These constructs were all in the GS backbone so that GS expression in the presence of MSX could be used as a selection agent. As well as generating stable cell pools for these DNA constructs, another variable was added to the experiment whereby the levels of MSX were varied in order to compare whether varying the concentration affected the recombinant Spike protein production. For each DNA construct, CHO-S stable cell pools were constructed, one with 25 µM MSX and the other with 50 µM

MSX in line with the concentrations regularly used in the laboratory at Kent. To be able to evaluate these stable cell pools, the culture viability and the viable cells were monitored during batch culture to generate growth curves. To determine the amount of recombinant spike protein produced, samples were taken on day 6 and then also at the end of the batch culture (EOC) with the supernatant and cell lysate expression investigated. The Spike transcript mRNA expression in these stable cell pools were also investigated using qRT-PCR to quantify the relative mRNA amounts to confirm that the transcript was being made and enable the comparison across the different stable cell pools.

The viability of the different stable cell pools (Figure 3.5.1) shows the cell pools maintained high culture viability until 168-hour (day 7) whereby the culture viability declined. For the Nat-FL(50) pool there was an aberrant pint at the 48-hour mark where viability was around 45% but this was out of line with other measurements. When comparing the 25 μM MSX cell pools to the 50 μM MSX cell pools, the culture viability declined quicker with the 25 μM MSX, dropping to under 20% viability by 216 hours compared to the 50 μM MSX cell pools which dropped to under 20% culture viability around the 264-hour mark. Overall, the 50 μM MSX stable cell pools maintained a higher level of culture viability for a longer period of time. To note, at the end of the growth curve, the 25 μM MSK stable cell pool cultures had a large presence of cell clumping compared to those with 50 μM MSX, which could have impaired the readings of both the culture viability and viable cells. Each stable cell pool viability and viable cell growth curve result is a single culture and not an average of triplicates.

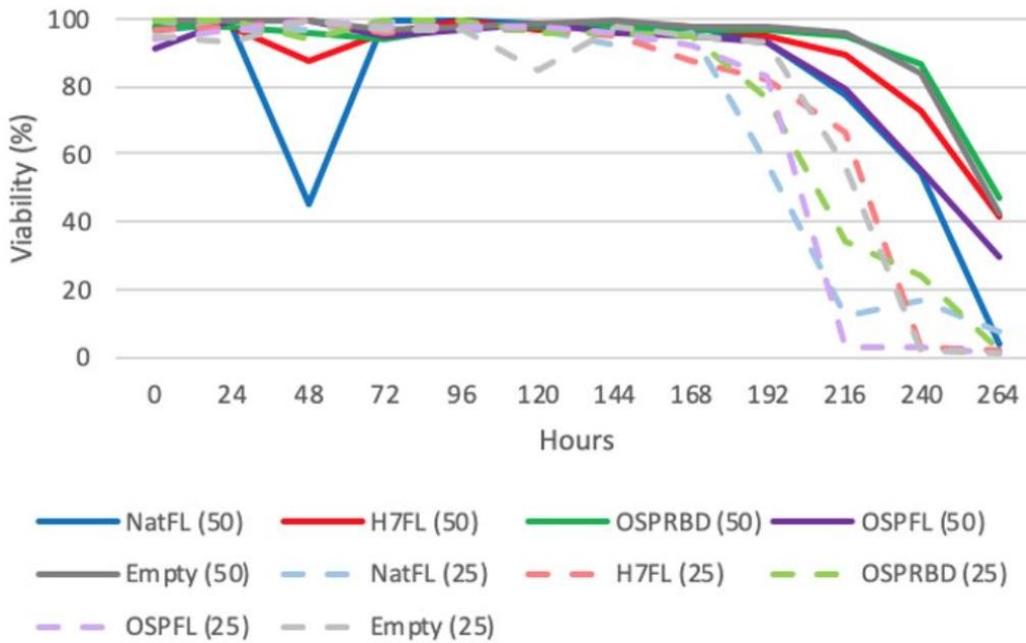**Figure 3.5.1:** *Growth curve outlining the culture viability (%) of the CHO-S stable cell pools with either 50 μM MSX (block lines) or 25 μM MSX (dotted lines). Cells were counted over an 11-day period. N=1.*
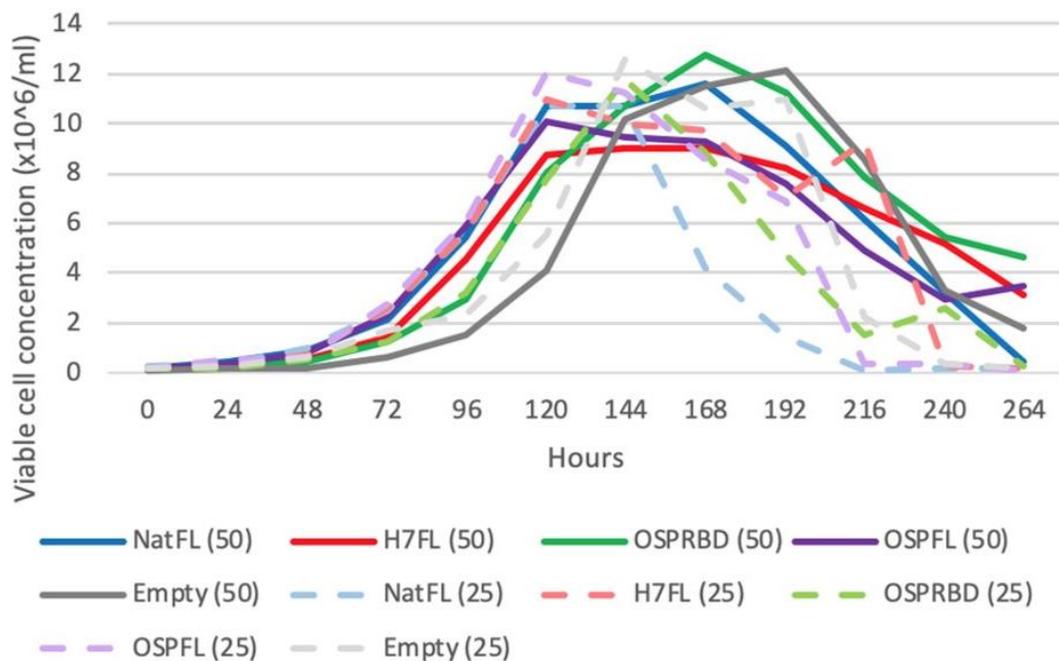
Figure 3.5.2: *Growth curve outlining the viable cells (x10⁶ cells/ml) of the CHO-S stable cell pools with either 50 μM MSX (block lines) or 25 μM MSX (dotted lines). Cells were counted over an 11-day period. N = 1.*

When considering the viable cell profiles (Figure 3.5.2) all the stable cell pools increased in the same pattern from 0 hours to 120 hours where at this point the 25 μM MSX stable cell pools started to drop and the viable cell count declined quicker to under 2×10^6 cells/ml by the 216 hour mark except for the H7FL(25) cell line which dropped to under 2×10^6 cells/ml by the 240 hour mark (day 10). The 50 μM MSX stable cell pools maintained their higher number of viable cells for much longer and started to decline later around the 168-hour mark, compared to the 120-hour mark for the 25 μM MSX stable cell pools. The OSPRBD(50) stable cell pool had the highest viable cell 168-hour region with over 12×10^6 cells/ml before they started to decline down to around 5×10^6 cells/ml by the 264-hour mark.

Overall, the results of the growth curves for the culture viability (Figure 3.5.1) and viable cells (Figure 3.5.2) indicate the 50 μM MSX stable cell pools to have better growth than the 25 μM MSX as they maintained higher culture viabilities for longer and achieved a higher level of viable cells for a longer period of time. However, this data gives no information on if any, and how much, recombinant Spike protein was produced from the different cell pools.

3.2.4 Investigating recombinant spike protein expression on Day 6 and at the end of culture (EOC) from the different stable pools

During the 11-day growth period of the cells, samples were taken on day 6 and at the end of culture (EOC) to run on western blots to investigate the recombinant Spike protein expression. Before the western blot was probed for the presence of Spike protein, a ponceau stain was applied to the blot to view any protein bands that may have transferred onto the nitrocellulose. The results of the ponceau stain blots are displayed in Figure 3.5.3.

The results of the ponceau stained blots (Figure 3.5.3) show the presence of protein on all of the blots. Both the Day 6 (Lysate) blots indicate a lot of protein present on the blot as would be expected showing successful transfer of protein. For the supernatants, the EOC (SP) blot showed strong bands whereas the Day 6 (SP) blot did not appear to have as much protein present on the blot as would be expected for a secretory blot earlier in culture, these proteins secreted accumulate over time. The stain is not protein specific and does not determine

presence of the different SARS-CoV-2 spike proteins but gives an indication as to the presence of protein after the transfer.
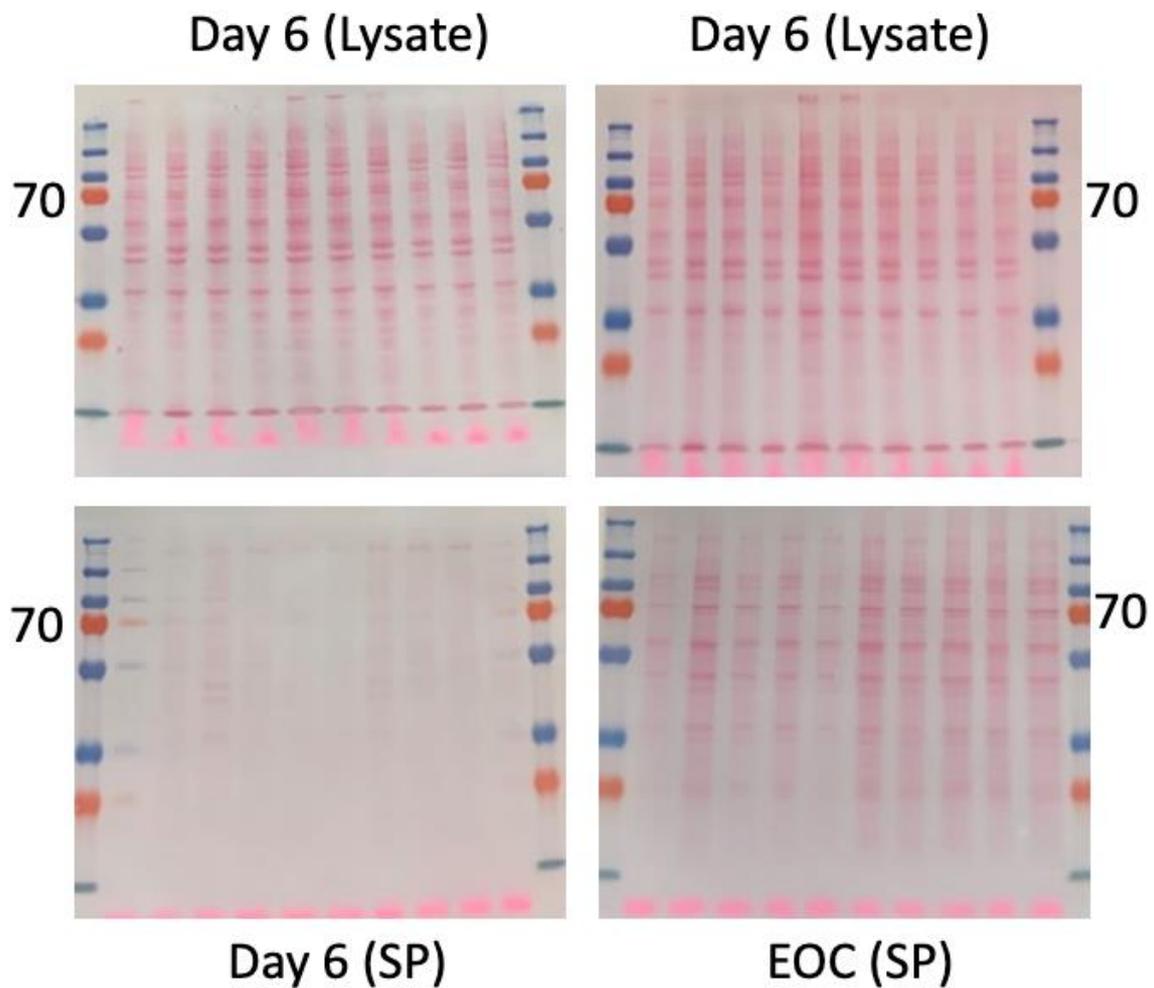
Figure 3.5.3: *Ponceau stain of nitrocellulose blot which highlights all protein bands present. Two Day 6 (Cell Lysate) blots are present with a Day 6 (Supernatant, SP) and EOC (End of Culture) (Supernatant, SP) from the stable cell pools.*

To determine the Spike protein expression and if the expected protein bands were present, western blots were developed for the supernatant Day 6 , End of Culture Day 6 and cell lysate Day 6 samples for all 10 samples (50 μM MSX and 25 μM MSX) using an anti-His antibody. The results are outlined in Figures 3.5.4, 3.5.5, 3.5.6 and 3.5.7.
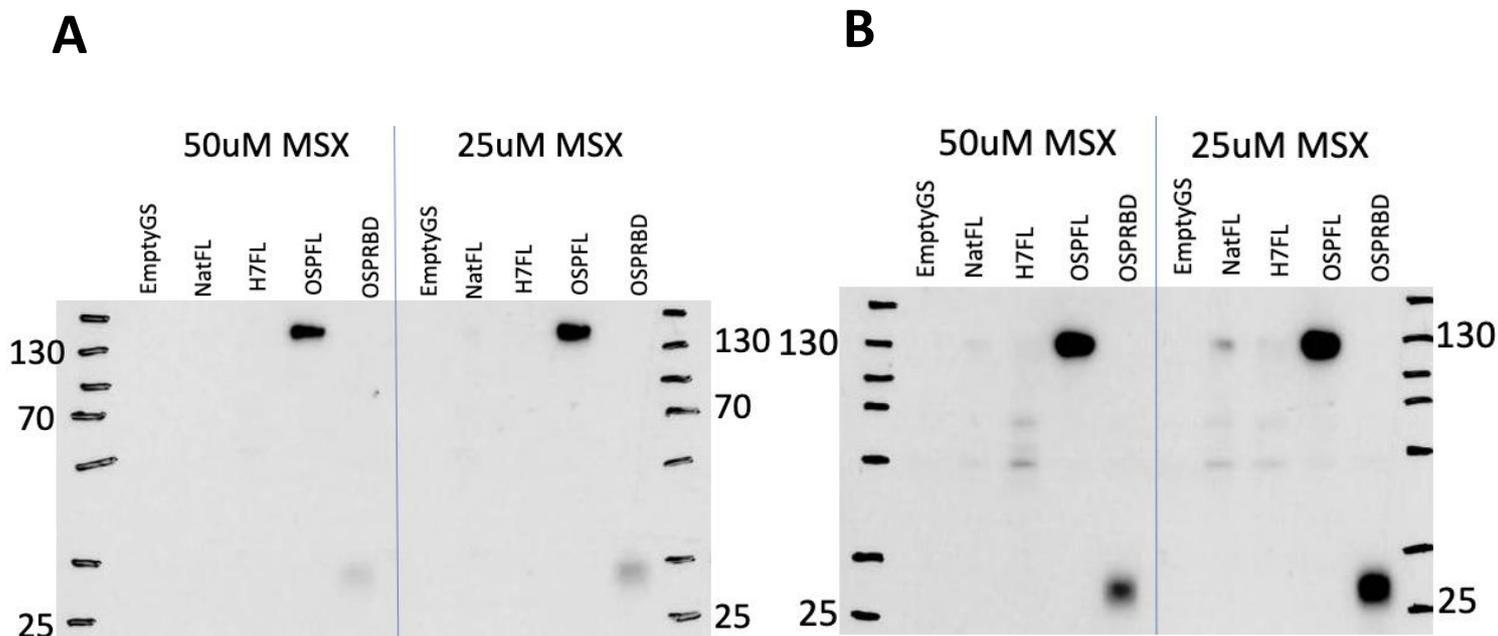
Figure 3.5.4: *Western blot displaying supernatant Day 6 protein sample expression for the EmptyGS, Nat-FL, H7-FL, OSPFL, OSPRBD with either 50 μM or 25 μM MSX. **(A)** Supernatant Day 6 blot developed for 7 seconds and probed with anti-His Ab. **(B)** Supernatant Day 6 blot developed for 20 seconds and probed with anti-His Ab.*

The results displayed in Figure 3.5.4 for the supernatant Day 6 samples indicate strong Spike protein expression for the OSP-FL and OSP-RBD samples in both 50 μM MSX and 25 μM MSX stable cell pools in the expected band region of around 130kDa and 30 kDa respectively. It is noted that these proteins are made using the native ER signal peptide which is not CHO cell codon optimised. When comparing the 50 μM MSX OSP-FL stable cell pool to the 25 μM MSX OSP-FL stable cell pool, visually there appears to be more protein in the 25 μM MSX OSP-FL (clearer in the 7 second (A) developed blot compared to the 20 second developed (B) blot). For the OSPRBD, not only was the expected band of around 30kDa present, but for the 25 μM MSX stable cell pool the protein expression was stronger compared to the 50 μM MSX OS-RBD stable cell pool (Figure 3.5.4 (B)). This indicates that the 25 μM MSX stable cell pools may produce more protein that the 50 μM MSX generated pools even though the growth profiles were not as good in these pools, perhaps due to the load imparted on the cell by a higher Spike protein production.

For the H7-FL, which utilises the H7 ER signal peptide, no protein bands of the correct size were present in supernatant day 6 protein samples. This could indicate an issue with the H7 signal peptide not correctly translocating the protein. For the Nat-FL stable cell lines, the 25 µM MSX stable cell line, as displayed in Figure 3.5.4 (B), a very faint band in the expected region of around 130 kDa was observed indicating the Native signal peptide that is CHO codon optimised is somewhat effective but very low levels of protein are being translocated. There was no presence of the Nat-FL in the 50 µM stable cell pools. For the EmptyGS 50 µM MSX and 25 µM MSX stable cell pools there was no protein bands as expected due to the vector containing no Spike gene.

End of culture supernatant spike expression
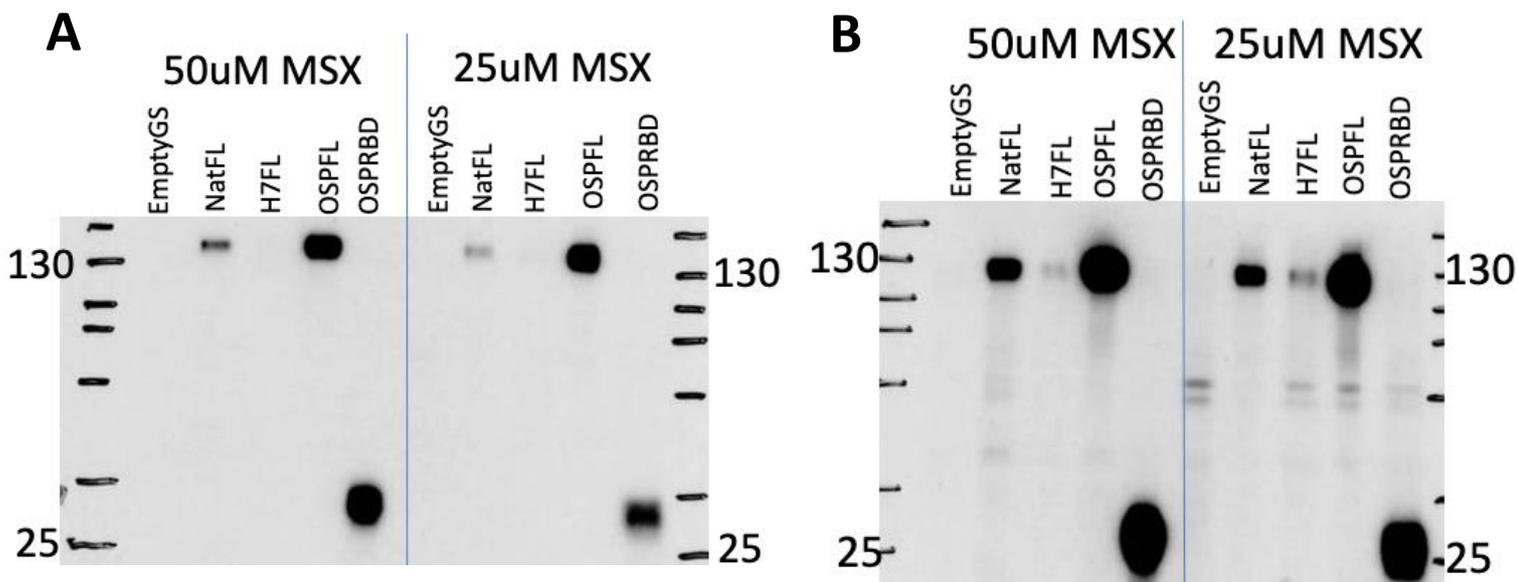
Figure 3.5.5: *Western blots displaying End of Culture (day 11) Spike protein expression analysis for the EmptyGS, Nat-FL, H7-FL, OSPFL, OSPRBD stable cell pools with either 50 µM or 25 µM MSX.* **(A)** *Supernatant End of Culture blot developed for 7 seconds and probed with Anti-His Ab.* **(B)** *Supernatant End of Culture blot developed for 20 seconds and probed with Anti-His Ab.*

End of Culture samples were also taken (Day 11) to compare to the Day 6 supernatant samples. The results (Figure 3.5.5) indicate strong protein expression of the expected protein band around 130 kDa in the OSPFL and 30kDa for the OSPRBD stable cell pools for both 50 µM MSX and 25 µM MSX pools. Although it is not possible to directly compare expression across different blots, for the OSPFL and OSPRBD proteins there is strong expression in the

end of culture supernatant samples (Figure 3.5.5) compared to the supernatant day 6 (Figure 3.5.5) samples.

For the Nat-FL 50 μM and 25 μM MSX stable cell pools, there was protein present in the expected protein band region of around 130 kDa. Compared to the day 6 supernatant samples (as displayed in Figure 3.5.4), there appear to be stronger protein bands with the 50 μM MSX Nat-FL stable cell pool displaying a more intense protein band than the 25 μM MSX stable cell pool. In Figure 3.5.5 (B), there was a faint protein band present for the H7-FL 50 μM MSX and 25 μM MSX stable cell pools in the expected protein band range of around 130 kDa. The 25 μM MSX H7-FL stable cell pool had a more intense protein band indicating higher levels of protein at the lower concentration of MSX. In the EmptyGS lanes for both 50 μM MSX and 25 μM MSX, there were no protein bands as expected.

Overall, Figure 3.5.5 analysis of Spike protein at the end of culture supernatant in the stable cell pool samples show that the OSP-FL and OSP-RBD protein bands are strong and that the OSP signal peptide is effective in translocating the protein. While there were faint bands for the Nat-FL and H7-FL stable cell lines, the data indicates lower levels of protein expression.

Cell lysate day 6 Spike Protein Expression
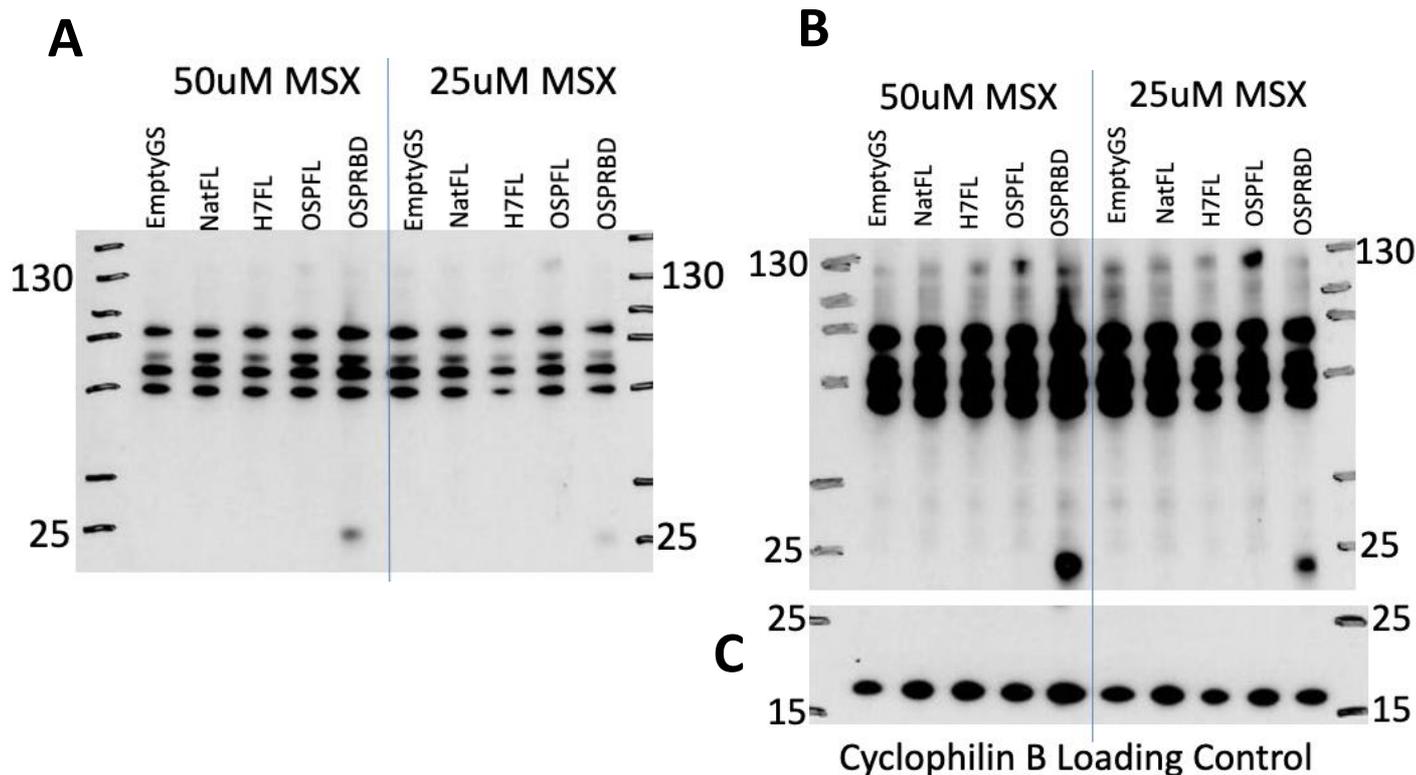
Figure 3.5.6: *Western blots investigating Spike protein expression in the cell lysate Day 6 samples with either 25 μM or 50 μM MSX for the various SARS-CoV-2 spike constructs including Nat-FL,H7-FL,OSP-FL, OSP-RBD and an Empty GS.* ***(A)*** *Western blot showing result of cell lysate Day 6 protein samples, film developed for 7 seconds and probed with Anti-His Ab.* ***(B)*** *Western blot showing results of cell Lysate Day 6 protein samples with film developed for 25 seconds and probed with Anti-His Ab.* ***(C)*** *Western blot showing loading control for cell lysate Day 6 samples probed with anti-cyclophilin B.*

The cell lysate samples from day 6 were also probed for Spike protein expression (Figure 3.5.6). There were ¾ bands observed in all lysate samples, including the empty GS where no Spike should be present, again confirming that bands around 70 kDa detected using the anti-His antibody detect cellular proteins. Only at the longer exposure time could any protein around 130 kDa be detected for the full-length spike pools whilst for the RBD pools a band around 25 kDa could be observed in both blots. The loading control of cyclophilin B, shown in Figure 3.5.6 (C), confirms that the protein loads across the samples are approximately equally loaded. Overall, the results of the western blot confirm that the amounts of intracellular detected protein appear to correlate with the amount of protein detected extracellularly in the supernatant at Day 6 of culture, Figure 3.5.4.

GS protein expression in day 6 samples
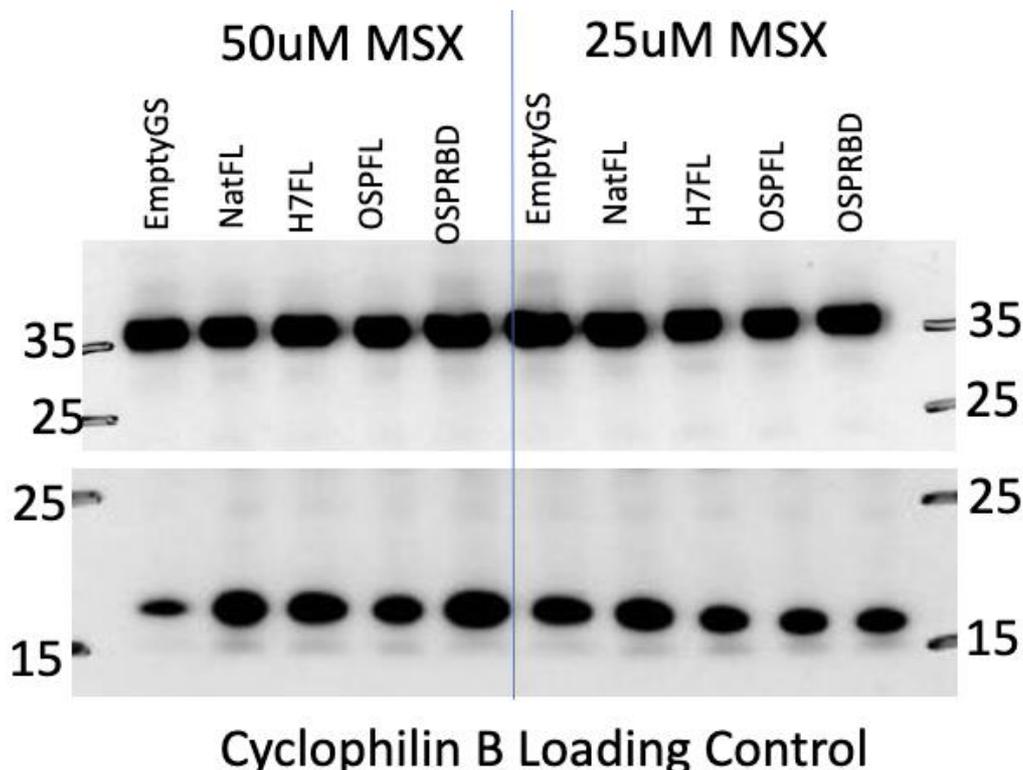
Figure 3.5.7: *Western blot of the GS expression in the cell lysate Day 6 samples(upper blot) for both 50 μM MSX and 25 μM MSX generated stable cell pools. The film was developed for 1 second and probed with an anti-GS Ab.*

As the stable cell pools were generated using GS selection in the presence of MSX, and the DNA constructs contained mGS within them so that the CHO-S cells could express exogenous GS to overcome the addition of MSX. It was therefore essential to check the expression of GS across the various stable cell pools with both 50 μM and 25 μM MSX. Figure 3.5.7 reports the results of a cell lysate day 6 western blot for the expression of the GS protein in which the stable cell pools have been probed with an anti-GS Ab. As shown in Figure 3.5.7, there were clear protein bands across all 10 stable cell pools (50 μM and 25 μM MSX) at the expected band size of 35 kDa for GS. This shows high levels of GS expression with little to no difference when comparing the 50 μM or 25 μM concentrations of MSX. A Cyclophilin B loading control was also run for all samples and this showed the wells had been loaded equally. In order to

further confirm exogenous GS was being expressed these samples were then run on a blot alongside a sample from a CHO-S non-transfected control sample (Figure 3.5.9).
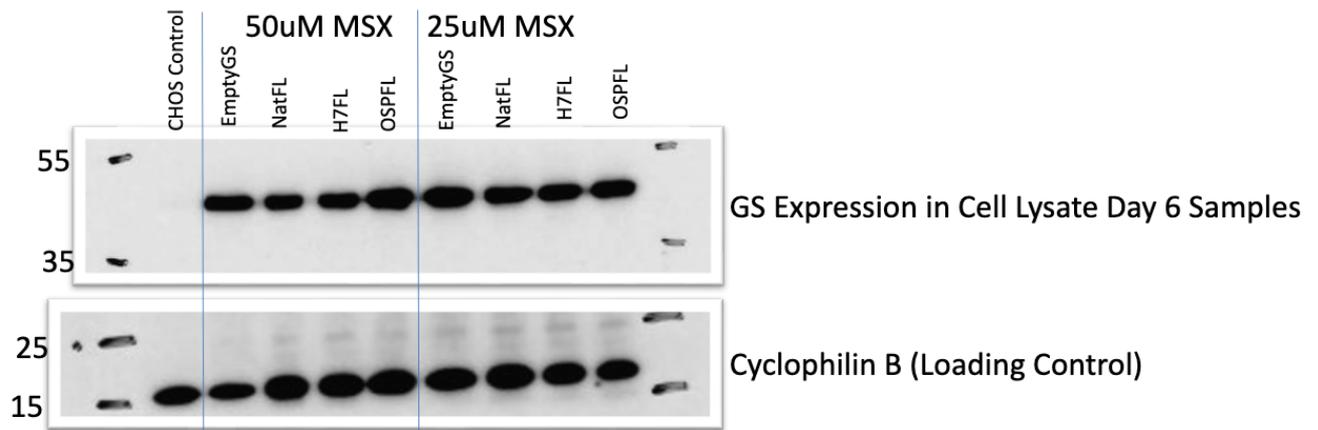
Figure 3.5.8: *Western blot comparing GS expression from the transfected CHO-S 50 μM and 25 μM MSX samples to a CHO-S control containing no transfected material to measure the over-expression of the GS in the transfected cells. The film was developed for 1 second and probed with Anti-GS Ab. A loading control for Cyclophilin B was used on the CHO-S stable cell pools with cell lysate day 6 GS samples.*

The non-transfected GS control lysate was therefore run alongside cell lysate day 6 samples that had been probed with anti-GS Ab (Sigma G2781-2ML) to check the expression of the GS was from the transfected cells and not endogenous from the CHO-S cells themselves. The results of Figure 3.5.8 show that the GS expression in the cell pools is from the transfected CHO-S as the GS expression was only detectable in the stable cell pools transfected with the DNA constructs containing the mGS gene. The CHO-S cell control lysate had no detectable GS expression. A Cyclophilin B loading control was also used to check the protein had been loaded equally (Figure 3.5.8).

Overall, by evaluating the protein expression of the 10 stable cell lines that had been constructed via western blots, the results (as shown in Figures 3.5.4, 3.5.5 and 3.5.6) indicated that the OSP signal peptide was much more effective in producing secreted recombinant Spike protein. The native and H7 signal peptide constructs showed only very faint band indicating issues with protein expression.

3.2.5 Analysis of spike transcript amounts in the different cell pools

To confirm that the differences in Spike protein expression between constructs was not due to differences in transcript amounts, qRT-PCR analysis was undertaken. RNA samples were taken on day 3 of the batch cultures and subsequently used for determining the relative amounts of transcript between samples. The results are reported in Figure 3.5.9 and shows that there are differences in the transcript amounts from the different Spike constructs. As shown in Figure 3.5.9, the OSPFL 25 μM and 50 μM MSX stable cell pools had the highest relative levels of mRNA expression with more in the 25 than the 50 μM MSX samples. This is concurrent with the protein expression displayed in the western blots (Figure 3.5.4 and 3.5.5) as they show stronger protein bands in the 25 μM MSX OSP-FL stable cell pools compared to the 50 μM MSX OSP-FL stable cell pool. This indicates that it may not be the OSP signal peptide itself that drives higher protein expression but that the presence of the nucleotide sequence that encodes for this either has a stabilising impact on the transcript so that more transcript is present or impacts transcription such that greater transcript amounts are made.
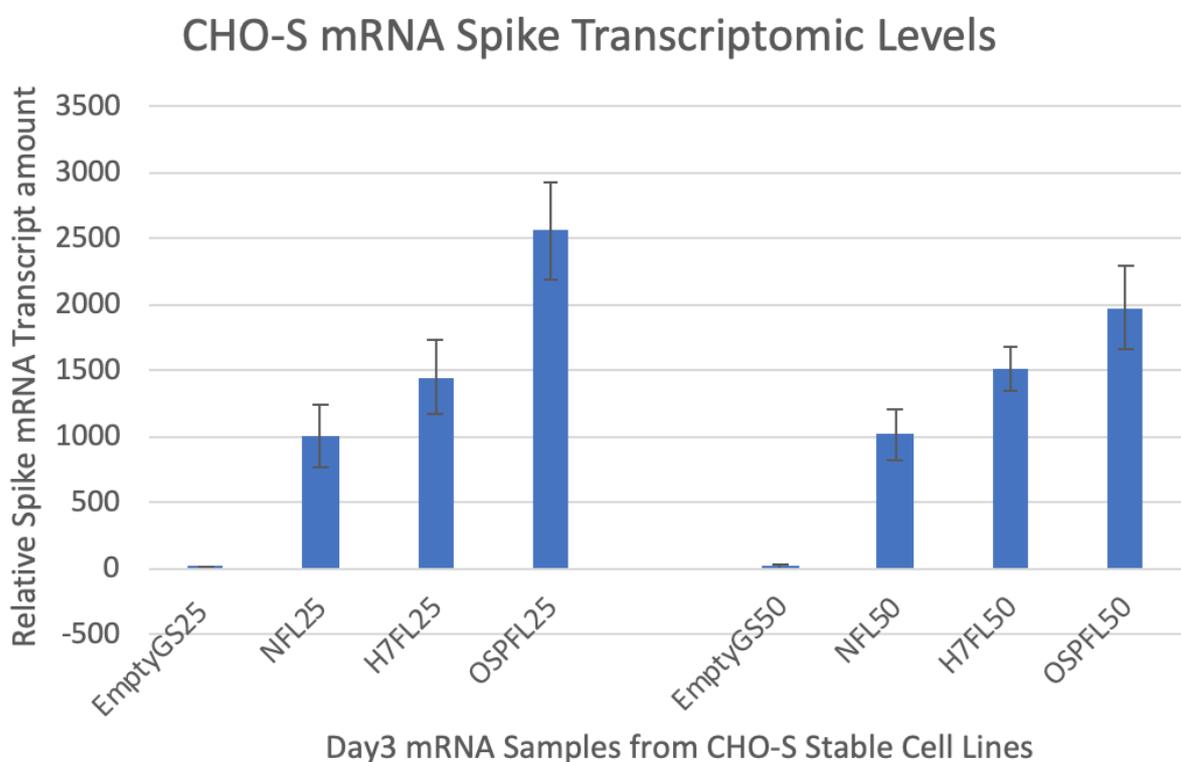
Figure 3.5.9: *Results of qRT-PCR analysis of the relative levels of Spike mRNA expression of the EmptyGS, Nat-FL (NFL), H7-FL and OSP-FL transcripts using Day 3 RNA samples for both 50 μM and 25 μM MSX CHO-S Stable cell pools. Data was normalised against β-actin.*

For the Nat-FL and H7-FL transcripts there was no difference in the amounts between the 50 μM and 25 μM MSX stable cell pools. This indicates the concentration of MSX did not influence the mRNA levels for the Nat-FL or H7-FL pools. The Nat-FL had the lowest levels of mRNA expression out of the four constructs measured. The lower mRNA expression levels will lead to lower protein expression and thus partially explains the results of the western blots (as shown in Figures 3.5.4 and 3.5.5) in which little to no protein was seen from the Nat-FL and H7-FL pools. The EmptyGS cell lines (both 50 μM and 25 μM MSX) had no spike mRNA transcript expression as expected.

Overall, the results of the qRT-PCR show that the OSP-FL sequence results in higher mRNA amounts than form the other constructs. By comparing the results of the qRT-PCR (Figure 3.5.9) to the western blots (Figures 3.5.4, 3.5.5 and 3.5.6), there was a relationship between the two. However, the signal peptide itself must still play some role as the decrease in transcript between the OSPFL and other constructs at the transcript level was not as great as the decrease in protein expression observed visually by western blot. Further, the concentration of the MSX in developing the stable pools also impacted the amount of OSPFL transcript observed.

3.3 Measuring antibody response of various Spike protein variants against human serum samples containing SARS-CoV-2 antibodies via ELISA

To measure antibody responses to various spike protein variants, an ELISA (see section 2.5.1) was run with two viral inactivated serum samples, one with Wuhan antibodies (patient infected with Wuhan strain) and one with B.1.1.7 antibodies measured against Nat-FL spike (Wuhan), B.1.1.7 spike and B.1.351 spike.

3.3.1 Wuhan patient serum against various spike variants

Figure 3.6.1: *(Wuhan LV094) Serum tested against Wuhan (Nat-FL), B.1.1.7 and B.1.351 spike protein showing antibody response with absorbance measured at 450 nm.*

The results of the spike protein variants measured against the Wuhan Serum Sample (LV094) are outlined in Figure 3.6.1. As the data indicates, there is the presence of a concentration – response relationship between the Wuhan Serum samples and the various spike protein variants with a higher absorbance signal observed for the lowest dilution factor, which is the most concentrated dilution. The B.1.1.7 OSP-FL spike protein variant had the overall highest

signal compared to the Native Wuhan Spike protein and the B.1.351 spike protein samples except for the 1:1000 dilution which had the lowest signal. Overall, there was no major difference between the responses of the different spike protein variants when measuring antibody responses of the Wuhan serum sample.

B.1.1.7 Sample Serum (Sondus 5730) tested against SARS-CoV-2 Spike Protein Variants
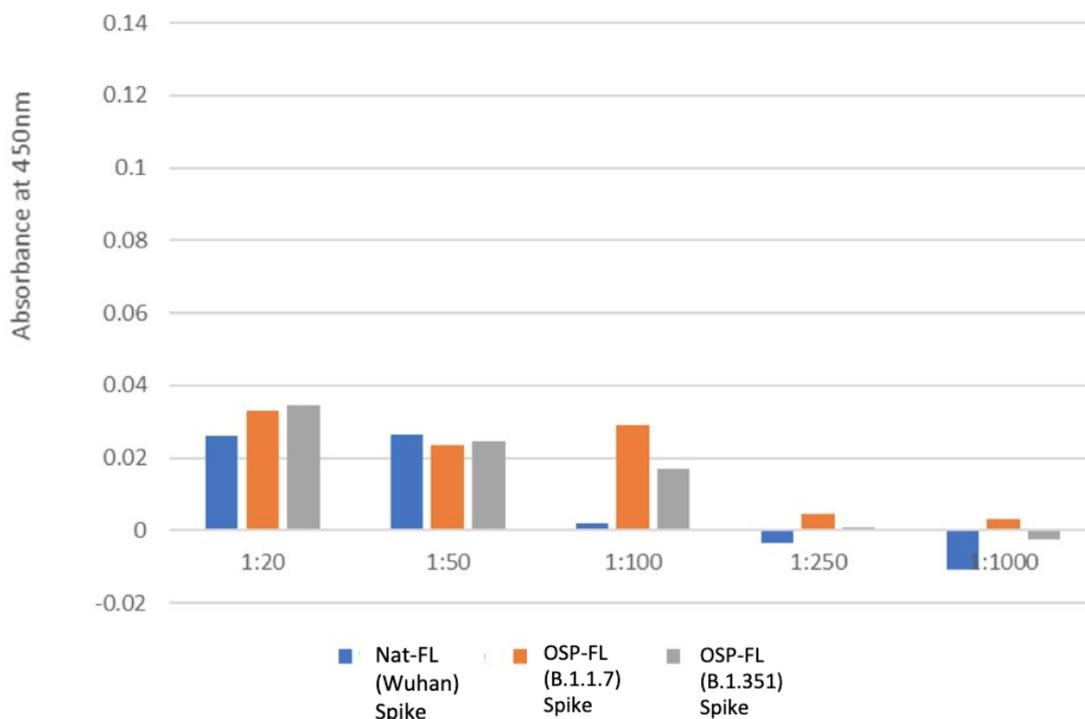
Figure 3.6.2: *B.1.1.7 Sample Serum (Sondus 5730) tested against Wuhan (Nat-FL), B.1.1.7 and B.1.351 spike protein showing antibody response with absorbance measured at 450nm.*

3.3.2 B.1.1.7 serum against various Spike variants

The next serum sample to be measured is the B.1.1.7 serum sample against the various spike protein variants. As displayed in Figure 3.6.2, there is little binding of any of the serum antibodies to the different spike protein variants in the sample. The most concentrated dilution of 1:20, the signals for the Nat-FL (Wuhan) spike, B.1.1.7 and B.1.351 spike protein variants reached a max of around 0.03. Overall, there was no observed relationship between the B.1.1.7 serum sample antibodies and the different spike protein variants.

Anti-Spike Antibody vs SARS-CoV-2 Spike Protein Variants

Legend:
- Nat-FL (Wuhan) Spike
- OSP-FL (B.1.1.7) Spike
- OSP-FL (B.1.351) Spike

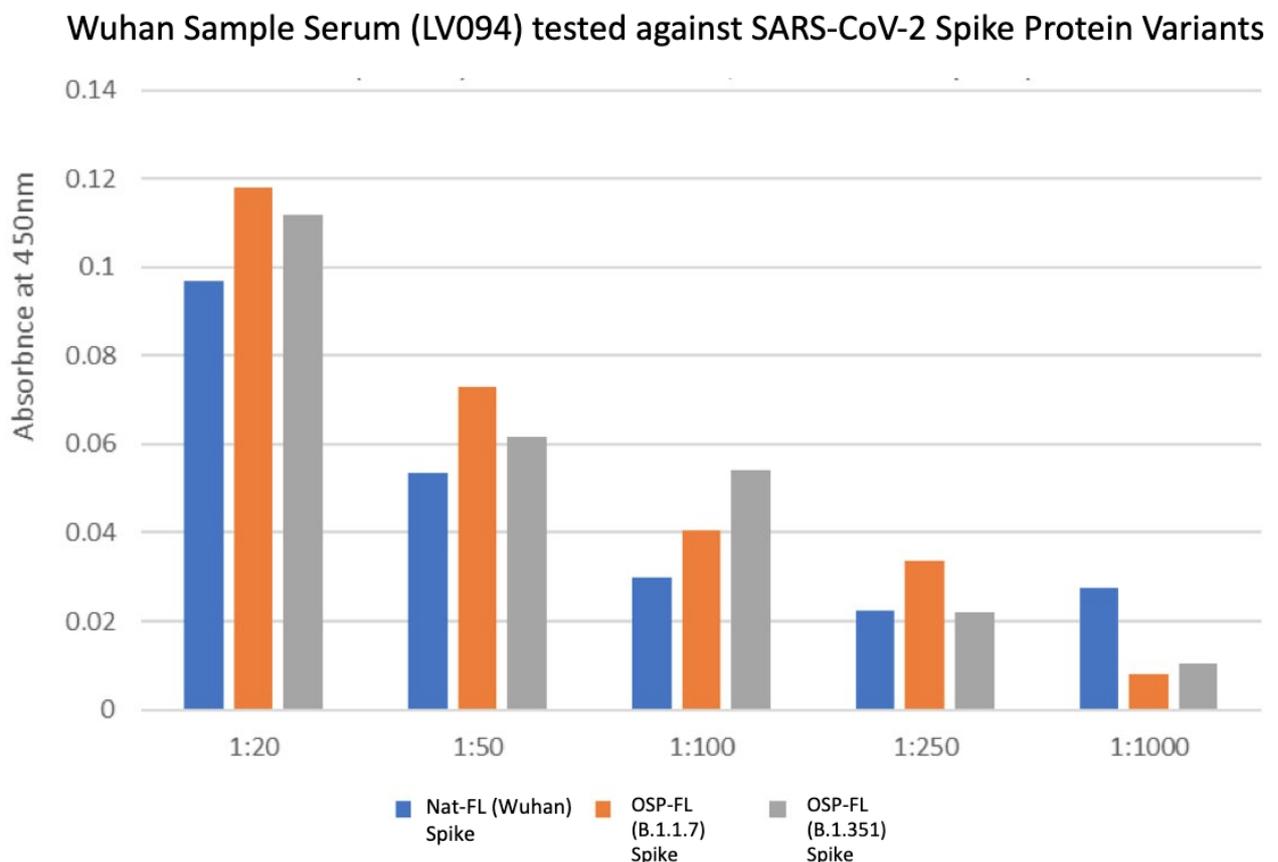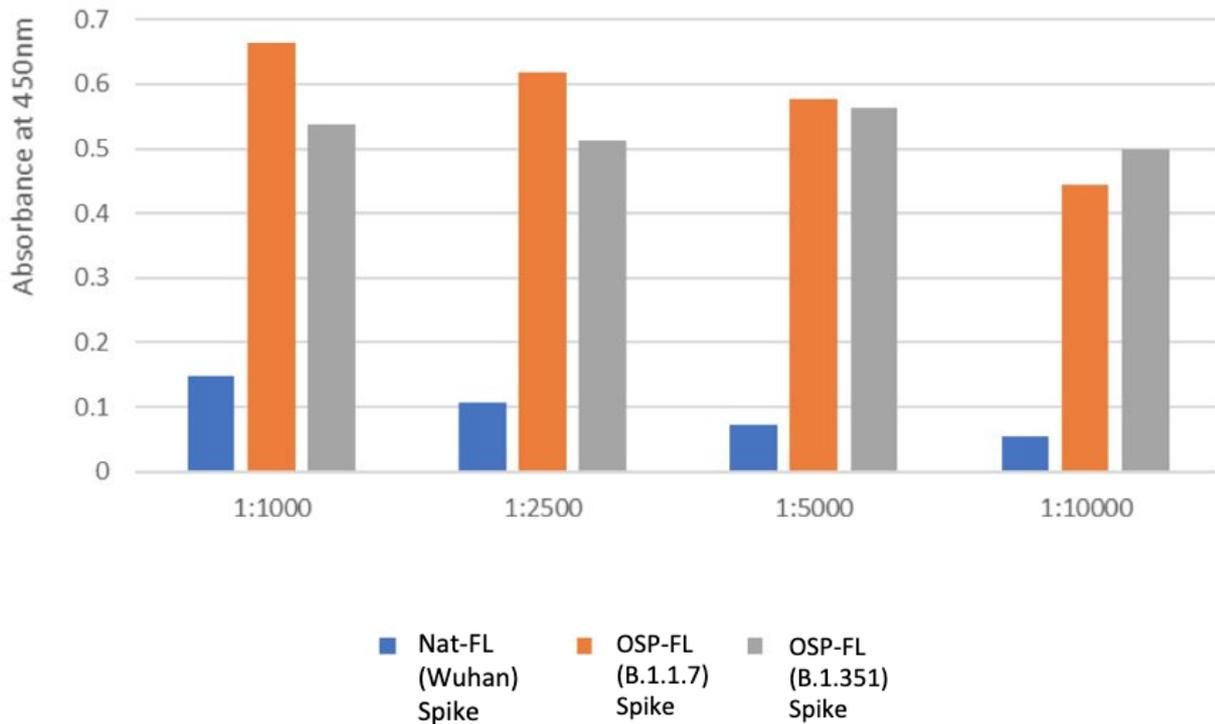**Figure 3.6.3:** *Anti-Spike Antibody vs Wuhan (Nat-FL), B.1.1.7 and B.1.351 spike protein showing antibody response with absorbance measured at 450nm.*

3.3.3 Anti-Spike antibody against various Spike variants

A positive control using anti-Spike antibody against the Nat-FL (Wuhan) , B.1.1.7 and B.1.351 spike variants for the ELSA was also carried out. The results displayed in Figure 3.6.3 show a differential response of the binding of the anti-spike antibody compared to the different spike protein variants. There was a signal of around 0.65 for the B.1.1.7 spike at the most concentrated dilution for 1:1000 and around 0.5 for the B.1.351 spike variant. For the Nat-FL (Wuhan) Spike protein however, the immunoreactivity was much lower with a signal around 0.1. This result was unexpected as the antibody was raised against the original Wuhan strain spike protein.

Overall, looking at the results from Figures 3.6.1 and 3.6.2., the immunoreactivity of the serum samples against the various spike variants gave a low signal. An explanation for this could be due to the secondary antibody that had been used as this was an anti-human IgG (Ɣ chain specific) antibody. Previous experiments have shown signals with gamma chain antibodies were much lower compared to whole molecule anti-human antibody.

Chapter 4

Discussion

4.1 DNA Subcloning

4.1.1 Cloning of the Spike gene into commercial vectors

The overall aim of this project was to generate constructs for the expression of the SARS-CoV-2 spike and variant thereof and investigate the different ways the spike protein production process could be optimised. To do this, the DNA constructs first had to be constructed using the process of subcloning. The various Spike constructs were ligated into commercial vectors such as pcDNA3.1hygro and pcDNA3.1mGS and transformed into DH5alpha *E. coli*. The construction of these plasmids was successful and Maxipreps were able to be produced that were high in concentration and purity. Initially, the Native FL and S1 spike genes were the first DNA plasmids to be successfully constructed and ligated into pcDNA3.1hygro commercial vectors. As shown in Figure 3.1.3, the ligations were successful for Nat-FL-His and Nat-S1-His with high concentrations of 975.4ng/μl and 857.2 ng/μl respectively. These constructs were then able to be reproduced with SpyTag this time as the SpyTag spy catcher system allows direct linking of recombinant protein to another structure, useful for some vaccine development strategies. As shown in Figure 3.1.5, these ligations were also successful into the pcDNA3.1hygro commercial vector as well as the Nucleocapsid-CHO insert also being ligated. All had high concentrations and the correct purities, as indicated in Table 3.1.3.

4.1.2 Generation of New Variants

The generation of new variant constructs as they emerged such as the B.1.1.7 and B.1.351 variants was also successful. The B.1.1.7 sequence was designed and commercially ordered and was then digested and ligated into the pcDNA3.1hygro and pcDNA3.1mGS commercial vectors. Figure 3.2.9 indicates the ligations were successful with these commercial vectors. For B.1.351, the mutations had to be defined via the monitoring of reported variants where sources were limited and not always reputable due to the constant updating of information of this variant. Using a variety of sources, a list of mutations was compiled as outlined in Table 3.1.8 and these mutations were chosen due to their consistency within the release of papers

and government documents. After being designed and commercially ordered, this B.1.351 spike gene was then successfully ligated into pcDNA3.1hygro and pcDNA3.1mGS commercial vectors with the ligations checked and shown via a restriction enzyme digest using *PmlI* as outlined in Figure 3.4.3.

The additional point mutations that emerged during the pandemic and the project were also successfully added to the relevant spike constructs including the B.1.1.7, Nat-FL-His and H7-FL-His. These mutations were D614G and E484K. The D614G mutation was the first to be added to the Nat-FL-His, B.1.1.7 and H7-FL-His spike constructs. Whilst the first attempt at SDM was not successful for B1.1.7 and H7-FL-His, there were colonies present for Nat-FL-His. This process was repeated again this time increasing the μl of template DNA added and this resulted in a successful SDM for all three of the spike constructs with multiple colonies which were then miniprepped and assessed for the highest quality ones considering their concentration and purity. As shown in Figure 3.3.3, the ligations were correct for these samples. As the B.1.1.7 (D614G) was the most prevalent and required spike construct at the time, this was sequenced to check the D614G mutation was present and no other mutations had also been picked up and the sequencing showed this to be correct, as shown in Figure 3.3.4. The E484K mutation was also successfully able to be added to the B.1.1.7 (D614G) spike construct via the process of SDM. Sequencing, as shown in Figure 3.3.6 shows the D614G and E4848K mutations present with no extra mutations picked up.

Creating these various spike constructs with variables such as different signal peptides, different tags and additional point mutations, allowed a bank of constructs to be generated for this project in which comparing these different variables in order to optimise the spike protein production process.

4.1.3 Addition of mGS into pcDNA3.1 vector

While the ligations and subcloning of the constructs were successful, other aspects of the DNA subcloning section of the project were not. Notably, the addition of mGS into the pcDNA3.1hygro vector as an alternative metabolic selection marker. As GS is required for stable cell line development, the addition of mGS into the commercial pcDNA3.1 vector replacing the hygro gene was important as a step to developing the spike protein production

process as stable cell line is a much more effective system of producing spike protein. The GS enables cells to be maintained when placed into medium lacking glutamine and containing MSX that inhibits the GHS enzyme, therefore ensuring the transfected CHO cells survive only when they have successfully integrated the plasmid DNA that also includes the target recombinant protein genes. Adding the mGS into the pcDNA3.1hygro vector required multiple steps and enzymes. The mGS was first released from a donor vector using *SmaI* and *PfoI* restriction enzymes which showed the expected band drop around 1100 bp as shown in Figure 3.1.6. However, the digest of the pcDNA3.1hygro vector was not successful. This was due to the expected band drop outs of around 1131 bp for the hygro gene being only faintly visible and therefore could not be extracted as the vector had not cut correctly, as shown in Figure 3.1.7. When investigating this, the issue seemed to be the optimal temperatures for the *SmaI* and *PfoI* enzymes being different. As they were being used together at a temperature that was not their optimum, it was decided to separate the enzymes and do two single digests with a clean up between the digests. This was not successful and resulted in a loss of yield. Even comparing the effectiveness for the double digests against the single digest, the gel was messy with inconclusive bands for the double digests and very faint bands for the single digests as outlined in Figure 3.1.8. The single digests were chosen to take forward for the ligation and as displayed in Figure 3.1.9, there were colonies for the mGS1, mGS2 and mGS3 samples as the expected band drop outs were present. However, when checking the ligations using the restriction enzymes *KpnI*, *BamHI* and *PvuI*, the expected band drop outs were not present especially when compared to the SnapGene™ agarose gel simulator, as shown in Figure 3.2.1, indicating the ligations had not worked even though the PCR Clone in Figure 3.1.9 suggested otherwise.

Overall, this experiment was not successful and did not contribute to the optimisation of the process of spike protein production as this vector was required for stable cell line development. As stable cell line development was one of the major aims to fulfil of this project when it came to developing spike protein, the pcDNA3.1mGS vector was commercially ordered and generated.

4.1.4 Addition of eGFP to SARS-CoV-2 constructs

Another aim of this project that was not able to be fulfilled due to limited time was the generation of Spike-eGFP fusion constructs for transient transfection. The generation of this construct would have allowed the location of the SARS-CoV-2 spike-GFP fusion protein to be viewed via confocal microscopy within the cells during transient expression (or stable expression) and allow the expression of the protein to be monitored. This opportunity to visualise the spike protein production could have also highlighted any other cellular issues in the spike synthesis such as aggregation and presence of stress granules that could limit spike protein yields. By replacing the his-tag of the spike constructs with eGFP, this would allow the expression of the spike protein in the cell to be identified and followed and the aim was to compare the location of the constructs on days 3 and 6, to gauge if there was a difference in cellular location and intracellular amounts. By using various ER signal peptides such as Nat-FL-, H7-FL, OSP-FL, OSP-FL (B.1.1.7), OSP-FL (B.1.351) and OSP-RBD with GFP tagged fusion, this would also allow a visual comparison of the efficiency of the different signal peptides to direct polypeptides to the ER and could have potentially provided an explanation for the reason of the poor translocation of the H7 and Native signal peptides as indicated by the results of the stable cell pool experiment. The correct restriction sites were able to be added to the spike vectors B.1.1.7, B.1.351, OSP-FL and H7-FL of *XbaI* and *ApaI* in which the his-tag was removed in preparation for the ligation of the eGFP tag. As the Nat-FL had a different set of restriction sites, the constructs had to be completely redesigned with the Nat-FL spike gene being ligated into the pcDNA3.1 vector in which the correct restriction sites could be added in preparation for the eGFP construct. The eGFP tags were also digested using their respective restriction enzymes depending on the spike vector. An overview of the spike vector and eGFP tag expected band drop outs are shown in Figure 3.2.6 showing that the vectors and inserts are correct and ready for ligation for a future eGFP study.

4.2 Protein Expression

4.2.1 Comparing spike protein expression in the CHO-S vs CHO-K1 host cell lines

A key section of this project was focused on the expression of the spike protein as the overall aim was to further optimise the spike protein production process and generate these coronavirus antigens for novel vaccine and diagnostic development. By transfecting a variety

of the spike protein constructs previously generated, the production process was able to be investigated through manipulation of a number of variables including comparing cell pools for their effectiveness, ER signal peptides and the ability to correctly translocate the spike protein such that it was secreted, and also the development of stable cell pools as this would allow continuous production of the spike protein in which more can be produced at a high quality and yield.

The CHO-S and CHO-K1 cell lines are two of the most commonly utilised CHO host cell lines for recombinant protein production due to their ability to produce human-like proteins, safety and stability (38). Comparing the expression of spike protein from these two cell lines was undertaken with a view to identifying the host that maintained high culture viability with the highest number of viable cells and highest production of the secreted spike protein. Transiently transfecting these cell lines with various spike protein constructs allowed the cell lines to be evaluated over days 4, 5 and 6 post-transfection. By looking at the culture viability and viable cell data of the CHO-K1 and CHO-S cell lines alone, as shown in Figure 3.4.4, the data indicated that the CHO-K1 cell line had greater culture viability remaining in the 80-99% bracket over the three-day period between days 4-6. CHO-S also had lower viable cells with an increase of $0.24x10^6$ for OSP-FL transfected cells over days 4,5 and 6 compared to CHO-K1 transfected cells with had an increase of $6.2x10^6$ cells. However, this data alone does not give an indication of how productive the cells are when it comes to spike protein synthesis.

By examining the protein expression of these cell lines via western blot, this allowed the most effective cell line to be determined and chosen to be carried forward with the project. Evaluating the cell lysate and supernatant samples of both the CHO-K1 and CHO-S cell lines, the data showed little evidence of the expression of the target proteins either intra or extracellularly. With the CHO-S cell line only showing evidence of expression in the supernatant for the OSP-FL sample, as displayed in Figure 3.4.6, it was decided that the CHO-S cell line should be carried forward. Even though the CHO-S cells displayed lower culture viability and viable cell concentrations compared to the CHO-K1 host post-transfection on the days analysed, it was the only cell line of the two to produce any expression of the spike protein at the expected size. An unexpected result of this experiment was the consistent 70 kDa bands present in many of the samples on the western blot. As previously stated, the

reason for this is unknown but a possibility could be due to the anti-His antibody detecting other cellular protein that are not related to the spike protein. In order to improve this, a different antibody such as anti-spike antibody which would directly bind to the spike protein instead of the his-tag as histidine is present in many cellular proteins, could be used in future (46).

4.2.2 Expression using the Native vs H7 signal peptide

Once the decision of the choice of the cell line was taken, the spike native and H7 ER signal peptides were then compared by transiently transfecting CHO-S cells that contained the spike with either one of these ER signal peptides. The reason for the use of the H7-signal peptide was that this has been shown to be effective for the recombinant production of heavy chain antibodies and was found to be the most effective ER signal peptide in a study of a number of ER signal peptides (45). By comparing the expression of the Native-FL-His and H7-FL-His spike constructs in CHO-S cells, this allowed evaluation of the effectiveness of the H7 signal peptide compared to the native signal peptide that had not been CHO optimised. The results of the cell lysate and supernatant blots, as shown in Figures 3.4.7 and 3.4.8 respectively, indicate little protein expression. While low levels of protein expression were expected for the cell lysate blot, this was not expected in the supernatant blot as spike protein in the expected 130 kDa region was not present and there was only a 70 kDa band in the Day 4 sample for H7-FL which showed no evidence of spike protein expression. For the H7-FL-His construct, low levels of expression could be explained due to the signal peptide being optimised for antibody production in CHO cells and not viral protein production. Due to the similar secretion pathway of co-translational translation pathway however (45)(47), this signal peptide should have been effective. As shown previously in the comparison of the CHO-K1 and CHO-S cell lines, the lack of expression of the native-FL-his protein was not expected and this is concurrent with the results of Figures 3.4.7 and 3.4.8 indicating a potential issue with the native signal peptide also. This signal peptide was not CHO optimised and therefore could be the explanation for the lack of translocation in the CHO-S cells. However, this signal peptide was used in stable cell line generation and did result in good amounts of secreted spike protein (see discussion below) and therefore this rules out this being the issue during transient expression of the protein.

4.2.3 CHO-S stable cell pool development

One of the main aims of this project was to develop various stably CHO cell expressing cell pools of different SARS-CoV-2 Spike constructs transfected into CHO-S cells. By developing stable cell lines, this allows continuous production from running batch or fed-batch cultures of the cells as and when protein is required. Compared to transiently transfected cells, stable cell lines can pass on all genetic modifications made to the cells to the next generation. The outcome of this results in larger amounts of protein being produced making the process far more productive (48). Stable cell line development is also far more convenient and profitable due to the steady onstream production of protein generation. As well as developing stable cell pools, various spike SARS-CoV-2 constructs were transfected containing various signal peptides also enabling comparison of the Native, H7 and OSP (CHO optomised) signal peptides. Another variable that was investigated within this experiment was the concentration of MSX added to the CD-CHO medium during selection to compare the effectiveness as recent studies found 50 μM MSX yielded CHO-S pools of spike protein that showed lower productivity than the 25 μM MSX transfected pools (49). The result of this study was reflected in the development of this stable cell line as for each spike SARS-CoV-2 construct, two stable cell lines were developed one with 50 μM and one with 25 μM MSX. In order to evaluate these stable cell pools, the culture viability and the viable cells of the cell pools were determined, the spike protein expression measured as well as the levels of spike transcript mRNA expressed.

The result of this experiment was concurrent with the results of the recent study published by Johari et al (2021) as the 25 μM MSX exhibited higher protein expression. The Day 6 supernatant spike expression western blot, as shown in Figure 3.5.4, depicts a stronger protein band in the 25 μM MSX for the OSP-RBD compared to the 50 μM MSX OSP-RBD sample in the expected band region of around 30 kDa. There was also a visually slightly stronger band for the 25 μM MSX OSP-FL as well as a faint band for the Nat-FL compared to the 50 μM MSX samples. The end of culture supernatant spike expression western blot, shown in Figure 3.5.5, shows the 50 μM MSX OSP-RBD protein band to be stronger than the 25 μM MSX OSP-RBD protein band. This could be due to the higher concentration of MSX repressing endogenous GS within the cells meaning it takes longer for cells to grow and

produce higher levels of spike protein compared to the 25 μM MSX which exhibits a lower pressure on the cells. When looking at the culture viability and viable cells of these stable cell lines, as shown in Figures 3.5.1 and 3.5.2, it is clear to see the 25 μM MSX stable cell pools had lower culture viabilities and lower numbers of viable cells which quickly declined around the 168-hour mark compared to the 50 μM MSX cell pools which declined at a slower rate around the 192-216-hour mark. This means that cell numbers and culture viability do not explain the difference in spike production observed

As stable cell pools require GS expression to survive, the GS expression for the 10 stable cell pools was measured and the results displayed in Figure 3.5.7 show clear GS bands in the expected band size for GS around 35 kDa regardless of the MSX concentration. By checking no GS was being expressed by the CHO-S cells, as shown in Figure 3.5.8, this confirmed the GS expression was only detectable in the stable cell pools that had been transfected with DNA plasmids containing the mGS gene.

When comparing the Native, H7 and OSP signal peptides in this experiment, the western blots produced showing protein expression clearly show the native OSP signal peptide to give the highest yields, with the strongest protein band intensity in the expected regions being present compared to the Native and H7 signal peptides which either had little to no evidence of spike protein expression, as shown in Figure 3.5.4 and Figure 3.5.5. The spike sequence used for this project was based on the design by Wrapp *et al* 2020 (50) in which the transmembrane and cytoplasmic domains of the native protein were removed. Two proline amino acids were also added to improve stability of the protein as well as the removal of cleavage sites which are often subjected to proteolytic cleavage (50). Budge et al 2021 (51) CHO optimised the spike sequence and substituted the native signal peptide with a mammalian signal sequence used in Lonza expression vectors to enable the secretion and sequence cleavage via signal peptidase in CHO cells (51). For this project, the CHO optomised Spike sequence was used but a CHO optomised SARS-CoV 2 signal peptide was used in place of the Lonza signal peptide, this cloning was carried out in house. By CHO optimising the signal peptide i.e. changing the nucleotide sequence to use the most abundant tRNAs as codons, this results in optimal translation (52).

To further confirm the effectiveness of the signal peptides and to compare the transcript amounts between them, qRT-PCR analysis was run. The results shown in Figure 3.5.9 show that the reason for the stronger spike protein expression of the OSP-FL is due to higher transcript amounts as both 50 μM MSX OSP-FL and 25 μM MSX OSP-FL had the highest transcript amounts. Looking at the transcript amounts in comparison to the levels of MSX also showed a difference as the OSP-FL 25 μM MSX mRNA sample exhibited higher transcript amounts compared to the OSP-FL 50 μM MSX mRNA sample . The H7-FL and Nat-FL cell pools exhibited much lower transcript amounts. This suggests that the different signal peptides impact the mRNA stability as the same promoter was used to drive transcription of all the signal peptide variants. The hypothesis here is that the native signal peptide sequence stabilises the mRNA and there is less turnover and hence higher transcript amounts which results in higher protein expression. This could be tested by measuring the mRNA half-life of the different transcripts.

4.3 Measuring Antibody Response

4.3.1 ELISA: Generating Antibody Response

The last aim of the project was to investigate the antibody response between the OSP-FL, B.1.1.7 and B.1.351 spike variants against hospital patient serums from either Wuhan or B.1.1.7 infected patients that should therefore have anti-spike antibodies to these variants. To do this, ELISAs were set up to compare the antibody response of the Wuhan serum vs the B.1.1.7 serum against the different spike proteins.

The results of this experiment showed there was an antibody response with the Wuhan serum sample against the spike proteins as expected when using the original spike protein and there was an expected antibody-spike protein concentration- response relationship. As the samples became more diluted, the lower the signal indicating a lower antibody response. The B.1.1.7 spike sample gave the highest signal at a dilution of 1:20, as shown in Figure 3.6.1, compared to the Nat-FL Wuhan spike. This was unexpected as the Wuhan infected patient serum antibodies would have been expected to be bound more directly to the Nat-FL (Wuhan) spike. Overall it was found there was no large difference between the responses of the spike protein variants against the Wuhan infected serum although further studies would need to be

undertaken to statistically confirm this. Due to the similarity of the variants with 8 mutations present in the B.1.1.7, as shown in Table 1.1.1, and the 10 mutations present in the B.1.351, as shown in Table 3.1.8, there is still a large degree of similarity by which the antibodies will still recognise these variants. It is worth noting that the E484K mutation that the B.1.351 variant possesses has reportedly the ability to evade neutralising antibodies (23). As shown in Figure 3.6.2, the B.1.1.7 infected patient serum showed little binding of serum antibodies to the various spike protein variants. This could be due to the anti-human IgG (gamma chain specific) instead of the whole antibody used for detection which could provide a better response. The positive control using anti-Spike antibody displayed the expected response, as shown in Figure 3.6.3, however this was not an ideal positive control due to the anti-spike antibody being raised in rabbit. Overall, the ELISA data showed a good antibody response in the Wuhan serum sample against the different various spike samples. This is also concurrent with recent studies showing the CV2-75, CV3-17, CV2-71 and CV3-25 cross-neutralising mAbs retain their neutralising activities against B.1.351 even with its neutralising mutations such as E484K (53) indicating that even the response of these antibodies across the different variants is of similar levels.

4.4 Limitations of this study

The limitations of this study were mainly due to time pressure and the restricted access to the laboratory due to the COVID pandemic throughout the project. Large portions of this project were catered towards collaborators such as the generation of new variants, notably the B.1.1.7 and B.1.351 in which the Maxipreps of the constructs needed to be generated in a short period of time to enable expression of the protein which could then be used in the collaborators vaccine and diagnostic studies.

Another limitation of the study was the stable cell pools were generated and run in singles and not triplicates. Due to restrictions in incubator space and quantity levels, triplicates were not able to be run. This meant the development of the stable pools were aimed at giving an overview as to what the stable cell line development process looked like generating spike protein for the first time. Further repeating of this experiment would improve the validity of the observed results.

Due to the limitations in lab access, the eGFP experiment was not able to be completed. However, while the whole experiment was not able to be completed, the components of the DNA constructs were able to be sub cloned ready for ligation and ultimately immunofluorescence studies.

4.5 Future Work and Ideas

After the stable cell pools had been developed, further research could be carried out to manipulate the culture conditions to increase the quantities of spike protein being produced. Temperature changes and small molecule chemical chaperone reagents that have been shown to influence recombinant protein production from CHO cells (54)(55) could be investigated to compare the most effective way of increasing spike protein production in the CHO-S cells further optimising the spike protein production process.

Interestingly, a study by Avello *et al* in 2017 (56) found that by modulating the cell cycle control points such as G1/S or G2/M by imposing mild hypothermia at 33°C and the addition of sodium butyrate, this impacted the metabolic and physiological state of CHO TF 70R cells by decreasing cell growth rate whilst improving glucose consumption efficiency. Overall, this increased productivity of the cells and increased protein production. These variables also increased doubling time of the cells slowing the cell cycle, thus more recombinant protein being produced (56). By applying the same variables of temperature and sodium butyrate, this could also have the desired effect on the stable CHO-S cell pools generating spike protein.

Another chemical reagent that has been found to increase protein levels in CHO cells is Valproic acid. This is a molecule that inhibits histone deacetylase activity, an enzyme that makes DNA less accessible to transcription factors. Valproic acid has previously been shown to increase levels of recombinant protein production in CHO cells. A study by Wulhard *et al* 2010 (57) found that Valproic acid increased transient recombinant protein yields with the increase in protein production being correlated with an increase in mRNA levels (57). While Valproic acid is more effective in transient transfections for CHO cells (58), it could be investigated for its effectiveness to increase spike protein expression. This would link to the observation that higher transcript amounts from the native signal peptide transcript resulted in higher secreted protein amounts. By comparing the effectiveness of such chemical

reagents alongside reducing the temperature the cells are kept at, this could in turn increase recombinant spike protein production which would then further optimise the spike protein production process.

Another future research idea would be completing the eGFP experiment in which the CHO-S cells were transfected with the various spike-eGFP constructs. The location of the spike protein could then be visualised in the cell and could be monitored for aggregation and stress granules. If it was found that the spike protein was being retained within the cells or aggregating, approaches to alleviate this could then be investigated.

As viruses depend on cellular translation machinery for replication and protein synthesis, when activated by the innate immune response, the stress kinase PKR inhibits translation in an attempt to block the replication of viruses. By doing this, this results in the formation of cytoplasmic granules, also known as stress granules containing all the translational machinery required until the cell recovers from stress (59). By transfecting CHO cells with various-spike-constructs, the potential for stress granules to form is there and should be investigated by tagging the spike constructs with eGFP to visualise the location of the protein in the cell. Indeed, a recent study by Zheng et al 2021 (60) found that the SARS-CoV-2 nucleocapsid protein impairs stress granule formation in order to promote replication of the virus. As the role of the nucleocapsid is RNA binding, critical for viral replication (61), the spike protein is less likely to prevent stress granules thus worth investigating by the means of eGFP tagging and confocal microscopy.

4.6 Conclusions

This project set out to generate coronavirus antigens for novel vaccine and diagnostic approaches and to further improve the spike protein production process from CHO cells. By developing various coronavirus antigens, notably various SARS-CoV-2 spike constructs, the antigens were able to be used within other studies involving novel vaccine and diagnostic approaches by collaborators. Through DNA subcloning, a bank of various Spike protein constructs within commercial vectors were generated ready for future studies. The B.1.1.7 and B.1.351 variant that emerged throughout the project were able to be synthesised and

cloned and quantities ready for transfection in CHO cells to express the spike protein required for collaborators research projects.

By comparing CHO-K1 and CHO-S cell lines, the CHO-S cell line was found to be the most effective which was taken forward when comparing the Native and H7 signal peptides. When developing the stable cell pools, this allowed data to be gathered enabling the spike protein production process to be enhanced with the native OSP signal peptide being the most effective signal peptide due to its spike protein expression and its high mRNA expression. It was also found that by lowering the concentration of MSX to 25 instead of 50 μM, this also increased spike protein expression thus optimising the process further. The ELISAs measuring antibody responses provided a concentration-response relationship with the Wuhan serum against various spike variants. Overall, the spike protein production process was able to be enhanced and with future research such as fed batch cell cultures of the CHO-S spike cultures, this could be optimised further to increase spike protein quantities.

Bibliography

1.   Araf, Y., N.A. Faruqui, S. Anwar, and M.J. Hosen. 2021. SARS-CoV-2: a new dimension to our understanding of coronaviruses. *Int. Microbiol.* 24:19–24.
2.   V'kovski, P., A. Kratzel, S. Steiner, H. Stalder, and V. Thiel. 2021. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* 19:155–170.
3.   Organiza-, T.W.H. 2021. WHO Coronavirus (COVID-19) Dashboard. .
4.   Perlman, S., and J. Netland. 2009. Coronaviruses post-SARS: Update on replication and pathogenesis. *Nat. Rev. Microbiol.* 7:439–450.
5.   Wiersinga, W.J., A. Rhodes, A.C. Cheng, S.J. Peacock, and H.C. Prescott. 2020. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA - J. Am. Med. Assoc.* 324:782–793.
6.   Sungnak, W., N. Huang, C. Bécavin, M. Berg, R. Queen, M. Litvinukova, C. Talavera-López, H. Maatz, D. Reichart, F. Sampaziotis, K.B. Worlock, M. Yoshida, J.L. Barnes, N.E. Banovich, P. Barbry, A. Brazma, J. Collin, T.J. Desai, T.E. Duong, O. Eickelberg, C. Falk, M. Farzan, I. Glass, R.K. Gupta, M. Haniffa, P. Horvath, N. Hubner, D. Hung, N. Kaminski, M. Krasnow, J.A. Kropski, M. Kuhnemund, M. Lako, H. Lee, S. Leroy, S. Linnarson, J. Lundeberg, K.B. Meyer, Z. Miao, A. V. Misharin, M.C. Nawijn, M.Z. Nikolic, M. Noseda, J. Ordovas-Montanes, G.Y. Oudit, D. Pe'er, J. Powell, S. Quake, J. Rajagopal, P.R. Tata, E.L. Rawlins, A. Regev, P.A. Reyfman, O. Rozenblatt-Rosen, K. Saeb-Parsy, C. Samakovlis, H.B. Schiller, J.L. Schultze, M.A. Seibold, C.E. Seidman, J.G. Seidman, A.K. Shalek, D. Shepherd, J. Spence, A. Spira, X. Sun, S.A. Teichmann, F.J. Theis, A.M. Tsankov, L. Vallier, M. van den Berge, J. Whitsett, R. Xavier, Y. Xu, L.E. Zaragosi, D. Zerti, H. Zhang, K. Zhang, M. Rojas, and F. Figueiredo. 2020. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* 26:681–687.
7.   Huang, Y., C. Yang, X. feng Xu, W. Xu, and S. wen Liu. 2020. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* 41:1141–1149.
8.   Watanabe, Y., J.D. Allen, D. Wrapp, J.S. McLellan, and M. Crispin. 2020. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science (80-.).* 369:330–333.
9.   Guan, W., Z. Ni, Y. Hu, W. Liang, C. Ou, J. He, L. Liu, H. Shan, C. Lei, D.S.C. Hui, B. Du, L. Li, G. Zeng, K.-Y. Yuen, R. Chen, C. Tang, T. Wang, P. Chen, J. Xiang, S. Li, J. Wang, Z. Liang, Y. Peng, L. Wei, Y. Liu, Y. Hu, P. Peng, J. Wang, J. Liu, Z. Chen, G. Li, Z. Zheng, S. Qiu, J. Luo, C. Ye, S. Zhu, and N. Zhong. 2020. Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* 382:1708–1720.
10.   López-León, S., T. Wegman-Ostrosky, C. Perelman, R. Sepulveda, P.A. Rebolledo, A. Cuapio, and S. Villapol. 2021. More than 50 Long-Term Effects of COVID-19: A Systematic Review and Meta-Analysis. *SSRN Electron. J.* 1–22.
11.   Marroquín, B., V. Vine, and R. Morgan. 2020. Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information . .
12.   Brooks, S.K., R.K. Webster, L.E. Smith, L. Woodland, S. Wessely, N. Greenberg, and G.J. Rubin. 2020. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *Lancet*. 395:912–920.
13.   Wang, P., M.S. Nair, L. Liu, S. Iketani, Y. Luo, Y. Guo, M. Wang, J. Yu, B. Zhang, P.D.

Kwong, B.S. Graham, J.R. Mascola, J.Y. Chang, M.T. Yin, M. Sobieszczyk, C.A. Kyratsous, L. Shapiro, Z. Sheng, Y. Huang, and D.D. Ho. 2021. Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7. *bioRxiv*. 2021.01.25.428137.

14. Kim, Y., E.J. Kim, S.W. Lee, and D. Kwon. 2021. Review of the early reports of the epidemiological characteristics of the b.1.1.7 variant of sars-cov-2 and its spread worldwide. *Osong Public Heal. Res. Perspect.* 12:139–148.

15. Rambaut, A., N.J. Loman, and O.G. Pybus. 2020. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. .

16. Starr, T.N., A.J. Greaney, S.K. Hilton, D. Ellis, K.H.D. Crawford, A.S. Dingens, M.J. Navarro, J.E. Bowen, M.A. Tortorici, A.C. Walls, N.P. King, D. Veesler, and J.D. Bloom. 2020. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*. 182:1295-1310.e20.

17. 2020. Short Article A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. .

18. Zhao J, Yuan Q, Wang H, et al. 2020. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. Clin Infect Dis. 2020;ciaa344. doi:10.1093/cid/ciaa344. *Clin. Infect. Dis.* 1–22.

19. Tegally, H., E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, S. Pillay, E.J. San, N. Msomi, K. Mlisana, A. von Gottberg, S. Walaza, M. Allam, A. Ismail, T. Mohale, A.J. Glass, S. Engelbrecht, G. Van Zyl, W. Preiser, F. Petruccione, A. Sigal, D. Hardie, G. Marais, N. yuan Hsiao, S. Korsman, M.A. Davies, L. Tyers, I. Mudau, D. York, C. Maslo, D. Goedhals, S. Abrahams, O. Laguda-Akingba, A. Alisoltani-Dehkordi, A. Godzik, C.K. Wibmer, B.T. Sewell, J. Lourenço, L.C.J. Alcantara, S.L. Kosakovsky Pond, S. Weaver, D. Martin, R.J. Lessells, J.N. Bhiman, C. Williamson, and T. de Oliveira. 2021. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*. 592:438–443.

20. Hagen.A (M.S). How Dangerous Is the Delta Variant (B.1.617.2)? *Am. Soc. Microbiol.*

21. Martinez, B., G. Whittaker, N. York, T. Alpha, W. Barclay, B. Delta, and K. Sato. 2002. The mutation that helps delta spread like wildfire. 2–3.

22. Korber, B., W.M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E.E. Giorgi, T. Bhattacharya, B. Foley, K.M. Hastie, M.D. Parker, D.G. Partridge, C.M. Evans, T.M. Freeman, T.I. de Silva, A. Angyal, R.L. Brown, L. Carrilero, L.R. Green, D.C. Groves, K.J. Johnson, A.J. Keeley, B.B. Lindsey, P.J. Parsons, M. Raza, S. Rowland-Jones, N. Smith, R.M. Tucker, D. Wang, M.D. Wyles, C. McDanal, L.G. Perez, H. Tang, A. Moon-Walker, S.P. Whelan, C.C. LaBranche, E.O. Saphire, and D.C. Montefiori. 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*. 182:812-827.e19.

23. Wise, J. 2021. Covid-19: The E484K mutation and the risks it poses. *BMJ*. 372:n359.

24. Weisblum, Y., F. Schmidt, F. Zhang, J. DaSilva, D. Poston, J.C.C. Lorenzi, F. Muecksch, M. Rutkowska, H.H. Hoffmann, E. Michailidis, C. Gaebler, M. Agudelo, A. Cho, Z. Wang, A. Gazumyan, M. Cipolla, L. Luchsinger, C.D. Hillyer, M. Caskey, D.F. Robbiani, C.M. Rice, M.C. Nussenzweig, T. Hatziioannou, and P.D. Bieniasz. 2020. Escape from neutralizing antibodies 1 by SARS-CoV-2 spike protein variants. *Elife*. 9:1.

25. Jangra, S., C. Ye, R. Rathnasinghe, D. Stadlbauer, H. Alshammary, A.A. Amoako, M.H. Awawda, K.F. Beach, M.C. Bermúdez-González, R.L. Chernet, L.Q. Eaker, E.D. Ferreri, D.L. Floda, C.R. Gleason, G. Kleiner, D. Jurczyszak, J.C. Matthews, W.A. Mendez, L.C.F.

Mulder, K.T. Russo, A.B.T. Salimbangon, M. Saksena, A.S. Shin, L.A. Sominsky, K. Srivastava, F. Krammer, V. Simon, L. Martinez-Sobrido, A. García-Sastre, and M. Schotsaert. 2021. SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. *The Lancet Microbe*. 2:e283–e284.

26. Organiza-, T.W.H. 2018. Vaccines work. *Nat. Commun.* 9:1666.

27. Czochor, J., and A. Turchick. 2014. Introduction. *Yale J. Biol. Med.* 87:401–402.

28. Iwasaki, A., and S.B. Omer. 2020. Why and How Vaccines Work. *Cell*. 183:290–295.

29. Wang, J., Y. Peng, H. Xu, Z. Cui, and R.O. Williams. 2020. The COVID-19 Vaccine Race: Challenges and Opportunities in Vaccine Formulation. *AAPS PharmSciTech*. 21:1–12.

30. Lessard, J.C. 2013. Molecular cloning. 1st ed. Elsevier Inc.

31. Celie, P.H.N., A.H.A. Parret, and A. Perrakis. 2016. Recombinant cloning strategies for protein expression. *Curr. Opin. Struct. Biol.* 38:145–154.

32. Kim, J.Y., Y.G. Kim, and G.M. Lee. 2012. CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. *Appl. Microbiol. Biotechnol.* 93:917–930.

33. You, M., Y. Yang, C. Zhong, F. Chen, X. Wang, T. Jia, Y. Chen, B. Zhou, Q. Mi, Q. Zhao, Z. An, W. Luo, and N. Xia. 2018. Efficient mAb production in CHO cells with optimized signal peptide, codon, and UTR. *Appl. Microbiol. Biotechnol.* 102:5953–5964.

34. Butler, M., and M. Spearman. 2014. The choice of mammalian cell host and possibilities for glycosylation engineering. *Curr. Opin. Biotechnol.* 30:107–112.

35. Dumont, J., D. Euwart, B. Mei, S. Estes, and R. Kshirsagar. 2016. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit. Rev. Biotechnol.* 36:1110–1122.

36. Durocher, Y., and M. Butler. 2009. Expression systems for therapeutic glycoprotein production. *Curr. Opin. Biotechnol.* 20:700–707.

37. Berting, A., M.R. Farcet, and T.R. Kreil. 2010. Virus susceptibility of Chinese hamster ovary (CHO) cells and detection of viral contaminations by adventitious agent testing. *Biotechnol. Bioeng.* 106:598–607.

38. Dyson, M.R. 2016. Fundamentals of expression in mammalian cells. *Adv. Exp. Med. Biol.* 896:217–224.

39. Ghaderi, D., M. Zhang, N. Hurtado-Ziola, and A. Varki. 2012. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnol. Genet. Eng. Rev.* 28:147–176.

40. Nakamura, T., and T. Omasa. 2015. Optimization of cell line development in the GS-CHO expression system using a high-throughput, single cell-based clone selection system. *J. Biosci. Bioeng.* 120:323–329.

41. Fan, L., I. Kadura, L.E. Krebs, C.C. Hatfield, M.M. Shaw, and C.C. Frye. 2012. Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. *Biotechnol. Bioeng.* 109:1007–1015.

42. Pereira, S., H.F. Kildegaard, and M.R. Andersen. 2018. Impact of CHO Metabolism on Cell Growth and Protein Production: An Overview of Toxic and Inhibiting Metabolites and Nutrients. *Biotechnol. J.* 13.

43. Reddington, S.C., and M. Howarth. 2015. Secrets of a covalent interaction for biomaterials and biotechnology: SpyTag and SpyCatcher. *Curr. Opin. Chem. Biol.* 29:94–99.

44. Tegally, H., E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, S. Pillay, E.J. San, K. Wibmer, B.T. Sewell, J. Lourenço, L. Carlos, J. Alcantara,

S.L. Kosakovsky, S. Weaver, D. Martin, R.J. Lessells, J.N. Bhiman, C. Williamson, and T. De Oliveira. 2020. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *Arghavan Alisoltani-Dehkordi*. 10:2020.12.21.20248640.

45. Haryadi, R., S. Ho, Y.J. Kok, H.X. Pu, L. Zheng, N.A. Pereira, B. Li, X. Bi, L.T. Goh, Y. Yang, and Z. Song. 2015. Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. *PLoS One*. 10:1–16.

46. Holeˇcek, M. 2020. Histidine in Health and Disease: Metabolism, Physiological Importance, and Use as a Supplement Milan. *Nutrients*. 12.

47. Duan, L., Q. Zheng, H. Zhang, Y. Niu, Y. Lou, and H. Wang. 2020. The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. *Front. Immunol.* 11:1–12.

48. Scientists, B.R. for. Stable Cell Lines. .

49. Johari, Y.B., S.R.P. Jaffé, J.M. Scarrott, A.O. Johnson, T. Mozzanino, T.H. Pohle, S. Maisuria, A. Bhayat-Cammack, G. Lambiase, A.J. Brown, K.L. Tee, P.J. Jackson, T.S. Wong, M.J. Dickman, R.B. Sargur, and D.C. James. 2021. Production of trimeric SARS-CoV-2 spike protein by CHO cells for serological COVID-19 testing. *Biotechnol. Bioeng.* 118:1013–1021.

50. Wrapp, D., N. Wang, K.S. Corbett, J.A. Goldsmith, C.L. Hsieh, O. Abiona, B.S. Graham, and J.S. McLellan. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *bioRxiv*. 1263:1260–1263.

51. Budge, J.D., R.J. Young, and C.M. Smales. 2021. Engineering of Chinese Hamster Ovary Cells With NDPK-A to Enhance DNA Nuclear Delivery Combined With EBNA1 Plasmid Maintenance Gives Improved Exogenous Transient Reporter, mAb and SARS-CoV-2 Spike Protein Expression. *Front. Bioeng. Biotechnol.* 9:1–14.

52. Mauro, V.P. 2018. Codon Optimization in the Production of Recombinant Biotherapeutics: Potential Risks and Considerations. *BioDrugs*. 32:69–81.

53. Jennewein, M.F., A.J. MacCamy, and Nicholas R. Akins. 2021. Isolation and Characterization of Cross-Neutralizing Coronavirus Antibodies from COVID-19+ Subjects. *bioRxiv*. 3:436684.

54. Masterton, R.J., A. Roobol, M.B. Al-Fageeh, M.J. Carden, and C.M. Smales. 2010. Post-translational events of a model reporter protein proceed with higher fidelity and accuracy upon mild hypothermic culturing of Chinese hamster ovary cells. *Biotechnol. Bioeng.* 105:215–220.

55. Johari, Y.B., S.D. Estes, C.S. Alves, M.S. Sinacore, and D.C. James. 2015. Integrated cell and process engineering for improved transient production of a "difficult-to-express" fusion protein by CHO cells. *Biotechnol. Bioeng.* 112:2527–2542.

56. Avello, V., B. Tapia, M. Vergara, C. Acevedo, J. Berrios, J.G. Reyes, and C. Altamirano. 2017. Impact of sodium butyrate and mild hypothermia on metabolic and physiological behaviour of CHO TF 70R cells. *Electron. J. Biotechnol.* 27:55–62.

57. Wulhfard, S., L. Baldi, D.L. Hacker, and F. Wurm. 2010. Valproic acid enhances recombinant mRNA and protein levels in transiently transfected Chinese hamster ovary cells. *J. Biotechnol.* 148:128–132.

58. Backliwal, G., M. Hildinger, I. Kuettel, F. Delegrange, D.L. Hacker, and F.M. Wurm. 2008. Valproic acid: A viable alternative to sodium butyrate for enhancing protein expression in mammalian cell cultures. *Biotechnol. Bioeng.* 101:182–189.

59. Miller, C.L. 2011. Stress granules and virus replication. *Future Virol.* 6:1329–1338.

60. Zheng, Z.Q., S.Y. Wang, Z.S. Xu, Y.Z. Fu, and Y.Y. Wang. 2021. SARS-CoV-2 nucleocapsid protein impairs stress granule formation to promote viral replication. *Cell Discov.* 7.

61. Cubuk, J., J.J. Alston, J.J. Incicco, S. Singh, M.D. Stuchell-Brereton, M.D. Ward, M.I. Zimmerman, N. Vithani, D. Griffith, J.A. Wagoner, G.R. Bowman, K.B. Hall, A. Soranno, and A.S. Holehouse. 2021. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* 12:1–17.
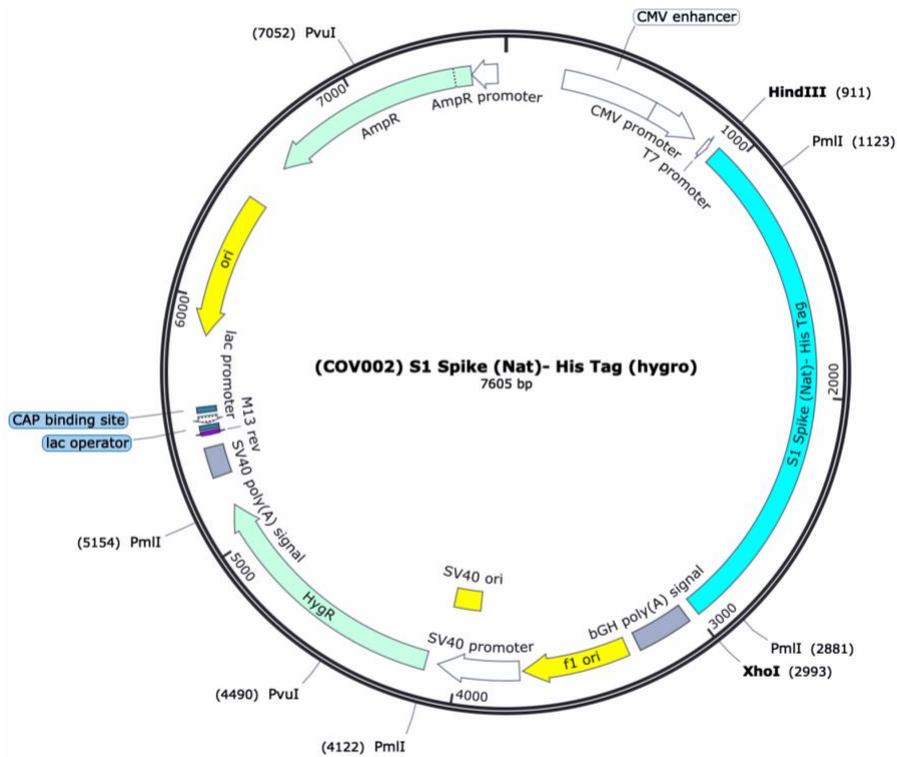
Appendices
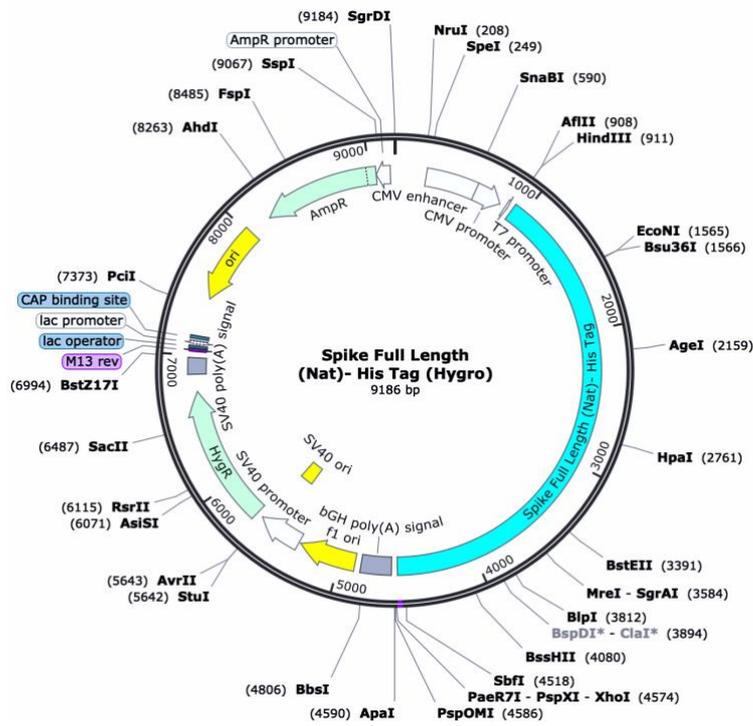
Appendix 1: *Plasmid Map showing the Nat-FL-His insert in pcDNA3.1hygro vector*
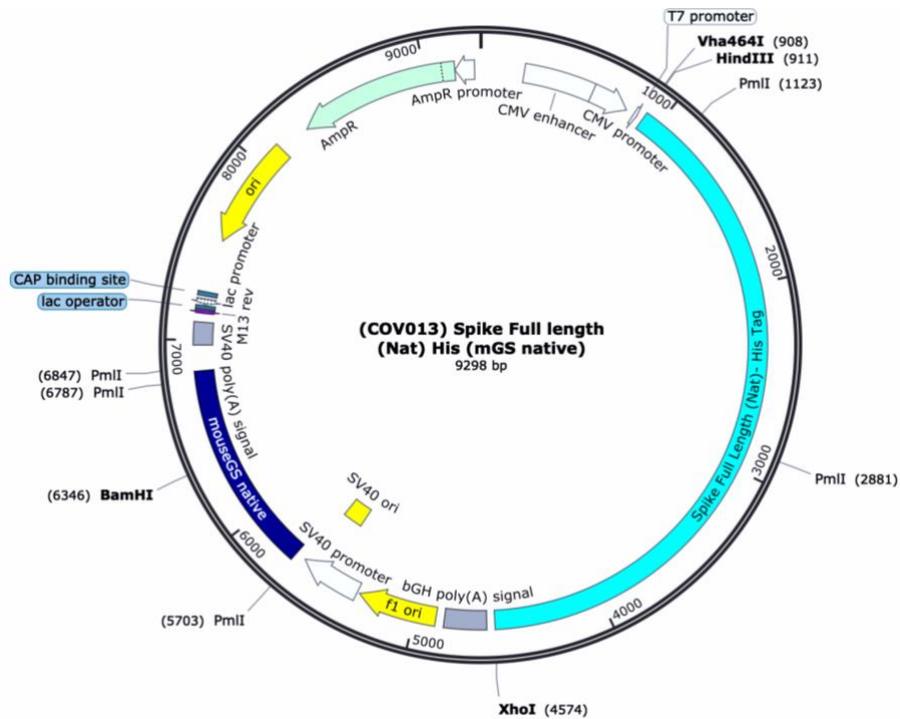
Appendix 2: *Plasmid Map showing the Nat-FL-His insert in pcDNA3.1mGS vector*

Appendix 3: *Plasmid Map showing the Nat-S1-His insert in pcDNA3.1hygro vector*
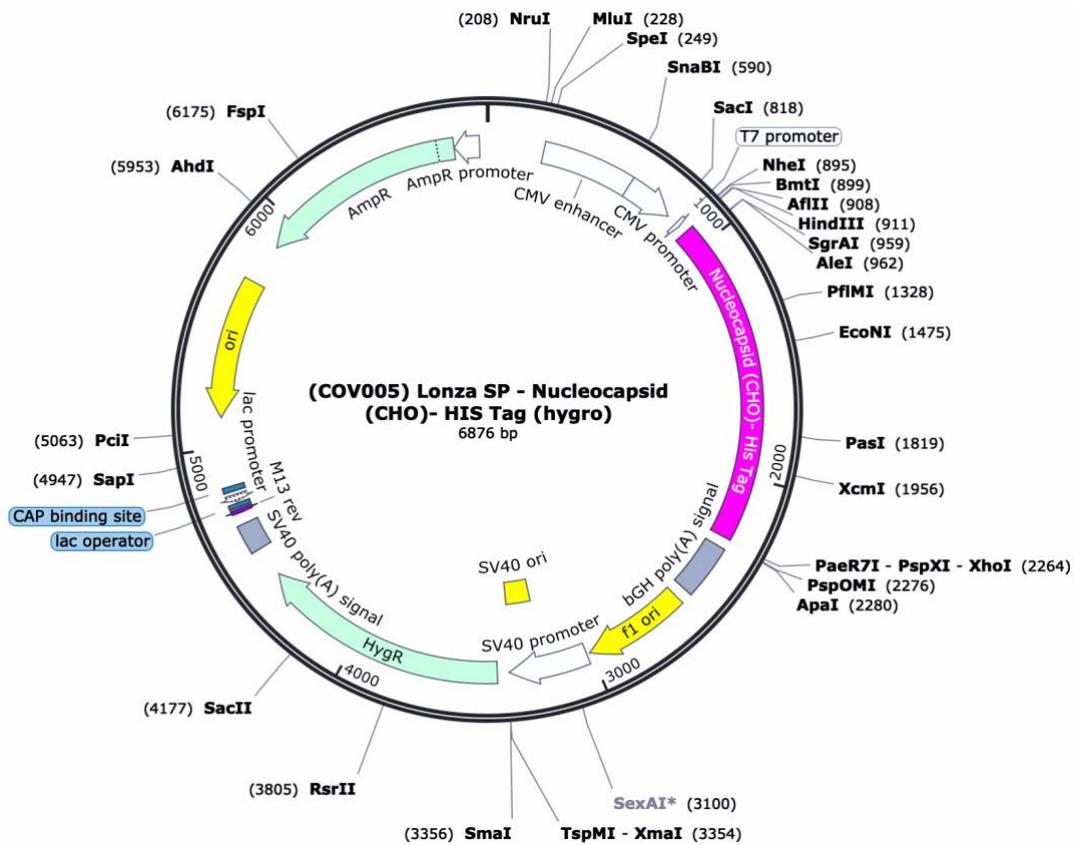
Appendix 4: *Plasmid Map showing the Nucleocapsid-CHO insert in pcDNA3.1hygro vector with Lonza-SP*
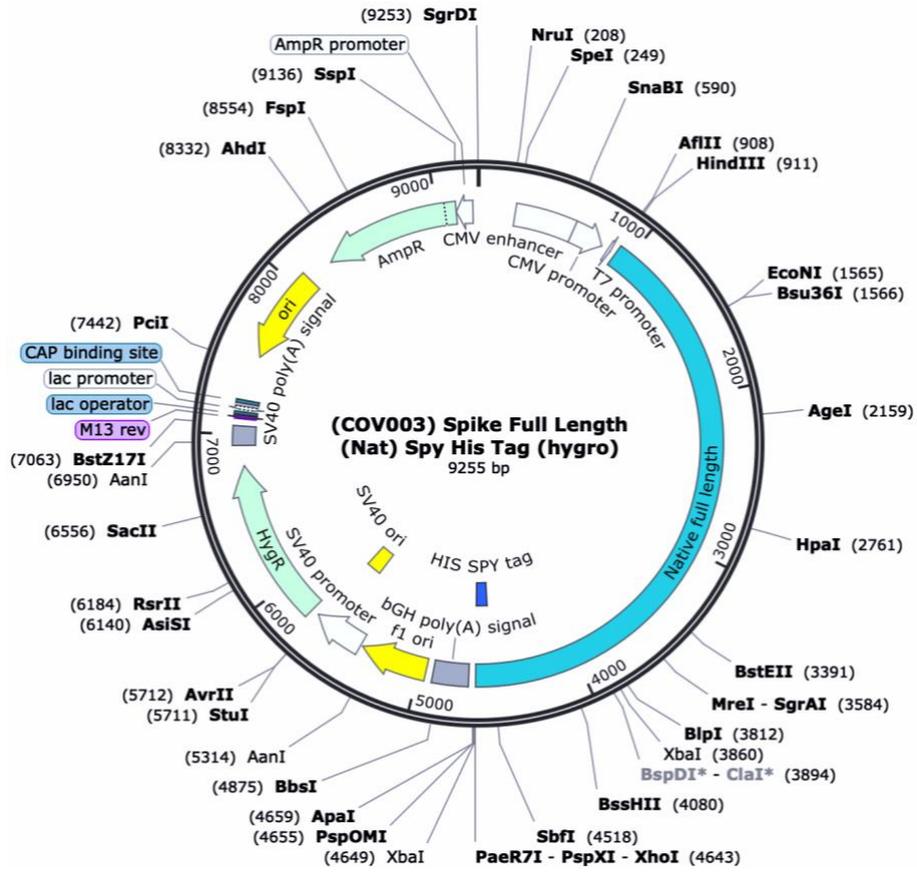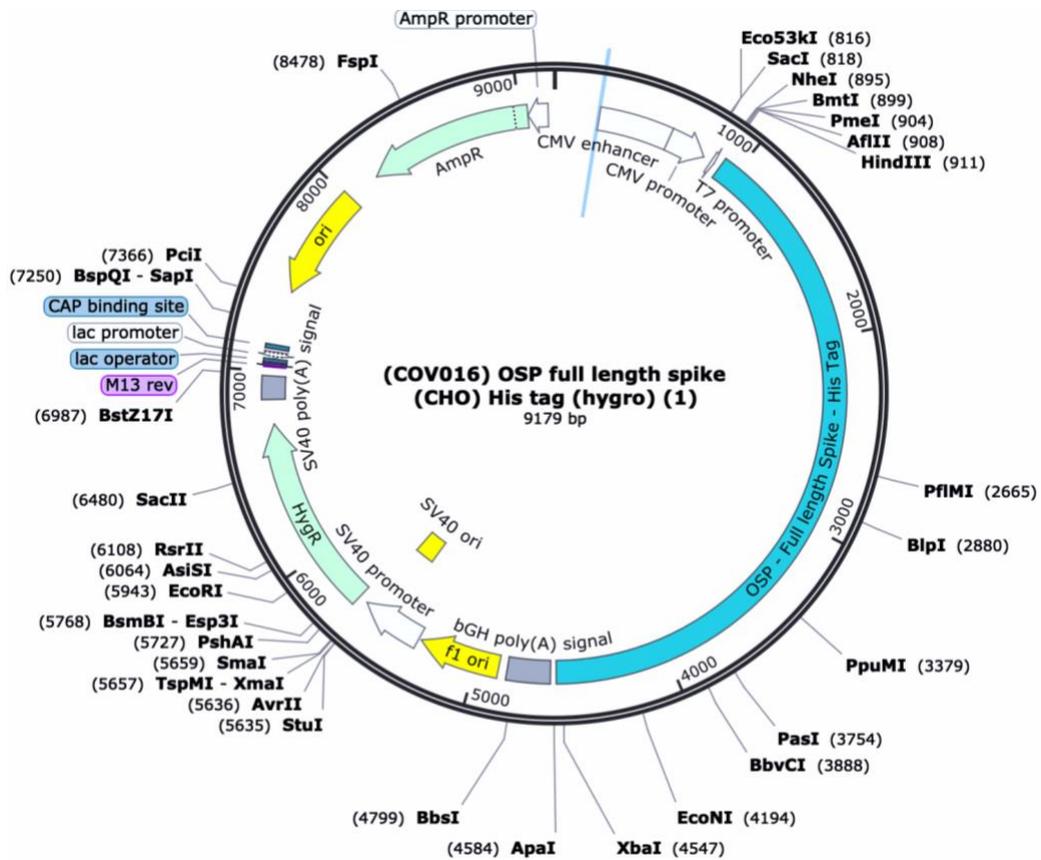
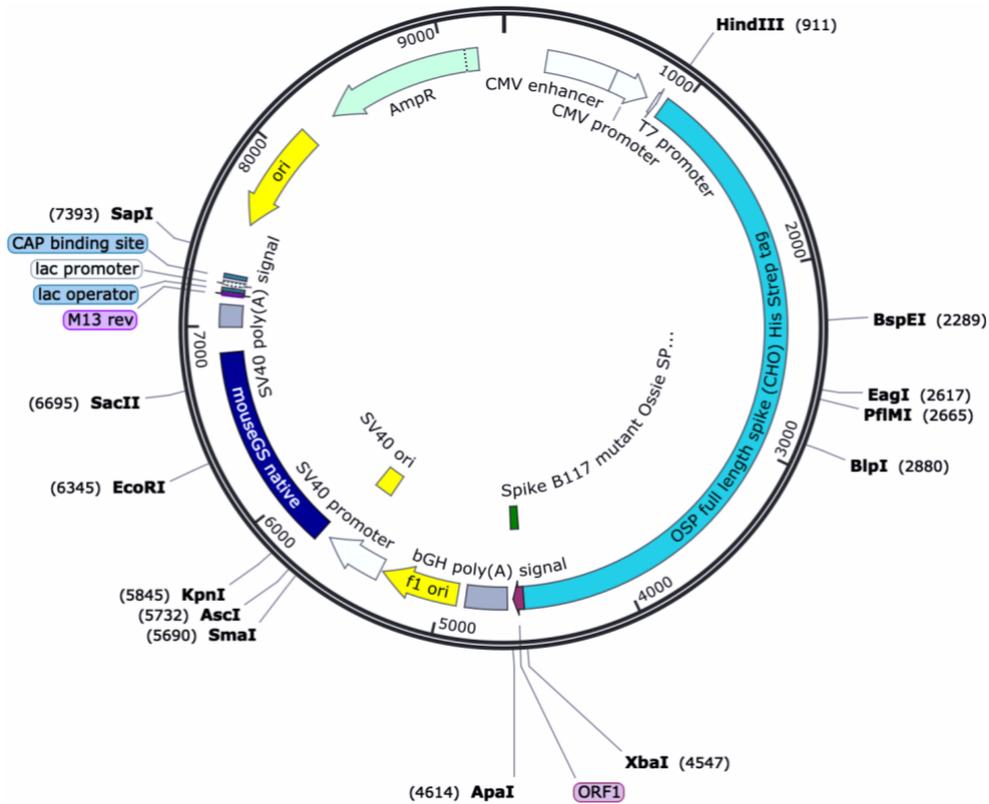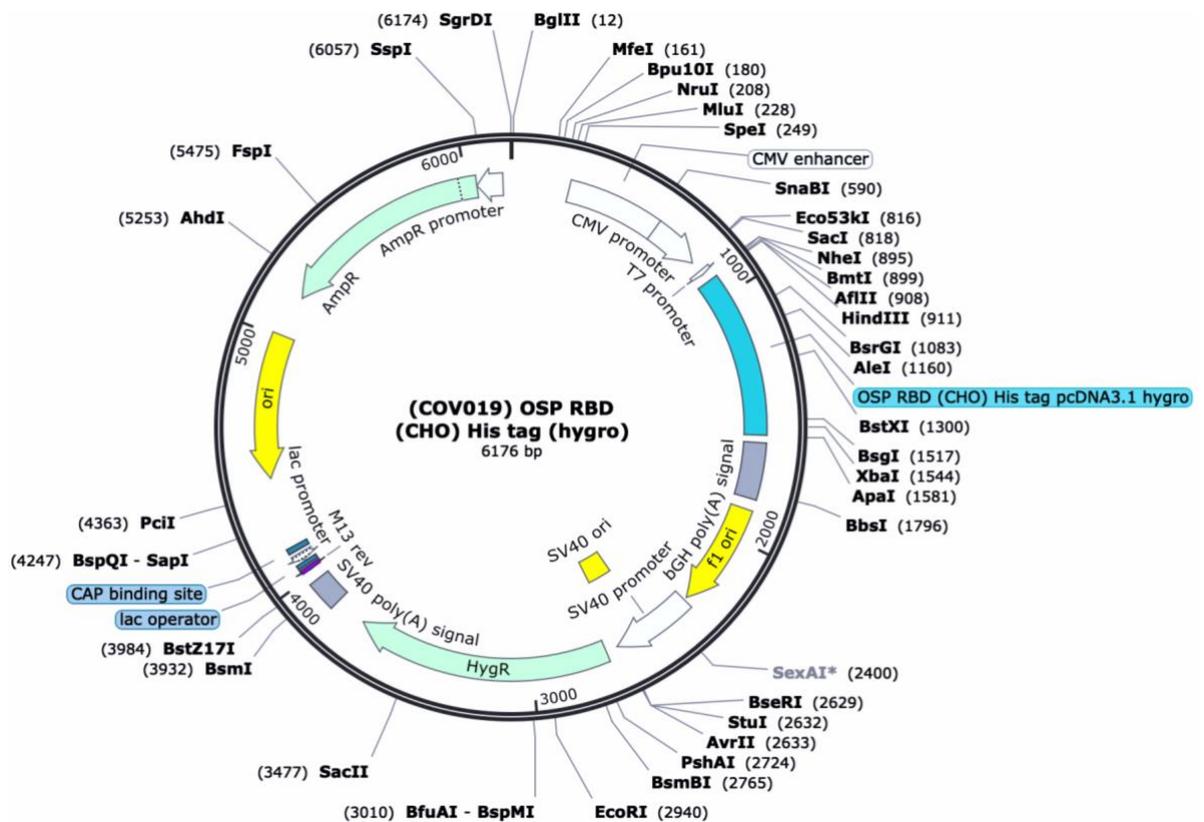**Appendix 5:** *Plasmid Map showing the Nat-FL-His-SpyTag insert in pcDNA3.1hygro vector*
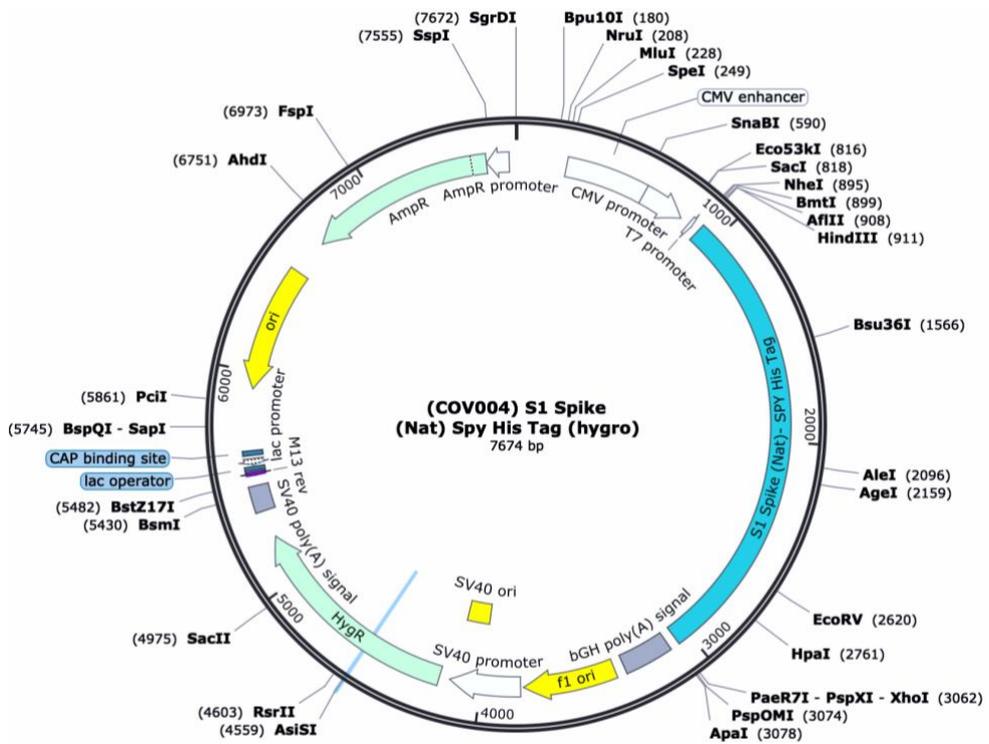
Appendix 6: *Plasmid Map showing the Nat-S1-His-SpyTag insert in pcDNA3.1hygro vector*
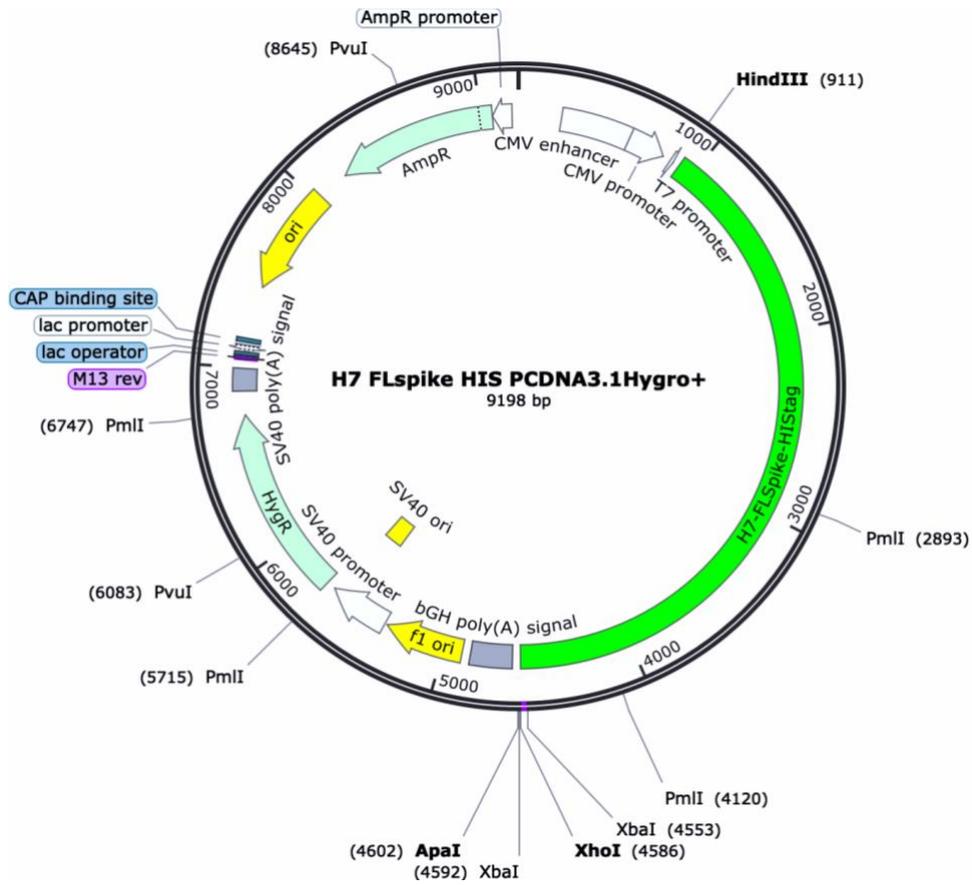
Appendix 7: *Plasmid Map showing the H7-FL-His insert in pcDNA3.1hygro vector*

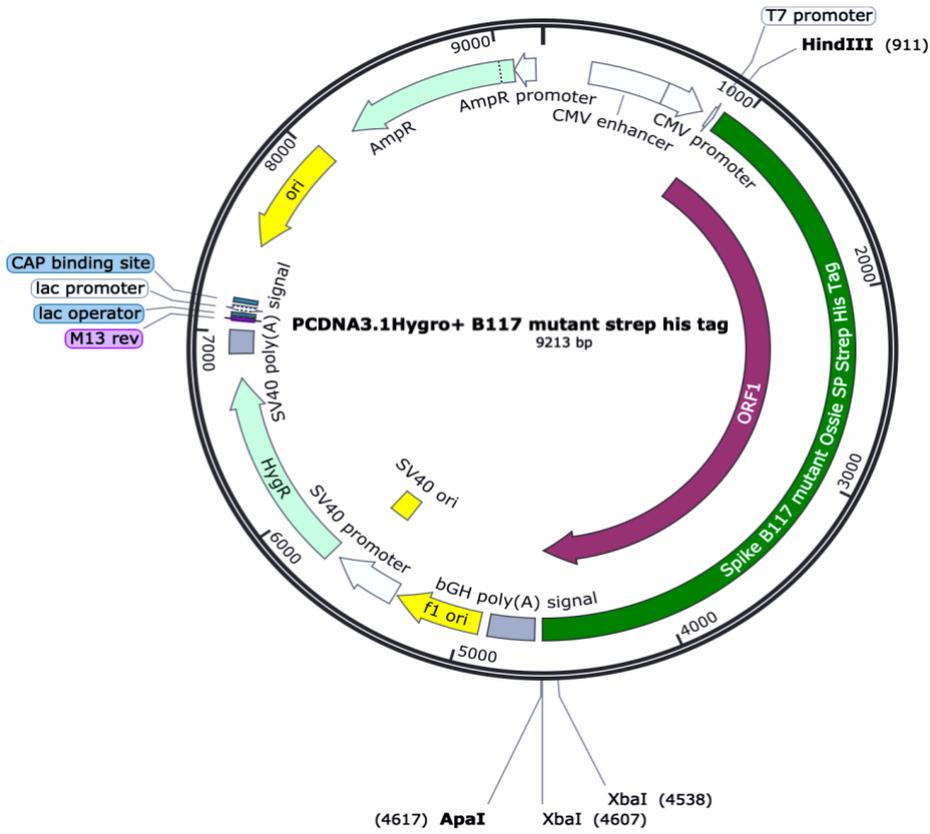**Appendix 8:** *Plasmid Map showing the OSP-FL-His insert in pcDNA3.1hygro vector*
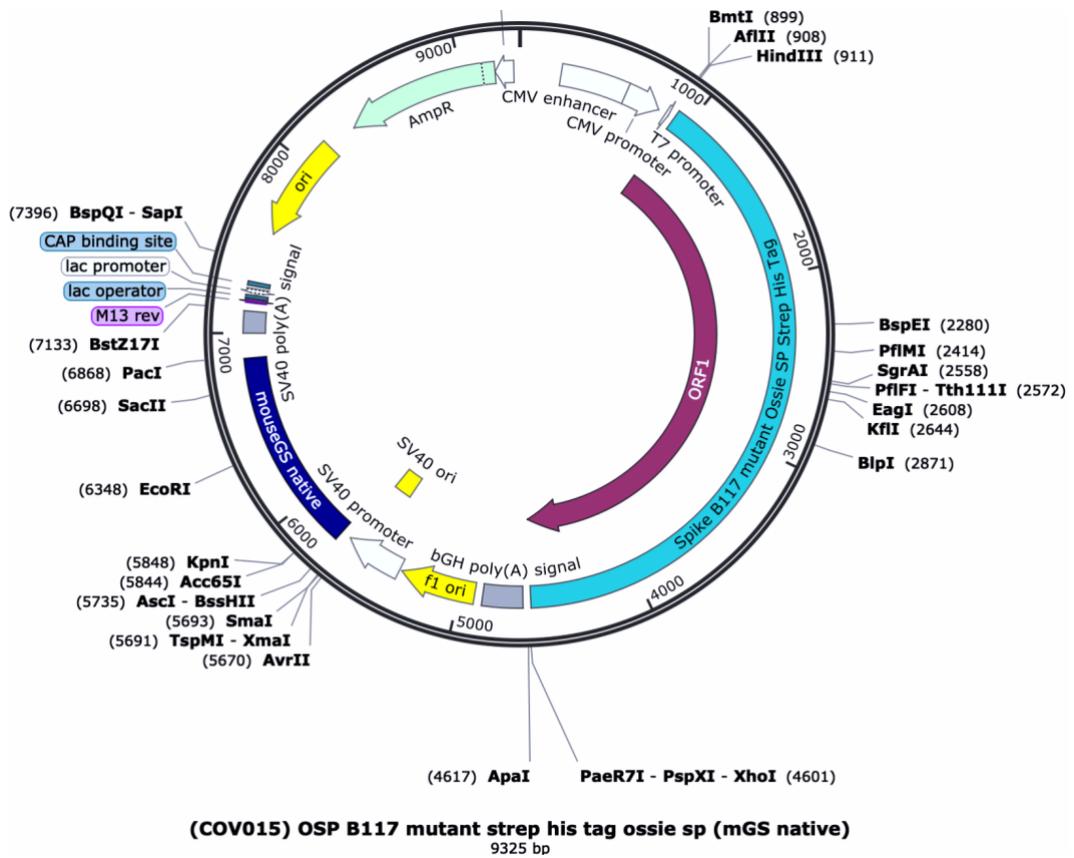
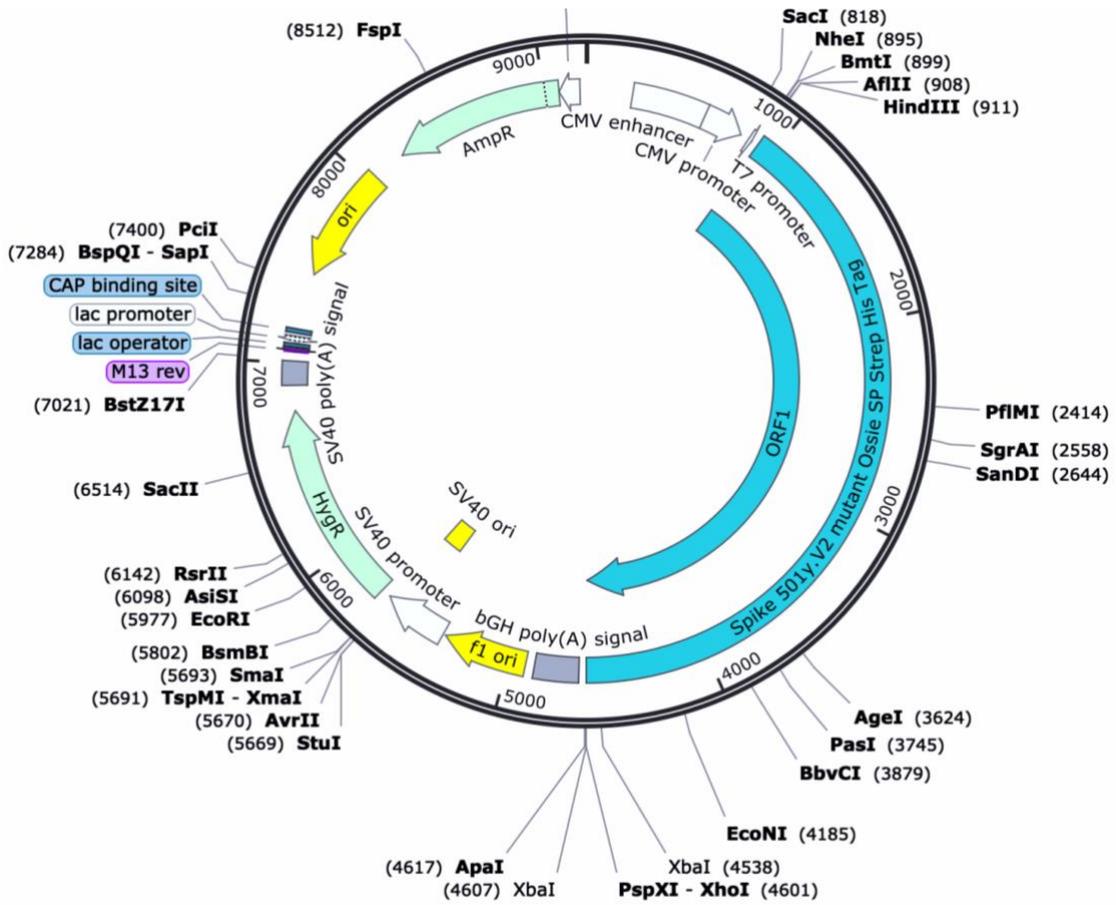**Appendix 9:** *Plasmid Map showing the OSP-FL-His insert in pcDNA3.1mGS vector*

Appendix 10: *Plasmid Map showing the OSP-RBD-His insert in pcDNA3.1hygro vector*

Appendix 11: *Plasmid Map showing the OSP-FL-His-Strep Tag (B.1.1.7) insert in pcDNA3.1hygro vector*

Appendix 12: *Plasmid Map showing the OSP-FL-His-Strep Tag (B.1.1.7) insert in pcDNA3.1mGS vector*

(COV038) OSP 501YV2 mutant His Strep tag (hygro) (1)
9213 bp

Appendix 13: *Plasmid Map showing the OSP-FL-His-Strep Tag (B.1.351) insert in pcDNA3.1mGS vector*

END