Towards Development of Novel Selection Systems that Utilise ER Signal Peptides for the Generation of Recombinant Protein Expressing Mammalian Cell Lines

Harry Seymour

Login ID: hts3

School of Biosciences
University of Kent

A thesis submitted to the University of Kent for the degree of Master of Research in Biochemistry

2021

Supervisors: Mark Smales & James Budge (University of Kent), Colin Jaques (Lonza Biologics)

Declaration

I confirm that the wording of this report is entirely my own. No part of the report has been copied from scientific journals, web sites or any other sources. For a detailed statement on plagiarism the student is referred to the guidelines for preparing project reports.

Signature of the student:

Date: 14/09/2021

Acknowledgements

I would like to thank every member of the Smales lab at the University of Kent for their unconditional support throughout the whole project. I would like to thank Professor Mark Smales and Dr James Budge for their fantastic time and effort during my time in the lab. Finally, I would like to thank Lonza Biologics for funding this research with special mention to Dr Colin Jaques and Dr Sandeep Ranpura who have also helped guide this research.

Abstract

It is estimated that around 30% of eukaryotic proteins enter the secretory pathway with the majority of these polypeptides being directed into the endoplasmic reticulum (ER) via ER N-terminal signal peptides. Such ER signal peptides are utilised for the production of high value biotherapeutic proteins produced in mammalian cells such as monoclonal antibodies. These are directed into the ER to attain the correct and human-like folding, assembly and post-translational modifications before being transited through the secretory pathway and out of the cell from which the product is harvested. ER signal peptides can however, function to do more than just direct polypeptides to the ER and some have secondary functions. In producing mammalian cell expressing cell lines, a key aspect is selection of high recombinant protein producing clones that is usually based upon selection with a metabolic marker whereby the cell cannot survive in the absence of the gene for the metabolic marker. Such systems do not directly select for recombinant protein production but for the activity of the metabolic marker. This study investigated utilisation of the viral Hepatitis C (HepC) ER signal peptide for localising of target polypeptides into the ER. The HepC signal peptide could subsequently be released to undertake other functions such as activate a metabolic marker, thus directly relating polypeptide synthesis of the target protein to the selection marker. Here, proof-of-concept studies of this idea were investigated using a split GFP system whereby two portions of GFP are made independently that are inactive without the presence of each other. In this case, the HepC ER signal peptide was fused to a portion of GFP termed GFP11 such that this would be 'released' and as a 'complementation' fragment (eGFP11) active eGFP1-10 that was separately transfected into cells. SDS-PAGE and western blotting suggested that the HepC signal sequence was able to direct cleavage of the polypeptide to release the GFP11/opt-HepC from an mCherry protein that was secreted out of the cell. Further, microscopy studies revealed that GFP expression was only observed in cells where mCherry was also present, confirming that the expression of GFP was directly linked to mCherry expression as would be desired for a recombinant target protein. Ultimately the complementation component (GFP11) would be replaced with a selection marker or a complement fragment of a selection marker such that cells could not grow in the absence of the activated selection marker. The work provides the basis for a new and novel approach to generating recombinant mammalian cell lines that directly links metabolic marker activity in a cell to the amount of recombinant biotherapeutic protein being produced. This should provide a stronger correlation between selection and recombinant product selection than current systems.

Contents

Declaration 2
Acknowledgements 3
Abstract 4
Abbreviations 7
1.0 Introduction 8
1.1 Compartmentalization and the Secretory Pathway 8
1.2 Entry of Polypeptides into the ER 9
1.3 Mammalian cell line development 11
1.4 The properties of ER signal peptides 15
1.5 ER signal peptide secondary roles and functions 18
1.6 The Hepatitis C virus ER targeting signal sequence 21
1.7 The split GFP system and mCherry reporter proteins 25
1.8 Project aims 27
2.0 Materials and Methods 28
2.1 Reagents for molecular biology procedures 28
2.1.1 Preparation of ampicillin stocks and LB plates with ampicillin 28
2.1.2 Preparation of calcium competent DH5α *E.coli* cells 28
2.2 Molecular cloning of genes of interest into mammalian expression plasmids 29
2.2.1 Cloning of Split GFP and mCherry into the 3.1 hygro and Lmm154 vectors 32
2.2.2 DNA agarose gel electrophoresis analysis and purification of DNA from agarose gels 33
2.2.4 DNA ligation and transformation of competent bacteria with plasmid DNA (constructs 1-10) 34
2.2.5 Golden gate assembly and transformation of competent bacteria with plasmid DNA (constructs 11-14) 35
2.2.6 Plasmid DNA purification from *E.coli* cultures 36
2.3 Tissue culture 37
2.4 Stable cell pool generation 38
2.5 Transient transfections 39
2.5.1 Harvesting transient transfections 40
2.6 Analytical Methods for Analysis of Protein Expression 40
2.6.1 SDS PAGE and Western Blotting 41
2.6.2 Confocal Microscope Slide Preparation 42
2.6.3 ZEISS LSM 880 Elyra Confocal Microscope Analysis 42
3.0 Results 44
3.1 Generation of split GFP and HepC signal peptide plasmid DNA constructs 44
3.1.1 Sanger sequencing of DNA constructs 48
..... 50

3.2 Batch culture growth analysis of stably expressing Split GFP1-10 CHOK1SV GS-KO® cells 51
3.3 Transient transfection of the split GFP constructs and eGFP11-HepC- signal peptide- mCherry constructs with eGFP1-10 constructs 52
3.3.1 Confocal microscopy image analysis of transient transfections 53
3.4 mCherry and GFP protein analysis following transfection by western blotting 61
4.0 Discussion 72
4.1 Confirmation of the successful generation of split GFP and HepC signal peptide plasmid DNA constructs 74
4.2 Batch culture growth analysis of stably expressing split GFP1-10 CHOK1SV GS-KO® cells 75
4.3 Confocal microscopy image analysis of transient expression of mCherry, GFP and the split GFP system 75
4.4 mCherry and split GFP protein analysis following transient expression by western blotting 79
4.5 Future work 81
4.6 Conclusions 84
5.0 References 86
7.0 Appendix 95

Abbreviations

ER – endoplasmic reticulum

CHO – Chinese hamster ovary

SP – signal peptide

SPPs – signal peptide peptidases

MTX – methotrexate

GS – Glutamine synthetase

GS-KO – Glutamine synthetase knock out

DHFR – dihydrofolate reductase

LCMV – Lymphocytic Choriomeningitis Virus

GP – glycoprotein

HIV – human immunodeficiency virus

HepC – Hepatitis C virus

IRES – Internal ribosome entry site

LB – Lysogeny broth

SOB – Super optimal broth

Hygro – hygromycin (antibiotic)

PCR – polymerase chain reaction

mAb – monoclonal antibody

1.0 Introduction

1.1 Compartmentalization and the Secretory Pathway

Compartmentalization is fundamental to all eukaryotic cells in the form of organelles, where specific functions of the cell are kept compartmentalized from each other. These specific conditions allow proteins/enzymes to perform their roles efficiently (1). The areas within the cell are separated by phospholipid bi-layer membranes that encompass and maintain the integrity of each compartment. However, for the cell to operate, molecules need to move in and out of these organelles such as mRNA, where these transcripts move from the nucleus to the cytosol where they are translated. Proteins and metabolites also move into and out of the nucleus, mitochondria, Golgi and endoplasmic reticulum (ER).

Entry of proteins into the nucleus, mitochondria and ER is initiated via signal peptides found within the immature or mature polypeptide/protein. Signal peptides recognised for entry into the nucleus are found on the surface of the mature protein and may be located anywhere within the protein sequence (2). On-the-other-hand, polypeptides/proteins destined for the ER or mitochondria are almost always located at the N-terminal, recognised by specific proteins/receptors that facilitate entry into the organelle and then tend to be processed/cleaved off the mature protein. For proteins that are to enter the secretory pathway and destined for the ER, Golgi, plasma membrane or to be secreted out of the cell, the first point of entry is the ER (3). The ER is an organelle that contains a specific subset of proteins, enzymes and redox potential that facilitates the oxidative folding and assembly of proteins, and initial N-glycosylation, that may be retained within the ER by attachment of a KDEL amino acid sequence at their C-terminal end or subsequently transported in vesicles to the Golgi for further processing and secretion (4, 5, 6). Indeed, biotherapeutic proteins made in mammalian cell expression systems, such as monoclonal

antibodies, are produced in these systems because the secretory pathway contains the required enzymes, proteins and environments to successfully fold, assemble, post-translationally modify and secrete such proteins at high yields (>5 g/L). These pathways produce mature proteins with human-like PTMs and activity (7).

1.2 Entry of Polypeptides into the ER

The entry of polypeptides into the secretory pathway is through the ER which can in most cases occurs co-translationally as the polypeptide is synthesised on the ribosome although can occur post-translationally. Indeed, it is estimated that around 30% of eukaryotic proteins enter the secretory pathway and this results in the so called rough-ER whereby actively translating ribosomes are ‘tethered’ to the ER as the polypeptides are co-translationally fed into the ER (3). The majority of these polypeptides are directed into the ER via ER N-terminal signal peptides; as the polypeptides are synthesised the N-terminal signal peptide is recognised by the signal recognition particle (SRP), which upon binding, attenuates/stops translation and hence further polypeptide synthesise. The SRP-then recognises and binds to the SRP receptor (SR) on the surface of the ER upon which the ribosome binds to the Sec61 translocon channel (in mammalian cells) allowing the nascent polypeptide to enter the channel. The SRP dissociates and polypeptide synthesis restarts, feeding the growing polypeptide into the ER through the channel where it is bound by chaperones such as the protein Bip until folded (this prevents mis-folding and aggregation). The signal peptide is then cleaved off to result in the mature protein; this can occur either co-translationally or post-translationally. The signal peptide is then often degraded but, in some cases, can undertake further roles in the cell (3). A more detailed review of the process, the protein players

involved and how N-linked glycosylation is initiated during this process is described in Gemmer & Forster (2020).

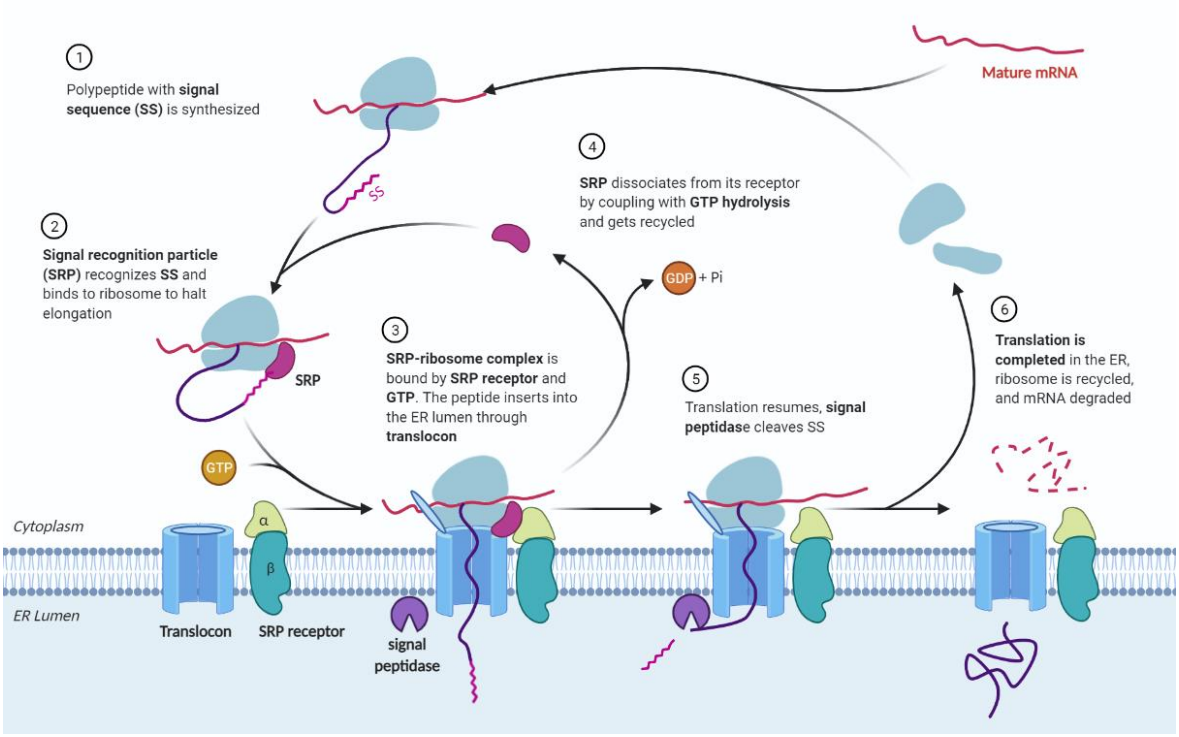


Figure 1.0 ER protein targeting pathway adapted from "Protein Targeting with Signal Recognition Particle" by BioRender.com (2021). Retrieved from https://app.biorender.com/biorender-templates application 1) Mature mRNA is translated by a ribosome, 2) the signal recognition particle (SRP) recognizes the signal sequence (SS) and binds to the ribosome, pausing translation, 3) the SRP receptor recognizes SRP, allowing the ribosome to bind the translocon and the polypeptide to translocate into the ER lumen, 4) the SRP dissociates from the complex, coupled with GTP hydrolysis, 5) translation resumes and the SS is cleaved by a signal peptidase, 6) translation completes in the ER.

The signal peptides themselves are short sequences (15-30 amino acids) found at the N-terminal of secretory proteins that direct the polypeptides into the secretory pathway and onto target organelles and membranes. As described above, an important feature of SPs is that they are cleaved by signal peptidases after they have delivered their cargo and therefore not part of the final, mature protein. Signal peptides were first reported by Blobel and Sabatini in 1971 (8), and throughout the next 30 years, their work earned them a Nobel prize in Physiology for the discovery of these ER signal peptides. As outlined above, a typical signal peptide has 15-30 amino acids but variation is seen in eukaryotes and viral signal peptides where

lengths of up to ~150 amino acid residues have been observed (9). Importantly for this work, other studies have shown that signal peptides can have secondary functions post cleavage (10).

The work in this thesis describes the development of a signal peptide selection system that might be harnessed to select for mammalian cells expressing recombinant or therapeutic proteins of interest. The work set out to demonstrate that ER signal sequences can be used to select for, and/or provide selection pressure for, recombinant CHOK1SV GS-KO® cells expressing a model reporter protein (GFP). This would then provide proof-of-concept of the utilisation of such signal peptides that could be further developed for selection of high expressing recombinant biotherapeutic expressing cells. The basis of the study was to identify and investigate if signal peptides could be used to complement and 'activate' an inactive eGFP molecules. Current favoured metabolic selective pressure strategies work by linking a metabolic enzyme to cell survival during cell line development whereby cells that have not taken up the metabolic gene of interest cannot survive in the absence of the metabolite the enzyme is involved in making. This means that selective pressure is for the transcription and translation of the selective metabolic marker, not the desired product. The product gene could be poorly expressed or progressively silenced without removing the proliferative advantage obtained from the selective marker and the product may fail to translate without reducing the proliferative advantage obtained from the selective marker.

1.3 Mammalian cell line development

Since Chinese hamster ovary (CHO) cells were first describe by Puck in 1957 (11) and subsequently development in the 1980's and beyond for recombinant biotherapeutic protein production, this expression system has become the dominant mammalian cell expression system for these molecules, particularly for the

manufacturing of monoclonal antibodies (mAbs) (12). There are several reasons why CHO cells have been applied so successfully to the manufacturing of biotherapeutic proteins; CHO cells can adapt and grow in suspension culture which is ideal for large scale culture (13), they can grow in chemically defined and serum and protein-free media ensuring reproducibility between different batches, they can undertake human-like post translational modifications to recombinant proteins (14), particularly N-glycosylation that is human-like, and does not contain the immunogenic α-galactose epitope (15), and the hosts and bioprocesses have been continually optimised such that yields in excess of 5 g/L are regularly reported, particularly for mAbs (7).

The most popular industrially applied cell line selection technologies are based on either the dihydrofolate reductase (DHFR) (16), or glutamine synthetase (GS) system (17). These systems harness the fact that cells grown in the absence of either (a) hypoxanthine and thymidine cannot survive with insufficient DHFR present, an enzyme that catalyses the conversion of folate to tetrahydrofolate, or (b) glutamine cannot survive with insufficient glutamine synthetase, the enzyme response for synthesis of glutamine from glutamate and ammonia (18, 19). When these metabolic selection systems were first used, both used a specific drug to inhibit the selectable enzyme marker essential for cellular metabolism: MTX inhibits dihydrofolate reductase (DHFR), and methionine sulphoximine (MSX) is the inhibitor in the GS system. Significant improvements to the GS system has been made, including the development of GS knockout cell lines such as the Lonza CHOK1SV GS-KO® cell line, which has the GS gene ‘knocked out’ in the genome and thus GS activity can be restored by the gene being reinserted on the same plasmid as the target protein, allowing the cells to grow in glutamine free medium (20, 17).

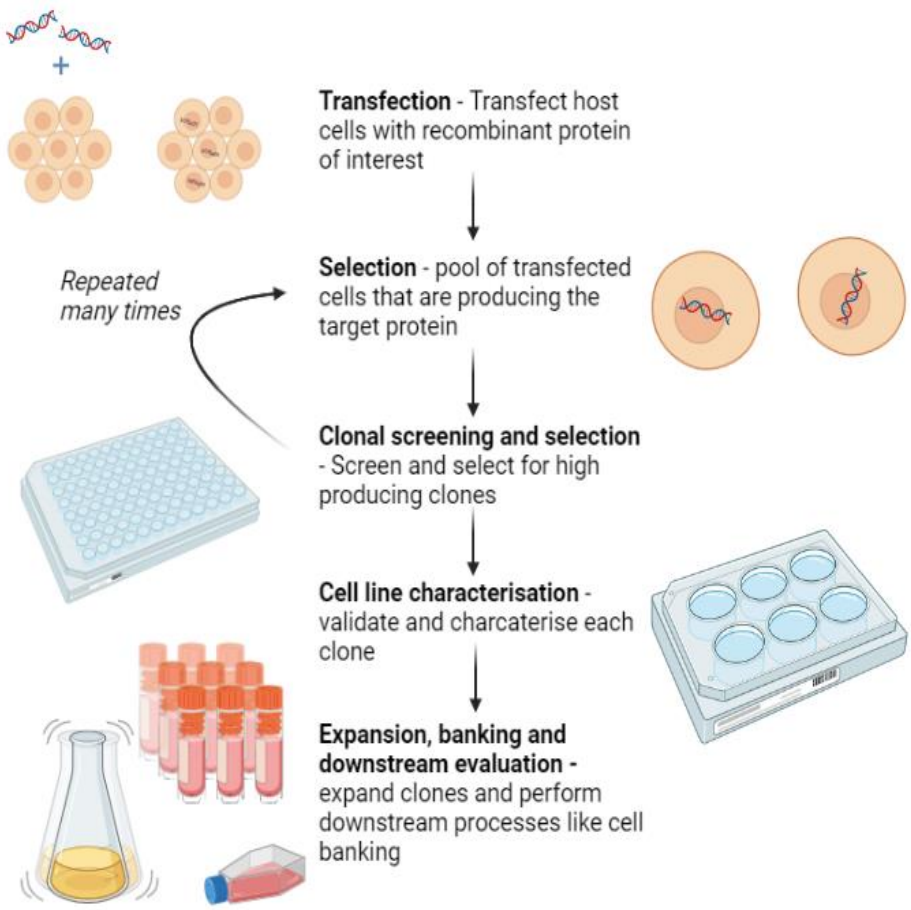


Figure 1.1 Diagram showing a standard process for developing cell line with high producing clones. Clonal selection and screening are undertaken several times to isolate a high producing and stable clone that can then be characterised and expanded for production. Created with BioRender.com.

Traditionally there has been random integration of plasmid DNA into the cell genome of plasmids carrying selection markers and the recombinant product genes of interest. As a result, transcriptional heterogeneity between clones can exist as the chromosomal environment has a strong influence on promoters and therefore the rate of transcription (21). High producing clones are therefore rare but essential for commercial application, and so many thousands of clones need to be screened which is expensive and time consuming. Although site specific integration systems are now available that give less diverse transcriptional activity across cells for the genes integrated into the target site, differences in the translation of the marker and product transcripts can mean the relationship between the amount of selection marker expression and the target molecule are not necessarily linear (22).

There have been some reports and technologies utilised such as Internal ribosome entry sites (IRES) to alleviate the problems with expression of the product and marker genes on the same vector, however these only marginally improve the reliability of selection of product expression, as the marker and product are still separately translated. An IRES sequence can be inserted between 2 genes, for example selection marker and target recombinant protein, and allows transcription of both genes to be dependent on the same promoter as only one transcript is produced and thus 2 mature proteins can be translated from the same, single, mRNA (23). However, the proteins are still translated as two independent polypeptides with different efficiencies and hence this does not address directly linking selection marker expression to target protein at the polypeptide level.

This study is ultimately aimed at improving the upstream selection process making selection and isolation of cells based on productivity directly linked at the polypeptide level. Whilst there have been various studies aimed at trying to improve the secretion efficiency of a recombinant biotherapeutic antibody by optimizing the ER signal peptides (24, 25, 26), and some of these studies have shown that some signal peptides are better at directing polypeptides to the ER than others, to our knowledge no one has looked at utilising the signal peptide to directly link selection with target protein polypeptide synthesis. The work here describes the investigation of the hypothesis that the ER signal peptide could not only be used for directing the target recombinant polypeptide to the ER, but could also be harnessed to activate a selection marker, directly linking selection marker activity with target recombinant product production.

1.4 The properties of ER signal peptides

A typical N-terminal ER signal peptide is divided into 3 main sections: The N-region: a positively charged domain, H-region: the hydrophobic centre, and C-region: the cleavage site (Figure 1.2). These regions are conserved in terms of their properties between ER signal peptides but there is little sequence conservation between ER signal peptides. The pro-region, after the C-region, is part of the final mature protein and not the SP, but does influence signal peptidase binding and therefore cleavage at the correct site to ensure the mature protein has the authentic and correct N-termini (27). This configuration is generalised for ER signal peptides, but the number of and specific residues, and the lengths of a SP, are highly variable and contribute to the capacity for polypeptide translocation.

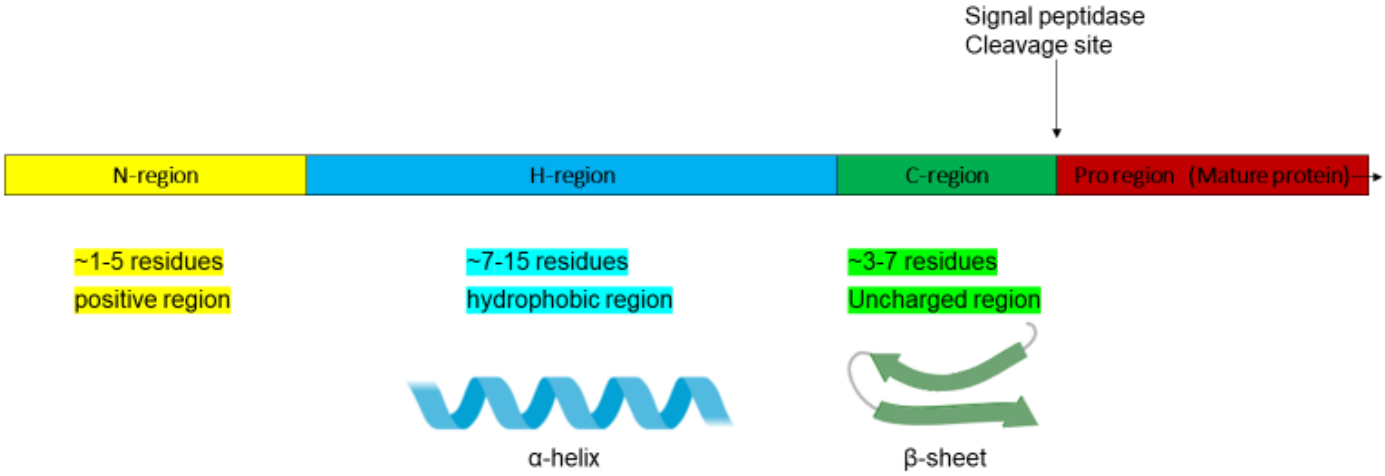


Figure 1.2 Schematic representation of a typical ER signal peptide. The positive N-region is shown in yellow, the hydrophobic H-region in blue, the uncharged C-region in green and the pro-region in red. The signal peptidase cleavage site is also indicated in the diagram as well as the features of each region such as the α-helix nature of the hydrophobic region and β-sheet nature of the C-region in the uncharged region. This schematic is of a typical signal peptide but there are examples of viral signal peptides that are over 100 residues in length which have an extended H-region.

The N-region of the ER signal peptide is at the N-terminal of the SP and is characterized by the occurrence of basic amino acids (K and R) and is usually around 5 residues in length. The basic residues means that the N-region is responsible for interactions with the anionic phosphate headgroups of the phospholipid bi-layer and is crucial for the translocation of polypeptides into the ER

membrane. The basic residues in the N-region create a positive charge compared to the rest of the signal peptide, controlling the balance of positive charge across the signal sequence and determines the orientation of the signal peptide, increasing or reducing translocation efficiency (28). In particular, the charge difference between the N- and C-region is important for efficient cleavage in the C-region and ensuring appropriate interaction with the signal peptidase. This is important to note, especially when using signal peptides for biotherapeutic production. Caspers *et al* (29) found that 3 point mutations in the AmyE signal peptide, which reduced the net charge from +3 to +2, increased quantities of cutinise (the product they were trying to produce) one- to two-fold compared to the wild type. The reduced net charge prevented the jamming of the secretion translocase and induction of stress genes (29, 26). In another study, Haryadi *et al* (24) investigated 5 top selling therapeutic antibodies and the impact of the signal peptides selected from a library on the final secreted productivity or yield of mAb. They showed that an 'optimized' signal peptide for one of the mAbs, Rituximab, was 2-fold better compared to its native signal peptides in terms of the yield. They also showed that substitution of a single amino acid in the optimized heavy chain signal peptide for another mAb, Avastin, reduced its production significantly. Thus, care must be taken when modifying ER signal peptides for alternative processes that this does not negatively impact upon yield or cleavage efficiency of the signal peptide by the signal peptidase.

The H-region of the ER signal peptide consists of a hydrophobic region within the SP that is necessary for spanning the phospholipid bi-layer. It is usually 7-15 residues long, but many viral signal peptides have extended H-regions. For example, $SP^{GP\text{-}C}$ is a signal peptide involved in the maturation of glycoprotein C and comes from the Lymphocytic Choriomeningitis Virus (30). It is comprised of 58 amino acid residues and contains an extended hydrophilic N-terminal region, two

hydrophobic regions, and a short C-terminal region. After cleavage by signal peptidase, SP^{GP-C} accumulates in cells. This LCMV SP^{GP-C} was shown as being an essential component of the glycoprotein complex and the different regions of SP^{GP-C} are required for virus maturation and infectivity (30).

The hydrophobic H-region of the signal peptide is also important in determining the conformation of the signal peptide and its orientation towards the cell membrane. Furthermore, it also affects cleavage, influencing the rate and efficiency of protein translocation. Usually as hydrophobicity decreases, so does the signal peptides protein processing ability. This was described in a study by Mordcovich *et al* where mutations in the H-region of pro-enterotoxin B signal peptide from *Staphylococcus aureus*, which led to a reduced hydrophobicity, abolished secretion of the mature protein (31). Moreover, increasing the hydrophobicity, as reported by Zhang *et al,* can result in increased secretion of recombinant proteins (32). However, it is not just as simple as this, as if hydrophobicity increases beyond a certain point, it can cause formation of immature proteins with improper folding jamming behind and blocking the translocase (27).

The C-terminal part of the ER signal peptide is called the C-region and usually consists of ~3–7 neutral or polar amino acids. This is the most conserved section of signal sequences compared to the other sections which can be highly variable not just in length but amino acids. Cleavage efficiency, and therefore protein secretion, is dependent on this C-region of the signal peptide. In addition, the sequence in the C-region needs to cleave efficiently to release a native polypeptide with no extra amino acids at its N-terminal end. Five residues are reportedly optimal for the highest signal peptide cleavage efficiency and if the length is increased to more than nine residues, this greatly hinders or abolishes secretion totally (33).

Increasing the C-region length means the cleavage site is further from the membrane surface and recognition by the peptidase is compromised. It is common for the C-region to be composed of residues forming an extended β-conformation, providing the binding site for a signal peptide peptidase. Position 1 and 3 prior to the cleavage site are considered the critical points of the C-region. These points are known as the (−3, −1) rule or AXA motif (34). This cleavage site rule is universal in prokaryotic, ER, and organellar signal peptides. The occurrence of AXA motif through analysis of 1877 eukaryotic, 475 prokaryotic signal peptides was 14.5% and 52.95% respectively (35), showing this motif is much more conserved in bacteria than eukaryotes.

It is also important to consider the pro-region of the ER signal peptide. This is not part of the signal sequence itself, but rather a few of the residues before or at the start of the mature polypeptide. These residues can influence protein export and the cleavage efficiency.

1.5 ER signal peptide secondary roles and functions

Signal sequences (signal peptides) play an important role in targeting and membrane insertion of secretory proteins in both prokaryotes and eukaryotes. But many studies have found that signal peptides are, as Martoglio puts it in his review, “more than just greasy peptides” (10). Section 1.4 outlines a typical ER targeting signal peptide structure, as well as some signal sequences with an unusual composition. These type of signal peptides, as well as typical structural signal peptides, often have secondary roles and contain specific information for performing distinct functions in targeting and membrane insertion and more importantly to this study, have roles post cleavage from the mature protein.

For example, the signal sequence of the HIV-1 envelope protein (p-gp160) is able to bind to calmodulin in a Ca^{2+} dependant manner due to a sequence in its N-region (Figure 1.3) (36). The N-regions of this signal peptide can form a basic amphiphilic α-helix and have a hydrophobic C-terminal region, they are like other calmodulin-binding domains (37). Calmodulin is a protein involved in the regulation of cellular processes controlled by Ca^{2+}-dependent signalling pathways (37). It is likely that the binding of p-gp160 signal peptide fragments to calmodulin influences Ca^{2+}/calmodulin-dependent cellular signal-transduction pathways (36).

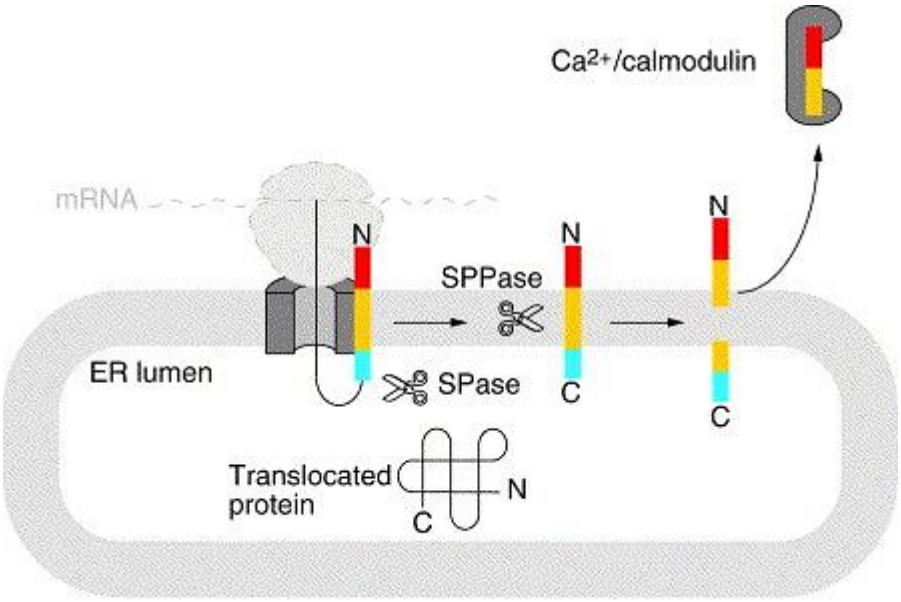


Figure 1.3 A figure adapted from Martoglio's review (10). The secondary function of the signal sequence of preprolactin from the HIV virus. The signal peptide is cleaved by signal peptidase (SPase) and further processed by signal peptide peptidase (SPPase). The N-terminal signal peptide fragment (yellow and red) is released into the cytosol where it interacts with Ca^{2+}/calmodulin.

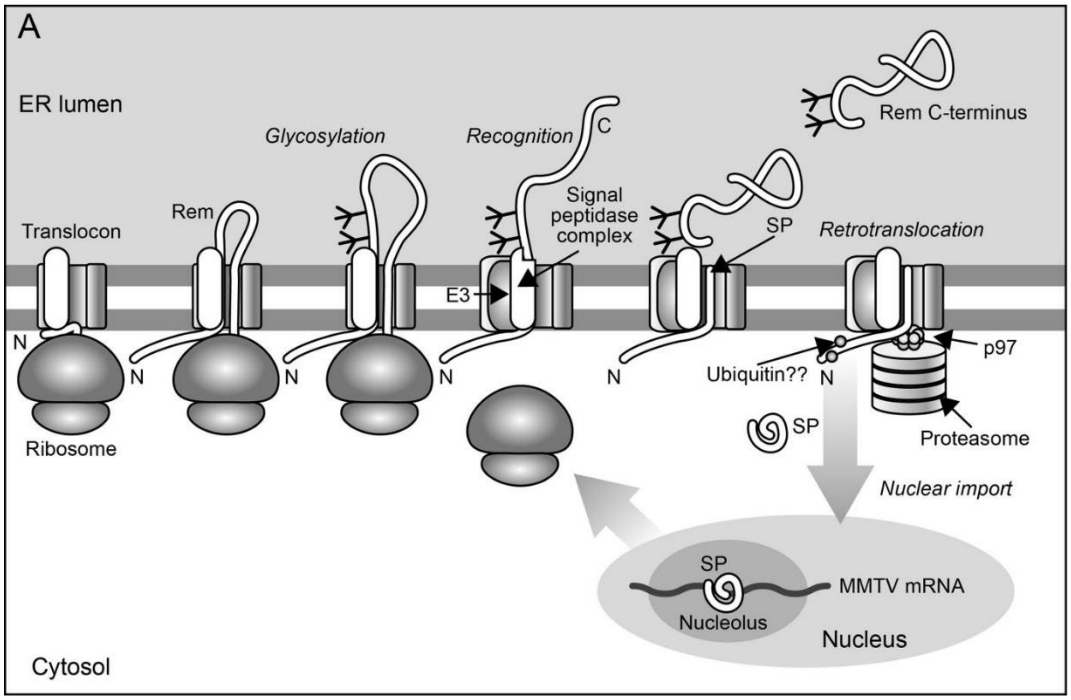


Figure 1.4 A figure adapted from Byun's study (38). The Rem processing and trafficking pathway. The misfolded state of Rem triggers association with the retrotranslocon. Cleaved MMTV-SP escapes proteasomal degradation to allow nuclear entry and targets to the nucleus prior to MMTV RNA binding. The Nuclear localisation signal within MMTV-SP provides a bridge between the viral RNA and the Crm1 export protein for nuclear export of viral RNA.

Another example of a post-cleavage secondary role is the signal peptide of the mouse mammary tumour virus rem protein (Figure 1.4) (39). The mouse mammary tumour virus alternative splice variant Rem (envelope protein) has an uncommonly long signal sequence, which contains a nuclear localization signal. The signal peptide facilitates targeting to the ER, where the C-terminal portion is translocated and glycosylated. The signal sequence is cleaved off, but then accumulates in nucleoli. This suggests a novel pathway whereby the signal sequence can be released from the ER and then fulfil a post-targeting function in the nucleus (39, 40) (38). The signal peptide is said to act as a nuclear export factor, binding to mRNA transcripts of Rem and Env (41).

1.6 The Hepatitis C virus ER targeting signal sequence

A signal sequence used by the Hepatitis C (HepC) virus was identified as a potential target in this project. It has unique features that could be utilised to create a direct link between selection pressure and target protein production. The HepC viral RNA genome is recognized by host ribosomes due to an internal ribosomal entry site (IRES) sequence located at the 5′ end to mediate viral genome translation. The translated polypeptide chain is about 3000 amino acids in length and is both co- and post-translationally processed by host and virus-encoded proteases at the endoplasmic reticulum (ER) to generate the individual mature forms of the viral proteins (Figure 1.5).

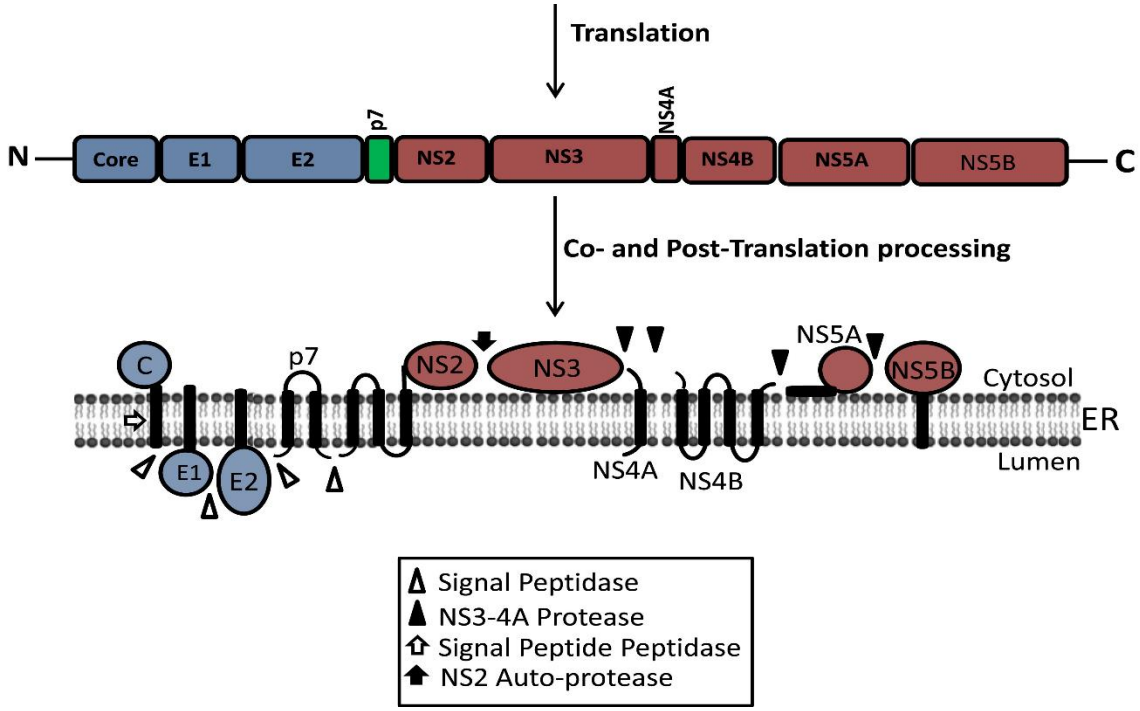


Figure 1.5 The Hep C virus in its polypeptide form adapted from Atoom's study looking at the p7 protein (42). The sections in blue represent the structural proteins and the red represent the non-structural proteins. The legend also indicates signal peptidases, signal peptide peptidases and proteases.

The core protein (C in Figure 1.5) at the N-terminal forms the capsid that encapsulates the HCV viral RNA. E1 and E2 are envelope proteins that are highly glycosylated and are important in cell entry (43, 42). The signal sequence is situated between the core and E1 proteins and therefore is known as the core-E1 signal peptide. This signal sequence is unique as it serves both the core protein and the E1 glycoprotein. The core-E1 signal peptide undergoes an initial cleavage by the cellular enzyme signal peptidase which releases the E1 protein into the ER. A second proteolytic event mediated by the intramembrane protease signal peptide peptidase (SPP) is necessary for the maturation of the core protein and its release from the ER membrane into the cytosol (Figure 1.6) (44).

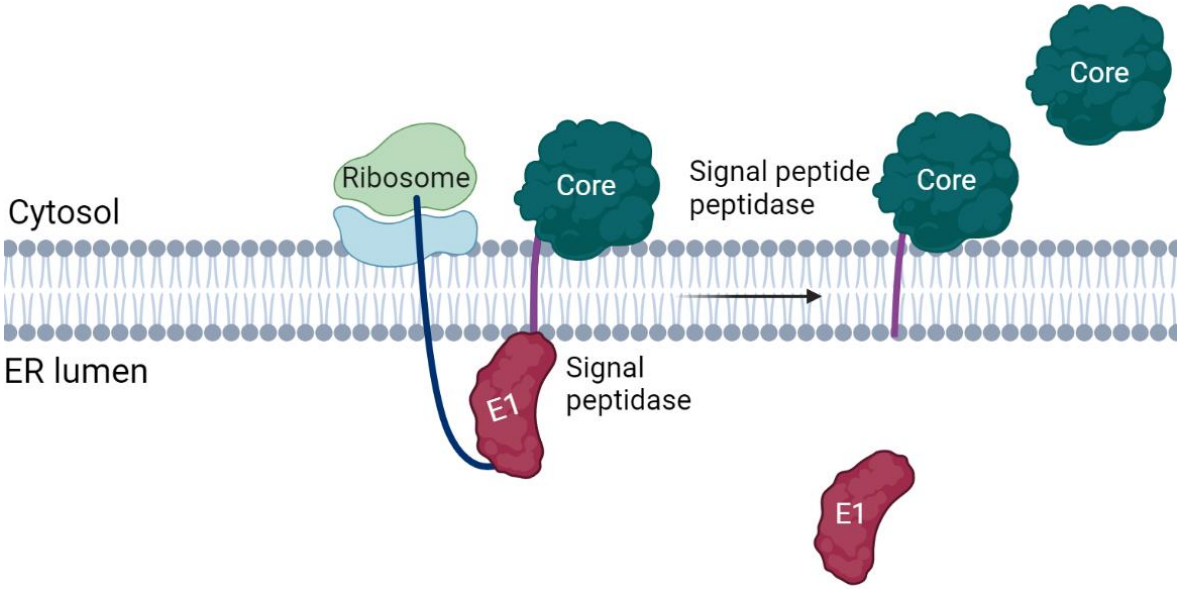


Figure 1.6 A schematic showing the Core-E1 signal peptide cleavage events. There is an initial cleavage event via signal peptidase that releases the E1 protein into the ER and then this is followed by a second cleavage event by the signal peptide peptidases that release mature the core protein and release it into the cytosol. The purple shows the core-E1 signal peptide anchored to the ER post signal peptidase action, ready for signal peptide peptidase cleavage (45).

The core-E1 signal sequence exhibits lipid-binding properties (binding to the ER), and this binding is necessary for its folding. It has a typical secondary structure for a signal sequence (Figure 1.7). Hence, the N-region between amino acids 171-174 is unfolded, and the segment between residues 175-186 forms an α-helix and is the hydrophobic H-region; the amino acids from 187-191 are more polar than the H-region and forms the cleavable C-region including the recognition site for a signal peptidase, which has a standard requirement of small neutral residues at positions −1 and −3 relative to the cleavage site (PASA) (43). Compared to classical signal peptides, the core-E1 signal peptide is shorter and does not have typical positively charged residues in the N-domain. These usually interact with the ER membrane to allow the signal peptide to be positioned within the membrane (43). The extended conformation propensities of the flanking N- and C-regions (located between amino acids 176 and 182) may overcome the limitation of the short Core-E1 signal peptide length, in order to facilitate sliding action within the membrane and therefore the accessibility of the cleavage site at positions 191-192 to the signal peptidase on the

luminal side of the ER membrane. Post signal peptidase processing, the sliding of the signal sequence in the membrane due to the lack of charged amino acids in the N-region may help the recognition of and interaction with the signal peptide peptidase (43).

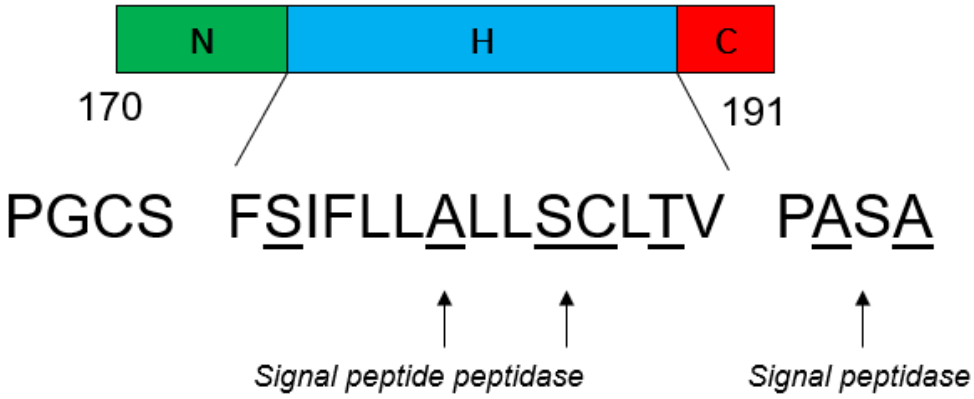


Figure 1.7 Amino acid sequence of the core-E1 signal peptide region from the hepatitis C virus. The amino acids underlined show important residues that affect the binding and cleavage of the signal peptidase and the signal peptide peptidase.

The H-region also contains polar residues that are atypical for signal sequences. The Ser/Cys doublet at positions 183 and 184 are conserved across all Hepatitis C virus genotypes. In addition Ser175 and Ala180 are also conserved, suggesting that these residues have a critical function (43). Further, the flexibility of the H-domain helix is an important characteristic for signal peptide peptidase cleavage, allowing the appropriate positioning of the signal peptide so signal peptide peptidases can access the cleavage site. The polar residues Ser175, Ser183, and Thr186 on the same helix face in the H-region might also be important for signal peptide peptidase recognition and directing the signal peptide to the enzyme's active site (43).

The core-E1 SP can direct proteins to the ER despite not being at the N-terminal and has a double cleavage event releasing two mature proteins. This would potentially allow a direct link between selection pressure and recombinant protein

produced, with the core protein for example replaced by the GS protein that would be released into the cytosol and the E1 protein replaced with the target recombinant polypeptide. In order to investigate this, this study used fluorescent maker proteins in place of actual recombinant protein targets (such as antibodies and selection markers like glutamine synthetase) in order to show that the ER targeting sequence could be used to direct polypeptides to the secretory pathway and release fragments that could complement an activate a reporter system. The core and E1 proteins would be replaced with the fluorescent markers that can be viewed under the microscope and also easily detected via SDS-PAGE and western blot.

1.7 The split GFP system and mCherry reporter proteins

Although the core protein of the HepC virus is moderate in size (21 kDa) (46), an even larger GFP or mCherry molecule, which are around 28 kDa in size, could affect the insertion into the ER membrane or the cleavage itself. This was remedied by using a split GFP molecule as outlined in the methods section.

Molecular engineering of the green fluorescent protein (GFP) has revolutionized the field of biosensor development and the use of fluorescent markers. GFP-based self-associating bipartite split-GFP systems have now been widely used to monitor expression *in vitro*, protein-protein interactions, localization, and trafficking of proteins (47)(48)(49)(50). A more recent class of split-GFP variants, tripartite split-GFP, relies on the self-assembly of three GFP fragments, and is also very well suited for the detection of protein-to-protein interactions (51, 52).

Split GFP (bipartite) works by generating two non-fluorescent parts of the GFP protein and recombing them so that the protein is active and fluoresces. The split GFP molecule has 2 parts: GFP1-10 and GFP11. GFP1-10 pertains to the first 10

beta sheets and GFP11 to the last (Figure 1.8). Interactions between these 2 parts form essential components of the chromophore which causes the green fluorescence. The chromophore involves three specific amino acid residues (Ser 65, Tyr 66, and Gly 67) that go through cyclisation, dehydration, and oxidation during maturation.

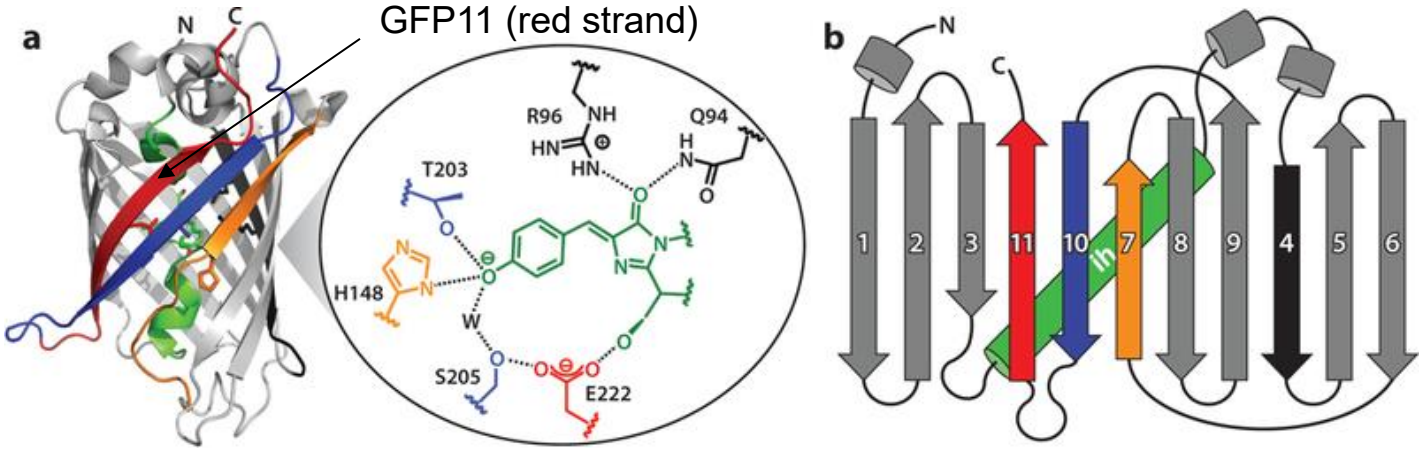


Figure 1.8 Green fluorescent protein (GFP) structure and topology adapted from Lin et al study (53). **(a)** The ribbon structure of GFP showing the chromophore environment. The internal α-helix that contains the chromophore is shown in green, while β-strands 4, 7, 10, and 11 are shown in black, orange, blue, and red, respectively. **(b)** Topology of GFP's 11 β-strands and internal α-helix (ih), GFP11 is shown in red.

Native superfolder GFP known as eGFP was split at the 10th and 11th strand to generate a split GFP molecule. This gave a GFP11 fragment of 80 amino acids, which is still considerably large and not well suited to recombination, therefore an optimised version of split GFP was also used throughout the study. The optimised GFP11 (GFP11opt) (51) is only 16 amino acids in length and specifically developed to be used as split GFP. It was hypothesised that this smaller sequence, when placed in front of the core-E1 HepC ER signal peptide, would be much less likely to affect the core-E1 signal peptide. The GFP1-10 part could be constitutively expressed in the background but requires the presence of GFP11 to generate the green fluorescence signal. By coupling the HepC core-E1 signal peptide with fluorescent proteins, we would be able to determine if the viral signal sequence is

capable of being transposed into a different system for the targeting of recombinant polypeptides into the ER and for release of the eGFP11 to complement and activate GFP1-10 in the cytosol.

Importantly, in a previous study by kakimoto *et al* they fused both an mCherry molecule, GFP1-10 molecule and a GFP11 molecule to membrane bound proteins in different organelles in the cell. The mCherry fluorescence was undisturbed due to the fusion and furthermore split GFP (GFP1-10 and GFP11) was able to recombine as fusion proteins as well. This means that GFP1-10 or GFP11 fusion proteins can still associate to give green fluorescence.

1.8 Project aims

The aim of this project was to investigate whether the HepC signal peptide described above could maintain its normal function of ER localising of a target polypeptide into the ER and release a 'complementation' fragment (eGFP11) to bind with eGFP1-10 and generate active eGFP. Alternately the complementation component could be replaced with a selection marker or a complement fragment of a selection marker such that cells could not grow in the absence of the activated selection marker. This would provide an alternative to the detection of high producing clones by making a direct link between the amount of selection and protein produced. In the proof of concept described here, fluorescent molecules were used to investigated whether the HepC signal peptide can release GFP (GFP11 fragment) and mCherry into the cytosol and ER, respectively. The GFP11 could then recombine with GFP1-10 in the cytosol whilst the mCherry is secreted out of the cell.

2.0 Materials and Methods

Where appropriate, the methods used during this investigation were undertaken aseptically, unless otherwise stated.

2.1 Reagents for molecular biology procedures

Lysogeny broth (LB) media was prepared containing 10 g bactopeptone, 5 g yeast extract, 10 g NaCl and made up to to 1 L with distilled water. LB agar plates were then prepared by adding 15 g agar per litre of LB.

Super optimal broth (SOB) was prepared with 20 g of bactopeptone, 5 g yeast extract, 0.5 g NaCl, 10 ml 0.25 M KCl, 5 ml 2 M $MgCl_2$ and made up to 1 L with distilled water. Super optimal broth with catabolite repression (SOC) was prepared with 10 ml SOB supplemented with 100 µl of 50% (w/v) glucose solution.

2.1.1 Preparation of ampicillin stocks and LB plates with ampicillin

A stock 50 mg/ml ampicillin solution was filter sterilised and stored as 1 ml aliquots at -20°C. 1 ml ampicillin was then added to 1 L of LB agar for preparation of agar ampicillin plates.

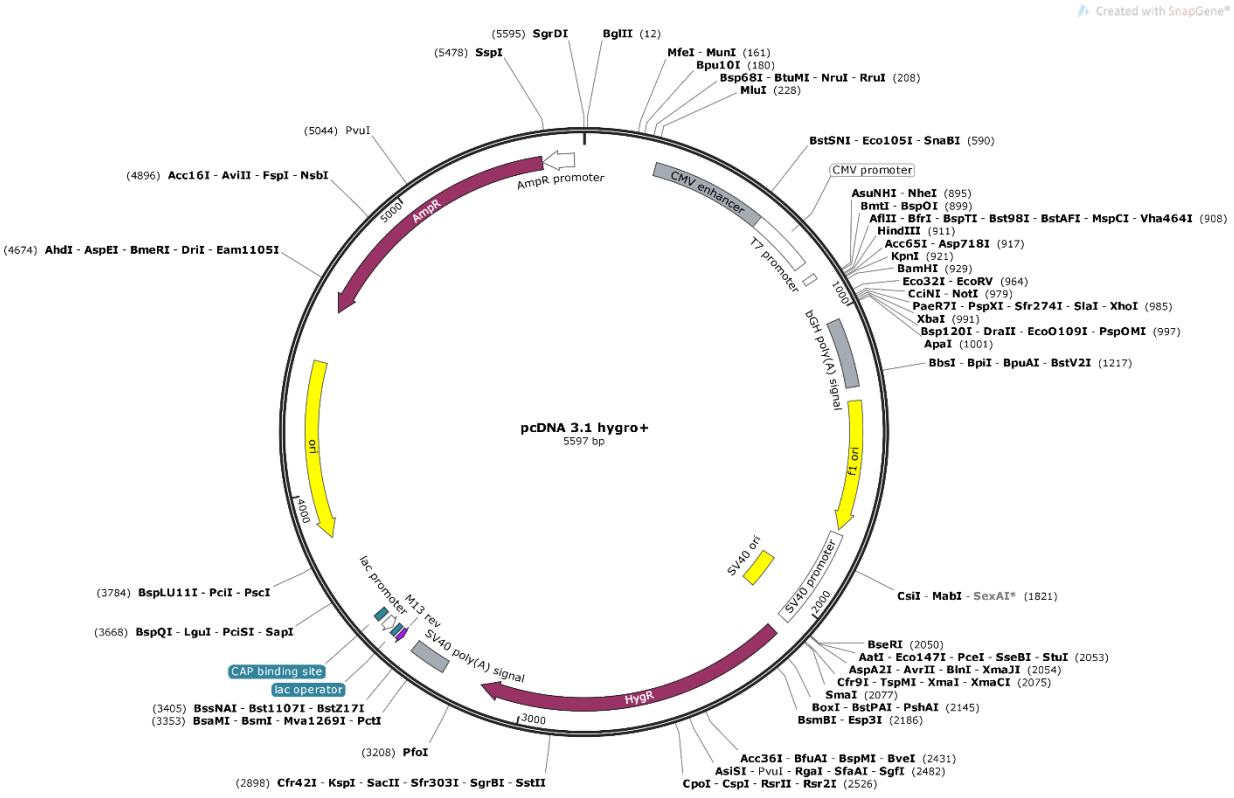
2.1.2 Preparation of calcium competent DH5α *E.coli* cells

DH5α cells were grown overnight by streaking from a glycerol stock on an LB agar plate. One colony was then picked and grown overnight in 5 ml of LB media in a 15 ml falcon tube at 37ºC in a shaking incubator at 150rpm. 50 ml of SOB was then inoculated with 1 ml of the 5 ml starter culture and grown until an absorbance of 0.4-0.6 at A_{600} was reached. Immediately upon reaching the desired absorbance the cells were chilled and centrifuged at 3500 rpm for 5 min at 4°C. The cell pellet was

then resuspended in 2 ml of 0.1 M $CaCl_2$ plus 1 ml of 80% glycerol (v/v). Cells were then dispensed as 100 µl aliquots into chilled cryotubes and flash frozen with liquid nitrogen before being stored at -80°C until required for experiments.

2.2 Molecular cloning of genes of interest into mammalian expression plasmids

A number of plasmid constructs (Table 2.1) were generated and used throughout this study. Constructs 1-4 containing split GFP sequences and a full length GFP (control) were cloned into a standard pcDNA3.1hygro backbone containing ampicillin (bacteria) and hygromycin (mammalian) resistance genes for selection. The remaining constructs were cloned into a proprietary Lonza vector backbone known as Lmm154. Lmm154 contains an ampicillin and glutamine synthetase (GS) resistance gene marker for selection.



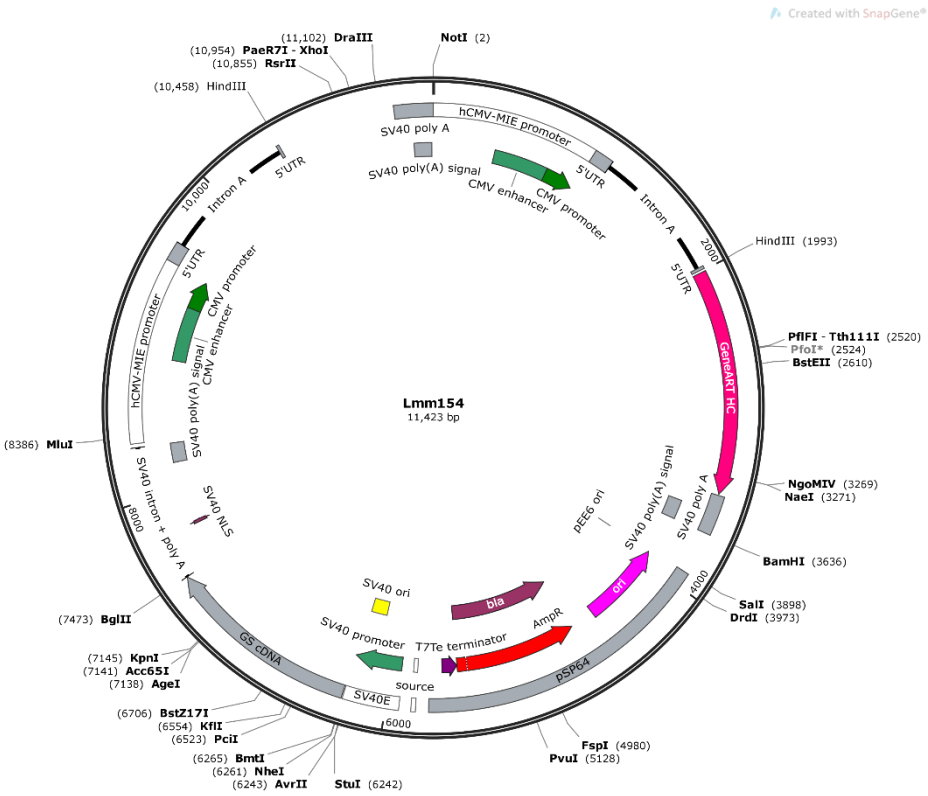


Figure 2.0. Plasmid backbone maps of the 3.1hygro and Lmm154 plasmids used throughout this study. The plasmid maps show the restriction sites that were available for use, the mammalian selection markers, hygromycin (3.1 Hygro) and GS (Lmm154), the ampicillin resistance gene and the promoters used to drive gene expression. Inserts were cloned into the 3.1Hygro vector using NheI and BamHI restriction sites and into the Lmm154 vector using HindIII and BamHI. In the case of the Lmm154 vector, this removed the gene already residing in the vector and was replaced with the new target insert.

Table 2.1 Plasmid DNA *constructs created and used throughout this study, with their designated name, plasmid DNA backbone, restriction enzymes used for cloning and insert size.*

	Name of Construct	Backbone Cloned into	Restriction Enzymes	Insert Size (bp)
1	*GFP1-10*	*3.1hygro*	*NheI/BamHI*	*498*
2	*GFP1-10$_{opt}$*	*3.1hygro*	*NheI/BamHI*	*654*
3	*H7GFP1-10*	*3.1hygro*	*NheI/BamHI*	*564*
4	*H7GFP1-10$_{opt}$*	*3.1hygro*	*NheI/BamHI*	*729*
5	*mCherry*	*Lmm154*	*NheI/HindIII*	*711*
6	*mCherryGFP11*	*Lmm154*	*NheI/HindIII*	*974*
7	*mCherryGFP11$_{opt}$*	*Lmm154*	*NheI/HindIII*	*779*
8	*H8mCherry*	*Lmm154*	*NheI/HindIII*	*785*
9	*H8mCherryGFP11*	*Lmm154*	*NheI/HindIII*	*1071*
10	*H8mCherryGFP11$_{opt}$*	*Lmm154*	*NheI/HindIII*	*876*
11	*GFP*	*3.1hygro*	*NheI/BamHI*	*720*
12	*GFP-HepC-mCherry*	*Lmm154*	*NheI/HindIII*	*1538*
13	*GFP11-HepC-mCherry*	*Lmm154*	*NheI/HindIII*	*1067*
14	*GFP11$_{opt}$-HepC-mCherry*	*Lmm154*	*NheI/HindIII*	*875*

A detailed description of each construct in Table 2.1 is as follows;

1. eGFP1-10; the split eGFP eGFP1-10 component
2. eGFP1-10opt; split eGFP using the optimised sequence component (M3 variant) from Cabantous *et al* (54)
3. eGFP1-10; as for 1 but containing a IgG heavy chain ER signal sequence to direct to the ER and a KDEL ER retention sequence to maintain in the ER
4. eGFP1-10opt; as for 3 but with optimised eGFP1-10 sequence also from Cabantous *et al* (54)
5. mCherry
6. mCherry fused at the 3’ end with the eGFP11 sequence
7. as for 6 with optimised eGFP11 sequence (54)
8. mCherry with an IgG heavy chain ER signal sequence to direct to the ER
9. as for 6 with an IgG heavy chain ER signal sequence to direct to the ER
10. as for 7 an IgG heavy chain ER signal sequence to direct to the ER
11. Intact eGFP

12. eGFP11 fused to a HEPC signal peptide to direct to the ER with a 3' mCherry
13. as for 12 but with eGFP11 optimised sequence from Cabantous *et al* (54)
14. As for 12 but eGFP11 replace with full length GFP

The required insert sequences for constructs 1-4 were synthesised commercially by ThermoFisher Geneart (55) as DNA strings with complementary restriction sites added at both 3' and 5' ends for cloning into the desired vector backbone. Construct 11 was sourced from within the laboratory plasmid library. The remaining required inserts for generation of constructs (6-10 + 12-14) were also synthesised commercially by ThermoFisher Geneart (55) as gene strings and cloned into the Lmm154 vector.

2.2.1 Cloning of Split GFP and mCherry into the 3.1 hygro and Lmm154 vectors

Oligonucleotide primers reported in Tables 2.2 and 2.3 for PCR amplification of the gene of interest described in Table 1 were designed and purchased using IDT oligo synthesis website (56). Oligonucleotides of approximately 20-30 bp complementary to the 5' and 3' ends of the DNA strings purchased were used in PCR reactions to amplify the DNA for cloning. DNA strings were amplified for cloning with 1 µl of Phusion polymerase (57), 20 µl 5x high fidelity buffer, 2 µl 10 nM dNTP's, 4 µl forward and reverse primer made up to 100 µl with water. The PCR was then run for 40 cycles with the following parameters; first denaturing stage at 98ºC for 10 seconds, annealing stage at the melting temperature of the primers -2ºC (58) (calculated using oligioanalyzer tool), finally, the extension stage at 72ºC for 30 seconds x kbp of gene string. After completion of the 40 cycles the cycles, the reaction was kept for 7 min at 72ºC before the amplified DNA was either used for immediately for cloning or stored at -20ºC. Before cloning using the amplified DNA,

the PCR products were analysed by agarose gel electrophoresis to check the band size of the product(s) obtained.

Table 2.2 *Primers designed for the amplification of split GFP genes into the 3.1 hygro and Lmm154 plasmids.*

Primer ID	***Sequence and melting temperature***
1.eGFP10.for	TAT GCT AGC ATG GTG AGC AAG GGC (62ºC)
1.eGFP10.rev	ATA GGA TCC TTA CTT CTG CTT GTC GGC CAT (62ºC)
2.eGFP10opt.for	TAT GCT AGC ATG TCC AAA GGA GAA GAA CTG TTT ACC (62ºC)
2.eGFP10opt.rev	ATA GGA TCC TCA TTT TTC ATT TGG ATC TTT GCT CAG GAC (62ºC)
3.H7.G10K.for	TAT GCT AGC AGT GAG TTT GGG CTG AGC (62.3ºC)
3.H7.G10K.rev	ATA GGA TCC TTA TCA CAG CTC GTC TTT CTT CTG C (61.8ºC)
4.H7.G10optK.rev	ATA AAG CTT TTA CAG CTC GTC TTT TTT TTC ATT TGG ATC TTT GC (62ºC)
5.mC.for	TAT GGA TCC ATG GTG AGC AAG GGC G (62.8ºC)
5.mC.rev	ATA AAG CTT TTA TTG TAC AGC TCG TCC ATG CCG (61.7ºC)
6.mC.G10.rev	ATA AAG CTT TTA CTT ATA TAG TTC GTC CAT GCC GAG AGT GAT C (62ºC)
7.mC11opt.rev	ATA AAG CTT TCA TGT AAT CCC AGC AGC ATT TAC GTA CTC (62.1ºC)
8.H8.mC.for	TAT GGA TCC ATG GAC CTC CTG CAC AAG (61.8ºC)

Table 2.3 *Primers designed for the insertion of HepC SP genes into Lmm154 plasmids.*

Primer ID	***Sequence and melting temperature***
for_GFP11Hep	TAT GGA TCC ATG AAC GGC ATC AAG GTG AAC TTC AAG ATC C (65ºC)
for_11optHepPrim	TAT GGA TCC ATG CGT GAC CAC ATG GTC CTT CAT G (65.1ºC)
for_GFPHep	TAT GGA TCC ATG AGT AAA GGA GAA GAA CTT TTC ACT GGA GTT GTC C (64.6ºC)
rev_HepC	ATA CGT CTC TAG CGG ATG CTG GCA CGG TAA (65.2ºC)

2.2.2 DNA agarose gel electrophoresis analysis and purification of DNA from agarose gels

In order to assess PCR products and DNA digests, agarose electrophoresis was used. Gels were prepared with 0.5 g agarose powder and 50 ml of 1xTAE solution

that was microwaved until clear. 7.5 µl of SYBR™ safe gel stain was then added, and the solution then poured into a DNA gel cast. When cool and set, the gel was covered in 1xTAE buffer until submerged. 10 µl of a 1 kB DNA ladder (Promega) (59) was loaded into one well of every gel as a size marker. Other lanes were loaded with 20 µl of sample. Gels were then run for ~ 1 h at 90 mV before being imaged using a Syngene U:Genius3 imaging system.

PCR products or digested DNA was excised from gels with a razor blade using a UV gel imaging lamp. The DNA was then purified from the gel using a Promega Wizard® SV Gel and PCR Clean-Up System following the manufacturer's instructions (60).

2.2.4 DNA ligation and transformation of competent bacteria with plasmid DNA (constructs 1-10)

To generate constructs 1-10 (Table 2.1) 1 ug of amplified target insert string DNA and plasmid were separately digested with 0.5 µl FastDigets® enzymes (Table 2.1) as per the restriction sites required to produce "sticky ends". Digests were prepared with 2 µl buffer and then nucleotide free water to give a total volume of 20 µl. The mixture was placed at 37°C for 30 mins for digestion to occur.

The plasmid and gene strings were then ligated together using 1 µl of Promega T4 DNA ligase (61), 1 µl DNA ligase buffer, 3 µl digested plasmid DNA and 5 µl digested gene string to give a total volume of 10 µl. Ligation reactions were then left to incubate overnight at room temperature. Following this, the ligation reaction (10 µl) was added to 50 µl of competent *E.coli* cells in a cryotube and incubated on ice for 30 min. The competent cells were heat shocked in a 42°C water bath for 90 sec and

then placed immediately back on ice for 2 min. 950 µl of SOC media was then added and the mixture incubated for 1 h at 37°C with shaking at 150 rpm. 150 µl of sample was then spread on an LB agar Amp+ plate and left overnight at 37°C.

2.2.5 Golden gate assembly and transformation of competent bacteria with plasmid DNA (constructs 11-14)

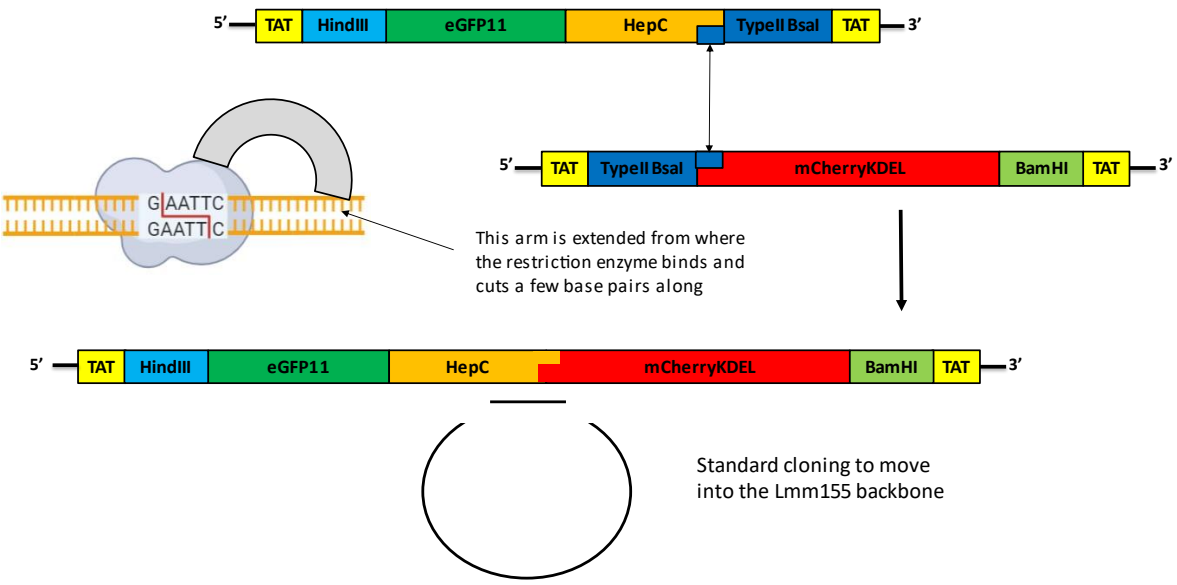


Figure 2.1. Schematic depicting the Golden Gate Assembly of HepC and mCherry strings. Type IIS restriction enzyme sites are shown in dark blue and type I in green and light blue. The diagram shows how the type II restriction enzymes cleave outside the recognition sequence indicated by the grey semi-circle shape. This type of cloning allows multiple DNA strings to be added directionally one after another in a seamless fashion. This type of cloning meant that the second string shown in red could be used multiple times saving money. Once the two strings have been ligated together, the new single DNA piece can be inserted into one of the plasmid vectors using the standard cloning previously mentioned in the methods.

A different cloning approach, known as Golden Gate Assembly, that utilises so called Type IIS restriction enzymes, was used to generate constructs 11-14 (Table 2.1). Figure 2.1 shows where type IIS restriction sites were added at the 3' end of the HepC DNA and 5' end of the mCherry DNA. Golden Gate uses the ability of Type IIS restriction endonucleases (Reases) to cleave DNA outside of the recognition sequence (62)(63).

As described in the previous traditional restriction site cloning (see section 2.2.1), commercially synthesised DNA strings were amplified using PCR. 100 ng of the front sequence (HepC) and 100 ng of the back sequence (mCherry) were added to a PCR tube containing 1 µl ligase buffer, 0.5 µl T4 DNA ligase, 0.5 µl BsmBI restriction enzyme and H_2O to give a total reaction volume of 10 µl. This mixture was placed in a PCR thermocycler for 10 min at 42°C, then programmed to cycle between 42°C and 16°C with 5 min each for 30 cycles. Finally, the sample was incubated at 55°C for 10 mins. The Golden Gate assembly products were then cleaned up using a PCR clean-up wizard kit (Promega) (60), and then subjected to standard PCR amplification to amplify the assembled sequence. The cloning methods previously described above were then used to ligate the newly generated and assembled insert into the Lmm154 plasmid backbone. All constructs were confirmed by both DNA agarose gel electrophoresis of restriction enzyme digests and by commercial sequencing to confirm the presence of the required insert without the introduction of any mutations.

2.2.6 Plasmid DNA purification from *E.coli* cultures

Plasmid DNA (10 ng in 1 µl) was added to 100 µl of competent *E.coli* cells in a cryotube and incubated on ice for 30min. The competent cells were heat shocked in a 42°C water bath for 90 sec and placed immediately back on ice for 2 min. 900 µl of SOC media was then added and incubated for 1 h at 37°C shaking at 150 rpm. 100 µl of sample was then spread on LB agar Amp+ plate to grow overnight at 37°C.

Single colonies picked from LB agar amp+ plates were inoculated in 100 ml LB and incubated overnight at 37°C in a shaking incubator at 150 rpm. The plasmid DNA was then purified using a QIAGEN plasmid maxikit as per the manufacturer's

instructions (64). The DNA concentrations of the isolated DNA for each construct were then determined using a Nanodrop spectrophotometer (Table 2.4) such that the volume of DNA required for transfections could be calculated.

Table 2.4 *DNA concentrations for the generated DNA constructs indicated calculated using the nanodrop spectrophotometer.*

	DNA Construct	DNA Concentration (ng/μl)
1	*GFP1-10*	*1332*
2	*GFP1-10$_{opt}$*	*510*
3	*H7GFP1-10*	*N/A*
4	*H7GFP1-10$_{opt}$*	*N/A*
5	*mCherry*	*1039*
6	*mCherryGFP11*	*771*
7	*mCherryGFP11$_{opt}$*	*941*
8	*H8mCherry*	*1198*
9	*H8mCherryGFP11*	*1071*
10	*H8mCherryGFP11$_{opt}$*	*876*
11	*GFP*	940
12	*GFP-HepC-mCherry*	911
13	*GFP11-HepC-mCherry*	860
14	*GFP11$_{opt}$-HepC-mCherry*	775

2.3 Tissue culture

Tissue culture refers to those experiments performed using the mammalian CHO cell lines, and therefore requires absolute aseptic techniques and working within biosafety cabinets to preserve cells and not introduce bacterial or fungal contamination. Throughout the work describe in this thesis, the Lonza Biologics proprietary CHOK1SV GS-KO® host cell line or the CHO-S ThermoFisher Scientific host cell line was used and cultured in chemically defined, protein free CD-CHO media (ThermoFisher Scientific) unless otherwise stated.

2.4 Stable cell pool generation

To generate stable cell pools with plasmid DNA integrated into the genome, 50 µg of plasmid DNA was first linearised using 2.5 µl of FastDigest® enzyme FSP1 for 6 hours then adding 10 µl sodium acetate and 250 µl ice cold ethanol. The DNA was then pelleted via centrifugation for 30 min at 13000 rpm and the pellet washed with 100 µl of 0.2µm filtered 70% EtOH. The DNA was then pelleted again via centrifuging at 13000 rpm for 5 min, the supernatant discarded, and the pellet allowed to airdry for 15 min. The pellet was then resuspended in 100 µl of 0.2 µm filtered 1xTE buffer and left for 1 h at room temperature before being mixed thoroughly after. The DNA concentration was then measured and adjusted to a xc concentration of 20 µg in 100 µl.

For transfection, $1x10^8$ viable CHO-S cells were suspended in 7 ml of CD-CHO media containing 6 mM L-glutamine and 0.7ml added to 100 µl of the 20 µg DNA solution in an electroporation cuvette. The mixture was then electroporated in a BioRad electroporator with a single pulse of 300 V and capacitance of 900 µF after which it was added to 8.2 ml of fresh CD-CHO containing 6 mM L-glutamine in a T25 flask. A further 1 ml of CD-CHO media was then used to wash the electroporation cuvette and added to the T25 flask which was placed at 37°C in 5% CO_2. 24 h after, 750 µg/ml hygromycin as added. Cell count and culture viability was then determined at day 13, 17 and 21 post-transfection. At 21 days post transfection a flask at $0.2x10^6$ viable cells/mL in CD-CHO 6 mM L-glutamine and 750 µg/ml hygromycin was prepared from the T25 flask in a total of 20 ml in a 125ml corning Erlenmeyer non-vented flask. The flask with gassed with 5% CO_2 and subsequently at each passage. The resulting culture was grown at 37°C, 140 rpm and passaged every 3-4 days, where at each passage the cells were seeded at $0.2x10^6$ cells/mL in 20 ml.

After 3-4 passages, when culture viability was >90%, the cell pools were cryopreserved. To achieve this, cells were frozen in growth media + 10% (v/v) DSMO at $1x10^7$ viable cells/mL in 1 ml aliquots in cryostat tubes. These were then placed in a Mr Frosty freezing container in a -80°C freezer for gradual cooling. After 24 hours the cells were moved to liquid nitrogen storage in a cryostat.

2.5 Transient transfections

Although stable cell pools for constructs 1-4 were generated, transient transfection was largely used throughout the project. A number of different transfections and co-transfections were undertaken as described in Table 2.5. All transient transfections were undertaken in CHOK1SV GS-KO® cells. 2 x 10^6 CHO cells were prepared in 9.8 ml of CD-CHO media. Plasmid DNA was also prepared at 20 µg in 100 µl of TE buffer for each construct used. 100 µl of DNA and 700 µl of prepared cells were then added to a 4 ml electroporation cuvette and the mixture electroporated in a BioRad electroporator with a single pulse of 300 V and 900 µF. The electroporated sample was then added directly to a 50 ml falcon tube containing 8.5ml of CD-CHO + 6 mM glutamine. The electroporation cuvette was washed with a further 700 µl of media that was also added into the 50 ml falcon tube. All samples were then gassed with 5% CO_2 for 8-10 seconds and left to grow for 48 hours at 37°C with shaking at 125 rpm before harvesting the supernatant and cells for analysis as described in section 2.5.1.

Table2.5 *Transient transfections undertaken in* CHOK1SV GS-KO® *cells.*

	Construct used for Transfection	Co-transfected with
1	*3.1 hygro (control)*	*N/A*
2	*Lmm154 (control)*	*N/A*
3	*H8-mCherry*	*N/A*
4	*mCherry*	*N/A*
5	*GFP*	*N/A*
6	*GFP1-10*	*N/A*
7	*GFP1-10$_{opt}$*	*N/A*
8	*GFP11-HepC-mCherry*	*N/A*
9	*GFP11$_{opt}$-HepC-mCherry*	*N/A*
10	*GFP-HepC-mCherry*	*N/A*
11	*GFP11-HepC-mCherry*	*GFP1-10*
12	*GFP11$_{opt}$-HepC-mCherry*	*GFP1-10$_{opt}$*
13	*GFP11-mCherry*	*GFP1-10*
14	*GFP11$_{opt}$-mCherry*	*GFP1-10$_{opt}$*

2.5.1 Harvesting transient transfections

After 48 hours post-transfection, cultures were analysed for viable cell numbers and culture viability and then the supernatant collected by centrifugation to pellet cells. Aliquots of cells to give pellets of 1 x 10^6 viable cells were also collected and the pellet used for intracellular analysis of protein expression. The supernatant samples were stored in a -20°C freezer. To the 1 x 10^6 cell pellets was added 300 µl of lysis buffer (consisting of- hepes 0.48g - NaCl 0.58 g - b-Glycerophosphate 0.216 g - triton 1 ml - pH to 7.2 with NaOH and for every 10 ml of this add 1 Roche inhibitor tablet plus 500 ul Naf and 50 ul Nav) in an Eppendorf and the sample then left to lyse on ice for 10 min. The cell lysate was then centrifuged at 10000 rpm for 5 min and the supernatant transferred into -20°C freezer for storage until analysed.

2.6 Analytical Methods for Analysis of Protein Expression

The following methods describe how protein expression after transient transfection was analysed.

2.6.1 SDS PAGE and Western Blotting

A Bradford assay was initially used to determine the total protein content of each cell lysate following transient transfection (intracellular material). Solutions were prepared consisting of 47 µl water and 3 µl of cell lysate sample and added to 1 ml of Bradford reagent. A standard curve with known concentrations of bovine serum albumin (BSA) was used to calculate protein concentration.

For SDS-PAGE analysis (5% stacking gel, 12% resolving gel), a volume of lysate to give a protein concentration of 20 µg was taken and mixed with 8 µl of 5x non-reducing sample buffer and then made up to a total volume of 40 µl with water. For cell culture supernatant samples (extracellular samples) 20 µl of supernatant was mixed with 8 µl of 5 x non-reducing sample buffer and then the sample made up to 40 µl with water. Samples were not boiled. 10 µl of each sample was then added to each lane of the gel. A protein marker was also run in one lane to allow estimation of protein size. SDS-PAGE gels and run for ~30 min at 100 mV and 1 h at 150 mV. For staining of proteins Coomassie blue was used.

For western analysis, following SDS-PAGE, proteins were transferred to nitrocellulose membranes, blocked with 5% milk powder (w/v) in TBS-Tween buffer and then incubated in the appropriate primary antibody overnight. Membranes were probed with anti-mCherry (ab167453), anti-GFP and anti-β-actin. Antibodies were diluted 1:5000 in 5% blocking buffer (10 ml T-TBS 1x containing 0.5 g dry milk powder). Anti-mouse secondary antibodies diluted 1:5000 in 5% blocking buffer were used with anti-GFP and anti-β-actin. Anti-rabbit secondary antibodies diluted 1:5000 in 5% blocking buffer were used with anti-mCherry antibody. To develop the blots, 2 ml of ECL reagent were added to the membrane, the membranes

transferred to a case and exposed to photographic film for a set amount of time. After exposure, the blots were developed using a x-ray film developer.

2.6.2 Confocal Microscope Slide Preparation

In order to image cells for fluorescence expression using confocal microscopy it was necessary to prepare cells on appropriate microscope slides. To achieve this, 13 mm round coverslips were placed in wells of a 24 well tissue culture plate and the coverslips then coated with 100 µl poly-L-Lysine 0.01% for 15 min. This allows the suspension cells to attach to the surface of the coverslip for analysis. The coverslip was then washed with 1 ml of sterile H_20 and left to air dry. Transfected cells for analysis (2 x 10^6) were then placed in an Eppendorf, pelleted at 1000 rpm for 5 min and then washed with 1 ml of PBS. The PBS was then removed after pelleting the cells again, after which the cells were fixed by addition of 300 µl of 4% (w/v) paraformaldehyde by gently mixing and then incubating for 20 min at room temperature. The cells were then pelleted again by centrifuged at 1000 rpm for 5 min to allow removal of the paraformaldehyde, after which the cells were washed twice with 1 ml of PBS. The cells in 1 ml of PBS were then added to the coverslips and the cells left for 1 h to attach before the PBS was removed. Finally, the coverslips were mounted onto glass slides with ProLong Gold Antifade Mountant ready for confocal microscopy analysis.

2.6.3 ZEISS LSM 880 Elyra Confocal Microscope Analysis

Lattice-SIM images were acquired on a Zeiss LSM 880 Elyra AxioObserver confocal microscope (65). The system used 2 laser lines on 2 channels, an argon laser to detect GFP green fluorescence at 488 nm and a Helium 543 laser to detect mCherry red at 561 nm. The argon laser used a standard PMT detector whereas the Helium 543 laser used the Zeiss airyscan detector. The third channel; T PMT-T2, was a

transmitted light detector, essentially a brightfield channel. Images were collected in tiles of 9 at a size of 209.85 x 209.85 µm. The Z-stack scans were performed in 0.22 µm intervals for a total of 49 images. All settings used are reported in an image in the appendix. Image processing was performed with the ZEN 3.0 SR (blue) and (lite) software.

3.0 Results

3.1 Generation of split GFP and HepC signal peptide plasmid DNA constructs

This project set out to establish if ER signal sequences, that direct polypeptides to the ER, but are subsequently cleaved from the mature polypeptide/protein, could be used to undertake alternative functions. Split GFP systems have previously been described whereby a helix forming peptide from the C-terminal of eGFP (labelled eGFP11 in this work) can associate with the remainder of the eGFP protein (labelled eGFP1-10) and result in the formation of active, fluorescing eGFP (54, 48). This split GFP system was used as a model in proof-of-concept studies to show that our novel approach to utilising ER signal peptides to complement other functions in the cell was possible. In particular, the novelty in the designed system is that the ER signal sequences can be used post-cleavage and after directing a target polypeptide to the ER to undertake other functions that directly relate ER signal peptide amounts to target secretory protein amounts.

The eGFP fragments used in this study were made as either the native sequence or a so termed ‘optimised’ sequence that has previously been reported to enhance the ability of the two fragments to associate and fluoresce (54, 48). A viral ER signal sequence, termed HepC here, that has previously been described (43) was also utilised to direct target polypeptides to the ER and to incorporate the GFP11 or full length GFP molecule into. In this way we could be established whether the GFP11 could be released from the HepC signal sequence to complement GFP1-10, or if the full length GFP molecule itself could be released, thus demonstrating the novel finding that ER signal peptides can be used to direct other functionalities in the cell.

To investigate this, initially split GFP (GFP1-10 and GFP11) and HepC DNA strings were commercially purchased (Table 2.1) and digested with restriction enzymes and then ligated into the relevant DNA plasmid; 3.1 hygro or Lmm154 (Figure 3.1), selected as they contain CMV promoters to drive recombinant gene expression and for stable expression in mammalian cells contain either a hygromycin resistance (3.1hygro) or glutamine synthetase (GS) selection marker. After ligation, the plasmids were transformed into competent bacteria cells (DH5α) and plated onto ampicillin agar plates which was used as a selection pressure. Both plasmids contain an ampicillin resistance gene so in theory only cells containing fully ligated plasmid survived. A restriction enzyme digest was then undertaken on colonies picked from these ligations. (Table 2.1) these confirmed the released a band of the expected size for the inserted DNA and that the remaining plasmid backbone was of the expected size. Thus, the same enzymes used to clone the DNA inserts into the plasmid DNA were used to ‘dropout’ a band of DNA corresponding to the target insert. the resulting agarose gel electrophoresis analysis are shown in Figure 3.2. It is noted that the inserts ligated into the 3.1 hygro plasmid used NheI and BamHI restrictions sites while for the inserts ligated into the Lmm154 backbone the HindIII and BamHI sites were used which results in release of the IgG heavy chain (HC) in this vector that was subsequently replaced with the new target insert.

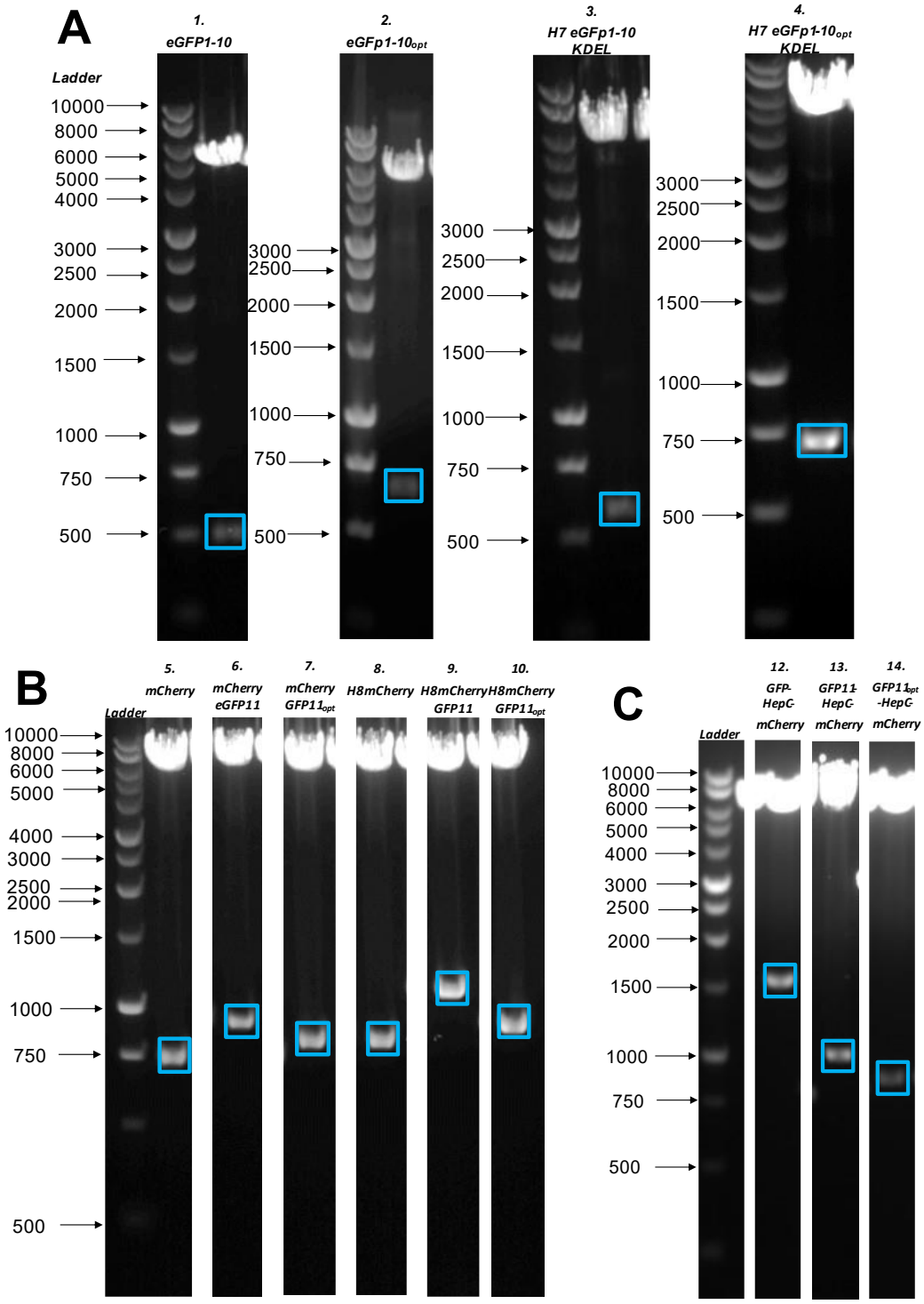


Figure 3.2. DNA agarose gel electrophoresis analysis of restriction enzyme digested and released split eGFP and HepC components from colonies after ligation confirming successful insertion of the required insert. Banding patterns highlighted in blue boxes on gels **(A), (B)** and **(C)** indicate bands of the expected size to be released by digest of successful ligations. Each well was loaded with 0.5 µg of purified plasmid DNA and run per the methods section. **(A)** Constructs 1-4 (table 2.1) show bands that should drop out at 498, 654, 564 and 729 bp, respectively. **(B)** Gel B of constructs 5-10 (table 2.1), part of the split eGFP set and a band dropout can be seen at the appropriate sizes; 711, 974, 779, 785, 1071 and 876 bp from left to right. Finally, Gel **(C)** shows constructs containing the HepC signal peptide and show band dropouts at sizes 1538, 1067 and 875 bp.

Analytical restriction enzyme digests of colonies following ligation were analysed by DNA agarose gel electrophoresis (Figure 3.2) for all constructs 1-14 (Table 2.1). The results for constructs 1-4 are shown on Figure 3.2, gel A, digestion of the eGFP1-10 construct variants showed bands at 498, 654, 654 and 729 base pairs, respectively, as expected (see expected sizes in Table 2.1). A difference of note between constructs is that the sequence for eGFP 1-10$_{opt}$ is larger than wild type eGFP1-10 as seen in figure 3.2. Conversely, the sequence for eGFP11$_{opt}$ is smaller than native eGFP11 and thus the band released is of a smaller size. This distinction is important as the size of eGFP11/$_{opt}$ can be used to identify the different constructs by DNA agarose gel electrophoresis.

Figure 3.2, gel B shows restriction digest analysis of constructs 5-10 of the split eGFP series. Bands that were dropped out by the restriction digests of a size corresponding to the insert have been highlighted with a coloured box in the figure. From left to right bands at approximately 750, just below 1000, above 750, above 750, above 1000 and between 750-1000 bp are observed. Table 2.1 show that these bands match that of the expected DNA dropout sizes for the respective inserts.

Figure 3.2, gel C shows restriction digest analysis of the final set of plasmids, containing the HepC signal peptide DNA. Construct 12 contains a full eGFP sequence and therefore is much larger as seen on gel C. The band for the insert is at about 1500 bp and matches the expected size as reported in table 2.1. Construct 13 and 14 both show the expected band sizes for the inserts in these vectors on the gel compared to the expected sizes reported in table 2.1.

Finally, construct 11 contains the eGFP sequence in the 3.1 hygro backbone but is not shown in the DNA agarose test digest gels. This plasmid containing eGFP was

gifted from another member of the laboratory and had already been analysed by restriction digest analysis and DNA sequencing.

3.1.1 Sanger sequencing of DNA constructs

The test restriction enzyme digests reported in Figure 3.2 provided evidence that the target DNA sequences had been inserted into the plasmid but only gives an indication that the insert matches based on size and not the actual sequence. All constructs (apart from 11 (eGFP) that had previously been sequenced) were therefore sent for commercial Sanger sequencing (66) to confirm that the DNA inserted was of the correct sequence and that no mutations were present in the sequence. This analysis showed that the sequence was correct for all constructs except for constructs 3 and 4.

For both construct 3; H7eGFP1-10KDEL, and construct 4; Construct 4 H7eGFP1-10optKDEL, the sequence was missing an initial ATG for a methionine residue. The bold text in Table 3.1 shows where the open reading frame (ORF) begins for the actual sequence for construct 3 and the correct sequence. As the ATG start codon is missing the translation start site would not be until the methionine approximately halfway through the sequence and a truncated eGFP1-10 protein would be made. Construct 4 had the same defect. Therefore, these constructs were not used as they were not considered necessary to remake and thus not investigated further.

Table 3.1 *Comparison of the amino acid sequence that would arise from the sequence in Construct 3 (sequenced) and the amino acid sequence that should have been coded by the DNA (correct sequence).*

Construct 3 H7eGFP1-10KDEL (sequenced)	REFGLSWVFLVALFRGVQCVSKGEELFTGVV PILVELDGDVNGHKFSVSGEGEGDATYGKLTL KFICTTGKLPVPWPTLVTTLTYGVQCFSRYPD H**MKQHDFFKSAMPEGYVQERTIFFKDDGNY** **KTRAEVKFEGDTLVNRIELKGIDFKEDGNILG** **HKLEYNYNSHNVYIMADKQKKDEL-**
Construct 3 H7eGFP1-10KDEL (correct sequence)	**MEFGLSWVFLVALFRGVQCVSKGEELFTGVV** **PILVELDGDVNGHKFSVSGEGEGDATYGKLT** **LKFICTTGKLPVPWPTLVTTLTYGVQCFSRYP** **DHMKQHDFFKSAMPEGYVQERTIFFKDDGN** **YKTRAEVKFEGDTLVNRIELKGIDFKEDGNIL** **GHKLEYNYNSHNVYIMADKQKKDEL-**

The remainder of the sequences from the constructs were reviewed with a multiple sequence alignment tool (*Clustal Omega < Multiple Sequence Alignment < EMBL-EBI*, 2020) and an open reading frame finder (68) and showed the DNA sequences to be correct. This sequencing data is available at request. A schematic of the different constructs generated and the makeup of these is shown in Figure 3.3. Details on these are provided in the methods section.

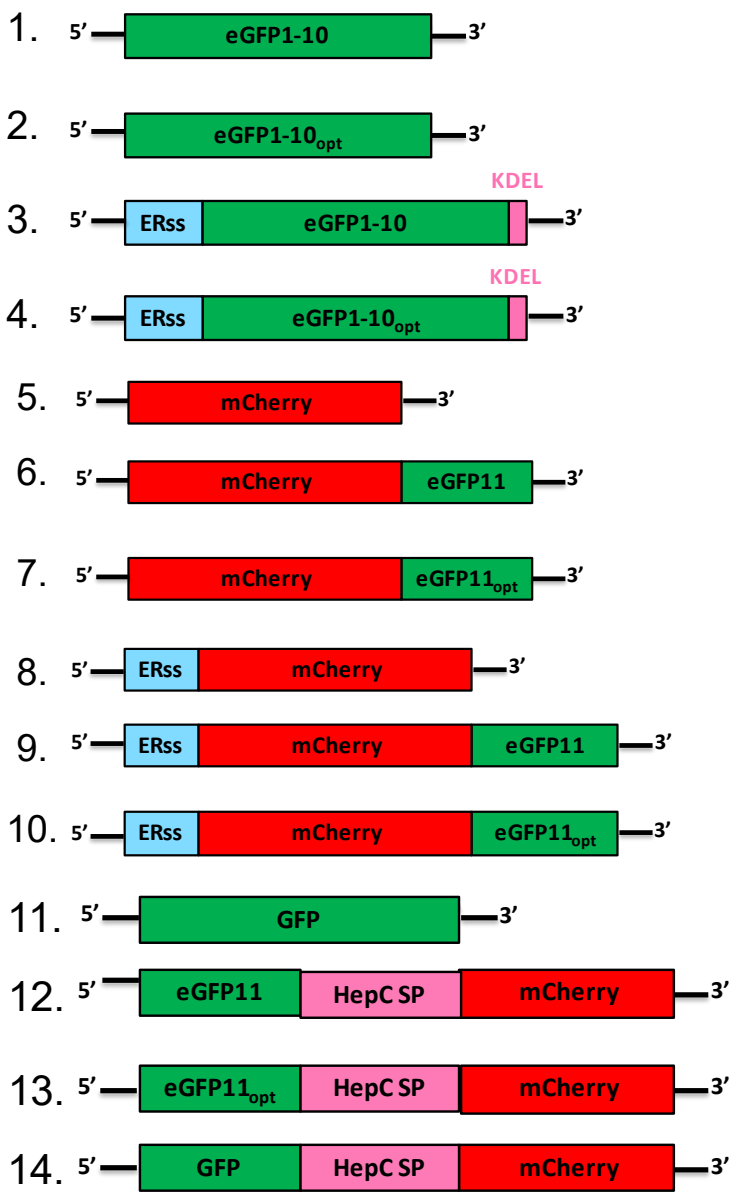


Figure 3.3. Schematic depicting the configuration of constructs 1-14 (see Table 2.1). The 5' and 3' prime ends are noted. Construct 6,7,9 and 10 do not show a linker that was placed between the mCherry and GFP11/$_{opt}$. The linker sequence is GGAGGTGGGTCAGGAGGGGGATCA. ERss = IgG heavy chain ER signal sequence (69); KDEL = ER retention sequence; HepC SP = HepC ER signal peptide. The specific details of each construct is as follows; 1. eGFP1-10; the split eGFP1-10 component; 2. eGFP1-10opt; split eGFP using the optimised sequence component from (54); 3. eGFP1-10; as for 1 but containing a IgG heavy chain ER signal sequence to direct to the ER and a KDEL ER retention sequence to maintain in the ER; 4. eGFP1-10opt; as for 3 but with optimised eGFP1-10 sequence from (54); 5. mCherry; 6. mCherry fused at the 3' end with the eGFP11 sequence; 7. as for 6 with optimised eGFP11 sequence from (54); 8. mCherry with an IgG heavy chain ER signal sequence to direct to the ER; 9. as for 6 with an IgG heavy chain ER signal sequence to direct to the ER; 10. as for 7 an IgG heavy chain ER signal sequence to direct to the ER; 11. Intact eGFP; 12. eGFP11 fused to a HEPC signal peptide to direct to the ER with a 3' mCherry; 13. as for 12 but with eGFP11 optimised sequence from (54); 14. as for 12 but eGFP11 replace with full length GFP.

3.2 Batch culture growth analysis of stably expressing Split GFP1-10 CHOK1SV GS-KO® cells

Constructs 1-2 were linearised using a restriction digest enzyme (FspI) and transfected by electroporation as per the methods section into CHOK1SV GS-KO® cells. Hygromycin was added to the CD-CHO media for selection along with 6 mM L-glutamine. The recovered cell pools were then grown in shake flasks in batch culture alongside the untransfected control CHOK1SV GS-KO® (which was grown in the absence of hygromycin). The culture viability (%) and number of viable cells (/ml x 10^6), were then measured using a Vi-Cell instrument that uses the Trypan Blue Dye exclusion method for 7 days over batch culture. The resulting growth profiles are shown in Figure 3.4.

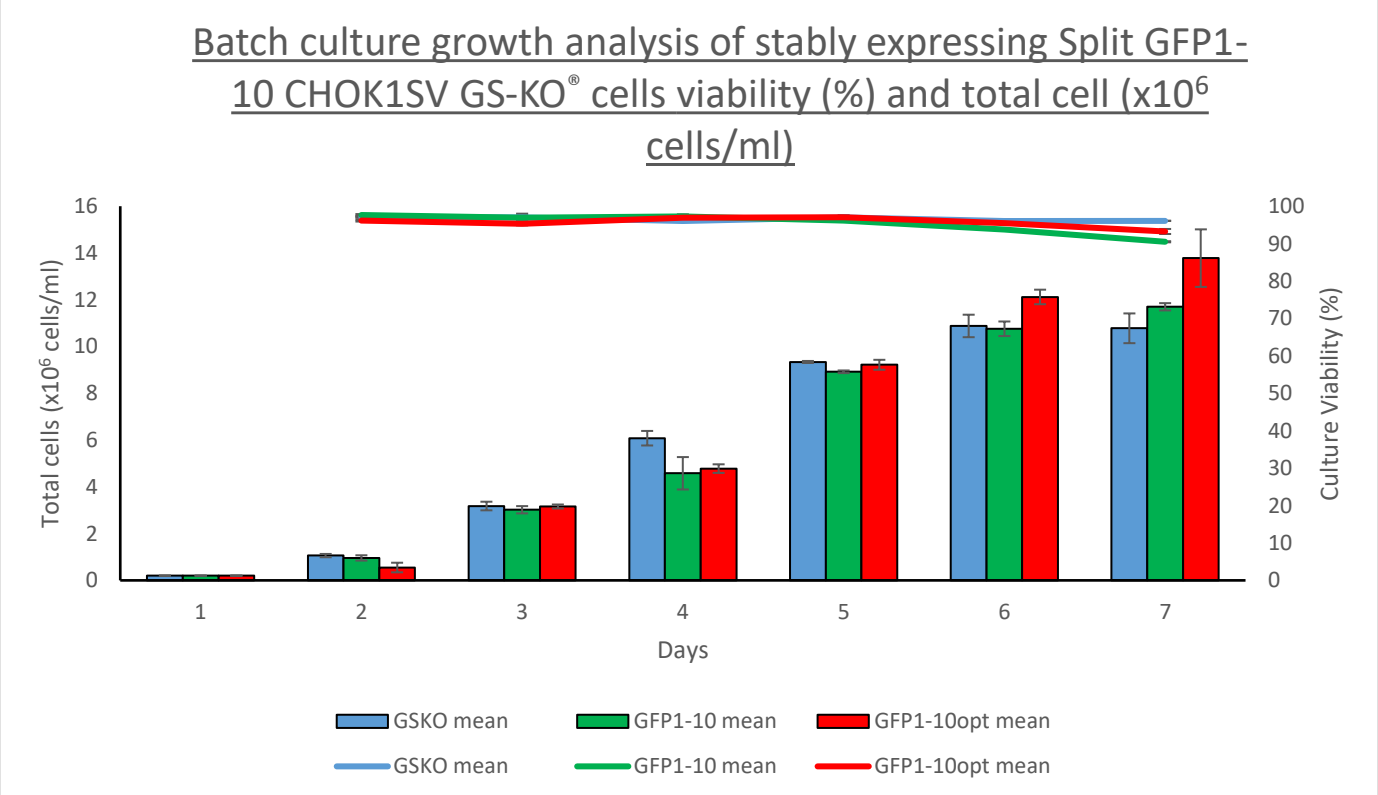


Figure 3.4 G*raph showing the culture viability (%) and viable cells (x10^6 cells/ml) over a 7 day batch culture of eGFP1-10 and control cell pools. The line graph shows culture viability (%) and the bars show viable cells. Error bars showing the standard deviation (n=2 biological duplicate experiments). Each cell pool was grown in duplicate and the average viability (%) and viable cells (x10^6 cells/ml) plotted.*

The growth of stably eGFP1-10 transfected cells across a 7-day batch culture showed that cell pools 1 and 2 had similar growth to each other and the original CHOK1SV GS-KO® host up until day 6 where the total number of cells started to plateau (Figure 3.4). At day 7 the pools with construct eGFP1-10opt reach a higher viable cell number than the control or eGFP1-10 pools although as the cultures were only undertaken in duplicate it was not possible to determine if this increase was significant from this data.

3.3 Transient transfection of the split GFP constructs and eGFP11-HepC- signal peptide- mCherry constructs with eGFP1-10 constructs

From the available 14 constructs and empty control plasmids (3.1 hygro and Lmm154 that do not contain any GFP gene sequences), 14 transfections/co-transfections were undertaken as outlined in Table 3.2. This was undertaken to confirm that the GFP1-10 alone or eGFP11 alone did not give green fluorescence, that full length GFP as a positive control did, that eGFP1-10 with eGFP11-mCherry resulted in green fluorescence and that the combination of the eGFP11-HepC-mCherry with eGFP1-10 resulted in green and red fluorescence.

Table 3.2 *Transient transfections of the different constructs generated in this study into CHOK1SV GS-KO® cells. Co-transfections are also shown on the table.*

DNA transfected	Co-transfected with
3.1 hygro (control)	*N/A*
Lmm154 (control)	*N/A*
H8-mCherry	*N/A*
mCherry	*N/A*
GFP	*N/A*
GFP1-10	*N/A*
GFP1-10$_{opt}$	*N/A*
GFP11-HepC-mCherry	*N/A*
GFP11$_{opt}$-HepC-mCherry	*N/A*
GFP-HepC-mCherry	*N/A*
GFP11-HepC-mCherry	*GFP1-10*
GFP11$_{opt}$-HepC-mCherry	*GFP1-10$_{opt}$*
GFP11-mCherry	*GFP1-10*
GFP11$_{opt}$-mCherry	*GFP1-10$_{opt}$*

Constructs without the HepC signal peptide were designed to show that the split GFP molecules; GFP1-10/$_{opt}$ and GFP11/$_{opt}$, can recombine *in vivo* and generate eGFP resulting in the detection of fluorescence. Constructs 11-14 use the HepC signal peptide to direct the mCherry to the ER but also contain the eGFP11 that can recombine with any eGFP1-10 present.

3.3.1 Confocal microscopy image analysis of transient transfections

The images collected after transfection from the confocal microscope detect both GFP and mCherry fluorescence. Green and red are used to represent the GFP and mCherry respectively. The images were collected and split into 4 lanes. Lane 1 – brightfield (light) to show the cells, lane 2 – GFP, lane 3 – mCherry and lane 4 – all 3 merged. A size marker is shown on all images. The resulting microscopy analysis results are shown in figures 3.5 to 3.9 below.

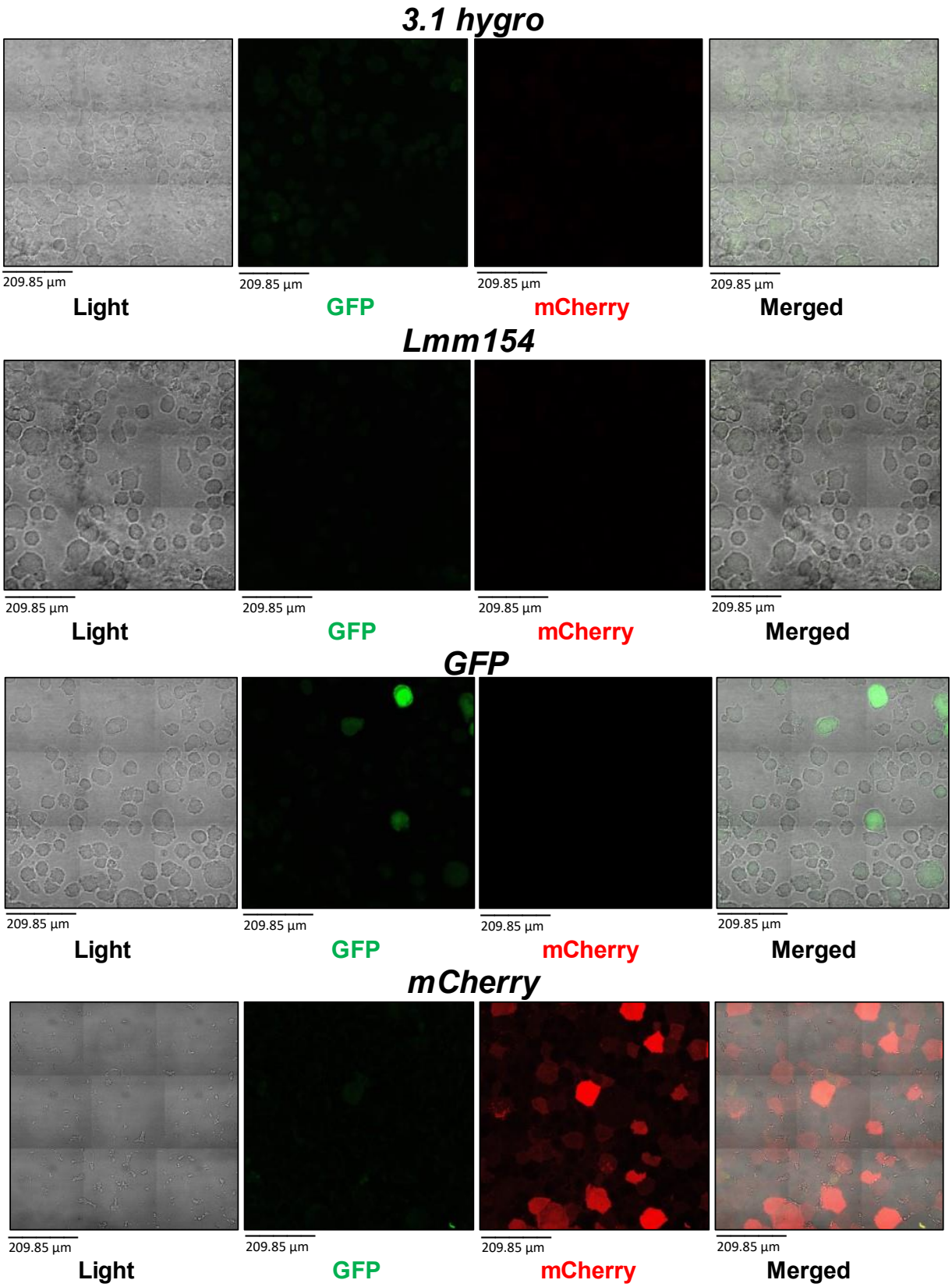


Figure 3.5. Confocal microscopy analysis after transfection of CHOK1SV GS-KO cells with control plasmids 3.1 hygro or Lmm154 (containing no eGFP or mCherry) or a positive control GFP or mCherry plasmid. Channel 1 – light channel, 2 – GFP Channel, 3 – mCherry Channel, 4 – merged. Images were collected in tiles of 9 at a size of 209.85 µm x 209.85 µm.

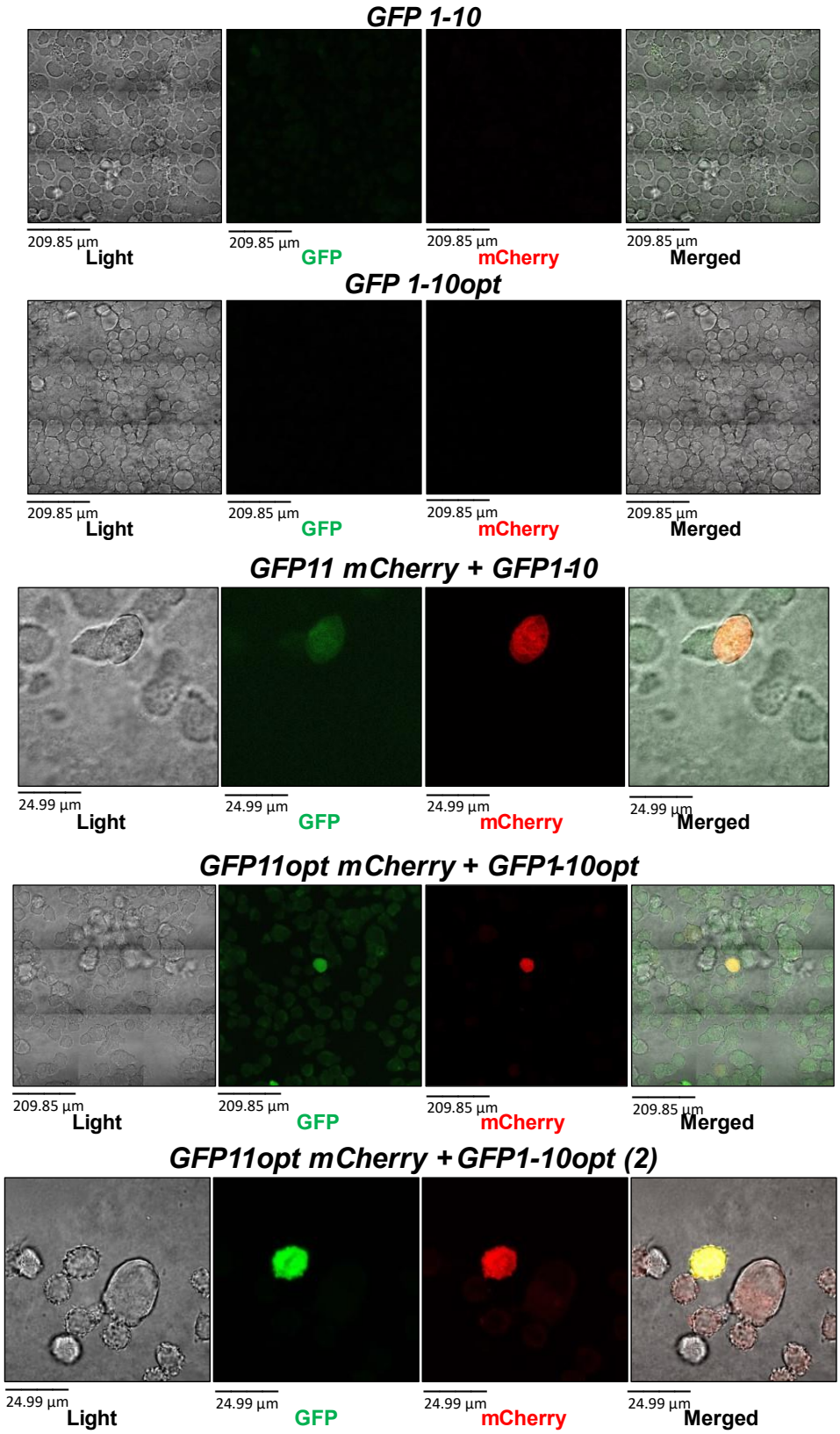


Figure 3.6. Confocal microscopy analysis after transfection of CHOK1SV GS-KO cells with different split eGFP containing constructs, including; eGFP1-10 alone, eGFP1-10$_{opt}$ alone, eGFP1-10 + eGFP11-mCherry, eGFP1-10opt alone, eGFP1-10opt + eGFP11opt-mCherry (bottom two panels). Channel 1 – light channel, 2 – GFP Channel, 3 – mCherry Channel, 4 – merged. Image sets 1,2 and 4 were collected in tiles of 9 at a size of 209.85 µm x 209.85 µm. Sets 3 and 5 were collected at a size of 74.97µm x 74.97µm.

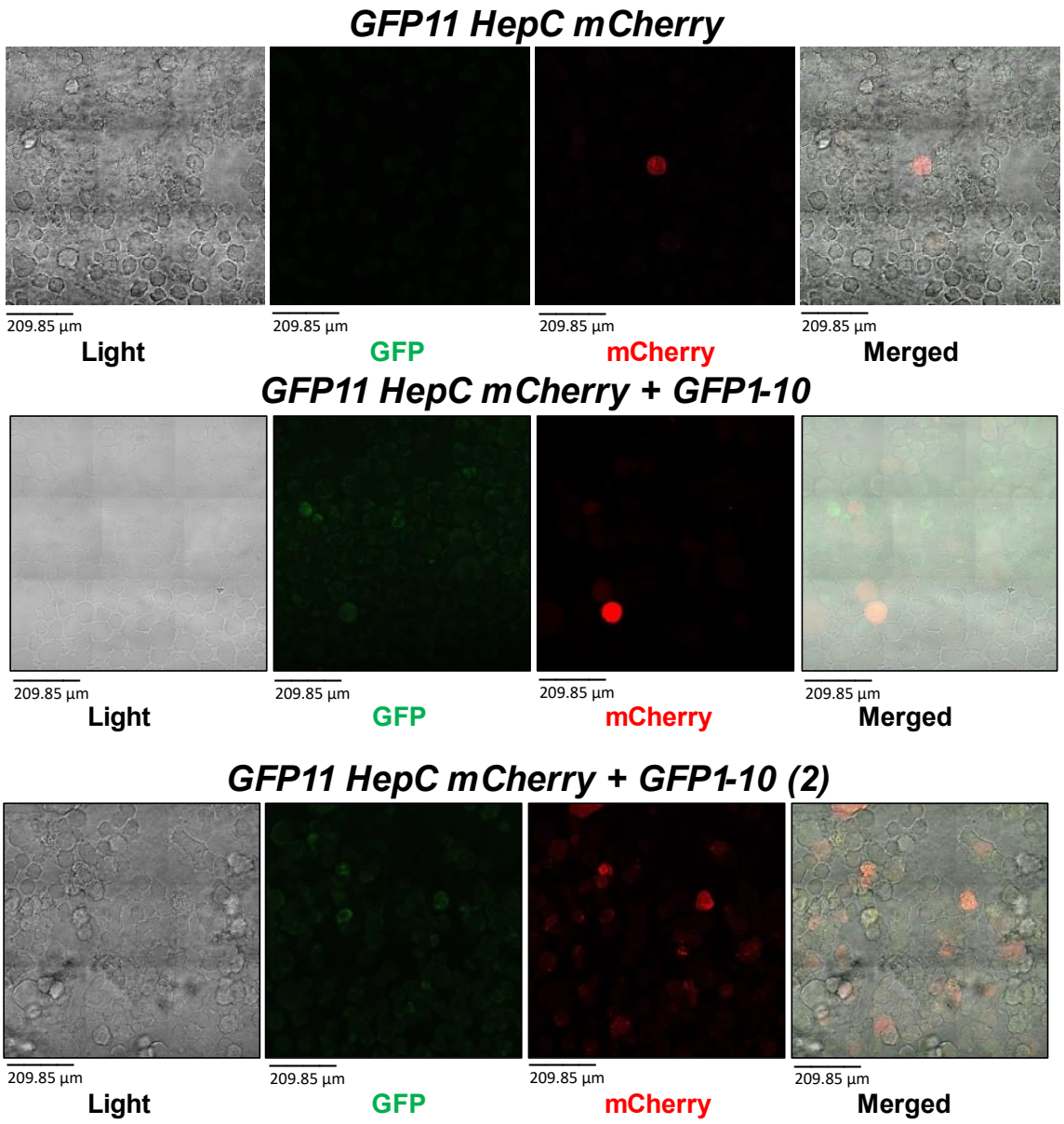


Figure 3.7. Confocal microscopy analysis after transfection of CHOK1SV GS-KO cells with different split eGFP containing constructs with the eGFP11-HepC ER signal peptide-mCherry construct, including; eGFP11-HepC-mCherry alone and eGFP11-HepC-mCherry + eGFP1-10 (bottom two panels). Channel 1 – light channel, 2 – GFP Channel, 3 – mCherry Channel, 4 – merged. Images were collected in tiles of 9 at a size of 209.85 µm x 209.85 µm.

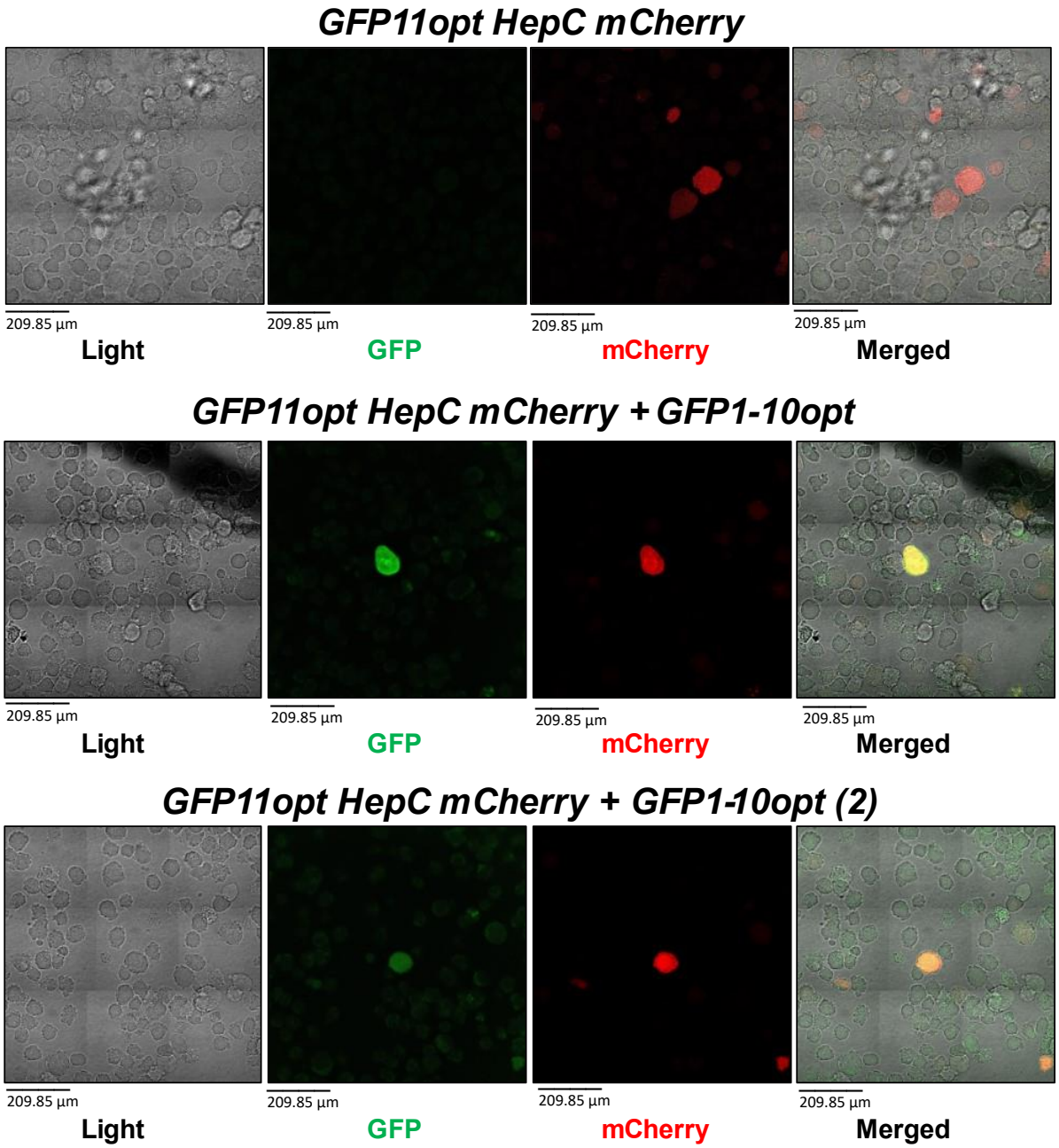


Figure 3.8. Confocal microscopy analysis after transfection of CHOK1SV GS-KO cells with different split eGFP containing constructs with the eGFP11opt-HepC ER signal peptide-mCherry construct, including; eGFP11-HepC-mCherry alone and eGFP11-HepC-mCherry + eGFP1-10opt (bottom two panels). Channel 1 – light channel, 2 – GFP Channel, 3 – mCherry Channel, 4 – merged. Images were collected in tiles of 9 at a size of 209.85 µm x 209.85 µm.

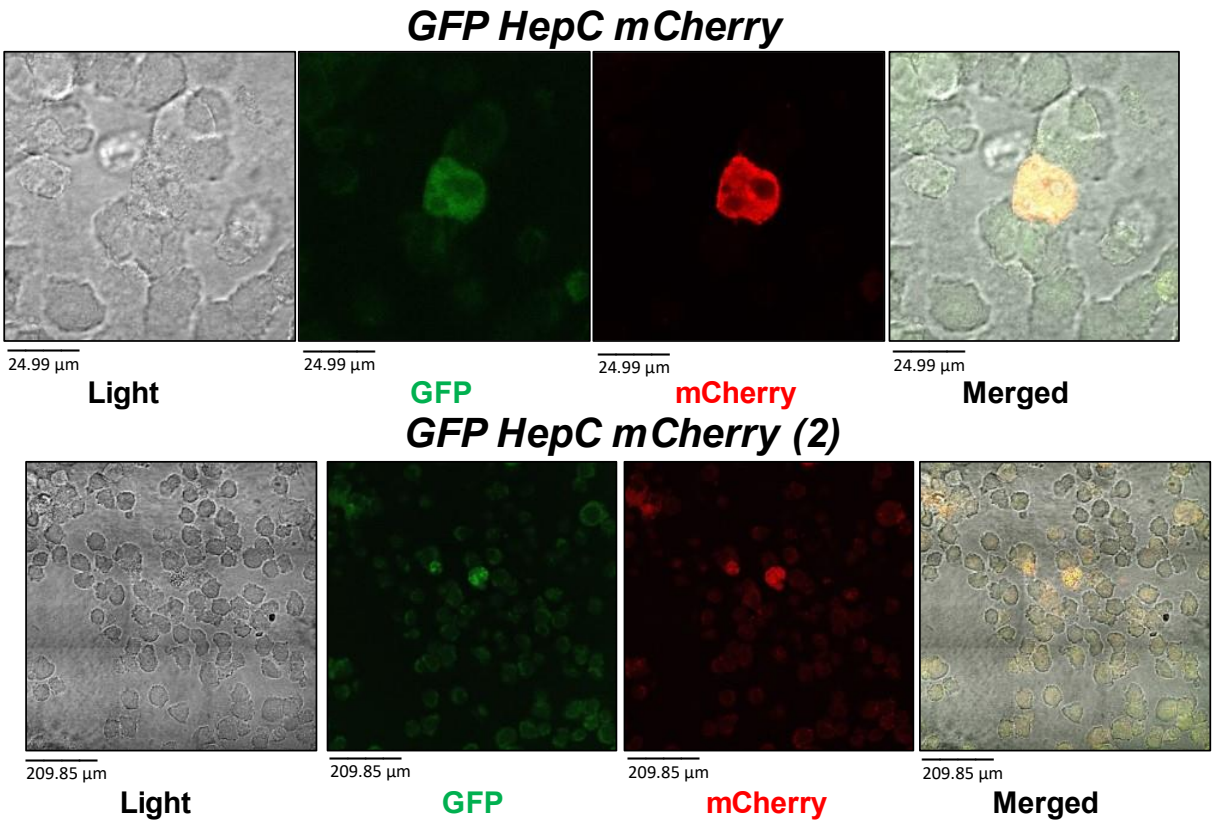


Figure 3.9. Confocal microscopy analysis after transfection of CHOK1SV GS-KO cells with the eGFP-HepC ER signal peptide sequence-mCherry construct. Channel 1 – light channel, 2 – GFP Channel, 3 – mCherry Channel, 4 – merged. The first set Images were collected at a size of 74.97 µm x 74.97 µm. The second set was collected in tiles of 9 at a size of 209.85 µm x 209.85 µm.

Figure 3.5 shows a set of controls confirming that transfection of the cells with the empty 3.1 hygro and Lmm154 plasmids resulted in no detectable fluorescence in the GFP and mCherry channel as expected. In the light channel there were clear outlines of cells. With the positive GFP and mCherry controls it was evident that only a few cells were expressing either GFP or mCherry, suggesting that the transfection efficiency was very low, likely due to the nature of transient transfections and the electroporation method used. The images show a 209.85 x 209.85 µm sized (9 tiles) area. The images shown are representative of the general slide. It is noticeable that throughout all the images cell outlines and shapes are visible and faint in the GFP channel. This can be attributed to very faint background auto-fluorescence of cells and was not considered as actual GFP fluorescence. The mCherry positive control (figure 3.5) and GFP positive control (figure 3.5) showed clear GFP and mCherry fluorescent protein present. Additionally, there was no evidence of mCherry fluorescence in the GFP control, or vice versa.

The next set of data show the fluorescence when GFP1-10 or GFP1-10_{opt} (Figure 3.6) constructs were transfected alone or with the corresponding eGFP11 (or eGFP11opt)-mCherry constructs to confirm the split GFP system could be reproduced in this work and GFP expression observed. Both transfection with either the GFP1-10 or GFP1-10_{opt} constructs (Figure 3.6) show neither GFP nor mCherry fluorescence, suggesting that the GFP11/$_{opt}$ (11^{th} beta sheet) is necessary for fluorescence. It is noted that all GFP11/$_{opt}$ molecules were delivered as fusions to an mCherry protein (previous work has shown that fused GFP11 can still associate with GFP1-10) (48). This serves two purposes; 1 the mCherry expression is a marker to show that GFP11/$_{opt}$ is present and 2, the fusion acts to stabilise the GFP11 which would otherwise be a small peptide and prone to rapid turnover. When the GFP11/$_{opt}$ containing constructs were co-transfected with GFP1-10/$_{opt}$ (figures

3.6), GFP fluorescence was detected. It was also clear to see from figure 3.6 that the optimised versions of GFP1-10 and GFP11 gave brighter and more intense GFP fluorescence. Figures 3.6 also shows that GFP11/$_{opt}$ without GFP1-10$_{opt}$ does not result in the detection of any GFP fluorescence and only mCherry fluorescence. These data confirm that GFP11/$_{opt}$ and GFP1-10/$_{opt}$ together are necessary to detect the green fluorescence and confirm that the eGFP11 and eGFP1-10 split GFP approach was established for further studies in the laboratory.

The next set of transient transfections looks to see if the use of the HepC ER signal sequence could direct mCherry to the ER and carry the GFP11 sequence such that this could associate with GFP1-10 and result in green fluorescence. In the HepC signal peptide transfection construct set, GFP-HepC-mCherry; a full length GFP was added to the front of the signal peptide. Figure 3.9 shows there was both an mCherry fluorescence signal and green fluorescence following transfection of the cells with this construct. The green fluorescence seen in almost all the cells across the 9 tiles is mostly background fluorescence as the power of the microscope was increased to see the actual fluorescence, which was weak. It was apparent whilst identifying cells under the microscope that transfection of this construct gave fluorescence in fewer cells than all other transfections, as this was the only identifiable cell with green and red fluorescence. This is unlikely to be because the transfection efficiency was worse than other samples, but rather may be related to problems with protein synthesis caused by the size of the full length GFP at the 5' end of the HepC signal peptide.

When the GFP11-HepC-mCherry + GFP1-10 constructs were co-transfected (figure 3.7) there was clear evidence of mCherry fluorescence. In all the images collected, where strong mCherry fluorescence was observed in such a co-transfection, then

there was only weak GFP fluorescence. This could be due to the low fluorescence generally from association of split wild type GFP, where a similar GFP fluorescence was observed (figure 3.6, GFP11-mCherry + GFP1-10). Additionally, this may signify that GFP11 on the HepC signal peptide is not cleaved well or is prevented from associating with GFP1-10 as discussed in the discussion section of this thesis.

In contrast, when the eGFP11 and eGFP1-10 optimised versions were used with the GFP11opt-Hepc-mCherry (Figure 3.8) there was clear evidence of the optimised split GFP system working from the HepC signal peptide format. This data suggests that it is the wild type split GFP recombination shortcomings rather than the HepC signal peptide preventing association of the wild type GFP11 with the GFP1-10. However, this data does not show if the mCherry was cleaved from the HepC ER signal peptide or if it was still intact, and it does not show if the GFP11 was cleaved from the HepC ER signal peptide or if it is still attached to it. In order to further address this question western blot analysis was undertaken to estimate the size of the products expressed.

3.4 mCherry and GFP protein analysis following transfection by western blotting

Although the fluorescence microscopy results reported in section 3.3 confirm the association of eGFP1-10 with eGFP11 when eGFP11 is fused to mCherry and the HepC ER signal sequence, this information gives no information on whether the HepC ER signal sequence is cleaved to liberate mCherry from the signal sequence and eGFP11. In order to investigate this, western blotting was used to probe for GFP and mCherry. Although SDS-PAGE is likely to destroy any interaction between eGFP1-10 and eGFP11, whether the eGFP11-HepC signal peptide-mCherry fusion is still intact can be evaluated based on size by SDS-PAGE followed by western

blotting for mCherry. The intact fusion would have an expected size of 30.7 kDa (70) whilst cleaved mCherry has a size of 26.7 kDa. Further, for the GFP-HepC-mCherry fusion an even greater size difference should be apparent between the fusion protein uncleaved and the mCherry cleaved from this.

SDS-PAGE followed by immunoblotting was therefore undertaken on samples from two separate transient transfections with the split GFP constructs. In the first set, the samples were from the same transient transfections as the samples used in the confocal microscopy imaging described in section 3.3. An additional set of transfections was also completed, using the same methods and the same transfections/co-transfections, for further analysis. Immunoblots reported are undertaken with mCherry as this was the primary target to which the GFP11 was fused and thus a change in size upon cleavage would be apparent. GFP probing was also undertaken but was largely unsuccessful. This is likely because only the native, full eGFP is likely to have been recognised by the antibody and any eGFP1-10-eGFP11 complex would most likely dissociate under the SDS-PAGE conditions used for analysis. Some protein bands on the GFP blot were visible, but the β-actin blot was still blank.

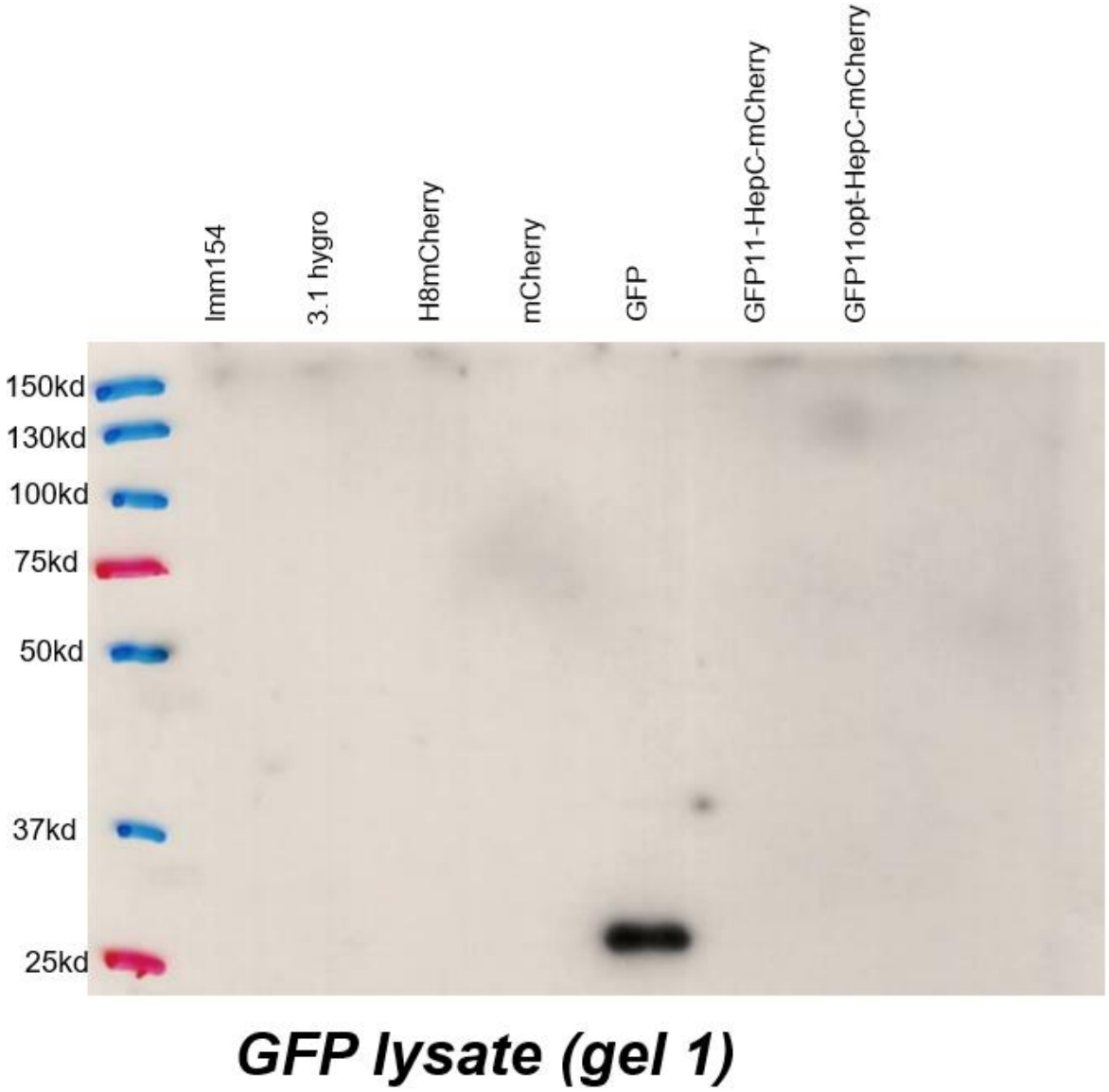


Figure 3.10 GFP lysate (Gel 1). Western blot analysis for GFP expression in cells transiently transfected/co-transfected with split GFP and HepC signal peptide constructs). The transfections were split across 2 gels (Figures 3.10 and 3.11) and contained intracellular protein. Gel 1: Lane 1 – ladder, Lane 2 – Lmm154, Lane 3 – 3.1 hygro, Lane 4 – H8mCherry, Lane 5 – mCherry, Lane 6 – GFP, Lane 7 – GFP11-HepC-mCherry, Lane 8 – GFP11$_{opt}$-HepC-mCherry. There is a clear band in lane 6 just above 25 kDa indicative of full length GFP (expected size of GFP approximately 28 kDa).

As expected, and shown in Figure 3.10, when probing for GFP here was no bands present in empty vector control transfections (Lmm154 and 3.1 hygro) or mCherry transfected samples. Further, there was no GFP band signal present when GFP11-HepC-mCherry constructs were transfected alone, as would be expected (Figure 3.10). However, in the positive control GFP transfected sample here was a band at approximately 28 kDa, the expected size of GFP (Figure 3.10).

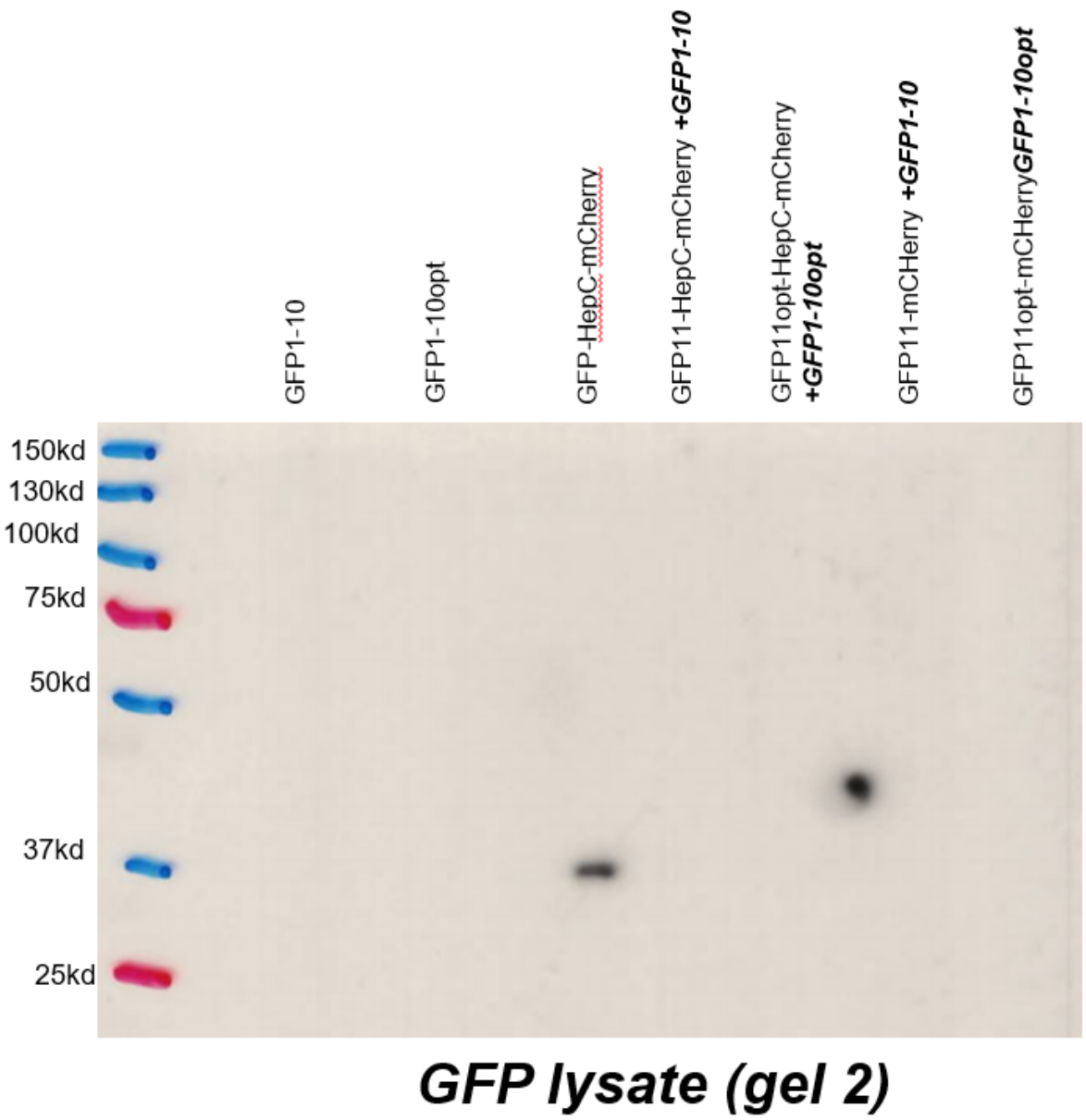


Figure 3.11 GFP lysate (Gel 2). Western blot analysis for GFP expression in cells transiently transfected/co-transfected with split GFP and HepC signal peptide constructs). The transfections were split across 2 gels (Figures 3.10 and 3.11) and contained intracellular protein. Gel 2: Lane 1 – ladder, Lane 2 – GFP1-10, Lane 3 – GFP1-10_{opt}, Lane 4 – GFP-HepC-mCherry, Lane 5 – GFP11-HepC-mCherry + GFP1-10, Lane 6 – $GFP11_{opt}$-HepC-mCherry + GFP1-10_{opt}, Lane 7 – GFP11- mCherry + GFP1-10, Lane 8 – $GFP11_{opt}$-mCherry + GFP1-10_{opt}. There is a small band in lane 4 at approximately 37 kDa. This blot was exposed for 10 min.

None of the co-transfections between GFP11 constructs and GFP1-10 constructs showed any GFP bands (Figure 3.11), further evidence that the antibody did not recognise the GFP1-10 on its own and that any GFP11-GFP1-10 complex disassociated under the SDS-PAGE conditions used here. In the full length GFP-

HepC-mCherry transfected sample a band at approximately 37 kDa was observed (Figure 3.11). This is too small to be the entire intact GFP-HepC-mCherry fusion

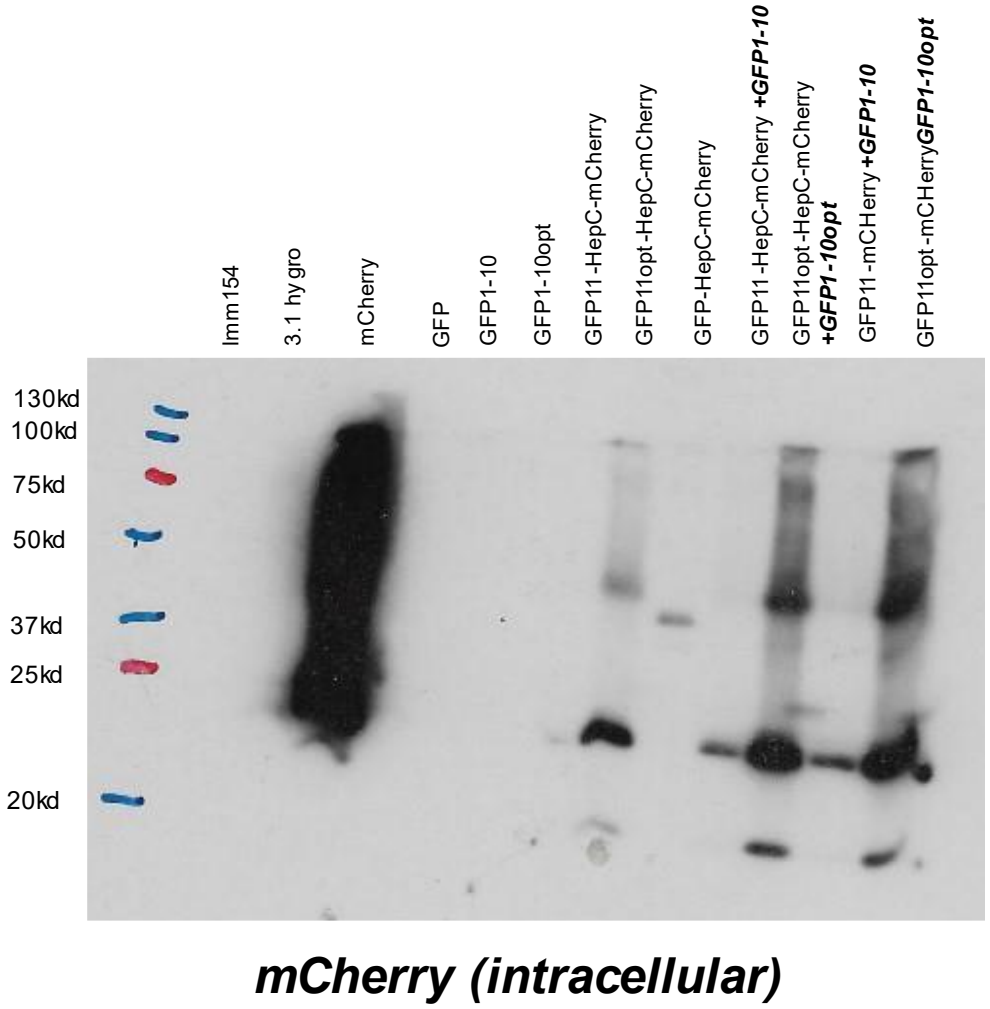


Figure 3.12 mCherry (intracellular). Western blot analysis for mCherry expression in cells transiently transfected/co-transfected with split GFP and HepC signal peptide constructs. The blot contains intracellular protein collected from the same transfections as the confocal images in Figures 3.5-3.9. Lane 1 – ladder, Lane 2 – Lmm154, Lane 3 – 3.1 hygro, Lane 4 – mCherry, Lane 5 – GFP, Lane 6 – GFP1-10, Lane 7 – GFP1-10_{opt} Lane 8 – GFP11-HepC-mCherry, Lane 9 – $GFP11_{opt}$-HepC-mCherry, Lane 10 – GFP-HepC-mCherry, Lane 11 – GFP11-HepC-mCherry + GFP1-10, Lane 12 – $GFP11_{opt}$-HepC-mCherry + GFP1-10_{opt}, Lane 13 – GFP11- mCherry + GFP1-10, Lane 14 – $GFP11_{opt}$-mCherry + GFP1-10_{opt}. There are multiple bands observed in lanes 4,8,10,11,12,13,14. The majority are at 25 kDa, mCherry has an expected size of 26.7 kDa.

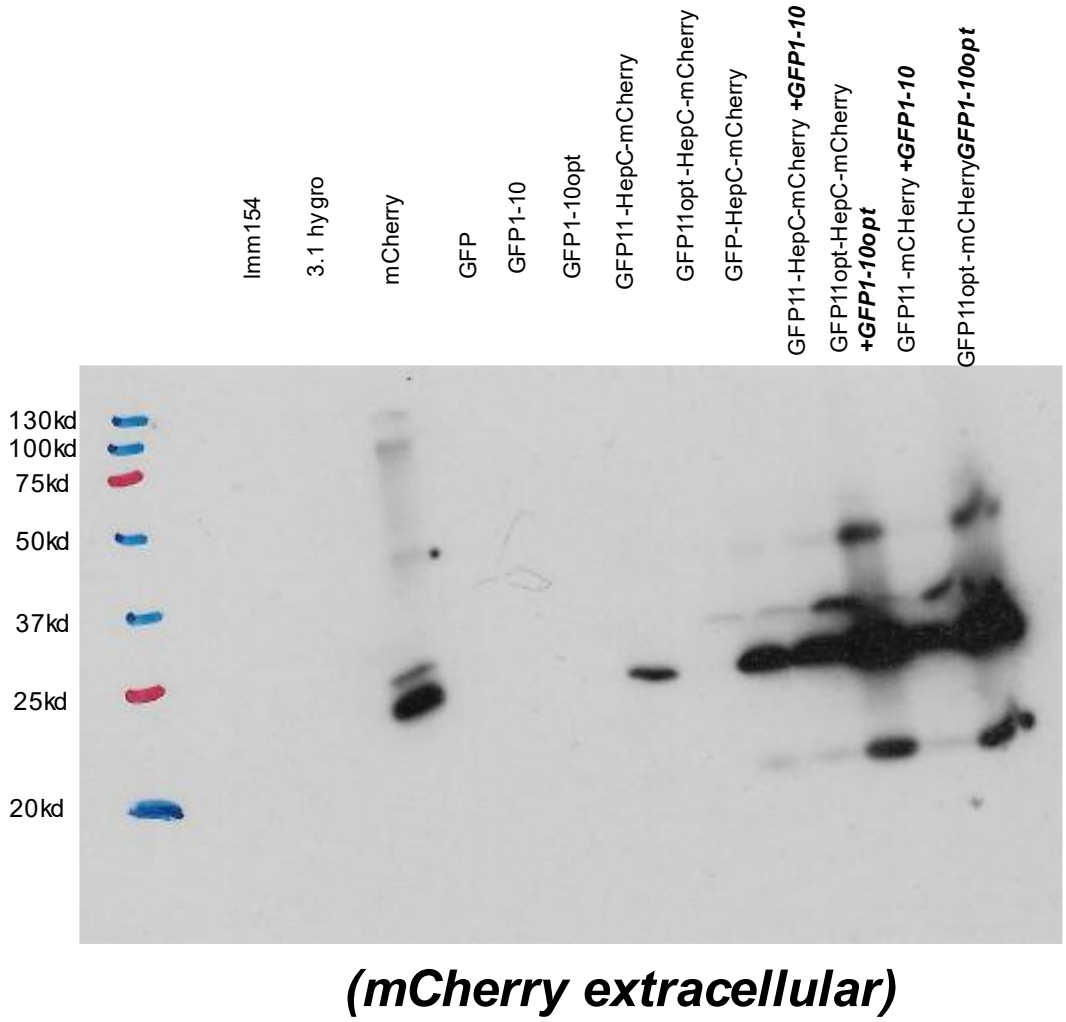


Figure 3.13 mCherry (extracellular). Western blot analysis for mCherry expression in cells transiently transfected/co-transfected with split GFP and HepC signal peptide constructs. The blot contains secreted/extracellular protein collected from the same transfections as the confocal images in Figures 3.5-3.9. Lane 1 – ladder, Lane 2 – Lmm154, Lane 3 – 3.1 hygro, Lane 4 – mCherry, Lane 5 – GFP, Lane 6 – GFP1-10, Lane 7 – GFP1-10_{opt} Lane 8 – GFP11-HepC-mCherry, Lane 9 – $GFP11_{opt}$-HepC-mCherry, Lane 10 – GFP-HepC-mCherry, Lane 11 – GFP11-HepC-mCherry + GFP1-10, Lane 12 – $GFP11_{opt}$-HepC-mCherry + GFP1-10_{opt}, Lane 13 – GFP11- mCherry + GFP1-10, Lane 14 – $GFP11_{opt}$-mCherry + GFP1-10_{opt}. There are multiple bands observed in lanes 4,8,10,11,12,13,14. The majority are at approximately 25 kDa, mCherry has an expected size of 26.7 kDa.

Figures 3.12 and 3.13 show the intracellular and extracellular mCherry protein content from the transient transfections/co-transfections from the same samples seen in section 3.31. The blots showed banding patterns consistent with the presence of mCherry of approximately 27 kDa in many samples along with both higher and lower molecular weight species. In figure 3.12, lane 4 there was a large smear across the whole lane. A one second exposure of film revealed a clearer

band (not shown) at around 25 kDa suggesting this is the mCherry protein as expected in this sample. In figure 3.12, lanes 9 and 11-14 contained a band at around 25 kDa, again suggesting this is a cleaved mCherry protein. These samples all come from GFP11-HepC-mCherry transfected cells and therefore the full fusion protein would give a mass higher than the approximately 27 kDa observed. Notably these lanes contain 2 higher bands, the first just above at 37 kDa is likely to be un-cleaved fusion proteins and the highest bands are likely mCherry dimers. Further, in Lane 10 there was a band at approximately between 25kDa and 37kDa which may be a full-length fusion protein although from this data it is difficult to confirm the identity of this band. Lanes 9,12 and 14 also contained bands at 20 kDa, indicative of sample degradation, possibly due to protease activity.

As the mCherry should be directed into the ER and then cleaved from the HepC signal peptide and then secreted out of the cell, the cell culture supernatant was also probed for the presence of mCherry (Figure 3.13). There were similar banding patterns to the intracellular blot (Figure 3.12), but with some subtle differences. The mCherry control (lane 4) is visible which contains an IgG heavy chain ER signal sequence and can clearly be seen secreted into the supernatant. There was one major band corresponding to mCherry and a fainter, higher molecular weight band that likely contains misprocessed signal sequence. This would need to be checked by mass spectrometry. Nevertheless, this acts as a good control to compare the other GFP11-HepC-mCherry transfected samples to. Importantly, in all the GFP11-HepC-mCherry construct transfected samples, except for the GFP11opt-HepC-mCherry sample, a strong band at approximately 25 kDa was observed (Figure 3.13). this is strong evidence that the HepC signal peptide successfully directs the mCherry to the ER even with the GFP11 fused to it and is correctly cleaved from the mCherry. However, there was also the presence of a fainter band of approximately

30 kDa (Figure 3.13) that may correspond to incompletely processed mCherry or mCherry with the HepC signal sequence still attached (but the GFP11 cleaved which is possible with the HepC signal sequence where cleavage occurs before and after the sequence). The lack of a signal in Lane 9 is an anomaly, likely a sample loading issue. The extracellular blot also had a similar pattern of lower molecular weight bands at approximately 20 kDa in lanes 12 and 14.

As the initial samples suggested secreted and cleaved mCherry from the HepC ER signal sequence and thus release of the GFP11, a further transfection was undertaken to confirm this and obtain a clearer picture of the protein banding patterns for the transient expression of the different constructs. These transfections did not directly relate to the confocal images presented in section 3.3, but the blot data collected gives a further insight into the cleavage patterns when using the HepC signal peptide to direct a polypeptide to the ER and look to utilise the sequence N-terminal of the cleaved HepC sequence to complement another protein (e.g. eGFP11 to complement eGFP1-10). The resulting western blots from this further transfection experiment are shown in figures 3.14-3.15.

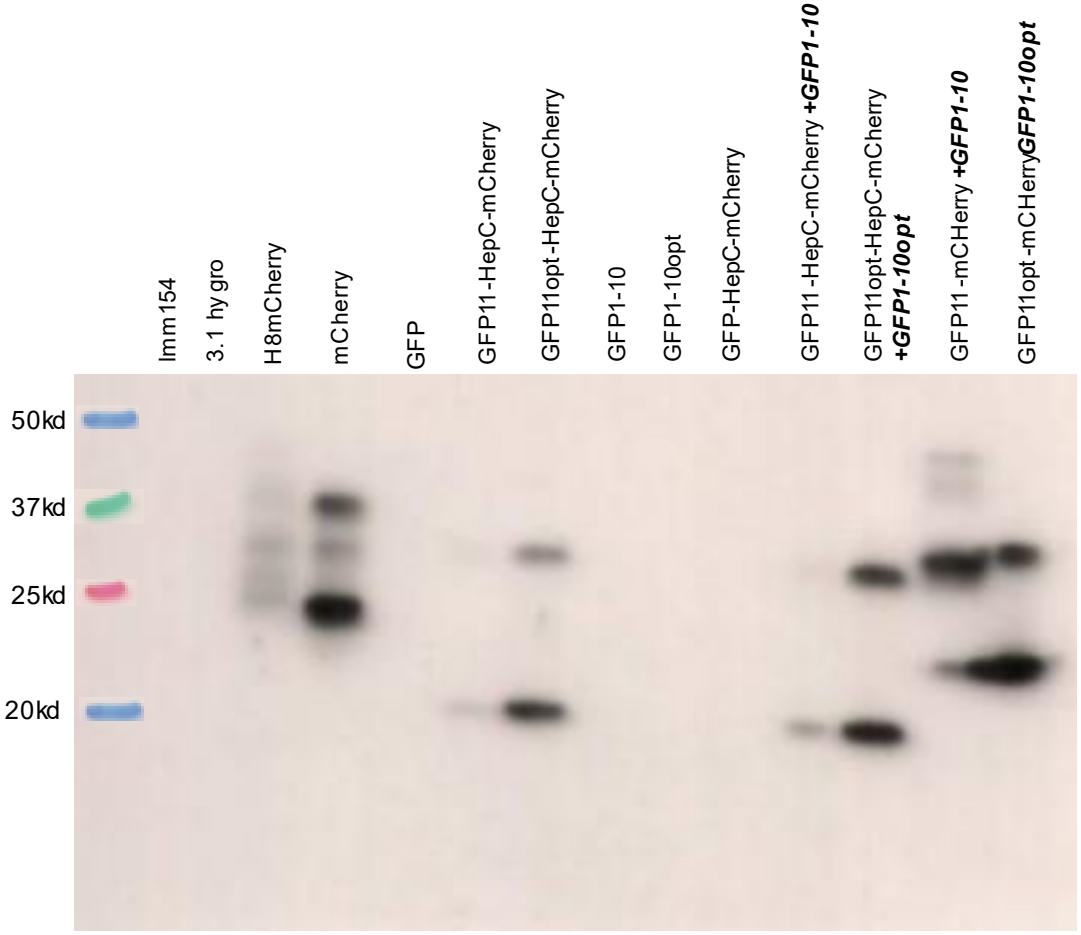


Figure 3.14 mCherry (intracellular). Western blot analysis for mCherry expression in a second experiment of CHOK1SV GS-KO® cells transiently transfected/co-transfected with split GFP and HepC signal peptide constructs. The blot contains intracellular protein. Lane 1 – ladder, Lane 2 – Lmm154, Lane 3 – 3.1 hygro, Lane 4 – mCherry, Lane 5 – GFP, Lane 6 – GFP1-10, Lane 7 – GFP1-10_{opt} Lane 8 – GFP11-HepC-mCherry, Lane 9 – $GFP11_{opt}$-HepC-mCherry, Lane 10 – GFP-HepC-mCherry, Lane 11 – GFP11-HepC-mCherry + GFP1-10, Lane 12 – $GFP11_{opt}$-HepC-mCherry + GFP1-10_{opt}, Lane 13 – GFP11- mCherry + GFP1-10, Lane 14 – $GFP11_{opt}$-mCherry + GFP1-10_{opt}. There are multiple bands observed in lanes 4,5,7,8,12,13,14.

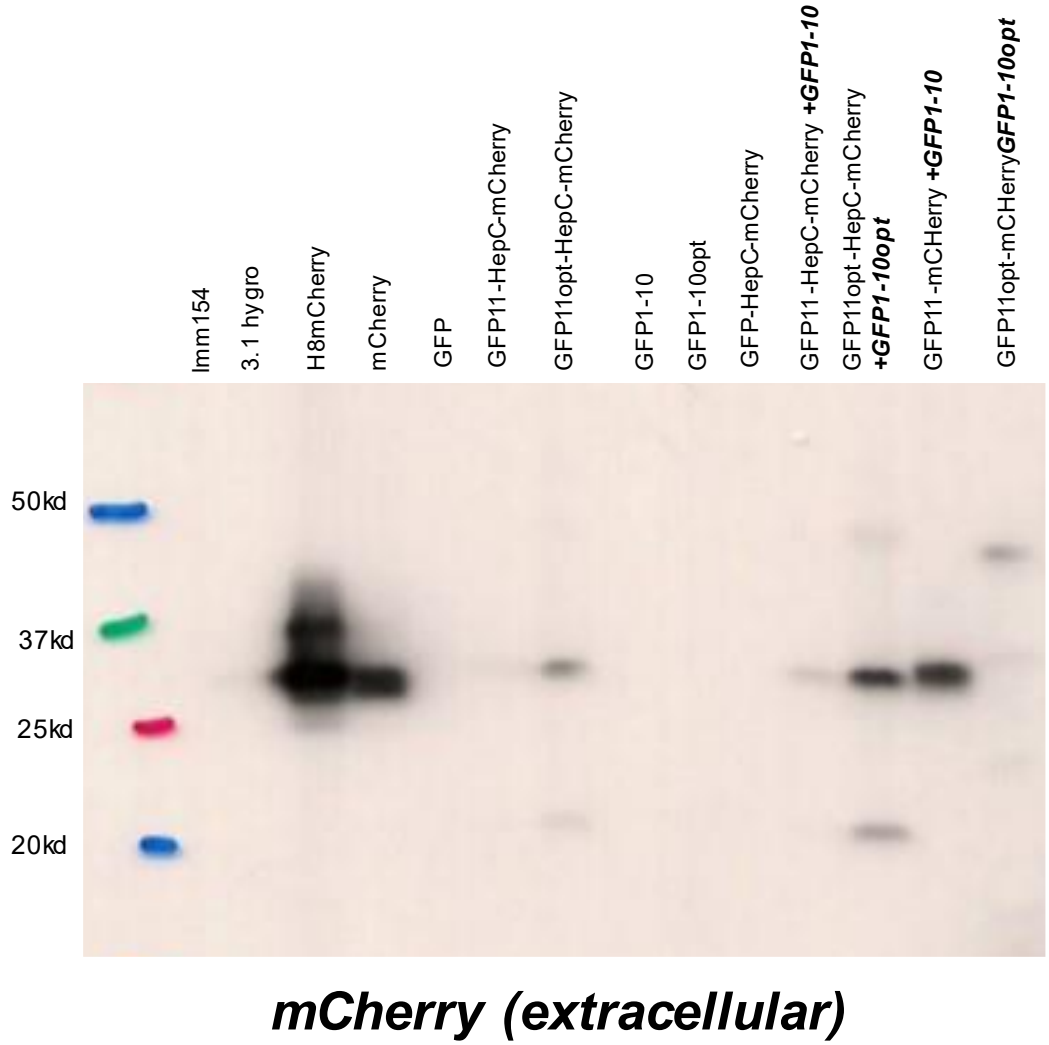


Figure 3.15 mCherry (extracellular). Western blot analysis for mCherry expression in a second experiment of CHOK1SV GS-KO® cells transiently transfected/co-transfected with split GFP and HepC signal peptide constructs. The blot contains secreted/extracellular protein. Lane 1 – ladder, Lane 2 – Lmm154, Lane 3 – 3.1 hygro, Lane 4 – mCherry, Lane 5 – GFP, Lane 6 – GFP1-10, Lane 7 – GFP1-10_{opt} Lane 8 – GFP11-HepC-mCherry, Lane 9 – $GFP11_{opt}$-HepC-mCherry, Lane 10 – GFP-HepC-mCherry, Lane 11 – GFP11-HepC-mCherry + GFP1-10, Lane 12 – $GFP11_{opt}$-HepC-mCherry + GFP1-10_{opt}, Lane 13 – GFP11- mCherry + GFP1-10, Lane 14 – $GFP11_{opt}$-mCherry + GFP1-10_{opt}.

The blots in figures 3.14 and 3.15 show the analysis for mCherry expression from the same 14 transfections as the previous western blots in figures 3.12-3.13. There was a clear distinction between the intracellular and extracellular protein banding in lane 4: H8mCherry and lane 5: mCherry with the secreted H8 version being more intense in the secreted sample than the intracellular as expected and this being reversed for the mCherry without an ER signal sequence. Again, the presence of some secreted H8mCherry containing unprocessed signal sequence was evident in the secreted protein samples (Figure 3.15).

Perhaps most interesting is the pattern of banding for the GFP11-HepC-mCherry samples. Transfections with the GFP11/opt-HepC-mCherry in lanes 7 and 8 revealed a band above 25 kDa in the approximate expected position for a signal peptide processed mCherry in both the intracellular and extracellular blots. It is noted that the optimised version of GFP11 (GFP11opt) seemed to give more mCherry expression and cleavage. There were also protein bands at 20 kDa which are clearer on the intracellular blot and which once again maybe protein degradation or incorrect cleavage. These bands are also seen in the other GFP11/opt-HepC-mCherry + GFP1-10/opt transfections in lanes 13 and 12 (Figure 3.14 and 3.15).

Figures 3.14 and 3.15 showed no evidence in either the intracellular or extracellular blot of mCherry protein in lane 11 (GFP-HepC-mCherry), in contrast to previous blots. Lanes 12-15 contain GFP11 containing constructs co-transfected with GFP1-10. In the intracellular blot there was a band around 25 kDa in lanes 12-13 from the GFP11-HepC-mCherry constructs which is likely to be due to the presence of mature mCherry protein cleaved from the GFP11-HepC before being secreted. In lanes 14-15 the intracellular protein bands from the GFP11-mCherry construct for mCherry were just above 25 kDa and were slightly larger in size than those in lanes 12-13. In the GFP11/opt-mCherry construct the GFP11 is connected/linked to mCherry via a small linker so that both parts of the molecule are translated in one hybrid fusion protein that is not cleaved. This means it is larger in size (by the size of GFP11 plus the linker) than the mCherry from the HepC cleaved constructs. As these bands from the GFP11-mCherry are larger than those of the GFP11/opt-HepC-mCherry (lanes 12-13) it strongly suggests that the HepC signal peptide is facilitating cleavage of the mCherry molecule correctly. Interestingly, in the extracellular secreted samples only a band corresponding to mature mCherry is

observed (Figure 3.15) once again suggesting that this material is being cleaved and processed properly.

4.0 Discussion

When using mammalian cell expression systems to produce biotherapeutic targets, the gene(s) encoding the target protein must be introduced into the cell and stably assimilated into the genome for long term expression (71). Introduction of the DNA is usually undertaken using plasmid DNA. A further consideration is the need to direct the polypeptide(s) of interest into the secretory pathway so that the correct folding and assembly of the target protein can be achieved. This requires directing of the polypeptide into the ER which is facilitated via the presence of an ER signal peptide (27). The signal peptide is then cleaved off the polypeptide to yield the mature polypeptide.

In only a small proportion of cells will the plasmid DNA randomly integrate into the genome (although targeted integration approaches can also be used (22)). As a result, and as described in the introduction to this thesis, selection systems are used to enrich/select for those cells that have integrated the plasmid DNA introduced into the cells. Traditionally this was achieved using an antibiotic selection marker on the plasmid of interest such that only cells that had the antibiotic selection marker survive in the presence of the said antibiotic. For biopharmaceutical manufacturing in mammalian cells, the GS (glutamine synthetase) system is commonly used as a metabolic selection marker with CHO cell expression systems. In GS knockout host cells, cells can only grow if given exogenous glutamine. However, if the GS gene is introduced into the cell along with the biotherapeutic target gene(s) of interest, those cells that have integrated and express the GS gene can synthesise their own glutamine in the absence of exogenous glutamine. This selection is often coupled

with the use of the GS inhibitor MSX to select for those cells expressing higher amounts of GS. Although this system is well established, the selection system is still an indirect election method, that is to say it does not directly select on the amount of polypeptide or product the cell can produce but rather on the metabolic selection marker as a proxy. There is not necessarily a direct correlation between GS expression and that of the recombinant biotherapeutic target (17, 19). This project therefore set out to develop proof-of-concept that the ER signal peptide could be harnessed as a reporter of recombinant polypeptide synthesis amounts as a more direct selection method than the use of metabolic selection markers.

In order to investigate this, this study explored the use of different ER signal sequences to direct polypeptides to the ER, and whether these could also subsequently, once cleaved from the target polypeptide, compliment activity of a non-active protein to restore its activity. For this purpose, the split GFP system was selected as GFP can be split into two components as outlined in the introduction, a large GFP1-10 section and a smaller GFP11 component (47). The recombination of split GFP, GFP11; a helix forming peptide from the C-terminal, can associate back with the remainder of GFP (GFP1-10) to produce an active, fluorescing eGFP, in both and so termed ‘optimised’ and ‘native’ version (47, 51). In this study the plan was to express the non-fluorescent GFP1-10 and then look to complement this with GFP11 that was included into an ER signal peptide and could thus be released upon cleavage from the ER targeted main polypeptide to generate an active, full GFP molecule. In particular, a viral ER targeting signal sequence from Hepatitis C was investigated (44) (72) that had the GFP11 sequence included within a GFP11-HepC-mCherry fusion protein. Cleavage of the HepC ER signal sequence would result in release of full length mCherry (that could be distinguished from mCherry still fused to the GFP11-HepC signal peptide by SDS-PAGE analysis) and the

GFP11 fragment that could associate with the GFP1-10 molecule. Indeed, a discussed below, when using SDS-PAGE and immunoblotting, there was evidence to suggest that the eGFP11-HepC signal peptide-mCherry fusion was cleaved correctly to release the molecule as two separate peptides, mature, full length mCherry and eGFP11.

4.1 Confirmation of the successful generation of split GFP and HepC signal peptide plasmid DNA constructs

In this investigation it was initially essential to confirm that the split GFP and HepC signal peptide DNA strings had been inserted into the plasmid backbone correctly. Purified plasmid DNA form transformed *E.coli* cells were analysed by agarose gel electrophoresis, after appropriate restriction digest experiments, to release each gene component as stated in section 3.1. From all of the restriction digests completed, the band dropouts observed in Figure 3.2A and B showed that a DNA fragment of the correct size had been inserted into the plasmid backbones. Subsequent Sanger sequencing (66) revealed that the H7GFP1-10KDEL and H7GFP1-10optKDEL (construct 3 and 4, table 3.1) did not have a ATG start codon and therefore these constructs were not available for further experiments in this study. It was concluded that although it would be possible to either design new primers with a start codon in the 5' forward primer or use site directed mutagenesis to generate these constructs, they were not essential to delivering the proof-of-concept data this project set out to achieve. The sequencing data for all the other constructs produced is provided at request and was analysed with a multiple sequence alignment tool ("Clustal Omega < Multiple Sequence Alignment < EMBL-EBI," n.d.) and an open reading frame finder (68) and were found to contain the desired sequences.

4.2 Batch culture growth analysis of stably expressing split GFP1-10 CHOK1SV GS-KO® cells

In order to utilise the split GFP system to evaluate the potential of using GFP11 containing ER signal sequences to complement the GFP1-10 portion of GFP, CHOK1SV GS-KO® cell pools stably expressing the GFP1-10 molecule were initially generated using hygromycin as the selection agent. Once these pools were established, the batch culture growth analysis was undertaken to assess whether the expression of GFP1-10 had any impact on CHOK1SV GS-KO® cell growth. It is noted that the CHOK1SV GS-KO® host cell line, where the GS gene has been knocked out, has been specifically developed to improve the efficiency of CHO cell line construction (19) (73). Figure 3.4 reports the total viable cells (x10^6 cells/ml) and culture viability (%) of the batch culture experiments and shows that the growth profile as determined by these parameters across days 1-6 of batch culture was similar for all 3 cultures, the CHOK1SV GS-KO® host cell line and the two GFP1-10 expressing pools (the native and optimised versions of GFP1-10). By day 7 the GFP1-10opt cell number was actually elevated compared to the CHOK1SV GS-KO® mean. This data suggests that the stably integrated plasmid DNA, and the subsequent expression of the hygromycin selection marker and GFP1-10 molecule do not negatively affect the growth of the cell pools. It is noted that the actual expression amounts of the hygromycin selection marker and GFP1-10 molecules in these pools at the RNA or protein level was not known.

4.3 Confocal microscopy image analysis of transient expression of mCherry, GFP and the split GFP system

As outlined previously above, in this study the split GFP system was utilised to investigate using ER signal peptides fused with GFP11 could reconstitute with GFP1-10 and generate an active GFP molecule. In this investigation, initially the

GFP11 helix sequence was fused with mCherry to give a fusion protein. Other studies have previously fused mCherry to the GFP11opt sequence and shown that the GFP11, even when fused to mCherry, could associate with GFP1-10 and result in an active, fluorescent GFP molecule (49). This provided confidence that the same could be achieved here. Furthermore, in this same study the authors fused GFP1-10opt to ER membrane proteins (49) and showed that GFP recombination was also viable when both split GFP parts were fused to other proteins. In this project, the identified HepC signal sequence in its native form in the hepatitis C virus has a large core protein at its N-terminal end (43),(see figure 1.5 and 1.6). In the current study this core protein was replaced with the GFP11 fragment or full length GFP. These fusions were developed to test the hypotheses that full length GFP could be released via the HepC signal sequence mechanism or the GFP11 to complement GFP1-10.

When cells were transiently transfected with the full length GFP-HepC-mCherry construct (figure 3.9) there was weak GFP expression observed. This was in stark contrast to when the GFP11-HepC-mCherry construct was co-transfected with GFP1-10, whereby the split GFP system gave clear green, fluorescent cells (Figure 3.8) where mCherry was also observed. The GFP11 fused to the HepC ER targeting peptide contains a much smaller molecule at the N-terminal end of the HepC signal sequence (the front) than the full length GFP. In the case of the GFP11 optimised version this is only 16 amino acids and appeared to work very successfully, suggesting that the size of the polypeptide/protein N-terminal to the HepC signal sequence may impact the expression and/or cleavage of the molecule. Furthermore, it is suggested that the core protein has residues near its C terminus that affect it cleavage and could explain the ineffectiveness of full length GFP cleavage (43). It is noted that mCherry was always observed in cells where GFP was observed as

expected with these constructs confirming that polypeptide synthesis itself was not hindered.

Excitingly, the fluorescent images captured by confocal microscopy revealed that the split GFP system, both the native and optimised sequences, when part of a GFP11-HepC-mCherry fusion co-transfected with the corresponding GFP1-10, facilitated the reconstitution of the GFP1-10/opt part and GFP11/opt to produce molecules that fluoresced. Further, the data showed that a fused mCherry-GFP11/opt and the GFP1-10/opt were both required to be present to give green fluorescence (figures 3.5-3.9), with images where the constructs were transfected alone resulting in no green fluorescence being observed. This confirmed that the GFP11 section of the GFP11-HepC-mCherry fusion was able to reconstitute with the GFP1-10 but not that the HepC cleavage had occurred and mCherry secreted out of the cell as would be required in a recombinant protein expression system. This was investigated by western blot analysis as discussed further below.

The image analysis (Figures 3.7, 3.8 and last 2 images of 3.6) also confirmed that the optimised versions of the GFP11 and GFP1-10 sequences when used gave more intense green fluorescence than the native sequences as expected from previous reports (51). The images in figure 3.7 and figure 3.8 show a strong mCherry signal but the difference between the images in figures 3.7 and 3.8 are in the intensity of the GFP images. The optimised versions of the split GFP gave a more intense green signal and based upon this data it would suggest further studies should use this version. The optimised GFP11 is also considerably shorter in length than the native GFP11, a further potential advantage in using this system.

The split GFP system has previously been described by Kim *et al* for cell line development (74) as have Sleiman *et al* (75). Kim used an IRES sequence to link a heavy and light chain antibody to the split GFP system. They hypothesised that the transcription level of the light and heavy chains is highly correlated with that of split GFP recombination and therefore speculated that high GFP-expressing clones are high antibody-producing clone. The key difference in this investigation is the addition of the novel Hepatitis C signal sequence instead of an IRES to link the split GFP molecule and the target recombinant protein. The IRES system results in the synthesis of two independent polypeptides whilst in the system described here just one polypeptide is produced and therefore a direct relationship between the GFP11 and mCherry should be observed. This led to the hypothesis that a direct 1:1 ratio of split GFP molecules and recombinant protein should be produced and therefore a high expressing GFP clone in this system is much more likely to be a high protein producing clone. Ultimately (as discussed further below) the system would seek to replace the GFP system with a selection marker whereby an inactive selection marker is reconstituted by the signal sequence released upon cleavage of the HepC fragment (e.g., GS), allowing only cells with the reconstituted selection marker to survive in a relationship directly proportional to the amount of recombinant polypeptide being made.

The fluorescent image analysis therefore provided the basis of the proof-of-concept data showing that the split GFP system combined with the HepC signal peptide system could be used to reconstitute active GFP and that GFP activity was only observed in cells where mCherry was observed (i.e., directly linking recombinant mCherry product expression to GFP expression). The data was not, however able to reconcile if the GFP11/opt fragment and mCherry protein had been cleaved and released from the HepC signal sequence. To investigate if the fluorescent products

seen under the microscope were the result of cleavage of the HepC signal sequence (and hence release of GFP11-HepC from mCherry) the size of the products was evaluated by SDS-PAGE and western blotting.

4.4 mCherry and split GFP protein analysis following transient expression by western blotting

Western blotting analysis would provide crucial evidence on the outcome of the cleavage of the fusion proteins at the HepC site by probing for GFP and mCherry. Although SDS-PAGE analysis is likely to destroy any interaction between eGFP1-10 and eGFP11, to determine whether the eGFP11-HepC signal peptide-mCherry fusion was still intact was evaluated based on the expected sizes of the full-length product or cleaved product by SDS-PAGE followed by western blotting for mCherry. The intact fusion protein (GFP11-HepC-mCherry) would have an expected size of 30.7 kDa whilst cleaved, mature mCherry has a size of 26.7 kDa, therefore a size difference of 4kDa would be seen on SDS-PAGE. Further, for the GFP-HepC-mCherry fusion an even greater size difference should be apparent between the fusion protein un-cleaved and the mCherry cleaved from this.

As expected, no ‘intact’ reconstituted GFP1-10/opt-GFP11/opt interaction was observed, and split GFP protein recombination was not visible on the western blot. This will be due to the non-covalent interactions between these and the presence of the SDS in the samples. A GFP control was observed which verified that the antibody was working correctly and recognised intact GFP. Other studies have used antibodies that detect split GFP fusion proteins by binding to the non-split GFP part such as that reported by Foglieni *et al* (76). They used an anti-β1 antibody binding domain in their split GFP fusion protein to detect the GFP1-10 part that was fused

to a TDP-43 protein. Such an antibody could be used in the future to try and detect the presence of the reconstituted GFP.

As an alternative the mCherry section was probed to determine if cleavage was occurring. In this study an anti-mCherry antibody (ab167453) was used to probe the fusion proteins. The first set of western blot analyses (Figures 3.12 and 3.13) are from the same samples as the microscopy images shown in section 3.3.1 and show mCherry detected at approximately 25kDa. However, further SDS-PAGE and western blotting was needed to obtain a clearer picture on whether the fusion protein was being cleaved resulting in 'mature' mCherry without the GFP11-HepC present.

A second set of transfections was undertaken, material harvested, and the resulting western blots are shown in figures 3.14 and 3.15. Note that these samples do not come from the same set of transfections as the microscopy images shown. These blots give an interesting insight into how well the HepC signal sequence is cleaved. Figure 3.14 is the intracellular protein blot. In the GFP11/opt-mCherry construct the GFP11/opt is connected/linked to mCherry via a short linker so that both parts of the molecule are translated in one hybrid fusion protein that is not cleaved (there is no HepC sequence present). This means that the fusion is larger in size (by the size of GFP11 plus the linker) than the native mCherry from the HepC cleaved constructs. As the bands from the GFP11-mCherry are larger than those of the GFP11/opt-HepC-mCherry (lanes 12-13) it strongly suggests that the HepC signal peptide is facilitating cleavage of the mCherry molecule correctly. This is a key piece of data that suggests that the HepC signal sequence is both able to be cleaved and release the mCherry molecule and the GFP11/opt fragment. The extracellular blots show mature mCherry of the expected size from the GFP11-HepC-mCherry as expected too, although the GFP11-mCherry also shows presence of material that is

marginally higher in size. This material does not contain an ER signal sequence so its release may come from lysis of cells although this would need to be confirmed.

The western data suggests that the mCherry protein could be replaced by another target biotherapeutic protein such as an antibody chain or a single chain target molecule. Studies on the Hepatitis C signal sequence refer to this as the Core-E1 signal peptide (77). An internal signal sequence between the core protein and the envelope protein E1 targets the premature polypeptide to the endoplasmic reticulum membrane for translocation of E1 into the ER. The signal peptide then inserts itself into the membrane, and the E1 protein is cleaved by a signal peptidase and released into the ER. The signal peptide is then further processed by the intramembrane-cleaving protease SPP that allows the release of core protein from the ER membrane(43)(72)(44). The data presented here suggests that the unique internal and double cleavage of the HepC signal sequence is transferable to other peptides fused to each side. The next step of this investigation would be to confirm the site of cleavage, using mass spectrometry to confirm the sequence at the termini of the mCherry and provide further evidence that both peptides are released by the HepC signal sequence.

4.5 Future work

The work reported in his thesis has shown proof-of-concept that a GFP11/opt polypeptide fused to a HepC ER signal peptide can associate with a GFP1-10/opt molecule and generate an active, fluorescing GFP. There is much potential to continue to work with both the split GFP system and the HepC signal peptide. Firstly, to make use of the stable transfections described in section 3.2, transient transfections of GFP11/opt-HepC-mCherry constructs into stably integrated GFP1-10 CHOK1SV GS-KO® would potentially allow more cells with the reassembled

GFP a co-transfection is not required and all cells in theory express GFP1-10. Sleiman *et al* (75) have shown a robust high-throughput screening method for the identification and isolation of high antibody-producing clones using two-colour intracellular fluorescence reporters and FACS. Here, we would sort high protein producing clones in a FACS machine based on GFP expression and determine if this directly relates to mCherry expression (or another recombinant product that replaces mCherry). Sleiman *et al* (75) have used an IRES for cap-independent translation of fluorescence reporters eYFP and eGFP to give a direct link between fluorescence and protein production such as the proposed ER signal peptide system described here. When using the ER signal peptide approach, the GFP expression results directly from the presence of the ER signal sequence and hence there would be a direct link between polypeptide synthesis of the target protein and eGFP expression. In this case, the constructs already made in this study could be utilised and the mCherry fluorescence intensity measured by flow cytometry to see if there is a linear relationship between levels of green and red fluorescence. Ultimately a non-fluorescence metabolic marker system would be used as described elsewhere whereby the ER signal peptide would activate a metabolic selection marker, a completely novel approach to cell line selection.

In section 3.1.1 the constructs 3 and 4 (H7-eGFP1-10/opt KDEL) were missing a start codon. Re-cloning these constructs correctly and transiently transfecting them into the CHOK1SV GS-KO® host with mCherry-HepC-GFP11/opt constructs would localise the GFP1-10 to the endoplasmic reticulum due to the H7 ER signal sequence and the KDEL sequence on the end. As the mCherry-HepC-GFP11/opt is also localised to the ER; it could be determined if recombination of GFP in the ER can happen.

There have been continual developments to increase the effectiveness of the split GFP recombination. A limitation found to the split GFP system is the slow chromophore formation after re-complementation (78). Lundqvist *et al* have found a solution to this problem by using a pre-matured GFP1-10 protein. They accomplished this via capture of inclusion body purified GFP 1–10 on a solid support containing the GFP11 partner, followed by acid elution of the matured GFP1–10 and protein refolding (78). The pre-maturated GFP1–10 provided up to 150-fold faster signal generation compared to the non-maturated version. In addition, pre-maturated GFP1–10 significantly improved the ability of discriminating between Chinese hamster ovary (CHO) cell lines secreting GFP11-tagged erythropoietin protein at varying rates. Incorporating pre-matured GFP1-10 could help in future work when using the GFP1–10 as a reporter, especially for ranking individual cell lines based on secretion fluorescence by FACS.

As previously mentioned in this section and shown in the results and other reports (49), split GFP recombination can occur if the molecules are fused to other proteins. To overcome the anti-GFP antibody not being able to bind due to the SDS-PAGE conditions, adding another antibody binding domain to the GFP1-10/opt or GFP11/opt to visualise the protein on an SDS-PAGE and western blot could show in the case of GFP11/opt if cleavage is occurring. This would be required to show the complete system is working as anticipated as the data presented only shows that the mCherry protein is cleaved from the signal sequence and that GFP11/opt is still functional. Further development of GFP1-10/opt and GFP11/opt specific antibodies would also overcome this limitation.

Finally, CHOK1SV GS-KO® host cells were used throughout the study, a glutamine synthase (GS) knockout CHO cell variant. Glutamine is added back to allow growth

of the cells but removed after transfection of plasmid DNA containing the GS gene to allow metabolic selection for cells whereby the glutamine synthase gene has been stably incorporated into the genome, allowing the cells to generate their own glutamine. Using the same approach as for the split GFP, metabolic markers could be investigated to see if they can be 'split' and recombined. Indeed, the metabolic marker DHFR has already been shown to be able to be split and recombine in this fashion (79). The DHFR molecule could thus be split and used as a direct metabolic selection approach linked directly to polypeptide synthesis amounts. A similar approach may be possible with GS whereby a major portion is engineered into the host to be expressed constitutively whilst a smaller portion is linked to the Hep-C and recombinant target. This unique approach, if successful, would allow a direct relationship between the GS activity and cell selection and recombinant polypeptide expression. In theory this should be a much stronger correlation than when the GS and recombinant product are produced independently.

The two domains of GS (that needs to form a hexamer to be active) contribute to the formation of the active sites, each in an adjacent monomer (80) (81). Fusing the smaller C-terminal GS domain to the N-terminal of the HepC signal sequence and continuing to have target protein at the C-terminal end of the HepC signal sequence (as shown in this study) with the other larger N-terminal GS domain expressed constitutively in the cell (like the GFP1-10) would be a possible such system.

4.6 Conclusions

In conclusion this work has been able to show that the optimised split GFP system can be linked to a viral ER signal sequence linked to a second, target (in this case mCherry) recombinant product. This system resulted in recombining of the two split GFP fragments. SDS-PAGE and western blotting suggests that the HepC signal sequence was able to direct cleavage of the polypeptide to release the GFP11/opt-

HepC from the mCherry. The work also shows that larger proteins at the N-terminal end of the HepC signal sequence are not as well tolerated in terms of cleavage of the HepC signal sequence and likely interfere with the cleavage. Overall, the work reports on a potentially novel mechanism of metabolic selection in mammalian cells that directly links the activity of the metabolic enzyme to that of the target recombinant product that may provide a stronger correlation between selection and recombinant product selection than current systems.

5.0 References

1. The Compartmentalization of Cells - Molecular Biology of the Cell - NCBI Bookshelf. .
2. Lu, J., T. Wu, B. Zhang, S. Liu, W. Song, J. Qiao, and H. Ruan. 2021. Types of nuclear localization signals and mechanisms of protein import into the nucleus. Cell Commun. Signal. 2021 191. 19: 1–10.
3. Gemmer, M., and F. Förster. 2020. A clearer picture of the ER translocon complex. J. Cell Sci. 133.
4. Akopian, D., K. Shen, X. Zhang, and S. Shan. 2013. Signal Recognition Particle: An essential protein targeting machine. Annu. Rev. Biochem. 82: 693.
5. Berks, B.C. 2015. The Twin-Arginine Protein Translocation Pathway. Annu. Rev. Biochem. 84: 843–864.
6. Natale, P., T. Brüser, and A.J.M. Driessen. 2008. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms. Biochim. Biophys. Acta - Biomembr. 1778: 1735–1756.
7. Budge, J.D., T.J. Knight, J. Povey, J. Roobol, I.R. Brown, G. Singh, A. Dean, S. Turner, C.M. Jaques, R.J. Young, A.J. Racher, and C.M. Smales. 2020. Engineering of Chinese hamster ovary cell lipid metabolism results in an expanded ER and enhanced recombinant biotherapeutic protein production. Metab. Eng. 57: 203.
8. Blobel, G., and D.D. Sabatini. 1971. Ribosome-Membrane Interaction in Eukaryotic Cells. In: Biomembranes. Springer US. pp. 193–195.
9. Lindemann, D., T. Pietschmann, M. Picard-Maureau, A. Berg, M. Heinkelein, J. Thurow, P. Knaus, H. Zentgraf, and A. Rethwilm. 2001. A Particle-Associated Glycoprotein Signal Peptide Essential for Virus Maturation and

Infectivity. J. Virol. 75: 5762–5771.

10. Martoglio, B., and B. Dobberstein. 1998. Signal sequences: More than just greasy peptides. Trends Cell Biol. 8: 410–415.
11. Tjio, J.H., and T.T. Puck. 1958. Genetics of somatic mammalian cells : ii. Chromosomal constitution of cells in tissue culture. J. Exp. Med. 108: 259.
12. Li, F., N. Vijayasankaran, A. (Yijuan) Shen, R. Kiss, and A. Amanullah. 2010. Cell culture processes for monoclonal antibody production. MAbs. 2: 466.
13. T, Lai., Y. Yang, and SK. Ng. 2013. Advances in Mammalian cell line development technologies for recombinant protein production. Pharmaceuticals (Basel). 6: 579–603.
14. JY, Kim., Kim YG, and Lee. GM. 2012. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. Appl. Microbiol. Biotechnol. 93: 917–930.
15. D, Ghaderi., Zhang. M, Hurtado.-Ziola. N, and Varki. A. 2012. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. Biotechnol. Genet. Eng. Rev. 28: 147–176.
16. SK, Ng. 2012. Generation of high-expressing cells by methotrexate amplification of destabilized dihydrofolate reductase selection marker. Methods Mol. Biol. 801: 161–172.
17. L, Fan., Kadura. I, Krebs. LE, Larson. JL, Bowden. DM, and Frye. CC. 2013. Development of a highly-efficient CHO cell line generation system with engineered SV40E promoter. J. Biotechnol. 168: 652–658.
18. RE, Kingston., Kaufman. RJ, Bebbington. CR, and Rolfe. MR. 2002. Amplification using CHO cell expression vectors. Curr. Protoc. Mol. Biol. Chapter 16.

19. Noh, S.M., S. Shin, and G.M. Lee. 2018. Comprehensive characterization of glutamine synthetase-mediated selection for the establishment of recombinant CHO cells producing monoclonal antibodies. Sci. Rep. 8: 5361.
20. 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.
21. AG, W., and F. P. 2005. Remote control of gene transcription. Hum. Mol. Genet. 14 Spec No 1.
22. Lee, J.S., T.B. Kallehauge, L.E. Pedersen, and H.F. Kildegaard. 2015. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. Sci. Reports 2015 51. 5: 1–11.
23. Martinez-Salas, E., R. Francisco-Velilla, J. Fernandez-Chamorro, and A.M. Embarek. 2018. Insights into Structural and Mechanistic Features of Viral IRES Elements. Front. Microbiol. 8: 2629.
24. 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.
25. Kober, L., C. Zehe, and J. Bode. 2013. Optimized signal peptides for the development of high expressing CHO cell lines. Biotechnol. Bioeng. 110: 1164–1173.
26. Low, K.O., N.M. Mahadi, and R.M. Illias. 2013. Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. Appl. Microbiol. Biotechnol. 97: 3811–3826.
27. Owji, H., N. Nezafat, M. Negahdaripour, A. Hajiebrahimi, and Y. Ghasemi. 2018. A comprehensive review of signal peptides: Structure, roles, and applications. Eur. J. Cell Biol. 97: 422–441.

28. Van Voorst, F., and B. De Kruijff. 2000. Role of lipids in the translocation of proteins across membranes. Biochem. J. 347: 601–612.

29. Caspers, M., U. Brockmeier, C. Degering, T. Eggert, and R. Freudl. 2010. Improvement of Sec-dependent secretion of a heterologous model protein in Bacillus subtilis by saturation mutagenesis of the N-domain of the AmyE signal peptide. Appl. Microbiol. Biotechnol. 86: 1877–1885.

30. Schrempf, S., M. Froeschke, T. Giroglou, D. von Laer, and B. Dobberstein. 2007. Signal Peptide Requirements for Lymphocytic Choriomeningitis Virus Glycoprotein C Maturation and Virus Infectivity. J. Virol. 81: 12515–12524.

31. Mordkovich, N.N., N.A. Okorokova, and V.P. Veiko. 2015. Structural and functional organization of the signal peptide of pro-enterotoxin B from Staphylococcus aureus. Appl. Biochem. Microbiol. 51: 641–648.

32. Zhang, L., Q. Leng, and A.J. Mixson. 2005. Alteration in the IL-2 signal peptide affects secretion of proteins in vitro and in vivo. J. Gene Med. 7: 354–365.

33. Suominen, I., P. Meyer, C. Tilgmann, T. Glumoff, V. Glumoff, J. Kapyla, and P. Mantsala. 1995. Effects of signal peptide mutations on processing of Bacillus stearothermophilus α-amylase in Escherichia coli. Microbiology. 141: 649–654.

34. von Heijne, G. 1984. Analysis of the distribution of charged residues in the N-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells. EMBO J. 3: 2315–2318.

35. Choo, K.H., and S. Ranganathan. 2008. Flanking signal and mature peptide residues influence signal peptide cleavage. In: BMC Bioinformatics. BioMed Central. pp. 1–11.

36. Martoglio, B., R. Graf, and B. Dobberstein. 1997. Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin. EMBO J. 16:

6636–6645.

37. James, P., T. Vorherr, and E. Carafoli. 1995. Calmodulin-binding domains: just two faced or multi-faceted? Trends Biochem. Sci. 20: 38–42.

38. Byun, H., N. Halani, Y. Gou, A.K. Nash, M.M. Lozano, and J.P. Dudley. 2012. Requirements for Mouse Mammary Tumor Virus Rem Signal Peptide Processing and Function. J. Virol. 86: 214–225.

39. Dultz, E., M. Hildenbeutel, B. Martoglio, J. Hochman, B. Dobberstein, and K. Kapp. 2008. The signal peptide of the mouse mammary tumor virus rem protein is released from the endoplasmic reticulum membrane and accumulates in nucleoli. J. Biol. Chem. 283: 9966–9976.

40. Müllner, M., B. Salmons, W.H. Günzburg, and S. Indik. 2008. Identification of the Rem-responsive element of mouse mammary tumor virus. Nucleic Acids Res. 36: 6284–6294.

41. Akhlaq, S., N.G. Panicker, P.S. Philip, L.M. Ali, J.P. Dudley, T.A. Rizvi, and F. Mustafa. 2018. A cis-Acting Element Downstream of the Mouse Mammary Tumor Virus Major Splice Donor Critical for RNA Elongation and Stability. J. Mol. Biol. 430: 4307–4324.

42. Atoom, A.M., N.G.A. Taylor, and R.S. Russell. 2014. The elusive function of the hepatitis C virus p7 protein. Virology. 462–463: 377–387.

43. Oehler, V., A. Filipe, R. Montserret, D. da Costa, G. Brown, F. Penin, and J. McLauchlan. 2012. Structural Analysis of Hepatitis C Virus Core-E1 Signal Peptide and Requirements for Cleavage of the Genotype 3a Signal Sequence by Signal Peptide Peptidase. J. Virol. 86: 7818–7828.

44. McLauchlan, J., M.K. Lemberg, G. Hope, and B. Martoglio. 2002. Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets. EMBO J. 21: 3980–3988.

45. McLauchlan, J., M.K. Lemberg, G. Hope, and B. Martoglio. 2002.

Intramembrane proteolysis promotes trafficking of hepatitis C virus core protein to lipid droplets. EMBO J. 21: 3980.

46. Gawlik, K., and P.A. Gallay. 2014. HCV core protein and virus assembly: what we know without structures. Immunol. Res. 60: 1.

47. Cabantous, S., T.C. Terwilliger, and G.S. Waldo. 2005. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat. Biotechnol. 23: 102–107.

48. Hansen, H.G., C.N. Nilsson, A.M. Lund, S. Kol, L.M. Grav, M. Lundqvist, J. Rockberg, G.M. Lee, M.R. Andersen, and H.F. Kildegaard. 2015. Versatile microscale screening platform for improving recombinant protein productivity in Chinese hamster ovary cells. Sci. Rep. 5: 18016.

49. Kakimoto, Y., S. Tashiro, R. Kojima, Y. Morozumi, T. Endo, and Y. Tamura. 2018. Visualizing multiple inter-organelle contact sites using the organelle-Targeted split-GFP system. Sci. Rep. 8: 1–13.

50. Cabantous, S., and G.S. Waldo. 2006. In vivo and in vitro protein solubility assays using split GFP. Nat. Methods. 3: 845–854.

51. Pedelacq, J.D., and S. Cabantous. 2019. Development and applications of superfolder and split fluorescent protein detection systems in biology. Int. J. Mol. Sci. 20.

52. Romei, M.G., and S.G. Boxer. 2019. Split Green Fluorescent Proteins: Scope, Limitations, and Outlook. Annu. Rev. Biophys. 48: 19–44.

53. Lin, C.-Y., J. Both, K. Do, and S.G. Boxer. 2017. Mechanism and bottlenecks in strand photodissociation of split green fluorescent proteins (GFPs). Proc. Natl. Acad. Sci. U. S. A. 114: E2146.

54. Cabantous, S., T.C. Terwilliger, and G.S. Waldo. 2005. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat. Biotechnol. 23: 102–107.

55. GeneArt^{TM} Project Manager. Online tool for obatining customized DNA constructs.
56. DNA oligo synthesis online tool from PCR primers to GMP | IDT. https://eu.idtdna.com/pages/products/custom-dna-rna/dna-oligos , Date accessed (15/04/2021) .
57. PCR Protocol Phusion® DNA Polymerase | NEB. m/protocols/0001/01/01/pcr-protocol-m0530 Date accessed (15/04/2021).
58. OligoAnalyzer Tool - primer analysis | IDT. https://eu.idtdna.com/pages/tools/oligoanalyzer?returnurl=%2Fcalc%2Fanalyzer, Date accessed (15/04/2021) .
59. Parkinson, N.J., S. Maslau, B. Ferneyhough, G. Zhang, L. Gregory, D. Buck, J. Ragoussis, C.P. Ponting, and M.D. Fischer. 2012. Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. Genome Res. 22: 125–133.
60. Removal of Ethidium Bromide and Calf Intestinal Alkaline Phosphatase Using the Wizard SV Gel and PCR Clean-Up System. .
61. Engler, M.J., and C.. Richardson. 1982. The Enzymes. New York: Academic Press.
62. Lee, J.H., P.M. Skowron, S.M. Rutkowska, S.S. Hong, and S.C. Kim. 1996. Sequential amplification of cloned DNA as tandem multimers using class- IIS restriction enzymes. Genet. Anal. - Biomol. Eng. 13: 139–145.
63. Padgett, K.A., and J.A. Sorge. 1996. Creating seamless junctions independent of restriction sites in PCR cloning. Gene. 168: 31–35.
64. QIAGEN Plasmid Kits. .
65. LSM 980 with Airyscan 2 – Your Confocal Microscope With Multiplex Mode. https://www.bioinformatics.org/sms2/orf_find.html, Date accessed

(15/04/2021)

66. Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74: 5463–5467.
67. Clustal Omega < Multiple Sequence Alignment < EMBL-EBI. Clustal Omega is a multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences, https://www.ebi.ac.uk/Tools/msa/clustalo/., Date accessed (15/04/2021)
68. ORF Finder. Open reading frame finder, tool that searches for openn reading frames in the sequence entered https://www.bioinformatics.org/sms2/orf_find.html, Date accessed (15/04/2021) .
69. 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.
70. Peptide and Protein Molecular Weight Calculator | AAT Bioquest. .
71. Khan, K.H. 2013. Gene Expression in Mammalian Cells and its Applications. Adv. Pharm. Bull. 3: 257.
72. Wu, J.Z. 2001. Internally Located Signal Peptides Direct Hepatitis C Virus Polyprotein Processing in the ER Membrane. IUBMB Life (International Union Biochem. Mol. Biol. Life). 51: 19–23.
73. 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.
74. Kim, Y.G., B. Park, J.O. Ahn, J.K. Jung, H.W. Lee, and E.G. Lee. 2012. New cell line development for antibody-producing Chinese hamster ovary cells

using split green fluorescent protein. BMC Biotechnol. 12: 24.

75. Sleiman, R.J., P.P. Gray, M.N. McCall, J. Codamo, and N.A.S. Sunstrom. 2008. Accelerated cell line development using two-color fluorescence activated cell sorting to select highly expressing antibody-producing clones. Biotechnol. Bioeng. 99: 578–587.

76. Foglieni, C., S. Papin, A. Salvadè, T. Afroz, S. Pinton, G. Pedrioli, G. Ulrich, M. Polymenidou, and P. Paganetti. 2017. Split GFP technologies to structurally characterize and quantify functional biomolecular interactions of FTD-related proteins. Sci. Rep. 7: 1–15.

77. Kapp, K., S. Schrempf, M.K. Lemberg, and B. Dobberstein. 2009. Protein Transport into the Endoplasmic Reticulum. 1st Editio. Boca Raton: Landes Bioscience.

78. Lundqvist, M., N. Thalén, A.L. Volk, H.G. Hansen, E. von Otter, P.Å. Nygren, M. Uhlen, and J. Rockberg. 2019. Chromophore pre-maturation for improved speed and sensitivity of split-GFP monitoring of protein secretion. Sci. Rep. 9: 1–8.

79. YS, Levray., Berhe. AD, and Osborne. AR. 2020. Use of split-dihydrofolate reductase for the detection of protein-protein interactions and simultaneous selection of multiple plasmids in Plasmodium falciparum. Mol. Biochem. Parasitol. 238.

80. Joo, H.K., Y.W. Park, Y.Y. Jang, and J.Y. Lee. 2018. Structural Analysis of Glutamine Synthetase from Helicobacter pylori. Sci. Rep. 8: 11657.

81. Llorca, O., M. Betti, J.M. González, A. Valencia, A.J. Márquez, and J.M. Valpuesta. 2006. The three-dimensional structure of an eukaryotic glutamine synthetase: Functional implications of its oligomeric structure. J. Struct. Biol. 156: 469–479.

6.0 Appendix

Images settings for the Zeiss LSM 880 Elyra AxioObserver confocal microscope

Image Dimensions	
Channels	4
Tiles	9 Tiles
Scaling (per Pixel)	0.05 µm x 0.05 µm (Edit)
Image Size (Pixels)	4288 x 4288
Image Size (Scaled)	209.85 µm x 209.85 µm
Bit Depth	16 Bit
Image Center Position	X: 3.60 mm, Y: -9.02 mm
ROI Center Offset	X: 0.00 µm, Y: 0.00 µm

Microscope	LSM 880 Elyra, AxioObserver
Objective	Plan-Apochromat 63x/1.4 Oil DIC M27
Filters	-2147483648 - -2147483648
Beam Splitter	eGFP: MBS : MBS 488/561/633 MBS_InVis : MBS -405 DBS1 : Mirror TRITC: MBS : MBS 488/561/633 MBS_InVis : MBS -405
Lasers	488 nm : 0.5 %
Calibration Marker Positions	Marker 1: 0.00 µm, 0.00 µm Marker 2: 0.00 µm, 0.00 µm Marker 3: 0.00 µm, 0.00 µm

	Track 1	Track 2
Contrast Method	Fluorescence	Fluorescence
Pinhole	6.10 AU	3.94 AU
Laser Wavelength	488 nm: 0.5 %	561 nm: 0.5 %
Scan Mode	Frame	Frame
Scan Zoom	1.8	1.8
Rotation	0°	0°
Pixel Time	1.37 µs	1.37 µs
Line Time	30.00 µs	30.00 µs
Frame Time	15.04 s	15.04 s
Scan Direction	Unidirectional	Unidirectional
Line Step		
Averaging	2	2

	Channel 1	Channel 2	Channel 3	Channel 4
Channel Name	Ch2 GaAsP-T2	T PMT-T2	ChA-T3	
Channel Description				
Channel Color				
Excitation Wavelength	488		561	
Emission Wavelength	510		579	
Effective NA	1.4	1.4	1.4	1.4
Detection Wavelength	496-524			
Depth of Focus		0.85 µm		
Binning Mode	1x1	1x1	1x1	
Detector Type	PMT		Airyscan	
Detector Gain	600.0	250.0	791.0	
Detector Offset	-100	-60	0	
Detector Digital Gain	1.0	1.0	1.0	
Airyscan Mode	Off	Off	Airyscan SR	