Preliminary Analysis of Vital Proteins of the Human Respiratory Syncytial Virus

BACHE, HELEN,CLARE

Preliminary Analysis of Vital Proteins of the Human Respiratory Syncytial Virus

Author: Helen Bache

Thesis for Postgraduate Master of Science by Research

Bioactive Chemistry Centre

Chemistry Department

Supervisors: Dr Ehmke Pohl and Dr Robert Paul Yeo

2010

This project received funding from One North East

Acknowledgements

I would like to thank my supervisors, Dr Ehmke Pohl and Dr Robert Yeo for all of their guidance during this project and in writing this thesis. The project would not have been possible without the funding received from One North East. Thanks go to the students and staff of the Bioactive Chemistry Centre for the help and advice I received from them. I am very grateful to everyone at Onyx Scientific who made my placement there enjoyable and useful, especially Dr Julian Northen and Dr Michael Hallett. Finally, I must thank my family, friends and boyfriend for their support and advice.

Abstract

Human Respiratory Syncytial Virus infects the vast majority of children under the age of two and reoccurs in adulthood. It is a serious global problem as severe infection and death can result in the very young or very old and in immunocompromised people. The related bovine Respiratory Syncytial Virus is a serious problem for the farming community. To date, there is no effective treatment for Respiratory Syncytial Virus infection, with drugs only available to reduce symptoms in the most severe cases. Human Respiratory Syncytial Virus produces 10 proteins carrying out a diverse array of roles for the infectivity of the virus. Despite this there is still a poor understanding of the three dimensional structure. Only two proteins have been resolved to atomic resolution, the Nucleocapsid protein and the Matrix protein. The objective of this thesis is to initiate work on a further two proteins, with the long term aim of crystallisation studies that may lead to high resolution structures. The Fusion transmembrane glycoprotein is responsible for the fusion of the virion membrane to the host cell membrane for viral entry and for the fusion of an infected cell membrane to membranes of healthy cells, spreading infectivity. Without the Fusion protein there is no infectivity and it is therefore a therapeutic target. Structural information such as a high resolution X-ray crystal structure is required to facilitate the design of inhibitors that could be future therapeutic agents. The second protein studied here is the M2-1 protein, which is a transcriptional elongation factor, involved in replication of the viral genome. Deletion of M2-1 has adverse affects on the replication of the virus. Therefore M2-1 could also make a strong target for drug design. This thesis explores what is currently known about each of these vital proteins, the advantages in inhibiting their respective activities and the possible routes to effective inhibitor design. Procedures for expression and purification of the M2-1 protein are detailed here, as are preliminary experiments into the cloning and expression of the Fusion protein in insect cells. Immunofluorescence experiments into the localisation of the Fusion protein within mammalian cells and the possible interaction of the Fusion protein with the Matrix protein are thoroughly described.

Table of Contents

List of Tables

List of Figures

List of Schemes

List of Abbreviations

Aa	amino acids
APS	ammonium persulphate
Asn	asparagine
BDV	borna disease virus
Boc	butyloxycarbonyl
BOP	benzotriazolyloxytris-(dimethylamino)phosphonium hexafluorophosphate
Bp	base pairs
bRSV	bovine respiratory syncytial virus
Bzl	benzyl
CT	cytoplasmic tail
DAPI	4',6-diamidino-2-phenylindole
DCC/DCCI	dicyclohexylcarbodiimide
DCM	dichloromethane
DKP	diketopiperazine
DMAP	4-dimethylaminopyridine
DMEM	Dubeccio's modified eagle's media, Gibco
DMF	dimethylformamide
DMJ	deoxymannojirimycin
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
EDCI	3-(ethyliminomethylideneamino)-N,N-dimethylpropan-1-amine hydrochloride
FCS	foetal calf serum
FITC	fluorescein isothiocyanate
F protein	fusion protein
Fmoc	9-fluroenylmethoxycarbonyl
FSol	soluble construct of the fusion protein
GST	glutathione S-transferase
HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3,-tetramethyluronium hexafluorophosphate
HPLC	high performance liquid chromatography
HPIV	human parainfluenza virus
HR	heptad repeat
hRSV	human respiratory syncytial virus
L protein	large protein
LC	liquid chromatography
LC-MS	liquid chromatography mass spectrometry

Leu	leucine
M protein	matrix protein
Met	methionine
MeV	measles virus
m.o.i	multiplicity of infection = the ratio of infectious virus particles to cells
mRNA	messenger ribonucleic acid
MuV	mumps virus
N	nucleocapsid protein
NDV	newcastle disease virus
NMR	nuclear magnetic resonance
NS protein	non-structural protein
OBzl	benzyl ester
ORF	open reading frame
P	phosphoprotein
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
Phe	phenylalanine
ppm	parts per million
PTFE	polytetrafluoroethylene
PIV5	parainfluenza virus 5 = SV5
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
RABV	rabies virus
RNA	ribonucleic acid
SDS	sodium dodecyl sulfate
Ser	serine
SF21	Spodoptera frugiperda 21
SH	small hydrophobic
SP	signal peptide
SPPS	solid phase peptide synthesis
SV5	simian virus 5 = PIV5
^{t}Bu	tert-butyl
TEMED	N,N,N,N -tetramethyl-ethylenediamine
TFA	trifluroacetic acid
THF	tetrahydrofuran
Thr	threonine
TLC	thin layer chromatography
TMD	transmembrane domain
Trp	tryptophan
Tyr	tyrosine
UV-vis	ultra violet visible light
VSV	vesicular stomatis virus
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
Z	benzyloxycarbonyl
ZnF	Zinc finger

Chapter One: Introduction

1.1 Introduction to Respiratory Syncytial Virus

Respiratory Syncytial Virus, RSV, affects many different species of vertebrates including humans (hRSV) and cattle (bovine RSV). By the age of two years of age most human infants are seropositive for RSV. Recurrent infection is common in later life although usually produces less severe infections. RSV is a leading cause of upper respiratory tract infections, with bronchiolitis and severe pneumonia in cases where infection spreads to the lower respiratory tract.[1] There are an estimated 64 million cases a year globally, but most healthy adults when infected only suffer mild infections with symptoms similar in severity to a cold. However in very young infants (< 6 months of age), the elderly and those who are immunocompromised, the virus can cause severe disease and 160,000 deaths a year world-wide are attributed to hRSV.[2] An extreme case is in those needing a bone marrow transplant- if infected by hRSV the mortality rate increases to 90 %.[2] Importantly bovine RSV results in large agricultural economic losses.[3] Therefore research into drugs or vaccines against both bRSV and hRSV is highly important. This includes research ongoing into the viral structure and in particular knowledge of the individual proteins hRSV produces. The various proteins of hRSV enable effective infection, and they also affect and interact with each other to facilitate replication. It is therefore pertinent that structural information on hRSV proteins is developed to enable an effective treatment or vaccine to be developed. Currently the only treatment available to control the symptoms is the drug Ribavirin, which only controls symptoms and is given only in the most extreme cases when other options have been exhausted. This is because the drugs developed so far are simply not effective enough, or there is little evidence to their efficacy other than control of symptoms.

There is a huge range of potential patients who could benefit from effective treatment directed against RSV if it were available, stretching from preventing the infection in at risk groups via vaccine, to stopping progression in already infected patients and improving the outcome of a severe infection. However, the lack of structural knowledge of most of the proteins of hRSV is potentially limiting the progress in drug development. With knowledge of the 3D structure, preferably from crystallographic data for full length proteins, potential drug binding sites can be identified and the mode of action of the

protein's functionality could be made clearer, leading to unambiguous information on drug mechanism of action.

1.2 RSV- A Member of the Paramyxoviridae

RSV is a member of the order *Mononegavirales*, family Paramyxoviridae, subfamily *Pneumovirinae* and genus *Pneumovirus*. It is therefore related to Newcastle Disease Virus, Measles, Human Parainfluenza Viruses and the increasingly threatening Nipah and Hendra viruses.[4]

1.2.1 Structure of hRSV

hRSV has a negative sense, single stranded non-segmented genome i.e. one RNA molecule makes up the viral genome of 15,222 nucleotides.[5, 6] All four taxonomic families in the order *Mononegavirales* have a similar arrangement but varying numbers of genes and sizes of genome, shown in Figure 1.1 below.[2]

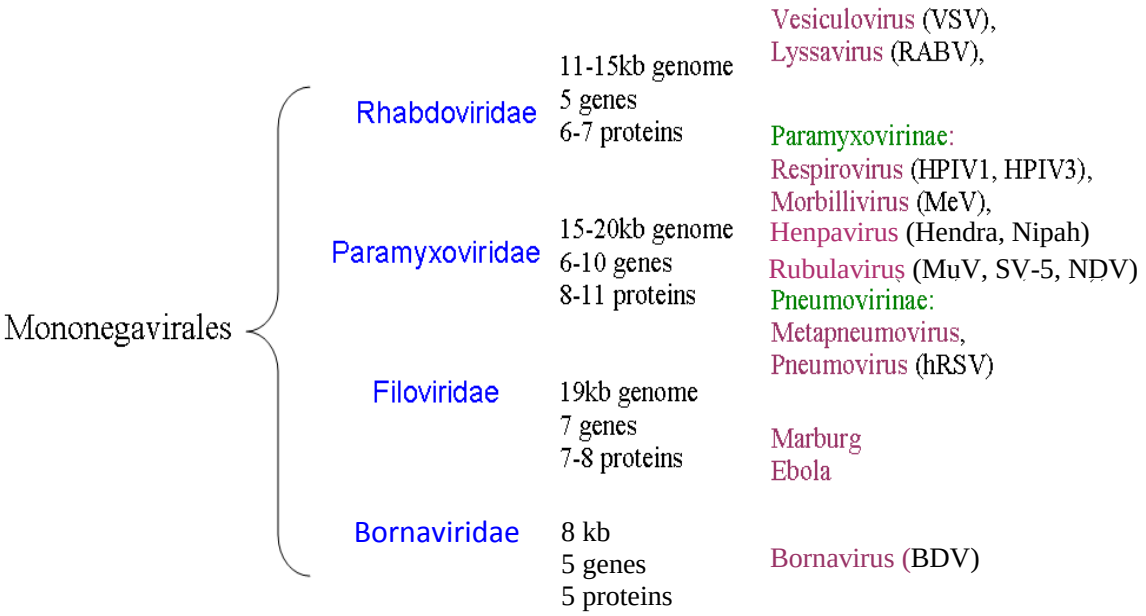


Figure 1.1: Diagram of the Order *Mononegavirales* including families (blue), approximate genome size ranges for each family, genera (purple) with subfamilies (green) and appropriate examples.

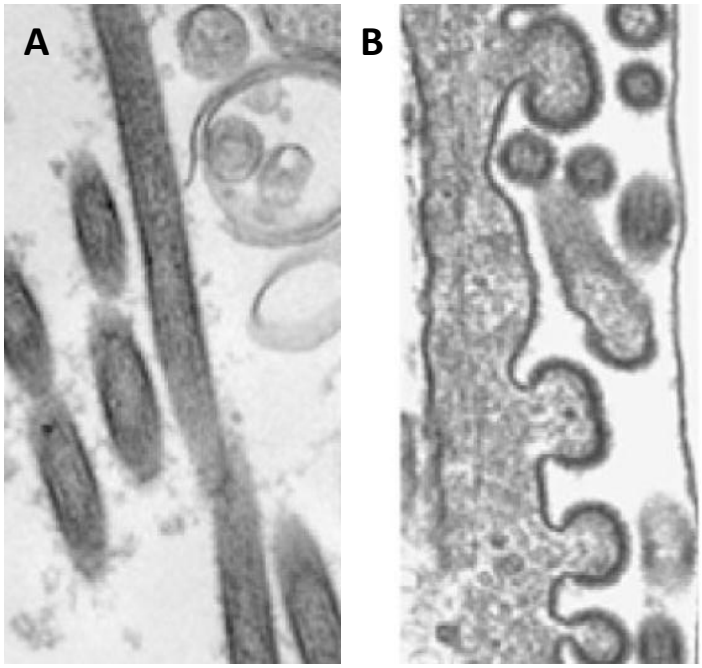


Figure 1.2: A) Viral filaments B) Spherical viral particles. Immunoelectron microscopy of the F protein in Vero cells infected with the Long strain of hRSV for 48 hours. The glycoproteins can be seen spiking out from the virions' surfaces. Adapted from T.L. Gower et al. 2005 Figure 6[7]

hRSV is an enveloped virus and contains a helical nucleocapsid. The virion, the extracellular virus particle, can be imaged using an electron microscope, showing both spherical and filamentous morphologies.[4]

1.2.1.1 The hRSV virion

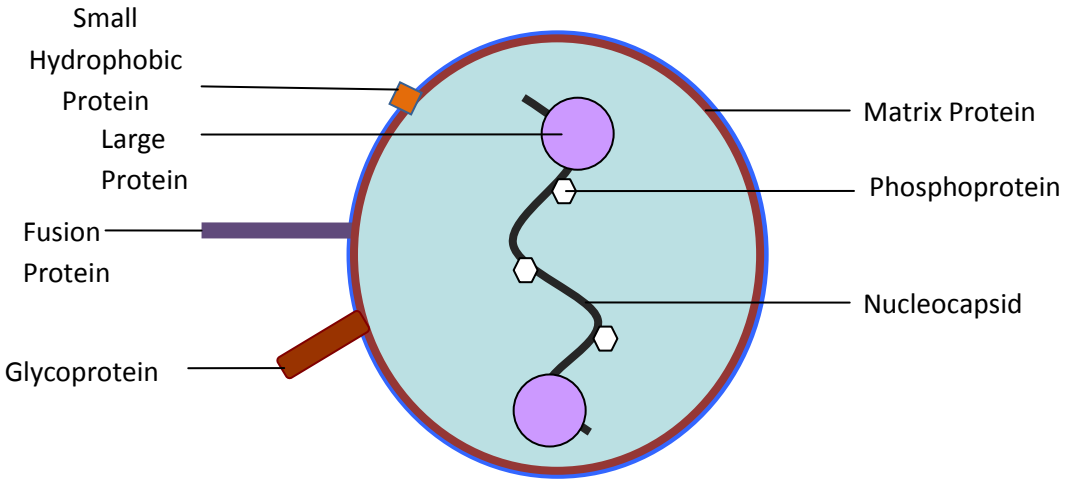


Figure 1.3: Diagrammatic representation of the virion of hRSV including locations of the viral proteins. The virion plasma membrane is shown in blue.

The basic structure of the virion, shown schematically in Figure 1.3, follows the basic structure of the order *Mononegavirales;* there are variation in sizes, shapes and individual components between virions from the families.[8] The RSV virion membrane is taken from the host cell and contains the viral glycoproteins, F, G and SH. An important feature in the diagram is the proteinaceous layer of M protein underneath the viral membrane. The presence of M is vital to viral assembly and morphology, and this layer also provides protection for the virion.[8] *Paramyxoviridae* virions are pleomorphic, occurring as spheres or filaments, but other members of the *Mononegavirales* are different, such as the bullet shape of the virions of the *Rhabdoviridae* or the filament virions of the *Filoviridae*.[8] The nucleocapsid consists of N protein encapsulating the viral genome and the holo-nucleocapsid is formed when L, P, M2-1 and M2-2 proteins associate with the nucleocapsid.

For the budding process from the host cell membrane the Matrix protein, the glycoproteins and the nucleocapsid are required but the individual interactions are not yet completely clear and are largely beyond the scope of this project. The interaction of F and M will be discussed however. It is worth noting, that F requires M in order to relocate from the basolateral membrane to the apical membrane and likewise M appears to need F in order to be stabilised in structures in the membrane such as lipid rafts.[9] The crux of studies into M's function in RSV virion assembly is that presence of M is an essential condition for viral assembly and the layer of M beneath the membrane forms the spherical or filamentous shape of the virion.[9]

1.2.1.2 The genome of hRSV

The viral replication process is similar for all (-) strand RNA viruses. The RNA polymerase (L and P proteins in the nucleocapsid) synthesises multiple mRNAs from the individual cistrons, commonly referred to as genes, of the non-segmented genome. The transcriptase recognises specific start and stop signals that flank the individual cistrons in order to create individual mRNAs that are identical to "normal cellular" mRNAs. The proteins are then translated from the mRNAs; some RNAs are translated to make more than one gene product by alternate translational mechanisms.[2] Glycoproteins are then processed in the ER and Golgi and transported to the plasma membrane. The genome's replication occurs by the synthesis of a complementary genome, the antigenome, to act as a template for synthesis of the genome. The genome and antigenome, unlike the viral mRNAs, are encapsidated in the nucleocapsid comprised mainly of N protein. Naked genome/antigenome RNA is non-functional even in the presence of the viral polymerase. Therefore the replication

process requires viral protein synthesis as it is dependent on the supply of N protein to form the nucleocapsid. When enough viral proteins have formed, the production moves from mRNA production to antigenome formation to produce templates for genomic (-) RNA synthesis.[4] The one difference between the replicase and the transcriptase is that the replicase does not contain the M2-1 protein. Other than this the proteins involved in formation of antigenome and the proteins involved in formation of mRNA are identical.[10] The viral genome enters into the host cell via a pore created by fusion proteins protruding from the virion surface, a vital process to infectivity and one which will be explored in detail in this introduction.

hRSV has important differences to many of the members of its own family. It encodes more genes than any other member of the *Paramyxoviridae* and also produces more proteins, see Figure 1.4 for the layout of the genome. However, most of these proteins have equivalents in other family members.

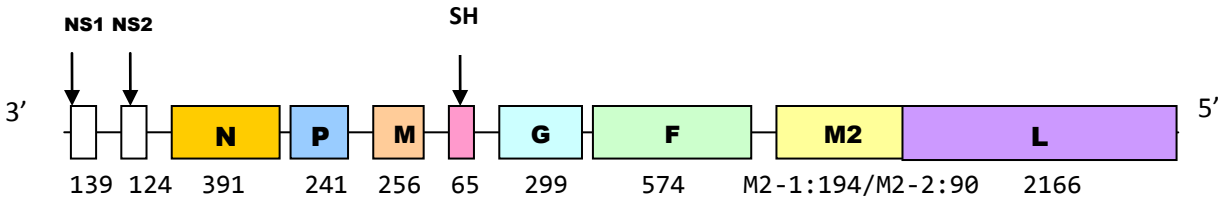


Figure 1.4: The genome organisation of hRSV, 3' to 5' due to the minus-strand RNA. The number of amino acids in each gene segment is shown underneath.

1.2.2 Proteins of hRSV

The 11 proteins produced by hRSV include three surface glycoproteins that project out from the membrane as spikes. The NS (non-structural) proteins have a role in suppression of the immune response.[2] The G protein (glycoprotein) is a transmembrane protein responsible for attachment to the host cell.[12] It has no sequence or structural homology to other *Paramyxoviridae* members' attachment proteins, such as the H (Haemagglutinin) from Measles Virus or HN (Haemagglutinin/Neuramidase) from Mumps Virus. A second transmembrane glycoprotein of the virus is the SH (Small Hydrophobic) protein - its role in viral replication is unknown. It is not needed for attachment or fusion and so is unimportant for infectivity *in vitro.*[13] *In vivo*, however, the SH is important in pathogenicity as it down regulates apoptosis in infected cells.[14] The third surface glycoprotein is the F (Fusion) protein which is in contrast vital for infectivity as it mediates fusion of the viral and cellular membrane to allow entry of the nucleocapsid. hRSV also codes for a few

transcription factor proteins (M2-1 and M2-2). N, the nucleocapsid protein, encapsulates the viral genome interacting with L (Large Protein that carries out the enzymatic activity) indirectly via P (the Phosphoprotein needed in RNA synthesis).[4, 15, 16] Together the L and P form the RNA-dependent RNA polymerase (RdRp). The M (Matrix) protein has been a focus of many recent studies showing that it forms a proteinous layer underneath, and in contact, with the viral membrane. M also has a role in inhibiting host cell transcription and viral RNA synthesis, before packaging occurs for virion formation, by associating with the nucleocapsid.[4]

Protein	**Location in the Virion**	**Function**
NS1 and NS2, Non-Structural 1 and 2	Host cell cytoplasm-not packaged in virion but synthesised in infected cells	Suppresses innate immune response
N, Nucleocapsid	Nucleocapsid	Encapsulates viral genome
P, Phosphoprotein	Nucleocapsid	Part of the polymerase
M, Matrix	Layer on inside of membrane	Involved in virion assembly and has a protection function
SH, Small Hydrophobic	Membrane	Unknown
G, Glycoprotein	Membrane	Attachment to host cell
F, Fusion	Membrane	Fusion of virion and host membrane and fusion of infected cells' membranes (syncytia)
M2-1	Nucleocapsid and polymerase complex (transcriptase)	Anti-termination factor
M2-2	Nucleocapsid and polymerase complex	Transcription factor
L, Large	RNA Polymerase complex	Enzymatic activity of the polymerase

Table 1.1: Proteins of RSV

1.2.2.1 Summary of Current Structural Knowledge

The atomic structure of the Matrix protein is relatively well understood compared to other proteins of RSV. The crystal structure of M was determined in 2009 to a

resolution of 1.6 Å by Money *et al*.[8] The crystal structure of the Nucleocapsid protein was determined in 2009 to a resolution of 3.3 Å by Tawar *et al*.[11]

In terms of potential drug targets, the F protein and the M2-1 protein would be the most effective. Inhibiting F would prevent viral entry, whilst inhibiting M2-1 could inhibit viral replication. However, these are two of the proteins for which there is very little structural information currently, as discussed below.

1.3 The Role of F: Fusion

The most useful protein for drug targeting is the third transmembrane glycoprotein- the F (Fusion) protein. The Fusion Protein is responsible for fusion of viral and host cell membranes. Without this step, viral entry into a host cell cannot occur. Therefore if an effective inhibitor to F can be found, and successfully developed into a drug, the virus could be stopped in its tracks. This has been a successful strategy with the Human Immunodeficiency Virus fusion protein gp41, as there is a fusion inhibitor drug (Enfuvirtide) used in antiviral treatment. Enfuvirtide consists of a peptide, T20 based on the C-terminal heptad repeat of gp41, which inhibits fusion.[17] Second and third generation fusion inhibitors based on T20, such as the more potent inhibitor T-1249, are now being developed to overcome problems of resistance as well as to increase potency. The development of Enfuvirtide and its analogues has demonstrated that fusion protein inhibition can be successful and lead to anti-viral drug development.[17, 18, 19]

The role of the F protein is to bring host and viral membranes into close proximity and induce fusion. Another function of F is to fuse an infected cell with neighbouring non-infected cells. This produces multinucleated structures of giant cells called syncytia, enabling the spread of infection without the production of extracellular virions.[20]

Out of all the proteins that hRSV codes for, the F is the one that is absolutely essential for viral entry into a host cell. Its function is not dependant on any other of the proteins, though it was previously thought that hRSV F would need activation by the attachment protein G.[5, 21] In other members of the *Paramyxoviridae* the attachment protein, H or HN, is needed for binding to cell receptors and in activating F.[22] However, this was found not to be the case for RSV. Studies on hRSV mutants that did not produce G or SH proteins, but did produce F, resulted in productive infections.[21, 23] But without the F, viral infection could not occur even in the presence of the other viral proteins.[23] However fusion was enhanced to maximum efficiency only when all three of the transmembrane glycoproteins were present.[12] It is also the F protein, along with the G

protein, that stimulates an immune response resulting in the symptoms of harbouring the virus.[24]

F, a type-1 membrane protein shown schematically in Figure 1.5, is anchored to the membrane at its C-terminal end,[25] has a mass of 67 kDa (574 amino acids) and is synthesised as an inactive precursor, F_0.[25] At the C-terminus, there is a cytoplasmic tail (discussed in section 1.3.3). In the mature, membrane-associated virus, the CT is located in the interior of the virion, and it is adjacent to the transmembrane domain (TMD) that spans the membrane. For the protein to mature, F_0 is cleaved into shorter polypeptides at two target motifs comprised of basic amino acid residues, The two cleavage products F_1 (50 kDa) and F_2 (10-22 kDa), are then linked by a series of disulphide bridges, whereupon the protein becomes functional due to the exposure of a hydrophobic peptide sequence, the Fusion Peptide, at the N-terminal of F_1.[25] The F_2 subunit has been observed to be 22 kDa, 16 kDa or 10 kDa (the latter only present in Vero cells), due to observed differences in N-linked glycosylation, discussed in section 1.3.4.[26] The two stage cleavage event, at amino acid residues 106-109 (site I) and 131-136 (site 2), in which a short intervening peptide is excised from F, is unusual as the norm in the *Paramyxoviridae* is for a single event to expose the fusion peptide.[27] The cleavage events in RSV, as for other paramyxoviridae F proteins, are performed by a furin-like protease in the trans-Golgi apparatus.[25] The cleavage occurs late in the transport to the membrane to prevent premature fusion of membrane components of the cell from killing the cell.

The peptide that is excised during the processing of F_1 and F_2 (residues 110-136, referred to as Pep27) is thought to have a role in stimulating the immune response to F and therefore hRSV.[22, 28]

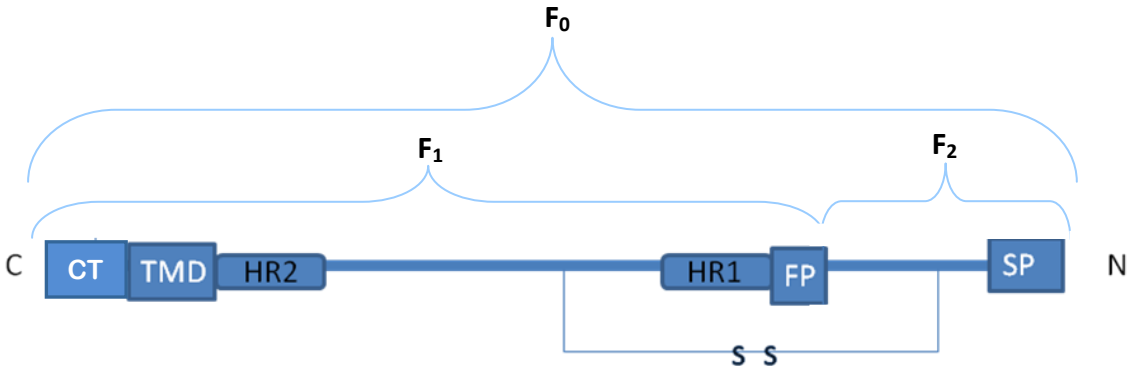


Figure 1.5: Diagrammatic representation of the F Protein, including the cytoplasmic tail (CT), the transmembrane domain (TMD), the two heptad repeat regions (HR1 and HR2), the fusion peptide (FP) and the signal peptide (SP).

As aforementioned, F_1 has a C terminal hydrophobic transmembrane domain (TMD) which is embedded in the viral membrane, anchoring the protein.[12] The rest of the protein, without the TMD or CT, makes up the extracellular domain.[21] A second hydrophobic region is the glycine-rich, 20 amino acid long, Fusion Peptide at the N-terminus of F_1, which fuses to the host membrane during viral attachment.[1, 12] The two heptad repeats (HR1 and HR2), present in all paramyxoviridae fusion proteins, are important in the process of maintaining F's conformation in the process of fusion (see section 1.3.6.1). F_2 holds a third heptad repeat from residues 53 to 100, although evidence so far suggests this is unimportant for functional integrity outside of the correct folding of the protein as a whole.[29] F_2 carries the Signal Peptide (SP) at the N-terminus, which controls translocation of the protein into the ER, where it is removed by signal peptidase.[26]

1.3.1 Disulphide Linkage of F_1 and F_2

In active form, post-cleavage, the RSV F protein has fifteen cysteine residues,[22] which is more than any other *Paramyxoviridae* member, for example, the closely related Measles virus fusion protein has 10 cysteine residues.[24, 31] Human Parainfluenza Virus-3 F has been used to model the molecular structure of hRSV F for study into their conserved cysteine residues. These are also conserved within *Paramyxovirinae*, such as NDV (which has 12 cysteines conserved between strains) and Sendai virus and other viruses of the *Pneumovirinae* such as Pneumonia Virus of Mice (PVM) and Avian Pneumovirus (APV).[25] It has been shown by N.D. Day et al. in 2006, that by mutating these residues individually to serine, that ten are crucial for proper folding and translocation of F to the cell surface.[25] The expression was reduced to 7 % of normal levels and the fusion activity was also greatly reduced or did not occur at all.[25] Three of the residues when mutated to serine did not alter folding or function. This was notable due to the conservation of these cysteines in different *Paramyxoviridae* viruses, which would indicate they are necessary to function. Two cysteine residues within F_2, residues 37 and 69, were shown to be extremely important in correct folding of F. When these were separately altered to serine the resulting fusion activity was negligible.[25] Cysteines 37 and 69 are therefore vital to folding of the protein. Cys439 and Cys37 are interestingly only conserved amongst pneumoviruses and not with F proteins from the other *Paramyxoviridae* subfamily the *Paramyxovirinae*. It is probable that the residues that form the disulphide linkages between F_1 and F_2 are Cys37 to Cys439, and Cys69 to Cys212.[25] This conclusion has come from the comparison of the mapping of disulphide bridges in Sendai Virus and the conservation of these cysteines within other pneumovirus fusion proteins.[25] If so, then the second disulphide bridge

(C69-C212) would be unique to the *Pneumovirinae* as two residues are not conserved amongst all *Paramyxoviridae* fusion proteins. Therefore the cysteines of the F protein are crucial to folding and consequently to function, and their inactivation can clearly significantly reduce, or even completely eliminate, fusion activity and so infection.

1.3.2 The Transmembrane Domain

The transmembrane domain of F has been found to have an apical targeting function that is independent of the protein's orientation with a role in directing the protein to the membrane after maturation. This role could extent to directing the virion for budding from the apical membrane.[5] However, it was shown that none of the glycoproteins that hRSV produces (G, F or SH) are needed for apical membrane targeting, although they may make a small contribution to it, nor are they needed for the virus's release from that membrane.[31] N- and O-linked glycans can also act as apical sorting signals and sites for these occur on the F and G proteins, with the apical sorting signal on the TMD of F being an independent one.[31] There was no effect on the targeting of the virus to the apical membrane when each glycoprotein was deleted in turn, and no effect either upon the release of the virus from the host cell.[31]

1.3.3 Importance of the Cytoplasmic Domain

The cytoplasmic tails (CT in Figure 1.5) of many of the fusion proteins from other paramyxoviruses have a role in fusion. Related to their function are their motifs for palmitoylation.[32] These are motifs where fatty acids attach to cysteine residues to make the protein more hydrophobi, thereby stabilising the membrane/protein interaction. This aids processing and translocation depending upon the nature of the protein in question- an example could be the F protein's transport to and insertion into the viral membrane.[33] Palmitoylation sites are sometimes also needed in viral assembly.[32, 33] The C-terminus of hRSV F, proximal to the transmembrane domain has 26 amino acids, including a cysteine which is palmitoylated but this cysteine has not been shown to be essential for processing or activity.[25, 34]

A separate study, by Oomens *et al.* (2006) illustrated that the entire cytoplasmic tail does have an important role in cellular localisation of the fusion protein. It was found that deleting the cytoplasmic tail reduced the number of infectious virions produced. It is possible that without the C-terminal domain, fusion occurs prematurely. In the absence of the CT the protein had a very even distribution on the cell membrane, whereas when the CT was present the F protein was localised on filamentous structures 2-10μm in length.[5] hRSV was modified to express the VSV (*Vesicular*

Stomatis Virus, a member of the *Rhabdoviridae* of the *Mononegavirales*) G protein with its CT replaced with that of RSV F. Compared to a recombinant virus with unmodified VSV G protein, the virus with a F CT domain produced more progeny and more/larger syncytia formed.[5] This demonstrated that the activity of the C terminus was independent of the F protein's ectodomain. The cytoplasmic tail also mediated the interaction of the protein with membrane lipids and this was important for viral assembly. This showed that the cytoplasmic tail is necessary for proper protein function, not for fusion itself, but to give the correct distribution at the cellular membrane and so maximum viral progeny.

1.3.4 Glycosylation sites

F has a number of N-linked glycosylation sites. This distinguishes it from the G protein that has high levels of O-linked glycosylation.[35] Therefore the effect of glycosylation on protein function can be studied separately for F and G. In unprocessed F_0, there are potential N-linked glycosylation motifs at Asn116, Asn120 and Asn126, close to the furin recognition sequences at the cleavage sites. In other fusion proteins such sites aid cleavage by furin-like proteases. However, in mature hRSV F these sites are not glycosylated, suggesting they do not have a role in processing of F.[35]

In the mature protein, F_1 has one N-linked glycosylation site whereas F_2 has two, which could be important for correct protein folding.[25] N-linked glycans, when formed at the glycosylation site in the ER, bond to nitrogen atoms on asparagine residues in either Asn-X-Ser or Asn-X-Thr (were X is any amino acid) motifs. These glycans attach to mannose cores which are then removed in the Golgi apparatus, being replaced with more complex glycans such as fucose (this is glycan maturation). F_1's sole glycosylation site is Asn500 and the glycosylation sites of F_2 are at Asn27 and Asn70. It was shown, by McDonald *et al.* (2006) that Asn500 has an important role in fusion mediation and that Asn27 and Asn70 could have a minor role in F_2's attachment ability. The N-linked glycosylation was not required for cleavage of the F protein, nor was it required for cell surface expression or for generation of progeny.[35] When Asn27, Asn70 and Asn500 were deleted in all different combinations, there was no reduction in cell surface expression or proteolysis of F.[36] They do, however, affect fusion in different ways. A mutation, Asn70Gln, increased fusion activity by 50%, but the same treatment for Asn27 had no affect on fusion. When glycosylation on Asn500 was inhibited, fewer syncytia formed as fusion activity was reduced 100-fold. Other members of the *Paramyxoviridae* have been shown to be dependant upon such glycosylation for infectivity. N-glycosylation and protein processing are closely linked in most of these

but hRSV F is an exception due to the lack of change in cell surface expression when the glycosylation sites are mutated.[36] The reason for the importance of Asn500 is due to its location in HR2, and it could be that the local structure in this region needs the glycan maturation in order to form properly.[35] It is also the case that bRSV and hRSV have different glycosylation needs as the equivalent bRSV Asn500 when mutated to alanine prevents the protein's expression. Compounds inhibiting glycan maturation could be useful in inhibiting the action of F as drug molecules, at least reducing the level and spread of infection.[35] The compound used in this study was DMJ, an inhibitor with broad effect. A more specific compound would be needed to be used as a first lead compound for design of effective drugs.[35]

hRSV F also has three other potential N-linked glycosylation sites (Asn120, Asn126 and Asn116) that differ from all other *Mononegavirales* fusion protein sequences. These are located adjacent to the cleavage site next to fusion peptide, in the Pep27 region, possibly to aid targeting of this cleavage site to form the mature peptide .[12] However, N-glycans have not been detected on this residues and when mutating them does not affect the glycosylation status of the protein.[36]

1.3.5 Current knowledge of 3D structure

Integral to the structure of F are the hydrophobic heptad repeat regions situated at either end of the F_1 subunit. This means that each F_1 subunit within the F trimer has two repeat elements distal to each other on the extracellular domain of F_1. A hydrophobic heptad repeat is a sequence of residues *abcdefg* where the amino acids at positions a and d are hydrophobic, see Figure 1.6. These are referred to as HR1 (or HRN) and HR2 (or HRC), as in Figure 1.5.[1] This primary sequence tends to fold into an α-helix. [37]

474 **I**I**N**F**Y**D**P**L**V**F**P**S**D**E**F**D**A**S**I**S**Q**V**N**E**K**I**N**Q**S**L**A**F**I**R**K**S**D**E**L**L**H**N**V**N**A**V**K**S**T**T** 523

Figure 1.6: The sequence of Heptad Repeat 2. Positions a and d are in green. Note residues in these positions: Isoleucine, valine, leucine, phenylalanine and proline are all very hydrophobic and take up the majority of the a and d positions. Serine and threonine are not hydrophobic but are also uncharged. Lysine is hydrophilic however only appears once.

Due to their hydrophobic nature, when two of these structures interact, their α-helical coils wind around each other creating a coiled-coil structure. The crystal structure of a complex of RSV F's two heptad repeat regions has been solved to a resolution of 2.3 Å and is shown schematically below in Figure 1.7.[1] The interaction of the two heptad repeats is integral to the function of F. Both HR1 and HR2 form trimers of their α-

helices, and therefore both form coiled-coils. The α-helical trimer of HR2 resides as a coiled-coil inside HR1 peptides which are arranged in an anti-parallel manner in the hydrophobic grooves of HR2.[1, 12]

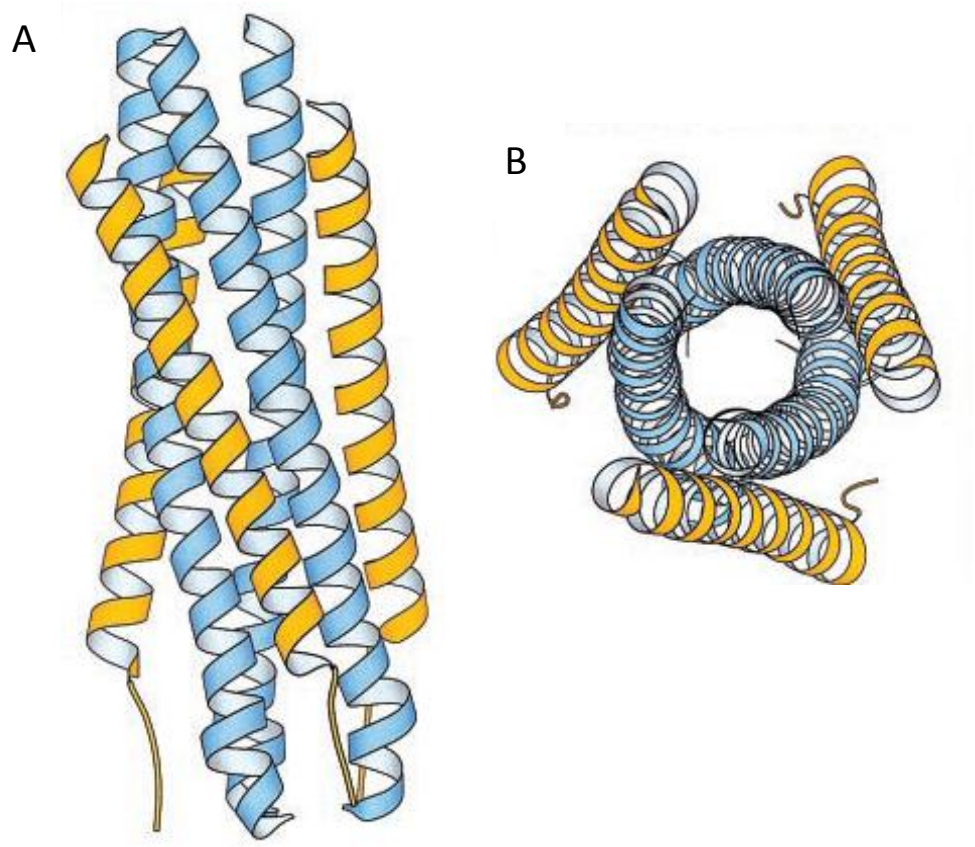


Figure 1.7: The 6-helix bundle of HR1 interacting with HR2 diagram adapted from Zhao *et al.,* (2000).[1] Blue represents HR1 α-helical trimer and yellow represents HR2 α-helical trimer

Together the heptad repeats form a 6-helix bundle that is approximately 68Å long and 27Å in diameter. This relates well to the structure of HIV's fusion protein gp41 and other viral fusion proteins.[12] Within the *Paramyxoviridae* the positioning of the repeats is highly conserved, with 270 residues between them.[1] This was confirmed and expanded upon by crystal structure determination of the two heptad repeats as the 6-helix bundle but bound to a potential fusion inhibitor, TMC353121, discussed in section 7.6.[38]

Some structural information for RSV F can be gained from studies of other members of the *Paramyxoviridae* such as the crystal structure of human parainfluenza's F protein, and from the high degree of conservation of the F proteins within the *Paramyxoviridae*.

Current knowledge of the hRSV F Protein's full structure consists mainly of the aforementioned structure of the two heptad repeat bundle, and a homology model. The homology model is based on the high resolution crystal structure determination, by X-ray crystallography, of the Newcastle Disease Virus F Protein combined with the heptad repeat hexameric bundle structure, see Figure 1.8.[39] This is because the sections of the NDV F Protein structure that could not be solved in this study were the heptad repeats. Therefore these two sections of information have been pieced together to give a model of what the hRSV F structure should look like. However, more recently the full structure of NDV-F has been solved, including the structures of the NDV F heptad repeats, and this can now be compared to the homology model and analogies made with RSV-F.[29, 40] Comparing the two NDV F structures, the major difference is the inclusion of the hexameric bundle but over the whole structure there is very little difference. The root mean square difference between them, referring to the α-carbon atoms of 340 residues, being 0.74 Å.[40] The small differences come in the loop regions.[40]

Likewise there are crystal structures that have been obtained for hPIV3 and SV5 fusion proteins, and due to their close relation with RSV F in terms of function, the structure and mode of action of RSV F can be partially elucidated from this information. Class 1 viral fusion proteins tend to be similar in terms of structure and modes of action but not as much so in terms of sequence, and thus such comparisons should be treated with caution.[41, 42, 43] For example, the sequence identity between NDV F and hPIV3 F is only 29 %, with 47 % sequence similarity, and with PIV5 these figures rise slightly to 34 % and 52 % respectively, which are reasonably high.[40] The full crystallographic structure of post-fusion NDV F (resolution to 3.5 Å, slightly lower than for the previous partial structure) shows many general structural similarities with that of hPIV3 fusion protein, with minor differences attributable to the differences in their sequences. The main part of the PIV5 fusion protein structure that corresponds well with NDV F and hPIV3 is that of the heptad repeat hexameric bundle structure, this protein core is well conserved in the viral fusion proteins and would seem to be extremely important for their functional integrity.[40, 43] Charge distribution studies comparing these three known structures indicated a huge similarity between them, which is again surprising given the low sequence homology. The stalks are generally negatively charged and the three trimer lobes in each are positively charged- the conservation of such a pattern indicates that

the function of the lobes lies in facilitating the interaction of the protein with the plasma membrane.[40]

Due to the large number of structural similarities between the current structures of class I viral proteins and the fact that NDV F was used to create homology model of RSV F, this review will concentrate upon NDV F. Out of the three viral fusion proteins that already have a solved structure, RSV F has moderately higher sequence identity with NDVF than that of PIV5, with 13.5% compared to 12.4%, and only slightly lower than that of hPIV3 at 17.2%.[44]

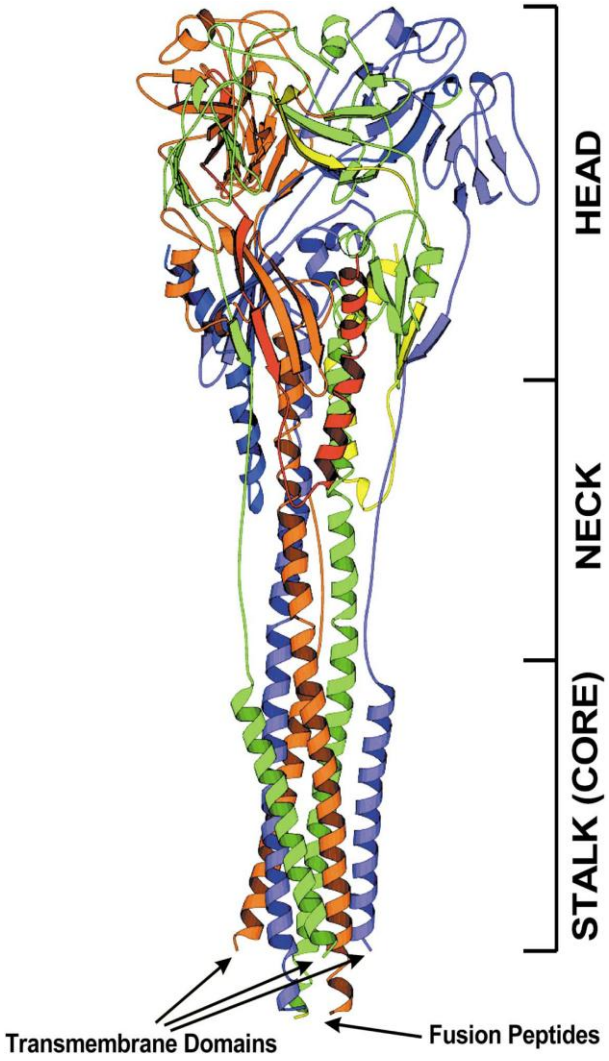


Figure 1.8: Homology model of RSV F based upon X-ray crystal structure of the hexameric heptad repeat bundle combined with partial X-ray crystallographic structure of NDV F. Adapted from Morton et al. 2003.[39]

The NDV F protein structure study revealed the protein's organisation to consist of three regions described as a head, a neck and a stalk. In the head, each monomer of F shows an immunoglobulin-type β-sandwich and highly twisted β-sheet. As with the

structure of the heptad repeat regions for RSV, the second crystallographic structure of NDV F confirmed the hexameric bundle structure to be present as the stalk. Interesting features to note are the three radial channels through the head and one central, 50Å long, central channel. These channels provide the method of exposure of the fusion peptide during conformational change and protein activation. When inactive, the fusion peptides are sequestered in the radial channels.[12]

The NDV protein F has identical function to hRSV F. Their morphology is likely to be highly similar as they have important structural and sequence features in common.[39] This implies a homology model would be a useful tool. However, the partial NDV F protein structure used for the model was only solved to a resolution of 3.3 Å and the sequence homology between NDV F and RSV is below 15 %, below the recommended threshold for creating a homology model.[12] Their sequence identity is only 13.5 %, as they are from different subfamilies. The homology model was also only of the active conformation of the protein. There are regions of the two F proteins where the secondary structure was predicted to be different and therefore the model is even more unreliable in these regions, particularly regarding disturbances to the predicted secondary structure of the heptad repeats. [12]

F protein presents as a homotrimer consisting of three units of the F_1-plus-F_2 entity.[25] It has been shown to be relatively thermostable because it is highly helical.[25] Alpha helices and especially coiled-coils are highly thermostable due to the large amount of intra- and interchain hydrogen bonding involved. Oligomerisation into a homo-oligomer occurs within the rough ER, before cleavage by proteases into F_1 and F_2. [12, 24]

1.3.6 Relating 3D structure to function

1.3.6.1 Fusion

It is likely that the mechanism by which F acts to bring the two membranes together is a conformational change largely involving the F_1 domain. This causes the protein to fold into a horseshoe conformation bringing the TMD (attached to the viral membrane) and the Fusion Peptide (attached to the host cell membrane) into juxtaposition (see Fig. 1.8).

Studies into the structure of the heptad repeat regions indicated that the HR1 and HR2 interact, forming a 6-helix bundle that stabilises the protein. It could be the case that for the protein to be fully functional, these two regions must bind making the active protein. This would occur only when they were sufficiently close, after the conformational

change has begun. It is thought that a number of fusion proteins in close proximity form a fusion pore though which the nucleocapsid can enter the host cell.[45]

Studies on other members of the type I fusion protein family (such as HIV-1 gp41, gp2 of *Ebola virus* and F_1 of SV5, HA of *Influenza virus*) support the hypothesis of the hairpin structure and the role of the HR stabilising the fusion complex. It is unknown whether the hairpin structure forms before fusion or simultaneously with it.[1] By analogy with HIV-1 gp41, the protein could take one of three conformations depending on the stage of fusion. There is great structural similarity between the protein cores of gp41 and RSV F. The inactive state is the protein where the heptad repeats are not bound in the hexameric bundle and the conformational change has not yet occurred, the pre-active state would be when the conformational change has occurred but HR1 and HR2 are not bound and the active conformation would be when the heptad repeat regions are fully bound in a heterodimer and the two hydrophobic regions are in the ideal position for the membranes to fuse (see Fig. 1.8). With gp41, in the pre-active conformation HR1 is exposed and the fusion peptide inserts itself into the host cell membrane.[1] The active conformation is the fusogenic hairpin that holds the two membranes together.[1] If either of the pre-active or inactive conformations could be prevented from progressing into the active conformation then membrane fusion and therefore viral entry could be also prevented. The deep groove of the HR1 trimer could be a promising drug target due to it having a hydrophobic cavity at its C-terminus and many hydrogen bond acceptors and donors. This would give strong binding ability for a drug molecule.[1]

This model of gp41 in different conformational stages during the fusion process agrees with the appearance of RSV F when observed by electron microscopy. Anchorless, full-length uncleaved F_0 presents as unaggregated cone-like structures anchorless. Fully cleaved and partially cleaved (at site I not yet at site II) mature F presents as aggregated rosettes of lollipop-shaped spikes.[27] It is only anchorless, cleaved F that presents as rosettes.[27] Cleaved mature F always presents as aggregated cones or aggregated lollipop spikes due to interaction between their stalk regions.[46] The protein only aggregates when it has been cleaved at both site I and site II. If the anchorless protein is cleaved with trypsin at both sites, it aggregates into rosettes of lollipop spikes. This is likely to be due to the interactions of the exposed fusion peptide trimer (as this is the only remaining hydrophobic domain). This change in shape reflects the change in activation of the F protein, with the unaggregated cones being the pre-active form and the lollipop spikes being the active form.[28, 47] The lollipop shape is achieved

via transference of the HR2 from the neck and upper stalk to the stalk's base during conformational change, diminishing the protein's width from cone to lollipop.[46]

The obvious problem with the current crystallographic structure studies on RSV F is that none have dealt with the full length protein, or even the anchorless protein. The only structure that has been successfully crystallised and solved is that of the two heptad repeat peptides bound to each other (see Figure 1.7).[1] This is not fully representative of the structure of the protein in its active conformation as most of the sequence is missing from the structure and this absence could greatly affect the mode of their binding. It is clear that realisation of the full protein structure would provide a much greater level of information to design potential F inhibitors. Also, solving the two bound repeats does not give any indication of the structure of the inactive conformation. So far this has only been inferred from the NDV-F homology model, EM images of F and analogy with gp41's mode of action. It would be most useful to obtain structures of all conformations of the protein (inactive, pre-active and active) in order to obtain a complete picture of its mode of action.

1.3.6.2 F stimulates the immune response

Neutralising antibodies act against F and G proteins but the two strains of hRSV, A and B, are defined by differences in the interaction of F with these antibodies. G is the same in both strains.[24] There are three or four antigenic sites located on F. Antigenic sites IV, V and VI are located close to HR2, next to the cysteine rich region. Antigenic site II (not to be confused with furin cleavage site II) is at the N terminal of F_1, at the opposite end of the cysteine rich region. Therefore this region may be important for antibody recognition.[24] These antigenic sites are similar to those of the F proteins of NDV, Sendai and parainfluenza.[24] The homology model based on NDV F does concur with known antigenic site positions of RSV F, which is another indication that it is a good fit. Site I is close to the crown of the head and IV, V and VI are at the pinnacle of the trimeric structure within the head.

1.3.6.3 F and species specificity

The hRSV F Protein is also one of the elements that decide the virus' species-specific entry into host cells.[5] It was expected that it would be G, in its job of first attachment, that would control species specificity. However it is likely that G simply attaches to surface glycosaminoglycans (GAGs) as it is not species specific.[22] It is in fact F_2 that is the determinant. GAGS do not have to be present for F to bind. Chimeric proteins of hRSV and bRSV F_1 and F_2 showed that only when F_2 of hRSV was present did the

virus infect human cell lines of peripheral blood lymphocytes. When F_2 of bRSV was present the virus infected bovine cell lines.[22] The chimeric proteins were still functional in the opposing cell line due to the homology of 83 % between the two F proteins, but were less efficient. Therefore another importance of F has been illustrated and another incentive to inhibiting its action found. It is likely that F interacts with more specific cell surface receptors than GAGs, after G has bound non-specifically.[22] Both bRSV and hRSV F can also bind heparan sulphate on the cell surface, and so F also contributes to the non-specific binding.[22] This could contribute to its ability to carry out viral entry without the presence of the G protein. hRSV can successfully enter non-human cell lines but cannot replicate in them. A study by P.J. Branigan et al. in 2005 on feline, equine, avian, bat, canine and other vertebrate cell lines showed that F can carry out fusion of non-human cell lines. However, GFP labelling of mRNA revealed that the virus does not replicate in any of these cell types and therefore specificity is mainly determined post-entry. The fusion protein may use less specific or highly conserved cell surface receptors, or the virus may be able to enter cells using a number of different mechanisms.[48] This considered together suggests that although F_2 does have a role in species specificity between bRSV and hRSV, there are other factors, dependant on species, controlling the ability of the virus to replicate inside the host cell.

1.3.7 The importance of lipid assemblies in fusion

The relative importance of the action of lipid assemblies versus the actions of proteins in fusion processes is subject to debate. The protein has the stored conformational energy, assembles the closed lipid bilayers, alters conformation, stabilises unstable lipid structures and forms the fusogenic conformation to allow fusion. Obviously the form of the lipid bilayers of host and viral membrane is an important aspect of fusion but whether a change in the bilayer assembly occurs first to allow membrane fusion (lipidic pore hypothesis), or whether a protein drives the change first (protein pore hypothesis) is uncertain.[49] The interaction of the F protein TMD and Fusion Peptide with the bilayer is, in either case, crucial to fusion. The fusion pore is created with the coming together of an unknown number of fusion proteins, when the fusion proteins are in their active conformation and fused with both the virion and the host cell membrane. They bring the two membranes together and at a point between the fusion proteins, the two membranes fuse creating a long “stalk” and then a flat hemifusion diaphragm, where the outer leaflets of the membranes fuse, with increasing tension. Finally, this increasing tension from the pulling of the fusion proteins causes the diaphragm to break apart resulting in the fusion pore via which the nucleocapsid can be transferred into the host cell.[45] Lipid rafts are important because the interaction between the Fusion

Protein and the plasma membrane occurs at lipid rafts in the membrane. F targets the virion to these lipid rafts for budding, and does not require any other surface protein to do so. The target signal within F that controls this is not located in the C terminal or TMD nor does it involve modification of residues with fatty acids. The domain is within the extracellular portion of F. It may be that assembly and budding occurs at lipid rafts as they provide a domain for the proteins to localise to.[50]

1.4 The Role of M2-1: Transcription

M2-1 proteins are present in all pneumoviruses. hRSV M2-1 is a 194 aa, 22,150 Da cofactor of the viral RNA polymerase complex that acts as an elongation factor, more accurately an anti-termination factor.[3] It is not involved in replication but is essential for transcription.[3] It enhances read-through at gene junctions and therefore promotes long mRNA synthesis. This means it prevents termination occurring at stop signals. It therefore enables transcription to fully complete and has been implicated in the transcription-replication switch, though its mere presence in the ribonucleocapsid complex with L and P does not facilitate this switch.[51] It is found in both phosphorylated and dephosphorylated forms.[52] It has a predicted zinc finger domain at its N terminus, essential for its function.[52]

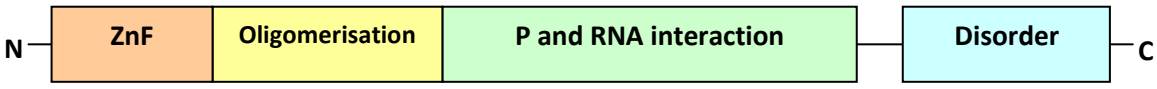


Figure 1.9: Functional domains of the M2-1 protein.

1.4.1 M2 mRNA encodes M2-1 and M2-2

M2-1 is encoded by one of two ORFs contained within the M2 mRNA (see Figure 1.10). ORF-1 encodes M2-1 and a second ORF overlaps and encodes M2-2. This evolutionary mechanism for producing the most number of proteins with the smallest possible genome RNA also occurs in PVM and APV.[53] In PVM, not only does this occur with the M2-1 protein but the gene for the P protein also contains a second ORF.[54] All other transcriptional units of RSV are monocistronic and have highly conserved start-stop sequences.[6]

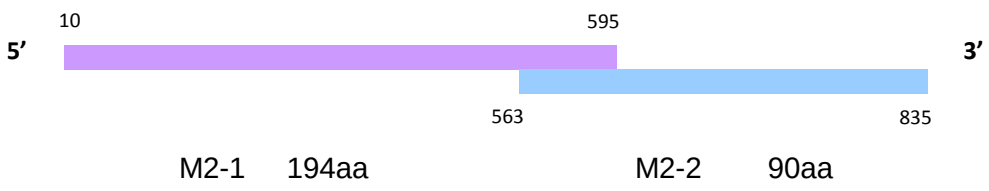


Figure 1.10: The two overlapping open reading frames of the M2 gene segment

M2-2 interestingly inhibits mRNA synthesis, acting as a transcriptional regulator, and has also been implicated in the replication-transcription switch but it is unnecessary for viral replication.[54] It has been noted that low levels of both M2-1 and M2-2 gives transcriptional processivity but high levels of both together results in inhibited RNA synthesis. Removal of M2-1 reduces yield of RNA (both genomic and anti-genomic), by significantly reducing replicative efficiency.[54] M2-1 takes up 70 % of M2's coding capacity. ORF 2 overlaps at the 3' end of M2 by 4, 6 or 10 amino acids dependant upon the initiation codon used in translation of M2-2.[6, 53] Relocating M2-2's initiation codon downstream of M2-1's termination codon resulted in failure of M2-2 expression- thus the size of the overlap is essential for regulating levels of M2-2.[6] M2-2 translation occurs by ribosomal reverse translocation. The ribosome translates M2-1 reaching its stop codon.[53] The overlap region redirects the ribosome back upstream to the start of the ORF2 start codon.[53] This process is significantly more effective for RSV M2-2 than that of PVM or APV.[53] The region that controls the level of expression has been shown to be a very stable region with a high degree of secondary structure, a pseudoknot, located 150 nucleotides upstream of the overlap.[53]

1.4.2 M2-1 is an RNA binding phosphoprotein

M2-1 interacts primarily with viral mRNA for transcriptional anti-termination activity, and the interaction between M2-1 and N is mediated by RNA binding.[3] M2-1 is able to bind RNA with high affinity but binds to long RNAs (with 700 to 1300 nucleotides) with little or no sequence specificity. [51, 52] This would indicate a structural function similar to the chromosomal histones, like the Nucleocapsid protein's protection function.[42] M2-1 is also able to bind the 5' end positive sense antigenomic RNA leader sequences (42 nucleotides long) on segments of short RNAs (up to about 80 nucleotides) with high specificity. The RNA binding ability is regulated by phosphorylation. [28, 51] The RNA binding domain of M2-1 is located at its N terminus, mainly between residues 59 to 85. Phosphorylatable residues Thr56, Ser58, not contained within the general binding region, and Ser61 have been implicated in the changes in M2-1's phosphorylation status and therefore in turn binding affinity for different RNAs.[51] Ser58 and Ser61 are located at the end of the oligomerisation domain, at the surface of the structure.[3] When Thr56 and Ser58 residues are mutated, the binding affinity of M2-1 to long RNAs decreases considerably.[51] This could be due to phosphorylation status but also may be down to alterations in the protein's conformation caused by the substitution. It was hypothesised that the differential phosphorylation alters binding affinity and that binding specificity for the different lengths of RNA and leader sequences may be due to

specific binding by different oligomers of M2-1.[51] A third potential contribution to differential binding could be the long RNAs holding more non-specific binding sites simply due to their length.[51] It has been experimentally determined that phosphorylation of M2-1 results in a 5 to 10-fold decrease in long RNA binding affinity, supporting the first hypothesis.[6] M2-1 is mainly dephosphorylated in infected cells with a minor proportion of phosphorylated form present.[3]

It is residues 59-85 that bind RNA and within these it has been shown that 59-62 may be sufficient for binding.[6, 51] Sequence conservation between residues 59 and 80, comparing human and bovine RSV M2-1, is 95% suggesting an important conserved function. The region within residues 70 to 80 is predicted to be a β-sheet, whilst residues 50-62 are predicted to form an α-helix. Such structures exist in other RNA binding domains of RNA binding proteins. Thr56 and Ser58 account for 96-98% of M2-1's phosphorylation, with other sites contributing a small amount. The residues 85 to 194 are not required for RNA binding.[51]

An important part of the structure of M2-1 is a C_3H_1 binding motif at the protein's N-terminus between residues 7 to 25.[6, 51] This is highly conserved within the *Pneumovirus* genus and in many other proteins this motif has been shown to bind zinc.[51] In some viral proteins including HIV gp41, Zn fingers interact with RNA.[51] However, when the putative Zinc Finger of M2-1 (the first 30 amino acids of the protein) is deleted it results in a surprising increase in binding affinity to long RNAs.[6] Direct binding to RNA by the zinc motif has not yet been found to be the case for any member of the *Paramyxoviridae.*[3] However, a role in regulation of M2-1's phosphorylation has been demonstrated. Deletion of this zinc finger motif results in inhibition of M2-1's phosphorylation therefore it is necessary for phosphorylation to occur, and therefore needed for proper function of M2-1.[3, 55] This also correlates with the relationship between increased phosphorylation and decreased RNA binding affinity. The Zn finger motif in HIV-1 gp41 has been shown to interact with the viral RNA in order to facilitate tight encapsidation, which could be the case for M2-1.[51]

M2-1 has four conserved cysteine residues and three of which are located within the Zn binding motif.[6] Replacing C7, C15 and C21 reduces mRNA processivity whereas replacing the cysteine residue not within the C_3H_1 motif, C96, does not have a notable effect.[6] C96 is therefore non-essential although it has been shown to stabilise the conformation of the RNA binding domain. This was shown during studies into electrophoretic mobility of wild type versus C96G M2-1 suggesting that the conformation was altered by the substitution and that C96 is involved in a disulphide

bond, in addition to the replication of a virus containing the M2-1 C96G mutant being decreased.[3, 6] The mutated protein was able to co-precipitate with N indicating that the interaction with N was unaffected.[6]

The N-terminal thirty amino acids of M2-1 were used to make a chimeric protein with the C-terminal domain from the PVM M2-1 protein.[56] This chimera maintained normal activity in hRSV but the opposing chimera (with 29 amino acids from the PVM M2-1 N-terminal with the C-terminal from RSV M2-1) had less than 5 % of normal activity; this showed unequivocally that the N-terminus is essential to function, at least in RSV.[56] The motif is necessary for read-through activity and phosphorylation status but its precise role is currently only vaguely defined.[3] Leu16 and Asn17 have also been shown to be essential for function.[3, 56] Some zinc fingers bind macromolecules, such as other proteins, instead of nucleic acids. For example, Ebola VP30 is an anti-termination factor like M2-1 that holds a Zinc finger not directly binding to RNA.[3] It could be that it aids formation of dimers and other oligomers or else it could aid in intramolecular interactions helping to correctly fold the protein and holding the protein in a stable conformation, as with the RNA binding domain.[3]

1.4.3 Interactions with proteins of the nucleocapsid

The Zinc finger is vital for the interaction of M2-1 with the N protein.[3, 57] Substituting C7, C15, and H25 showed that this motif is necessary for interaction with the nucleocapsid.[3, 51] The interaction with the nucleocapsid is moderated by M2-1-viral RNA interactions.[3] It should be noted that phosphorylation status does not affect the nucleocapsid interaction.[3] In infected cells, M2-1 is seen to associate in inclusion bodies containing N, P and L and this interaction with proteins of the holonucleocapsid complex is vital for M2-1's function in transcription.[3]

Phosphorylation of P at Thr108 releases the M2-1 protein from P; consequently M2-1's read-through activity is lost.[3, 58] M2-1 interacts with P via a helical domain between residues 53 and 177, the same region within which M2-1 binds viral RNA.[3] If RNA is present, M2-1 will preferentially bind RNA over P. Since P and RNA are both negatively charged, this suggests a competitive binding mechanism for positively charged sites on M2-1, these sites having a higher affinity for RNA. Alternatively, a conformational change could occur once M2-1 has bound RNA meaning P is unable to bind. It has been shown by Tran *et al.* (2009) that without P being expressed, M2-1 is largely phosphorylated. Comparing this with the usual state of dephosphorylation in infected cells indicates that P interacting with M2-1 may affect phosphorylation.[3] P protein is

important because it interacts with the L protein, which needs P to act as a cofactor in order to act as a polymerase during RNA synthesis. L is able to interact with the nucleocapsid using interactions L-P-N-RNA. N cannot encapsidate RNA during replication without P interacting with it.[58] The switch between replication and transcription may be down to the phosphorylation state of P at Thre108. When phosphorylated, the protein is in the viral RdRp complex and cannot bind M2-1 and replication is carried out. When P Thre108 is dephosphorylated, M2-1 can interact with P, therefore M2-1 is now also in the complex, promoting transcriptase activity instead of replicase activity.[58] P residues L101, Y102 and F109 are needed for the interaction with M2-1.[58]

Furthermore, M2-1 also directly interacts with the final 110 amino acids of the Matrix protein's N terminus.[3, 59] M associates with N and P in inclusion bodies only if M2-1 is present.[57, 59] There is potential for creating inhibitors of the M2-1-M interaction in order to prevent viral assembly.[59]

1.4.4 The C-terminal domain of M2-1

The heterogeneous nature of M2-1 proteins' C termini does indicate that this region has little importance for function. Deletion of 47 amino acids from the C-terminus of M2-1 prevents full mRNA synthesis and results in a failure to produce viable viral progeny.[6] C-terminal domains originating from different viral M2-1 proteins have different lengths and the shortest (that of PVM) is 17aa shorter than the A2 strain of hRSV; these 17 aa can be removed with little consequence to the protein's read-through function but reducing efficiency of viral replication by up to 60%.[6, 56]

1.4.5 3D structure knowledge

Knowledge of the 3D structure of M2-1 is, just as for to the F protein, very limited. It is simple to define the M2-1 protein in four sections: zinc finger, oligomerisation domain, globular α-helical domain (binds P and RNA) and the C terminus, see Figure 1.9.[3] Sections of secondary structure have been predicted. Circular dichroism results indicate that M2-1 is mostly α-helical. M2-1 is known to form globular 7.6 nm diameter tetramers of approximately 89 kDa.[3] Oligomerisation occurs via residues in the region 32-63: deleting this region gives greatly reduced replicative efficiency. The region 35-58 is highly conserved within the family and consists of regular hydrophobic and polar residues on the same face of the helices. It is indicative of an amphipathic coiled-coil. Even with a GST or His tag attached, M2-1 will form oligomers. This indicates that

oligomerisation of M2-1 is vital for full functional ability.[3] Phosphorylation of M2-1 does not affect its ability to oligomerise.[3]

1.5 Conclusion

There has been a great deal of information collected on the viral proteins encoded by hRSV, their biological activity and on the viral genome, its replication and transcription. Their purposes are well defined and yet their structures remain unknown. Structural information is available on related proteins so analogies can be made and preliminary structural studies have provided an insight as to their possible mechanisms. However, without full structures to a high resolution of a viral protein, designing inhibitors is much more difficult, especially given the trial and error nature of drug development, the only option being to construct a screen based on its biological activity. Another hurdle to gaining the full structure of F is the hydrophobic fusion peptide in the middle of the sequence, which would have to be synthesised chemically and characterised separately due to difficulties in purification. Preferably it would be incorporated into the rest of the protein. Therefore if eventually the structure of the RSV F is solved, it may merely be an important part of the jigsaw puzzle and it may well take many years to directly design an inhibitor to fit the structure and many more years to develop it. The same can be said of M2-1, and with its multifunctional mechanisms this structure may be even more complex to decipher. The full complexity of either protein's mode of action cannot be completely understood until full crystallographic structures are obtained.

Chapter Two: Production of the Fusion Protein

2.1 Cloning of FSol using the Baculovirus Expression System

Work on the fusion protein sought to clone, express and purify the soluble construct FSol. FSol is the extracellular domain of the F_1 component of the Fusion Protein, which does not include the TMD or the Fusion Peptide. Previously, bovine RSV F protein has been successfully expressed; including using a baculovirus vector in Sf9 cells.[60] A baculovirus system was used as early as 1989 to express the Fusion protein without the C-terminus, but has not been successfully purified.[61] The protein is only partially soluble and therefore the aim of this project was to create a soluble construct to be used in further analysis and crystallisation. A second aim was to begin synthesis of the Fusion Peptide separately by chemical methods. The work on the synthesis of the Fusion Peptide is described in Chapter Five.

The original cloning and expression strategy aimed to clone the full FSol ORF and express the protein in the Invitrogen Baculovirus Expression System.[62] Using Sf21 insect cells to express the protein should result in most post-translational modifications, such as glycosylation, that normally occur to give functional F protein. The strategy has two stages. First we need to construct a vector expressing FSol fused at the amino terminus to a honey bee melittin export peptide which will promote the secretion of FSol into the extracellular medium. The cloning will also result in the addition of a histidine tag to the carboxy terminus to facilitate downstream purification. This protein construct will then be transferred to a baculovirus shuttle vector ready for insertion into a baculovirus. This would involve initially cloning the FSol ORF into the vector pMIB/V5-His, in frame with the melittin and His tags, and then subcloning into the FastBac™ vector. Whilst the pMIB/V5-His vector can itself be transfected into insect cells for protein expression, expression levels are up to 10-fold higher using the Baculovirus expression system.[63] The system of choice is the Bac-to-Bac cloning kit.[62] In this the generation of a recombinant virus genome is performed in bacteria by homologous recombination between a bacmid (baculovirus genome modified with a plasmid origin of replication and antibiotic resistance marker to allow replication in the cell as a plasmid/cosmid) and a shuttle vector with our target sequence.

2.2 Materials and Methods

General Materials

All restriction enzymes and buffers were obtained from Fermentas with the exception of *Sma*1 and its appropriate buffer Buffer J, which were obtained from Promega. All polymerase enzymes (Platinum Pfx, Pfu, GoTaq) and buffers were obtained from Invitrogen. TOPO cloning kits were obtained from Invitrogen. The pFastBac^{TM} vector and baculovirus stock were obtained from Invitrogen. All primers were obtained from Eurofins MWG Synthesis GmbH and diluted with deionised sterile water to a concentration of 100pmol/µl before use. Blue/Orange 6x agarose gel DNA loading dye was obtained from Promega. 1kb Generuler DNA ladder was obtained from Fermentas. NuPage LDS 4x sample buffer (NP0007) protein loading buffer was obtained from Invitrogen. The Biorad Precision Plus Unstained protein ladder was used throughout. T4 DNA ligase and buffer were obtained from Invitrogen. Liquid S.O.C. medium (2 % Tryptone, 0.5 % yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, 10 mM magnesium sulphate, 20 mM glucose) was obtained from Invitrogen.

LB growth media: 10 g Tryptone, 10 g Yeast and 5 g NaCl in 1 L of distilled water. This was autoclaved for sterility.

Ampicillin was employed at a final concentration of 100 µg ml^{-1} (stock 100 mg ml^{-1} in water). Chloramphenicol was employed at a final concentration of 34 µg ml^{-1} (stock 34 mg ml^{-1} in ethanol)

Instrumentation

Boeco M-24OR Microcentrifuge
Beckman Coulter Avanti^{TM} J-20XP centrifuge
Techne TC-312 Thermocycler
BDH TFX-40M UV gel visualiser
Olympus IX71 immunofluorescent microscope, 10x objective lens

2.2.1 LB/Agar Plates

5 g Agar was added to 300 ml LB media and autoclaved to dissolve the agar for pouring, and for sterility. The appropriate amount of the necessary selecting antibiotic was then added when the agar solution was cooled below 50 °C and plates poured.

2.2.2 Growth and maintenance of SF21 cells

Sf21 insect cells were maintained from stock in TC100 medium (Gibco), 10 % FCS and antibiotic penicillin/streptomycin (50 U penicillin and 50 µg streptomycin per millilitre) and incubated at 26 °C. Cells were routinely spilt 1 in 10 when they reached confluence. As Sf21 are largely non-adherent, the flask was tapped until the cells detached from the surface of a flask into 10 ml of fresh medium, 1 ml was added to a fresh flask containing growth medium. To prepare frozen stocks for long term storage: cell were detached from a 75 cm^2 flask as before, pelleted at 300 x g for 3 minutes and resuspended in 10 ml TC100 + 10 % v/v DMSO. 1 ml was pipetted into each of 10 cryo-tubes and frozen over night at -70 °C before long term storage at -150 °C

2.2.3 Transformation of chemically competent cells

The primary expression vector, pMIB/V5-His (Invitrogen) was transformed into chemically competent cells and plated onto ampicillin selective plates as detailed below. The secondary expression vector for expression by baculovirus, the pFastBac™ vector, was transformed into chemically competent cells and plated onto ampicillin selective plates in the same manner.

All steps were undertaken on ice and quickly where possible. Competent cells (TOP10, Invitrogen) were defrosted on ice. 2.5 µl (0.25 ng) plasmid DNA was placed in 1.5 ml Eppendorf tube on ice and 50 µl competent cell solution added. This was incubated for 30 minutes on ice, then heat-shocked for 90 seconds in a water bath at 42 °C. The tube was then left for 2 minutes on ice and then 1 ml LB (without antibiotics) was added. This was incubated for 45 minutes at 37 °C then 100 µl and 200 µl of the solution were spread onto LB/agar plates. These were incubated at 37 °C overnight. Background plates were incubated for comparison.

2.2.4 Agarose gels

All agarose gels were made with 1 % (w/v) agarose in TAE (242 g Tris base, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA (pH 8); made up to 1 L with distilled water) buffer stained with ethidium bromide. All samples were loaded using 1 volume of Promega blue/orange 6X loading buffer to 5 volumes of sample. They were run at 120 V, constant current, for 30 minutes.

2.2.5 Primer design

The primers were designed as shown in Appendix B and contained restriction enzyme sites (in bold). Synthesis was undertaken by MWG (Germany). When received the primers were resuspended in deionised water at a final concentration of 100 pmol μl^{-1}.

2.2.6 PCR amplification

1 µl containing ca. 20 pg of template DNA was added to a reaction mix consisting of 5 µl 10X Buffer, 1 µl reverse primer, 1 µl forward primer, 1 µl dNTP mix and 35 µl deionised water. Finally 1 µl (ca 1 unit) Phusion DNA polymerase was added and the reaction tube placed in a thermocycler (Techne). The programme used for the PCR was as follows: Preheat to 95 °C; Initial denaturation 95 °C 1 minute; followed by 30 x cycles of Denaturation 95 °C 1 minute, Annealing 60 °C 30 seconds, Extension 72 °C 2 minutes; Final Extension 72 °C 5 minutes. Reactions were maintained at 4 °C until they could be stored at -20 °C or analysed on a 1 % agarose/TAE gel.

2.2.7 PCR product purification

PCR product purification was performed using the QIAquick PCR Purification kit.[64] Briefly, the protocol used was as follows: 5 volumes of Buffer PB were added to 1 volume of the PCR product and mixed. The sample was transferred to a QIAquick spin column in a 2 ml collection tube and centrifuged at 14,000 x g for 1 minute in a microfuge. The flow-through was discarded and 0.75 ml Buffer PE (wash buffer) was added and centrifuged for 1 minute. The flow-through was discarded and the column was centrifuged again for 1 minute to dry. The column was placed in a clean microcentrifuge tube and DNA eluted by the addition of 50 µl Buffer EB or 50 µl deionised water. This was left to stand for 1 minute, and centrifuged for 1 minute to acquire the eluted DNA.

2.2.8 Mini-plasmid purification

Bacterial cells from a 5-10 ml overnight culture were harvested by centrifugation for 3 minutes at 5000 rpm in a benchtop centrifuge at room temperature and plasmid DNA isolated using the QIAprep Spin Miniprep Kit (Qiagen).[65] Briefly, the pelleted bacterial cells were resuspended in 250 µl Buffer P1 and transferred to a microcentrifuge tube. 250 µl Buffer P2 (lysis buffer) was added and the tube was inverted 6 times to mix until a homogenously coloured solution is obtained. 350 µl Buffer N3 (neutralising solution) was added and the tube was inverted 6 times to mix and the solution became cloudy. The cell debris was pelleted by centrifuging at 13,000 rpm for 10 minutes. The supernatant was transferred to a QIAprep spin column and centrifuged for 1 minute. The flow-through was discarded. The column was washed with 0.75 ml Buffer PE and centrifuged for 1 minute, the flow-through was discarded and the tube was again centrifuged for 1 minute to ensure all wash buffer was removed. To elute the DNA the QIAprep column was placed in a clean microcentrifuge tube and 50 µl Buffer EB or 50

µl deionised water was added to the column. This was left to stand for 1 minute and centrifuged for 1 minute to acquire the eluted DNA.

2.2.9 Restriction enzyme digestion of DNA

Restriction of DNA (plasmid or PCR fragment) was performed as follows: Assuming a 20 µl reaction volume to DNA in water, 2 µl of the appropriate 10 x buffer was added and 2 µl (5-10 units) of the enzyme or enzymes (1 + 1 µl), if a double digest was being performed, added. Deionised water was added to adjust to a final volume of 20 µl. For different volumes the amounts were adjusted so that a 1 x buffer concentration was maintained as well as an appropriate enzyme to DNA ratio. The reactions were incubated for 3 hours at the recommended temperature, often 37 ºC. Plasmids used in cloning experiments were dephosphorylated by the addition of CIAP (calf intestinal alkaline phosphatase) during the restriction digest

2.2.10 Gel extraction for restriction product purification

Restricted DNA fragments were purified using a gel extraction kit from Qiagen.[64] DNA was resolved by agarose gel electrophoresis and excised using a scalpel whilst visualised under UV light.

The DNA-gel fragment was weighed and place in a microcentrifuge tube. To the fragment 3 gel volumes (100 mg gel ~ 100 µl) of buffer QC (solubilising buffer) was added and this was incubated at 50 °C for 10 minutes until the gel was completely dissolved. 1 gel volume of isopropanol was added and the tube mixed. The solution was transferred to a QIAquick spin column in a 2 ml collection tube, and centrifuged for 1 minute. The flow-through was discarded and 0.5 ml Buffer QG was added to the column. This was centrifuged for 1 minute and the flow-through discarded. 0.75 ml of Buffer PE (wash buffer) was added to the QIAquick column and this was centrifuged for 1 minute. The flow-through was discarded and the column centrifuged for another minute to ensure all wash buffer was removed. The column was placed into a fresh microcentrifuge tube and to elute DNA, 50 µl Buffer EB or 50 µl deionised water was added and the tube centrifuged for 1 minute to obtain the DNA.

2.2.11 Alternative gel extraction protocol

The DNA was excised from the agarose gel using a scalpel, ensuring full coverage of hands, neck and face due to use UV for visualisation and the fragments were dissolved in 600 µl Binding Buffer (6 M sodium perchlorate, 50 mM Tris-HCl pH 8.0, 10 mM EDTA). To this, 10 ul 166 mg ml^{-1} silica in water was added and this was incubated at

room temperature for 30 minutes. The suspension was then centrifuged at 12, 000 x g for 30 seconds. The supernatant was removed and the pellet was re-suspended in 200 µl Binding Buffer. This was centrifuged again and the pellet was re-suspended in 750 µl Wash Buffer (400 mM NaCl, 20 mM Tris-HCl ph 7.5, 2 mM EDTA and 50% (v/v) ethanol). This was repeated. Then the suspension was centrifuged, supernatant was removed and the pellet left to dry in the air for 30 minutes. The pellet was re-suspended in 15 µl deionised water and this was incubated at 37 °C for 15 minutes. The sample was then centrifuged at 12,000 g for 2 minutes and the supernatant containing DNA was removed and stored at -20 °C.

2.2.12 Ligation

Into each of two PCR tubes was pipetted 10 µl restricted pMIB/V5-His (ca. 100 ng) and 4 µl 5X Ligase Buffer. 5 µl (50 ng) insert DNA was added, giving a insert: vector ratio between 3:1 and 5:1. To the control tube (background vector ligation) 5 µl H_2O was added in place of the insert DNA. The reactions were heated to 65 °C then cooled to < 30 °C at which point 1 unit of of T4 Ligase was added and the reaction incubated at 14 °C for 16 hrs.

2.2.13 Ethanol precipitation

This was only necessary if using electroporation for transformation of electro-competent cells to concentrate and desalt the ligated DNA in order to avoid an electrical arc.

The ligated DNA was transferred to a 1.5 ml Eppendorf tube and 1/10 vol Sodium Acetate (pH 5.5) was added. This was mixed and 2.5 volumes of 100 % ethanol were added. 1 µl seeDNA (GE Healthcare) was added in order to visualise the DNA when pelleted. The solution was incubated at -20 °C for at least one hour followed by centrifugation at 13,000 rpm for 30 minutes at 4 °C.. The supernatant was removed and the DNA was visualised as a red pellet. The pellet was washed with cold 70 % ethanol and centrifuged at 13, 000 rpm for 3 minutes. The liquid was removed and the DNA allowed to air dry. The DNA was then suspended in deionised water or appropriate buffer if needed.

2.2.14 Electrotransformation

All steps were performed on ice. Electrocompetent *Escherichia. coli* XL-1 blues (a gift from Dr R. Dorazi), were thawed on ice for 15 minutes. 4 µl ligated DNA in deionised water was added to a new 1.5 ml cold Eppendorf tube to which 40 µl of electo-

competent cells was added. This was transferred to a cold electroporation cuvette (0.2 gap, BioRad), placed in the electroporator and an electric pulse with the following settings: voltage 2.5 kV, capacitance 25 μF and resistance 200 Ω. A pulse time of 5.8 msec indicated a successful pulse. 1 ml warm S.O.C. media was added to the cuvette and transferred, with the transformed bacteria, to a universal and shaken for 1 hr at 37 °C then 100 µl and 200 µl were spread onto LB/agar plates with ampicillin. These were incubated at 37 °C overnight until colonies developed.

2.2.15 Colony PCR

Colonies obtained after transformation were screened for successful ligation of insert into the plasmid by colony PCR. A 500 µl PCR reaction solution was made up: 250 µl 2x GoTaq polymerase enzyme reaction mix (includes dye for visualisation on agarose gels), 230 µl H_2O and 10 µl of each primer and split into 25 µl aliquots in PCR tubes. A numbered grid was drawn on a fresh LB/agar ampicillin selective plate. Colonies to be tested were numbered similarly and a pipette was used to streak each colony onto the appropriately numbered square and then spot them into the appropriately numbered PCR tube. The ligated DNA product was used as a positive control. The samples were subjected to PCR programme was as follows: Primary denaturation 5 minutes 95 °C; 30X cycles: denaturation 30 seconds 95 °C, annealing 30 seconds 58 °C to 43.5 °C (temp decrease by 0.33 °C/cycle), extension 1 minute 72 °; final extension 10 minutes 72 °C. Reactions were maintained at 4 °C until they could be analysed on a 1 % agarose/TAE gel.

2.2.16 Topoisomerase cloning of FSol

2.2.16.1 Ligation and transformation with the Invitrogen Champion TOPO Kit [66]

To facilitate TOPO cloning (a ligase free system) PCR (1 µl) products were added to 3 µl deionised water, 1 µl salt solution, and 1 µl TOPO vector. This was left for 5 minutes at room temperature. 3 µl was transferred to a tube containing chemically competent TOP10 cells (Invitrogen) on ice for 30 minutes followed by heat-shocking at 42 °C for 45 seconds. This was then left on ice for 2 minutes to which was added 1 ml warm LB and incubated at 37 °C for 45-60 minutes. 100 µl was then spread onto a LB ampicillin plate.

2.2.16.2 Ligation and transformation with the Zero Blunt® TOPO® PCR Cloning Kit for Sequencing [67]

This kit did not require the specific end on the primer necessary for the previous Champion kit. The protocol in the manual for using the Invitrogen Zero Blunt TOPO PCR Cloning kit was used.[67] For a 6 µl reaction: 4 µl PCR product, 1 µl salt solution, 1 µl Zero Blunt pCR4®Blunt-TOPO® vector were mixed and this was incubated at room temperature for 5 minutes. This solution was transformed into commercial DH5α-T1R One Shot® Chemically Competent cells from the kit, a summary of the protocol follows. 6 µl of reaction mixture was transferred to a tube containing One Shot® chemically competent cells and mixed. This was incubated on ice for 20 minutes. The cells were heat-shocked at 42 °C for 30 seconds then transferred immediately to ice. 250 µl pre-warmed LB medium was added and the tube was shaken for 1 hr at 37 °C. 200 µl and 20 µl were spread onto 50 µg/ml kanamycin selective LB/agar plates, which had been pre-warmed to 37 °C and coated in X-gal for blue/white colony selection.

2.3 Results and Discussion

The initial conventional cloning strategy was repeated several times (PCR amplification, digestion, gel extraction, ligation, transformation and colony PCR), by the methods detailed in section 2.2. Primers used can be found in Appendix B along with the sequence of the full 574 amino acid long F protein. FSol was RT-PCRed from RSV (strain A2) genome RNA and cloned via engineered EcoRI (5') and HindIII (3') into pSVZeo2+, by Dr R. P. Yeo FSol is the 499 amino acid long construct of F that does not contain the Signal Peptide or the TMD.

Cloning stages analysed on 1 % agarose gels were always repeated successfully (see Figure 2.1 and Figure 2.2) until the ligation and transformation stages, where few positive colonies resulted or (in the case of the ligation into pMIB/V5-His) few colonies selectively grew at all. The FSol fragment is 1722 base pairs in length. The pMIB/V5-His plasmid is 3596 nucleotides long.

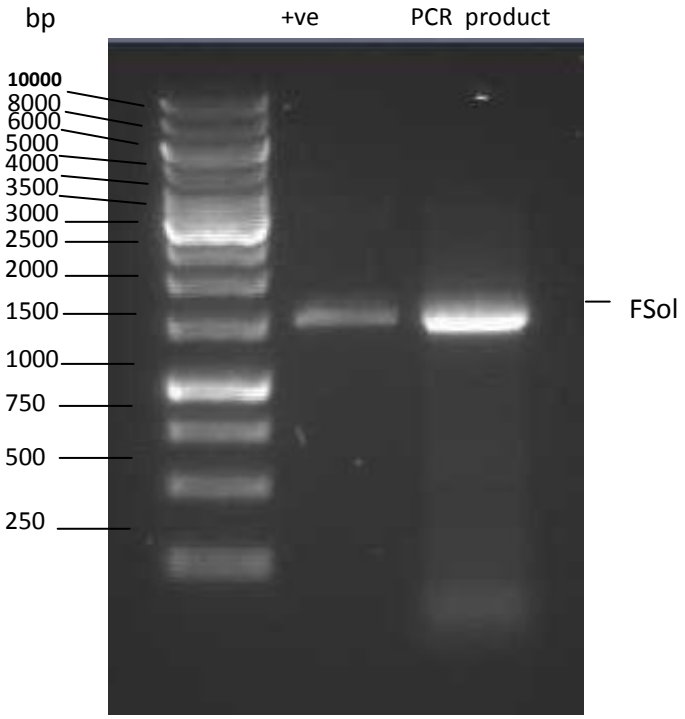


Figure 2.1: FSol fragment as a positive control and the PCR amplified sample of FSol run on a 1% agarose gel

The purified pMIB/V5-His plasmid and FSol fragment were restricted and purified by gel extraction, see Figure 2.2.

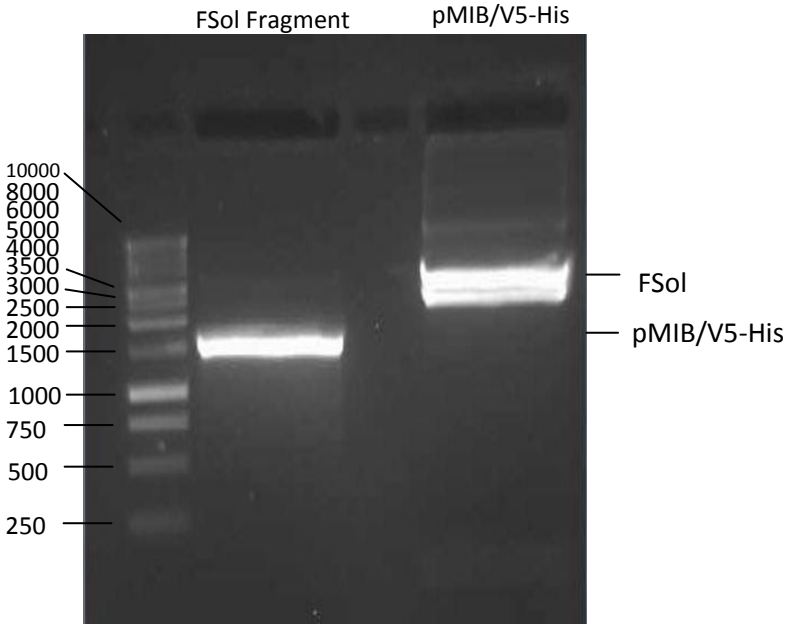


Figure 2.2: pMIB/V5-His plasmid and FSol DNA fragment restricted with *Hind*III and *XBa*1, run out on a 1% agarose gel for gel extraction and purification.

The pMIB/V5-His plasmid was restricted in order to unwind the supercoil for viewing on a DNA gel and check it is the correct size using 10 µg plasmid, 10X Buffer 1 µl, 3 µl H_2O, 5 µl *Pvu*1 restriction enzyme, 1 µl CIAP.

The transformation stage was repeated using commercial chemically competent TOP10 cells from Invitrogen but this did not give any positive colonies upon colony PCR.

It is not possible to analysis the success of the ligation stage on an agarose gel due to the low quantity of DNA used. Therefore effort was made to optimise these stages, using TOPO cloning techniques and commercial cells. Two Invitrogen TOPO cloning kits (Champion and Zero Blunt) were used in order to facilitate the ligation of the construct into pMIB/V5-His.

2.3.1 Topoisomerase Cloning

The principle of TOPO cloning is based upon the activity of Topoisomerase 1, an enzyme from Vaccinia virus that cleaves the phosphodiester backbone of one strand after the motif 5'-CCCTT. To conserve the energy from the breaking of this covalent

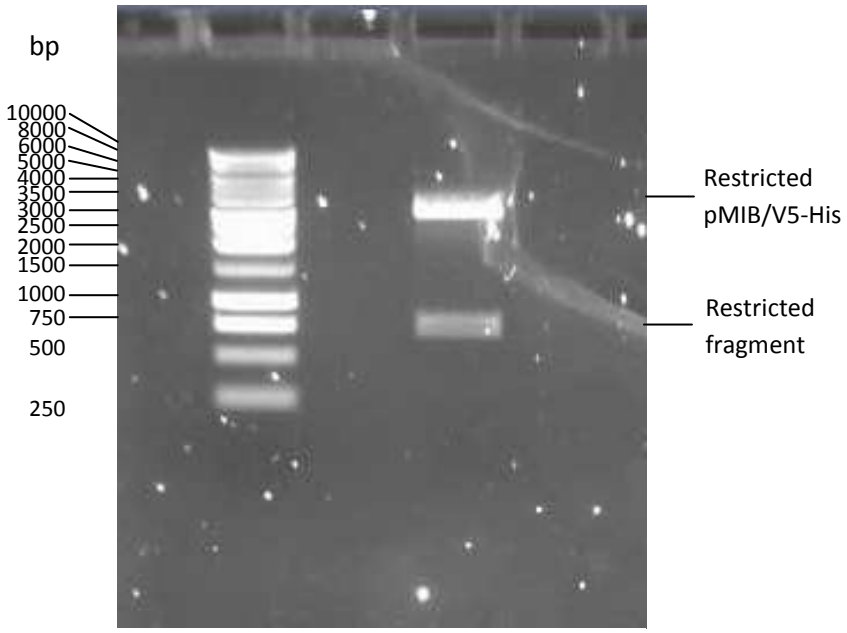


Figure 2.3: Restriction of pMIBV5His with *Pvu*1 to check the plasmid size

bond, one is formed between the 3' phosphate of the cleaved strand and the Tyr-274 of topoisomerase 1. This bond between DNA and enzyme is then attacked by the 5' hydroxyl group of the cleaved strand, releasing TOPO isomerase and reversing the reaction recreating the bond in the phosphodiester backbone. This is used in TOPO cloning to insert in a PCR product to a TOPO vector without the use of ligase.[62, 63]

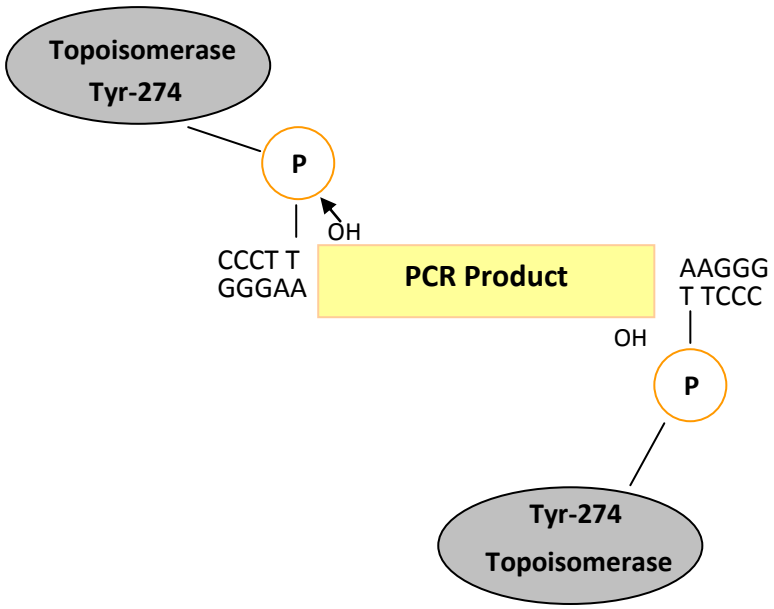


Figure 2.4: Representation of the 5' OH of the cleaved strand attacking the DNA-enzyme bond to form the backbone between PCR product and vector for cloning of a PCR product. Adapted from the Champion™ pET Directional TOPO® Expression Kits User Manual.[66]

2.3.1.1 Directional Cloning: Invitrogen Champion pET Directional TOPO Expression Kit [66]

Directional TOPO cloning, using the Champion pET Directional TOPO Expression Kit, is achieved by addition of an overhang on the TOPO-DNA of GTGG and of a CACC sequence on the 5' end of the forward primer required for PCR of the FSol fragment. The cloning vector overhang of GTGG invades the 5' end of the PCR product, anneals and holds the PCR product in the correct orientation for directional cloning. The same reverse primer as above was used but the forward primer employed here was therefore altered to **CACC**AAGCTTACAAAACATCACTGAAGAATTTTATC (see Appendix B).

This commercial cloning kit from Invitrogen was used, as detailed in 2.2.16.1, as an alternative method of cloning, (the object being to sub-clone into pMIB/V5-His and then express the protein in insect cells) as it should be quick (the ligation takes only 5 minutes to react) and highly efficient for the cloning of the FSol blunt ended PCR product. It also does not require a ligase or use conventional ligation, a procedure which was thought to be a problem for FSol cloning. For this method a new forward primer had to be designed, in order to include the motif CACC on the 5' end of the primer for directional cloning. However, attempts at using this protocol and commercial

kit with commercial TOP10 cells failed, shown by agarose gels of colony PCR negative results (see Fig. 2.5) and the fact there were few colonies produced (under 10 per attempt).

2.3.1.2 Cloning: Invitrogen Zero Blunt® TOPO® PCR Cloning Kit for Sequencing 67

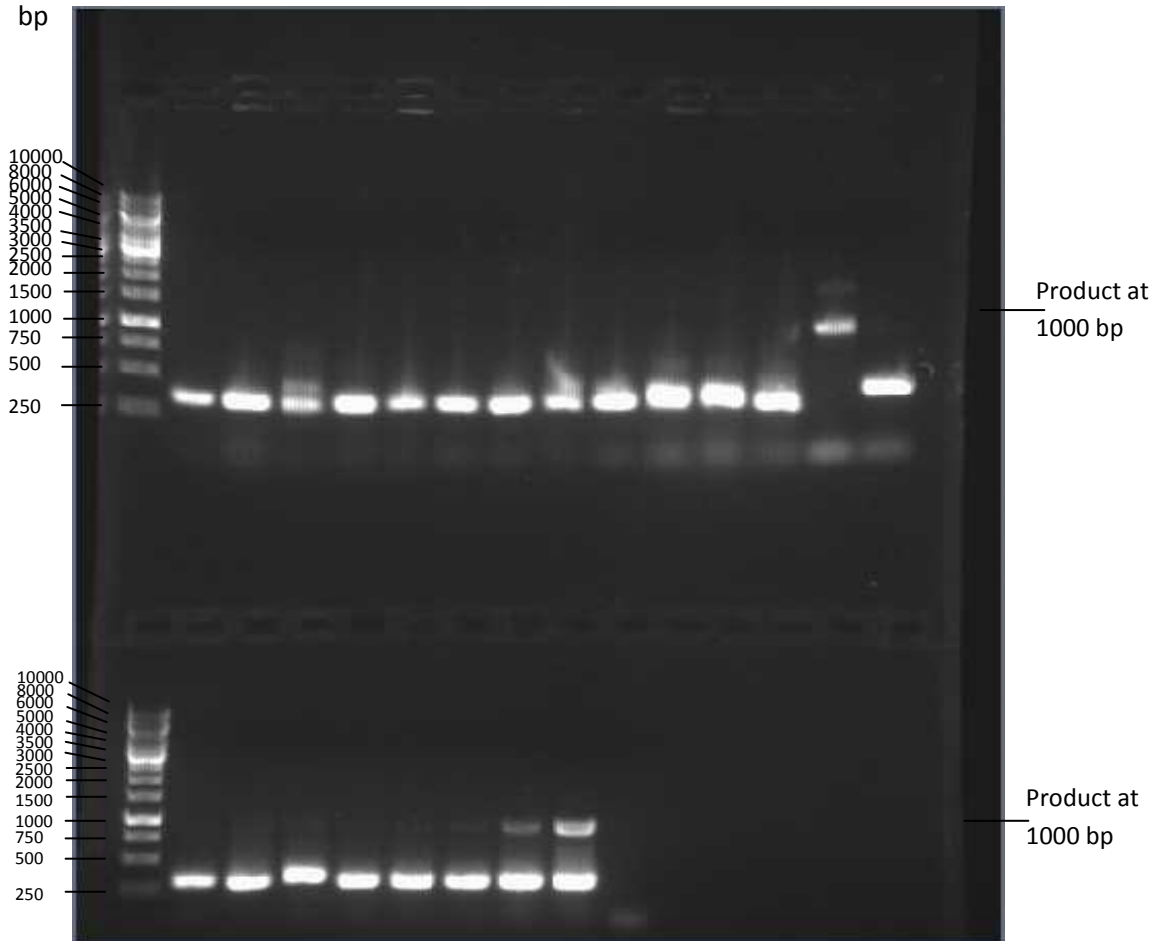


Figure 2.5: Results of colony PCR on 22 colonies from TOPO Zero Blunt cloning and a negative control analysed on a 1 % agarose gel.

The likely reason of the failure in cloning FSol up till this point was due to the fact that ligating a PCR product alone is more difficult than when the desired gene is ligated in another plasmid, as it is more difficult for the ends to be located by the ligase than if anchored in a large plasmid. Cloning of a PCR product is notoriously difficult. Therefore this cloning kit was used, as detailed in section 2.2.16.2, as it optimises the ligation of blunt-ended PCR products to the appropriate site in the Zero Blunt pCR4®Blunt-TOPO® Vector. Then, this can be restricted using the appropriate restriction enzymes (in this

case *Hind*III and *Xba*I) in order to sub-clone the gene into pMIB/V5-His. This can then be transferred to the Baculovirus expression system as planned. The maximum amount of PCR product, 4 µl, was used in order to force the reaction to complete and give the maximum amount of annealed PCR fragment in vector for transformation. More reaction mixture than recommended (the whole 6 µl of reaction mixture rather than 2 µl) was used for transformation for this reason. X-gal was used to screen for positive colonies as the vector supported blue/white colony screening. X-gal is a non-toxic chromogenic substance that can detect the activity of the enzyme β-galactosidase (that hydrolyses lactose into glucose and galactose) in *E. coli* as X-gal is converted by the enzyme into an insoluble blue compound- this allows Lac+ and Lac- colonies to be distinguished.[68]

A band was obtained for two tested colonies on the agarose gel analysing the colony PCR from Zero Blunt cloning, see Figure 2.5. However this band appeared at 1000 base pairs and not at the expected 1500 base pairs. This could be down to mis-priming. Therefore an additional test was conducted on the DNA obtained.

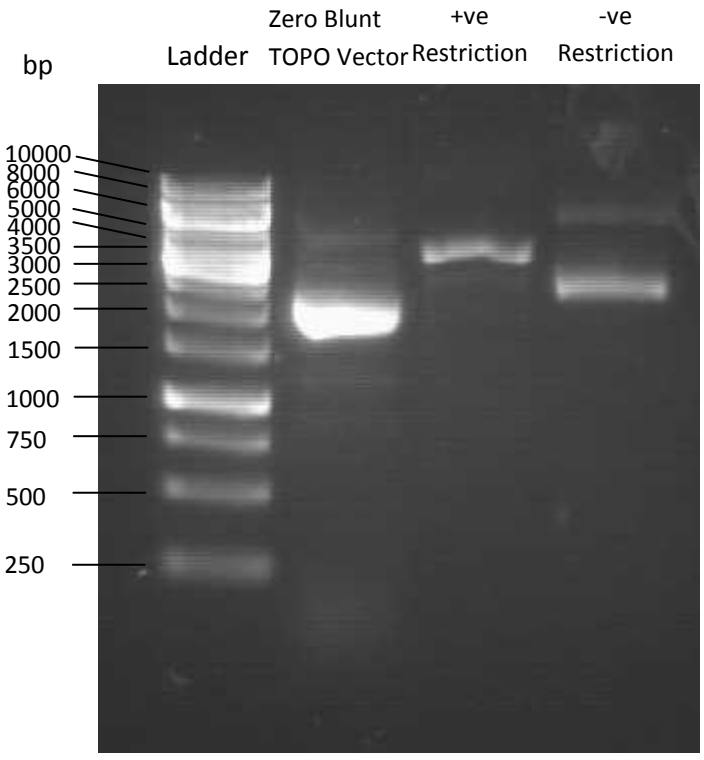


Figure 2.6: Restriction of the recombinant pCR® 4Blunt-TOPO® vector with *Nde*1

The purified recombinant plasmid was tested for the desired gene FSol by restriction with an enzyme (Nde1) that would restrict the insert in one place but not affect the

plasmid. It was seen on a 1 % agarose gel that the restricted plasmid from the positive colony appeared to contain the fragment as the DNA has been cut and linearised and is no longer supercoiled (appearing slightly higher up on the gel). See Figure 2.6.

Therefore it was decided to continue to the next stage of cloning whilst awaiting sequencing results, and the original restriction enzymes *Hind*III and *Xba*1 were used to restrict the recombinant pCR® 4Blunt-TOPO® plasmid and the pMIB/V5-His plasmid for subcloning.

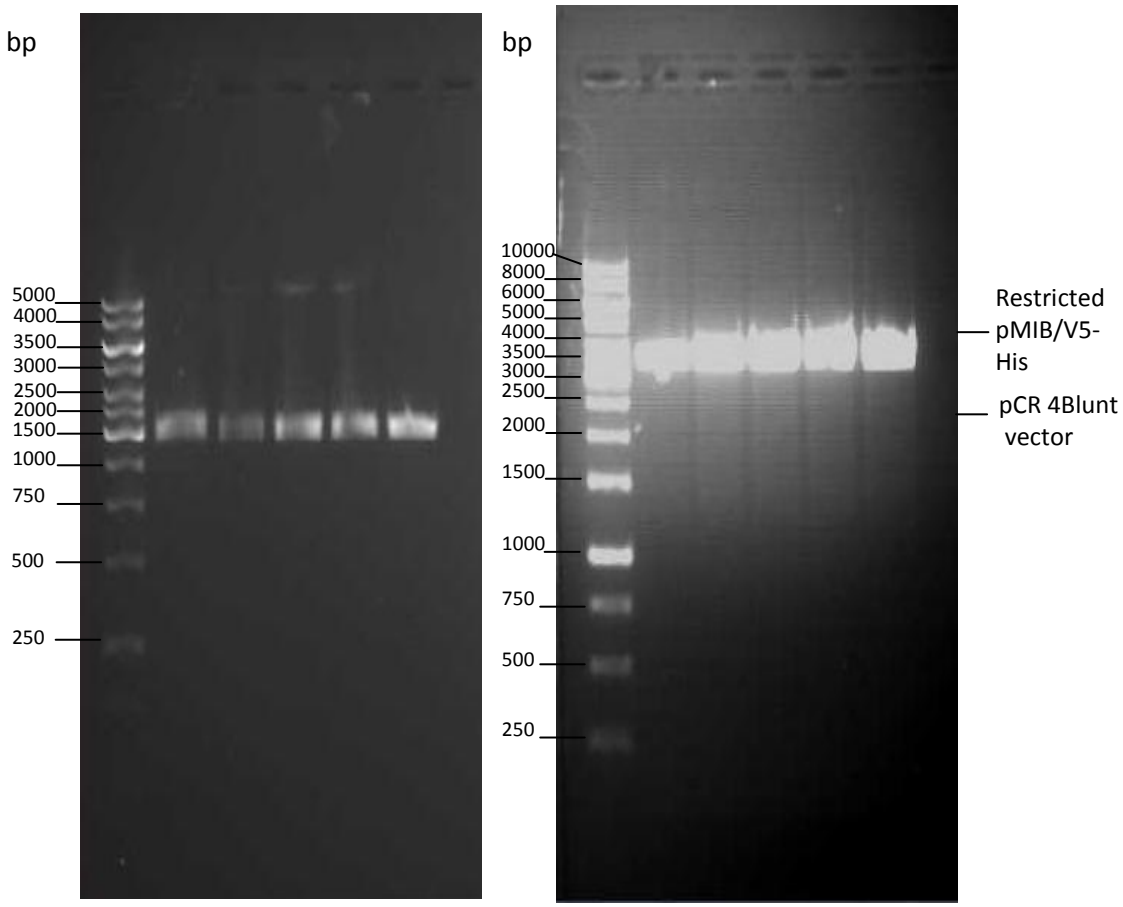


Figure 2.7: Restriction of recombinant plasmid and pMIB/V5-His with *Hind*III and *Xba*1 A) 1 % agarose gel of 5 fractions of cloned and restricted FSol in the pCR® 4Blunt-TOPO® vector B) 1 % agarose gel of 5 fractions of restricted pMIB/V5-His plasmid

The positive clones were sequenced using the vector M13 forward and backward primers. Unfortunately, the result bore no resemblance to the sequence of the fusion protein; this was despite the positive result of the *Nde*1 restriction- obviously the undesired product contained this restriction site. Due to time constraints a new cloning procedure could not be attempted but suggestions for further work are outlined in Chapter Seven.

Chapter Three: Retargeting and Immunofluorescent Imaging of F

3.1 Introduction and Aims

The aims of immunofluorescence experiments were to image the localisation of various F constructs inside and on the surface of lung cells, and to ascertain at what point the Matrix protein and F interact with each other.

3.2 Materials and Methods

Materials

α-bovine F_{ITC} antibody was obtained from Sigma Aldrich. C5335 bovine α-RSV F was a gift from Geraldine Taylor, Institute of Animal Health, Compton, UK. Murine α-RSV M antibody was a gift from Dr R.P. Yeo, University of Durham, UK and was diluted 1/3000 for use. AlexFluor goat anti-bovine IgG was obtained from Invitrogen.

3.2.1 Preparation of a working baculovirus stock

Sf21 cells in a T75 flask were infected with BacF (a baculovirus expressing the full length F protein, a gift from P. Yeo). Typically the cells' medium was removed from the cells and 10 ml fresh medium added. BacF was added to the cells at an m.o.i. of 5-10 and allowed infect for 30-60 minutes with an occasional tilt to ensure the cells did not dry out. A further 10 ml fresh TC100 medium was added to the flask. After 6-7 days, most of the media was removed and centrifuged at 5000 r.p.m to pellet cells and other debris. The supernatant, which is the virus stock, was aliquoted (1 ml) and stored at -70 °C. Titres of 1 x 10^7 ml^{-1} could routinely be produced

3.2.2 Subculturing of A549 cells

The cell medium (2 % FCS in DMEM, Gibco) was poured off and few mls of trypsin were added to the monolayer of cells and swirled over in order to wash the remainder of the medium away. This is in order to remove the Bovine Pancreatic Trypsin Inhibitor (BPTI) for the trypsin to be active. The trypsin was removed and a few more mls added, then versene was added to a ratio of 1:4 trypsin: versene. Versene contains EDTA that aid in the dissociation of cells from the flask. The flask was placed in the incubator at 37 °C for a few minutes. The solution was opaque as the trypsin began to

remove the cells from the flask and at this point the flask was removed from the incubator. 50 ml fresh media was added to the cells, dispersing them evenly throughout the medium. The cells were then split between fresh flasks as needed.

3.2.3 Immunofluorescence assays

From a stock of A549 cells approximately 1 x 10^5 cells in 1 ml was pipetted into each well of a 12 well plate. Each well also contained a 13 mm cover slip. This was incubated at 37 °C until the cells reach 50% confluency. The medium was removed using a pipette and 200 µl fresh media was added to each well. 20 µl human respiratory syncytial virus (m.o.i 1) was added to each well. This was left for 30-60 minutes with an occasional tilt. 2 ml fresh media were added to each well and this was left 33 °C.

Experiment	*Primary Antibody and Dilution*	*Secondary Antibody and Dilution*
Imaging of F in A549 cells	Bovine α-RSV 1/1000	AlexFluor goat anti-bovine IgG diluted 1/250
Imaging of varying F constructs in A549 cells for localisation studies of F	Pure C5335 bovine α-RSV F 0.6 $mgml^{-1}$ diluted 1/1000	α-bovine F_{ITC} diluted 1/250
Transfection of A549 Cells	Pure C5335 bovine α-RSV F 0.6 $mgml^{-1}$ diluted 1/1000	α-bovine F_{ITC} diluted 1/250

Table 3.1: Antibodies and dilutions used in immunofluorescent imaging studies of F

After 24-48 hours the media was carefully removed and the cells washed gently with PBS (x3). The cells were fixed with 0.15 % formaldehyde in PBS containing 0.05% w/v sucrose (19 ml PBS, 1 ml 30% formaldehyde, 0.2 g sucrose) for 10 minutes at RT. This was removed and washed with PBS as before. 0.5 ml permeabilising solution (19 ml PBS, 1 ml Tritonx-100, 2 g sucrose) was added to each well and left for 10 minutes. This was removed by three washes with PBS + 1 % FCS and then left for 30 minutes. Antibodies used and their dilutions are tabulated in Table 3.1. The primary antibody was added and left for 1 hr. Unbound primary antibody was removed by washing three times with PBS. The secondary (usually F_{ITC}-conjugated) antibody was added to each well and left for 1 hour after which the cells were washed with PBS three times. 100 µl PBS + DAPI (to label nuclei) diluted 1/1000 was added, left for 5 minutes and removed.

PBS was added to just cover the cells. The cells in the wells were viewed under an immunofluorescence microscope.

3.2.4 A549 cell transfection

A549 cells were split and transferred 1 ml into each well of a 24 well plate. Overnight these grew to 75 % confluence. The cells were then infected with *Vaccinia* TF7.5 virus by pipetting 15 µl of virus in 100 µl Optimem solution into each well. They were left to infect for one hour then washed with PBS. Transfection mix for 24 wells was made, 200 ng of plasmid and 0.6 µl Fugene was required for each well: 100 µl Optimem per well was pipetted into a sterile 1.5 µl Eppendorf tube and then Fugene was added and then the DNA. The construct used varied as follows with 4 wells transfected with each F construct: F_{WT}, F_{STOP}, F_{ER}, F_{ENDO}, F_{GOLGI}, $F_{SURFACE}$ (see Table 3.2). 6 wells (2 of each F construct well) were also transfected with the plasmid DNA coding the Matrix protein. This was left for 30 minutes at room temperature. The correct amount of transfection mixture was added to each well carefully dropwise. This was left for 6-10 hours and then washed with PBS and fixed with formaldehyde affixing solution as in section 3.2.3. The plate was now safe to work on out of the Microbiological Safety Cabinet. Half of the cells were permeabilised with permeabilising solution and labelling with antibodies was carried out as in 3.2.3.

3.3 Results and Discussion

3.3.1 Surface Expression

An initial experiment was successfully undertaken to visualise the mature F protein on the surface of A549 (lung epithelial) cell infect with RSV, see Figure 3.1.

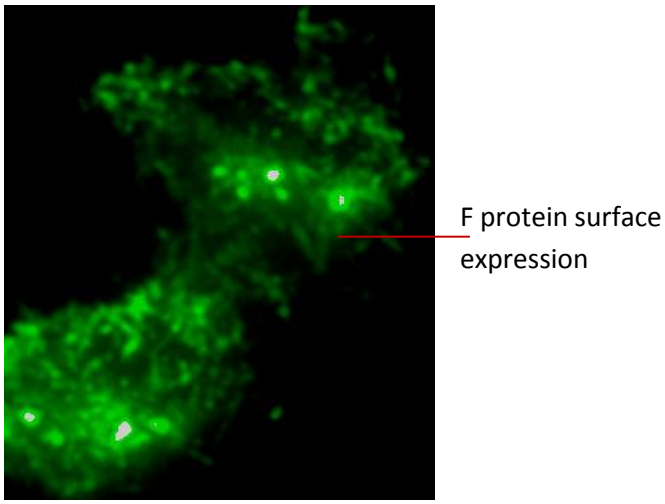


Figure 3.1: Immunofluorescent image of A549 cells infected with RSV and labelled with anti-F.

3.3.1 Retargeting F to specific cellular organelles

Constructs were obtained from Dr R.P. Yeo. These were designed to target specific organelles and therefore give compartmentalisation of F by including a signal sequence in the cytoplasmic tail of each F construct.

Imaging experiments carried out by myself and Dr R.P Yeo were successful in showing the localisation of the F constructs to the different organelles they were carefully designed for. The F_{STOP} construct was full-length F protein without the C-terminal domain in order to ascertain whether it is the C-terminal domain that allows interaction with lipid rafts in the membrane. Primers were designed to insert the retention signals in Table 3.2. The anti-sense primers have 18 bases at the 3' end to anchor to the target sequence in the early stages of PCR then the retention signal (for example KKXX, where X is any amino acid) motif, a stop codon and a restriction enzyme site. The additional 9 bases at the 5' end are there to enhance the efficiency of the restriction enzyme.

Label	Construct
$F_{SURFACE}$	KKAAAA
F_{ER}	KKAA
F_{ENDO}	YKGL
F_{GOLGI}	CPSDSEEDEG
Negative Control F	KKAAAA

Table 3.2: Constructs of the F gene used in transfection of A549 immunofluorescence experiments

The preliminary experiments used long exposure times (4 seconds) for the α-bovine F_{ITC}, compared to the DAPI (100 ms) and therefore the relative fluorescence levels seen on the pictures are unreliable and hazy. It may be that the antibody had been exposed to light on a previous occasion and underwent photobleaching. The preliminary results do illustrate that the expressed constructs were correctly locating either on the surface of the cell or within the cell but specific locations and organelles cannot be identified. See Figure 3.2.

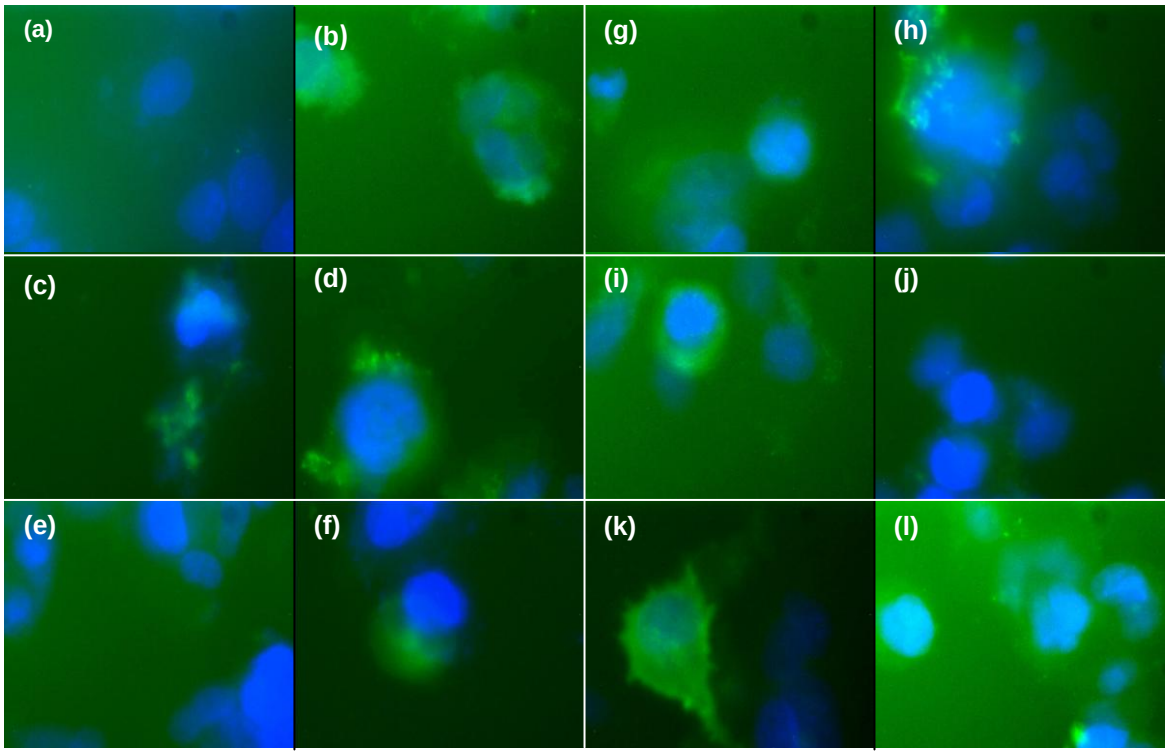


Figure 3.2: Preliminary immunofluorescence of A549 cell transfections, either surface-stained (S) or permeabilised (P) as follows: (a) F_{ENDO} S (b) F_{ENDO} P; (c) F_{ER} S (d) F_{ER} P; (e) F_{GOLGI} S (f) F_{GOLGI} P; (g) F_{STOP} S (h) F_{STOP} P; (i) $F_{SURFACE}$ S (j) $F_{SURFACE}$ P; (k) F_{WT} S (l) F_{WT} P . DAPI was used in order to stain cell nuclei, seen in blue. F protein can be seen in green.

Further experiments carried out using an alternative batch of antibody were successful and gave clear results where the localisation of the F constructs was evident. See Figure 3.3. From Figure 3.3d it can be seen that the construct F_{STOP}, which does not contain the cytoplasmic tail, is still localising to the membrane.

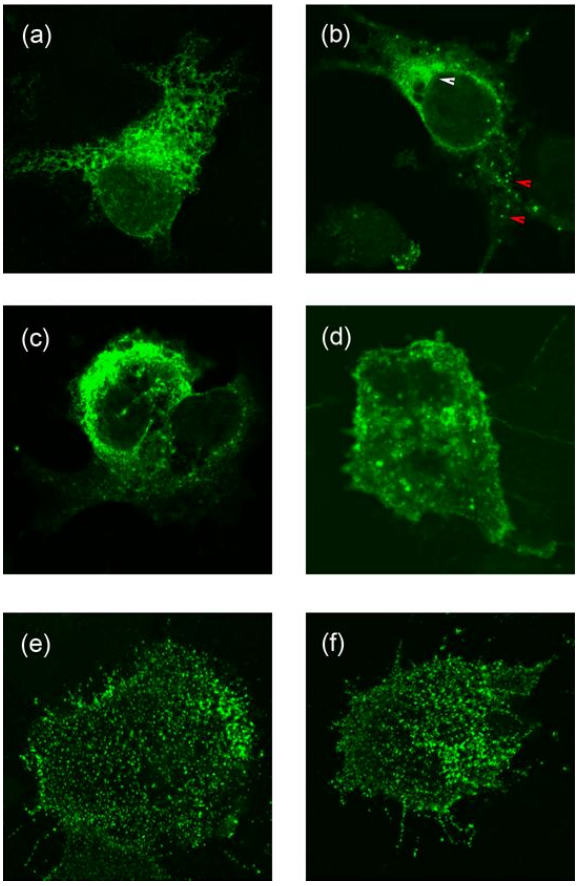


Figure 3.3: Immunofluorescence images of mammalian cells infected with (a) F_{ER}; (b)$F_{GOL-END}$ (F is recycled to the Golgi via the endosomes); (c) F_{GOLGI}; (d) F_{STOP} (e) F_{WT} (f) F Negative control Labelled F protein can be seen in green. No DAPI staining was used.

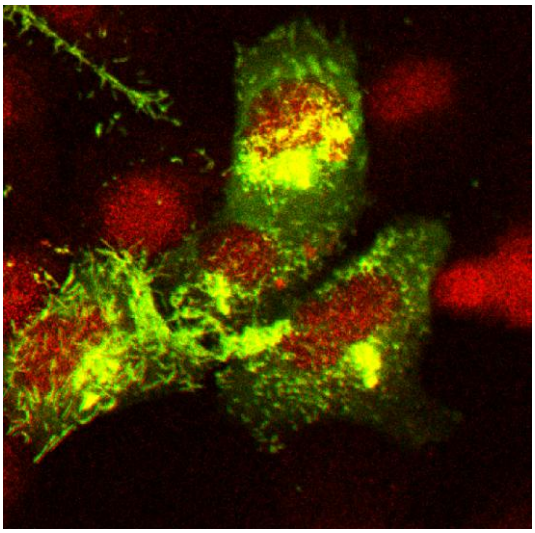


Figure 3.4: Cells infected with RSV and labelled for F for showing viral filaments in yellow. The red signal is a nucleus stain called propidium iodide.

3.3.3 Localisation of F and M in A549 cells

It can be seen that F did indeed interact and co-localise with M. However, it is obvious in Figure 3.5F that this caused massive damage to the cells when the localisation occurred at the membrane. In Figure 3.5H it can be seen that F and M co-localised at the Golgi apparatus. Therefore a new approach will be necessary, outlined in Chapter Seven.

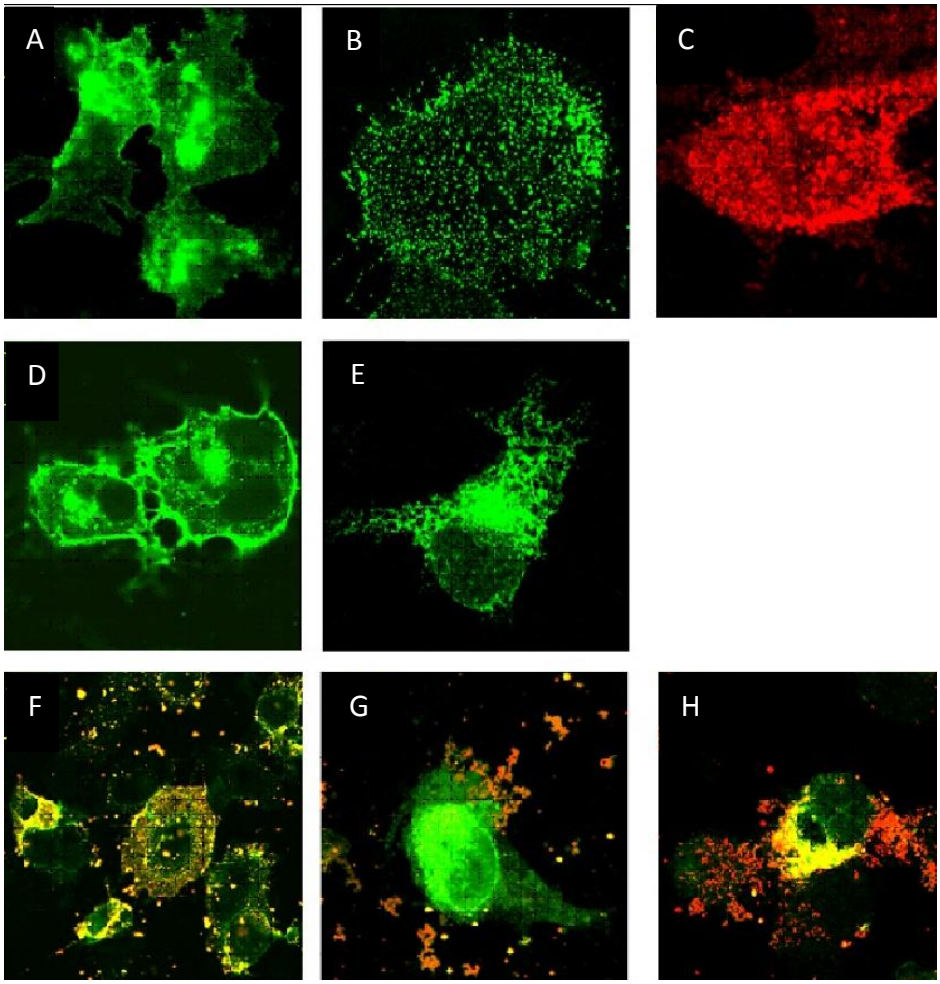


Figure 3.5 Immunofluorescence images of A549 cells viewed with a Ziess confocal microscope transfected with A) F_{WT} P B) F_{WT} S C)M D)F_{ENDO} E)F_{ER} F) F_{WT} and M G) F_{ER} and M H)F_{ENDO} and M. F protein is seen in green, M protein is seen in red and colocalisation of F and M can be seen in yellow.

Chapter Four: Optimisation of the Expression and Purification of the M2-1 Protein in Bacteria

4.1 Introduction and Aims

The aim of this project was to express and purify full length M2-1 protein in preparation for structural analysis by circular dichroism and X-ray crystallography. The sequence of M2-1 can be found in Appendix B2.

4.2 Materials and Methods

Materials and Instrumentation

Standard transfer unit for Western Blotting: semi-dry blotting apparatus from BioRad.

Glutathione-sepharose beads and a 5 ml GSTrap^{TM} FF (Fast Flow) column were obtained from GE Healthcare Sciences.

Thermo Scientific NanoDrop Spectrophotometer 230nm-600 nm

4.2.1 SDS PAGE

To make 5 ml of resolving gel mix (12.5% polyacrylamide) mix 1.75 ml H_2O, 1.25 ml 4 x resolving buffer (1.5 M Tris pH8.8, 0.4% w/v SDS), 2.00 ml 30% w/v polyacrylamide, 25.0 µl 10% w/v APS, 5 µl TEMED.

For 2 ml of stacking Gel: 1.15 ml H_2O, 0.50 ml 4 x stacking buffer (0.5M Tris pH 6.8, 0.4 % w/v SDS), 0.335 ml 30 % w/v polyacrylamide, 15.0 µl 10 % w/v APS, 2.5 µl TEMED

A Biorad Protean III mini-gel system was employed using gel plates with 0.75 mm spacers. Initially the resolving gel mix was poured into assembled glass plates until approximately 1.5 cm from the top. A layer of propan-2-ol was used overlaying the setting gel. This resulted in a more even junction and excluded oxygen which can inhibit the polymerisation reaction. Once set a layer of stacking gel was added along with the comb to create wells and this was allowed to polymerise.

Samples in NuPage LDS sample buffer (Invitrogen) were boiled for 3 minutes before being pipetted into the appropriate wells, the first well containing unstained protein ladder (Promega). Electrophoresis was performed at 200 V for 45-50 minutes. The gels

were dissembled and the gel washed with deionised water and stained for one hour with Coomassie Blue staining solution (0.25g Coomassie Blue, 75 ml Glacial acetic acid, 500 ml methanol, water to 1L), then de-stain solution (methanol 200 ml, 100 ml glacial acetic acid, water to 1 L) was applied until the protein bands were apparent..

4.2.2 Western blotting

1X Western Blot Transfer Buffer: 5.82 g Trizma Base and 2.93 g Glycine were dissolved in H_2O and 200 ml methanol was added. This was made up to 1 L with distilled water. A 10X solution was made.

Nitrocellulose membrane and two pieces of extra thick blotting card were cut to the size of the gel. These were soaked in 1X transfer buffer. Onto the standard transfer unit was placed in order first a piece of blotter, followed by the membrane, then the gel and then the second piece of blotter. A small corner was cut off the membrane for orientation. After assembling the semi-dry blotting apparatus (BioRad) the transfer was performed for 30 minutes at 15 V with a limit of 3 mA. The membrane was then transferred to a clean container and washed with a 2 % milk protein/PBS solution. The membrane was then blocked in 4 % milk protein/PBS for 1 hr followed by washing three times with PBS + 0.2 % v/v Tritonx-100 solution (PBS-T). The primary antibody (anti-M2-1 Mab 312 from mouse) diluted 1/1000 in PBS-T was added- and this was left to bind for 1 hr. The membrane was again washed x3 with PBS-T solution. The secondary antibody, anti-mouse IgG Peroxidase antibody produced in goat, in PBS-T was added and left on the shaker for 1 hr. This was washed x3 for 5 minutes each in PBS-T.

4.2.3 Detection by chemiluminescence

The two components that constitute the detection reagent (Amersham ECL) were mixed together just before they were added to just cover the membrane and incubated for 1 minute and then it was blotted dry gently. In the dark room, the membrane was exposed to X-ray film (X-Ray Retina XBM blue-sensitive film) for 5 minutes, ensuring the membrane is the correct side facing the X-ray film. This was placed in developer solution for a couple of minutes and then placed in Fixer solution for a couple of minutes. Both fixer and developer solutions were Champion Photochemistry RG Universal Ready to Use solutions. Finally the developed film was washed in water and dried before recording image via a digital scanner.

4.2.4 Production of Chemically Competent Cells

A scrape of frozen BL21 Codonplus (BL21C+, Novagen) cells in glycerol was obtained with a loop and put into a tube with 10 ml LB with chloramphenicol. This was incubated in a shaker overnight at 37 °C. 0.5 ml of the overnight culture was added to 50 ml 2xYT media (16 g Trytone, 10 g Yeast Extract, 5 g NaCl, pH 7.2) with chloramphenicol. This was allowed to grow for 3 hours at 37 °C with shaking. 100 mM $CaCl_2$ (2.94 g in 200 ml water) was autoclaved then put on ice. The culture was transferred to a 50 ml Falcon tube and centrifuged for 3 minutes at 5000 rpm at 4 °C. The medium was poured off and the cell pellet resuspended in 25 ml ice cold 100 mM $CaCl_2$. This was incubated on ice for 15-30 minutes then pelleted again in the centrifuge. The supernatant was poured off and the pellet resuspended in 20 ml $CaCl_2$, then this was left on ice again for another 15-30 minutes. This was spun down, the supernatant poured off and the pellet resuspended on ice repeatedly in 10 ml, 5 ml and 2.5ml $CaCl_2$.

4.2.5 Transformation of BCL21C+ cells with GST-M2-1-pGEX-6.

300 µl of competent cells were transferred into a chilled 1.5 ml Eppendorf tube on ice. 1 µl DNA was added. This was incubated for 30-60 minutes on ice then heat-shocked at 42 °C for 3 minutes. This was then left on ice for 3 minutes. Then 700 µl 2xYT was added and this was left to grow for 1 hr then 100 µl was plated on ampicillin plus chloramphenicol selective plates and incubated at 37 °C overnight. 100 µl was grown in 10 ml 2XYT with ampicillin and chloramphenicol overnight.

4.2.6 IPTG induced expression

1 L 2 x YT media with ampicillin was inoculated with a 10 ml overnight culture of transformed cells. This was left to grow for 3 hrs at 37 °C in a shaker. A 1 ml sample was retained for analysis. Isopropyl-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.2 mM -1 mM, this was permitted to grow for another three hours at 28 °C. A sample was taken then the culture was centrifuged at 6000 rpm for 10 minutes at 4 °C. The pellet was stored at -70 °C.

4.2.7 M2-1 Purification

Purification Buffers prepared:

- Lysis Buffer: 50 mM Tris pH, 7.5, 1 M NaCl, 2 % Triton x-100, 100 mM $MgCl_2$, 1 mM $CaCl_2$, 1 $mgml^{-1}$ lysozyme, RNAase A and protease inhibitor tablets (Roche, Mannheim, Germany). Triton x-100 was added last after all other components to facilitate dissolution.

- PreScission Protease Cleavage Buffer: 50 mM Tris HCl pH 7.8, 150 mM NaCl, 1 mM EDTA, 1 mM Dithiothreitol (DTT)

- Elution Buffer: pH 8: 50 mM Tris, 10 mM reduced glutathione

4.2.7.1 Purification using a GSTrap™

All steps were performed on ice or at low temperature where possible. The pellet was fully defrosted and resuspended in 20 ml lysis buffer (this was altered to 50 ml lysis buffer for purification of larger pellets) and incubated on ice for one hour. A lysate was prepared by sonicating the bacterial solution and removing the cell debris by centrifugation in a JA 25.50 rpm rotor at 20000 rpm (ca 40, 000 x g) for 15 minutes at 4°C. The clarified lysate was filtered successively through 1.2 µm, 0.8 µm and finally 0.2 µm syringe filters. The GSTrap FF column was equilibrated with 6 column volumes of deionised water followed by 6 column volumes of lysis buffer. The filtered lysate was loaded onto the column through the 0.2 µm filter at a rate of 4 ml min^{-1}. The through-flow was collected and reloaded on to the column to ensure all tagged protein had adhered to the column. The column was washed with 10 column volumes of lysis buffer followed by 10 column volumes of cleavage buffer. 20 µl PreScission Protease diluted in 5 ml cleavage buffer was loaded onto the column and this was left at 4 °C overnight. 5 ml cleavage buffer was loaded onto the column and 4 fractions of 1 ml were collected. Elution of protease and GST was done using 10 column volumes of elution buffer. The column was washed using 10 column volumes each of lysis buffer, water then 20 % v/v ethanol in water.

4.2.7.2 Purification using glutathione-sepharose beads

All steps were performed on ice or at low temperature where possible. A lysate was prepared as described above. Glutathione beads (GE Healthcare Sciences) were washed three times with 3 ml PBS to remove the storage ethanol containing buffer: 1.5 ml/litre of starting culture beads were added to the clarified bacteria lysate and allowed to bind for 30 minutes at 4 °C on a carousel. The glutathione beads containing bound proteins were collected by low speed centrifugation for 5 minutes. The supernatant was carefully pipetted off (as beads easily resuspend) and the beads were washed with an excess volume of lysis buffer (x20 volume of beads) three times followed by three times in cleavage buffer. At this stage a 12.5 % polyacrylamide gel was run on samples collected throughout the purification stages in order to check the tagged protein was attached to the beads. 20 µl PreScission Protease was added in 5 ml cleavage buffer. This was left overnight at 4 °C. The beads in solution were centrifuged at 2000 rpm for

two minutes. The supernatant containing the protein was pipetted into a separate tube and the beads were washed in cleavage buffer, collecting the supernatants for analysis.

4.2.8 Anion exchange chromatography

Buffer A: 1 L: 5.01 g 1-methyl-piperazine (volatile liquid needs a beaker with 100 ml water on a balance and added), 0.05 M Bis-Tris (10.46 g), 0.025 M Tris 3.03 g

Buffer B: 1 L: 0.1 M HCL (3.65 g)

Buffer C: 1 L: H_2O

Buffer D: 1 L: 2 M NaCl (116.4 g)

To determine the most effective condition for anion exchange chromatography a scout technique was employed using a GE Healthcare ÄKTA Explorer FPLC system. Using the basic buffers above the system will mix each in various combinations, automatically, to produce buffers of varying pH and salt composition. To start with protein in a low salt buffer is added to the column and sequentially the effect of changing the pH (6, 7, 8) on binding and elution, during a increasing salt gradient was tested. Chromatograms were obtained. Fractions obtained were analysed on 12.5 % polyacrylamide gels.

4.2.10 Dialysis

One 1.5 ml tube of protein in cleavage buffer was dialysed into the required buffer such as low salt buffer (50 mM Tris pH 7.5, 25 mM KCl, 5 mM $MgCl_2$), PBS or deionised water, using dialysis buttons or dialysis tubing for volumes over 1 ml. The buffer was changed twice and left overnight.

4.2.11 Protein solution concentration

The 14 ml of the M2-1-GST containing fraction was dialysed from elution buffer into cleavage buffer using dialysis tubing and clamps. The solution was changed to fresh cleavage buffer after 3 hours and again before leaving overnight. Still in the dialysis tubing, this was dried gently with paper towel and placed on a pile of Polyethylene glycol in a deep plastic tray and then covered with more PEG. This was left for half an hour. Damp PEG was replaced regularly and fresh PEG was added if necessary at half hour intervals. This reduced the volume to 3 ml which was pipette into two Eppendorf microcentrifuge tubes and a sample was run on a 12.5% polyacrylamide gel.

4.3 Results and Discussion

Before expression and purification of the protein, the M2-1-pGEX-6-P1 clone obtained from Dr R.P. Yeo was checked by restriction with the original restriction enzyme, *Sma*I, in order to establish the presence of the correctly sized M2-1 gene fragment, as in section 2.2.9, see Figure 4.1.

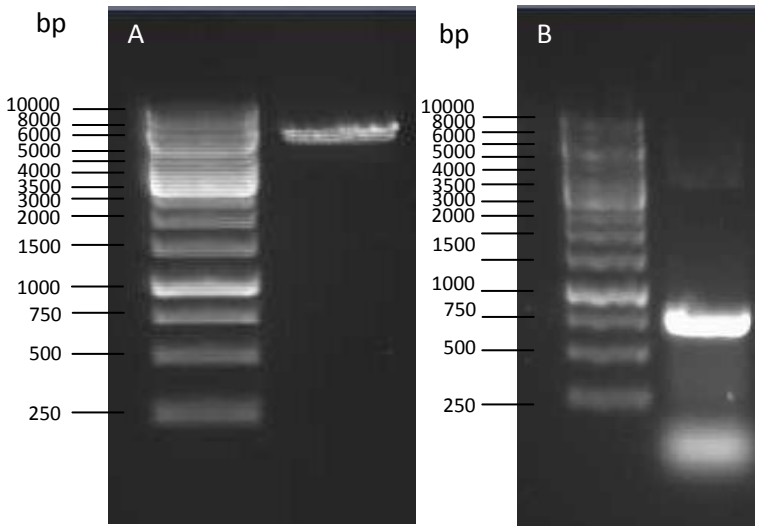


Figure 4.1: A) M2-1-pGEX-6-P1 before restriction
B) M2-1-pGEX after restriction with *Sma*1

Two approaches were used to optimise the conditions that would lead to the production of high quality protein suitable for high resolution structural studies such as X-ray crystallography. The first was to use a pre-packed GSTrap FF column and the second was to use glutathione sepharose beads in a batch-type preparation. The M2-1 protein ORF had previously been cloned into the pGEX-6-P1 vector in frame with the GST protein (M2-1-pGEX-6-P1). The GST and M2-1 moieties were separated by a linker containing a PreScission protease cleavage site, a highly specific protease based on the Rhinovirus 3C protease. In both approaches protein was loaded onto the column/beads, washed to remove non-specific binding proteins and this was followed by cleavage with the PreScission protease. This process released the M2-1 component leaving the GST tag on the solid phase. The sepharose bead method overall produced the purest protein, although a combination of the two techniques also produced relatively pure protein. This could be because the GST-protein attached more firmly to the column, needing a high concentration of glutathione to be eluted and cleavage was less efficient on the column than in solution- possibly due to reduced access to the cleavage site or aggregation of the protein on the column. Protease cleavage on the beads was simple and yeilded protein in the supernatant without GST (which was still

attached to the glutathione sepharose beads). It was also possible to put the beads on a carousel overnight during cleavage, in order to facilitate mixing and access to the protein, rather than allowing the beads to settle with consequent restricted access to the protease in the buffer. Another advantage of the beads was they were more suitable for using on large amounts of bacterial cell lysate, whereas in the column it was necessary to dilute the pellet material down with lysis buffer in order not to block the syringe filters.

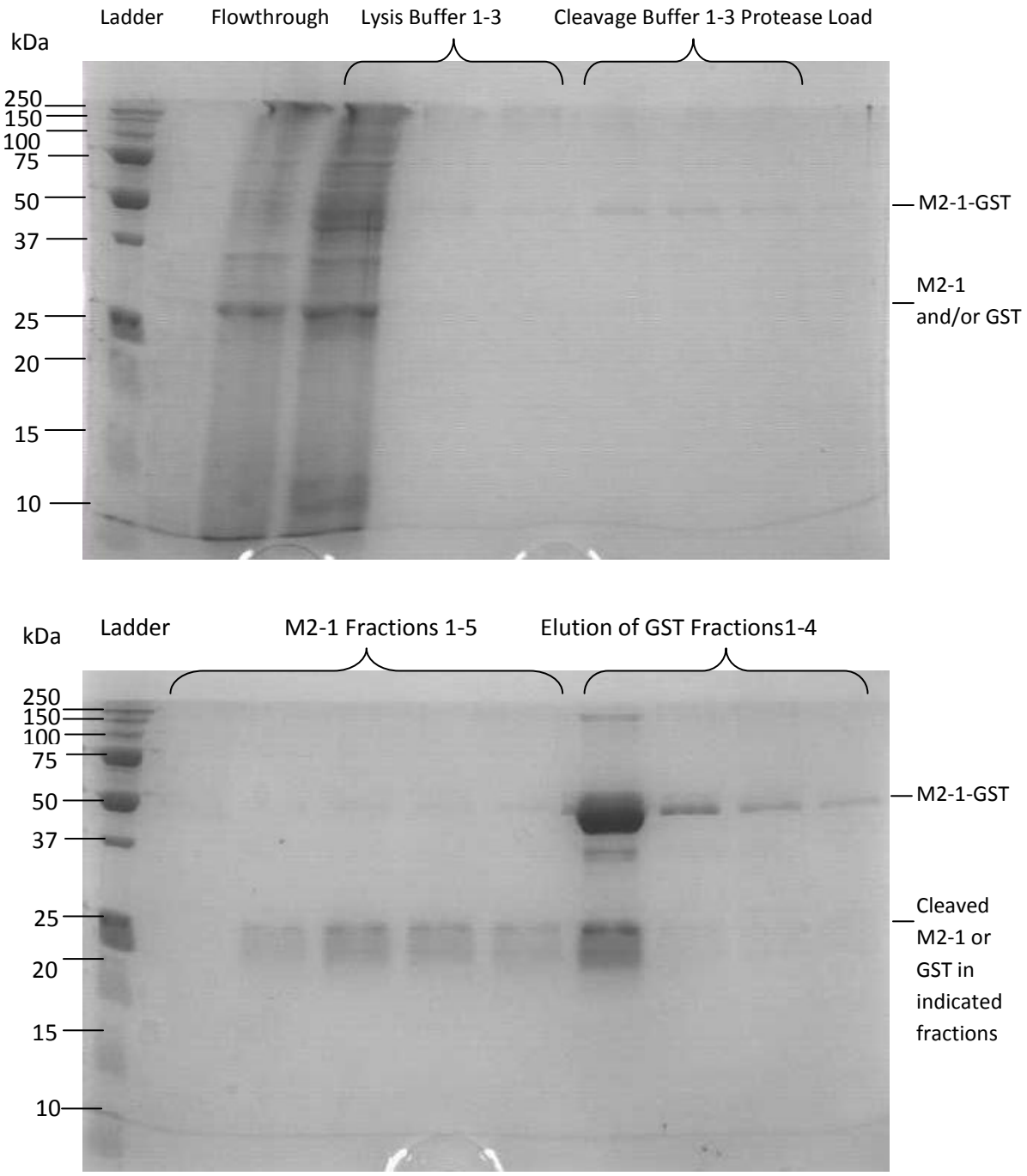


Figure 4.2: Samples taken from stages of purification of M2-1 using a GSTrap™

There was initial difficulty in obtaining pure protein that did not include amounts of GST-M2-1 uncleaved and in eluting the GST using the column, see Figure 4.2.

This was attributed to the amount of protease used and therefore it was increased in a subsequent purification. An alternative to the commercial PreScission Protease, 3C protease (a gift from Dr J. Flint, Institute for Cell and Molecular Biosciences, Newcastle University) was tried but had the same results. In addition, the culture growth time was extended from 3 hours 37 °C, 3 hours IPTG-induced 28 °C to 6 hours 37 °C, 16 hours (overnight) IPTG-induced 28°C. This gave much larger pellets containing potentially greater amounts of protein. At one stage, a great deal of protein was produced as half of the product obtained was re-cleaved with another 50 µl of protease and half was left alone- both fractions after 2 days displayed one band on an agarose gel indicating full cleavage had occurred, see Figure 4.3.

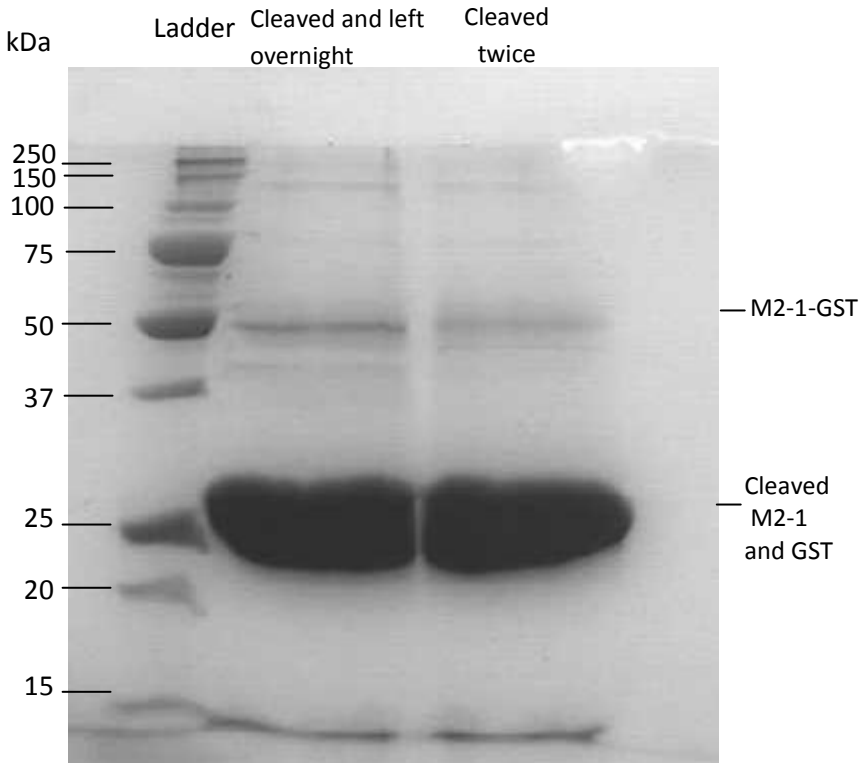


Figure 4.3: Fully cleaved M2-1 and GST

In an attempt to further purify the protein to ensure all GST was removed the GSTrap™ was used but with the following alterations to protocol:

3.00 ml protein in cleavage buffer was loaded and a 5.00 ml fraction was collected. 5.00 ml cleavage buffer was put through the column. 10.0 column volumes of elution buffer were put through. The column was washed as normal. A gel was run on the collected fractions- the higher bands on the previous gel were removed during

purification and a great deal of protein appeared to remain but the dimer GST-M2-1 was still present in the elution fraction (some could be viewed on the comparison of the cleaved and uncleaved so this was to be expected). See Figure 4.4.

Concentrations of each fraction were obtained using the NanoDrop spectrophotometer as follows: 1) 0.316 mg ml^{-1} 2) 0.320 mg ml^{-1} 3) 0.451 mg ml^{-1} 4) 0.427 mg ml^{-1} 5) 0.438 mg ml^{-1}. Four of the fractions underwent FPLC anion exchange chromatography purification and one was dialysed into a low salt buffer for analysis by LC electrospray mass spectrometry.

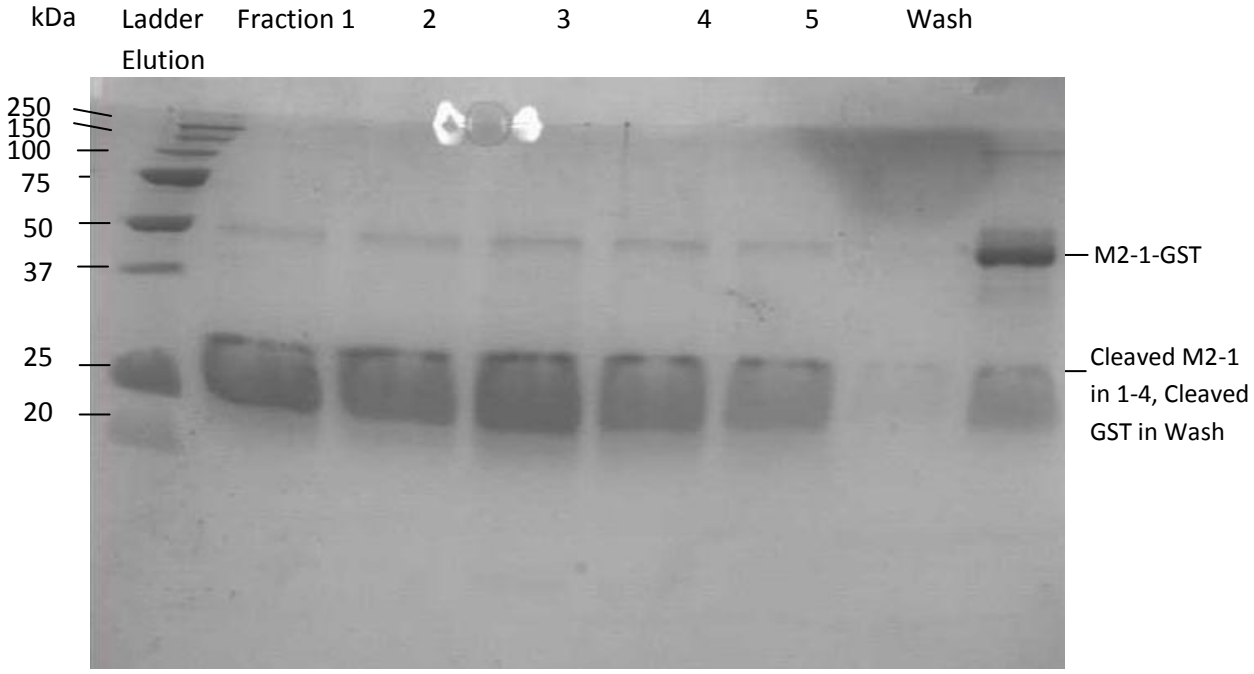


Figure 4.4: Second purification of the cleaved protein from remaining GST

However, purification by anion exchange chromatography was unsuccessful. The sample in LSB did not stick to the column and initial attempt with a small amount of the other four pooled fractions showed the protein did stick to the column and eluted at the highest salt concentration, favouring pH 8. The remainder of the samples were injected and run at pH 8; however the protein remained stuck to the column and could not be recovered.

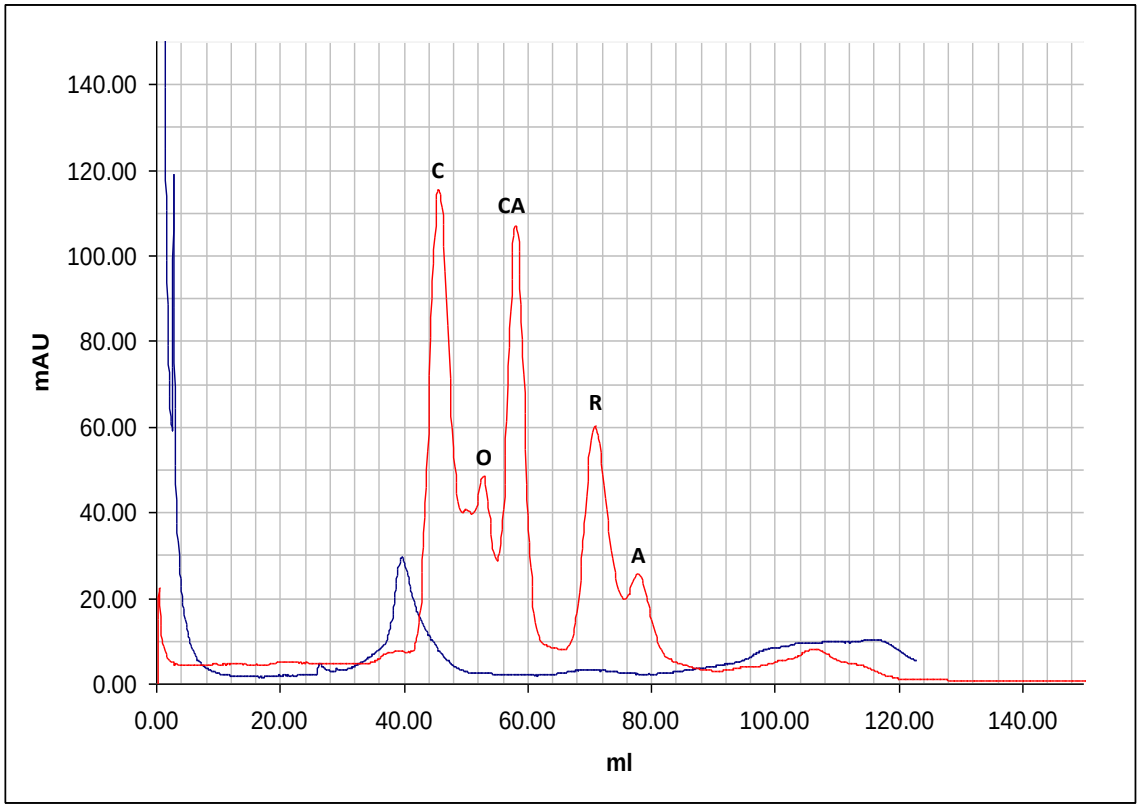


Figure 4.5: Gel filtration chromatography results using the HiLoad™ 16/60 Superdex Prep Grade gel filtration column. Blue: Chromatographic curve for retention times of protein solution. Red: Chromatographic separation of standard proteins. C: ConAlbumin 75kDa, O: Ovalbumin (43 kDa), CA: Carbonic anhydrase (29kDa), R: Ribonuclease A (14 kDA), A Aprotinin (6.5 kDa).

Purified M2-1 was analysed on a HiLoad™ 16/60 Superdex Prep Grade gel filtration column showing that it was present as a tetramer. This was shown by comparison of a calibration curve carried out with the low molecular weight (LMW) gel filtration calibration kit (Amersham Pharmacia Biotech Inc.), against a graph of retention times of the M2-1 protein solution, both shown in Figure 4.5. On the chromatogram for the M2-1 solution, it can be seen there is a peak at 40 minutes. The peak at 46 minutes on the protein standard chromatogram corresponds to ConAlbumin, a protein with a molecular mass of 75 kDa. As it eluted from the column at a time extremely close to but prior to that of ConAlbumin, M2-1 is present as a tetramer of around 89kDa (based on four monomers of 22.15 kDa). To completely confirm this is a tetramer and not larger, a protein standard higher than 90 kDa would be needed.

Finally, the purification protocol was done more quickly, by not waiting to run an acrylamide gel to check the protein was present before adding protease (as this was known to work). Instead, 50 µl 3C protease was added to the beads immediately and left for a longer period of time overnight. This produced immediately pure protein, with

no other dimer or GST-Protein bands remaining on the gel see Figure 4.6. Completely pure M2-1 protein was obtained and made suitable for initial crystallisation studies and

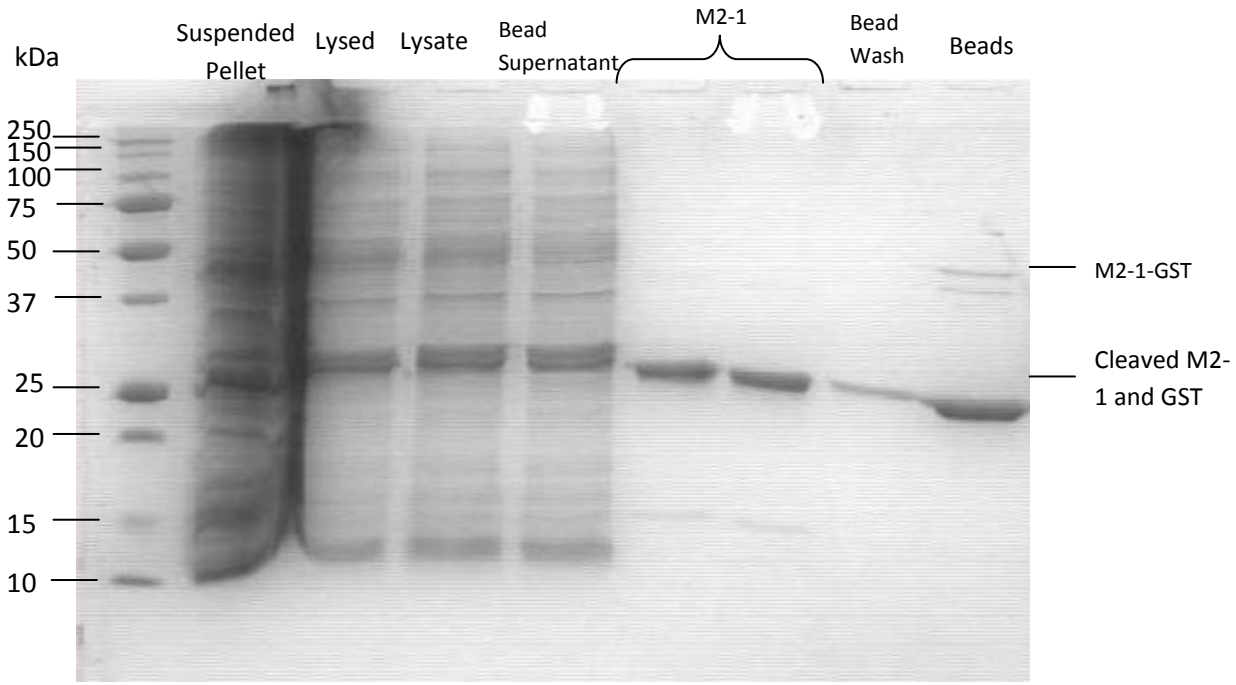


Figure 4.6: Optimised purification of M2-1 on glutathione-sepharose beads

structural analysis, specifically circular dichroism analysis.

It is worth noting whilst viewing the above gel photographs that the M2-1 protein is 22.15 kDa, 3C protease is 22 kDa and GST is 27 kDa. Not enough 3C protease was used for it to be visible on a gel. Therefore on some of the acrylamide gels the GST and M2-1 protein are very difficult to tell apart. It should also be noted that the disadvantage of using homemade 3C protease instead of commercial PreScission protease is that PreScission also contains a GST tag which would leave it attached to the GST Trap or to glutathione sepharose beads. Western blot analysis had to be utilised in order to verify the presence of M2-1.

Western blot analysis was carried out using α-M2-1 antibody. This produced three bands on the film, relating to monomer, dimer and tetramer of the protein, see Figure 4.7. It should be noted again that on an SDS page gel the dimer would appear at the same point as the uncleaved GST-Protein and therefore be difficult to identify. Western blotting is therefore necessary to determine whether GST or M2-1 is present, or both. A second Western blot was carried out on the same sample but this time testing for the presence of GST and repeated and both were negative.

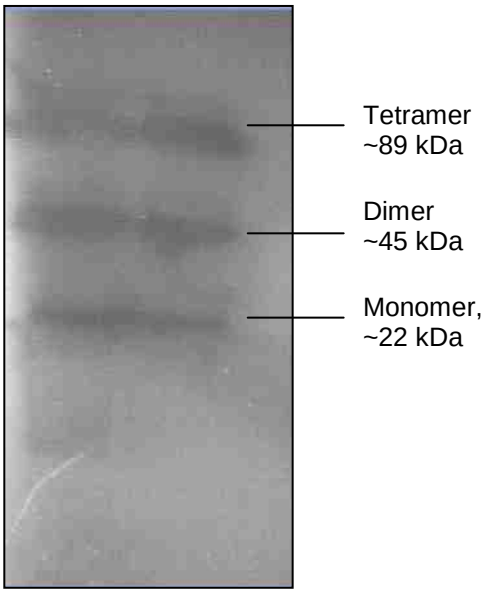


Figure 4.7: Photograph of X-ray film after detection by Chemiluminescence of a Western blot analysis nitrocellulose membrane

Chapter Five: Placement in Industry: Onyx Scientific and Peptide Synthesis

5.1 Introduction to Onyx Scientific

During this project a placement in a North East of England chemical company was undertaken. The partner company was Onyx Scientific, a medium sized medicinal chemistry business involved in the production of small molecules mainly for larger pharmaceutical companies. The small molecules produced at Onyx Scientific are usually produced there because they are difficult to synthesise, requiring first small scale research into the most suitable synthesis pathway. Within the company, the production is scaled up in stages until kilograms may be produced to a high level of purity.

5.2 Aims and Objectives

The aims of the time spent at Onyx Scientific, in relation to the research project, were to learn the principles and practice behind successful solution phase peptide synthesis, via the production of a dipeptide. This was with a view to planning and carrying out the synthesis of the full peptide. The dipeptide to be produced consisted of the first two amino acids in the sequence of the Fusion Peptide (amino acid sequence: FLGFLLGVGSAIASGVAVS). This is the hydrophobic peptide of the Fusion Protein that cannot be easily made and purified via recombinant DNA methods.

A second and equally important aim was to gain experience of working in an industrial (rather than an academic) setting and to learn about the chemicals industry and science in business rather than in academic research.

5.3 Introduction to peptide synthesis

5.3.1 The stepwise approach versus fragmentation condensation

Basically, there are two approaches to peptide synthesis. The first is to take a stepwise approach, adding on one amino acid at a time onto the peptide chain. The second method is to make short fragments of two or three amino acids via stepwise synthesis and then to condense them together forming a longer peptide. It is often the case with long peptides (>20 residues) that fragments are made using solid-phase synthesis techniques and then condensed together using solution-phase chemistry.

A tripeptide can be made in three ways: [1 + 2 + 3] versus [(1 + 2) + 3] or [1+ (2 + 3)]. Obviously the number of possible ways to make a peptide increases with its length. Therefore instead of arbitrarily assigning fragments to piece together or forging ahead with a stepwise approach and seeing how far one gets, there are a number of considerations to take into account. Firstly, that if the fragment condensation approach is taken there are advantages in terms of parallel synthesis- more than one of the same peptide but with subtle differences, such as in mutation studies, can be made in parallel. Likewise, if the strategy and reaction conditions are uncertain several could be tried on a small scale at the same time. Stepwise synthesis is, naturally, sequential and therefore slow, as each stage should be characterised. Also, bringing small fragments together to make a larger peptide means that isolation of the desired product will be easier at this stage, as there will be a larger difference between product and starting materials (compared to the starting material being only one amino acid shorter than the product).

A second consideration is that some residues on the end of peptide fragments will allow fragment condensation more easily than others, for example glycine on the end of a peptide fragment would mean no steric hindrance for another amino acid binding. The presence of cysteines means disulphide bond formation is likely and must be planned around- preferably leaving them until the last possible point to add to the peptide and protecting them (see 5.3.2). Met or Trp should similarly be added last if possible, as methionine's thioester side-chain and tryptophan's indole ring may be attacked by electrophiles in acidolysis.[69] Protection group choices (see 5.3.2) and limitation of racemisation (see 5.3.6) also make the matter more complex. The protection strategy used for making the fragments has to be compatible with the conditions used in condensation. Finally, it is always a good objective to aim for one final deprotection step, in which all of the side chain protecting groups are removed and, if solid phase synthesis is used, the peptide can also be cleaved from the resin cleanly.

5.3.2 Protection group strategies

There are a great many options to consider for developing the most appropriate protecting group strategy. The α-carboxy protecting group should be orthogonal to the α-amino protecting group used, meaning one protecting group can be removed from the molecule using a method that will not affect the second protecting. Orthogonality must also be considered when choosing side-chain protecting groups, if they are necessary. Ease of cleavage is an issue, as obviously it is preferable to use the mildest

conditions possible- a reason one might choose to use benzyl esters rather than ethyl esters, for example, as the latter requires strong cleavage conditions (see 5.3.2.3). Certain protection methods can even make racemisation less likely due to their mode of action.[69]

Because amino acids have carboxy groups that are much more nucleophilic than their amino groups, peptides can be synthesised in theory without any protection at all. However, usually a maximal protection method is used. This is because protecting groups can also offer additional advantages in terms of increasing the solubility of the peptide and facilitating its isolation and purification.

This introduction to protection strategies will focus on the synthesis of the Fusion Peptide and the advantages of the strategy used, as a broader discussion of protecting groups is beyond the scope of this project.

5.3.2.1 Fmoc α-amino protection

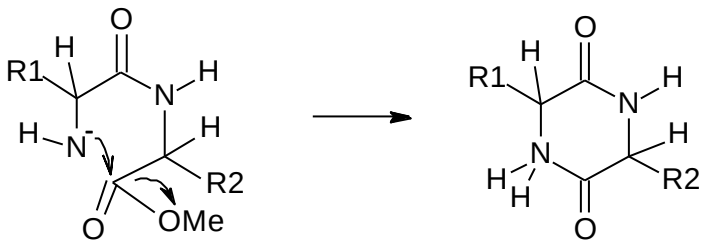
Fmoc (9-fluroenylmethoxycarbonyl) is stable to acidic conditions and is cleaved by bases, such as 20% piperidine in DMF. Using piperidine at room temperature to cleave Fmoc from the α-amino group of an amino acid on the end of a peptide chain takes a matter of seconds, though there are alternative cleavage reagents. The elimination reaction occurs by a $E1_{CB}$ (elimination conjugate base) mechanism via a dibenzocyclopentadienide intermediate anion. Dibenzofulvene produced as a result reacts with the piperidine giving a co-product in addition to the deprotected amino acid.[69] Fmoc deprotection wouldn't affect a Boc or a Z group on another amino acid as they are orthogonal groups (see 5.3.2.2). Fmoc can also be cleaved by catalytic hydrogenolysis.

5.3.2.2 Boc α-amino protection

In contrast to Fmoc, the Boc (Di-tert-butyl dicarbonate) group is stable in basic conditions and in the presence of nucleophiles but is very labile to acids. It is more labile to acids than an alternative protecting group, the Z (benzyloxycarbonyl) group, but is considerably more stable to base.[69] Therefore, Z and Boc are orthogonal, which is important as this means one can be cleaved without affecting the other at different stages of the synthesis. Boc is removed by treatment with acid, such as TFA or HCl in dioxane and ether flooding. In workup of reactions involving a Boc group, the use of strong acids in acidification should be avoided so citric acid is usually chosen as it is not extracted easily into the organic phase from the aqueous.

5.3.2.3 Ethyl ester α-carboxy protection

Ethyl (and methyl) esters are good protecting groups as they remain uncleaved by most coupling conditions and most deprotection conditions, and α-amino protecting groups can be removed selectively. However, one problem that can occur is with the use of piperidine to remove an Fmoc group. This is because dipeptide methyl or ethyl ester free bases can cyclise to DKPs (diketopiperazines) readily (Scheme 5.1), and this will occur under basic cleavage conditions used to remove the Fmoc protecting group.[69]



Scheme 5.1: DKP formation under basic conditions

The use of methyl esters is less favoured than using ethyl esters, as they can be difficult to remove when it is time to do so, needing strong conditions.

5.3.2.4 Benzyl ester α-carboxy protection

Preferably, the use of benzyl esters for protection should be considered instead of methy; or ethyl esters. Although dipeptide benzyl esters may still form DKPs under basic conditions, they are easier to remove than ethyl or methyl esters by saponification (forming an alcohol plus carboxylic acid salt under basic conditions) or hydrazinolysis.[69] Benzyl esters cannot be cleaved with acids like TFA and therefore are orthogonal with Boc α-amino protection.

5.3.3 Activation and Coupling

In this synthesis, the coupling of two amino acids to make a dipeptide was undertaken by the activation of the ester group formed as the α-carboxy protecting group on phenylalanine. A phenol ester makes for a better leaving group than an ethyl ester, and this is important for the aminolysis of the ester, which occurs via collapse of a tetrahedral intermediate. Active esters are normally synthesised using DCCI or HBTU coupling reagents (see below). Side reactions and racemisation are not a significant

problem as active esters are relatively inactive before the intermediate is produced. However, if they are made directly and the carbonyl component is vulnerable then racemisation is likely, usually occurring if a base is present (see section 5.3.6). More reactive active esters are used more in solid phase synthesis. Active esters give excellent results during salt couplings (as in this dipeptide synthesis).[69]

The most popular coupling reagent is dicyclohexylcarbodiimide (DCCI/DCC) for activating carboxy groups, such as making active esters and can be used as a direct coupler. The activating event is the formation of O-acylisourea, a strong acylating agent formed by addition of the amino acid carboxy group to the carbodiimide group, and this leads immediately to the peptide by aminolysis. However, the reaction via this carbodiimide group often results in partial racemisation. Direct coupling using DCCI needs only an organic solvent and equimolar amounts of each component.[69]

Another class of coupling reagents are the phosphonium and “uronium” salts, such as the two used in the dipeptide coupling of this project- PyBop and HBTU. HBTU, O-(Benzotriazol-1-yl)-N,N,N’,N’-tetramethyluronium hexafluorophosphate, does not actually contain a uronium component as first thought, despite its name, but in fact has a guanidinium structure.[70] The use of HBTU reduces racemisation to a minimum, and reactions using this coupling agent are very fast, under one hour.[69] PyBOP, a phosphonium coupling reagent, is an analogue of the coupling reagent BOP (benzotriazolyloxytris-(dimethylamino)phosphonium hexafluorophosphate). BOP became almost as popular as DCCI despite the fact that using it runs the risk of severe racemisation occurring during the coupling. In this respect, HBTU is clearly superior. The advantage of PyBop over Bop is that it minimises the production of a carcinogenic side-product, hexamethylphosphoramide.[69] There are now more analogues of PyBop available that have this advantage but that produce higher yields of coupled product, and likewise there are analogues of HBTU although these have not surpassed the original in yield or racemisation minimisation.[70]

5.3.4 Solution phase synthesis versus solid phase synthesis

Synthetic peptides were first produced using solution phase synthesis. The first fully biologically active peptide to be synthesised chemically using solution phase techniques was oxytocin (uterine contracting pituitary hormone). This was achieved by Vincent de Vigneaud in 1954; this beginning of modern peptide synthesis was recognised by his Nobel Prize the following year.[71] Examples such as oxytocin peptide hormone and also porcine gastrin 1 are of particular significance to the field in terms of

development of protection strategies and fragment condensation strategies. The production of human insulin achieved the synthetic formation of a peptide consisting of two chains held together by disulphide bonds.

Solid phase peptide synthesis (SPPS) was developed by Bruce Merrifield in the 1960s and is now the focus of much peptide synthesis research. The initial test of the method was done by synthesising a tetrapeptide in 1963.[72] Within a year, the technique had been finely tuned for the synthesis of the nonapeptide bradykinin.[73, 74] The "solid" part of the technique's title refers to the resin on which the peptide is made, which consists of swollen gels made by solvent and solute molecules contained inside a polymeric matrix.[72] Solid phase synthesis methods are much quicker and cleaner than solution phase methods, and the length of peptides being produced since the 1960s has greatly increased due to the development of this method. It is worth noting that the most successful solution phase peptide chemist, de Vigneaud, had no trouble recognising the benefits of SPPS, and indeed his laboratory adopted Merrifield's new technique to synthesise vasopressin and deamino-oxytocin, which they famously co-authored a paper on in 1968.[71]

There are two main SPPS methods - the original Merrifield technique or the Sheppard technique. Merrifield's original concept has been finely tuned for modern synthesis. Polystyrene cross-linked by divinylbenzene is chloromethylated and the target peptide C-terminal amino acid residue (amino-protected with a Boc group) is attached to the matrix by nucleophilic displacement. Deprotection of the Boc group with mild acid allows coupling of the next amino acid by DCCI. This is repeated until the desired peptide is formed.[71] Side chains are protected where necessary with groups that are not labile to the mild acid needed for Boc deprotection. Washing occurs regularly to remove co-products and excess reagents (the peptide-polymer matrix is insoluble). Strong acid (HF) at the final step removes all of the protecting groups and the peptide from the resin.

Sheppard's argument in the early 1970s was that if the polymeric matrix was chemically similar to the peptide being synthesised then both could be used in an open gel system in dipolar aprotic solvents, with free access to reactive sites.[75] In Merrifield's synthesis, the solvent used to swell the matrix (dichloromethane) cannot also solvate the peptide chain and so this encourages aggregation and, as a result, reactive site access is lower.[75] Therefore polystyrenes were often replaced by polyamide carriers. Sheppard synthesis is usually carried out using Fmoc α-amino protection instead of Boc and the final deprotection can be under much milder conditions due to the strategy

of α-Fmoc combined with ω-mild acid labile groups. Merrifield α-Boc with ω-strong acid labile tactics are also not completely orthogonal as with the Sheppard protection.[69]

The only problems with solid phase synthesis are firstly, that certain peptides form aggregates around the resin, therefore sterically hindering any further synthesis. Secondly, the level of reaction completeness needed at each stage to make a long peptide at high yield was simply too high. The end product can contain truncated sequences (if a residue remains unreacted) or deletion mutants (if an unreacted residue is coupled in the next coupling reaction). For example, the Fusion peptide requires 19 amino acids, so 18 coupling reactions. If each is 99 % complete and no truncated sequences occur then the end product will be comprised of approximately 19 % impurities reflecting the 19 possible deletion sequences. However, it is possible to push completion to 100 % with an excess of acylating reagent and doing each coupling more than once.

Modern technology such as the use of HPLC for separation and microwaves to speed up synthesis (not only by temperature but also by electromagnetic agitation of the chain to prevent aggregation), along with automation of the solid-phase process, has greatly increased the speed and yields of peptide synthesis compared to solution phase techniques, and also made it much easier for the non-chemist to synthesise a peptide needed for biological characterisation. This is the reason why it is advisable to carry out solution phase synthesis first, at least in part, if the chemistry is to be understood. It is most useful to be able to programme a computer to carry out the synthesis but this can be to the detriment of knowledge and understanding.

5.3.5 Considerations for individual residues

In the case of this particular dipeptide synthesis only phenylalanine and leucine need be considered in detail. These amino acids fortunately have limited side-chain functionality and therefore do not need to be protected. The only problem might come if, to cleave a protecting group, catalytic hydrogenolysis was used as phenylalanine could reduce to cyclohexylalanine.[69] There are no such issues with leucine

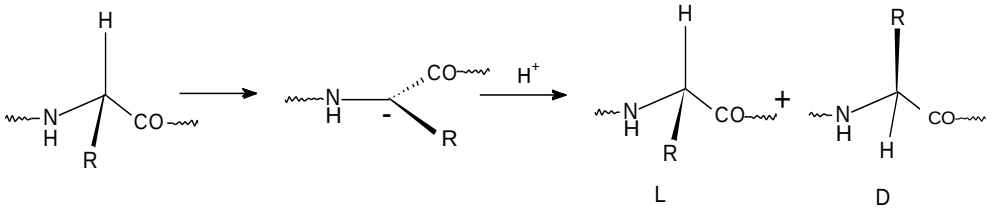
Other residues can cause a great deal of difficultly and their side chains will need to be protected during synthesis. The other residues of the complete Fusion Peptide are unlikely to cause difficulty but some complications are possible. Unhindered access to glycine residues in a peptide allows side-reactions to occur-if a peptide had a relatively high content of glycine then it may be insoluble and difficult to purify. The hydroxyl group of serine is able to react with acylating agents so would usually be protected (for

example using benzyl ether subsequently cleaved by strong acidolysis) but if mild carboxy activation is used it may be left unprotected.[69]

5.3.6 Consideration and control of racemisation

For peptide synthesis this involves the change from one enantiomer to a mixture of enantiomers or partial epimerisation (where a mixture of diastereomers that differ at one chiral centre forms). Often, the desired product is only one diastereomer or enantiomer (especially in medicinal chemistry), which makes purification a problem and analysis slightly trickier if a racemic mixture has been formed.

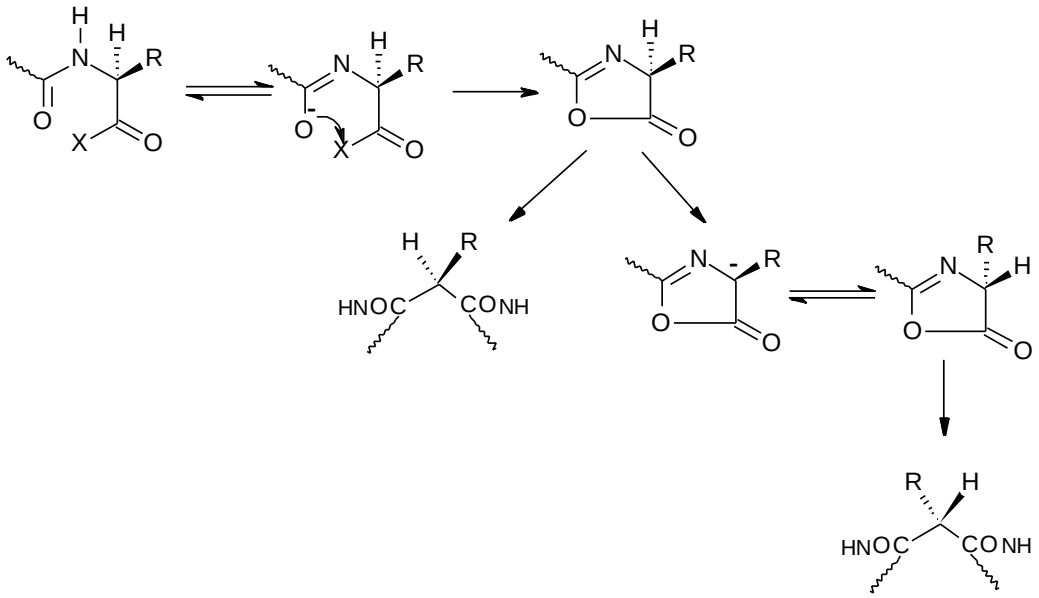
There are two main mechanisms of racemisation that occur during peptide synthesis. The first is direct enolisation (Scheme 5.2),[69] where the α-carbon is deprotonated and the carbanion can then protonate on either side:



Scheme 5.2: Direct enolisation

The rate is dependent upon the base strength, the nature of the solvent, and the electron withdrawing effect of groups at the chiral centre. Racemisation is faster when strong electron withdrawing groups are present and with sterically unhindered strong bases in dipolar aprotic solvents, such as DMSO.[69] However, the rate alone does not determine the amount of racemisation that occurs. It is a balancing act between the rate of enolisation and the rate of coupling and, usually, this mechanism only causes significant racemisation when coupling is slow.

The second mechanism of racemisation is the formation of oxalones (Scheme 5.3). These form due to activated acylamino acids cyclising in basic conditions. Oxalones are prone to aminolysis, reacting with amino acids inducing peptide bond formation. But they also racemise via stabilised anions and this is unfortunately usually at a faster rate than coupling can occur.[69]



Scheme 5.3: Oxalone formation and racemisation

This fast rate can be slowed if the protecting group at the amino group is an alkoxycarbonyl such as Boc or Fmoc (Figure 5.1), as these oxalones are quicker to aminolyse and racemise compared to acylamino acid oxalones.[69]

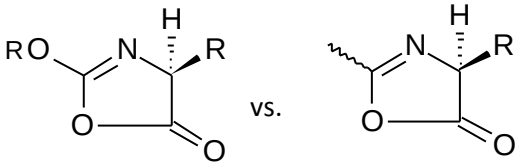


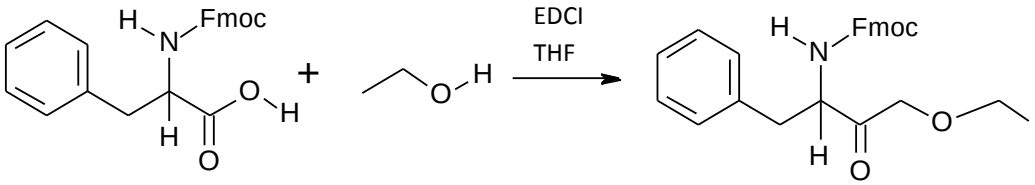
Figure 5.1: An oxalone with an alkoxycarbonyl group versus an oxalone with an alkyl group

This means that Boc and Fmoc protection strategies have a huge advantage as activation and coupling of Boc or Fmoc protected amino acids are not affected by racemisation in this way.

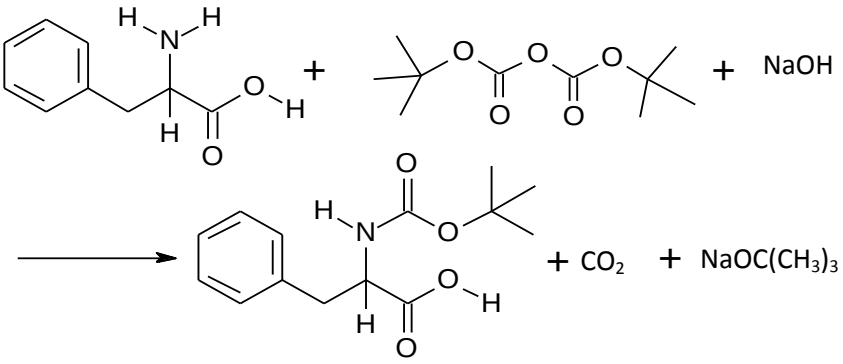
5.4 Synthesis of the dipeptide OBzl-Phe-Leu-Fmoc

As the RSV F protein Fusion peptide contains no troubling residues, the strategy decided was to use Fmoc α-amino-protected amino acids in a stepwise synthesis. The length of the peptide is the only hurdle to an efficient synthesis and therefore reasons as to the decision to use solid phase synthesis to make the entire peptide, but solution phase chemistry to synthesis the first dipeptide, will be described in section 5.3.4 below.

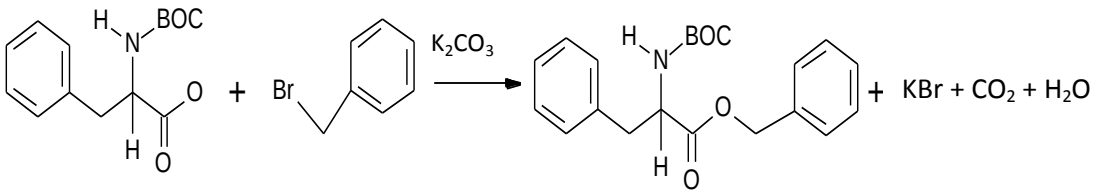
At all stages, DL amino acids were used (as an enantiomerically pure product was not necessary at this preliminary stage) and therefore a racemic mixture of the dipeptide would be produced as the end product. Therefore the stereoscopic configurations of the molecules are not noted in the reaction schemes.



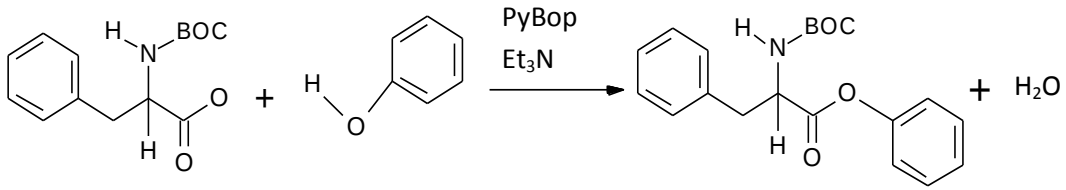
Scheme 5.4: α-carboxy protection of Fmoc-Phenylalanine



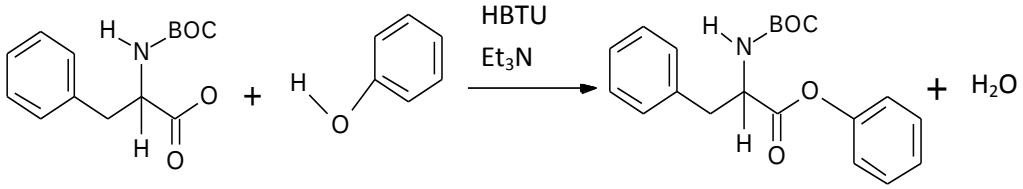
Scheme 5.5: α-amino Boc protection of Phenylalanine



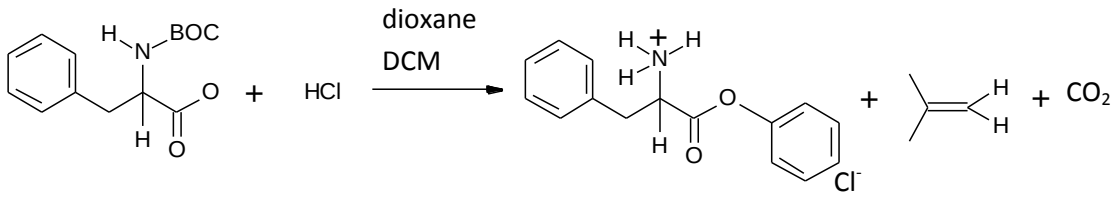
Scheme 5.6: α-carboxy protection of Boc-Phe by formation of the Benzyl ester using benzyl bromide



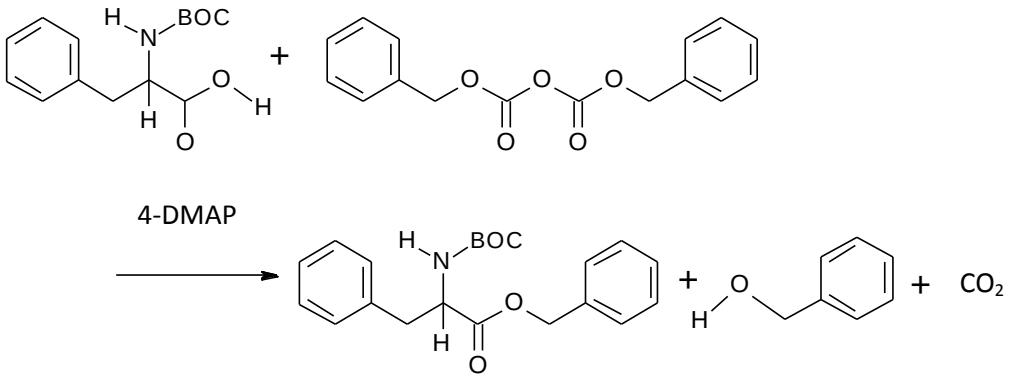
Scheme 5.7: α-carboxy protection of Boc-Phe by formation of the Benzyl ester using Phenol and PyBop



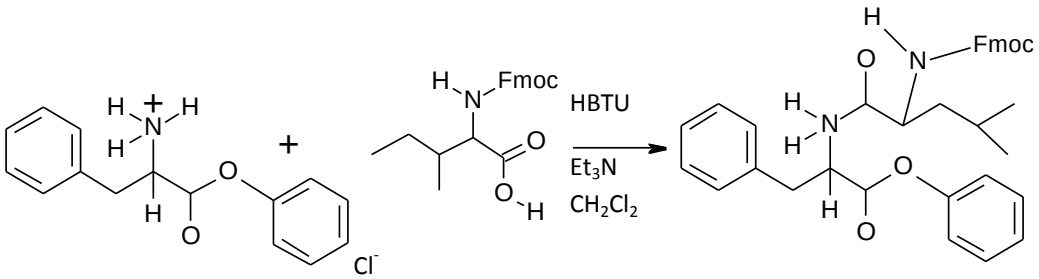
Scheme 5.8: α-carboxy protection of Boc-Phe by formation of the Benzyl ester using Phenol and HBTU



Scheme 5.9: Deprotection- removal of the Boc group from Boc-Phe-OBzl on a small scale



Scheme 5.10: α-carboxy protection of Boc-Phe by formation of the Benzyl ester using dibenzyl dicarbonate



Scheme 5.11: Coupling of Phe-OBzl and Fmoc-Leucine

5.5 Materials and Methods

All reagents were obtained from Fischer Scientific.

5.5.1 Experimental data analysis

Analysis of reaction progress was carried out using LC-MS (liquid chromatography mass spectrometry), HPLC (high performance liquid chromatography) in MeOH/H_2O, NMR and TLC (MeOH in DCM). NMR was carried out by the Eclipse 270 spectrometer. Chemical shift data is reported in ppm (parts per million) with the protons of the solvent, $CDCl_3$, as an internal reference at 7.25 ppm. *J*-coupling values are given in Hz. Mass spectra were recorded on the Esquire 3000 Plus mass spectrometer. HPLC was recorded on the 4.6 x 50 mm Zorbax XDB-C18 column.

5.5.2 Experimental Method: α-carboxy protection of Fmoc-Phe by formation of the ethyl ester

100 mg Fmoc-Phe was dissolved in 3.00 ml 100 % ethanol and 1.00 ml THF. 49.8 mg EDCI coupling reagent was added and the reaction was stirred to completion with slowly bubbling nitrogen. The reaction mixture was concentrated to 3.00 ml *in vacuo* then 50.0 ml H_2O was added and the solution was extracted with 100 ml ethyl acetate. 50.0 ml citric acid in water was added and then the extract was washed with 50.0 ml saturated brine solution. The solution was dried with $MgSO_4$ and filtered then an attempt was made to obtain the product *in vacuo*. However, because the citric acid was added after the first separation the product moved into the aqueous layer. Therefore the combined aqueous layers were extracted with 3 x 50.0 ml ethyl acetate. The solution was washed with brine, dried over $MgSO_4$ and the solvent evaporated off *in vacuo*. 50.0 ml ethyl acetate was added to the sticky, glassy solid obtained and the solution was washed with 20.0 ml saturate potassium carbonate, and then washed with 20.0 ml brine. This was again dried over $MgSO_4$ and filtered then the product was obtained *in vacuo*.

5.5.3 Experimental Method: α-amino Boc protection of phenylalanine

2.00 g DL-Phe was dissolved in 23.8 ml 1, 4-dioxane. 12.1 ml 1N NaOH and 12.1 ml H_2O were added under ice cooling and the temperature was monitored to ensure no excessive temperature rise occurred. 3.08 ml di-tert-butyl dicarbonate was added. This was allowed to stir for about 20 minutes and then the ice bath was removed and the reaction allowed to reach room temperature. The reaction was stirred to completion

(approximately 21 hours) then concentrated on a rotary evaporator to an oil of 5.0 ml. 20.0 ml ethyl acetate and 20.0 ml heptane were added and evaporation continued. A white powder was obtained. [^{1}H NMR (269.72 MHz; $CDCl_3$) $\boldsymbol{\delta_H}$ 7.34-7.28 (m, aromatic, 3H), 7.25-7.22 (m), 7.17 (d, *J* = 8.9, aromatic 2H), 5.99 (s, OH, 1H), 4.91 (d, *J* = 8.1), 4.62 (d, *J* = 8.1), 4.39 (s, CH, 1H), 3.15 (qd J = 5.4, 10.8, 5.4, 10.8, CH_2, 2H), 1.43 (s, $(CH_3)_3$, 9H), 1.30 (s), 1.17 (s)]. See Appendix C1.

5.5.4 Experimental Methods: α-carboxy protection of Boc-Phe

5.5.4.1 Formation of the benzyl ester using benzyl bromide

0.800 g Boc-Phe and 2.09 g potassium carbonate were dissolved in 90.0 ml acetone. The reaction was stirred and heated using an oil bath with temperature monitoring and bubbling nitrogen. 0.700 ml benzyl bromide was added. The solution was heated to 61 °C and stirred to completion, for approximately 20 hours. This reaction began as a heterogeneous solution but progressed overnight to a homogenous saturated solution. Another equivalent of benzyl bromide and another equivalent of potassium carbonate were added to force the reaction to completion (the LC chromatogram indicated a small amount of product was being formed). After an additional 48 hours, the reaction solution was Buchner filtered until no potassium carbonate remained. The solvent was taken off *in vacuo* using a PTFE pump due to the HBr co-product.

5.5.4.2 Formation of the benzyl ester using phenol and PyBop

4.71 g PyBop was dissolved in 30.0 ml DCM and 2.31ml triethylamine. 2.00 g Boc-Phe and 0.851 g phenol were added and the amber solution was left to stir with nitrogen bubbling. 40.0 ml H_2O was added the solution was extracted then the organic layer washed with 50.0 ml brine, dried with $MgSO_4$ and filtered. The solvent was evaporated off under vacuum producing amber oil. This was purified via column chromatography, with solid phase silica and mobile phase 100 % DCM. The product came off in the first ten fractions. The column was flushed with 200 ml DCM, then 5 % MeOH in DCM, followed by a 1:1 MeOH:DCM flush and finally 100 % MeOH.

[^{1}H NMR (269.72 MHz; $CDCl_3$) $\boldsymbol{\delta_H}$ 7.38-7.28 (m, aromatic, 8H), 6.98 (d, *J* = 8.1, aromatic, 2H), 5.05 (s), 4.80 (d, *J* = 8.1, NH, 1H), 3.22 (d, *J* = 5.4, CH_2, 2H), 2.16 (s), 1.55 (s, $(CH_3)_3$, 10H), 1.43 (s), 0.05 (s)] **LCMS** m/z 364.2 $[M + Na]^+$, 242.0$[M - Boc]^+$, 705.4 $[2M + Na + H]^+$. See Appendix C2.

5.5.4.3 Formation of the benzyl ester using Phenol and HBTU

3.43 g HBTU was dissolved in 30.0 ml DCM and 2.31 ml triethylamine. 2.00 g Boc-Phe and 0.851 g Phenol were added and the solution was left to stir with nitrogen bubbling. The solution was extracted with 50 ml H_2O followed by 50.0 ml potassium carbonate and a brine wash. The solvent was evaporated off *in vacuo* and a white solid was precipitated. The product was purified via column chromatography (solid phase silica and mobile phase DCM), and then elution was with 50.0 ml DCM, 50.0 ml 1 % MeOH then 50.0 ml 2 % MeOH. Fractions 11-19 were combined and concentrated *in vacuo* to give 1.52 g white solid product. [^{1}H NMR (269.72 MHz; $CDCl_3$) $\boldsymbol{\delta_H}$ 7.39-7.28 (m), 7.19 (s, 8H aromatic), 7.00 (d, *J* = 8.1 aromatic, 2H), 6.97 (s), 5.29 (s), 5.11 (d, *J* = 8.1), 4.82 (q, *J* = 5.4), 3.23 (d, *J* = 8.1, CH_2, 2H), 1.56 (s), 1.45 (s, $(CH_3)_3$ 9H), 0.06 (s)]. See Appendix C3. **LCMS** 241.9 $[M - Boc + H]^+$, 363.9 $[M + Na + H]^+$, 705.2 $[2M + Na + H]^+$. For LC see Appendix C4.

5.5.4.4 Formation of the benzyl ester using dibenzyl dicarbonate

1.00 g Boc-Phe was dissolved in 15.0 ml THF. 0.0460 g DMAP was added carefully. Finally, 1.08 g dibenzyl dicarbonate was added. The reaction was left to stir to completion under bubbling nitrogen overnight. The solution was extracted with 10.0 ml ethyl acetate and 10.0 ml saturated potassium carbonate solution, then ethyl acetate again. The solution was dried with $MgSO_4$ and filtered before concentrating *in vacuo*.

5.5.5 Deprotection: Removal of the Boc group from Boc-Phe-OBzl

100 mg Boc-Phe-OBzl was dissolved in 0.340 ml DCM in a small reaction tube. 0.230 ml 4N HCl in 1, 4-dioxane was added. The reaction was left to stir under bubbling nitrogen over the afternoon and overnight. 0.0200 ml 4N HCl in 1, 4-dioxane was added to force the reaction to completion and the reaction was left to stir to completion. 2.00 ml diethyl ether was added to precipitate the product and this was filtered under vacuum. 20.2 mg white powder product was obtained. The product was analysed by LC, see Appendix C5.

5.5.6 Coupling of Phe-OBzl and Fmoc-Leu

20.0 mg Phe-OBzl and 0.0380 g HBTU were dissolved in a solution of 0.300 ml DCM and 0.0200 ml triethylamine. 0.0350 g Fmoc-Leu was then added and the reaction was stirred to completion for 4 hours. A phase separator was used to wash the extract the solution with potassium carbonate and 10.0 ml DCM followed by a wash with brine. The solution was dried over $MgSO_4$, filtered and concentrated to 10.0 mg oil *in vacuo*.

[^{1}H NMR (269.72 MHz; $CDCl_3$) δ_H 5.30 (s), 3.48 (s), 2.79 (s), 0.06 (s)] **LCMS** m/z 577.1 $[M^+ + H]^+$, 599.1 $[M^+ + Na + H]^+$. See Appendix C6.

5.6 Results and Discussion

5.6.1 Synthesis of Fmoc-Phe-OEthyl

TLC analysis indicated that the product obtained was either very impure or that the EDCI had altered the Fmoc group. It was decided to continue the synthesis using a Boc protection strategy, as the cleavage of the Boc protecting group is orthogonal to the cleavage of an ester. If Fmoc amino acids continued to be used, it would be necessary to carboxy protect using ^{t}Bu, however the reagents were unavailable.

5.6.2 Synthesis of Boc-Phe

Comparison of an NMR sample of product with an NMR taken from a commercial sample of Boc-Phe confirmed the product as Boc-Phe as the spectra are identical.

5.6.3 Synthesis of Boc-Phe-OBzl

Comparing the four separate routes undertaken to synthesise Boc-Phe-OBzl, the purest product with good yield that was obtained was via the route using the coupling reagent HBTU with phenol. The NMR spectrum of the product obtained is a clear fit, and the liquid chromatography result confirms the purity of the product with only one peak at 5.08 minutes. It was concluded that the product should appear at 5.08 minutes from the information provided by the LCMS of experiment 5.5.4.2 (the same synthesis except PyBop was used as the coupling reagent). This LCMS indicated a significant peak at 242.00, corresponding to the Boc group having been removed from the product. The removal of the Boc group was caused by the spectrometry and therefore it was safe to conclude that a small amount of product was present from reaction 5.5.5.2. This indication that the peak at 5.08 was the desired product lead to its purification in reaction 5.5.5.3. The reaction using PyBop only produced a small amount of product and additional impurities, therefore was not as successful in terms of yield of product or as clean as the reaction using HBTU, but was useful in terms of this information. The formation of Boc-Phe-OBzl was most successful using HBTU as in Scheme 5.8. LC chromatograms for both reactions showed a highly polar peak at 1.6 minutes, which was likely to have been the ester of the coupling reagent used. This was not washed out successfully in reaction 5.5.5.2 as no potassium carbonate wash

was done but this was corrected for work up of reaction 5.5.5.3. 1.52 g product was obtained from 2.00 g of starting material, giving a reasonable 59.1% yield.

The first method attempted, using benzyl bromide, was slow to progress and formed a large proportion of a very polar product, likely to involve the acetone used. The problem with using acetone as a solvent is that it is also able to react. This was successfully removed during evaporation under vacuum leaving two peaks on the LC at 4.7 and 5.2 minutes, neither being starting material and the second peak only appeared on the chromatograms on the sixth analysis. At this point it was obvious from the LCMS from 5.5.5.2 that neither peak represented the desired product, and the yield was poor.

The last method attempted was by using dibenzyl dicarbonate, reaction 5.5.5.4. LC confirmed a substantial amount of dibenzyl dicarbonate and starting material remained unreacted, however a peak at 5.106 minutes indicated some product had formed but the reaction did not go any further when left overnight and therefore workup was completed to attempt to recover the starting material.

5.6.4 Deprotection of Boc-Phe-OBzl

This reaction fully completed, with no starting material remaining. Three peaks were visible on the chromatogram before purification. The isolated product was confirmed to be pure by LC.

5.6.5 Coupling of Phe-OBzl to Fmoc-Leu

The coupling of the first two amino acids of the Fusion peptide produced a number of products and would have been difficult to purify. In the first instance, it was expected that a mixture of diastereomers would be produced, as the protected phenylalanine originated from DL-phenylalanine but the leucine was purely in L configuration from a commercial source. However, on the LC chromatogram, where a pair of peaks was expected there were two pairs of peaks. To understand this result, after purification, an LCMS was taken. The NMR showed little apart from the presence of a few aromatic hydrogen atoms. The LCMS clearly confirmed the presence of the desired product with the $[M^+]^+$ and $[M^+ + Na]^+$ peaks shown, see Appendix C6.

5.7 Conclusions

Overall, the formation of the dipeptide was a success with clean reactions on the pathway to the product, despite there not being a sufficient quantity of product to continue on with to make the full 19 amino acid long fusion peptide. The main aim was

to carry out reactions with a view to learning the main principles and practice of solution phase peptide synthesis, and this was largely achieved. The numerous methods that were attempted to α-carboxy protect Boc-Phe with the benzyl ester group gave a clear illustration of how many options there are in peptide synthesis, and this was simply to add on one protecting group. The options multiply when one considers other protecting groups and alternative protecting group strategies, even for one case of a specific peptide. It was always clear that attempting to make the full peptide by solution phase chemistry would take months, and that solid phase synthesis would be the obvious method of choice, unfortunately not available within Onyx. Other principles, that did not need to be taken into account during the specific reactions encountered, such as control of racemisation and residue side-chain protection were considered in theory regarding the reactions being undertaken. Residue side-chain protection was not an issue due to the amino acids required. In fact, very few problems would occur in this regard during the remainder of the synthesis, as none of the residues involved would be problematic.

5.8 Further Work

The obvious next step would be synthesis of the Fusion Peptide using stepwise SPPS methods. This will be carried out on a resin using a microwave synthesiser within Durham University, which will greatly speed up the rate of protection, coupling and deprotection to a few days to complete the peptide. Synthesis from 3’ to 5’ would be advantageous as useful experimental labels could be readily added to phenylalanine such as a fluorescent label, in order to view the peptide’s interaction with a synthesised membrane and/or with cell membranes. Methods of characterisation will be undertaken; including mass spectrometry to check the mass of the peptide is correct and circular dichroism to assess the level of secondary structure. It is predicted from the sequence that this hydrophobic peptide will form an α-helix. This could create problems during solid phase synthesis as the helix will begin to form at a certain point (it is likely this will occur at the LeuLeu section, if synthesised from 3’ to 5’). At this point, half of the partially made peptide will be taken forward whilst it is analysed by mass spectrometry to assess whether steric hindrance is affecting the process. In fact, at three intervals during the synthesis the peptide will need to be checked by mass spectrometry.

Residue specific considerations are only an issue for the serine residues in the sequence, whose side chains may need to be protected in order to prevent them forming disulphide bonds.

In contrast to the solution peptide chemistry performed above, where only the principles of peptide chemistry were the concern, in the synthesis of the full peptide maintaining stereochemical integrity will be an important issue. The two mechanisms of racemisation are outlined above in section 5.3.6. If there is 1 % conversion from L to D configuration at each coupling stage for the Fusion peptide then this would mean 19 % of the products would be D configured. HPLC may be used to check for D and L diastereomers (they will appear as pairs of peaks, as in the coupling carried out to make the dipeptide above).

Chapter Six: Overall Discussion of Results

M2-1 protein was made ready for crystal trials. This was indicated by gel filtration chromatography analysis to be in the form of a tetramer, but this analysis should be repeated to include a protein standard over 90 kDa. This was sent for crystal trials by the Durham University in-house service. Optimisation of the purification protocol was a success despite difficulties in cleaving the protein from GST and in purifying M2-1 by FPLC anion exchange chromatography. As during the last attempt at crystallisation there was difficulty finding the right conditions for crystal growth, a scale up of the expression and purification is recommended to give more protein to work with, which could easily be achieved with a larger volume of glutathione sepharose beads and a higher concentration of salt (1 M up from 150 mm in the cleavage buffer) may give a higher yield of protein.[3] Circular dichroism has been carried out previously to obtain information on the secondary structure of M2-1 in solution and the data was published in July 2009.[3] It would be useful to obtain data on the protein obtained from this project for comparison and in order to repeat the experiment, expecting to confirm the results. This previous study found the far-UV CD spectrum was similar to other highly structured proteins with estimations of secondary structure obtained as 38 % α-helices, 12 % β-sheets, 40 % random coils and 10 % turns. Therefore it would be expected that a repeat of this experiment would show the protein to be highly α-helical. The methods of expression and purification method were subtly different from the 2009 paper, but this should not affect protein integrity. Comparing the two methods, the major difference is the use of a thrombin cleavage site between GST and M2-1 instead of using 3C or PreScission protease. PreScission protease has the advantage that it is most active at 4 °C and therefore cleavage can occur at a temperature at which the protein will remain stable. Thrombin is most active at 22 °C and also it gives more unspecific cleavage, in general, when compared with PreScission. In a comparison of GST-Fusion proteases PreScission protease was the most effective in terms of yield of protein by a sizeable margin when viewed on an SDS gel.[76] It should be noted that this study was carried out for the company that sells PreScission protease, however all of the proteases compared in the study were sold by the same company.

It was an aim of this project to successfully clone and express a soluble form of the Fusion Protein using the Baculovirus Expression System from Invitrogen. Unfortunately, initial cloning of FSol into a primary vector proved difficult and has not yet been achieved. Attempts were made via conventional ligation dependant cloning and cloning using the topoisomerase enzyme with a commercial kit but all were to

similar levels of success. All materials used were re-checked for activity, such as the restriction enzymes used and the topoisomerase cloning kits as detailed in Chapter Two. There are many variables to be considered that may have contributed to the difficulty in cloning the soluble fragment of F. It is possible that UV damage occurred to the plasmid and/or fragment during extraction of the bands from the gel. Possible UV damage to DNA includes pyrimidine dimer formation, destruction of bases, backbone breakdown and cross-linking. It is possible that this could interfere with the efficiency of ligation. If this occurred to the plasmid during gel extraction then this would explain why ligation into pMIB/V5-His did not work during any attempt at the initial strategy or during subcloning into this plasmid after cloning into a TOPO vector (see below). The gel was only exposed for a short amount of time, however this has still been shown to cause significant damage at around 300 nm UV and be a factor in reduction of cloning and transformation efficiency.[77] It has been shown that addition of 1 mmol/L guanosine in the running buffer protected the DNA from damage to a significant extent and this study determined that exposure of 20-45 seconds using a standard visualiser could leave as little as 1 % of the DNA fully intact.[77] A wavelength of around 254 nm causes the most damage compared to those of higher wavelengths.[78] There have been reports of UV damage being a likely cause of downstream cloning problems as early as the 1970s but the problem is often still thought to be a minor consideration.[78] Relatively recent development of alternative visualisation technology (for example the Clare Chemical Dark Reader which uses visible light and filters) means this is not necessary and the use of UV, or at least shortwave UV light could be avoided.

The problem appeared to occur at the point of ligation and/or transformation. If the problem was with the ligation step, this would explain why a couple of positive colonies were obtained when topoisomerase cloning was carried out and none with conventional ligation-dependant cloning. This was still a very low level of success. The TOPO cloning procedures were attempted on other DNA constructs for separate experiments successfully and therefore there is no fault with the topoisomerase enzyme and as using this method did not solve the problem it is safe to assume there is also no problem with the original DNA ligase used for the conventional cloning technique. Therefore there is either fault in the design of the ligating ends, and therefore problems at the restriction stage, or there are problems in transforming the product into competent cells. The commercial competent cells were double checked by transforming them with a separate construct, the gene YcaO from *E. coli* (a gift from Dr R. Dorazi, University of Durham), in the Champion vector successfully. Therefore the problem most likely arose in ligation.

The difficulty with cloning F is most likely to have come from the ability of the ligase to ligate the ends of the DNA fragment into the appropriate restricted sites of the new vector. Also, during the ligation if the molar ratio of the plasmid to target DNA is too high then the ligation may give a high number of circular empty plasmids and if the ratio is too low then an excess of linear and circular homo and heteropolymers of varying sizes, orientations and compositions will be produced. The presence of 5' phosphates recircularising the DNA has been accounted for by using Calf Intestinal Alkaline Phosphatase. Options for improvement and further work on cloning and expression of FSol are detailed in Chapter Seven.

The synthesis of the 19 amino acid long Fusion Peptide should be straightforward by solid phase synthesis techniques and could take as little as a week (Dr S. Cobb, personal communication), with structural characterisation to follow after purification, as discussed in Chapter Five. Although the heptad repeats have been characterised by crystallography before, this has not been the case for the fusion peptide, which is predicted to form a hydrophobic α-helix.

Experiments targeting F to specific cellular organelles were successful, showing that the constructs do localise to the correct respective organelles. The activity of the construct F_{STOP} was of interest as the images (see Fig. 3.3(d)) showed that absence of the C-terminal domain did not affect localisation of the Fusion protein to the membrane. Therefore another region of the protein must be responsible, likely the Transmembrane Domain itself. The Fusion protein imaging experiments to analyse at what point F interacts with M in order for virion assembly to occur were moderately successful, showing that F does not first interact with M in any of the organelles but rather at or before reaching the plasma membrane. Massive damage was caused to the cell structure during localisation of F and M using wild type F. It was noted in the construct that was designed to transport throughout the organelles in the normal order of glycoprotein processing that F finally localised with M at the Golgi apparatus, but as this construct was recycled it is likely that the interaction occurred at the plasma membrane and M was taken with F through the cycle. M was not localised at the Golgi apparatus with an F construct that was designed to localise to the Golgi apparatus only. It has not been possible as yet to ascertain whether this interaction occurs at or before reaching the plasma membrane. The Golgi apparatus is the last organelle that the protein locates to and is processed by before locating to the membrane.

Relating all aspects of these results to each other, progress has been made towards optimised production of the two protein constructs and the fusion peptide, even the

difficulties found in expressing a soluble form of the F protein only adds to the knowledge of how and how not to clone this particular fragment, and will aid the formulation of a new strategy. The localisation experiments with F and M proteins are a part of characterising the activity of these vital viral proteins within the cellular environment, as the main aim is indeed structural information and crystal structures of F and M2-1 but this line of research also aims to obtain useful biochemical information about their activity, specifically from localisation studies and experiments on lipid-protein interactions.

Chapter Seven: Conclusions and Recommendations

The aims of this project on preliminary analysis of viral proteins of human respiratory syncytial virus were to clone, express and purify the soluble form of the Fusion protein; to express, purify and begin crystallography trials on M2-1 protein; to synthesis the Fusion peptide by chemical methods; and to visualise the co-localisation of the F protein with M in mammalian cells. The conclusions made from this project are detailed below, along with recommendations to further this work and the potential use of structural information on F and M2-1 is described.

7.1 M2-1 Expression and purification

The M2-1 protein can now be expressed and purified successfully and at a level suitable for crystallographic trials. The purification protocol was optimised and a procedure utilising glutathione sepharose beads was found to be the most effective in terms of yield of protein and purity. Mass spectroscopic analysis has yet to be carried out on M2-1 obtained to confirm accurately the molecular mass of the product obtained. Purification on a MonoQ anion exchange column was not a success as the protein failed to elute from the column. Therefore a resin with lower affinity for M2-1 may improve this. It is also possible that the monomer and oligomers of M2-1 have different affinities for the MonoQ column resin, and therefore whilst the initial sample purified well at pH 8 the next and larger fraction passed straight through the column or stuck to the column at the same pH. However, purification by size using a gel filtration column was a success and M2-1 was obtained as a single species, in this case a tetramer confirming previous observations.[3] On a previous occasion when pure protein was obtained, there were difficulties in determining the optimum conditions for crystallisation (Dr V. Money, personal communication). In addition, the solution containing M2-1 was quite viscous and this meant difficulties in crystal trials, but this is not the case from the results of purifications carried out during this project. Initial crystal trial conditions were attempted using an Opti-Salts Suite kit from Qiagen.[79] To date, no crystals have been obtained but the full range of crystal forming conditions have not yet been exploited. M2-1 protein will also undergo circular dichroism analysis, in order to obtain data to compare against references and gain some knowledge of M2-1's secondary structure, and this will be compared against previous data obtained in 2009.[3]

7.2 FSol cloning

Cloning of the FSol fragment, or rather the extra-cellular F_1 domain, proved to be difficult, despite attempting two separate strategies that usually produce excellent results were utilised. Due to insufficient time, a new strategy was not attempted after topoisomerase I cloning. The commercial cloning kits were double checked by successful cloning of a different non-related target, the gene YcaO from E. coli (a gift from Dr R. Dorazi, University of Durham). There could be an incompatibility for cloning in the design of the construct and/or primers used as. Recommendations for future attempts include double-checking every small aspect of the cloning strategy, alteration of the FSol construct (even if this means obtaining a mutated protein) and attempting an alternative strategy, possibly with a commercial kit to give reliability and reproducibility. It may be a good idea for future work to simply attempt restriction using different restriction sites and enzymes, to create different ends. Other research groups have managed to clone constructs of F, but notably not a full-length soluble form. For example, a 1999 paper on the interaction of F with RhoA[80] successfully produced the extracellular domain and deletion mutants thereof for interaction studies. They did this using restriction sites *Eco*R1 and *Bam*H1 of the pAS2-BD vector from Clontech.[80] Alternative procedures are discussed below.

Alternative ligation independent cloning (LIC) methods (by recombination or producing and annealing overhangs) could be attempted, that are therefore independent of the insert sequence.[81] These methods, of which topoisomerase cloning is one, often add extra complexity to the procedure, for example requiring additional sequences at the 5' end of the primers that must match the vector sequences. For cloning with no enzyme mediation, these methods often also require four pairs of primers in order to replace conventional ligation with PCR mediated cloning.[82] There is also enzyme-mediated LIC, where after the additional sequences are designed on the 5' of the primers, the fragment is treated with one of a number of enzymes (such as T4 DNA polymerase, exonuclease III or Uracil DNA glycosylase) to give sticky ends in order to insert it into a suitable vector by hybridising with the complementary sticky ends of the vector.[82, 83] A broad review of ligation independent cloning methods along with details of a highly efficient enzymatic cloning procedure describing the In-Fusion™ enzyme (Clontech) can be found in Berrow et al. 2007,[81] and in the Clontech In-Fusion cloning handbooks.[84]

In order to increase the chance of ligation of all available fragments and vectors in the reaction, it could be an idea to add the original restriction enzymes, *Hind*III and *Xba*I, to

the ligation reaction. This would be advantageous as, providing that after ligation the restriction sites were not usually reformed, then the presence of the restriction enzymes would restrict any self-ligated plasmid and increase the number of recombinants. The equilibrium would be driven to produce the desired recombinant plasmid. Circular DNAs produce more transformants than linear DNA and therefore the colonies successfully transformed with the recombinant plasmid should outnumber those with plasmid lacking the desired cloned insert. Despite the advantages in terms of cost and of not needing to add special sequences to the ends of the primers, it is true that blunt-end cloning is generally less efficient than sticky-end cloning and there is an increase in the number of vectors self-ligating.[82] Sticky-end cloning may be the option that would maximise the efficiency of this cloning reaction. However, the topoisomerase cloning using the Invitrogen Champion Cloning kit uses a sticky end approach in order to give directional cloning and this failed to product positive results. Sticky end approaches also have problems with the ability of the restriction enzymes to efficiently cut the DNA at the required sites, especially if those sites are at the 5' end of double stranded PCR product. However, one way of avoiding this would be to use a technique called Autosticky PCR (AS-PCR) that uses two primers containing tetrathydrofuran-derivative abasic sites that stall DNA polymerase, giving PCR products with different single stranded 5' overhangs at both ends- giving sticky ends for two restriction sites. These products are then directionally cloned into appropriately restricted plasmids.[82]

It could be more effective to use synthetic linkers 8-16 bases long encoding restriction sites, ligating them to the blunt ended fragment to give cohesive ends for cloning.[68] A detailed protocol for this can be found in Sambrook and Russell, 2001.[68] However, this relatively old technique is usually only used now when a necessary restriction site is missing from the DNA fragment to be cloned, or when the sequence of the DNA is unknown.

Further ahead, research questions about the mode action of the fusion protein include ascertaining how many Fusion Proteins are required to form a functional fusion pore and also how many fusion pores are required to give full infectivity. In depth mathematical modelling in combination with lipid membrane interaction experiments would be necessary to answer these questions.

7.3 Co-localisation of F and M

The IFA imaging experiments were successful, despite the preliminary experiments requiring high exposure times and therefore giving poor images. It was demonstrated that the constructs that targeted or retained the F protein in sub-cellular compartments other than the plasma membrane expressed and functioned as predicted. When co-expressed with the M protein we only observed co-localisation at a later stage of the transport of F i.e. post Golgi body. We can conclude this as the M protein did not co-localise with F when F was retained in the ER or Golgi. The F and M did co-localise at the plasma membrane and within endosomes and Golgi as a result of recycling from the membrane. We can tentatively conclude failure to co-localise at the early stage was not due to the modification of the cytoplasmic tail as the later constructs were similarly modified, notably the F construct with the ER retention signal disrupted acts as unmodified F. Co-localisation of F and M was also not dependent on the presence of the cytoplasmic tail; however we did not test if the cytoplasmic tail minus construct was directed to lipid rafts as was the wild type F.

However, it has not been possible as yet to ascertain whether the F/M interaction occurs at or before reaching the plasma membrane via the normal transport pathway. This could be done if a larger time frame was available in order to undertake electron microscopy studies of the transport of proteins between the membrane and the Golgi apparatus. The Golgi is the last organelle that the protein locates to, where it is modified by proteolytic cleavage to mature F to before locating to the membrane. This would involve high pressure freezing of cell samples (to preserve ultrastructure without creating ice crystal damage) before labelling with gold-labelled antibodies specific to M and to F. To distinguish between M and F antibodies the gold labels would be of different diameters for F and for M, which could be distinguished using an electron microscope. A project such as this would be complex and take a considerable length of time.

One significant observation was that when F and M were co-expressed the levels of M produced were significantly greater than if M was expressed alone. This was not due to the levels of DNA transfected when two plasmids were used as an empty vector was usually added to the M only mix to normalise DNA levels. The observed effect was of significant damage to the membranes resulting in cell disruption. Even when there was no observed interaction between F and M at the ER or Golgi, the result was the same, significant damage to the membrane. Thus it would be inappropriate to fully dismiss

any interaction between ER or Golgi trapped F and M, but currently we cannot provide an adequate explanation of what that interaction is.

Further work using an alternative expression system, such as a lentiviral based one, with inducible promoters to control expression of the proteins and integration of the genomic DNA into the target cell genome giving stable expression, could be employed.[85] Changing the transmembrane domain should also be considered as cytoplasmic tailless F interacted with M.

7.4 Chemical synthesis of the Fusion Peptide

Synthesis of the Fusion Peptide using solid phase techniques should be straightforward as detailed in Chapter Five. Future studies on the activity of the fusion protein should include crystallisation and studies into the peptide's interaction with lipids in bilayers, i.e. membranes. The question to answer would be is the fusion peptide able to locate and interact with the cell lipid membrane on its own or does it require other parts of the whole protein to do this? It would be useful to conduct experiments on both artificially synthesised lipid membranes, liposomes and on mammalian cells using immunofluorescence studies and electron microscopy.

7.5 Validity and reliability of results

A point on the validity of any results obtained from studies on FSol and the fusion peptide is that as they are separate constructs, any results could not be wholly applied to conclusions about the full protein, although they would provide a very good indication of the full structure. It would be ideal to be able to crystallise the protein in both inactive and active conformations. These have been issues with previous studies described in Chapter One, for example, the NDV homology model of F and the crystal structure of F core. It should be noted however that if F protein was able to be expressed in Baculovirus in insect cells (once cloning was successful), then this should reproduce most post-translational modifications that would occur in a mammalian cell, such as cleavage of F_1 from F_2. Expressing F protein in mammalian cells could be investigated but first cloning needs to be achieved. Any structural information gained or knowledge on their activity with lipids will be at the least an additional piece of information, but an important piece for biological understanding and future inhibitor development. Any structural information gained from crystallography of M2-1 would indeed be valid as the protein was obtained as a whole, with no hydrophobic sections or disulphide bridges to make purification difficult. M2-1 does form oligomers and of course this would need to be taken into account during crystallography studies.

7.6 In the Future: Routes to inhibition of the F and M2-1 proteins functionality

Concerning studies into the inhibition of the F protein's fusion ability, it was found that using peptides that represent the heptad repeats results in the most success in terms of specific, targeted and complete inhibition of F activity and therefore viral infectivity.[38] Peptides synthesised from the two heptad repeats have been shown to have an inhibitory affect on the protein's activity. Peptides from HR1 can also prevent syncytia formation.[15] The mode of action is proposed to be by interference by the peptide on the interaction between HR2 and HR1 during the conformational change that occurs to facilitate drawing together of opposing membranes, which is crucial for infectivity. It is likely this interference would occur between the peptide and the α-helical N terminus of HR2.[15] There are a number of potential molecules developed that could inhibit F's fusion activity or its processing.[15] As a proof of principle in developing molecules that inhibit fusion two peptides synthesised from HIV-1 gp41's (the equivalent of F) heptad repeats, DP-107 and DP-178, are potent inhibitors of HIV-1.[86] The sequences they were synthesised from correspond to the heptad repeat regions of the protein. Therefore the study was extended to DP-like peptides synthesised from other related viruses, including hRSV, as structurally the proteins share a common mechanism in mediating membrane fusion. It was found that the DP-178-like inhibitors (based on HR1) were highly potent in blocking syncytia formation, and were very specific for their virus of origin. Only the inhibitors of hPIV3 were cross reactive with hRSV and MuV but this required thirty times the usual concentration.[86] The most important example of this is however the peptide T20, which has been developed from the C-terminal heptad repeat of gp41, into a successful drug, Enfuvirtide, against HIV-1 (marketed also as Fuzeon). Studies into this mode of inhibition could be the most likely route to drug design, especially considering this strategy has been a success against HIV-1's gp41 fusion protein.[17, 18] It only accentuates the importance of obtaining an accurate crystal structure of hRSV F as exact knowledge of the protein's conformation and intra-domain interactions between HR1 and HR2 would aid in the design of the inhibitor molecule. In this respect, it would of course be ideal to aim to obtain not only the crystal structure of hRSV F, but also that of F with inhibitors bound. There is one such structural study which showed an inhibitor, derived from the heptad repeats of RSV F (TMC353121), bound to the hexameric core.[38] This inhibitor did not prevent the interaction between the protein's heptad repeats directly, but instead altered the conformation of HR2.[38] However, this structure is limited to the inhibitor molecule and to the heptad repeat peptides and not the complete protein.[38]

A recent study into M2-1 inhibition concentrated upon the zinc finger motif. Compounds targeting a zinc finger have successfully inhibited other viral nucleocapsid proteins, including HIV-1 and MuLV proteins. In general, it is better to inhibit an intracellular protein than an extracellular one as altering surface proteins risks the host immune response to the virus being adversely altered also. However, this study involved the use of a very general chemical inhibitor that cross-links cysteine residues, which are obviously present in many proteins of a cell not only the M2-1 protein.[87] Therefore, with few studies into specific inhibition of the M2-1 protein available, structural information would be an excellent start.

Small steps have been made in this project towards structural characterisation and therefore towards successful design of inhibitor molecules targeting the F and M2-1 proteins of hRSV. The suggestions for further work in this chapter should be carried out in order to not only achieve cloning and expression but especially to succeed in the aim of obtaining crystal structures for both proteins. Research into hRSV has been going on for decades without positive results for a vaccine or effective drug molecule, making accurate structural information on the vital F and M2-1 proteins very important indeed for careful inhibitor design.

Appendix A: Plasmid Maps

A1. pMIB/V5-His

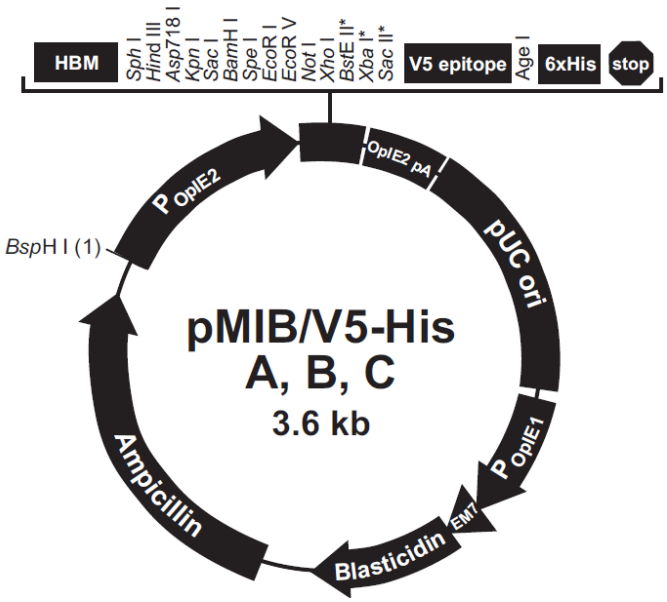


Figure A1: Features of the pMIB/V5-His plasmid including restriction sites of the multiple cloning site. Taken from the Invitrogen pMIB/V5-His A, B, and C Vector Kit User Manual.[88]

HBM	Honeybee melittin secretion signal
V5 epitope	V5 epitope
6xHis	Polyhistidine (6x) tag
Stop	Stop codon
EM7	EM7 promoter
P_{OpIE1}	OpIE1 Promoter
pUC ori	pUC origin
P_{OpIE2}	Reverse priming site
P_{OpIE2} pA	Polyadenylation sequence
Ampicillin	Ampicillin resistance gene
Blasticidin	Blasticidin resistance gene

A2. pET100/D-TOPO®

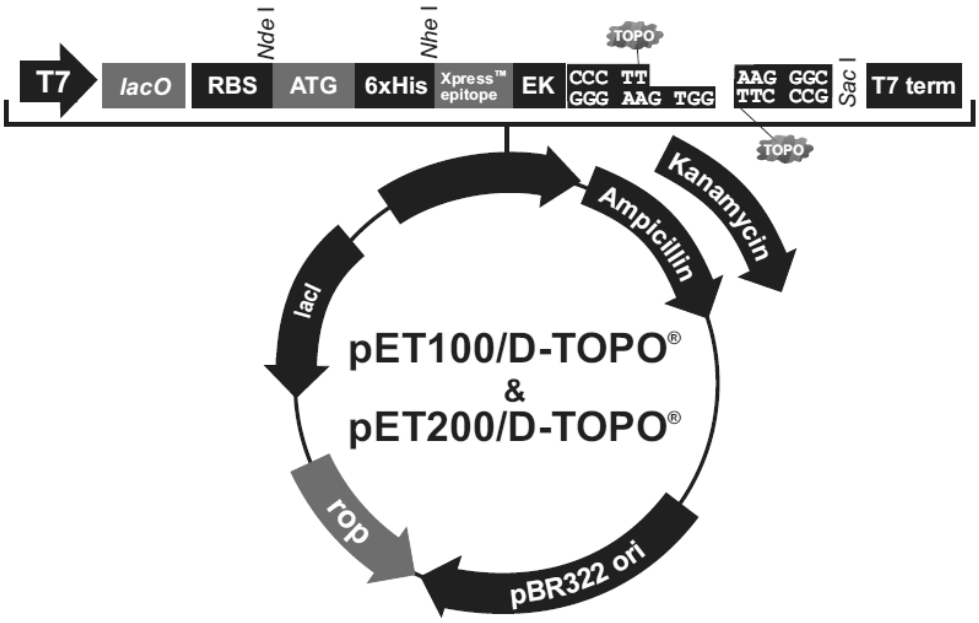


Figure A2: Features of the pET100/D-TOPO plasmid from the Invitrogen Champion pET Directional TOPO Expression Kit. Taken from the Champion™ pET Directional TOPO® Expression Kits User Manual.[66]

lacO	*Lac* operator
RBS	Ribosome binding site
ATG	Initiation codon ATG
6x His	Polyhistidine (6x) tag
T7	T7 promoter and priming site
Xpress™ epitope	
EK	EK recognition site
T7 term	T7 transcription termination region
lacI	lacI origin of replication (complementary strand)
Ampicillin	Ampicillin resistance gene
Kanamycin	Kanamycin resistance gene
pBR322 ori	pBR322 origin
rop	ROP origin of replication (complementary strand)

A3. pCR® 4Blunt-TOPO®

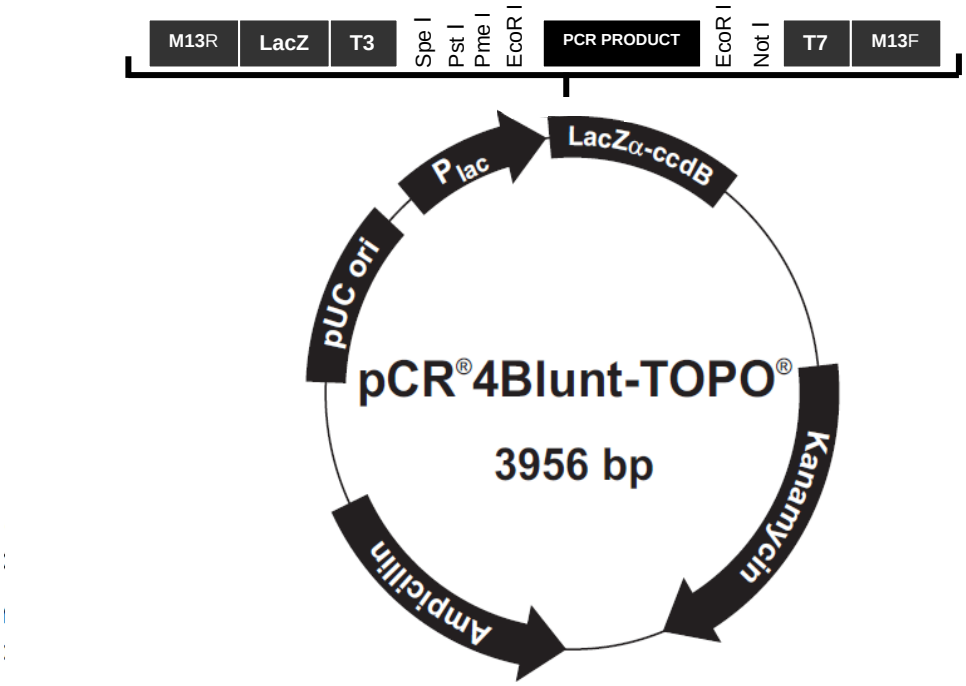


Figure A3: Features of the pCR 4Blunt-TOPO vector from the Invitrogen Zero Blunt® TOPO® PCR Cloning Kit for Sequencing, including restriction sites. Adapted from Zero Blunt TOPO PCR Cloning Kit for Sequencing User Manual.[67]

M13R	M13 Reverse Priming site
LacZ	LacZα initiation codon
T3	T3 priming site
T7	T7 priming site
M13F	M13 Forward priming site
LacZα-ccdB	LacZα-ccdb gene fusion. The expression of this is disrupted if a blunt ended PCR product is ligated into the vector, ensuring growth of only positive colonies.
pUC ori	pUC origin
P_{Lac}	Lac promoter
Ampicillin	Ampicillin promoter and resistance gene
Kanamycin	Kanamycin promoter and resistance gene

Appendix B: Sequences and Primers

B1. Sequence of F with the Fusion Peptide in purple, the signal peptide and transmembrane domain in teal:

1 MELLILKANAI **11** TTILTAVTFC **21** FASGQNITEE **31** FYQSTCSAVS
41 KGYLSALRTG **51** WYTSVITIEL **61** SNIKKNKCNG **71** TDAKIKLIKQ
81 ELDKYKNAVT **91** ELQLLMQSTP **101** ATNNRARREL **111** PRFMNYTLNN
121 AKKTNVTLSK **131** KRKRRFLGFL **141** LGVGSAIASG **151** VAVS KVLHLE
161 GEVNKIKSAL **171** LSTNKAVVSL **181** SNGVSVLTSK **191** VLDLKNYIDK
201 QLLPIVNKQS **211** CSISNIETVI **221** EFQQKNNRLL **231** EITREFSVNA
241 GVTTPVSTYM **251** LTNSELLSLI **261** NDMPITNDQK **271** KLMSNNVQIV
281 RQQSYSIMSI **291** IKEEVLAYVV **301** QLPLYGVIDT **311** PCWKLHTSPL
321 CTTNTKEGSN **331** ICLTRTDRGW **341** YCDNAGSVSF **351** FPQAETCKVQ
361 SNRVFCDTMN **371** SLTLPSEVNL **381** CNVDIFNPKY **391** DCKIMTSKTD
401 VSSSVITSLG **411** AIVSCYGKTK **421** CTASNKNRGI **431** IKTFSNGCDY
441 VSNKGVDTVS **451** VGNTLYYVNK **461** QEGKSLYVKG **471** EPIINFYDPL
481 VFPSDEFDAS **491** ISQVNEKINQ **501** SLAFIRKSDE **511** LLHNVNAVKS
521 TTN IMITTII **531** IVIIVILLSL **541** IAVGLLLYCK **551** ARSTPVTLSK
561 DQLSGINNIA **571** FSN

B2. Sequence of M2-1:

1 MSRRNPCKFE **11** IRGHCLNGRR **21** CHYSHNYFEW **31** PPHALLVRGN
41 FMLNKILKSM **51** DKSIDTLSEI **61** SGAAELDRTE **71** EYALGIVGVL
81 ESYIGSINNI **91** TKQSACVAMS **101** LLIEINSDD **111** IKKLRDNEEP
121 NSPKIRVYNT **131** VISYIESNRK **141** NNKQITHLLK **151** RLPADVLKKT
161 KNTLDIHKS **171** ITISNPKEST **181** VNDQNDQTKN **191** NDITG

B3. Primers:

FSOLFORHINDIII: GGC ATG CT**A AGC TT**A CAA AAC ATC ACT GAA GAA TTT TAT C

FSOLREVXBA1: C GAA GGG CCC **TCT AGA** ATT TGT GGT GGA TTT AAC AGC

B4. Topoisomerase cloning forward primer:

TOPOFSOLFORHINDIII: **CAC C**AA GCT TAC AAA ACA TCA CTG AAG AAT TTT ATC

Appendix C: Experimental analysis spectra

C1. Boc-Phe

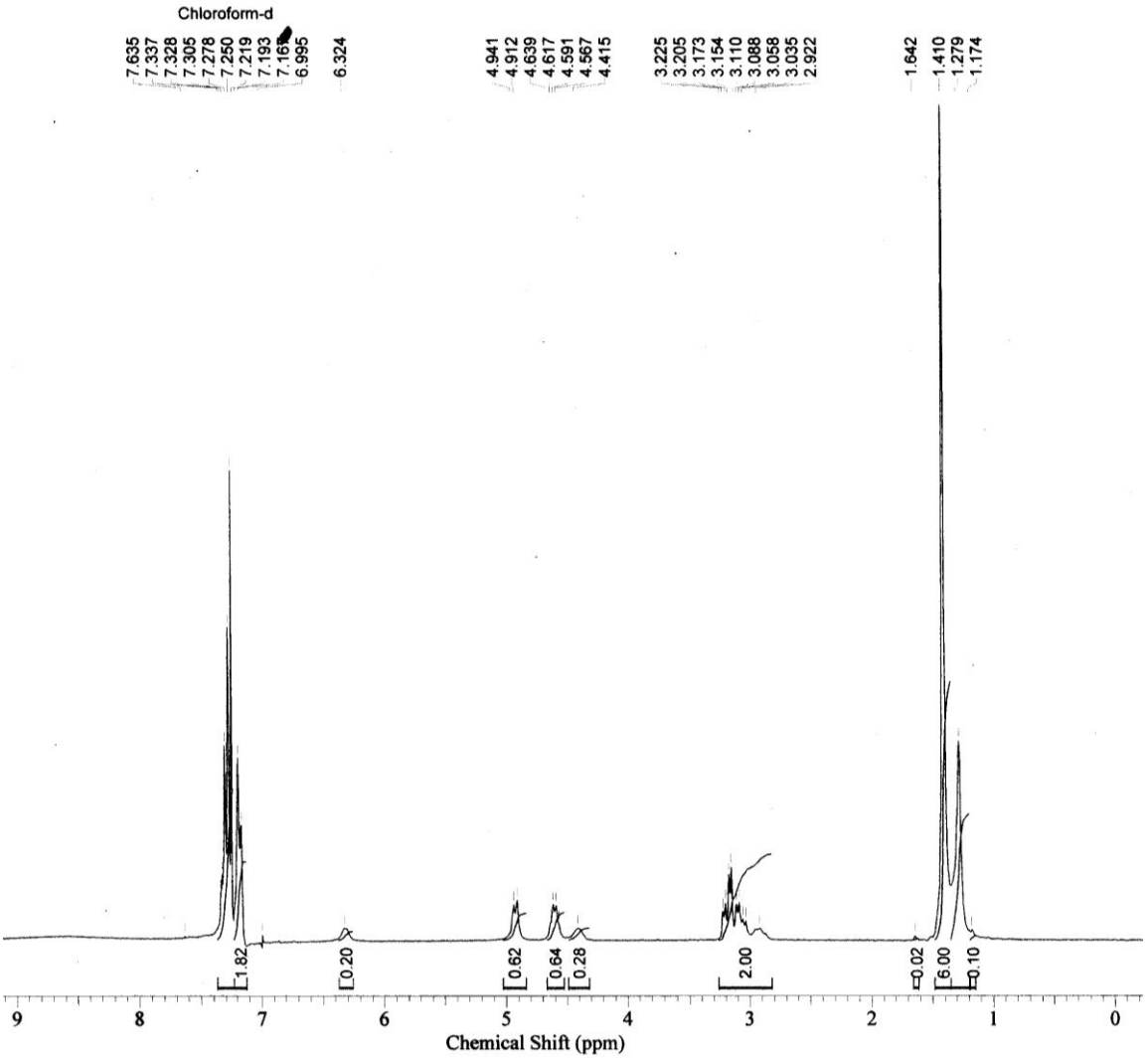


Figure C1: Boc-Phe ^{1}H NMR (269.72 MHz; $CDCl_3$) $\boldsymbol{\delta_H}$ 7.34-7.28 (m, aromatic, 3H), 7.25-7.22 (m), 7.17 (d, *J* = 8.9, aromatic 2H), 5.99 (s, OH, 1H), 4.91 (d, *J* = 8.1), 4.62 (d, *J* = 8.1), 4.39 (s, CH, 1H), 3.15 (qd J = 5.4, 10.8, 5.4, 10.8, CH_2, 2H), 1.43 (s, $(CH_3)_3$, 9H), 1.30 (s), 1.17 (s)

C2. Boc-Phe-OBzl produced via Scheme 5.7

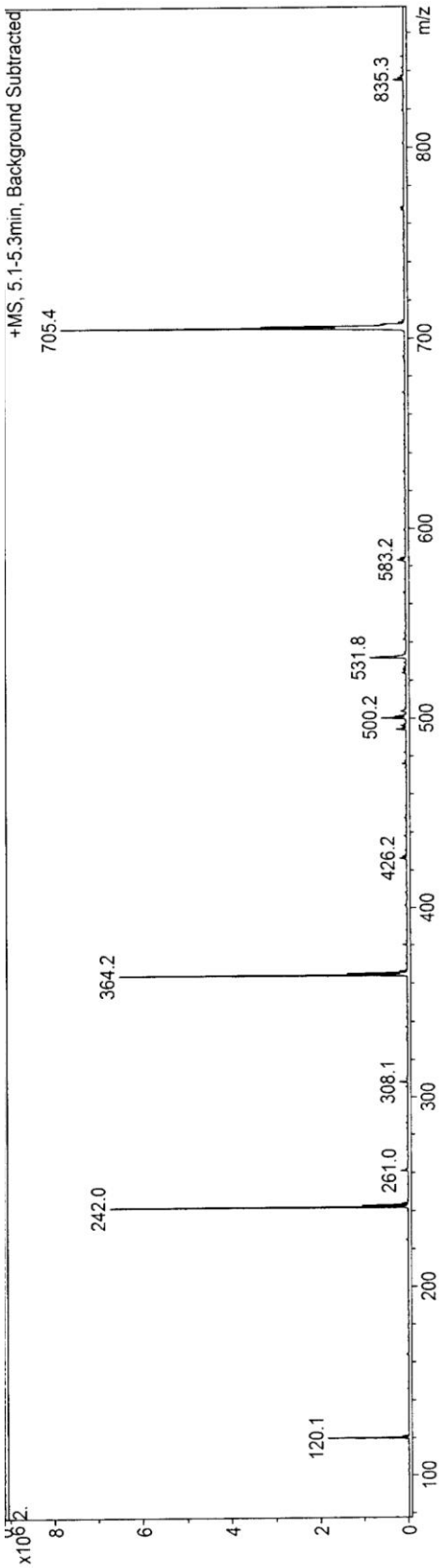


Figure C2: Boc-Phe-OBzl produced via scheme 5.7: LCMS m/z 364.2 $[M + Na]^+$, 242.0$[M - Boc]^+$, 705.4 $[2M + Na + H]^+$

C3. Boc-Phe-OBzl produce via Scheme 5.8

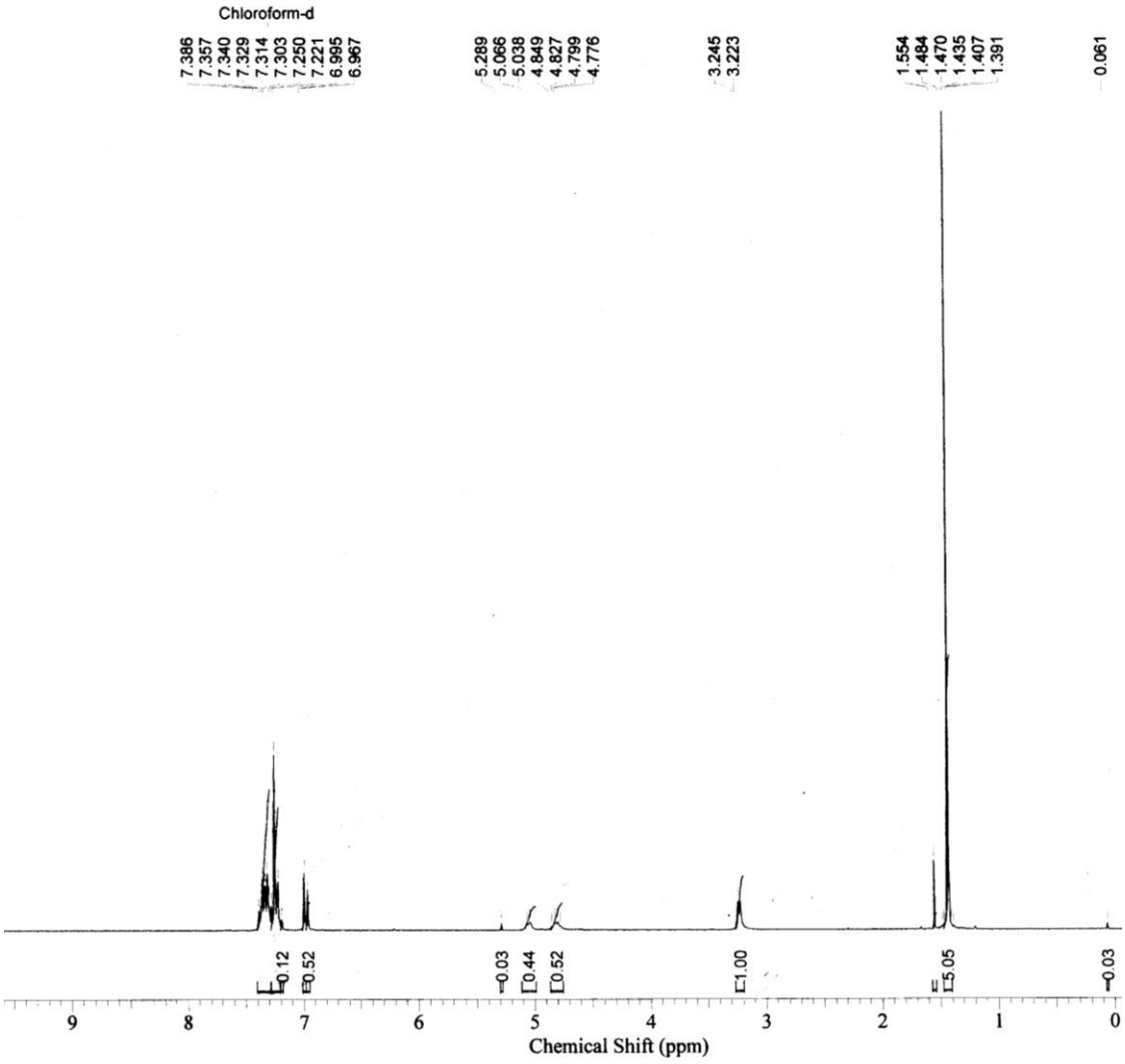


Figure C3: Boc-Phe-OBzl produce via scheme 5.8: ^{1}H NMR (269.72 MHz; $CDCl_3$) $\boldsymbol{\delta_H}$ 7.39-7.28 (m), 7.19 (s, 8H aromatic), 7.00 (d, *J* = 8.1 aromatic, 2H), 6.97 (s), 5.29 (s), 5.11 (d, *J* = 8.1), 4.82 (q, *J* = 5.4), 3.23 (d, *J* = 8.1, CH_2, 2H), 1.56 (s), 1.45 (s, $(CH_3)_3$ 9H), 0.06 (s)

C4. Boc-Phe-OBzl produced via Scheme 5.8

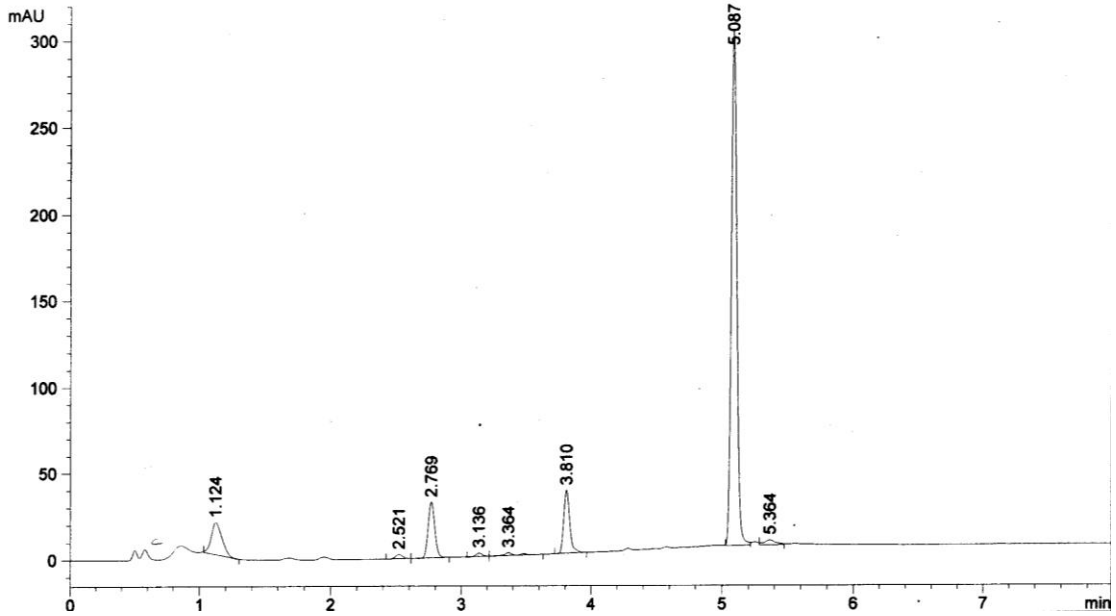


Figure C4: Boc-Phe-OBzl produced via scheme 5.8: Liquid Chromatography chromatogram showing completeness of reaction and purity of product

C5. Deprotection of Boc-Phe.

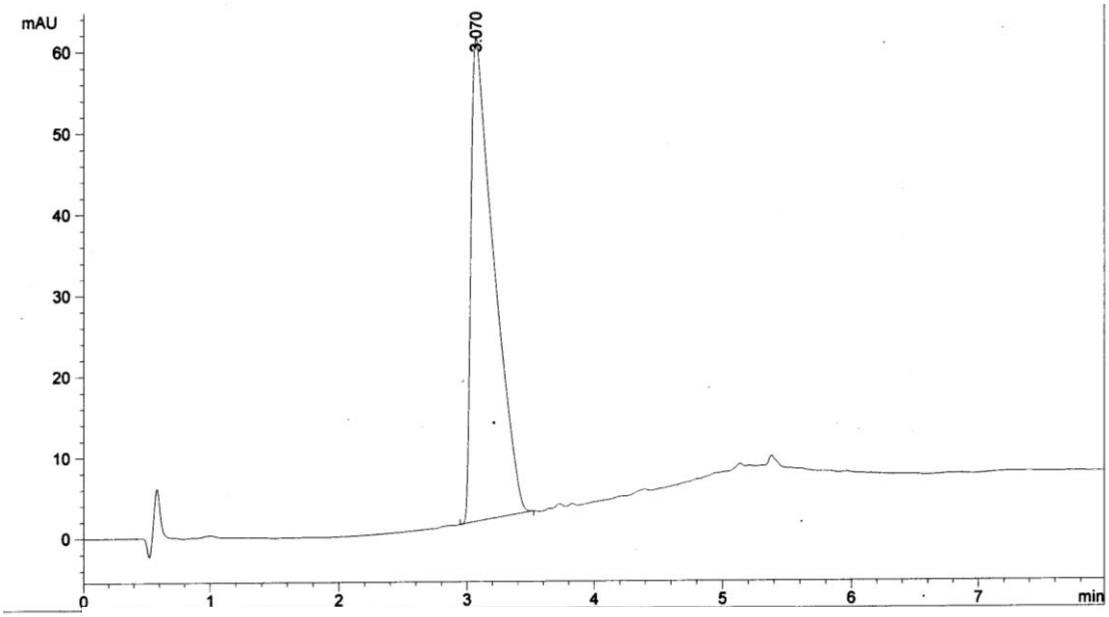


Figure C5: Deprotection of Boc-Phe: Liquid chromatography chromatogram showing completeness of reaction and purity of product

C6. OBzl-Phe-Leu-Fmoc

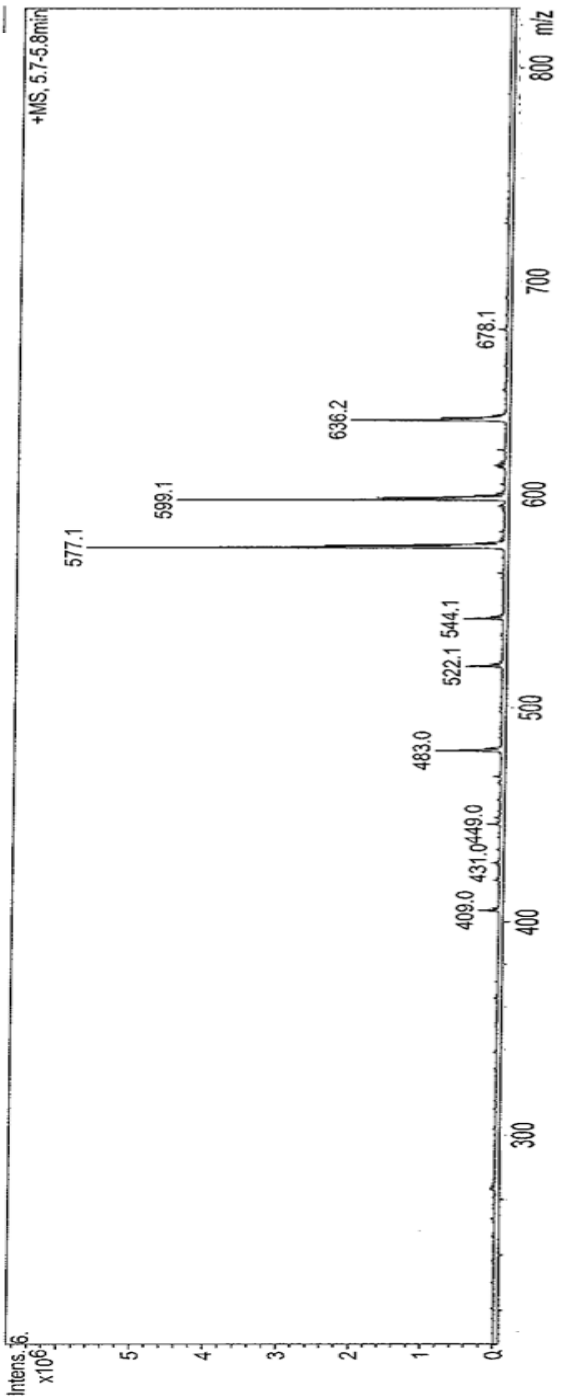


Figure C6. OBzl-Phe-Leu-Fmoc: LCMS m/z 577.1 $[M^+ + H]^+$, 599.1 $[M^+ + Na + H]^+$

Bibliography

1. X. Zhao, M. Singh, V.N. Malashkevich, P.S. Kim, Structural characterisation of the human respiratory syncytial virus fusion protein core, Proc. Natl. Acad. Sci. U. S. A., 2000, 97, 14172-14177
2. Strauss J.H. and Strauss E.G., Viruses and Human Disease, Academic Press, 2002, 123-143
3. T. Tran, N. Castagne, V. Dubosclard, S. Noinville, E. Koch, M. Moudjou, C. Henry, J. Bernard, R.P. Yeo, J. Eleouet, The respiratory syncytial virus M2-1 protein forms tetramers and interacts with RNA and P in a competitive manner, J. Virol., 2009, 83, 6363-6374
4. I. Cuesta, X. Geng, A. Asenjo, N. Villanueva, Structural phosphoprotein M2-1 of the Human Respiratory Syncytial Virus is an RNA binding protein, J. Virol., 2000, 74, 9858-9867
5. A.G.P. Oomens, K.P. Bevis and G.W. Wertz, The cytoplasmic tail of the Human Respiratory Syncytial Virus F Protein plays critical roles in cellular localisation of the F Protein and infectious progeny production, J. Virol., 2006, 80, 10465
6. R.S Tang., N. Nguyen, X. Cheng, H. Jin, Requirement of cysteines and length of the human respiratory syncytial virus M2-1 protein for protein function and virus viability, J. Virol., 2001, 75, 11328-1133
7. T.L. Gower, M.K. Pastey, M.E. Peeples, P.L. Collins, L.H. McCurdy, T.K. Hart, A. Guth, T.R. Johnson, B.S. Graham, RhoA signalling is required for respiratory syncytial virus-induced syncytium formation and filamentous virion morphology, J. Virol., 2005, 79, 5326-5336
8. V.A. Money, H.K. McPhee, J.A. Mosely, J.M. Sanderson, R.P. Yeo, Surface features of a Mononegavirales matrix protein indicate sites of membrane interaction, Proc. Natl. Acad. Sci. U. S. A., 2009, 106, 4441-4446
9. G. Henderson, J. Murray, R.P. Yeo, Sorting of the respiratory syncytial virus matrix protein into detergent-resistant structures is dependant on cell surface expression of the glycoproteins, Virology, 2002, 300, 244-254
10. P.L. Collins, M.G. Hill, J. Cristina, H. Grosfeld, Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus, Proc. Natl. Acad. Sci. U. S. A., 1996, 93, 81-85
11. R. G. Tawar, S. Duquerroy, C. Vonrhein, P. F. Varela L. Damier-Piolle, N. Castagné, K. MacLellan, H. Bedouelle, G. Bricogne, D. Bhella, J. Eléouët, F.A. Rey Crystal Structure of a Nucleocapsid-Like Nucleoprotein-RNA Complex of Respiratory Syncytial Virus, Science, 2009, 326, 1279-1283
12. B.J. Smith, M.C. Lawrence, P.M. Colman, Modelling the structure of the fusion protein from human respiratory syncytial virus, Protein Eng., 2002, 15, 365-371
13. A. Bukreyev, S.S. Whitehead, B.R. Murphy, P.L. Collins, Recombinant respiratory syncytial virus from which the entire SH gene has been deleted grows efficiently in cell culture and exhibits site-specific attenuation in the respiratory tract of the mouse, J. Virol., 1997, 71, 8973-8982
14. S. Fuentes, K.C. Tran, P. Luthra, M.N. Teng, B. He, Function of the respiratory syncytial virus small hydrophobic protein, J. Virol., 2007, 81, 8361-8366
15. D. S. Yeo, R. Chan, G. Brown, L. Ying, R. Sutejo, J. Aitken, B. Tan, M.R. Wenk, R.J. Surge, Evidence that selective changes in the lipid composition of raft-membranes occurs during respiratory syncytial virus infection, Virology, 2009, 386, 168-182
16. C. Carromeu, F.M. Simabuco, R.E. Tamura, L. E. Farinha Arcieri, A.M. Ventura, Intracellular localisation of human respiratory syncytial virus L protein, Arch. Virol., 2007, 152, 2259-2263
17. B.E. McGillick, T.E. Balius, S. Mukherjee, R.C. Rizzo, Origins of resistance to the HIV gp41 viral entry inhibitor T20, Biochemistry, 2010, 49, 3575-3592
18. A.M.T. Martins do Canto, A.J.P. Carvalho, J.P. Prates, L.M.S. Loura, Structure and conformation of HIV fusion inhibitor peptide T-1249 in presence of model membranes: A molecular dynamics study, J. Mol. Struct.:Theochem, 2010, 946, 119-124
19. D. Eggink, J.P.M. Langedijk, A.M.J.J. Bonvin, Y. Deng, M. Lu, B. Berkhout, R.W. Sanders, Detailed mechanistic insights into HIV-1 sensitivity to three generations of future inhibitors, J. Biol. Chem., 2009, 284, 26941-26950

20. N. L. Kallewaard, A.L. Bowen, J.E. Crowe Jr., Cooperativity of actin and microtubule elements during replication of respiratory syncytial virus, Virology, 2005, 331, 73-81
21. D. Martin, L. J. Calder, B. Garcia-Barreno, J.J. Skehel, J.A. Melero, Sequence Elements of the fusion peptide of human respiratory syncytial virus fusion protein required for activity, J. Gen. Virol. 2006, 87, 1649-1658
22. J. Schlender, G. Zimmer, G. Herrler, K. Conzelmann, Respiratory Syncytial Virus (RSV) Fusion Protein Subunit F2, Not Attachment Protein G, Determines Specificity of RSV Infection, J. Virol., 2003, 77, 4609-4616
23. R.A. Karron, D.A. Buonagurio, A.F. Georgiu, S.S. Whitehead, J.E. Adamus, M.L. Clements-Mann, D.O. Harris, V.B. Randolph, S.A. Udem, B.R. Murphy, M.S. Sidhu, Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication *in vitro:* Clinical evaluation and molecular characterisation of a cold-passaged, attenuated RSV subgroup B mutant, Proc. Natl. Acad. Sci. USA, 1997, 94, 13961
24. J.A. Lopez, R.Bustos, C. Orvell, M. Berois, J.Arbiza, B. Garcia-Barreno, J.A. Melero, Antigenic Structure of Human Respiratory Syncytial Virus Fusion Glycoprotein, J. Virol., 1998, 72, 6922
25. N.D. Day, P.J. Branigan, C. Liu, L.L Gutshall, J. Luo, J.A. Melero, R.T. Sarisky, A.M. Del Vecchio, Contribution of cysteine residues in the extracellular domain of the F protein of human respiratory syncytial virus to its function, Virology Journal, 2006, 3, 34
26. H.L. McL. Rixon, C. Brown, G. Brown, R.J. Sugrue, Multiple glycosylated forms of the respiratory syncytial virus fusion protein are expressed in virus-infected cells, J. Gen. Virol., 2002, 83, 61
27. L. Gonzalez-Reyes, M.B. Ruiz-Arguello, B. Garcia-Barreno, L. Calder, J.A. Lopez, J.P. Albar, J.J. Skehel, D.C. Wiley, J.A. Melero, Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion, Proc. Natl. Acad. Sci. U. S. A. 2001, 98, 9859-9864
28. M. B. Ruiz-Arguello, L. Gonzalez-Reyes, L.J. Calder, C. Palomo, D. Martin, M.J. Saiz, B. Garcia-Barreno, J.J. Skehel, J.A. Melero, Effect of proteolytic processing at two distinct sites on shape and aggregation of an anchorless fusion protein of human respiratory syncytial virus and fate of the intervening segment, Virology, 2002, 298, 317-326
29. A.S. Yunus, T.P. Jackson, K. Crisafi, I. Burimski, N.R. Kilgore, D. Zuomplis, G.P. Allaway, C.T. Wild, K. Salzwedel, Elevated temperature triggers human respiratory syncytial virus F protein six-helix bundle formation, Virology, 2010, 396, 226-237
30. T.F. Wild, J. Fayolle, P. Beauverger, R. Buckland, Measles Virus Fusion: Role of the cysteine-rich region of the fusion glycoprotein, J. Virol. 1994, 68, 7546-7548
31. M. Batonick, A.G.P. Oomens and G.W. Wertz, Human Respiratory Syncytial Virus glycoproteins are not required for apical targeting and release from polarized Epithelial cells, J. Virol. 2008, 82, 8664-8672
32. P.J.Branigan, N.D. Day, C. Liu, L.L. Gutshall, J.A. Melero, R.T. Sarisky, A.M. Del Vecchio, The cytoplasmic domain of the F Protein of Human respiratory syncytial virus is not required for cell fusion, J. Gen. Virol., 2006, 87, 395-398
33. J. Basu, Protein palmitoylation and dynamic modulation of protein function, Current Science, 2004, 87, 212-217
34. R.G. Arumugham, R.C. Seid, Jr., S. Doyle, S.W. Hildreth, P.R. Paradiso, Fatty acid acylation of the fusion glycoprotein of respiratory syncytial virus, J. Biol. Chem., 1989, 264, 10339-10342
35. T.P. McDonald, C.E. Jeffree, P. Li, H.W. McL. Rixon, G. Brown, J.D. Aitken, K. MacLellan, R.J. Sugrue, Evidence that maturation of the N-linked glycan of the respiratory syncytial virus (RSV) glycoproteins is required for virus-mediated cell fusion: The effect of α-mannosidase inhibitors on RSV infectivity, Virology, 2006, 350, 289-301
36. G. Zimmer, I. Trotz, G. Herrler, N-Glycans of F Protein differentially affect fusion activity of human respiratory syncytial virus, J. Virol., 2001, 75, 4744-4751
37. M.K. Lawless-Delmedico, P. Sista, R. Sen, N.C. Moore, J. B. Antczak, Jonathan M. White, R. J. Greene, K.C. Leanza, T. J. Matthews, and D.M. Lambert, Heptad-repeat regions of Respiratory Syncytial Virus F_1 Protein form a six-membered coiled-coil helix, Biochemistry, 2000, 39, 11684-1695
38. D. Royman, H.L. De Bondt., E. Arnoult, P. Geluykens, T. Gevers, M. van Ginderen, N. Verheyen, K. Hidong., R. Willebrords, J. Bonfanti, W. Bruinzeel, M.D. Cummings, H.

van Vlijimen, K. Andries, Binding of a potent small-molecule inhibitor of six-helix bundle formation requires interactions with both heptad repeats of the RSV fusion protein, Proc. Natl. Acad. Sci. U. S. A., 2010, 107, 308-313

39. C.J. Morton, R. Cameron, L. J. Lawrence, B. Lin, M. Lowe, A. Luttick, A. Mason, J. McKimm-Breschkin, M.W. Parker, J. Ryan, M. Smout, J. Sullivan, S. P. Tucker, P.R. Young, Structural characterization of respiratory syncytial virus fusion inhibitor escape mutants: homology model of the F protein and a syncytium formation assay, Virology 2003, 311, 275–288

40. K. Swansom, X. Wen, G.P. Leser, R.G. Paterson, R.A. Lamb, T.S. Jardetzky, Structure of the Newcastle Disease Virus F protein in the post-fusion conformation, Virology, 2010, 402, 372-379

41. H.S. Yin, R.G. Paterson, X. Wen, R.A. Lamb, T.S. Jardetzky, Structure of the uncleaved ectodomain of the paramyxovirus hPIV3 fusion protein, Proc. Natl. Acad. Sci. U. S. A. , 2005, 102, 9288-9293

42. P. Yuan, T.B. Thompson, B.A. Wurzburg, R.G. Paterson, R.A. Lamb, T.S. Jardetzky, Structural studies of the parainfluenza virus 5 hemagglutinin-neuraminidase tetramer in complex with its receptor, silalyllactose, Structure, 2005, 13, 805-815

43. K.A. Baker, R.E. Dutch, R.A. Lamb, T.S. Jardetzky, Structural basis for paramxyovirus-mediated membrane fusion, Mol. Cell, 1999, 3, 309-319

44. B.J. Smith, M. C. Lawrence, P.M. Colmann, Modelling the structure of the fusion protein from human respiratory syncytial virus, Protein Eng., 2002, 15, 365-371

45. K.J. Cross, L.M. Burleigh and D.A. Steinhauer, Stalk-pore hypothesis for membrane fusion mediated by influenza haemaggluttinin (HA), Expert reviews in molecular medicine, 2001, Cambridge University Press, Figure 6

46. P.J. Branigan, C. Liu, N.D. Day, L.L. Gutshall, R.T. Sarisky, A.M. Del Vecchio, Use of a novel cell-based fusion reporter assay to explore the host range of human respiratory syncytial virus F protein, Virol. J., 2005, 2, 54-66

47. L.J. Calder, L. Gonzalez-Reyes, B. Garcia-Barreno, S.A. Wharton, J.J. Skehel, D.C. Wiley, J.A. Melero, Electron microscopy of the human respiratory syncytial virus fusion protein and complexes that it forms with monoclonal antibodies, Virology, 2000, 271, 122-131

48. P.J. Branigan, C. Liu, N.D. Day, L.L. Gutshall, R.T. Sarisky, A.M. Del Vecchio, Use of a novel cell-based fusion reporter assay to explore the host range of human respiratory syncytial virus F protein, Virol. J., 2005, 2, 54-66

49. B.R. Lentz, V. Malinin, M.E. Haque, K. Evans, Protein machines and lipid assemblies: current views of cell membrane fusion, Curr. Opin. Struc. Biol., 2000, 10, 607-615

50. E.H. Fleming, A.A. Kolokoltsov, R.A. Davey, J.E. Nichols, and N.J. Roberts, Jr., Respiratory Syncytial Virus F envelope protein associates with lipid rafts without a requirement for other virus proteins, J. Virol., 2006, 80, 12160–12170

51. I.G. Cuesta, A. Xuehui Asnenjo, N. Villanueva, Structural Phosphoprotein M2-1 of the Human Respiratory Syncytial Virus is an RNA Binding Protein, J. Virol., 2000, 74, 9858-9867

52. Tran T., Castagné N., Dubosclared V., Noinville S., Koch E., Moudjou M., Henry C., Bernard J., Yeo R.P., Eléouët J., The Respiratory Syncytial Virus M2-1 Protein forms tetramers and interacts with RNA and P in a competitive manner, J. Virol., 2009, 83, 6363-6374

53. P. S. Gould and A.J. Easton, Coupled translation of the second open reading frame of M2 mRNA is sequence dependant and differs significantly within the subfamily pneumovirinae, J. Virol. 2007, 81, 8488-8496

54. O. Dibben, L.C. Thorpe, A.J. Easton, Roles of the PVM M2-1, M2-2 and P gene ORF2 (P-2) proteins in viral replication, Virus Res., 2008, 131, 47-53

55. R.W. Hardy, G.W. Wertz, The Cys_3-His_1 Motif of the Respiratory syncytial virus M2-1 protein is essential for protein function, J. Virol., 2000, 74, 588-5885

56. H. Zhou, X. Cheng, H. Jin, Identification of amino acids that are critical to the processivity function of Respiratory Syncytial Virus M2-1 Protein, J. Virol. 2003, 77, 5046-5053

57. J. Garcia, B. Garcia-Barreno, A. Vivo, J.A. Melero, Cytoplasmic inclusions of respiratory syncytial virus-infected cells: Formation of inclusion bodies in transfected cells that coexpress the nucleoprotein, the phosphoprotein and the 22K protein, Virology, 1993, 195, 243-247

58. A. Asenjo, E. Calvo, N. Villanueva, Phosphorylation of human respiratory syncytial virus P protein at threonine 108 controls its interaction with the M2-1 protein in the viral RNA polymerase complex, J. Gen. Virol. 2006, 87, 3637-3642
59. D. Li., D.A. Jans, P.G. Bardin, J.Meanger, J. Mills, R. Ghildyal, Association of respiratory syncytial virus M protein with viral nucleocapsids is mediated by the M2-1 protein, J. Virol. 2008, 82, 8863-8870
60. K. Walravens, J. P. Matheise, I. Knott, P. Coppe, A. Collard, C. Didembourg, F. Dessy, R. Kettmann, and J.J. Letesson, Immunological response of mice to the bovine respiratory syncytial virus fusion glycoprotein expressed in recombinant baculovirus infected insect cells, Arch. Virol., 1996, 141, 2313-2326
61. M. W. Wathen, R. J. Brideau, D. R. Thomsen, Immunization of Cotton Rats with the Human Respiratory Syncytial Virus F Glycoprotein Produced Using a Baculovirus Vector, Source: J. Infect. Dis., 1989, 159, 255-264
62. Bac-to-Bac® Baculovirus Expression System An efficient site-specific transposition system to generate baculovirus for high-level expression of recombinant proteins, User Manual, Version E, January 2009, Invitrogen
63. D.C. Tessier, D.Y. Thomas, H.E. Khourib, F. Lalihert and T. Vernet, Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide, Gene, 1991, 98, 177-183
64. QIAquick® Spin Handbook, Second Edition, March 2001, Quiagen, p. 19-20, 3-25
65. QIAprep® Miniprep Handbook, Second Edition, December 2006, Quiagen, p.22-23
66. Champion™ pET Directional TOPO® Expression Kits User Manual, June 2010, Invitrogen
67. Zero Blunt TOPO PCR Cloning Kit for Sequencing: Five-minute cloning of blunt-end PCR products for sequencing User Manual, Version N, July 2004, Invitrogen
68. J. Sambrook and D.W. Russell, Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory Press, New York, 3rd edn., 2001
69. Amino Acids and Peptide Synthesis (Oxford Chemistry Primers), Oxford University Press, 1992
70. S-Y Han and Y-A Kim, Recent developments of peptide coupling reagents in organic synthesis, Tetrahedron, 2004, 60, 2447-2467
71. M. Manning, Impact of the Merrifield Solid Phase Method on the design and synthesis of selective agonists and antagonists of oxytocin and vasopressin: A historical perspective, Biopolymers, 2008, 90, 203-212
72. R.B. Merrifield, Solid Phase Peptide Synthesis I: The Synthesis of a Tetrapeptide, J. Am. Chem. Soc., 1963, 85, 2149-2154
73. R.B. Merrifield, Solid-Phase Peptide Synthesis II: The synthesis of Bradykinin, J. Am. Chem. Soc., 1964, 86, 304-305
74. R.B. Merrifield, Solid-Phase Peptide Synthesis III: An Improved Synthesis of Bradykinin, Biochemistry, 1964, 3, 1385-1390
75. E. Atherton, D.L.J. Clive and R.C. Sheppard, Polyamide Supports for Peptide Synthesis, J. Chem. Soc.,1975, 97, 6584-6585
76. C. Dian, P. Baráth, R. Knaust, S. McSweeney, E. Moss, M. Hristova, V. Ambros, and D. Birse, Overcoming protein instability problems during fusion protein cleavage, Life Science News 10, 2002, Amersham Biosciences
77. D. Gründemann and E.Schömig, Protection of DNA During Preparative Agarose Gel Electrophoresis Against Damage Induced by Ultraviolet Light, 1996, BioTechniques, 21, 898-903.
78. Brunk, C.F. and L. Simpson, Comparison of various ultraviolet sources for fluorescent detection of ethidium bromide-DNA complexes in polyacrylamide gels. Anal. Biochem., 1977, 82, 455-462.
79. EasyXtal® and NeXtal® Protein Crystallization Handbook, January 2010 Qiagen
80. RhoA Interacts with the Fusion Glycoprotein of Respiratory Syncytial Virus and Facilitates Virus-Induced Syncytium Formation,1999, J. Virol, 73, 7262-7270
81. N.S. Berrow, D. Alderton, S. Sainsbury, J. Nettleship, R. Assenberg, N. Rahman1, D.I. Stuart, and R. J. Owens, A versatile ligation-independent cloning method suitable for high throughput expression screening applications, Nucleic Acids Res., 2007, 35, e45
82. M.J. Breaugh, PCR Cloning Protocols, Humana Press, 2nd edn., 2002
83. C. Aslanidis and P.J. de Jong, Ligation-independent cloning of PCR products (LIC-PCR), Nucleic Acids Res., 1990, 18, 6069-6074

84. In-Fusion® Advantage PCR Cloning Kit User Manual, Clontech, August 2010
85. R. Zufferey, T. Dull, R.J. Mandel, A. Bukovsky, D. Quiroz, L. Naldini, D. Trono, Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery, J Virol., 1998, 72 9873-80.
86. M. Lambert, S. Barney, A.L. Lambert, K. Guthrie, R. Medinas, D.E. Davis, T. Bucy, J. Erickson, G. Merutka, S.R. Petteway Jr, Peptides from conserved regions of paramyxovirus fusion (F) proteins are potent inhibitors of viral fusion, Proc. Natl. Acad. Sci. U. S. A., 1996, 93, 2186-2191
87. M.S. Bouckhvalova, G.A. Prince, J. C.G. Blanco, Inactivation of syncytial virus by zinc finger reactive compounds, Virology J., 2010, 7, Article 20
88. pMIB/V5-His A, B, and C Vector Kit For the Selection of Transfected Cells and Stable Expression of Secreted Heterologous Proteins in Lepidopteran Insect Cell Lines, User Manual, Version D, July 2008, Invitrogen