PROBING PROTEIN-PROTEIN INTERACTIONS FROM HUMAN RESPIRATORY SYNCYTIAL VIRUS.

HAYLEY JOY LUMB

PROBING PROTEIN-PROTEIN INTERACTIONS FROM HUMAN RESPIRATORY SYNCYTIAL VIRUS.

HAYLEY J LUMB

THESIS SUBMITTED TO THE DEPARTMENT OF CHEMISTRY FOR THE PARTIAL FULFILLMENT OF REQUIREMENTS OF THE UNIVERSITY OF DURHAM FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

MAY 2015.

DEDICATION

To my late Dan Dan Lumb and Nana Lumb, you always knew I could do this and saw the potential in me from a young age. I'm sorry you couldn't see me following my dreams, but I know you would be proud.

To my late Father, your eternal words of encouragement and support still linger on. Thank you for believing in me, you will never be forgotten.

"*Gone with the wave*"

Bernard J.J Lumb

1958-2013

DECLARATION

All work described in this thesis was undertaken at Durham University between October 2011 and April 2015. All work presented is the Authors unless otherwise stated. No part of this work presented herein has previously been submitted for a degree at this, or any other university.

CONTENTS

ACKNOWLEDGEMENTS

I would like to thank several people on this page, most importantly my supervisor, Dr V A Money. Her door was always open and without her constant guidance and support throughout the years at Durham, I certainly wouldn't be where I am today.

To all the various Durham academics I have worked with, Dr J.M Sanderson, Dr R.P Yeo and Dr S Cockerhill especially. Many thanks to Dr J.M Sanderson for the data analysis in Chapters 3 & 4. I never knew where we were going next and it definitely made the projects exciting and interesting to work on.

To Lab 209/229, I know I'm not the tidiest person and thanks for all of the help and times in the pub 'talking through a problem'. You made me feel welcome and I will miss you all. Special thanks to Ian Edwards, your help with various techniques and crystallisation trays was most appreciated.

Thanks to N. J. Tatum for helping and tutoring me through learning various computer programs. You made the process much smoother and your advice for starting to dock ligands was most appreciated.

To O. Neale, E. Antonio, R. G Stewart, H. Yamada and L. Argent, thank you for your hard work and dedication in your projects. Your results formed the basis of many experiments and chapters contained in this thesis.

To my Mum, Alan and my sisters, thanks for your never-ending support and house moving skills. I know there will always be a glass of wine and a good chat if I ever needed it.

To Rob, thanks for the coffee, support and the 'odd' pint or 2 - let's see what the next 12 months bring and beyond!

To Matt, I love you and you really have ridden this rollercoaster with me! Thank you so much for your support and strength.

And to my Dad, I miss you and thank you for your support while you were here. I hope I've made you proud.

ABSTRACT

Human respiratory syncytial virus (hRSV) is the leading cause of viral respiratory tract infections in children.[1] In 2013 there were over 60 000 infants hospitalised in the USA alone and the virus resulted in approximately 2.1 million outpatient visits by children under 5 years old.[2] The virus leaves the lungs weakened and open to further secondary infections such as bronchitis and pneumonia, especially in immunocompromised and elderly patients. Infections can lead to long-term effects and up to 70% of infants are left with respiratory problems for up to 10 years following hRSV bronchiolitis.[3] For otherwise healthy children the treatment is usually supportive, using hydration and additional oxygen.[1] Previous antiviral agents and decongestants have been used to little positive effect, leading to the need for new antiviral agents.

In the work described in this thesis, key protein targets have been identified from the 11 protein hRSV Genomes. It has been found that F, M2-1, M and N are integral for structural roles, transcription and virus maturation.[4] M2-1 protein is a transcription factor and has long been a target for structural biologists due to its integral role in the virus. Chapter 2 describes the characterisation of the protein using biophysical techniques to obtain data on its stability and overall fold. Further work to crystallise the protein is also included.[5]

Chapter 3 builds on this work introducing the interaction between matrix protein and M2-1. The interaction is characterised through Surface Plasmon Resonance (SPR) and Fluorescence Anisotropy (FA), which gave equilibrium constants and thermodynamic pathways for the first time.

Chapter 4 introduces the interaction between hRSV matrix protein and hRSV fusion protein. Previous work has characterised the fusion protein and identified that the cytoplasmic tail is the key subunit which binds to matrix protein.[6] Through mutagenesis of a single residue, F572 was identified as the key binding residue.[7] The terminal six amino acids were used to form a peptide which formed the basis for the experiments in this thesis. The work described here presents *in vitro* data which will build on the published literature to further understand the interaction. SPR and FA provide experimental evidence for a hypothetical binding model. This binding model and the resulting equilibrium constants are discussed and evaluated.

The X-ray crystal structure of hRSV N protein is known[8,9] and this information can be used to design ligands with the aim of disrupting the protein's function. The work outlined here aimed to evaluate the binding of a series of ligands that were identified by Arrow Therapeutics Ltd as having antiviral activity.[10] Thermal shift assays were used to understand the stabilisation of the protein in the presence of the ligand. Dynamic light scattering showed the formation of aggregates of protein when exposed to various concentrations of ligand. Cryo-electron microscopy shows association of the ligand to the protein and the consequent aggregates which form. This provided an indication of how the ligands could interact with the protein and suggests possible modes of action.

In Chapter 6, hRSV M protein's interaction with hRSV F was further investigated. The interaction is key for the maturation of the virus[11] and if the binding could be diminished, the infectivity of the virus would suffer. A pocket at the dimer interface was identified, which opened up an opportunity for computational docking studies. Through the application of biophysical techniques and computational methods,[12] novel ligands were identified and the initial characterisation of their effect on protein stability was probed.

Key protein interactions which are involved in the maturation of the virus and its fusion to the host cell have been characterised and evaluated in this thesis. The opening chapters begin within basic protein characterisation and build to show protein interactions *in vitro*. In the later chapters, ligand subsets are identified with the aim of disrupting protein interactions or the protein folding. Characterisation of the protein ligand interactions showed protein changes and destabilisation, leading the way for further studies.

LIST OF FIGURES

CHAPTER 1. INTRODUCTION TO VIRUSES

CHAPTER 2. CHARACTERISING M2-1 PROTEIN

CHAPTER 3. PROBING MATRIX PROTEIN AND M2-1 INTERACTION

CHAPTER 4. INTERACTION OF THE MATRIX PROTEIN WITH CYTOPLASMIC TAIL OF THE FUSION PROTEIN

CHAPTER 5. NUCLEOCAPSID PROTEIN INTERACTION WITH RSV-604 AND RELATED COMPOUNDS

CHAPTER 6. IDENTIFICATION AND DOCKING DRUG LIKE MOLECULES INTO THE MATRIX PROTEIN OF HRSV

LIST OF TABLES

LIST OF ABBREVIATIONS

Borna Disease Virus	BDV
Clathrin-coated vesicles	CCVs
Circular Dichroism	CD
Covalent circle DNA	CCC-DNA
C Terminal domain	CTD
Dynamic Light Scattering	DLS
Deoxyribonucleic acid	DNA
Ebola Virus	EBOV
Effective dose for 50% population	EC50
Ethylenediaminetetraacetic acid	EDTA
Electron microscopy	EM
Fusion Protein	F
Fluorescence resonance energy transfer	FRET
Viral attachement glycoprotein	G
Ebola attachment proteins 1,2	GP 1,2
Glutathione sepharose Transferase	GST
Histidine tagged matrix protein	M(His6)
Human immunodeficency virus	HIV
Human Metapneumovirus	hMPV
Human Respiratory Syncytial Virus	hRSV
Inductively coupled plasma mass spectrometry	ICP-MS
International taxonomy of virus	ICTV
Isopropyl β-D-1-thiogalactopyranoside	iPTG
Association constant	ka
Dissociation constant	kd
L protein, viral RNA polymerase	L
Lethal dose for 50% of population	LC50
Matrix protein	M
Matrix assisted laser desorption ionisation –time of flight mass spectrometry	MALDI-TOF MS
Messenger RNA	mRNA
Mass Spectrometry	MS
Molecular weight	MW
Viral nucleocapsid protein	N
Newcastle disease Virus	NDV
Non structural protein 1&2	NS 1&2
Nitrolotriacetic acid	NTA
N Terminal domain	NTD
Open reading frame	ORF
Viral phosphoprotein	P
Phosphate buffered saline	PBS
Protein data bank	PDB
Root mean squared	RMS
Root Mean Squared Deviation	RMSD
Ribonucleic acid	RNA
Negative sense RNA	(-)RNA
Ribonucleocapsid (RNA complexed with proteins)	RNP
Self-assembled monolayer	SAM
Sodium dodecyl sulphate	SDS
Sodium Dodecyl sulphate polyacrylamide gel electrophoresis	SDS PAGE

non structural membrane protein	SH
Surface Plasmon Resonance	SPR
Transmission Electron microscopy	TEM
N,N,N',n'-tetramethylethylenediamine	TEMED
Tris(hydroxymethyl)aminomethane	TRIS
Unitied States of America	U.S.A.
Viral Hemorrhagic Fever	VHF
Ebola transcription factor protein	VP30
Matrix protein from Ebola or MArburg virus	VP40
Vesicular stomatitus Virus	VSV
World heath organisation	WHO
(2,3-Bis-(2-methyl-4-nitro-5-sulphonpheyl)-2H-tetrazolium-5-carboxcindide) assay	XTT

UNITS

Degrees of arc	°
Degrees celcius	°C
Daltons	Da
Kilo bases	kb
Kilo Daltons (10^3 Daltons)	kDa
Centimetres (10^{-2} metres)	cm
Millimetres (10^{-3} metres)	mm
Micrometres (10^{-6} metres)	μm
Nanometres (10^{-9} metres)	nm
Ångstroms	Å
Litres	L
Millilitres (10^{-3} Litres)	mL
Microlitres (10^{-6} Litres)	μL
Molarity (Moles in a litre – mol/dm^3)	M
Molarity (Millimoles in a litre – mol/dm^3)	mM
Molarity (Micromoles in a litre – mol/dm^3)	μM
Seconds	s
Volts	V

AMINO ACIDS

CHARGED:

Arginine	Arg - R
Lysine	Lys - K
Aspartic acid	Asp - D
Glutamic acid	Glu - E

POLAR (MAY PARTICIPATE IN HYDROGEN BONDS):

Glutamine	Gln - Q
Asparagine	Asn - N
Histidine	His - H
Serine	Ser - S
Threonine	Thr - T

Tyrosine	Tyr - Y
Cysteine	Cys - C
Methionine	Met - M
Tryptophan	Trp - W

HYDROPHOBIC (NORMALLY BURIED INSIDE THE PROTEIN CORE):

Alanine	Ala - A
Isoleucine	Ile - I
Leucine	Leu - L
Phenylalanine	Phe - F
Valine	Val - V
Proline	Pro - P
Glycine	Gly - G

1. INTRODUCTION TO VIRUSES.

The scope of this thesis covers the biophysical investigation of key structural proteins from human Respiratory Syncytial Virus (hRSV) with a view to understanding interactions and functions. This chapter will summarise the history and significance of viruses, primarily focusing on the *Mononegavirales* order, of which hRSV is a member, and the negative sense RNA viruses.[13] An overview of the relevance of human Respiratory Syncytial Virus (hRSV) will bring context to the following work and the prevalence and impact of the virus will be discussed. The key structural proteins studied interact within the virus and without these interactions the infectivity of the virus and even its shape would be different.[14,15] The replication of enveloped viruses has been poorly understood in comparison to capsidated viruses which can hinder critical research into the interaction of their structural proteins.[14] Each protein studied will be introduced later in this chapter for further familiarisation.

Following initial biophysical testing the complex task of solving the X-ray crystal structure of the protein must take place to find key residues which could be critical for the interaction of proteins within the virus.[16,17] Further chapters will investigate the interaction of key structural proteins by in depth biophysical studies. From this stage the three-dimensional structures can be used to identify new anti-viral drug target pockets and novel scaffold targets. This computational technique uses complementary small drug like fragments to target the interface and so disrupt the formation of the complex and is outlined in Chapter 6.

1.1 VIRUSES

Viral infections can range from the common cold to deadly diseases such as Ebola or AIDS. Virology is a relatively young science; only 100 years ago the first virus was filtered and identified as being smaller than bacteria.[18] Within the past 60 years the compositions of a number of viruses have been described and their electron microscope images have become known.[18] With advances in modern biotechnology the structure, molecular biology of viral replication and the interactions that take place between the virus and host are being characterised and understood. Improved sanitation, safer water supplies and the development of antiviral drugs have dramatically reduced the threat from many of these viral agents especially in the developed world.[13] In 1998 in excess of 3 million deaths occurred as a result of acute respiratory diseases, many of which were caused by viruses.[19] Many viruses have no vaccine and continue to be of great medical significance.[18]

1.2 NEGATIVE SENSE RNA VIRUSES

The virus classification system is key to understanding how the proteins interact in the virus.[20] The Baltimore classification classes viruses according to their type of genome (RNA or DNA) and replication method.[13,21] Usually within a family, the proteins are similar in size and function and comparisons can be drawn across members.

Class V, of the Baltimore classification system, is a broad class of viruses, which consists of two orders and eight families.[21] There are also a number of unassigned species and genera. Many have different morphologies e.g. *Filoviridae* are filamentous and *Paramyxoviridae* are usually pleomorphic or round.[18] The order *Mononegavirales,* which belongs to Class V, consists of four families as shown in Table 1.2.1.[21]

Order - *Mononegavirales*	
Family	**Genus**
Bornaviridae	Bornavirus
Filoviridae	Ebola Virus, Marburg Virus
Paramyxoviridae	Measles Virus, Mumps Virus, Nipah Virus, Hendra Virus. hRSV. Metapneumovirus (hMPV)
Phabdoviridae	Phabdovirus

Table 1.2.1 Families of the order *Mononegavirales*[1,4]

There are also unassigned families included in Class V.Family *Arenaviridae,* which contains Lassa Virus. The Lassa virus causes viral hemorrhagic fever (VHF) in human and non-human primates. The family *Orthomyxoviridae*[22] which contains the influenza virus is also unclassified, but linked to the *Mononegavirales* order. The virus cannot make its own mRNA so uses cellular RNA as primers for initiating RNA synthesis in the cell.[18]

1.3 HUMAN RESPIRATORY SYNCYTIAL VIRUS (HRSV)

Human Respiratory Syncytial Virus is the leading cause of viral respiratory tract infections in children.[1] The World Health Organisation estimates an annual mortality rate of approximately 250,000 deaths worldwide in 2010.[23] Updated figures show that hospitalisation of ~60 000 infants a year in the USA is caused by hRSV and in 2013 alone the virus resulted in 2.1 million outpatients visits by children under 5 years old.[2] The virus, and subsequent secondary bacterial infections are also the second leading cause of death in elderly patients.[24] It has been estimated that 87% of children by the age of 18 months have been infected with hRSV and by the age of 3 nearly all have been infected.[19]

hRSV was discovered in 1955, though thought at the time to be a childhood disease the virus is now known to be a serious threat to both young and old.[25] In 2001 human metapneumovirus (hMPV) was discovered and began to be characterised, this was found to be another important cause of bronchiolitis and produces similar symptoms.[26] hMPV infects children up to the age of 5 years old and co-infection with hRSV can produce a more severe disease.[27] Symptoms of hRSV include rhinitis, cough and coryza, as the disease begins in the upper respiratory tract. In 25%-40% of cases the disease travels to the lower respiratory tract[18] and can be characterised by feeding difficulties in the infant amongst the visual signs of respiratory distress.[25] If bronchiolitis in the lower respiratory tract begins, a wheeze could be present which is characterised by a crackle in the chest.[28] Bronchiolitis can lead to acute respiratory failure with severe bronchospasm, moderate hypoxia and carbon dioxide retention.[25] Death from bronchiolitis is rare without underlying immunological conditions and most infected children recover fully.[27] Some infections can lead to long-term effects; up to 70% of infants are left with respiratory problems for up to 10 years following hRSV bronchiolitis and in studies this has been linked to the child developing asthma.[28] In healthy children treatment of hRSV symptoms is mainly supportive.[1] In the past bronchodilators, antiviral agents and decongestants have been used to little effect.[29] Hydration and oxygenation are now considered to be the first line of support. Where the child has bronchiolitis; dehydration, fever and poor feeding due to difficulty breathing are problematic. Intravenous fluids may be needed if the case is particularly severe.[25]

Two approved drugs are available for treatment of hRSV.[30] Palivizumab is a humanised monoclonal antibody[31] to the viral fusion (F) protein and is used as a preventative

treatment for high risk infants such as the immunocompromised and premature. Palivizumab reduces the need for hospital admissions but is costly, between £3000-£5000 to treat each child.[31] In the USA Palivizumab is recommended for premature babies born below 35 weeks' gestation and for infants under two who have chronic lung disease.[31] The health gains versus the cost are minimal and do not have a real impact on the complications which arise from hRSV. Ribavirin is also licensed for severe infections only and is potentially toxic to health care workers.[32] New research has now given evidence to suggest it is also teratogenic.[33] The drug is nebulised and inhaled but has limited therapeutic affect in patients. Ribavirin has a broad spectrum range of antiviral activity *in vitro* and obstructs the replication of hRSV, influenza, parainfluenza, adenovirus and measles.[32]

The lack of available treatments highlights the problems that are faced in the quest for antiviral drugs. High costs and a lack of relative effectiveness are some of the challenges posed. Many problems arise from the lack of knowledge of the virus in question, which highlights the need for basic research to find potential drug target leads. It is known that virus structure and protein interactions are key for function. To understand the structure-function relationship, clues may be found in related viruses that reside in the same family or subfamily and these will be discussed in detail later in this chapter.

1.3.1 STRUCTURE AND REPLICATION.

hRSV is a member of the *Paramyxoviridae* family as described in section 1.2. It is an enveloped, non-segmented, negative-stranded RNA virus in the order *Mononegavirales* as shown below in Figure 1.3.1.1.[1]

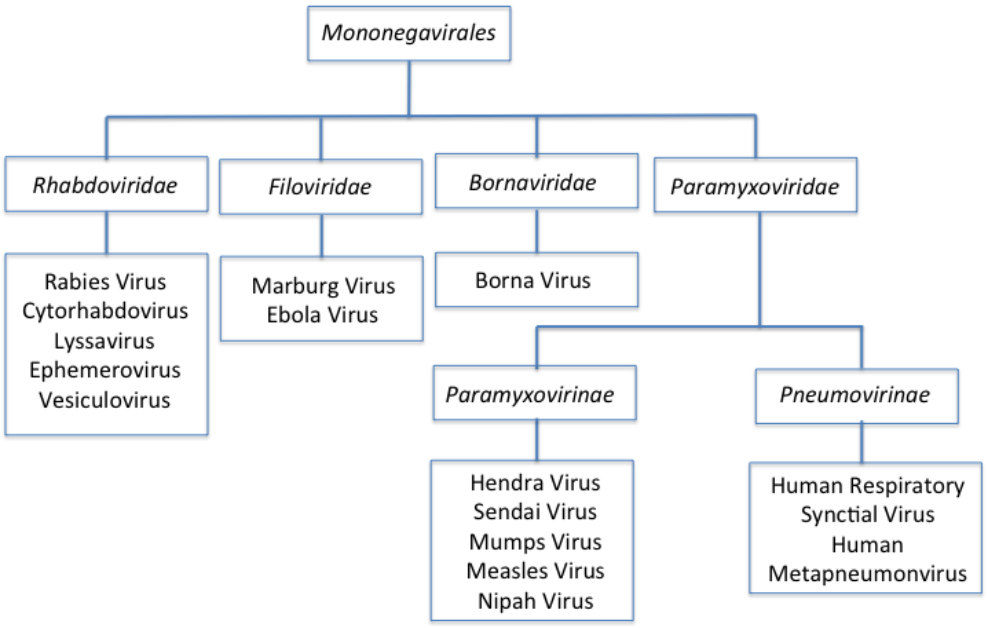


Figure 1.3.1.1. – *Mononegavirales* family tree[18]

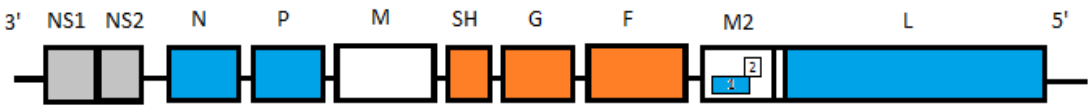


Figure 1.3.1.2. – Cartoon representation of the genome of hRSV[18,34]

The hRSV virion has a spherical or filamentous form and contains a helical nucleocapsid within the envelope, which is derived from host cell phospholipids and proteins.[18] The hRSV RNA genome encodes 11 proteins[11] as shown in Figure 1.3.1.2. The N protein forms the nucleocapsid, this is associated with phosphoprotein P, large polymerase subunit L and the transcription elongation factors M2-1 and M2-2. The viral envelope is studded with three transmembrane envelope glycoproteins, SH, G and F. F is the fusion protein, which is involved in viral attachment with the target cell. [34,6] The glycoprotein G has been shown to be non-essential for fusion and is heavily glycosylated.[35] SH, the small hydrophobic protein is a membrane protein, which is also studded into the side of the envelope.[35] The matrix protein (M) is a peripheral membrane protein that binds the viral envelope and nucleocapsid, and so holds them together forming a layer in between as shown in Figure 1.3.1.3.[36]

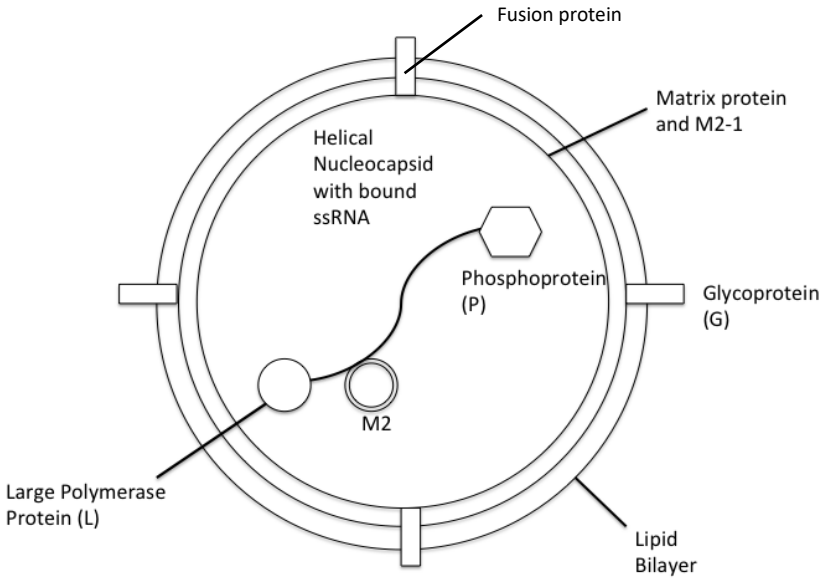


Figure 1.3.1.3.– Representation of hRSV[3]

The ultrastructure of the virus was characterised in the helical form through freeze-etching of virus filaments as shown in Figure 1.3.1.4 below. The core of the structures areexposed and shows the concave, inner surface of the viral envelope.[15]

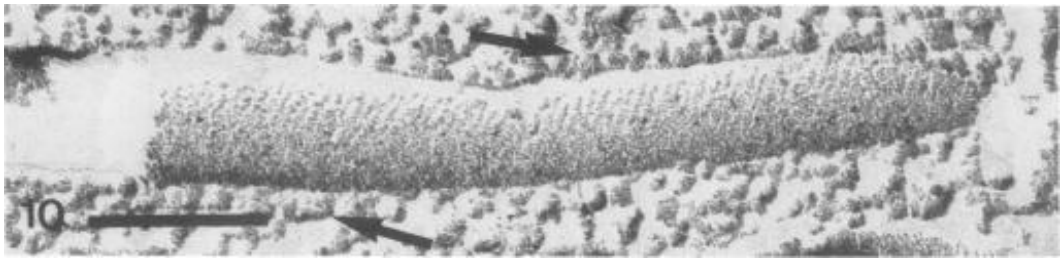


Figure 1.3.1.4 Filamentous virus filaments[15]

1.4 ENVELOPED VIRUSES

1.4.1. STRUCTURE AND REPLICATION CYCLE

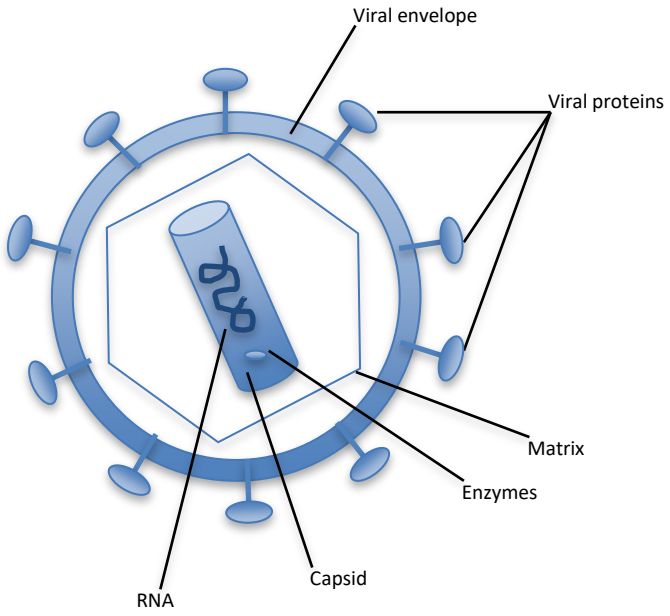


Figure 1.4.1.1 Structure of an enveloped virus.[37]

Enveloped viruses generally consist of three main parts; the genetic material, the capsid and the envelope containing other associated proteins as seen in Figure 1.4.1.1.[18] The primary function of the capsid is to protect the viral genome and ultimately the viral proteins.[13,15,20] The capsid also has other functions; these can include binding to the genome to aid the assembly of virions. This 'capsid-genome' complex is called the nucleocapsid.[18] In enveloped viruses the nucleocapsid may also interact directly with the envelope or proteins, which are bound to the envelope.[1,4,37]

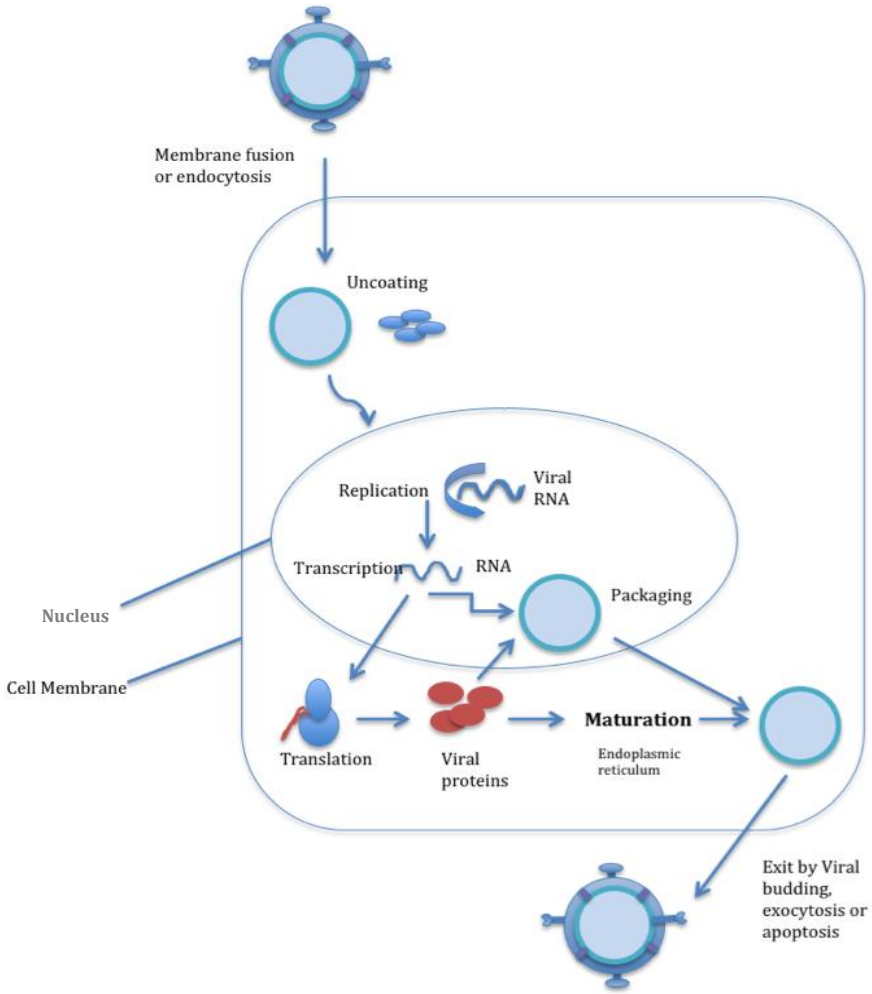


Figure 1.4.1.2 Replication cycle of an enveloped virus [18,38]

Viral entry is the first process during replication, this is achieved by binding to a cell surface receptor and inducing fusion between the envelope and the host cell membrane.[38] In hRSV, viral proteins (F and G) anchored in the envelope or the nucleocapsid assist this process.[39] This binding and fusion are followed by the injection of the viral genome into the cell. Enveloped viruses use membrane fusion of the envelope with the host plasma membrane.[38] One or more viral membrane proteins promote the various fusion steps as discussed below.[6]

1. **Prefusion** – The fusion protein is in prefusion conformation with fusion peptides in close proximity. The virus docks into a cell membrane receptor.
2. **Extended intermediate** – The fusion protein opens up and the fusion peptide interacts with the target bilayer.
3. **Collapse of the extended intermediate** – The C terminal segment of the fusion protein folds back, away from the core and the deformed membrane.
4. **Hemifusion** – The collapse of the intermediate brings the virus and host cell membrane bilayers into contact, and they merge by hemifusion. This is where the outer bilayers merge, but the inner stay intact.
5. **Fusion Pore formation** – The hemifused bilayers open into a fusion pore and the final segmentation of the C-terminal ectodomain of the fusion protein snaps the refolded trimer into a fully symmetric, post-fusion conformation preventing the pore from resealing. The genomic material enters into the cell through this pore.

Endocytosis, which is used to complete viral entry, usually transports fluids and small cell particles in animal cells, and there are multiple mechanisms available.[40] Host cell receptors have been shown to facilitate different mechanisms of viral entry by endocytosis. Once inside the cell the nucleocapsid is broken down through degradation by viral or host proteases, this leaves the virus free to replicate inside the host cell.[40] RNA viruses replicate in the cytoplasm or endoplasmic reticulum of a cell whereas the DNA virus will continue to the nucleus.[9]

The mechanism of formation of new virions is different in each class of virus, but there are common steps in all. First the DNA or RNA is replicated by polymerase enzymes, mRNA is then transcribed and used to produce viral proteins using the host cell's ribosomes.[18,39] The new viral proteins and genetic material can then be assembled to form progeny virions, which gather near the inside of the cell membrane.[18,41]

Instead of cell lysis, which is the mode of exit for non-enveloped viruses, the enveloped viruses use budding to exit the host cell.[14,42] This requires viral proteins to obtain part of the host-derived membrane to form an external envelope. The mechanism begins when the viral core components are incorporated into membrane vesicles, which also contain viral transmembrane proteins.[14] The nucleocapsid formation induces curvature of the host cell membrane, to form the distinctive 'bud' that wraps up the viral proteins that is

excised from the host membrane to release the enveloped particle.[39,41,43,44] Finally, the new virus infects another host cell and the cycle is repeated. The replication cycle is key to understanding viral infectivity.

1.5 hRSV MATRIX PROTEIN

1.5.1 STRUCTURE

The matrix protein (M) is a non-glycosylated phosphorylated protein of 256 amino acids.[45] It is a structural protein, which plays a fundamental role in virion assembly by hindering viral transcription and promoting budding. The protein also binds to the viral ribonucleocapsid and to the envelope, the latter of which is an important stage in virion release.[6] The binding is stabilised by viral glycoproteins and co-localisation with the F protein and SH proteins.[46]

The matrix protein is smaller than those of other paramyxoviruses, at 256 amino acids in length compared to 335-375 amino acids.[6] There is little sequence homology between the proteins in the family, but like its rhabdovirus equivalents, i.e. Rabies Virus Matrix, hRSV Matrix protein can bind RNA.[47] The RNA binding region is located between amino acid residues 120-170. There is a number of residues in this region specifically: lysine residues 121, 130, 156 and 157 which are conserved throughout the pneumovirus M proteins and which are known to be essential for RNA contact shown in Figure 1.5.1.1 below.[47]

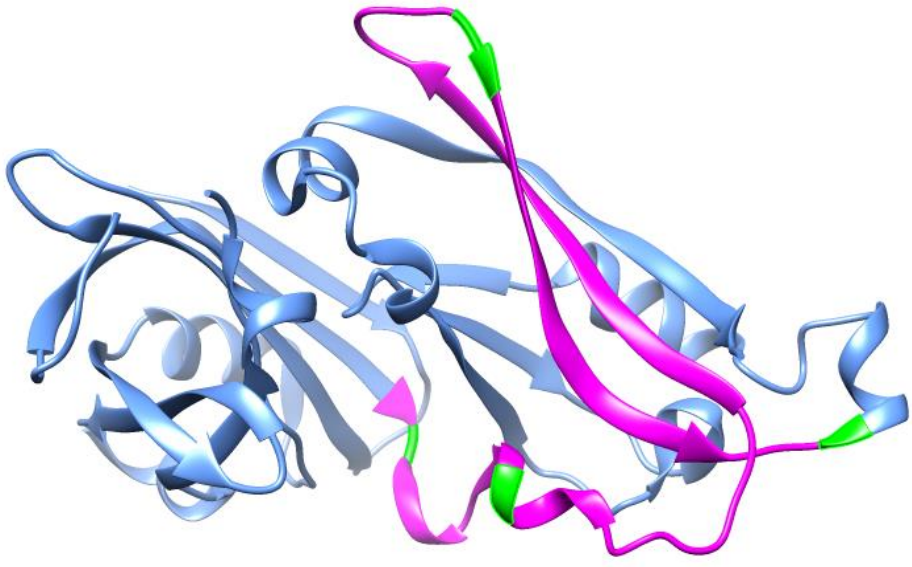


Figure 1.5.1.1 - Matrix protein monomer[48] (PDB code 2VQP) RNA binding region highlighted in magenta, specifically conserved residues highlighted in green, Lysine 121, 130, 156 and 157.

There are two domains per monomer, which are connected by a 13-residue flexible linker region and an unstructured loop region.[48] The N-terminal domain (residues 1-126) comprises a twisted β-sandwich, which consists of two β-sheets which are positioned nearly perpendicular to each other. The overall topology of the N-terminal domain is a curved horseshoe-like arrangement with 1 β-sheet forming the concave outer face.[48] The C-terminal domain (residues 140-255) consists of a flattened β-barrel, which is made up of two three-stranded anti-parallel β-sheets. The regions linking strands 2 and 3 and strands 5 and 6 are mostly helical in nature.

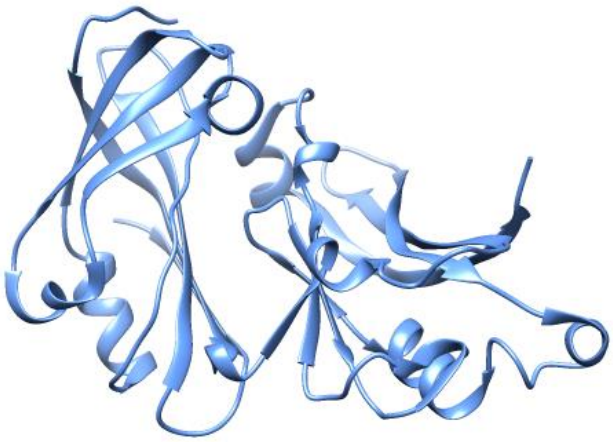


Figure 1.5.1.2. Ribbon diagram of hRSV Matrix protein.[48] (PDB code 2VQP)

The crystal structure of the protein was solved in 2009 and interpreted as a monomeric structure. Recently the structure for the hRSV Matrix protein was determined to be a dimer by Förster *et al*.[49] shown in Figure 1.5.1.3. This new structure gives definition to previously flexible regions and a clear dimerization interface. This dimerization interface is the same as other dimeric M structures from the *Mononegavirales* including Newcastle disease Virus (NDV)[39] and human Metapneumonovirus (hMPV).[50]

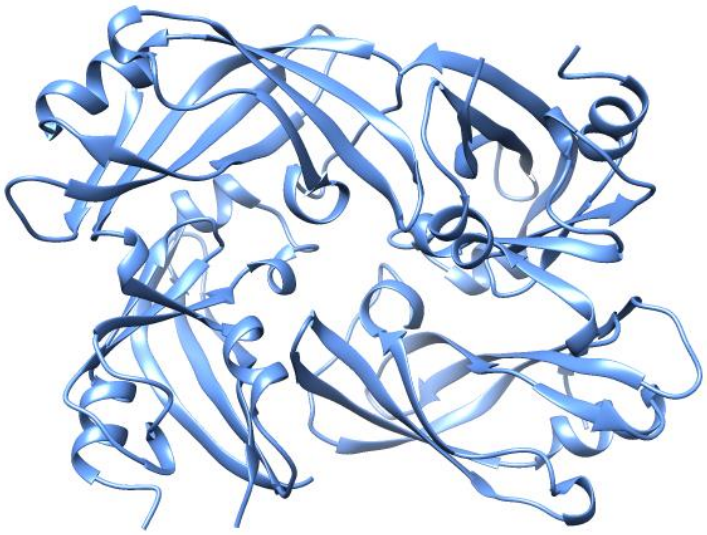


Figure 1.5.1.3 hRSV matrix protein dimer[49] (PDB code: 4D4T)

1.5.2. COMPARISONS TO OTHER MONONEGAVIRALES MATRIX PROTEINS.

In the order *Mononegavirales,* a number of matrix protein X-ray crystal structures have been solved.[39,45,51,52] This section will introduce structurally similar matrix proteins and discuss the similarities and differences between them and hRSV Matrix protein. Key conserved areas and differences could be the basis to targeting residues for antiviral agents to inhibit protein-protein interactions. The proteins discussed are from across the *Mononegavirales* order and include; VP40 from the Ebola (EBOV) virus, Borna (BDV) Virus M, Vesicular Stomatitis Virus (VSV) M and Human Metapneumonavirus (hMPV) M.

1.5.2.1. VP40 EBOLA VIRUS (EBOV)

VP40, the Ebola matrix protein, has much the same function as hRSV M, as it is also involved with virus assembly and budding. The C-terminal domain has similar membrane targeting properties to those of hRSV M, this is linked to its function within the budding of the virus.[51,53] The N-terminal domain is thought to be responsible for the hexamerisation of the protein into a lattice *in vivo,* much like that of HIV matrix protein in low resolution electron microscopy images.[54] The EBOV primary sequence comparisons with matrix proteins from other families in the order *Mononegavirales* show that the there is no significant homology between the proteins.[48,50,51] *Paramyxoviridae* is the only other family that contains a matrix protein of a similar size to Ebola virus VP40. Other families in the order are generally approximately 100 amino acid shorter.

VP40 has been crystallised as a monomer[51] and also a dimer but it can oligermerise *in vitro* to form octamers when it binds RNA.[55] The oligomerisation state of the protein *in vivo* is dependent on the sub-type of Ebola from which it originates whether that be Zaire or Sudan.[56]

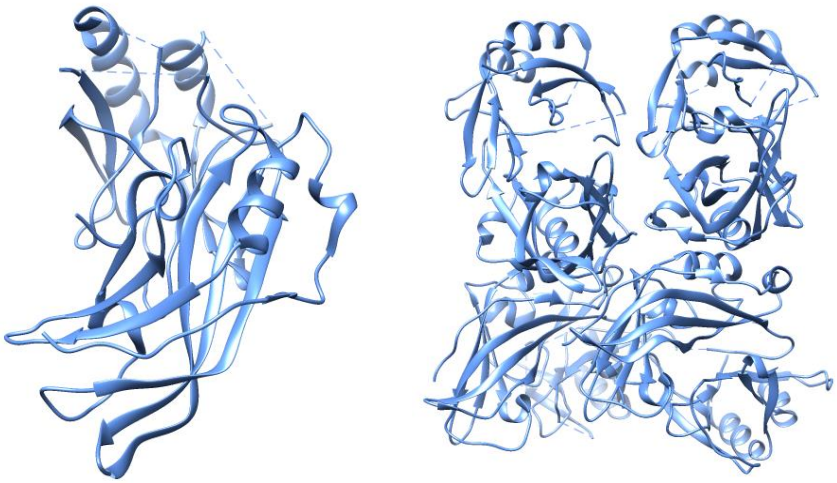


Figure 1.5.2.1.1. Left, Ebola Virus VP40 Matrix protein monomer (PDB code: 3TCQ).[51]

Right, Dimer Ebola Virus VP40[56] (PDB: 4DLB)

1.5.2.2. BORNA VIRUS (BDV)

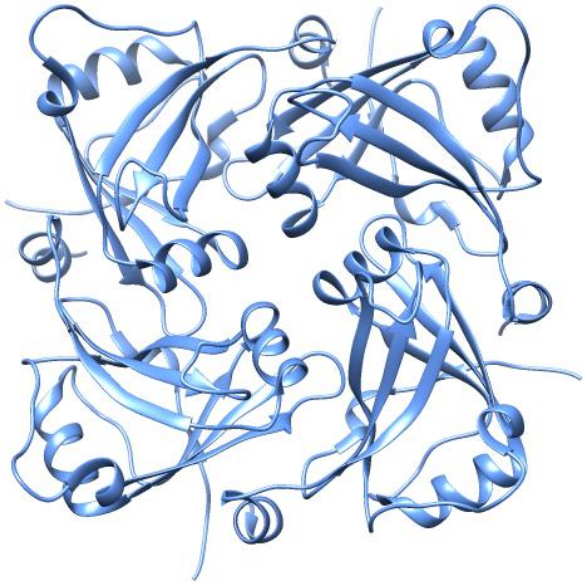


Figure 1.5.2.2.1. Structure of matrix protein from Borna Virus[57] (PDB: 3F1J)

The BDV-M protein is the smallest M protein at 16.2 kDa[57] in the order *Mononegavirales*. It forms oligomers *in vivo* and *in vitro* where a tetramer is the preferred oligomerisation state as shown in Figure 1.5.2.2.1. The Borna Virus M has some structural analogy with hRSV M, but is most similar to the N-terminus of VP40 from Ebola as shown in Figure 1.5.2.1.1. The overall fold has a L-shaped β-sandwich formation containing six antiparallel strands[57] that is similar to the overall fold of hRSV M, especially at the N-terminus, which has been shown in hRSV M to directly bind M2-1.[58]

1.5.2.3. VESICULAR STOMATITIS INDIANA VIRUS (VSV)

Vesicular Stomatitis virus, in the family *Rhabdovirus*, has been used as the prototype for budding within the order.[52] The VSV M protein is 26.6 kDa in size and the core 80 amino acids of the protein were first crystallised in 2002.[52] The whole protein fold is similar to hRSV M where there are two distinct domains, connected by a 20 amino acid linker region. In hRSV M, a slightly smaller 13 amino acid linker region is present.[52] The N-terminal region of VSV M is positively charged and contains a five-stranded anti-parallel β-sheet with two α-helices.[52] The C-terminal region has two small anti-parallel β-sheets and one α-helix. The stable core of the structure is shown below in Figure 1.5.2.3.1 containing amino acids 40-120.

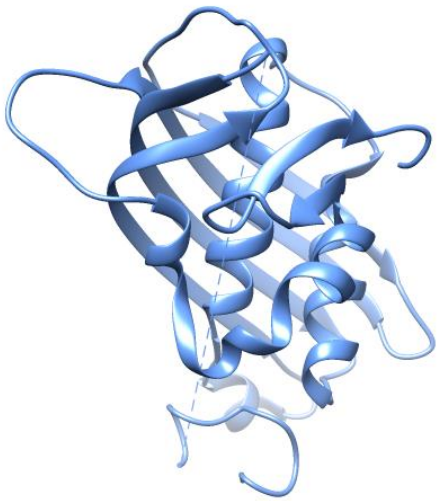


Figure 1.5.2.3.1. Structure of the core of VSV M.[52] (PDB: 1LG7) Containing amino acids 40-120

1.5.2.4. HUMAN METAPNEUMONAVIRUS (HMPV)

Human Metapneumonavirus (hMPV) belongs to the same subfamily as hRSV. The crystallised form of the hMPV matrix protein is a dimer much like hRSV M. The Borna virus matrix proteins have been shown to form dimers and tetramers with a similar quaternary diamond shape.[50] Unlike hRSV M, the hMPV matrix protein also has calcium ions bound in the crystal structure.[50]

Each monomer of hMPV M is comprised of two similarly folded domains joined by a 14 amino acid linker, an arrangement that has been found across the order in various sizes from 13 amino acids to 20 amino acids. The dimeric interface contains contacts between the N-Terminal domain and the C-Terminal domain. Each N-Terminal domain contains a twisted β sandwich, much like those in hRSV M.[48,50] The C-Terminal domain has disorder in solution and in the crystal form residues 180-188 and residues 208-2018 are not visible in the electron density.

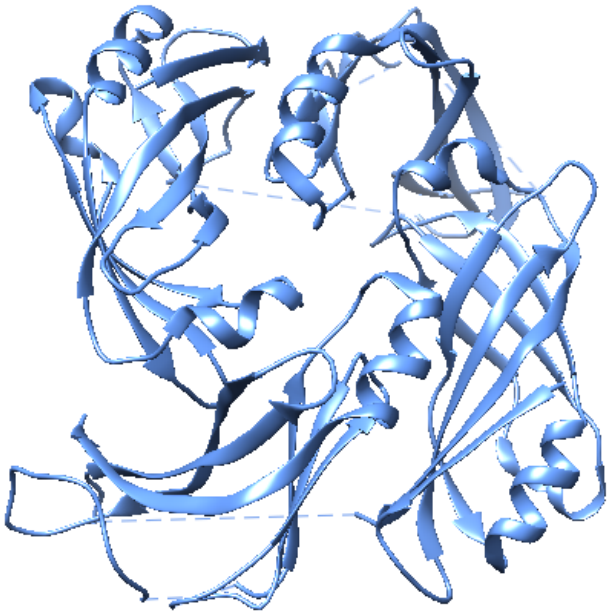


Figure 1.5.2.4.1. The dimer structure of hMPV M.[50]

1.6 M2-1 PROTEIN

1.6.1 STRUCTURE AND FUNCTION

The M2-1 protein is a transcription anti-termination factor which is key for transcription of full length mRNAs.[59] The M2 gene codes for a mRNA containing two open reading frames (ORFs). ORF1 codes for a 194 amino acid structural protein, M2-1, the second ORF codes for M2-2 a 90 amino acid protein.[60] M2-1 enhances the synthesis of polycistronic mRNAs and is an essential cofactor of the viral RNA polymerase complex. It has also been found in a mono-phosphorylated and unphosphorylated forms in infected human cells.[46] The protein is only present in the families *Pneumovirinae* and *Filoviridae* i.e. hRSV, Metapneumovirus and Ebola virus, which makes the protein a potential target for a specific antiviral agent.[61]

hRSV M2-1 forms a tight tetramer of 89 kDa in solution and has been found to be highly α-helical by circular dichroism.[62] Through dynamic light scattering M2-1 is a 5.4S tetramer and approximately 7.6 nm in diameter.[62] The N-terminus contains a Cys3-His1 motif, where the 3 cysteine residues and one histidine residue coordinate to a single zinc ion. This zinc-binding domain is essential for M2-1 function within the virus, this is also shown in the equivalent Ebola VP30 protein and is highly conserved throughout the genus.[60,62,61] The zinc-binding domain has been found to support the secondary structure and aids in the protein's ability to form a tetramer in solution.[62] Mutation of the zinc coordinating residues of the Cys3-His1 motif removes the effectiveness of M2-1 as a transcription factor.[61]

M2-1 protein has been shown to bind in a sequence specific manner to short RNAs which contain a 5'-end leader sequence.[5] The protein was also shown to bind long RNAs without any sequence specificity at a K_d of approximately 30 μM.[5,63] To determine the specificity of M2-1 binding, short RNAs were assayed and yielded a K_d of 90 nM.[5] The region between residues 40 and 80 in the M2-1 protein has been shown to be in direct contact with the RNA when bound.[60] Further studies have identified short A-rich RNA as being a preferable binding partner.[63]

As previously mentioned M2-1 has phosphorylated andun-phosphorylated forms. The phosphorylated form has been shown to interact with the nucleocapsid (N) protein. The phosphorylated residues in the protein were identified to be serines 58 and 61[64] and are

required for the efficient transcriptional function of M2-1, as their mutation, causes a severe decrease of antitermination activity.

M2-1 also binds to the phosphoprotein (P) in the hRSV genome in a competitive manner in the core region of amino acids 59 to 178.[65] hRSV P protein is known to bind with the nucleocapsid forming a large protein-protein complex, which has been found in inclusion bodies when all three proteins (N, M2-1 and P) are present in an unbound state.[63] The co-localisation of M2-1 with N and P was found to be indicative of its function as a transcription factor in the ribonucleoprotein (RNP) complex.[65]

A solution NMR structure of M2-1 core region (residues 58-177) was solved in 2012.[63] The core structure has been shown to be a monomer due to it lacking the oligomerisation domain as shown in Figure 1.6.1.1.[62] The core domain is structurally homologous to the C-terminal domain of Ebola virus VP30 and contains a singular globular domain from residues G75 and I171.[63] The NMR result was superseded by the determination of the full length M2-1 structure[5] by X-ray crystallography in 2014.

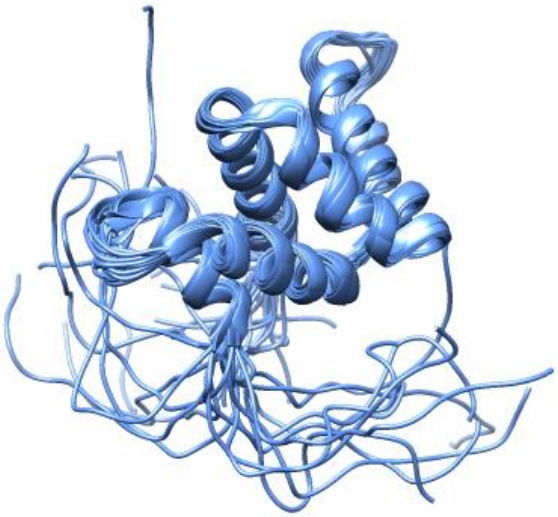


Figure 1.6.1.1. The solution NMR structure of M2-$1_{58\text{-}177}$ shows a monomer with large regions of unstructured loop (PDB Code: 2L9J)

M2-1 was crystallised in its native tetrameric form. The structure shows a disk-like assembly with a four-helix bundle in the centre. The helix responsible for tetramer formation is linked to the core domain that has been shown to be responsible for RNA binding.[63] The RNA binding region includes two critical serine residues, S58 and S61[64] that are phosphorylated during infection which are shown on a flexible linker region.

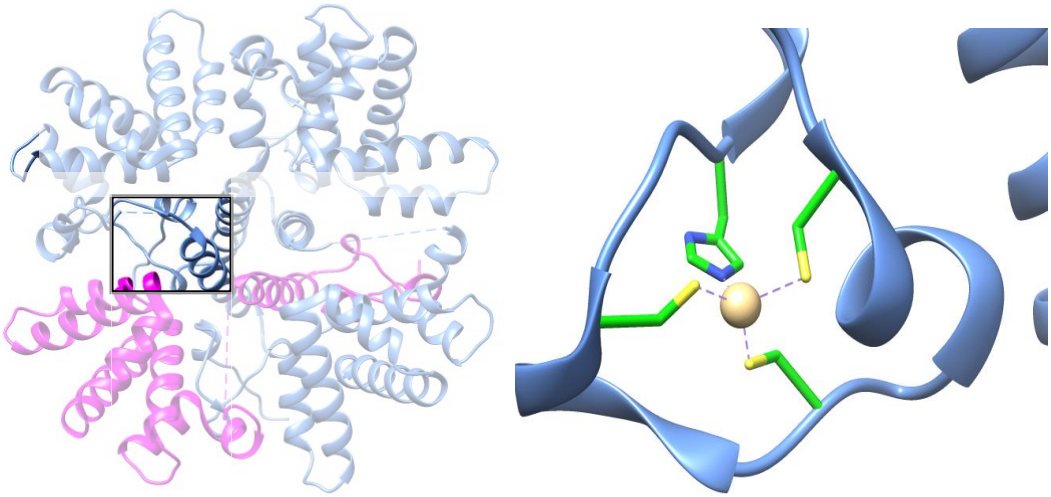


Figure 1.6.1.2. Left, Tetrameric RSV M2-1, Monomer unit shown in magenta. Right, close up of zinc finger domain with the zinc ion shown as a sphere bound to 3 Cysteine and 1 Histidine residues.[5] (PDB code: 4C3B.)

1.6.2. COMPARISONS TO OTHER TRANSCRIPTION FACTORS

1.6.2.1 M2-1 METAPNEUMOVIRUS. (HMPV)

hMPV M2-1 crystallised in the same space group and oligomerisation state as hRSV M2-1.[50] The structure shows an asymmetric tetramer where three monomers are in the globular formation, which mimics hRSV M2-1. The fourth monomer shows an open confirmation where the core region (residues 58-177) is twisted away by approximately 60 Å. This open formation is in contrast to the hRSV M2-1 tetramer which has all monomers in a closed formation forming a disc shape as shown in Figure 1.6.1.2 above.[5]

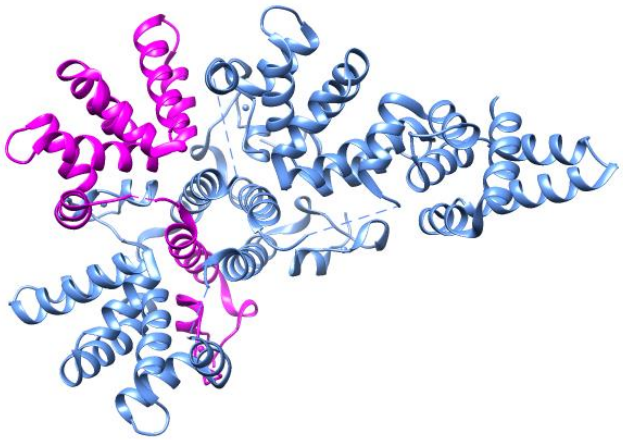


Figure 1.6.2.1.1. The structure of the Human Metapneumonavirus M2-1 tetramer.[50] (PDB code: 4CS7)

1.6.2.2 EBOLA VIRUS VP30

VP30 protein in the Ebola virus is essential for transcription as M2-1 is in hRSV. Currently a full-length structure of the VP30 protein has not been elucidated but a structure of the C-terminal domain has been resolved ($VP30_{CTD}$).[66] The C-terminal domain folds into a dimeric helical shape independent of the N-terminal domain.[66] The full length protein has been found in virions to form hexamers, due to the presence of the oligomerisation domain in the N terminal region.[66]

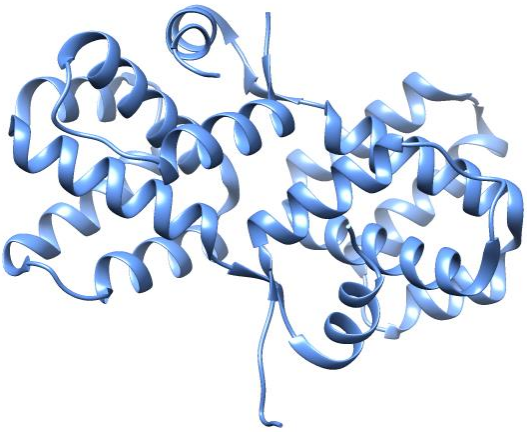


Figure 1.6.2.2.1 Crystal structure of dimeric $VP30_{CTD}$.[66](PDB Code:2I8B)

1,7 FUSION PROTEIN (F)

1.7.1 STRUCTURE AND FUNCTION

The fusion protein is a glycoprotein essential for viral entry and has been identified as a key drug target for vaccine development.[43] hRSV contains three surface glycoproteins in total but the fusion protein is the most highly conserved among the different strains in contrast to the attachment protein (G), which is highly variable.[35] The fusion protein alone has been found to be essential for the attachment of the virus to the host cell, if disrupted or an inhibitor is used, fusion and entry of the virus are limited.[6,37]

The hRSV fusion protein is a class I viral fusion protein, this means the protein conforms to a structure and interaction pathway indicative of this type.[14] The final form of hRSV F is a trimer, where each monomer of F protein is synthesised as a 67 kDa precursor, known as F0, which is modified by N-linked glycans.[43,67] The resulting F protein has a N-terminal F2 protein linked by disulphide bonds to the F1 trans-membrane domain protein.[68]

F protein is anchored into the membrane of the virion in a metastable, prefusion form. The protein is in two key parts, the fusion head and the cytoplasmic tail.[67] The fusion head is triggered into undergoing a conformational change, which inserts its hydrophobic fusion peptide into the target cell membrane.[68] The three F1 monomers then fold back on themselves to bring the viral and cell membranes together and initiate fusion.[68] A cytoplasmic domain is attached by a linker to the fusion head at one end and at the other the peptide interacts with the cytoplasm.[7,43,67]

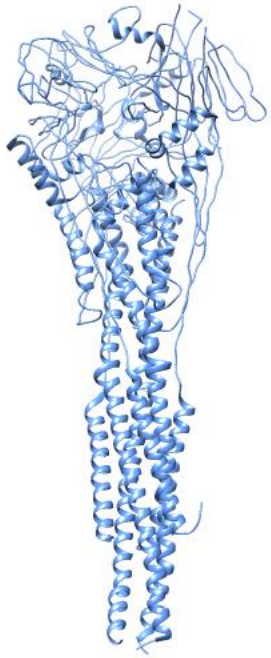


Figure 1.7.1.1 Postfusion hRSV F protein.[43](PDB Code: 3RRT)

1.7.1. COMPARISONS TO OTHER FUSION PROTEINS.

1.7.1.1 EBOLA FUSION PROTEIN. (GP1,2)

The Ebola fusion protein ($GP_{1,2}$) is also a trimeric fusion protein which is anchored to the surface of the virion and is solely responsible for membrane attachment and fusion.[69] The protein is expressed by the virus as a single peptide, which, in this case, is cleaved by furin instead of the protolytic cleavage seen in hRSV.[69] This cleavage yields the subunits GP1 and GP2 that associate to form a homotrimer structure, which is structurally similar to that seen in hRSV.[44,69]

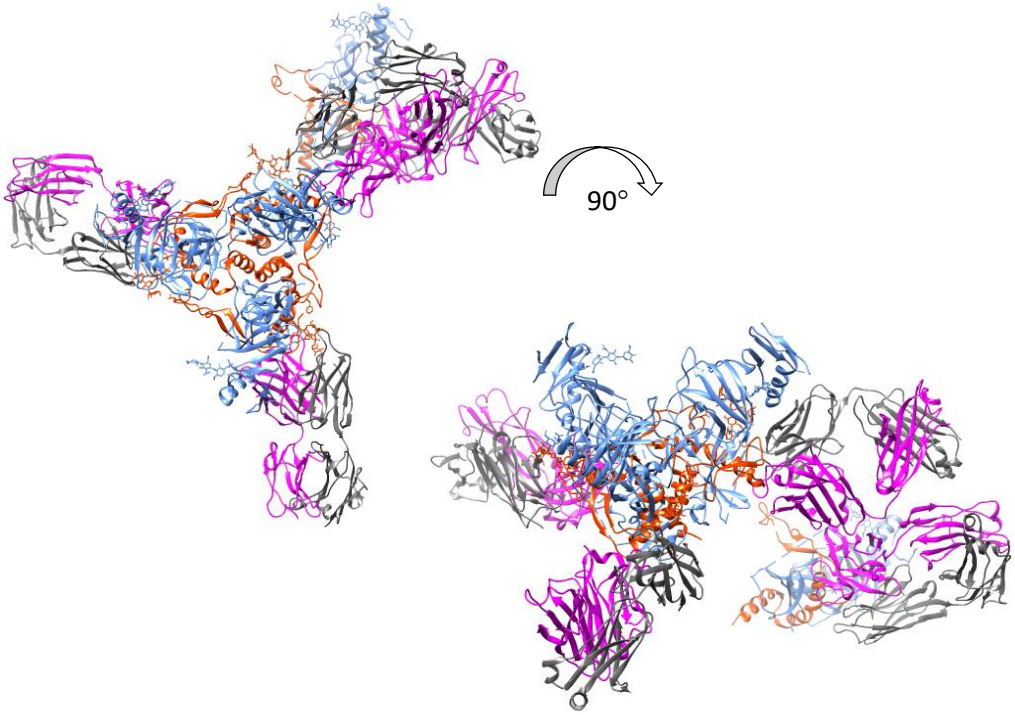


Figure 1.7.1.1.1 The crystal structure of Ebola fusion protein (GP). The homotrimer (GP protein 1 in Blue and GP 2 in orange) has been crystallised bound to a human antibody Fab KZS2. (magenta)[69] (PDB Code: 3CSY)

1.7.2.2 METAPNEUMOVIRUS FUSION PROTEIN (HMPV-F)

Metapneumovirus as detailed above in Section 1.3.1 is in the same subfamily as hRSV and so it follows that the two fusion proteins would be very similar as they carry out the same functions.[43,70] The fusion protein on hMPV is the same class I viral fusion protein and as in hRSV F the protein folds into a metastable, pre-fusion conformation. On activation the protein undergoes refolding concurrent with membrane fusion.[70] The prefusion state (F0) is shown below in Figure 1.7.2.2.1. The two proteins cannot be structurally compared as hRSV F is in the post fusion state and hMPV-F is the pre fusion state. However this does help us understand how each protein might look in the pre or post fusion form in a related virus.

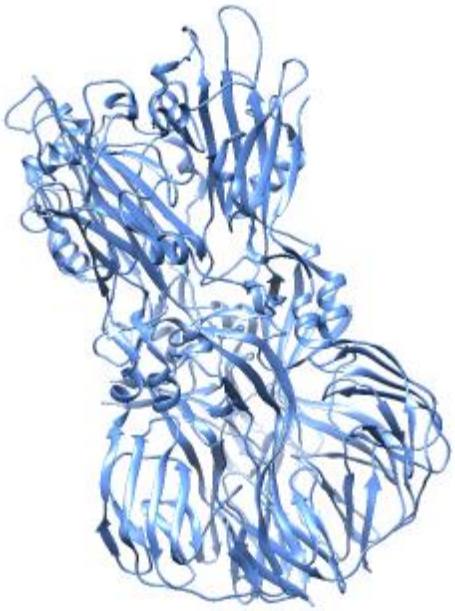


Figure 1.7.2.2.1 hMPV-F protein, F0 domain.[70] (PDB Code: 4DAG)

1.8 NUCLEOCAPSID PROTEIN.

1.8.1 HRSV NUCLEOCAPSID PROTEIN.

In 2009 the hRSV nucleocapsid (N) protein was crystallised and the structure determined to 3.3 Å resolution.[8]

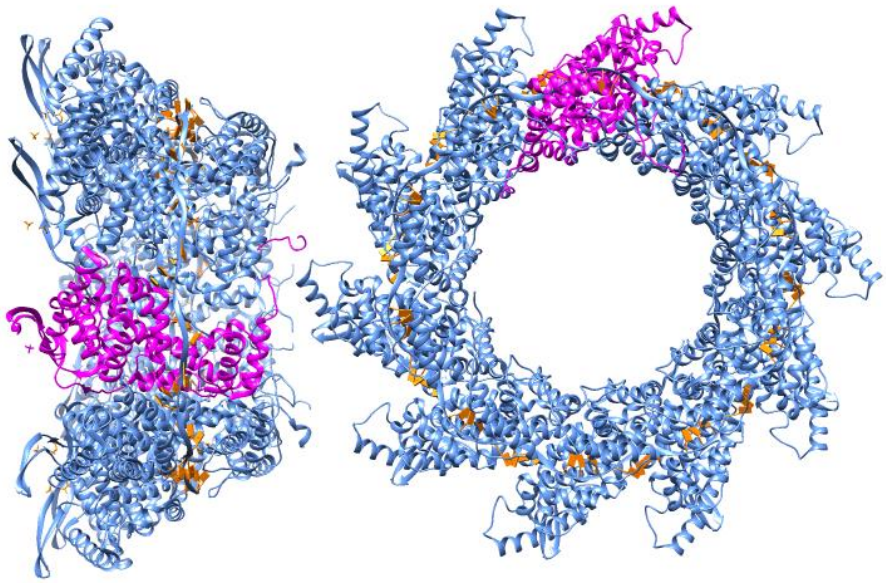


Figure 1.8.1.1 hRSV nucleocapsid protein shown in a decameric form. The monomer is highlighted in magenta. RNA is intergral to the structure.[8] (orange) (PDB:2WJ8)

The protein has been shown to form decameric and undecameric rings in solution, as illustrated above.[8] The monomers in the decamer were found to form a ring system. Inter-ring contacts between the residues suggest a basis for helical packing.[9] This packing helps stabilise the protein by mainly ionic interactions, thus the protein is stable as a ring in solution, which is key for the formation of a large nucleocapsid.[9] In this 2009 structure there was some ambiguity as to the location of the RNA in the ring as the electron density in the crystal structure was not well defined.[8] Due to the essential role that hRSV N has in genome protection and as a replication template, the N protein is a key drug target and various compounds such as RSV604 have shown this to be a viable approach.[71] Lead clinical candidate RSV-604 is discussed in depth in Chapter 5.

In addition the protein structure was solved without hRSV phosphoprotein to 3.6 Å resolution. In this solution clear electron density was observed corresponding to RNA in

Figure 1.8.1.2 unlike the earlier structure by Tawar *et al.* 2009.[8] There are seven ribonucleotides in contact with each protein monomer, this is in contrast to the VSV nucleocapsid[72] and Rabies nucleocapsid[73] proteins where the RNA has been located inside the ring and nine ribonucleotides are bound per monomer. The RNA found in the proteins is arranged in rows of interchanging four and three stacked bases which alternate between exposed and buried within a protein groove as shown below in Figure 1.8.1.2.[9]

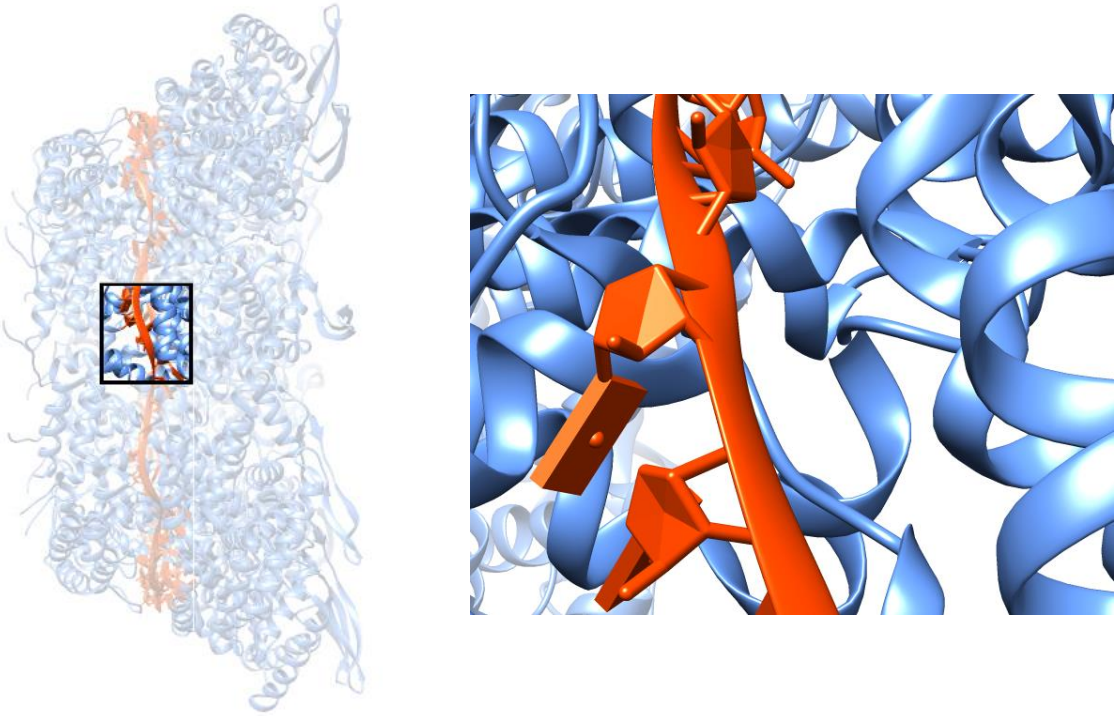


Figure 1.8.1.2 RNA shown in the cleft of hRSV N (left) and a close up view[9] (right) (PDB Code: 2YHM)

The monomer units of hRSV N are composed of the usual N-terminal domain (NTD) and C-terminal domains (CTD) which have extensions called the 'N-arm' and 'C-arm', these arms were found to make contact with the adjacent monomer and seem to give stability to the ring structure.[9] The buried surface area between each of the monomers in the ring can vary from 2500 to 2600 Å^2 and the interactions between them consist of salt bridges, hydrogen bonds and hydrophobic interactions.[9]

When overlaid, as shown in Figure 1.8.1.3 all three structures overlap as expected with a RMSD (Root Mean Squared Deviation) of 0.448 Å. These differences show that there is

flexibility in the decameric ring and that the individual protein monomers can undergo small rigid-body movements but still keep the ring's higher order structure.

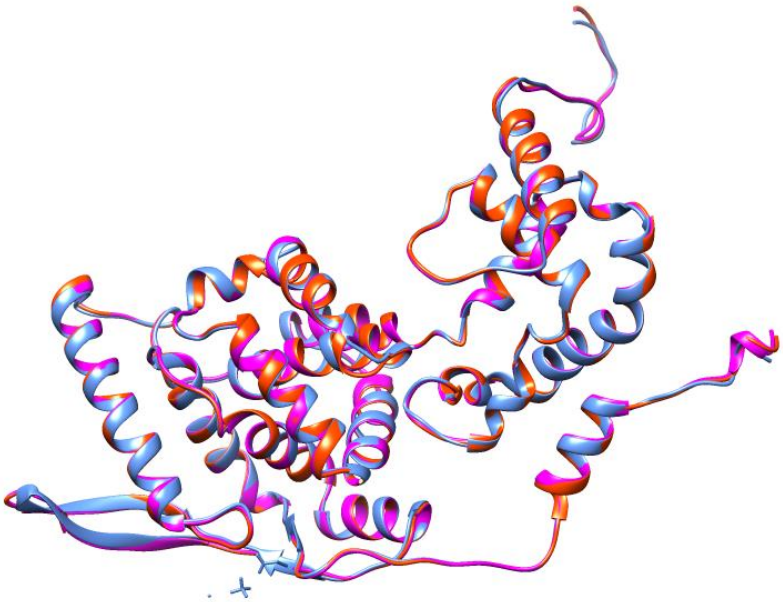


Figure 1.8.1.3 Overlay of each structural monomer of hRSV Nucleocapsid, Tawer *et al.* [8](orange, PDB:2WJ8), El Omari *et al* Orthorombic structure [9](blue, PDB Code:2YHM) and Triclinic form[8] (Magenta, PDB Code: 2V5V)

1.8.2 COMPARISONS TO OTHER MONONEGAVIRALES NUCLEOCAPSID PROTEINS

1.8.2.1 VESICULAR STOMATA VIRUS (VSV) NUCLEOCAPSID

Vesicular Stomata Virus (VSV), as described previously, belongs to the family *Rhabdoviridae*, its genome is only 11 kb in length and codes proteins that are similar to those of hRSV.[72] The VSV nucleocapsid also forms decameric rings and its structure has been determined a number of times with different co-crystal substituents. In 2006 Green *et al.*[72] crystallised the nucleocapsid protein to a resolution of 2.9 Å and found it to have near 10-fold rotational symmetry when crystallised with RNA in complex to the protein, in a similar way to that of hRSV N. It was found that the N-RNA complex consisted of 10 molecules of recombinant VSV N protein and 9 nucleotides of bacterial RNA,[72] this is in contrast to hRSV N which only contains 7 nucleotides in contact. Each N-RNA oligomer of VSV N was found to have a bi-lobed structure with RNA bound tightly in the cavity, which was found between the interface of the N and C- terminal helical lobes[74] as shown in Figure 1.8.2.1.1. It was also found that the VSV N protein could shield bacterial RNA from ribonuclease action by close contacts with key amino acid residues of the N protein in the middle of the protein.[75]

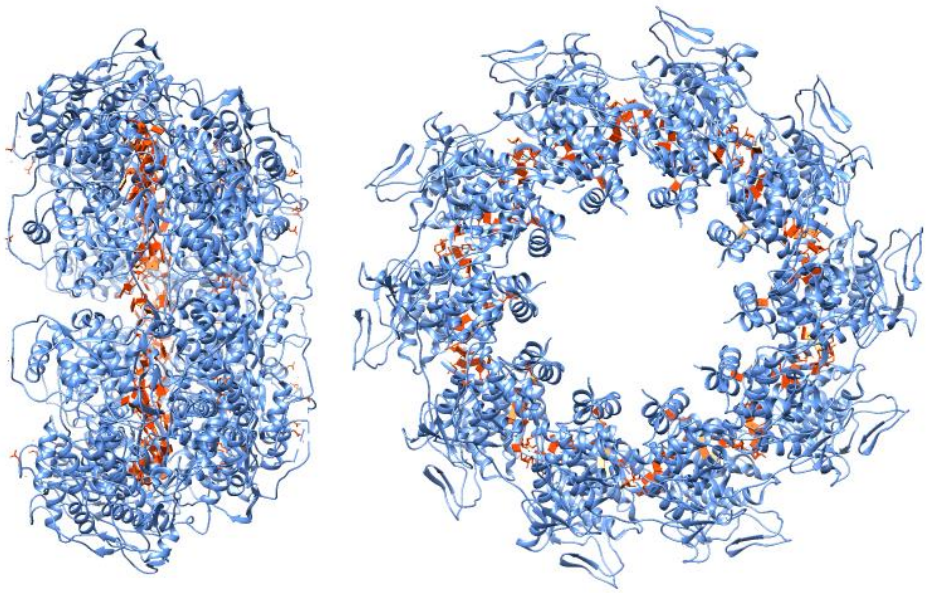


Figure 1.8.2.1.1 VSV nucleocapsid protein.[72] (PDB Code: 2GIC)

1.8.2.2 RABIES NUCLEOCAPSID PROTEIN.

Rabies virus nucleocapsid protein acts in the same way as the N proteins of hRSV and VSV by protecting the genomic RNA and providing a template for replication.[73] The structure of rabies N protein was solved in complex with viral RNA in 2006 by Albertini *et al.*[73] to 3.5 Å resolution in an undecameric ring as shown below in Figure 1.8.2.2.1. This oligomerisation has been seen in hRSV N protein where one of the crystal forms was an undecameric ring, but VSV only crystallised as a decamer.

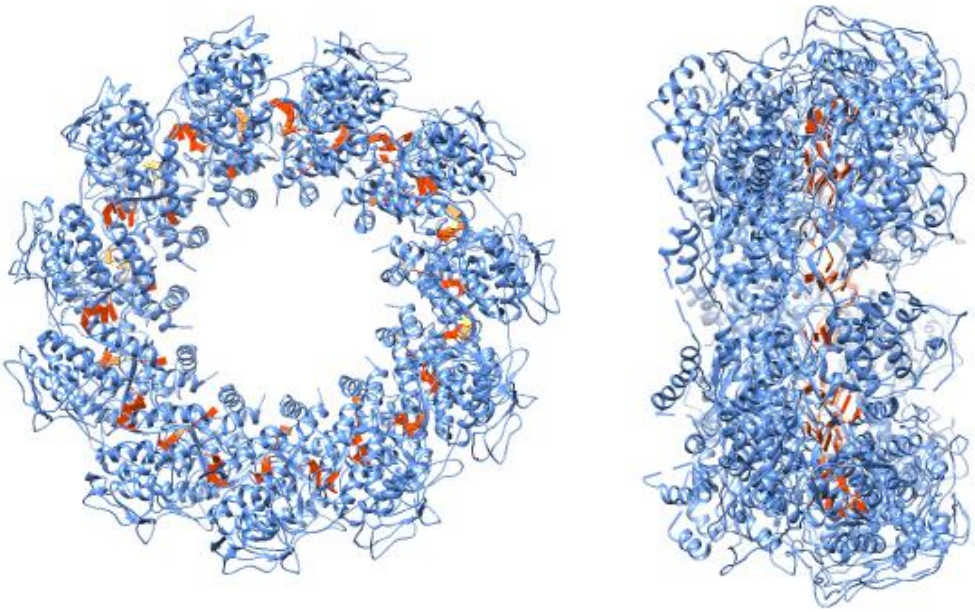


Figure 1.8.2.2.1 Rabies nucleocapsid protein[73] (PDB Code:2GTT)

The small changes in the monomer and the composition of the ring packing, allow an extra subunit to fit and form an undecameric ring.[73] The groove around the central belt where the RNA strands lie, which is pronounced in the hRSV N and still visible in the VSV protein, is almost lost in the Rabies protein. The RNA in the structure seems to intertwine with the ribbons of protein structure rather than sit inside the groove in the structure. Along with the VSV N protein, the rabies N has an important role in the switching between RNA transcription to replication by encapsulating the newly synthesised genome RNA.[76]

1.9 SCOPE AND AIMS OF THIS RESEARCH.

Many of the individual proteins encoded by the hRSV genome have been crystallised and characterised, but this only gives a small understanding of how these proteins interact together as an infectious agent. hRSV M2-1 is known to interact with both M and P[63,77] but until 2013[78] it was thought that M2-1 was only found in the host cell cytoplasm for the synthesis of mRNA. It is now known that it is involved in the formation of the ribonucleocapsid, where it is sandwiched between the nucleocapsid and the matrix protein.[36] This discovery is key to the research in this thesis as the location of the M2-1 in the virus directly affects the interactions that the protein can make. The interaction of M2-1 with the P protein has been characterised and the scale of the interaction has been reported[60,62] unlike its interaction with hRSV M. Chapter 3 in this thesis strives to gain insight into the interaction between these two proteins M and M2-1, through determination of the thermodynamic pathway.

hRSV M protein was first crystallised in 2009[48] and since then other matrix protein structures from the order *Mononegavirales* have become available.[50–52,57] Many have been crystallised in the dimeric form, including hRSV M which has recently been crystallised and solved.[79] The matrix proteins' interactions with the fusion protein are key for viral budding and assembly.[34] The cytoplasmic tail of the fusion protein is known to interact with the matrix protein, and the key residues have been identified in this interaction.[7] This research will begin to probe the interaction using biophysical techniques i.e. thermal shift assays, Förster resonance energy transfer and surface plasmon resonance and begin to build an understanding of the binding.

The nucleocapsid protein has long been considered a target for antiviral drugs as its functions are key to the encapsidation of RNA and as a replication template for the emergence of a new virus. In Chapter 5 high-throughput screening ligands by Arrow Therapeutics Ltd are introduced[10] and a number of novel scaffolds are described and validated by *in vivo* assays. In this chapter the aim is to find the mode of action of these ligands against the hRSV N protein. If this is known, it would provide an excellent platform to further characterise the interaction and direct structure-activity relationship studies.

The investigation of the interaction of matrix and fusion proteins is continued in Chapter 6, where the key interaction with phenylalanine 572[7] is the basis for a series of potential

ligand compounds. The aim of the chapter is to identify a pocket in the protein so docking of a ligand can take place.

GOLD[80,81] is a program developed by the Cambridge Crystallographic Data Centre (CCDC) and enables the user to identify favourable interactions in the protein and dock large subsets of ligands into the identified pockets. The aim is to dock ligand compounds to the protein *in silico* and validate these hits by biophysical techniques.

Overall this research aims to investigate key interactions in the virus' replication cycles that include those between matrix and M2-1 and matrix and fusion peptide. These interactions will be studied to understand basic stoichiometry and thermodynamic pathways. Further work to identify small molecule ligands that can bring a destabilising effect to the protein-protein interaction between N and M-F, conclude the thesis. The small molecules ligands are new targets and could lay the foundations for new anti-viral drugs in the future.

CHAPTER 2. CHARACTERISING M2-1 PROTEIN.

The work described in this chapter used biophysical techniques, including dynamic light scattering and circular dichroism, to confirm the known structural data on M2-1. Further characterisation using ICP-MS and MALDI-TOF also confirmed the amino acid sequence and zinc content of the protein which has been previously characterised by Tran *et al.*[62] For full structural characterisation, an X-ray crystal structure is required and the use of thermal shift assays to determine optimum buffer conditions for crystallisation is outlined. M2-1 is a core transcription protein from hRSV which is found in the ribonucleocapsid complex.[59] The M2 gene codes for mRNA containing two open reading frames where ORF1 codes for a 194 amino acid structural protein, M2-1. As discussed in Chapter 1.6, M2-1 enhances the synthesis of polycistronic mRNAs which are essential cofactors of viral RNA polymerase.[4] In 2009, M2-1 was relatively structurally unknown. It was known that M2-1 formed a tight tetramer of 89 kDa and the structure was found to be primarily alpha helical by circular dichroism.[62] In early 2012 a solution NMR structure was published[63] as shown in Figure 1.6.1.1 of the core region between residues 58-177 but no X-ray crystal structure was published until 2014.[5]

The work described in the following chapter aimed to successfully over-express and purify M2-1 in order to facilitate future crystallisation. Expression trials were used to obtain the largest yield of protein from an *E. coli* system building on previous work by Tran *et al* [62] which favoured a Glutathione-S-Transferase (GST) tagged protein to improve solubility.

2.1. PROTEIN EXPRESSION AND PURIFICATION.

M2-1 *pGEX-6p-1*[82] was successfully cloned by Dr. R.P. Yeo. Expression, purification and further analysis were carried out by the author. A pGEX vector was chosen as it provided a glutathione-S-transferase tag (GST tag), a high affinity to glutathione enabled purification from other cell debris and *E. coli* cytoplasm.[83] The protein can be isolated by the addition of glutathione beads, where the GST tag on the protein binds to the free glutathione on the surface of the beads. The native untagged protein can be harvested after purification; by cleavage of the linker between the tag and the protein by Precission™ Protease enzyme. This cuts between a glutamine and a glycine residue in the recognition site (Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro).[83–85] Protein molecular weight and purity were confirmed and monitored by SDS-PAGE as shown in Figure 2.1.1.[86]

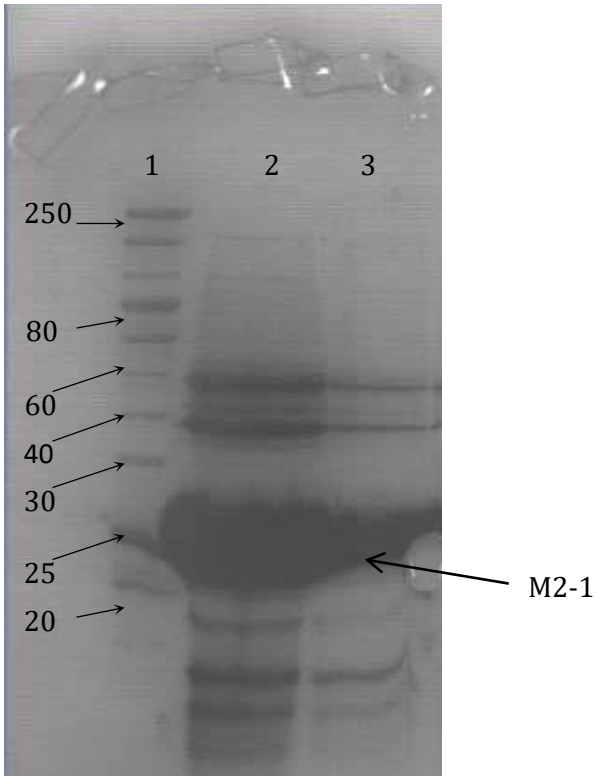


Figure 2.1.1. 15% SDS-PAGE showing cleaved hRSV M2-1 protein at ~22kDa. Lane 1. Ladder of standard protein masses (10-250 kDa). Lane 2 and 3 untagged crude protein after cleavage from glutathione beads.

The protein was further purified and characterised through size exclusion chromatography (SEC). Purification was required as extra bands, which were not the protein of interest, were present on the gel as shown above in Figure 2.1.1. The protein oligomerisation state could be determined by running known proteins, called size markers, through the system and monitoring their retention times. This is a standard calibration and quantifies the size of proteins in solution. A variety of proteins with known retention time were used from thyroglobulin, 97 kDa to vitamin B12, 1.3 kDa. The protein, M2-1 was eluted after injection of 30 mL of buffer, this gave an indication that the protein was a tetramer (88 000 Da) in solution as previously reported.[62]

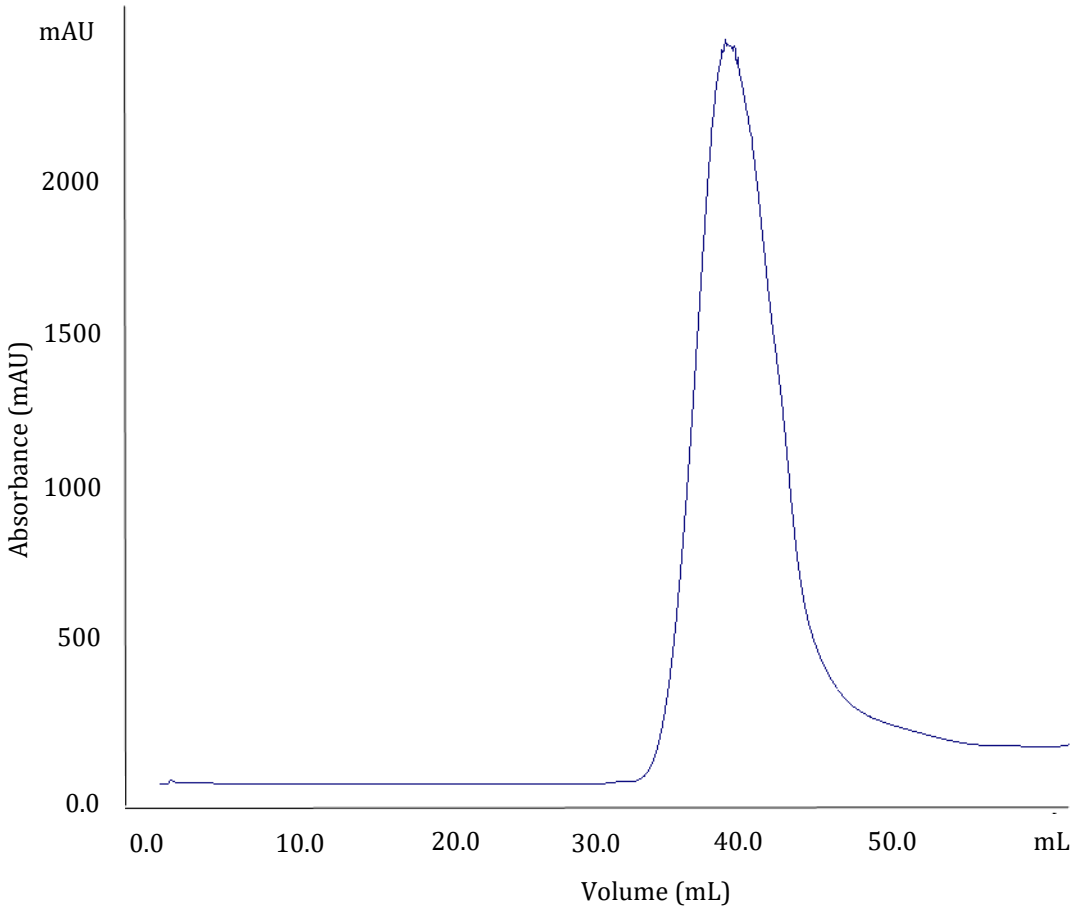


Figure 2.1.2. Size exclusion Chromatography raw trace of M2-1 tetramer elution as shown from the Unicorn Software.[87]

2.2. ANALYSIS BY MASS SPECTROMETRY.

2.2.1. MALDI-TOF MS

Confirmation of the identity of the protein shown on SDS PAGE was carried out by tryptic digest and MALDI Mass spectrometry. Dr A.P. Brown, (School of Biological and Biomedical Sciences, Durham University) conducted the experiment. The tryptic digest of the protein was carried out by extraction from a SDS PAGE gel band. Firstly the correct band was cut from the larger gel and the coomassie stained piece was destained using ammonium bicarbonate/acetonitrile (1:1 v/v). The protein gel piece was purified by the addition of formic acid/water/2-propanol (1:3:2 v/v/v) which removed the SDS. The gel was then suspended in a buffer containing trypsin and incubated at 37 °C overnight, where the protein was digested and then extracted from the gel using (1:2 v/v) 5% formic acid/acetonitrile. The extracted protein fragments were mixed with matrix solution, 0.02-0.03 M alpha-cyano-4-hydroxy-cinnamic acid in 0.1% TFA/acetonitrile (1:2 v/v). The mixture was highly acidic and deactivates the trypsin and a small sample was transferred onto a MALDI sample plate and allowed to dry. The sample was analysed in the spectrometer and the output was run though a BLAST (Basic Local Alignment Search Tool)[88] search and a positive identification of the M2-1 protein was obtained. Full fragment data are found in Appendix A.

2.2.2. ICP-MS

ICP-MS (Inductively coupled plasma mass spectrometry)[89] was used to confirm the binding capacity of the zinc finger in the purified M2-1 protein. The presence of a zinc finger in the N-terminus of the protein has been extensively characterised[61] and a zinc ion was bound to the crystallised protein.[5] The technique uses known concentrations of metal ions to produce calibration curves shown in Appendix A. Each curve plots the metal signals against their concentrations. This allows the user to test the unknown sample and receive the corresponding metal signal. This can be plotted onto the calibration curves to reveal the concentration of the metal ion in question.

Purification of M2-1 traditionally includes the addition of zinc ion when the cells are induced and also in the lysis buffer when the purification process begins.[5,64] To test the dependence of M2-1 on zinc ion, the protein was purified with no extra zinc in the lysis buffer nor were zinc ions added to the culture when the cells were induced. This sample

was found to contain 2.5% zinc by mass and the expression levels were found to be lower than when the zinc ions were added to the growth media on induction. The metalation of M2-1 was tested following the addition of $ZnSO_4$ at induction, the protein was found on average to be 67% metalated. After size exclusion chromatography the protein was found to be on average 40% metalated. Removing the zinc entirely from M2-1 had a particularly harsh effect on the protein and precipitation was seen after the addition of EDTA (10 mM) and incubation at 4 °C.

2.3. CIRCULAR DICHORISM. (CD)

Circular dichorism (CD) spectroscopy is a form of UV spectroscopy, which uses circularly polarised light to determine the secondary structure content of a protein. The polypeptide backbone of a protein is optically active and will absorb circularly polarised light to give a characteristic signal.[90] The signal is the difference in extinction coefficients between the right and left handed polarised light in units of cm M^{-1}

$$\Delta\varepsilon=(\varepsilon_L - \varepsilon_R) = (A_L - A_R)/cl$$

Where $\Delta\varepsilon$ is the overall change in in the extinction coefficient, ε_L is the signal from the left handed polarised light. ε_R is the signal from the right handed polarised light.

This is often shown in terms of ellipticity, θ, where

$$\theta = 33(A_L - A_R)$$

or molar ellipticity:[35]

$$[\theta] = 100\theta/cl$$

where c –concentration (mol^{-1}), l – pathlength (cm)

The ellipticity at different wavelengths of UV light depends on the secondary structure of the protein; α-helices, β-sheets and random coil each give distinctive CD spectra as shown in figure 2.3.1.

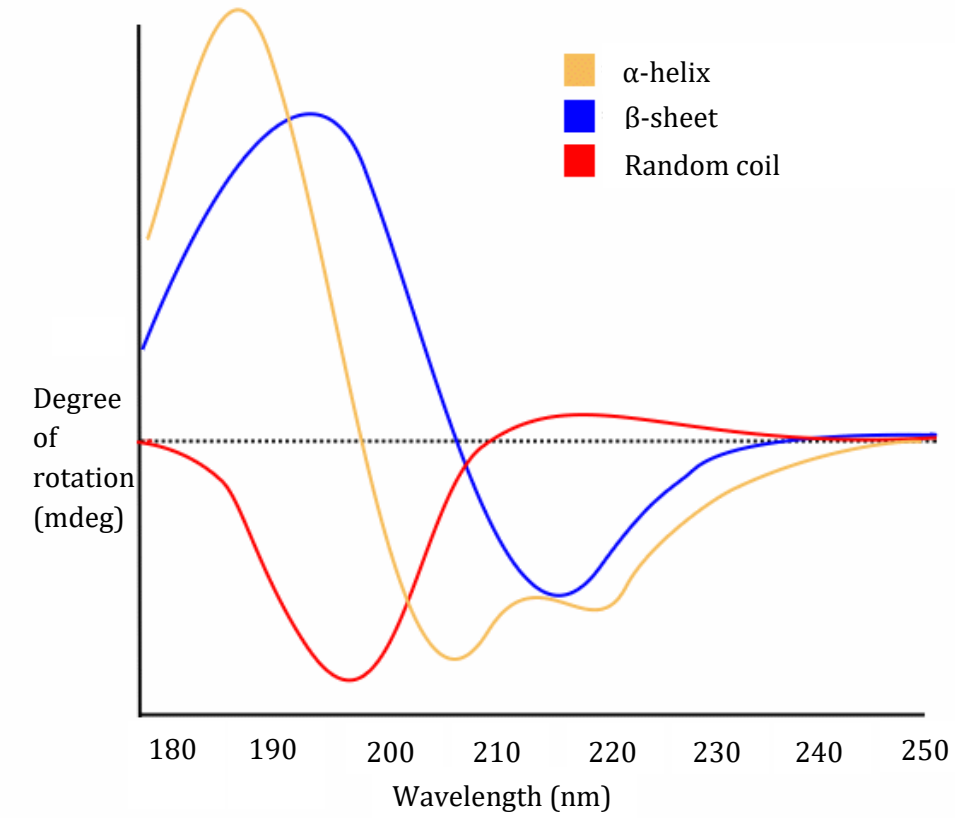


Figure 2.3.1. A typical CD spectra showing the different types of secondary structure.[87]

Several empirical algorithms have been developed to estimate the secondary structure of a protein from the corresponding CD spectrum. They work by analysing the experimental data in relation to the spectra of proteins with known structure. Largely these can be accessed online *via* the CDPro software package.[91] All the algorithms rely on the assumption that the contribution to the signal from different components i.e. alpha helices, beta sheets within the protein are independent and can be added linearly to produce the net spectrum C_{λ}.

$$C_{\lambda} = \Sigma f_k B_{k\lambda}$$

Where f_k is the fraction of secondary structure *k*, and $B_{k\lambda}$ is the value of the secondary structure component.

The CDpro package[91] was used to determine an approximation of the secondary structure of M2-1 using known protein structures. The known structures have been tested by the method and their spectra are contained in data sets which are used as a standard reference. The software contains 10 data sets, 6 for soluble proteins, 2 of which represent membrane proteins and the final 2 contain a mixture of the two different types.[91] M2-1 was suspended in 50 mM phosphate buffer for protein stability and tested

as detailed in Methods 2.8 and the raw data curves are shown in Figure 2.3.2. The signal is calculated as the different in the absorption of the left handed and right handed circularly polarised light. The first graph which plots circular dichroism (CD) against wavelength, shows the overall signal. The second HT is the amount of voltage used across the wavelength range. At lower wavelengths the protein, air and buffer solution will absorb light. Less light will then hit the detector which causes the HT voltage to increase as seen below.

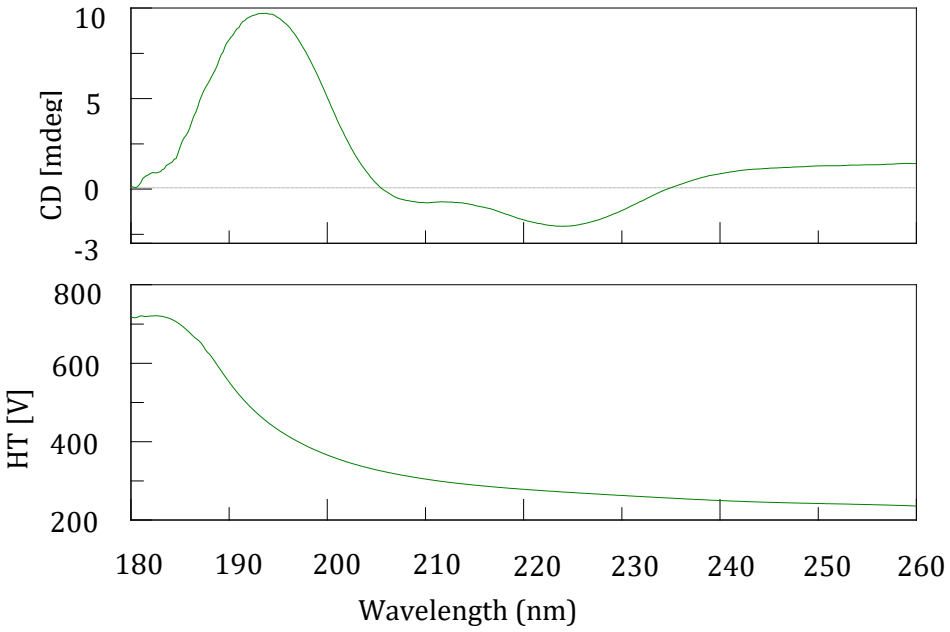


Figure 2.3.2. Raw data curve of M2-1 in 50 nM phosphate buffer solution.

Data set 9, which contains 50 proteins in the reference set, of which 13 are membrane proteins, was found to obtain the best fit to the data. The reference set and experimental data were then analysed using 3 different algorithms (CONTILL/LL-SELCON/CDSSTR)[91] supplied in the CDpro software. The algorithms analyse the data in subtly different ways, there has been no identification of which one gives the better result.[91] It has been found that the CDSSTR algorithm performs better with a smaller reference set and large wavelength range whereas CONTILL/LL prefers a larger reference set and smaller wavelength range.[90,91] The program analyses wavelengths from 180 to 240 nm where the protein has absorbed the circularly polarised light. The data set obtained from M2-1, had a shorter wavelength range but by using reference set 9, a larger protein set was obtained. CCDSTR could not be used as it gave an inconclusive result for the data.

CONTILL/LL and SELCON analyses overlaid on the experimental data are shown in Figure 2.3.3.

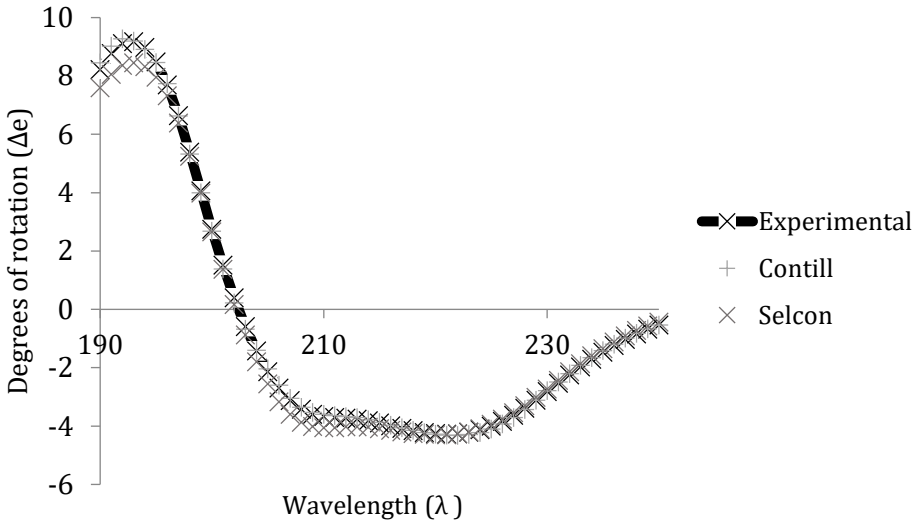


Figure 2.3.3. CD spectrum of M2-1 in 50mM Phosphate buffer (190-240 nm) fitted using CONTIN/LL and Selcon 3 algorithms in data set 9 in the CDPro package.[91]

From the graph it can be seen that the two data sets gave very similar results, but the CONTIN/LL system fits the experimental data from 190-200nm whereas the Selcon 3 curve deviates more significantly from the line. The secondary structure fractions from the analysis are shown below in Table 2.3.3.

	α-Helix	**3_{10}-Helix**	**Strand**	**Turn**	**Irregular**
% of total	53	11	4	6	26

Table 2.3.3. Secondary structure approximation of M2-1 based on CONTILL/LL algorithm from CDpro.

2.4. DYNAMIC LIGHT SCATTERING (DLS)

This technique is known by many different names such as Photon Correlation Spectroscopy and is a popular method of defining the size of particles in solution. A laser shone onto a solution causes a Doppler shift when the light hits the moving particle, the size of the shift is dependent on the size of the particle. This method has several advantages such as a small volume is required and there is a short time frame per measurement. There are some general assumptions which underlie the experiment's theory, the first is mentioned above where the particles must be in Brownian motion. This is where the particles are suspended in random motion in a fluid. The second is that the particles are spherical in solution as shown below.[92,93]

$$D = k_B T / 6\pi\eta a$$

Where a is the radius of the particles, k_B is the Boltzmann constant, T is the temperature in Kelvin and η is the viscosity of the solution and D the diffusion constant of particles in solution.

Size determination of the protein in solution was key to understanding the oligomerisation state. It has been previously stated that the tetramer of M2-1 forms a 15 nm ± 4 nm structure.[62] The sample from the size exclusion chromatography column was tested by DLS and was found to be 16 nm ± 7 nm in diameter as shown in Figure 2.4.1.

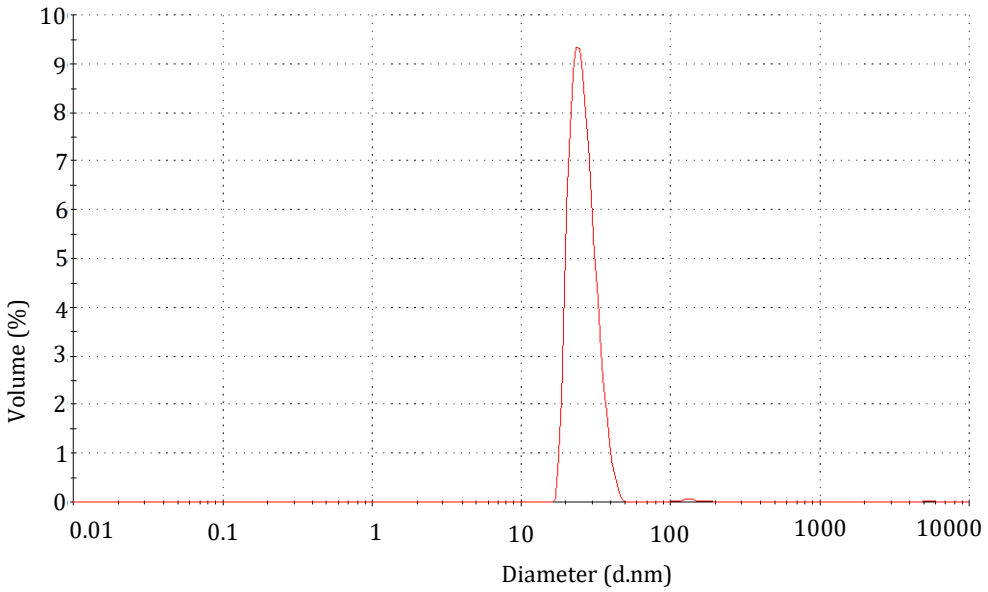


Figure 2.4.1 Dynamic Light Scattering trace showing an M2-1 size peak of approximately 16 nm in solution shown in Malvern Instruments Zetasizer Software.[94]

2.5. THERMAL SHIFT ASSAY (DIFFERENTIAL SCANNING FLUOROMETRY)

Thermal Shift assays are used to investigate the overall stability of the protein.[95] An in-house salt screen containing 96 conditions was performed in triplicate and the results averaged.

A dye, Sypro Orange, was added to the protein and salt screen mixture. The 96 well plate was heated from 24 °C to 96 °C with a reading of fluorescence taken at 1 °C intervals. As the protein unfolds, the dye binds to the hydrophobic residues in the protein backbone producing a higher fluorescence signal.[96] Every protein has some disorder in its structure where hydrophobic residues are exposed before any heat is added. This allows the dye to bind before he experiment starts and is known as a background fluorescence signal. The melting curve of the protein is seen through a large increase in fluorescence, where the 50% point of the curve is the melting temperature. The raw data melting curve of M2-1 in water is shown below in Figure 2.5.1 and the melting temperature was determined to be 47.9 °C in water through the NAMI software.[95]

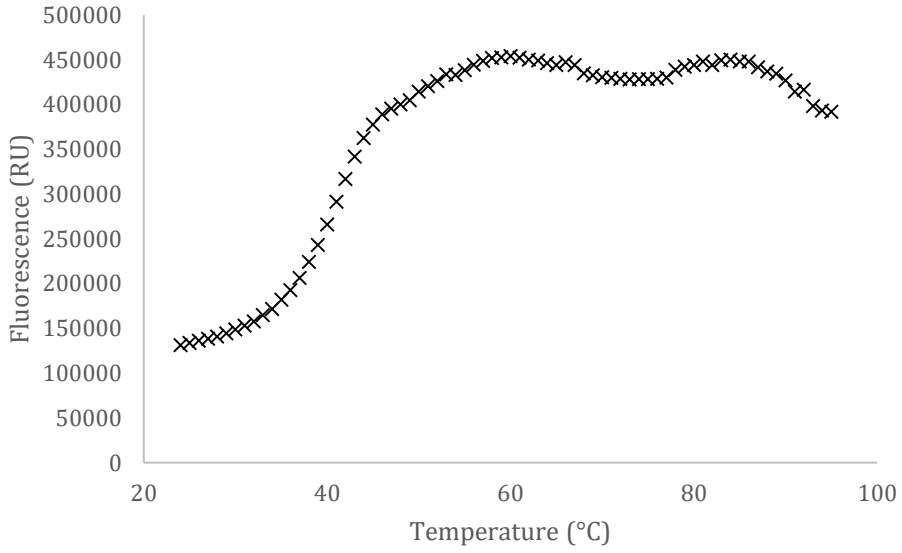


Figure 2.5.1 Raw data melting curve of M2-1 in water.

The difference in melting temperature which was taken as significant is 1°C, this is based on the sensitivity of the machine. In Figure 2.5.3 the output table of averaged results from NAMI[95] program is shown. NAMI is user interface software developed at Durham

University to calculate the 50% point of the curves and accurately determine the melting temperature. The program is driven by a numerical approach using an algorithm which identifies the melting curve to a defined window. The data in each window are then fitted by linear regression which results in a slope value (a) and a correlation coefficient (r). To calculate the precise melting temperature (T_m) the *UnivariateSpline* function from *SciPy* function[95,97] is used to interpolate the experimentally obtained data points. If no clear T_m is found the program automatically labels the data N/S for no signal. Up to two inflection points can be calculated in the program if more than one melting curve can be seen.

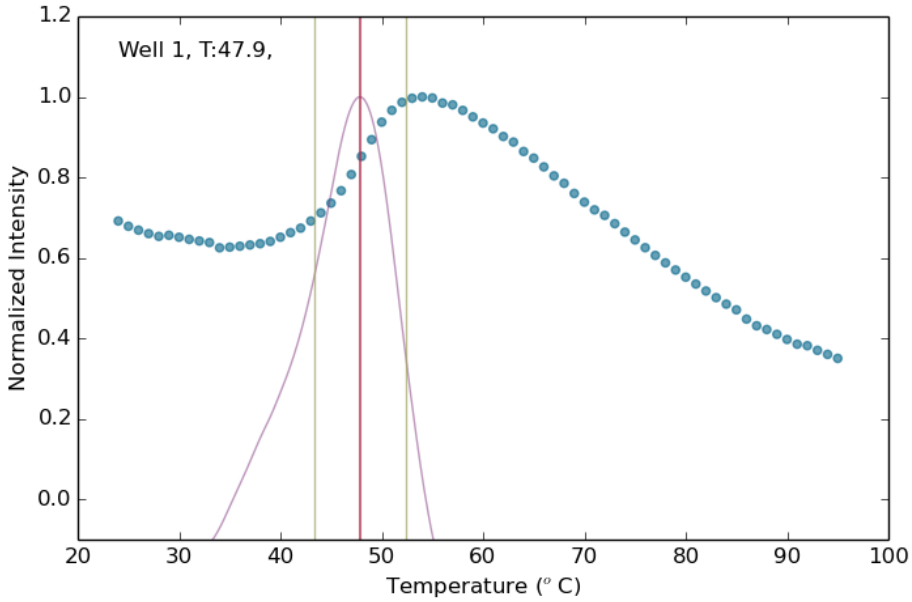
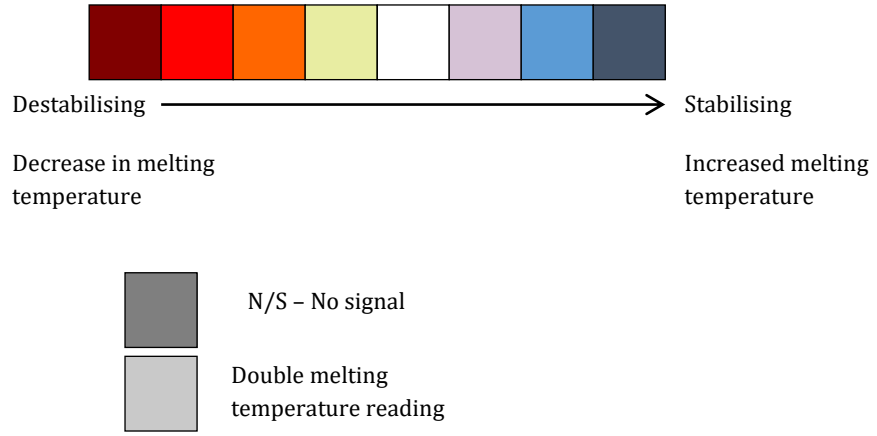


Figure 2.5.2. NAMI program representation of the melting temperature of M2-1. The melting temperature was found to be 47.9 °C and is represented by the red line.[95] The gold lines represent the melting window showing the beginning and the end of the curve. The purple curve represents the interpolated rate of change of the data.



	1	2	3	4	5	6	7	8	9	10	11	12
A	47.9	46.2	46.5	N/S	47.9	41.0	N/S	29.8 36.6	37.5	41.0	44.6	47.9
B	61.0 68.7	56.0 69.3	54.2	51.6	50.4	47.4	58.9	56.5	56.4	54.2	53.1	51.3
C	54.5	51.2	50.4	50.2	49.5	48.9	52.3	51.3	50.9	51.1	50.8	50.5
D	55.1	54.7	53.2	51.5	50.3	56.4	55.7	53.6	51.5	47.8	49.5	48.9
E	48.0	48.2	49.4	48.6	48.4	48.0	52.5	52.2	47.4	48.1	39.8	46.7
F	48.2	47.9	44.9	48.3	44.7	42.9	52.2	45.1	43.9	47.9	47.9	47.1
G	47.3	43.2	N/S	52.3	29.0 51.7	N/S	52.6	51.3	N/S	50.6	51.3	53.6
H	48.3	42.8	48.2	48.2	54.2	48.0	48.8	52.4	49.3	52.7	37.7 52.9	51.4

Figure 2.5.3. Table of salt screen melting temperature of M2-1 shown in NAMI.[95]

NAMI illustrates the change in melting temperature relative to the control sample using a colour code. This colour code coordinates the shift in temperature, small changes under a degree remain unchanged and are shown as white. Anything above a 1 °C, shows in light purple which gets darker and turns blue with each increase. Similarly a destabilising condition is shown in yellow and gradually turns red. Some conditions were shown to have two peaks but on examination of the raw data the protein was found to be denatured under these conditions and the two peaks were found to be a result of the program's, attempt to fit the data.[95] The protein was stabilised by various buffer conditions and additives, primarily these were used to change buffer conditions during purification and extraction of the protein in order to optimise yield and stability, and are shown in Table 2.5.5. The full conditions tested are detailed in Appendix A.3.

M2-1 is a structural protein, as discussed in Chapter 1, and it binds with a number of proteins within the virus structure.[5,36,61,98] The results of the thermal shift assay show that, as might be expected, M2-1 is stabilised primarily by amino acids as shown in Figure 2.5.4 below. These results were used to guide the experiments discussed in Chapter 2.6, where the aim is to crystallise M2-1. Addition of stabilising compounds in crystallisation screens can be used to enable successful crystallisation.

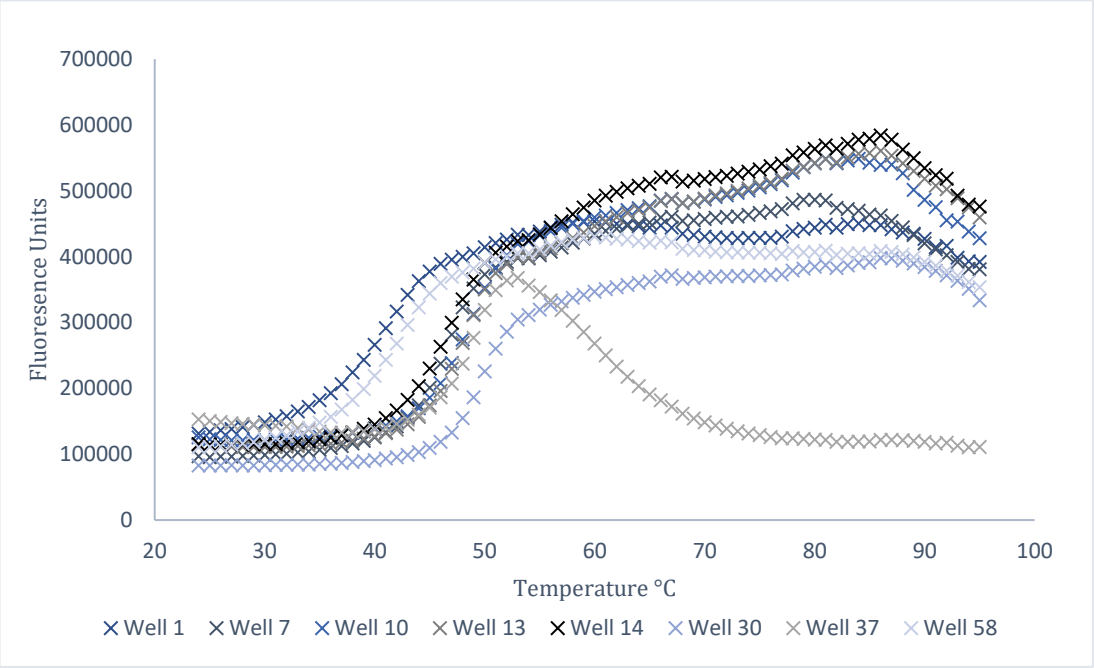


Figure 2.5.4. Raw data from hRSV M2-1 thermal shift assay showing compounds that exert a stabilising effect. Well contents are described in Appendix A.3 and Table 2.5.5.

Well	Compound	Final Concentration	Melting Temperature (T_m) (°C)
1	No compound	N/A	42
7	L-Histidine	50 mM	48
10	L-Arginine-ethyl ester	50 mM	49.3
13	Gly-Gly	50 mM	48.2
14	Gly-Gly-Gly	50 mM	47.3
30	Ethlydiamine	50 mM	50.2
37	Trichloroethylene hydrochloride	50 mM	49.2
58	Ammonium sulphate	50 mM	48.2

Table 2.5.5. Table showing the most stabilising compounds from the additive screen.

2.6 CRYSTALLISATION

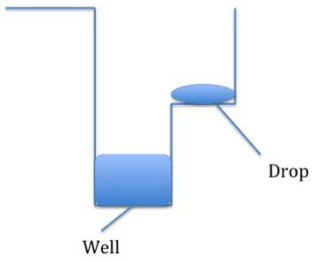


Figure 2.6.1. Sitting drop vapour diffusion crystallisation[16]

For full protein characterisation a crystal structure is required and at the beginning of the author's period of study this was not available and the main aim was to crystallise full length M2-1. High-throughput 96-well sitting drop screens were used to test a large number of conditions using the smallest volume of protein possible. The trays use the technique of sitting drop vapour diffusion as shown in Figure 2.6.1, containing a well consisting of reagent and two drops, which are a mixture of protein sample and reagent.[16] If the protein drops are at a lower concentration than the well of reagent, water vapour leaves the drop and enters the well. As water leaves the contents of the drop become more concentrated and an equilibrium is reached when the reagent concentration is approximately the same as the reservoir.[16]

If the concentration of one of the components is too low, then the drop will be clear. Once in the metastable zone, some amorphous precipitate may form and may lead to crystals. Once the protein shows thick precipitate one of the components is too high in concentration. The drops were scored using the Hampton Research Monograph Series as shown in Table 2.6.2.[99]

0	Clear Drop
1	Drop contains non-protein particles e.g Dust on drop, Fibres from clothes.
2	Drop is mostly clear but contains precipitated protein with no edges
3	Protein is fully precipitated with a dark colour
4	Gelantinous protein precipitate – usually transparent
5	Phase Separation
6	Spherulites – droplets or Small structures with defined edges.
7	Crystal growth in one dimension
8	Crystal growth in two dimensions
9	Crystal growth in three dimension

Table 2.6.2. Hampton Research scoring guidelines for crystal identification.[99]

Scores 3 to 6 describe the protein as having no defined edges, in many ways these are negative results but still provide information about which conditions the protein preferred.[99] The pH range of the drops should be monitored, this can greatly affect the results of each drop. If the pH is particularly high or low then the condition can be avoided in future. A gelatinous precipitate (Score 4) can be a good starting point for optimisation and phase separation. Where (Score 5) is visible as droplets within the drop or as an 'oil' film on the surface of a drop. It can be difficult to mount crystals from this stage as the phase is of unknown composition.[99]

From Score 6 to 9 the precipitate has definite edges. Score 6 are spherulites (transparent droplets or clusters), this stage can be a good start for crystallisation optimisation. Streak seeding can be used to optimise the drop conditions.[100] Also additives e.g. $MgCl_2$

or inhibitors can be added if optimisation conditions do not yield crystals. Score 6 can also be micro-crystals where edges are visible in high magnification. Score 7 is crystal growth in one dimension which can be seen as needles. Seeding could be used to try and grow larger crystals in optimised conditions. Score 8 is crystal growth in two dimensions, optimisation can be achieved through the use of additives and seeding. Score 9 is three dimensional, these should be checked if salt crystals are forming in the wells. Buffer conditions can induce salt crystal formation, for this reason phosphate buffers are to be avoided. To ensure the crystals are not salt, the crystals are mounted to check diffraction quality which determines if they are protein or salt crystals.[16]

There are many commercial screens available which cover a wide range of conditions. In-house at Durham University, the Molecular Dimensions 96 well plates were preferred and in stock. Five screens of 96 well plates; Pact Premier[101], Standard Screen 1+2[102], JCSG+[101,103], Standard Strategy and Standard screen 2[102], containing various high throughput conditions were used.

The first two trials at 30 μM and 60 μM M2-1 produced more clear screens (0) than precipitate. The general protocol provided through Molecular Dimensions states that approximately 50% of the drops must have a score of 2 (The drop is mostly clear but contains precipitated protein with no edges) or higher.[99] If this is not the case then the concentration of the protein should be increased. The concentration of the sample was increased from 60 μM to 120 μM of protein. This gave some light and heavy precipitate across the 96 well screens after 1 week at 4 °C, and some trays began to show small single crystals.

In Standard Screen 1+2 a number of conditions showed a range of results. E3 – 2 % v/v Dioxane, 0.1 M Bicine, 10 % w/v PEG 20,000 – Score 4 and 9. Drop C8 – No Salt, 0.1 M Tris, 8.5 pH and 2.0 M ammonium sulphate. – Score 4 and 9. Drop C8 contained ammonium sulphate, sulphates are known to form salt crystals as they interact with metal ions in the system. This was confirmed by mounting the crystal onto the goniometer and irradiating the sample with a beam of monochromatic X-rays. There was no regular repeating pattern of reflections seen. The crystals were determined to be salt. The single structures in drop E3 were not followed up as they had the same shape as the salt crystals in C8. A higher concentration of protein, 150 μM was used and the same conditions were re-tried. The standard 96 well sitting drop trays were constructed using the same commercial screen range as before. After 5 days a drop started to show needles

as shown in Figure 2.6.3. The drop conditions were from Pact Premier plate and found in D10 – 0.2 $MgCl_2$, 0.1 mM Tris-HCl, pH 8.0 with 20% [w/v] PEG 6000.[101]

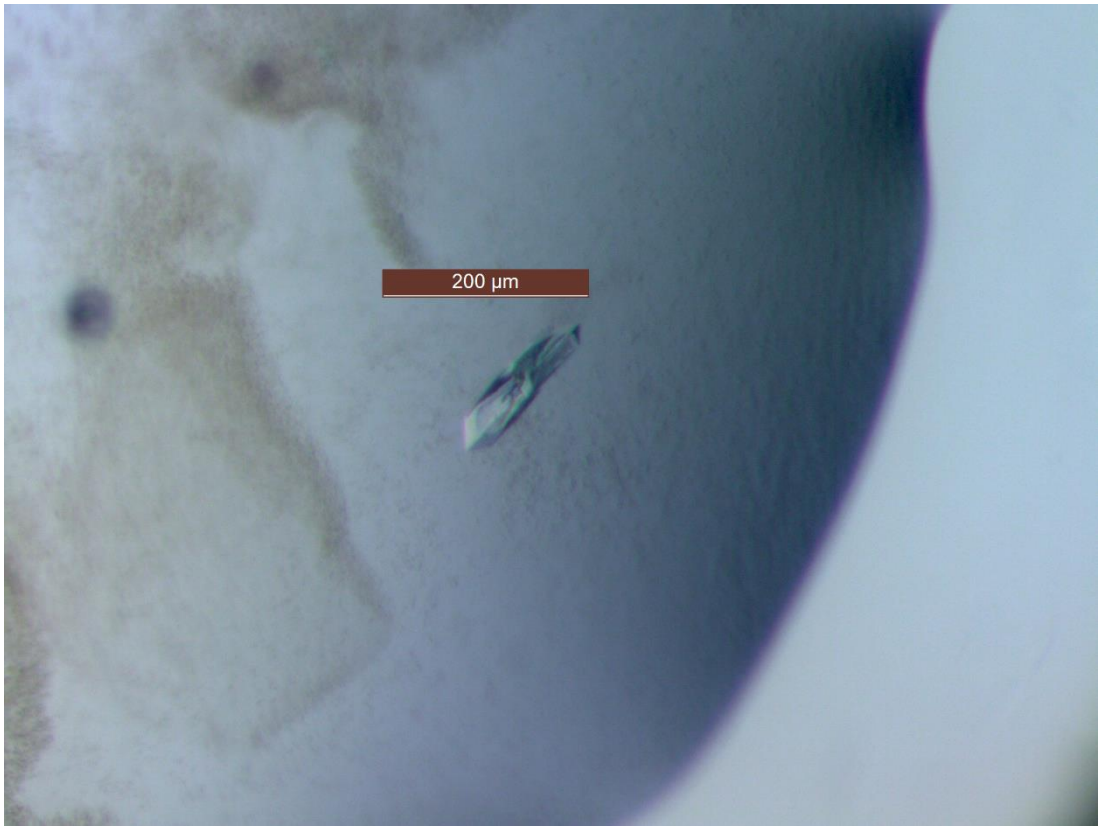


Figure 2.6.3 Flat needles formed by 150 µM M2-1 in 0.2 $MgCl_2$, 0.1m Tris-HCl, pH 8.0 with 20% [w/v] PEG 6000.

Optimisation is used to test different buffer conditions around the hit condition to try and improve the shape, size and stability of the initial hit. The hanging drop method works in much the same way as a sitting drop, with a vapour equilibrium between the well reservoir and drop being formed.[16] This method is popular for crystallisations of macromolecules as the crystals formed can be easily accessed and the drops resealed quickly. Multiple drops, each with a different protein to buffer ratio can be tested in the same well. The needles were tested using the in-house diffractometer and no diffraction was obtained, this could be due to a non-diffracting protein crystal or an amorphous object. An increase in concentration may further help to form the 3rd dimension of the crystal that is lacking in 2D needles.

Further screens at 195 μM M2-1 concentration were tested. These gave dipyramidal crystals of approximately 50 μm in length as shown below in Figure 2.6.4. The conditions in the drop were found in Pact Premier, E5, 0.2M calcuim acetate, 0.1M Tris at pH 7.5 and 25% PEG 2000 MME.[101]

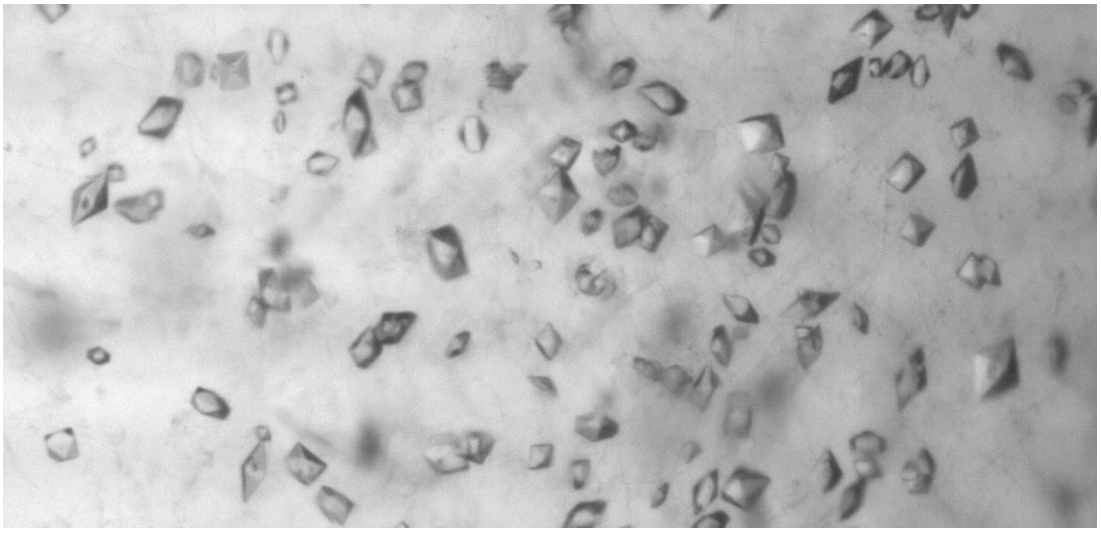


Figure 2.6.4. 195 μM M2-1 in 0.2M Calcuim Acetate, 0.1M Tris at pH 7.5 and 25% PEG 2000 MME. Crystal length was approximately 50 μm in length.

Seeding was used to try and grow larger crystals as the initial screen yielded small fragile crystals which would be susceptible to radiation damage during data collection. The crystals for seeding were crushed *in situ*, and then a cat whisker was coated in crushed crystals. The cat whisker was used in a straight line through the new optimisation drops to leave small amounts of crushed crystals in the well. This formed small nucleation sites where new crystals could form and the tray was incubated at 22 °C.

After 2 weeks the trays had grown the same crystals and to the same size. These were harvested and frozen using liquid nitrogen and sent for analysis at Diamond, Oxfordshire, UK. Cryo-protectant of 30% glycerol was used to protect the crystal from cracking during freezing. P.D Townsend tested the crystals and no diffraction was found. Unfortunately the structure was solved by Tanner *et al.*[5] in 2014 before more crystals could be grown and tested.

2.7 CONCLUSIONS

Overall, this chapter summarises the characterisation of M2-1 through biophysical techniques. The aim of the chapter was to understand the structure of the protein so further binding studies with other structural proteins in hRSV are possible. Without a full length structure it is extremely difficult to understand where possible contacts and binding may occur.

Firstly the protein was cloned by R. P. Yeo at Durham University in 2010. DNA sequencing confirmed that the protein ORF was full length and no mutations had occurred in the cloning process. The Author then began protein expression trials using *E.coli* to produce a high enough yield for biophysical characterisation. Expression through *E. coli* was preferred for the relative cost of the research and the yield of protein produced. A GST affinity tag had been used in the cloning process, which allowed for a clean purification process. At the N terminus of M2-1, a zinc finger was a key structural finger which had been characterised previously.[61] Further research showed a more stable protein was produced when zinc ion was added to the expression broth.

MALDI-TOF analysis confirmed the sequence identity of the expression protein, M2-1. This was followed by ICP-MS to determine if the zinc finger was full occupied and to understand the stability of the protein in the presence of zinc ions.

Dynamic light scattering and Circular Dichorism data were already available for M2-1, but this was repeated to ensure the protein was folded in the correct state. The results for these experiments correlated with the known values, here the diameter for the tetrameric M2-1 was found to be approximately 16 nm and it was observed to be 53% alpha helical in structure.[62]

Thermal Shift assays were used to determine stabilising buffer conditions through 96 well screening.[95] The screen contained many different conditions from metal ions through to amino acids and varying salt conditions. The most stabilising compounds were used as additives in the buffers for purification and crystallisation. Stabilising compounds were found to be L-Histidine, L-Arginine-ethyl ester, Glycine-glycine dipeptide, Gly-Gly-Gly tripeptide and diethylamine.

In 2011 at the beginning of the Author's period of study, no protein crystal structure for M2-1 was available. Many efforts were tried as detailed in the chapter, unfortunately Tanner *et al.* achieved diffracting crystals and subsequently published a full length structure in 2014[5] and at this point crystallisation efforts were abandoned.

2.8 METHODS

2.8.1 EXPRESSION AND PURIFICATION OF M2-1 PROTEIN.

A pGEX-6 vector containing the ORF of M2-1 protein, obtained from R. P. Yeo, was transformed into *E. coli* BL21-Codonplus and grown on ampicillin LB agar plates overnight at 37 °C. One colony was picked from the resulting plate and was incubated in 0.2 L of LB broth (Tryptone 10 g/L, sodium chloride 10 g/L, yeast extract 5 g/L; Melford Laboratories Ltd., Suffolk, UK, 25 g/L) overnight at 37 °C. 0.01 L of the starter culture was used to inoculate 1 L of LB Broth in a 2.5 L conical flask, the bacteria were grown in the presence of ampicillin (100 mg/mL) and carbenicillin (100 mg/mL) at 37 °C for 5 hours or until optical density at 600 nm reached 0.6-0.8. Expression was induced using isopropyl-β-D-1-thiogalactopyranoside (IPTG, 0.8 mM; Sigma Aldrich, UK) and left to shake overnight at 28 °C at 180 rpm. The bacterial cells were harvested by centrifugation at 4000 × g for 20 minutes in a Beckmann Coulter ™ J-20 XP centrifuge, using rotor JLA 8000 and the cell pellet was stored at -80 °C.

Protein extraction was achieved through resuspending the pelleted cells in 0.02 L of lysis buffer (50 mM Tris,1 M NaOH, 5 mM $CaCl_2$, 5 mM $MgCl_2$ pH 7.6). Lysozyme (1 mg/mL), DNase (1 µg/mL), RNase (1 µg/mL) and protease inhibitor tablet (cOmplete Mini™, EDTA-Free, Roche Diagnostics Ltd, West Sussex, UK) were added once the pellet was resuspended. The cells were lysed by sonication for 240 seconds at 20 kHz and 3-[(3-cholamindopropyl)dimethylammonio]-1-propanesulphonate hydrate (CHAPS, final concentration 8 mM in ddH_2O) was added to the mixture and centrifuged at 50,000 x g for 30 minutes Beckmann Coulter ™ J-20 XP centrifuge, using rotor JA 25.50 to remove cellular debris.

The supernatant was treated with pre-equilibrated glutathione-agarose beads (50% suspension, 500 µl/ mL) and mixed by inversion for 2 hours at 4 °C. This was then centrifuged in a Beckman Coulter™ Allegra™ X-22R centrifuge for 20 minutes at 4 °C to sediment the beads and the supernatant discarded. The beads were resuspended in 10 mL cleavage buffer (50 mM Tris,1 M NaOH, 1 mM DTT, pH 7.4)

The mixture was then loaded onto a clear plastic column. The buffer was allowed to flow through under gravity until just above the bead line, whereupon a 3 M NaCl high salt wash was added. The beads were washed a further 2 times with Precission cleavage buffer. The protein was eluted from the beads by the addition of Precission™ Protease (5 µL/ mL of lysate, GE Healthcare, UK). The column was stored at 4 °C for 48 hours. The

protein was eluted with successive washes of Precission Cleavage buffer (3 mL total volume), fractions were analysed by SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis)

Fractions containing the protein were loaded onto a Superdex 16/60 200 size exclusion Fast Protein Liquid Chromatography column on an AKTA prime system (G.E. Healthcare, Buckinghamshire, UK) and were eluted in 50 mM Tris, 200 mM NaCl , pH 7 at 0.8 mPa pressure. Where protein was present, a UV-absorbance was shown. These fractions were analysed again by SDS-PAGE and the protein containing fractions were combined.

2.8.2 THERMAL SHIFT ASSAYS (TSA)

Fluorescence data was collected on an Applied Biosystems 7500 FAST Real-Time PCR System with an excitation range of 455-485 nm. 8 µL of M2-1 (1mg/mL) was pipetted into each well of a 96-well plate Semi-Skirted FAST trays (Starlabs, Suffolk, UK) and 10 µL different known buffers (detailed in Appendix A.3) were added into each well. 2 µL of SYPRO Orange (5 fold concentration) was also added to give 20 µL in each well.

The fluorescence emission signal at 567-596 nm was used for data analysis. The temperature was held for 1 minute per degree from 24 to 95 °C. *NAMI*, a Python programme written by M. Grøftehauge and N. Hajizadeh, was used to analyse the data collected.[95]

2.8.3 SDS-PAGE (SODIUM DODECYL SULPHATE POLYACRYLAMIDE GEL ELECTROPHORESIS)

The purified protein was analysed using a 15% acrylamide resolving gel: 5 mL ProtoFlow 30% acrylamide gel (Flowgen Bioscience Ltd., Nottingham, UK), 2.5 mL buffer (1.5 M Tris, 0.4 % (w/v) SDS pH 8.8), 2.3 mL water, 0.1 mL 10 % SDS (Aldrich, Poole, UK), 0.1 mL 10 % (w/v) ammonium persulphate (APS) solution in water and 0.004 mL N,N,N',N'-tetramethylethylenediamine (TEMED; Fluka). The stacking gel was made from: 1.7 mL ProtoFlow 30 % acrylamide gel (Flowgen), 2.5 mL buffer (0.5 M Tris, 0.4 % (w/v) SDS pH 6.8), 5.55 mL water, 0.1 mL 10 % SDS (Aldrich,), 0.1 mL 10 % (w/v) APS solution in water and 0.01 mL TEMED (Fluka). Protein samples were prepared by addition of protein loading buffer (2.5 mL 1 M Tris-HCl pH 6.8, 0.5 mL ddH_20, 1.0g SDS,

0.8 mL 0.1% bromophenol blue, 4 mL 100% Glycerol, 2 mL 14.3 M β-mercaptoethanol (100% stock) and adjusted to 10 mL using ddH_20, 5 µL in each 15 µL sample) and heated to 95 °C for 5 minutes then loaded onto the gel with a protein ladder molecular weight marker in the first well (10 – 250 kDa; New England Biolabs., Hertfordshire, UK). The gel was run at 200 V for 1 hour and stained using InstantBlue™ (Expedeon, Cambridge, UK)

2.8.3. DETERMINATION PROTEIN CONCENTRATION

2.8.3.1 NANODROP 1000

The concentration of the protein was determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Kent, UK) by measuring absorbance at 280 nm where 1 IAU = 1 mg / ml. The protein concentration was increased using a Vivaspin 6 (MWCO 10,000; GE Healthcare Life Sciences). Samples were centrifuged at 4000 g in a Beckman Coulter™ Allegra™ X-22R centrifuge for 20 minutes at 4 °C, tested using the NanoDrop 1000 spectrophotometer and repeated until the desired concentration was achieved.

2.8.3.2 BRADFORD ASSAY

Protein concentration was also verified by Bradford assay. Standard concentrations between 5 and 100 µM of bovine serum albumin were prepared and allowed to equilibrate at room temperature for 30 minutes. Each sample was diluted using MilliQ water and Bradford reagent added. Each concentration was measured at 595 nm using UV-visable spectrophotometer (Multskan Go, Thermo Scientific, UK) and the absorbance was plotted against the concentration to give a linear plot. The analyte protein was diluted and prepared in the same way and measured at 595 nm. The resulting absorbance was plotted on the standard graph and the concentration was found on the x-axis.

2.8.4 DYNAMIC LIGHT SCATTERING

A Malvern Zetasizer µV DLS system (Malvern Instruments Ltd., Worcestershire, UK) was used to conduct all DLS measurements in this work. The sample was held at 25 °C and

allowed to equilibrate for 60 seconds prior to analysis. Each size measurement was determined from 13 runs of 10 seconds in 50 µL cuvettes. (Starlabs, Suffolk, UK)

2.8.5 CRYSTALLISATION SCREEN METHODS

2.8.5.1 96-WELL SITTING DROP TRAYS

Crystallisation conditions were screened using the sitting drop vapour diffusion method, together with JCSG+™,[101] Pact premier™,[101] Structure screen 1™,[103] Structure screen 2™ and Clear Strategy screen™[101] 96-well MRC crystallisation plate (Molecular Dimensions, Suffolk, UK) was filled with a mix of well solution from the above screens. Protein (30 nL and 60 nL) was dispensed by a crystallisation robot (Innovadyne Screenmaker 96+8, Solve Scientific, Australia) onto the sitting wells using a predetermined 20 psi method at Durham University. The well contents were then added to the drops to a total volume of 100 nL. The drops were sealed with ClearVue sheets (Molecular Dimensions, Suffolk, UK) and placed at temperatures ranging from 4 °C to 22 °C until growth was observed.

2.8.5.2. 24-WELL OPTIMISATION SCREEN

24 well clear plastic XRL™ (Molecular Dimensions, Suffolk, UK) were used to contain the desired well solution. The well solution (1 mL total volume) was a variation of the 96-well sitting drop tray hit to further optimise conditions. The protein and well solution were pipetted onto 24 mm siliconized round cover slips (Molecular Dimensions, Suffolk, UK) in various concentrations and sealed with sealing grease (Molecular Dimensions, Suffolk, UK). The trays were placed at temperatures ranging from 4 °C to 22 °C until growth was observed.

2.8.5.3. STREAK SEEDING

Small Crystals previously grown in other 96 or 24 well trays were crushed *in-situ* by a small crystallisation trowel (Molecular Dimensions, Suffolk, UK). A cats whisker was coated with the crushed crystals and moved through a new protein-buffer drop on a 24 mm siliconized round cover slip (Molecular Dimensions, Suffolk, UK) from left to right. The slip was sealed using sealing grease (Molecular Dimensions, Suffolk, UK) and placed at temperatures ranging from 4 °C to 22 °C until growth was observed.

2.8.5.4 CRYSTAL MOUNTING.

A cryoprotectant solution was produced by supplementing the mother liquor with glycerol (30% v/v). The crystals were harvested in round litholoops (Molecular Dimensions, Suffolk, UK) then the crystal and loop were covered in cryoprotectant solution before flash freezing in liquid N_2.

CHAPTER 3. PROBING THE INTERACTION BETWEEN MATRIX PROTEIN AND M2-1.

Previously it was thought that M2-1 was located only in the cytosol of the virus and in the host cell cytoplasm, but through cryotomography, it has been seen that M2-1 is likely to be a major structural protein.[78] This newly discovered role of M2-1 reveals an interaction with the matrix protein and a possible interaction with the nucleocapsid and ribonucleocapsid formations. The work completed by Li *et al.*[58] using hRSV M protein in hRSV-infected cells shows that the matrix protein inhibits viral transcription and colocalises to M2-1 in the cytoplasm of cells.[58] Further to this, nucleocapsid protein and matrix protein were shown to directly interact with M2-1 in a cell-free binding assay.[58] Li *et al.* have also shown that it is the N-terminal 110 amino acids of the matrix protein which interact with M2-1.[58] This interaction of M2-1 and M, with a focus on the binding N-terminal region of M, could be a target for antiviral research. Disruption of this interaction has the potential to block the interaction of M with the nucleocapsid and inhibit virus assembly. There are no reported data on the binding domain of M2-1 responsible for this interaction, in contrast to its interaction with the phosphoprotein which has been extensively characterised.[61]

This chapter discusses the preliminary investigations into the interaction of M2-1 with matrix protein. Firstly, direct binding will be shown and confirmed *in vitro* by the use of pulldown assays where SDS PAGE is used to show the proteins involved in binding. Once the binding was confirmed the interactions were investigated using Transmission Electron Microscopy to visualise the aggregates which formed.

Further characterisation of the interaction using Surface Plasmon Resonance and Fluorescence Anisotropy was used to gain an insight into the thermodynamic pathway and to understand the stoichiometry of the proteins involved. This work is key to further understanding how the viral proteins interact and may give an insight into potential mechanisms to hinder that interaction.

3.1. PULLDOWN ASSAY.

A pulldown assay is an *in vitro* method, which can be used to indicate the binding of two proteins. The method works by using an immobilised bait protein, in this case M2-1 bound to glutathione beads. The protein with the suspected interaction (a prey protein) was then passed over the beads.[104] Systems in which the prey and bait proteins interact will see the former, in this case M2-1 bound to the beads and if the second protein interacts with M2-1 binds and is eluted with the beads. If there is no interaction then in the subsequent wash steps the second protein is washed from the mixture. The results of this experiment are shown in Figure 3.1.1 where the SDS-PAGE shows that M2-1 binds the matrix protein in the experiment. Lane 2 shows the GST tagged M2-1 protein and also the free GST tag, this appears as a smear on the SDS PAGE gel as cell lysate was used as the crude source of GST-M2-1.

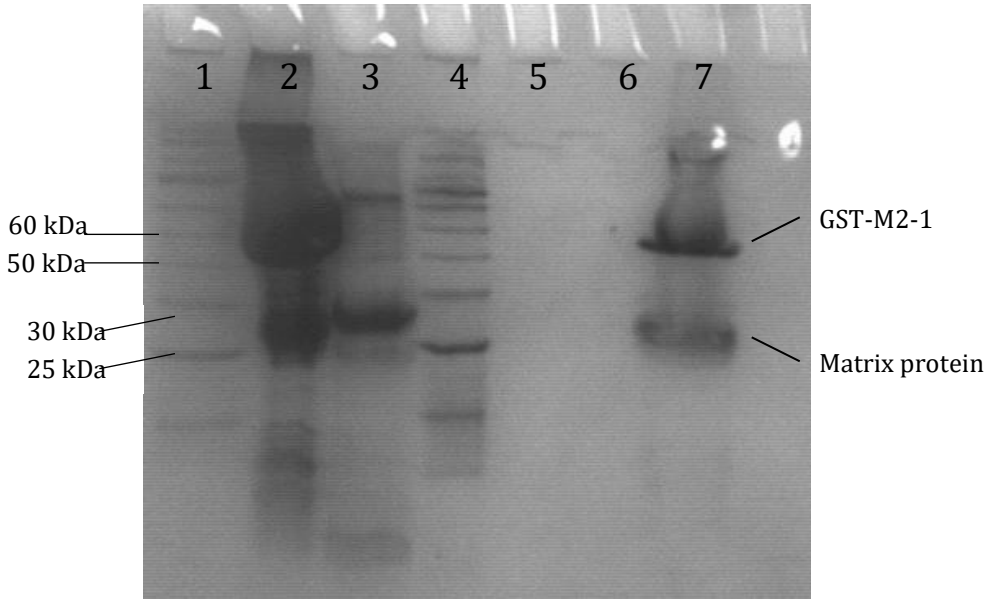


Figure 3.1.1. 15% SDS-PAGE gel, Lane 1 – NEB Protein Ladder (10 kDa-250 kDa) Lane 2 – GST fused M2-1, Lane 3 – Matrix protein. Lane 4 – NEB Protein Ladder (10 kDa-250 kDa), Lane 5 – Blank GST slurry, Lane 6 – Blank beads and matrix protein incubation, Lane 7 – GST fused M2-1 and matrix protein.

3.2. TRANSMISSION ELECTRON MICROSCOPY (TEM).

A Transmission Electron microscope (TEM) uses a beam of electrons instead of light to view the required sample. This is achieved through a high voltage electron beam that is focussed by electromagnetic lenses onto a sample that has been fixed to a graphite matrix film by glutaldehyde crosslinking.[105] Glutaldehyde cross-linking protects the structures formed from radiation damage by the electron beam and vacuum. Fixed samples were stained using uranyl acetate and air dried before being imaged in the microscope.[106] A negative image, as seen below, was produced where the darker colour is the background and the lighter image is the protein structure. This contrast is relative to the concentration of the dye used so this was held constant for the different sets of proteins that were tested.

Initially each individual protein alone was fixed to the film as a negative control to the complexed proteins. These images showed white shapes on a black background. The individual protein molecules were too small for a high resolution image to be obtained. The images showed molecules of M and M2-1 approximately ~20 nm in diameter as shown in Figure 3.2.1, which agrees with the previously reported size of the M2-1 tetramer in solution[62] and matrix protein in solution.

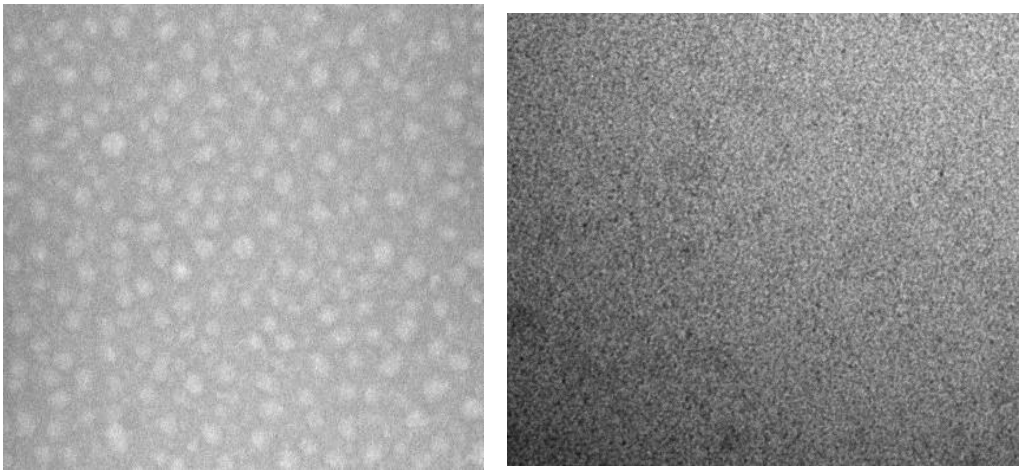


Figure 3.2.1 Left, M2-1 protein ~12 nm, 120,000 × zoom. Right, Matrix protein TEM image, ~7 nm 100 000 × zoom.

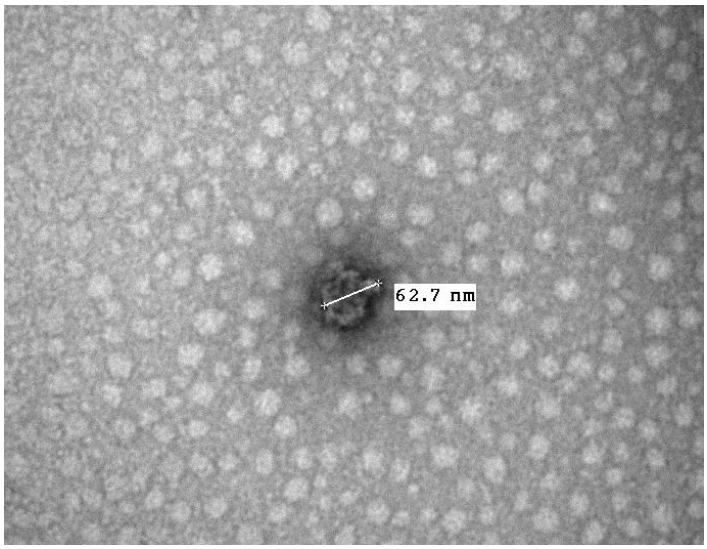


Figure 3.2.2. TEM image of a protein aggregate at ~63nm.

The image of the mixture of the two proteins M and M2-1, is shown in Figure 3.2.2. A number of different aggregates could be seen as shown in Figure 3.2.3, indicating a number of different conformations in the solution. Each of the aggregates has a dark centre, indicative of a point on the film where there is no protein. This could show that the complex has a ring style structure with a centre which is hollow.

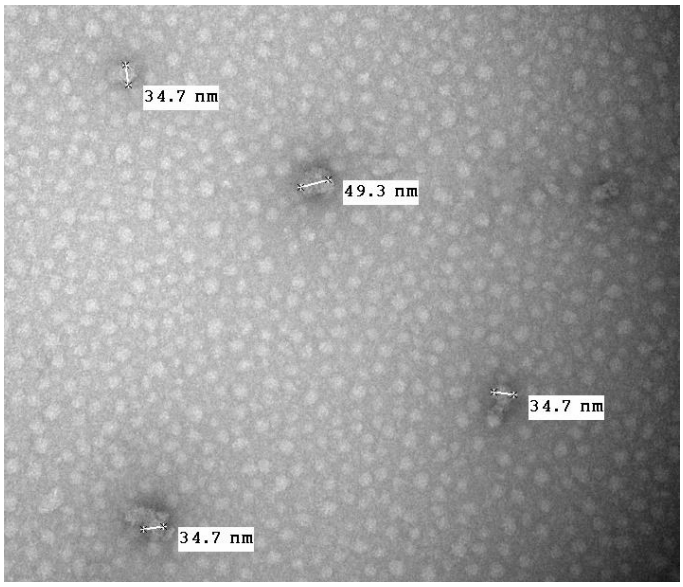


Figure 3.2.3 Smaller aggregates seen on the TEM at 120 000 × zoom.

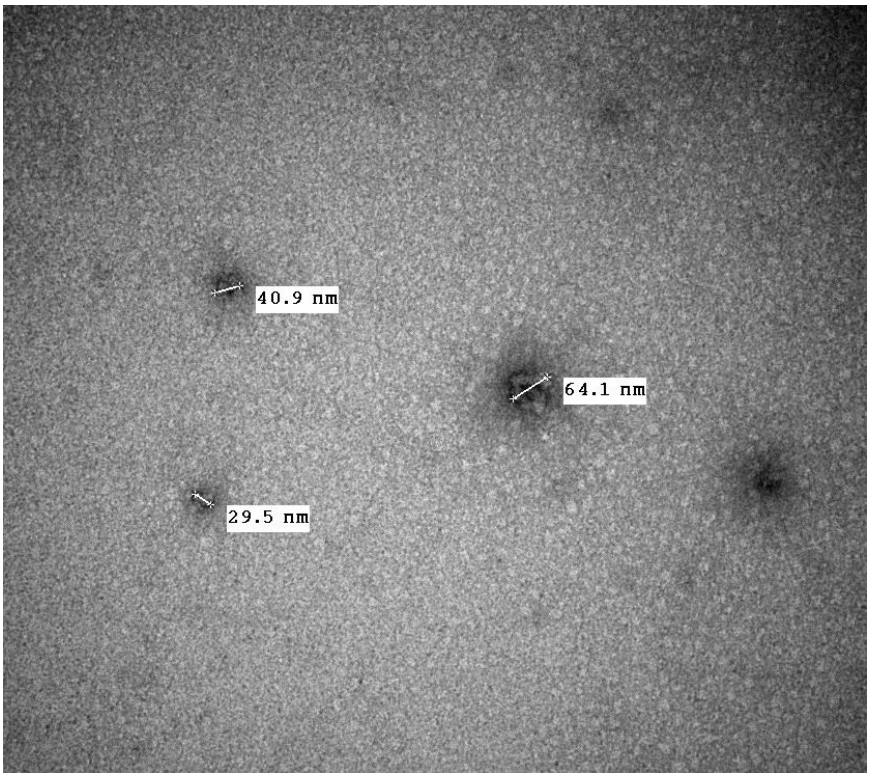


Figure 3.2.4 Aggregates at ~ 40 nm and 60 nm at 120 000 × zoom.

The hMPV M2-1 was crystallised in a 'open' formation as discussed in Chapter 1, where one monomer from the tetrameric model was swung 60 Å away from the other subunits.[50] This is a mechanism of interaction that cannot be ruled out for hRSV M2-1, but the aggregates do not seem to have an obvious protrusion showing a monomer unit. There is, however, a large dark centre in the ~ 60 nm aggregate, as this is a negative stain this points towards a hole in the centre of the larger aggregate, indicating a ring like structure.

From the results shown in Figures 3.2.2 and 3.2.3, two main species have been identified, one at approximately 40 nm and one at 60 nm in diameter. This is in comparison to the crystal structures of the two proteins which show M2-1 to be ~ 80 Å (8 nm) and M diameter in ~ 72 Å (7 nm), from point to point. The structures are similar in shape and dimensions so the comparisons could extend to the binding of M2-1 and M in the hMPV, where it could be predicted that the proteins would bind in a similar way. The monomer size corresponds well to the published dynamic light scattering data of 15 nm ± 4 nm[62] and the initial TEM images correlate well to the published data so the new species sizes would be expected to be indicative of the complexes found.

The new complexes that have been imaged in this chapter show that an interaction between the two proteins does give new complexes. Unfortunately, high resolution data are not obtainable using this technique, however further investigation through cryo-electron microscopy would allow us to reevaluate the 3D structure of the complexed species. The technique is not available at Durham University and the experiment could not be completed in the Author's research time.

3.3 SURFACE PLASMON RESONANCE

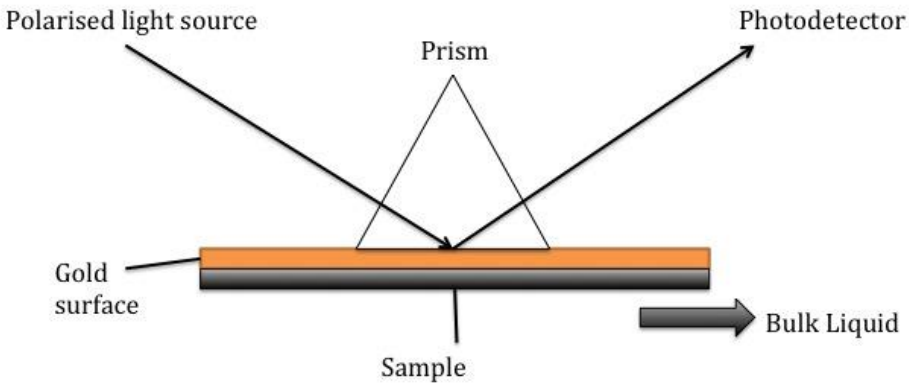


Figure 3.3.1 Surface plasmon resonance.[107]

This technique can be used to determine thermodynamic data for the binding between two species using a flow method, where one species is attached to an active surface as shown in Figure 3.3.1. The technique uses Maxwell's theory where the free electrons of the gold surface on the chip are present as plasma.[108] Plasma is an electron liquid of high density,[107] and density fluctuations of the surface liquid are called plasmons.[107] The signal given is based on the excitation of surface plasmons when the polarised light hits an electrically conducting gold layer. At a specific angle the excitation occurs and the resulting reflected light is of reduced intensity.[107] A change in the interface, i.e. when a sample is bound, will change the refractive index and so the thickness of the sample film leads to change in signal.[107]

There are multiple ways that protein can be attached to the chip, these vary between manufacturers. The chip is commonly a glass surface which is coated with gold, then a self-assembled monolayer (SAM) is constructed on the gold surface.[108] The SAM are inert organic molecules which are used as a linker layer to which non-crosslinked carboxymethylated dextran can adhere.[107]

The dextran is a chain of 1,6 – linked glucose molecules which has very low non-specific absorption and a head region which can be activated to covalently bind ligands to the surface of the chip by amine-, thiol- or aldehyde chemistry dependant on the type of

chip.[108] A CM5 chip is a general chip containing only dextrans on the surface, and is appropriate for the interaction analysis of all types of biomolecules from small organic molecules to proteins. The head region of the dextran chain is activated through carboxyl groups and functional groups on the ligand can then be coupled to the active carboxyl groups.[107] Initially this chip was utilised as a starting point for analysis of the interaction between M and M2-1 so each individual protein could be tested attached to the chip. To activate the dextran chain activation buffers of (1-ethyl-3-[3-dimethyaminopropyl]carbodiimide hydrochloride (EDC) and *N*-hydroxysuccinimide (NHS) were injected through the machine to give an active surface to which the protein could bind.[108]

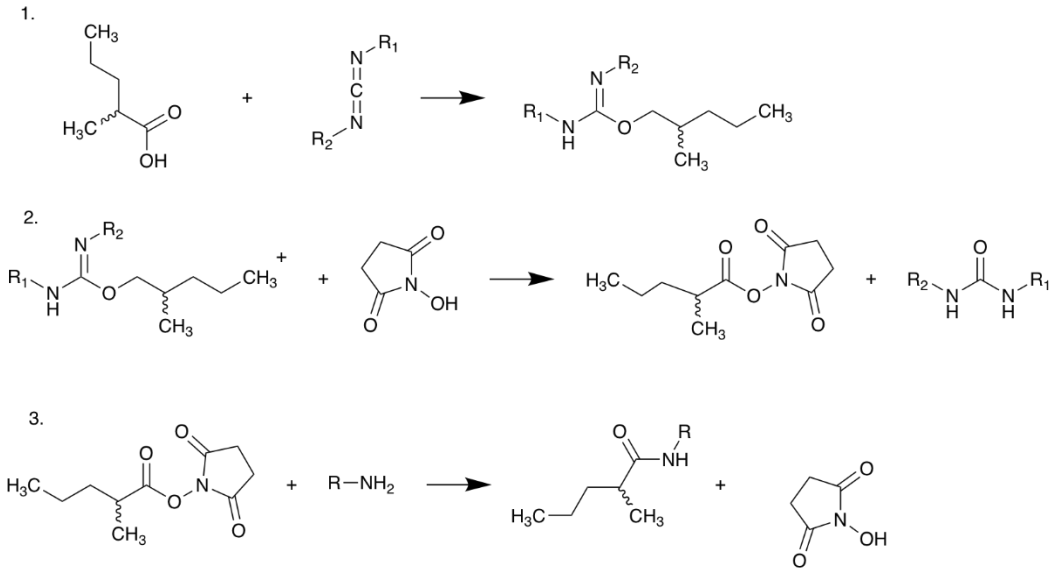


Figure 3.3.2 1 and 2. Activation of the carboxyl group with 1-ethyl-3-[3-dimethyaminopropyl]carbodiimide hydrochloride (EDC) and *N*-hydroxysuccinimide (NHS). 3. Covalent attachment of the ligand by a primary amine. [108]

It was found that M2-1 did not attach successfully to the chip, where a response of -5 RU was seen after deactivation of the free unattached head group of the dextran strands. The deactivation is a key step to ensure no background signal is recorded from the unspecific binding that can occur through protein groups.[108] Through multiple optimization steps it was decided that this type of chip was unsuitable for this experiment. An NTA chip was the secondary choice with the same dextran chains on the chip but this time the active head consisted of NTA (nitrilotriacetic acid). The head region was activated by chelating with nickel, this has a high affinity ($\sim 10^8$ M) towards a 6× histidine (His_6) tag on the protein of interest. The immobilization using this chip via

His_6 has the benefit of orientating the ligand molecules in a homogenous way.[107]

3.3.1 MATRIX PROTEIN CONCENTRATION SCOUTING.

As M2-1 did not attach successfully to the CM5 chip, the NTA chip was used. The matrix protein was previously cloned with a His_6 tag needed for the NTA chip. There is no hard and fast rule on how to attach the protein of interest to the chip, this is entirely dependent on the size and shape of the protein in question.[107] The quantity of protein bound to the chip should be sufficient to give a response on the sensogram and also be sufficiently stable to give a steady baseline. A ten-fold concentration range was advised for a concentration scout by the G.E Healthcare NTA Chip user manual.[109] Conditions for optimum loading were determined over a broad concentration range and were found to be 2 μM to 20 μM of M2-1 as shown in Figure 3.3.1.1 below.

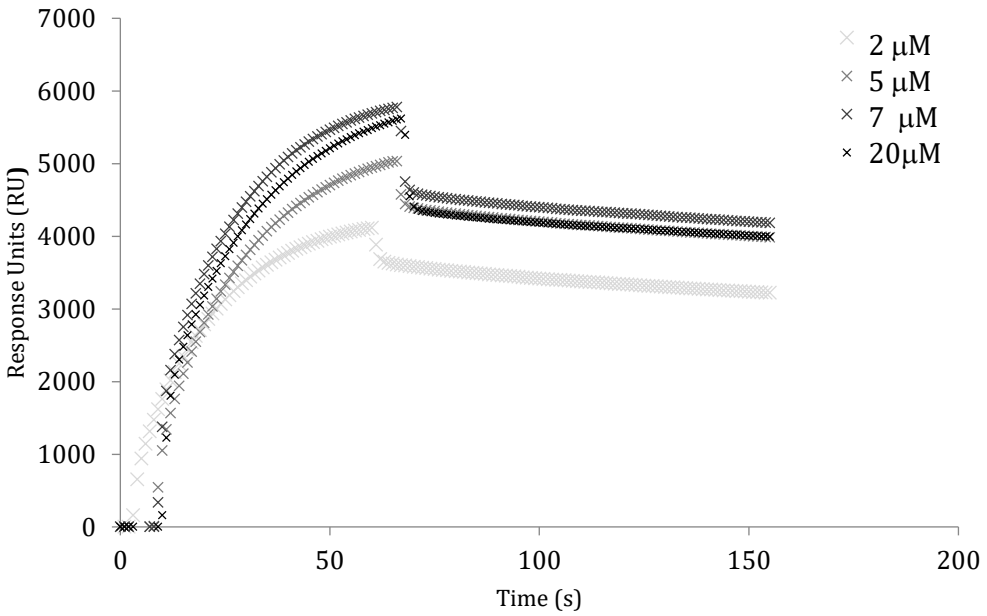


Figure 3.3.1.1 His_6-Matrix protein concentration scout. The dissociation curve is shown from 70 s onwards.

It can be seen that above 5 μM there was no higher response (RU) from injecting higher concentrations of protein. This suggests that the protein saturated the chip and all activated NTA dextran was bound. In using the NTA chip there must be a compromise between the stability of the protein on the chip and the response. If too much protein binds the decay of the signal from the chip could give an unstable wandering baseline for the addition of the second protein.[108] 3 μM protein of His_6-tagged matrix protein was used in all subsequent experiments.

3.3.2. ADDITION OF M2-1

Matrix protein was added at 3 μM to the chip, so a 2.5-20 μM range of M2-1 was tested initially. M2-1 protein solution was flowed across at 30 μL/min. The final sensogram is shown in Figure 3.3.2.1 where each step is labelled A-D. The chip has to be activated first so the NTA group can bind nickel as shown in Step A. Step B includes the addition of a set concentration of matrix protein and step C is the addition of M2-1 in the chosen concentration range. Step D is the regeneration of the chip where the EDTA solution deactivates the NTA group and removes any bound material. The process is repeated where only step C changes to test the whole M2-1 concentration range. After normalising the data and subtracting a blank run the final data are shown in Figure 3.3.2.2.

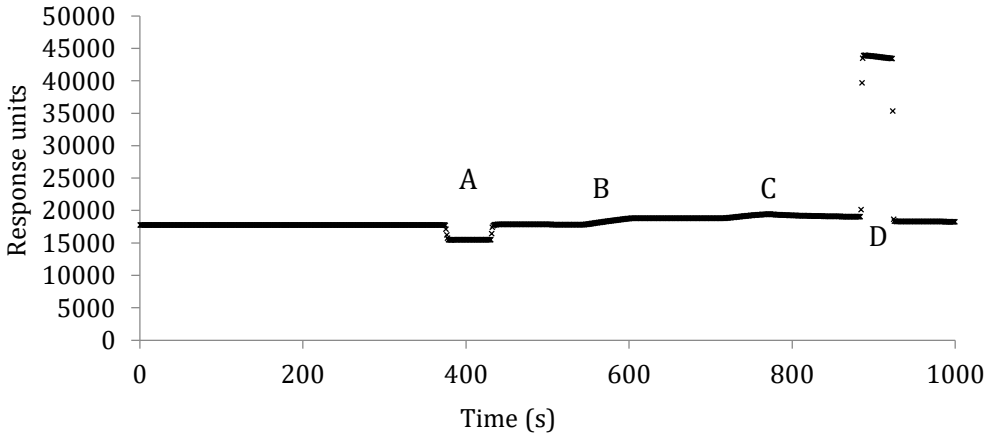


Figure 3.3.2.1 Sensogram of Biacore 3000 run. A) $NiCl_2$ activation B) His Matrix protein, 3 μM. C) M2-1, varing concentrations. D) EDTA regeneration step.

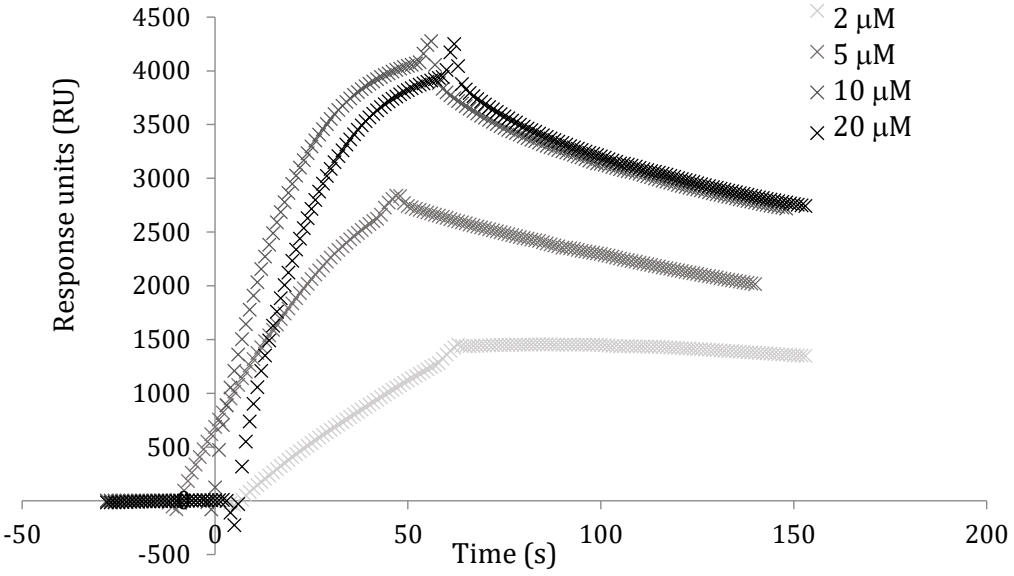


Figure 3.3.2.2 M2-1 binding on the immobilised matrix protein NTA chip. Concentrations used from 2.5 μM to 20 μM as shown in the legend.

Longer injections were used to achieve saturation of the association curve of each concentration. On analysis this did not change the overall fit of the curves and so smaller injections were subsequently made to preserve protein.

3.3.3 DATA ANALYSIS

Using the BIAEvaluation software, supplied by G. E Healthcare, analysis of the different binding curves could be completed. The software analyses data using a simple 1:1 binding mode where each curve was analysed using two separate algorithms, local and global fitting. In the local fitting each curve was treated as a separate experiment with its own k_{obs} rate constant. The averaged equilibrium constants and association curve fitting are shown below.

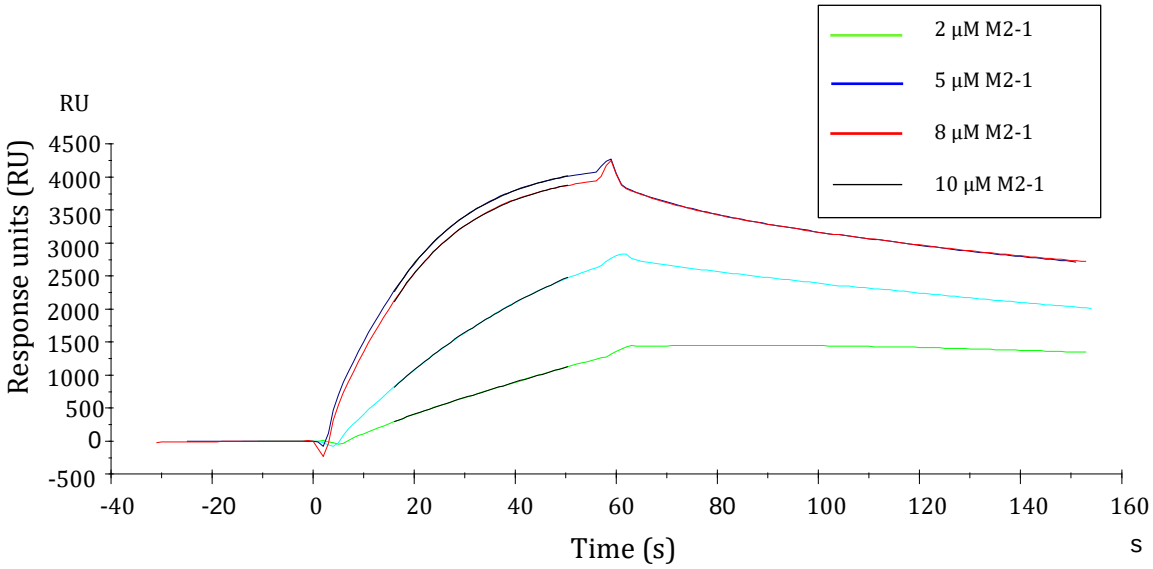


Figure 3.3.3.1 Association data local fitting curve for M2-1 addition to the immobilised matrix protein, shown in BIAevaluation software.[110]

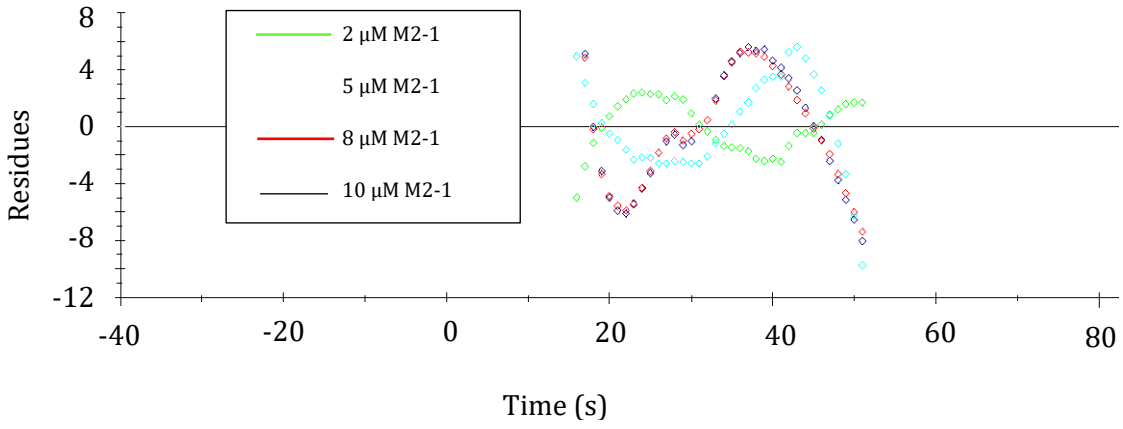


Figure 3.3.3.2 Residual plots of the curve fitting of M2-1 addition to the immobilised Matrix protein as shown in BIAevalution software.[110]

A residual (e) is the difference between observed value of the dependant variable (y) and the predicted value (ŷ). Plotting the values of each curve shows the statistical difference and the similarity of the fit, for example a perfect fit would be a horizontal line at residual value 0. In Figure 3.3.3.2 above, the maxima is +4 and the minimum of the curves at -8 residuals, which are low values and fit with a 1:1 interaction. These are used in comparison to the global analysis of the same data to understand which fit is the most appropriate and this is discussed later in this chapter.

k_a	k_d	χ^2
$3.66 \times 10^3\ M^{-1}\ s^{-1}$	$2.73 \times 10^{-3}\ s^{-1}$	14.1

Table 3.3.3.3 M and M2-1 interaction. Where k_d is the dissociation rate constant of the two proteins, k_a is the association rate constant and χ^2 is the fitness score.

The interaction was found to be 1:1 binding by analysis of local fit and the χ^2 value, which is a fitness score, is low in accordance with a good fit. The 1:1 binding is in contrast to the initial indications from the TEM where two major species were found, which could indicate multiple modes of action. This could suggest that one of the two species seen in TEM could be undergoing a rearrangement to allow further binding in one form as shown below in Figure 3.3.3.4. or that this species does not bind to the chip.

A)

$$L + P \underset{k_d}{\overset{k_a}{\rightleftharpoons}} PL$$

B)

$$L + P \underset{k_d}{\overset{k_a}{\rightleftharpoons}} P_nL_n^* \underset{k'_d}{\overset{k'_a}{\rightleftharpoons}} P_nL_n$$

Figure 3.3.3.4 A) Simple 1:1 Binding, where L is a Ligand and P is protein B) A rearrangement step shows the formation of a new intermediate species.

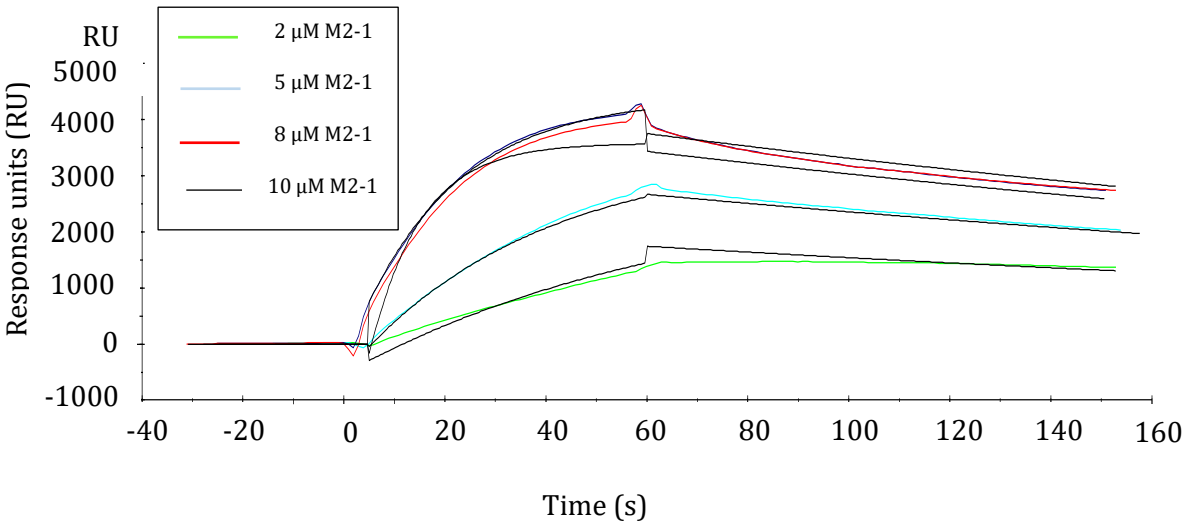


Figure 3.3.3.5 Global fitting curve for the experimental curve of M2-1 addition to the immobilised matrix protein as shown in BIAevalution software.[110]

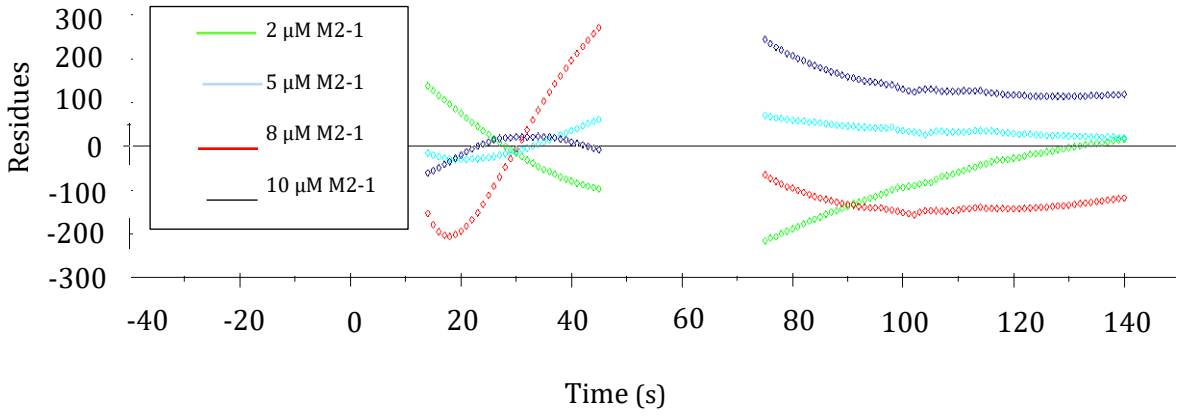


Figure 3.3.3.6 Residual plots for curve fitting using a global method of 1:1 interaction of M2-1 to the immobilised matrix protein as shown in BIAevalution software.[110]

The global analysis curves and the residual plots show a larger difference in the fitting of the curve to the experimental data than the local analysis plots. This is seen where the curve fitting graph shown in Figure 3.3.3.5 shows a large difference in the goodness of

fit parameters for the curves shown above in Figure 3.3.3.6. The residual plots vary from -200 to above 300, which is in contrast to the small variance shown in the local fit.

In the curve fitting in Figure 3.3.3.5 there is another feature that could contribute to the poor fit of the data. The fitted curves have a large drop where the injection stops and the dissociation curve begins. This is not seen in the experimental data and could be attributed to the inadequacy of the model of fitting.

k_a	k_d	χ^2
4.8×10^3 M^{-1} s^{-1}	3.11×10^{-3} s^{-1}	1.15×10^4

Table 3.3.3.7. Kinetic data for M and M2-1 using a global fit. Where k_d is the dissociation rate constant of the two proteins, k_a is the association rate constant and χ^2 is the fitness score.

Using both the local and global fittings, the equilibrium constant has been shown in this experiment to be 0.64 μM for a 1:1 interaction. It is known that M2-1 is a tetramer in solution[62] at these concentrations. In this case where the M protein is anchored by the His_6 tag dimerization may not be possible which could impact on the binding of the two proteins. The activated dextran binds to one of the two available N-terminal His_6 tags in the hRSV M dimer. This method of binding would mean that only one of the two available sites on matrix protein dimer would be used and one monomer would be anchored only by the interaction on the dimer interface. A high flow rate on the chip could disrupt the associated proteins and distort the results of this experiment. Further experimental data are needed using a technique where all the proteins are in solution, this could eliminate the problems seen in the SPR experiments where one protein is immobilised.

3.4. FLUORESCENCE ANISOTROPY

Fluorescence anisotropy is a technique which is used for many biophysical applications. These range from quantifying protein denaturation, measuring the dynamics of proteins in solution through to quantifying protein association with other macromolecules. This last use is particularly useful and will be applied to initially characterise the interaction between matrix protein and M2-1.

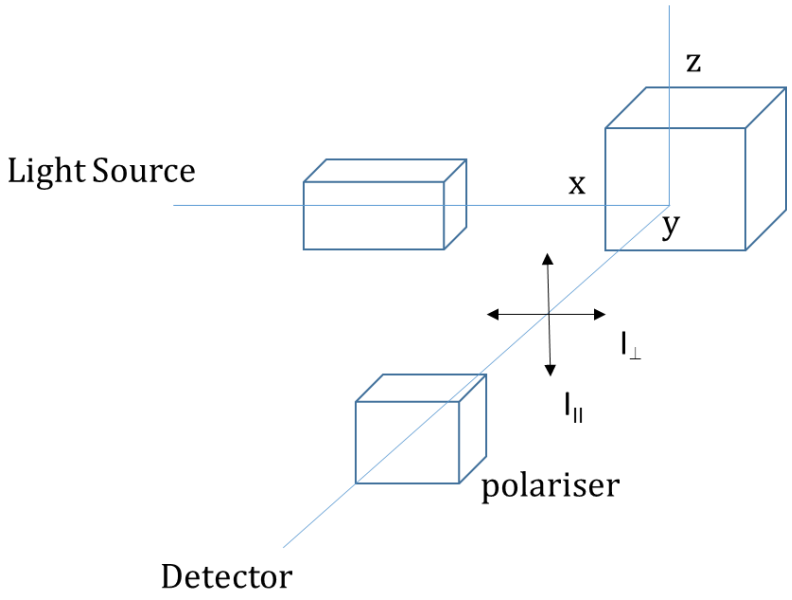


Figure 3.4.1. The components of a basic fluorescence anisotropy experiment.[111]

A basic fluorescence anisotropy experiment will comprise of the components shown above in Figure 3.4.1. The method begins when polarised light is shone on the fluorophore, an electron is excited into a higher energy state and extra energy is absorbed.[75] Fluorescence Anisotropy probes this transition moment at a specific orientation from the absorption of energy to the emission of the photon.[111,112] The difference between the transition moments is seen as depolarisation and can only occur when the light is located on a particular axis on the molecule and the emitted photon will have a specific polarisation.[112] Light is emitted at an angle dependant on the speed at which the molecule tumbles, whereas the rate of tumbling is related to the particle's size.[111,112]

For a signal to be seen and plotted, the sample is required to have a fluorophore bound. In this case the matrix protein was reacted with to Rhodamine Red succinimidyl ester, a

protocol for this in the VAM/JMS group was known and a highly labelled protein was produced. Previous labelling experiments by the group with M2-1 gave less than ideal results where a poor yield of protein was recovered. Rhodamine red (RR) X-succinamidyl ester is a dye shown in Figure 3.4.2, it is an amine specific fluorophore that attaches to the primary amine chains in lysine residues.

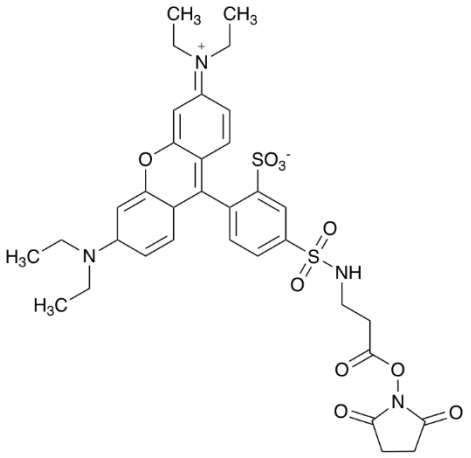


Figure 3.4.2. Rhodamine Red (RR) X- succinimidyl ester.

When the polarised light in the polarising fluorimeter is shone on to randomly oriented fluorophores, most of the energy will be emitted at a range of angles. The fluorophore can freely change orientation in solution before the emission of a photon and the degree of polarisation of this light will be reduced relative to the incidence beam. The difference in polarisation between the incident and emitted light depends on a number of factors such as the length of the fluorescence lifetime as shown below.

$$r = \frac{r_0}{1 + \frac{\tau}{\phi}}$$

Where r is the observed anisotropy, r_0 is the intrinsic anisotropy of the molecule, τ is the fluorescence lifetime and ϕ is the rotational time constant.[111]

In Fluorescence anisotropy, the fluorescent molecule involved in binding exists in the solution in one of two states, bound or free. Each state will have a unique anisotropy value and is summerised

$$A = F_f A_f + F_b A_b$$

Where $F_f + F_b = 1$, A= observed anisotropy value, F_f = fraction of fluorescent lignd that is free, F_b = fraction of fluorescent ligand that is bound, A_f = Anisotropy of the free fluorescent ligand, A_b = Anisotropy of the bound fluorescent ligand [113]

For protein-protein binding, the fluorescent signal (degree of polarisation) should be low at the start of the experiment, where Rhodamine Red (RR) bound matrix protein is turning unhindered in solution and no titrant is present.[112] The binding is indicated by an increase of the polarisation signal as the fluorophore is less able to change orientation due to the larger bound protein.[111] It is assumed that there is no quenching of the fluorophore occurring, as the fluorophore is not directly involved in the binding event. The resulting graph of the change in relative fluorescence is shown in Figure 3.4.3 below.

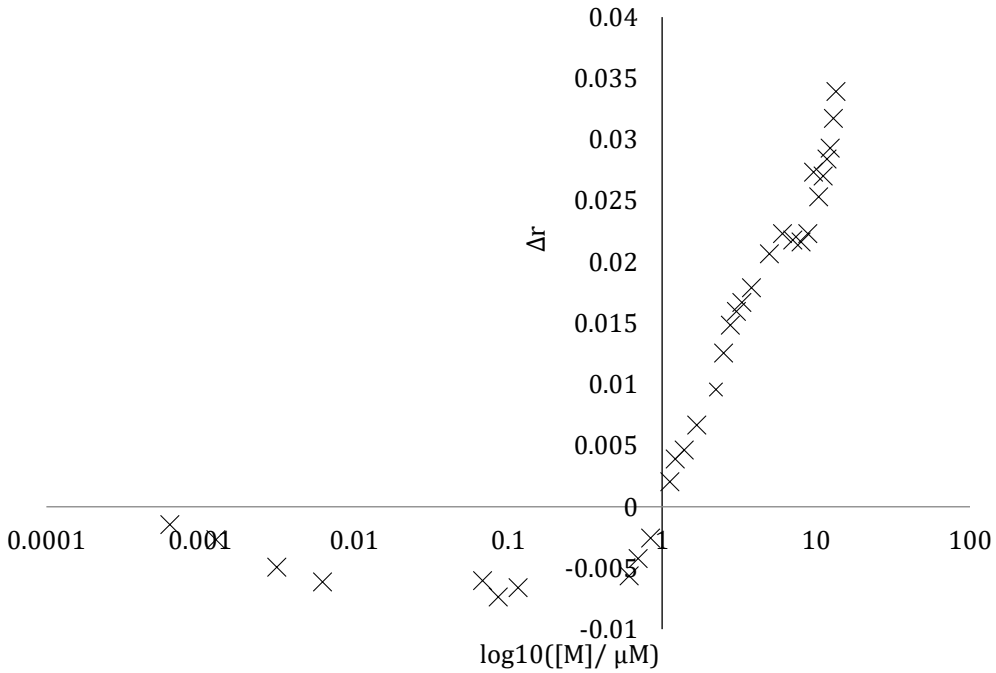


Figure 3.4.3 RR-matrix protein and M2-1 fluorescence anisotropy raw data.

There is a clear increase in Δr from approximately 0.7 μM through to 10 μM where the FR-matrix protein is rotating more slowly when bound to M2-1. Another binding curve can be seen above 10 μM, but further investigations were hindered due to protein precipitating out of solution.

The data were analysed by the author and J. M. Sanderson using a theoretical thermodynamic pathway, where the oligomerisation state of matrix protein was in question.

3.4.1 M OLIGOMERISATION MODEL

The thermodynamic model proposed, shown in Figure 3.4.1.1 includes all species thought to contribute to the raw data curve shown in Figure 3.4.3.

Scheme 3.4.1.1 Thermodynamic pathway for the binding equilibria of matrix protein to M2-1. Where M_n is the multimer of M which is in equilibrium with M as a monomer shown as K_m. M:M2-1 bind in a monomer:tetramer formation. This then leads to another M2-1 binding to M which is K_2.

Previously H. K. McPhee, Durham University[114] demonstrated that the matrix protein can be identified as a dimer, tetramer and a hexamer by chemical crosslinking and mass spectrometry analysis. Using the model above takes into account the potential for M oligomerisation, whilst allowing for interaction of monometic M with M2-1. If the model above in Scheme 3.4.1.1 is correct, it provides support to the binding model, where monomeric M binds to M2-1. Also another term where 2 M2-1 were bound to matrix protein allows for modelling of the formation of 2:1 stoichiometry potentially seen in the electron microscopy. The binding isotherm is shown below in Figure 3.4.1.2, which details all species in the mixture that contribute to the signal. Each [protein]/[M_{tot}] term determines the fraction of signal from each species relative to the total signal.

$$r_{calc} = r_M \frac{[M]}{[M_{tot}]} + r_n \frac{[M_n]}{[M_{tot}]} + r_{free}$$

Figure 3.4.1.2 Binding isotherm for M oligomerisation, M and M2-1 and M $(M2-1)_2$. M_n is the concentration of M in the oligomeric form and n is variable. Where r_{calc} is the calculated response from all components from the reaction. r_{free} is signal in absence of fluorophore. M2-1 is not labelled so does not contribute any signal.

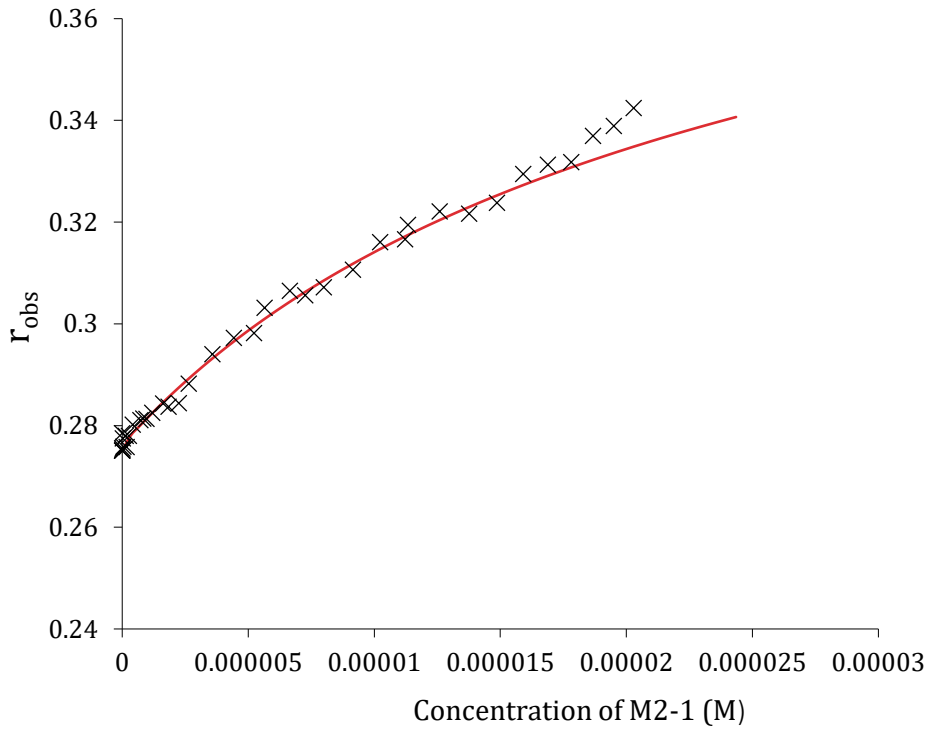


Figure 3.4.1.3 Binding curve of M and M2-1 showing the fit of the model to the data obtained. Concentration of M varies from 10×10^{-10} M to 2×10^{-5} M. Imported from a

The binding curve and fit is shown in Figure 3.4.1.3 and the concentrations of each component in the equilibrium binding were then calculated. It was noticed that the data above 10 μM did not fit with the above model, it was noticed at these higher concentrations that a small amount of precipitate began to form. The variance of the data (s^2) was found to be 5.32×10^{-5} and the standard deviation is 7.2×10^{-3} for calculated and experimental concentrations. A low value indicates that a close fit is observed using this thermodynamic pathway despite the variation at the higher concentrations. The final association constants are shown below in Table 3.4.1.4.

K_1	K_2	K_m
1.20×10^5 M^{-1}	2.88×10^4 M^{-1}	2.27×10^1 M

Table 3.4.1.4 Association and dissociation constant values from the oligomerisaton M fit.

The data as shown above indicate a 13 fold increase in the equilibrium constant in comparison to the previous SPR experiments shown in Table 3.3.3.3 and 3.3.3.7. Overall K_{spr} = 0.64 μM in comparison to $K_{1(fa)}$ = 8.3 μM. In K_2, where a second M2-1 binds to the matrix protein as shown in Figure 3.4.1.1, is weaker than the first addition of M2-1.

3.5. FUTURE WORK.

3.5.1 ISOTHERMAL TITRATION CALORIMETRY (ITC)

ITC directly measures the change in heat of the reaction between the two interactants and is considered one of the definitive techniques for understanding physical or chemical reactions. The Surface Plasmon Resonance experiments gave an K_d of the binding, but ITC would help characterise the interaction further. ITC is a highly sensitive technique and direct thermodynamic characterisation of the association of the proteins can be obtained.[115] The relationship between the binding sites, if there are indeed multiple sites, could be identified to understand the cooperativity of the reaction.

3.5.2 CRYO-ELECTRON TOMOGRAPHY (CRYO-ET)

Cryo-Electron Tomography uses the 2D image slices of a frozen sample to form a 3D image.[78] In this chapter the protein interactions were investigated using electron microscopy. This technique is limited by the resolution of the microscopy and the degradation of the sample. By using Cryo-Electron Tomography the limitations are removed as the structure of the sample is preserved from electron damage by freezing the sample in liquid nitrogen and placing in a vacuum chamber.[36] The technique preserves the biological structure in the vacuum, which would be ideal for a non-covalently linked complex such as these. Unlike Transmission Electron Microscopy there are no chemical fixing agents or stains used. This again preserves the structure and also prevents warping of the sample when exposed to the stain. This technique would be preferred to small angle X-ray scattering due to a lack of purification techniques for the complex. Cryo-EM could image the complex for shape, and thus a 3D structural shape of the complex in solution can be established.

3.5.3 BINDING REGION MUTATIONS

From previous papers[116,7] it is known that the N-terminal domain of the matrix protein interacts with the M2-1 protein. Binding region mutations could begin to pinpoint the exact residue/residues that are involved in the interaction. Firstly the gene would be recloned with small sections i.e. five amino acids each time from the N Terminal region removed. After optimisation and purification the truncated proteins would be tested using the techniques used in this chapter. The truncation that contains the interaction residues would show decreased or no binding in the experiments.

.

3.6 CONCLUSIONS

The interaction between M and M2-1 proteins has been indicated through work by Li *et al.*[58] using cytoplasmic inclusion bodies in hRSV infected cells. The proteins co-localise together during the experiment and on further investigation by cell free binding assays it was shown that M2-1 and M bind together. Further characterisation is needed to determine the scope of the interaction and how important this could be for future viral targets.

Initial pulldown assays were used to determine if the *in vitro* the interaction could be mapped and if indeed the two proteins do interact in the absence of the nucleocapsid protein. Once the interaction was confirmed, transmission electron microscopy was used to try and image the aggregate of the two proteins. This was successful and two major species were seen on the negative stain. The two species were unique sizes, which were ten-fold larger than the respective M2-1 tetramer and M dimer.

Further binding was observed and characterised by the use of surface plasmon resonance and fluorescence anisotropy. The former gave a K_d of 5 mM, where the data were fitted to a 1:1 Langmuir interaction model. This was a true fit for concentrations of M2-1 below 10 μM but above this, the 1:1 fitting deviated from the experimental data. This could be due to a second species being present above a certain concentration, which could be a rearrangement that mirrors that of the hMPV tetramer.[50]

To begin to understand the thermodynamic pathway of the reaction, fluorescence anisotropy experiments were conducted. In the oligomerisation model, the fitted curve was found to be a close fit up to 10 μM above this concentration the fit was poor. As matrix protein has been found to form up to a hexamer in solution so the dimerization form of the protein could not be relied upon. The variance between the curve and the observed polarisation indicated that this was the better statistical model for the data but further characterisation work is needed to determine if the thermodynamic pathway is a true representation.

Further analysis through Cryo-Electron Tomography could yield a more conclusive image of the proteins binding. In this technique protein slices are imaged which can be pieced together to form a 3D structure of the bound proteins.[78]

In future work, Mutations of the N-Terminal of M, which has been previously detailed as forming the binding region[117] could help to understand the binding and stoichiometry as well as employing Isothermal Titration Calorimetry to fully characterise the interaction.

3.7 METHODS

3.7.1 EXPRESSION AND PURIFICATION OF M PROTEIN.

A pET-16-b vector containing the ORF of M protein, obtained from R.P. Yeo, was transformed into *E. coli* BL21-DE3 and grown on ampicillin LB agar plates overnight at 37 °C. One colony was picked from the resulting plate and was incubated in 0.2 L of LB broth (Tryptone 10 g/L, sodium chloride 10 g/L, yeast extract 5 g/L; Melford Laboratories Ltd., Suffolk, UK, 25 g/L) overnight at 37 °C. 0.01 L of the starter culture was used to inoculate 1 L of LB Broth in a 2.5 L conical flask, the bacteria were grown in the presence of ampicillin (100 mg/mL) at 37 °C for 5 hours or until optical density at 600 nm reached 0.6-0.8. Expression was induced using isopropyl- β-D-1-thiogalactopyranoside (IPTG, 0.8 µM; Sigma Aldrich, UK) and left to shake overnight at 28 °C at 180 rpm. The bacterial cells were harvested by centrifugation at 4000 × g for 20 minutes in a Beckmann Coulter ™ J-20 XP centrifuge, using rotor JLA 8000 and the cell pellet was stored at -80 °C.

Protein extraction was achieved through resuspending the pelleted cells in 0.02 L of lysis buffer (50 mM Tris,1 M NaOH, 5 mM $CaCl_2$, 5 mM $MgCl_2$ pH 7.6). Lysozyme (1 mg/mL) and protease inhibitor tablet (cOmplete Mini, EDTA-Free, Roche Diagnostics Ltd, West Sussex, UK) were added once the pellet was resuspended. The cells were lysed by sonication for 240 seconds at 20 kHz and 3-[(3-cholamindopropyl)dimethylammonio]-1-propanesulphonate hydrate (CHAPS, final concentration 8 mM in ddH_2O) was added to the mixture and centrifuged at 50,000 × g for 30 minutes Beckmann Coulter ™ J-20 XP centrifuge, using rotor JA 25.50 to remove cellular debris.

The supernatant was then passed through a pre equilibrated nickel affinity His-trap column (GE Healthcare Life Sciences, Buckinghamshire, UK). After the supernatant had passed through the column, Tris buffer (50 mM Tris, 300 mM NaCl pH 7.6) was run through, followed by an increasing concentration of imidazole (50 mM Tris, 300 mM NaCl, 50 mM to 1 M imidazole pH 7.6). The protein eluted at 200 – 500 mM imidazole and the fractions were analysed for M protein by SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis)

The protein was buffer exchanged overnight into Tris buffer (50 mM Tris, 200 mM NaCl) or PBS (Gibco® PBS tablets, Life Technologies, Renfrew, Scotland) by dialysis (MWCO >

12,000 Da) at 4 °C. The His-tag was removed by cleavage with Factor Xa overnight, as per the manufacturers instructions (Novagen, Merck Chemicals Ltd., Nottingham, UK). The matrix protein was then stored and used at 4 °C.

3.7.2 SDS-PAGE (SODIUM DODECYL SULPHATE POLYACRYLAMIDE GEL ELECTROPHORESIS)

The purified protein was analysed using a 15% acrylamide resolving gel: 5 mL ProtoFlow 30% acrylamide gel (Flowgen Bioscience Ltd., Nottingham, UK), 2.5 mL buffer (1.5 M Tris, 0.4 % (w/v) SDS pH 8.8), 2.3 mL water, 0.1 mL 10 % SDS (Aldrich, Poole, UK), 0.1 mL 10 % (w/v) ammonium persulphate (APS) solution in water and 0.004 mL N,N,N',N'-tetramethylethylenediamine (TEMED; Fluka). The stacking gel was made from: 1.7 mL ProtoFlow 30 % acrylamide gel (Flowgen), 2.5 mL buffer (0.5 M Tris, 0.4 % (w/v) SDS pH 6.8), 5.55 mL water, 0.1 mL 10 % SDS (Aldrich,), 0.1 mL 10 % (w/v) APS solution in water and 0.01 mL TEMED (Fluka). Protein samples were prepared by addition of protein loading buffer (2.5 mL 1 M Tris-HCl pH 6.8, 0.5 mL ddH_20, 1.0g SDS, 0.8 mL 0.1% bromophenol blue, 4 mL 100% Glycerol, 2 mL 14.3 M β-mercaptoethanol (100% stock) and adjusted to 10 mL using ddH_20, 5 µL in each 15 µL sample) and heated to 95 °C for 5 minutes then loaded onto the gel with a protein ladder molecular weight marker in the first well (10 – 250 kDa; New England Biolabs., Hertfordshire, UK). The gel was run at 200 V for 1 hour and stained using InstantBlue™ (Expedeon, Cambridge, UK)

3.7.3 DETERMINING PROTEIN CONCENTRATION

3.7.3.1 NANODROP 1000

The concentration of the protein was determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Kent, UK) by measuring absorbance at 280 nm. The protein concentration was increased using a Vivaspin 6 (MWCO 10,000; GE Healthcare Life Sciences). Samples were centrifuged at 4000 × g in a Beckman Coulter™ Allegra™ X-22R centrifuge for 20 minutes at 4 °C, tested using the NanoDrop 1000 spectrophotometer and repeated until the desired concentration was achieved.

3.7.3.2 BRADFORD ASSAY

Protein concentration was also verified by Bradford assay. Standard concentrations between 5 and 100 µM of bovine serum albumin were prepared and allowed to equilibrate at room temperature for 30 minutes. Each sample contained the diluted

protein sample using MilliQ water to the appropriate concentration at 800 μL and Bradford Reagent (Sigma Aldrich, UK) 200 μL. Each concentration was measured at 595 nm using UV-visable spectrophotometer (Multskan Go, Thermo Scientific, UK) and the absorbance was plotted against the concentration to give a linear plot. The protein was diluted and prepared in the same way and measured at 595 nm. The resulting absorbance was plotted on the standard graph and the concentration was found on the x-axis.

3.7.4 PULL DOWN ASSAY

M2-1 protein was prepared according to chapter 2.8.1 until crude protein was obtained. 50 μL of protein (~30 μM) was decanted into microcentrifuge tubes and kept on ice for use later.

Glutathione sepharose 4B bead slurry (G.E Healthcare, Stevenage, uk) was decanted into a microcentrifuge tube (Starlabs, Milton Keynes, UK) and sendimented by centrifugation at 500 × gfor 5 minutes. The supernatant was decanted and 1 mL of 1 × PBS was added to resuspend the beads to form a 50% slurry.

20 μL of the 50% slurry of glutathione sepharose 4B to each centrifuge tube containing protein and gently mix for 1 hour at 4 °C. The slurry and protein centrifuged and the supernatant discarded. 100 μL of 1 × PBS was added, the tube was vortexed briefly and centrifuged for 10 seconds to sediment the sepharose beads. The supernatant was discarded and the wash was repeated 3 times.

To elute the protein from the beads, 10 μL of glutathione elution buffer (0.154 g reduced glutathione in 50 mL of 10 mM glutathione pH 8.0) was added to the protein slurry mixture. The mixture was vortexed to suspend the sepharose beads and the elution buffer was incubated for 5 minutes at RT. The slurry was centrifuged for 5 minutes to sediment the sepharose beads and the supernatants to fresh labelled tubes.

To identify the proteins in the mixture, 10 μL aliquots of each sample run on a SDS page gel, as described in chapter 3.7.2.

3.7.5 TRANSMISSION ELECTRON MICROSCOPY

All images obtained via transmission electron microscopy using a modified negative staining protocol were with assistance of A. C. Richardson. A JEOL 2100F FEG TEM with a GATAN GIF tridiem with 4 megapixel Ultrascan™ 1000 CCD camera was used for imaging. 20 µL of a protein solution (either single or mixture) was pipetted to formvar-carbon coated 400-mesh copper grid, and left to dry for 30 s at room temperature. Any excess was removed with standard filter paper. The grid was negatively stained with 1% uranyl acetate aqueous pH 4.0 for 30 s. The excess was removed from the grid, again with filter paper and allowed to dry briefly before imaging.

A standard TEM protocol for focussing and using the microscopy was followed and carried out by A. C. Richardson.

3.7.6 SURFACE PLASMON RESONANCE

All experiments were conducted using G.E Healthcare Buffers and chips. A Biacore 3000 machine was supplied by G.E Healthcare, UK.

3.7.6.1 MATRIX PROTEIN CONCENTRATION SCOUT.

The NTA Chip was equilibrated using 1 × HBS-EP buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20) in the machine for 20 minutes before use. Activation was achieved by injection of $NiCl_2$ (0.02M, G.E Healthcare) for 60 s at 10 µL/ min The standard NTA Biacore protocol (G.E Healthcare, UK) was then followed which detailed a 10 fold concentration range injected for 60 s at 10 µL/ min. The chip was regenerated using EDTA (0.01M, G.E Healthcare, UK) injected for 60 s at 30 µL/ min.

3.7.6.2 BLANK EXPERIMENT.

Lane 1 on the NTA chip was used for blank experiments, no other experiments were conducted on this lane. Activation was achieved by injection of $NiCl_2$ (0.02M, G.E Healthcare) for 60 s at 10 µL/ min, the non-tagged protein was then injected at 60 s at 30 µL/ min at desired concentrations. The chip was regenerated using EDTA (0.01M, G.E Healthcare, UK) injected for 60 s at 30 µL/ min.

3.7.6.3 SECOND PROTEIN ADDITION.

Activation was achieved by injection of $NiCl_2$ (0.02M, G.E Healthcare) for 60 s at 10 µL/ min The standard NTA Biacore protocol (G.E Healthcare, UK) was then followed which detailed the injection of $M_{(His6)}$ protein at the desired concentration for 60 s at 10 µL/ min. The second protein (no tag) was injected for 60 s at 30 µL/ min at the desired concentration. The chip was regenerated using EDTA (0.01M, G.E Healthcare, UK) injected for 60 s at 30 µL/ min.

This was repeated with 5 different non-tagged protein concentrations, which were pre determined using the guidelines detailed in G. E. Healthcare NTA Biacore protocol.

3.7.6.4 BIAEVALUATION ANALYSIS.

A Standard protocol from G.E Healthcare, UK was followed to analyse each data set using the provided software. Each concentration was normalized and blank subtracted. The final curves were exported in an ASCII file and plotted in Microsoft Excel.

3.7.7 FLUORESCENCE ANISOTROPY

The concentration of stocks containing RR-M peptide were predetermined by the author and stored at -80 °C. RR-M peptide was added to a standard quartz cuvette and diluted to 500 µL with ddH_20. All measurements were taken on a Cary Eclipse Fluorescence Spectrophotometer (Aligent Technologies, UK) using the supplied Eclipse ADL program. The excitation and emission wavelength (570 nm/580 nm) was added to the program and the voltage was changed to reflect the sample. The voltage was changed until the reading voltage stood at approximately 700 m.A.U. The blank experiment, of only the fluorophore in buffer was taken and the G Factor noted.

Predetermined stocks of M2-1 protein were titrated in to a quartz cuvette containing the fluorophore and measured. Each measurement was plotted in Microsoft Excel and the analysis conducted in the same program.

4. INTERACTIONS OF THE MATRIX PROTEIN WITH CYTOPLASMIC TAIL OF THE FUSION PROTEIN.

As described in Chapter 1.7, hRSV F is key to the attachment of the virus to the host cell.[11,37] The protein is a trimer, where the F2 proteins are linked together and disulphide bonds link these to the transmembrane domain F1.[43] The transmembrane domain, or cytoplasmic tail of F interacts with the matrix protein and this interaction is thought to be key for maturation and filament formation.[7] Previously, work has been conducted which shows that the cytoplasmic tail of the fusion protein is key to interaction between F and the matrix protein.[7,118] This chapter expands on the published data[7] and previous unpublished work from VAM/JMS/RPY, Durham University to try and understand the nature of the interaction. All work reported in this chapter was carried out alongside the undergraduate students: O. Neale, E. Antonio and L. Argent and in collaboration with Dr. J. M. Sanderson, Durham University.

4.1. SEQUENCE AND PURIFICATION.

4.1.1 MATRIX PROTEIN.

The matrix protein was initially cloned by Dr. R. P. Yeo at Durham University. Subsequent batches of the protein were purified by the author and undergraduate students E. Antonio, O. Neale and L. Argent.

4.1.2. F PEPTIDE.

This six amino acid long peptide comprises the final C-terminal amino acids of the F protein with the sequence NIAFSN shown Figure 4.1.2.1. This unstructured region, is known to bind to the matrix protein, but the details of the interaction between the two proteins have not been fully elucidated. Phenylalanine 572 has been demonstrated to be critical for mature virion formation *in vivo*.[7] The peptides used in this work were

purchased on solid support from Almac Sciences and cleaved, labelled and purified by E. Antonio and L. Argent.

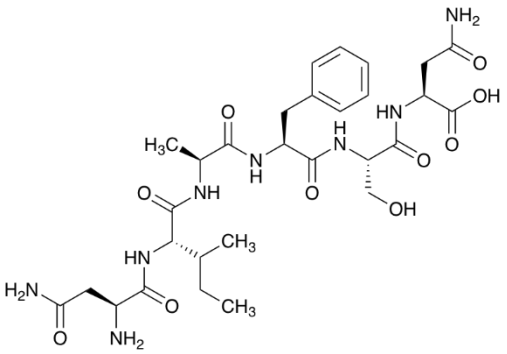


Figure 4.1.2.1 F peptide amino acid structure (NIAFSN)

4.1.3. MUTANT F PEPTIDE.

To test the interaction and confirm that it does indeed hinge on the phenylalanine, a mutant peptide was designed to eliminate this interaction with matrix protein. The mutated peptide contains an asparagine (N) in place of the phenylalanine, this was chosen due to solubility and is shown in Figure 4.1.3.1. Previous work had shown that replacement of the phenylalanine with alanine gives rise to solubility difficulties, which were incompatible with the required experimental conditions.

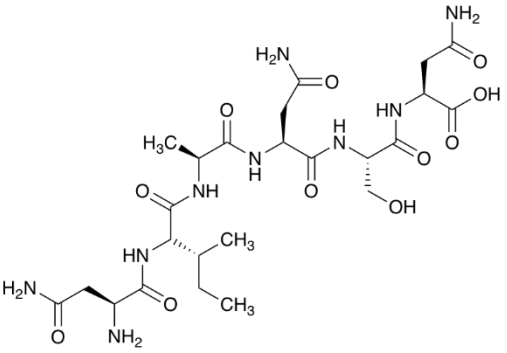


Figure 4.1.3.1 Mutant F peptide structure (Fmut) (NIANSN)

4.2 THERMAL SHIFT ASSAY (DIFFERENTIAL SCANNING FLUORIMETRY, DSF)

This technique was used to determine the stability of the protein and if there was an interaction with the peptide. The raw data is shown below and the NAMI[95] program was used to visualise the curves in Figure 4.2.1

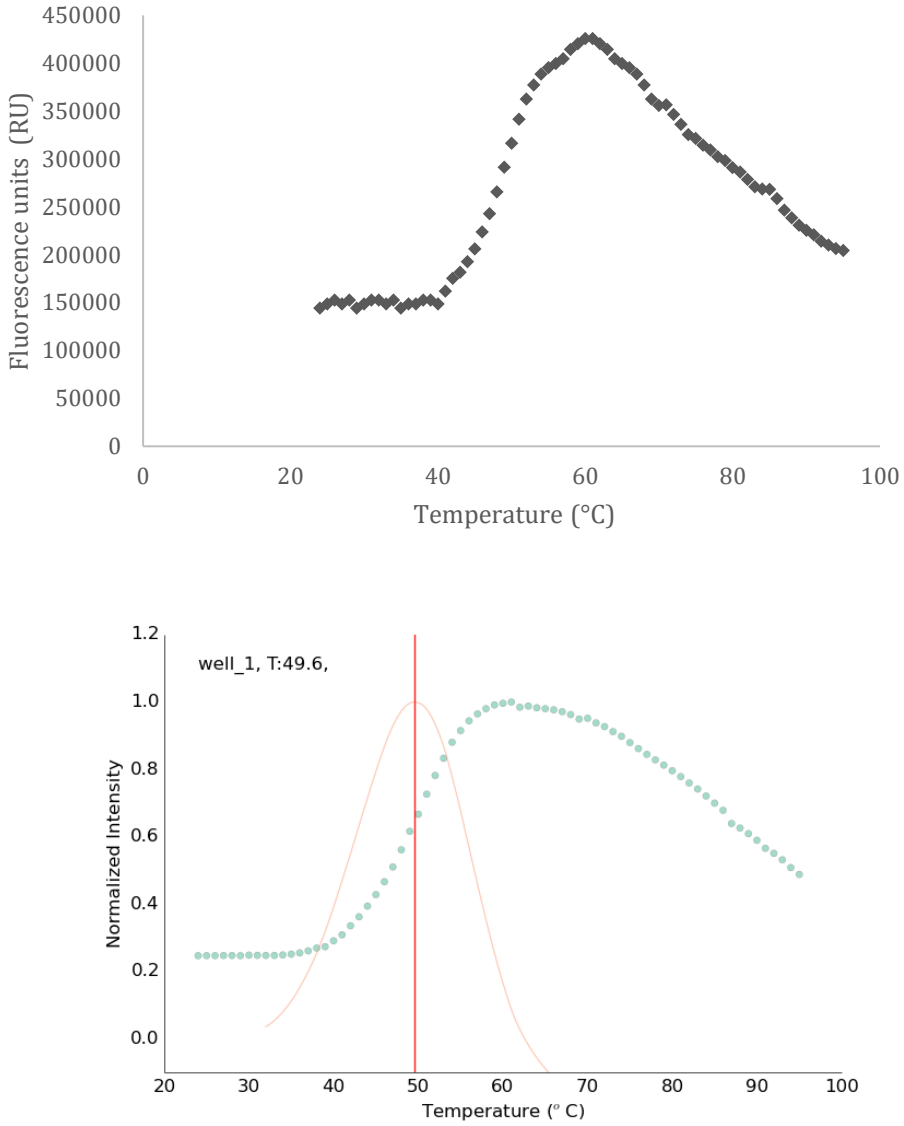


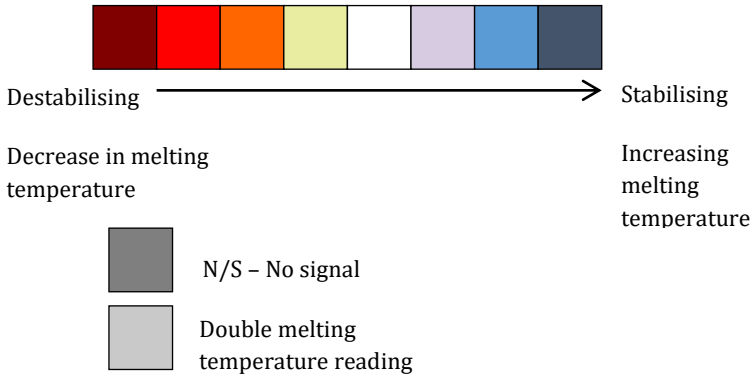
Figure 4.2.1. Raw data curve showing the melting temperature for M protein in water to be 49.6 °C shown in NAMI software.[95]

Firstly a salt screen was analysed against matrix protein itself. The aim was to stabilise the protein and to find buffer conditions to help co-crystallisation. For co-crystallisation

to occur the proteins must be in a meta-stable state, this is greatly affected by the buffering system in the crystallisation screen.[17,103] Many conditions that are included in this screen are present in commercial crystallisation screens. 96 conditions and concentrations were tested and the full list is detailed in the Appendix A. The full screen melting temperatures (T_m) results are shown in Figure 4.2.2. Many conditions stabilised the protein, so only those compounds which gave an increase in T_m of over 3 °C are shown in Table 4.2.3.

	1	2	3	4	5	6	7	8	9	10	11	12
A	49.6	49.5	49.6	DEN	31.0	44.5	58.6	56.5	55.2	53.8	50.7	46.8
B	50.1	47.4	47.5	46.9	47.1	47.6	49.4	49.1	46.6	46.0	45.9	46.6
C	55.7	54.8	53.1	50.6	48.3	46.4	49.7	30.0	29.4	30.0	36.7	42.5
D	51.1	50.1	48.8	48.0	48.2	54.0	51.9	51.3	48.4	48.2	51.0	48.5
E	44.5	46.5	51.1	49.4	47.9	47.4	44.8	47.4	45.6	47.6	33.7	47.5
F	47.4	47.4	50.0	46.0	41.1	45.8	48.5	48.5	49.0	48.4	45.8	49.0
G	41.9	49.2	59.0	53.7	40.0	46.7	48.4	41.2	46.7	53.3	37.1	36.2
H	45.2	38.1	46.5	47.2	49.2	50.3	40.3	52.2	50.8	51.9	51.0	50.0

Figure 4.2.2. NAMI representation of salt screen detailed in Appendix A.



Compound	Melting Temperature (°C)	Stabilising Temperature shift (°C)
2 mM $La(NO_3)_3$	59.0	9.40
1.0 M Gu-HCl	58.6	9.00
0.8 M Gu-HCl	56.5	6.90
1.5 M NaCl	55.7	6.10
0.6 M Gu-HCl	55.2	5.60
1.0 M NaCl	54.8	5.20
1.0 M $NaSO_4$	54.0	4.40
0.4 M Gu-HCl	53.8	4.20
2 mM $PrCl_3$	53.7	4.10
2 mM $HoCl_3$	53.3	3.70

Table 4.2.3. Matrix protein stabilising ligands over a 3 °C threshold,

4.3. SURFACE PLASMON RESONANCE

As previously described in Chapter 3, Surface Plasmon Resonance can be used to understand binding across a number of concentration levels. $M_{(His6)}$ protein was again bound to an NTA chip. Through the previous concentration scouting in Chapter 3.3.1, experimental parameters were already known.

4.3.1 $M_{(His6)}$ AND FUSION PEPTIDE.

Initially, the matrix protein was attached to the chip at 3 μM concentration in solution, as in previous experiments, this gave a satisfactory response and the fusion peptide was then flowed over the chip in various concentrations. The corrected results are shown below in Figure 4.3.1.1

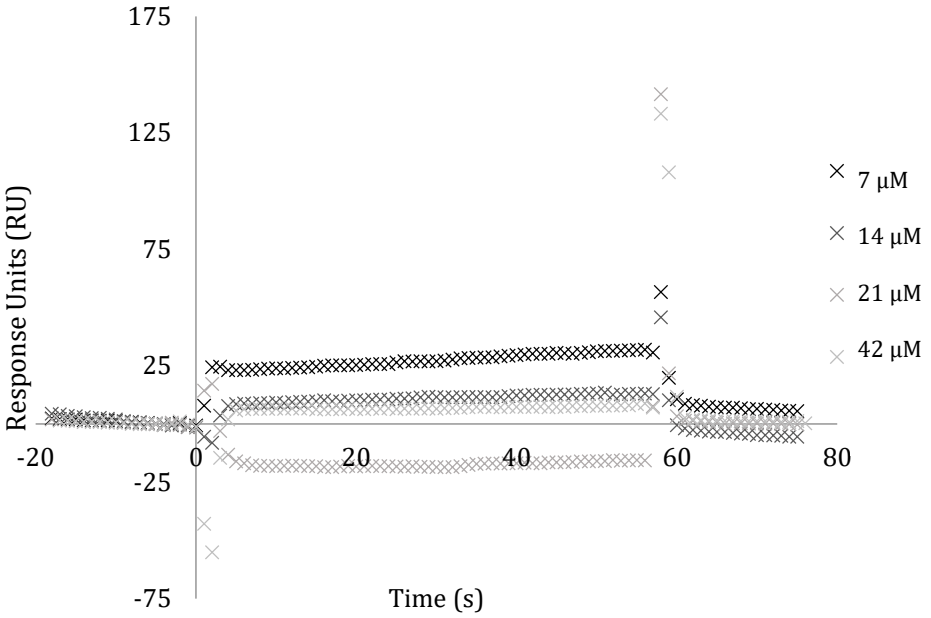


Figure 4.3.1.1. Sensogram curves of F peptide (NIAFSN) response on $M_{(His6)}$.

In comparison to the previous experiment in Chapter 3, the response in binding shows a very small amount of binding and the K_d could not be accurately determined from the curves using the BIAevaluation software[107] as the dissociation curve did not fit the models available. However, the association curves fit with the binding model and gave a preliminary k_a was recorded to be 10^3. It is known that the matrix protein in solution is a dimer[49] and this graph raises questions as to the experimental problems of using this technique where one protein, of unknown oligomerisation is attached to a surface. This could be further investigated using a technique that uses both proteins in solution i.e. Fluorescence Anisotropy.

4.3.2. $M_{(His6)}$, MATRIX PROTEIN AND FUSION PEPTIDE

The 6× Histidine (His_6) tag was cleaved with Factor Xa to yield matrix protein with no tag and this was flowed across the $M_{(His6)}$ protein attached on the NTA chip at 5 µM concentration. The same concentrations of fusion peptide were flowed over the chip and the resulting blank subtracted graph is shown below in Figure 4.3.2.1.

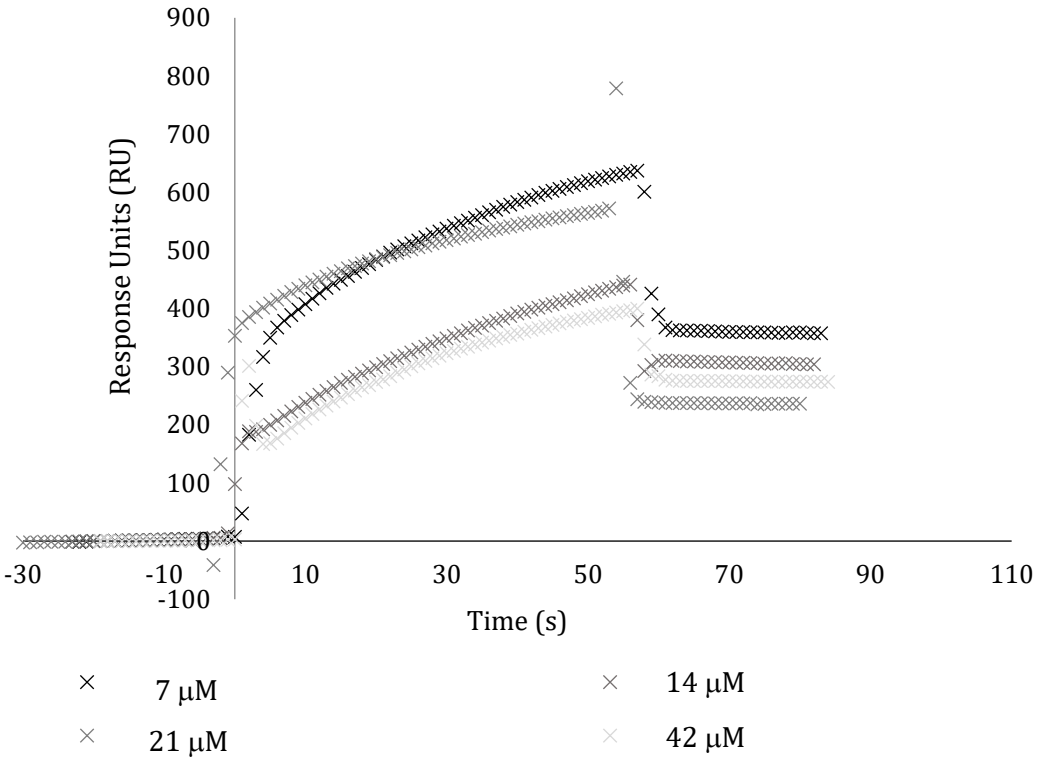


Figure 4.3.2.1 Binding response of F peptide to M protein.

The difference in the two graphs is notable, clear binding is shown in Figure 4.3.2.1 and thermodynamic analysis could be carried out using the BIAevaluation software.[110] The K_d obtained by a local fit was approximately 400 μM. The difference between the first and second graphs would suggest that the binding is integral to matrix protein being a dimer or a higher order oligomer. Further work using fluorescence studies could confirm the interaction and the stoichiometry, and give a degree of understanding of the complexities of the thermodynamic pathway.

4.4. FLUORESCENCE STUDIES.

4.4.1 FÖRSTER RESONANCE ENERGY TRANSFER (FRET).

Förster resonance energy transfer (FRET) is a distance-dependent interaction between the electronic excited states of two dye molecules in which excitation is transferred from a donor molecule to an acceptor molecule without emission of a photon.[119,112] The technique has been used to investigate an assortment of biologically important binding interactions and will provide information on the thermodynamics and stoichiometry of the binding of M and F proteins. The protein and peptide can easily be fluorescently tagged and binding should bring the two together in close enough proximity for the technique to give a signal.[120,121] Another consideration is that the absorption spectrum of the acceptor must overlap with the fluorescence emission spectrum of the donor so the choice of fluorophores is key. This feature allows the transfer of energy from the donor chromophore to the acceptor chromophore through a non-radiative dipole-dipole interaction.[120] Fluoroscein was chosen for the donor molecule and Rhodamine as the acceptor.

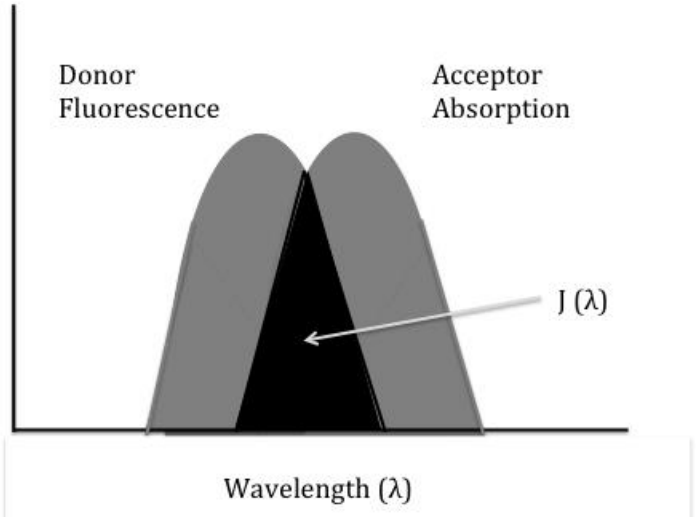


Figure 4.4.1.1. Overlap of the excitation and donor curves.[122]

Another important consideration in the experiment is the Förster Radius (R_0), this is the distance where the energy transfer is 50% efficient (where 50% of the excited donors

are deactivated by FRET). The size of R_0 is dependent on the properties of the different dyes in use in the experiment, but it is typically in the order of 50 Angstroms.[122] The distance is important in this technique as the interaction has to bring together two fluorophores within the Förster radius for energy transfer to occur.[122]

4.4.1.1 FUSION PEPTIDE (NIAFSN)

Initial investigations and experiments using this peptide were conducted by E. Antonio.[123] The protein M was tagged with a donor chromophore, in this case fluorescein was chosen and referred to as Fl-M. F peptide was tagged with an acceptor, Rhodamine Red now referred to as RR-F hereafter. Both of the fluorescent tags are known to be a FRET pair, with a Förster radius of 55 Å.[124] Fluorescein behaves as a donor (emitting at 525 nm) and Rhodamine Red behaves as an acceptor (emitting at 585 nm). A preliminary titration was carried out by E. Antonio to confirm that FRET was a suitable technique to investigate the interaction between Fl-M and RR-F, the resulting graph is shown below in Figure 4.4.1.1.1

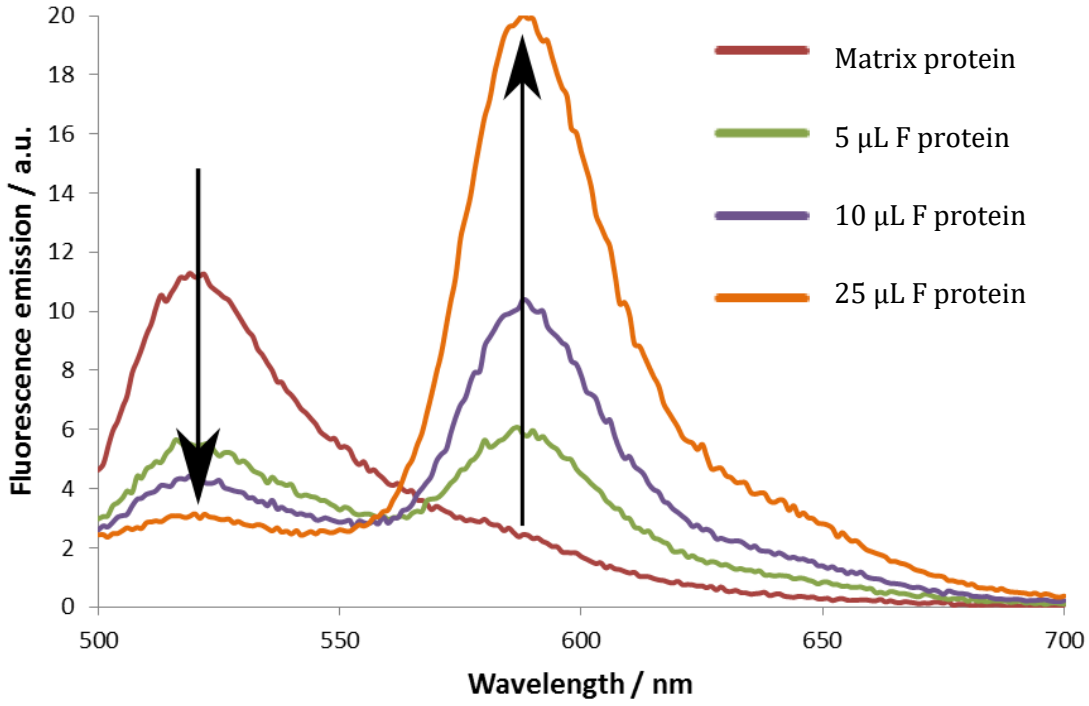


Figure 4.4.1.1.1 Emission spectra by E. Antonio for the addition of RR-F peptide to Fl-M protein[123]

A decrease in the energy seen in the donor (525 nm) and the increase shown in the acceptor at (585 nm), shows a clear FRET interaction. There were problems with this initial experiment due to a contribution being seen from the Fl-M in the emission wavelength, where the RR-F peptide shows a small signal in the absorption spectra where ideally there would be no signal of RR-F in the donor wavelength or Fl-M in the emission. The second problem is the large change in fluorescence emitted from the donor at 525 nm, it would be expected to have decreased linearly, with each addition of RR-F and could be due to a more complex system i.e. self-quenching.

E. Antonio completed a series of control titrations to determine which factors were involved in the non-linear change in Fl-M on the additions of RR-F.[123] One cause of the non-linear relationship could be self-quenching, where the fluorophores non-radiatively transfer energy to other identical fluorophores. Both of the fluorophores in the experiment are known to self-quench and this could be problematic due to oligomerisation and conformational changes.

The RR-F is a short peptide and should not have any concentration dependent quaternary structure and this was tested by a concentration dependant fluorescence graph by E. Antonio. The resulting graph should be a proportional linear relationship where the signal is proportional for each addition of peptide. This was confirmed and the RR-F peptide appears to be a monomer and not the source of the decreased signal by self-quenching.

E. Antonio then tested the Fl-M in the same manner to confirm that the matrix protein oligomer is responsible for the decrease in signal after the addition of RR-F.[123] Matrix protein could be multiply labelled by fluorescein on lysine residues. There are multiple residues on each side of the linker region where self-quenching could occur due to the flexibility of this region. The same concentration dependant fluorescence graph showed a non-linear trend and confirmed that at high concentrations in the assay (> 5 μM) Fl-M shows changes in its quaternary structure. Under 4 μM it was found that the graph remained linear.

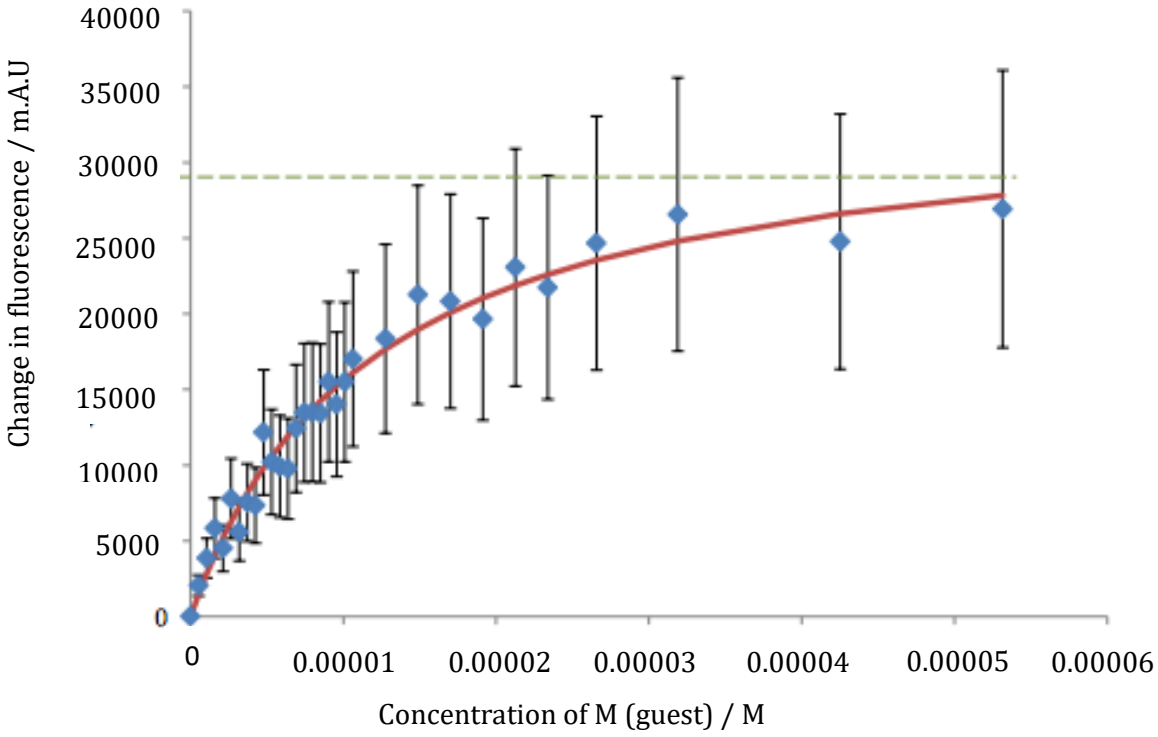


Figure 4.4.1.1.2. Binding curve attained from change in fluorescence upon addition of Fl-M and RR-F, Where RR-F was host and Fl-M was guest by E. Antonio.[123] Errors in the data are calculated to be approximately 20% by E. Antonio.

E. Antonio and J. M Sanderson fitted the data to a rectangular hyperbola fit model.[123] Rectangular hyperbola fit models are frequently used in biological systems to model processes,[125] they are typically of the form

$$V = (V_{max} S)/(K_m + S)$$

Where V is the biological rate, S is the concentration of the substrate in the process, the fitting constants V_{max} and K_m are the maximal rate at substrate concentration and subtsrate concentration at half the maximal rate.[125]

The concentration of the host was known and the other concentrations were calculated as a fraction of the original concentration which is given by the ratio of the fluorescence[119]. The concentration of bound guest ([B]) was determined through a ratio of change in fluorescence at a given concentration against the change in fluorescence associated with fully saturated binding. The concentration of free guest ($[G_F]$) can be

determined via a ratio of fluorescence of the free guest ($F_{observed}$-$F_{backgroundhost}$-F_{bound}) against the fluorescence of the bound guest ($F_{backgroundguest}$-F_{bound}).[123] From the upper asymptote, shown as the green line on Figure 4.4.1.1.2 the maximum change in fluorescence is observed when the guest is maximally bound .[123]

The binding constant (K_a) is given as the equation below, where K_{diss} is the reciprocal of the binding constant. [122]

$$K_a = \frac{[B]}{[G_F]([H_T]-[B]} = \frac{1}{K_{diss}}$$

E. Antonio and J.M Sanderson proposed a binding model to fit the data as shown below in Figure 4.4.1.1.3. Note there is no MF + M intermediate state, this is due to matrix protein favouring the dimer in solution.

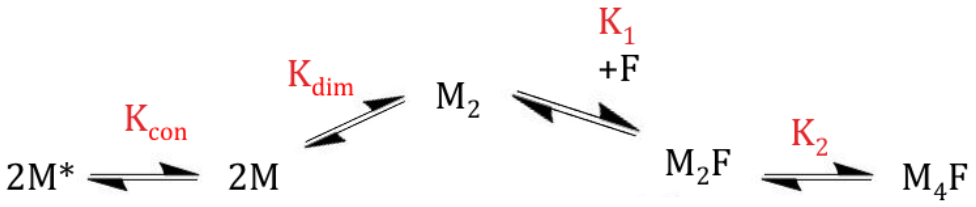


Figure 4.4.1.1.3 Reaction scheme of Fl-M and RR-F.[123]

Further experiments and analysis where the concentration of M was restricted, were undertaken using Klotz, Hill and Scratchard plots by E. Antonio.

The Klotz plot shown in Figure 4.4.1.1.4, where the binding data is plotted as concentration of bound guest ([B]) against concentration of log free guest ($[G_F]$). A generalised logistic function curve was expected where the binding constant is determined by the negative reciprocal of the dissociation constant. The curves were found to show positive cooperativity, where binding in one site increases the affinity of the other binding.

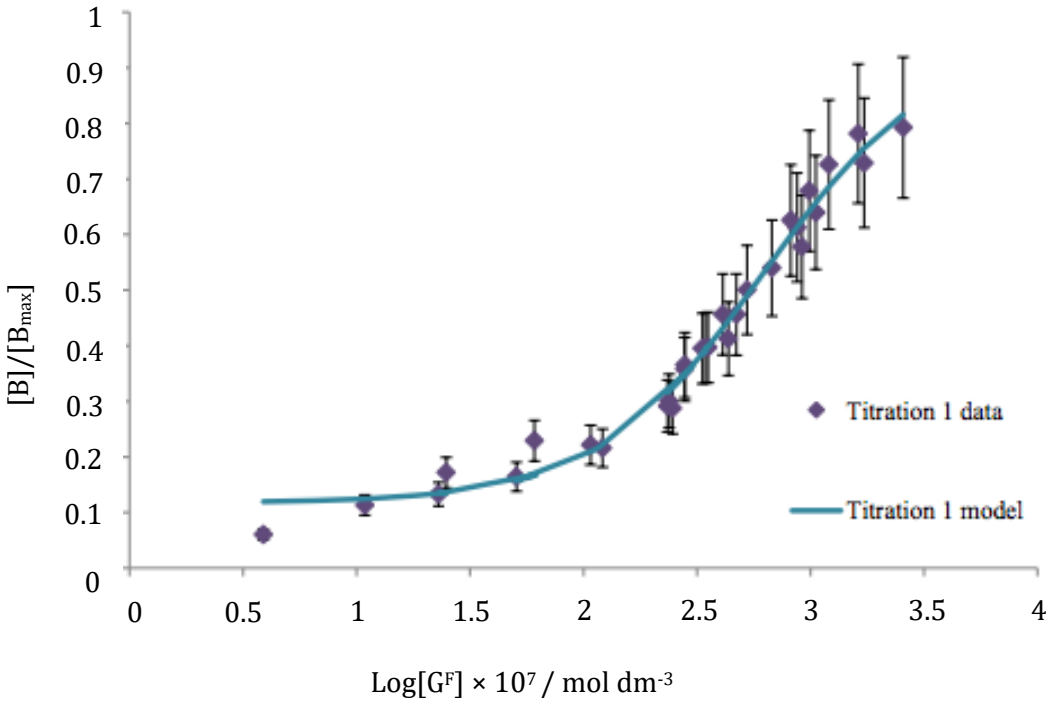


Figure 4.4.1.1.4 Titration curve modelled by a generalised logistics function shown as the blue curve by E. Antonio. Errors in the data were approximated to be 16% by E. Antonio. [123]

Insufficient data were obtained to complete the curve at the higher concentrations of free guest ([G_F]) so the binding constants obtained were only an approximation. The binding constants were approximately determined from the fitted logistical function curve and found to be 16 μM.

Scatchard plots calculate the affinity constant of a ligand by plotting the bound/free ligand against the bound ligand. Rearrangement of the equation shown on page 104 gives a linear relationship and a binding constant shown below.

$$\frac{B}{G_F} = \frac{-B}{K_{diss}} + \frac{H_T}{K_{diss}}$$

The binding constant was determined from the reciprocal of the gradient that is given by ($-1/K_{diss}$) and a linear gradient is expected [125] as shown in Figure 4.4.1.1.5.

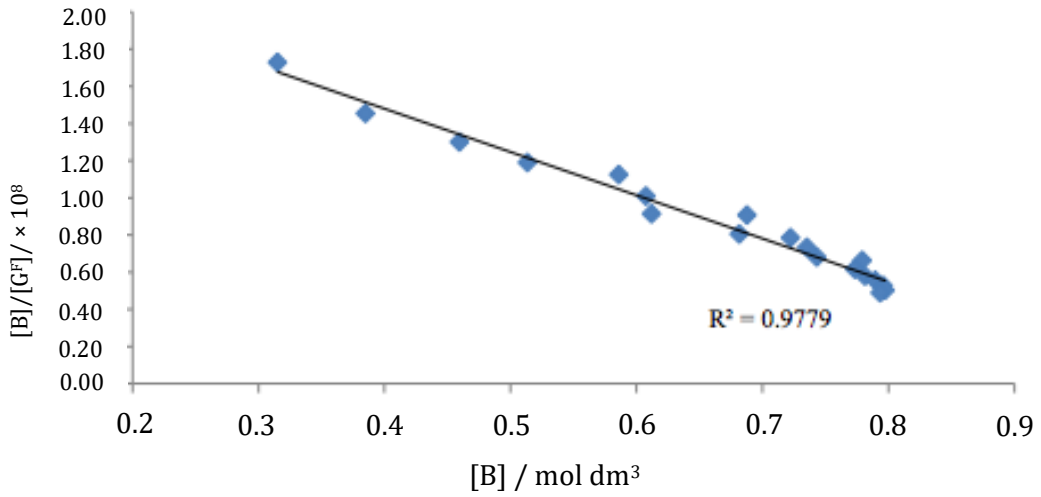


Figure 4.4.1.1.5. Scatchard plot for the addition of F peptide to M protein by E. Antonio.[123]

A linear Scratchard plot was found and the lower stoichiometry binding was found to be non-cooperative. The data were then analysed using a Hill plot to further investigate if the binding in the reaction is cooperative. The model does not assume that all sites are independent or equivalent, but that there is fractional occupancy (*f*) of the sites on the host as shown below

$$f = \frac{G_F^n}{K_{diss} - G_F^n}$$

Where (*f*) is fractional occupance, G_F^n is fractional occupancy of the guest, K_{diss} is the dissociation constant.[126]

Where n=1 is indicative of a class of single binding sites, n= integer >1 would suggest multiple binding sites and non integer values n<1 and n>1 suggest negative and positive cooperativity. The value of n is shown by the gradient of the Hill plot from the equation below

$$\log\left[\frac{B}{H_{T-B}}\right] = n[\log(G_F)] - \log(K_d)$$

Where B is bound species, G_F is Guest fractional occupancy.[126]

The Hill plot obtained, shown in Figure 4.4.1.1.6, demonstrates that the binding of M protein to the F peptide is indicative of a complex system. The result is similar to that of the Klotz plot where there are insufficient data points at higher concentrations to fully understand binding.

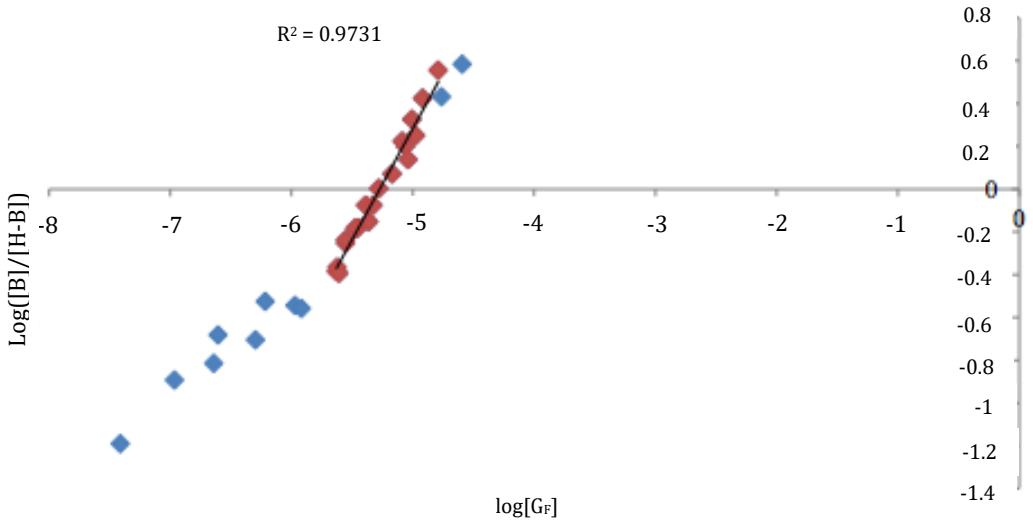


Figure 4.4.1.1.6 Hill plot for M protein binding to F peptide. The blues series shows data obtained, the red series where the guest was between 20% and 80% bound by E. Antonio.[123]

Higher concentrations of Fl-M should be added to further investigate binding, but due to self-quenching of the fluorophores, this could be a challenging issue. The binding constant was determined from the red region of data where a consistent gradient was obtained and found to be approximately 1. This confirms the hypothesis and indicates non-cooperative binding and the binding constant is detailed in Table 4.4.1.1.7.

Analytical model		**Further analysis**			
***K*₁**	***K*₂**	**Hyperbola**	**Klotz**	**Scatchard**	**Hill**
1.8 nM	28 nM	35 nM	24 nM	28 nM	39 nM

Table 4.4.1.1.7 Binding constants of the interaction of RR-F and Fl-M.

4.4.1.2. MUTANT FUSION PEPTIDE (NIANSN)

Further FRET experiments were conducted by Dr V.A Money to determine if binding would be diminished without the phenylalanine residue (F572) present in the peptide. The same experimental procedure was followed as detailed by E. Antonio.[123] As it can be seen in comparison to Figure 4.4.1.1.2 there is no clear binding when the F572 is replaced. Self-quenching is again another consideration for the higher concentrations especially. The relative fluorescence does not follow a set pattern and cannot be analysed in the same way as the wild-type peptide, which confirms the hypothesis that the F572 residue is key to the interaction.

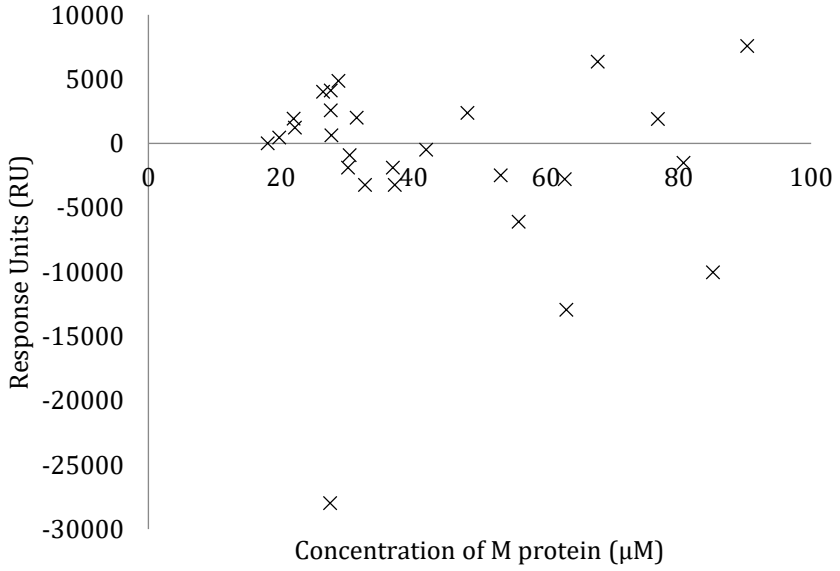


Figure 4.4.1.2.1 Binding curve obtained from RR-Fmut and Fl-M.

4.4.2 FLUORESCENCE ANISOTROPY (FA)

This technique was used to obtain thermodynamic values for the various interactions to compare to the K_d obtained through the Surface Plasmon Resonance experiments. As explained in Chapter 3, the degree of polarisation is low when the fluorescent peptide is changing orientation quickly. When binding occurs the fluorescent peptide changes orientation more slowly and the polarisation signal increases.

4.4.2.1 FUSION PEPTIDE (NIAFSN)

In this experiment the wild type peptide (NIAFSN) was tagged again with Rhodamine Red to form RR-F that was placed in the cuvette and non-tagged matrix protein was then titrated in. Various scouting experiments were conducted to understand the concentration range needed, previously the initial FRET experiments described above in 4.4.1 had indicated a nanomolar K_d and this was used as a starting point. The resulting curve is shown below in Figure 4.4.2.1.1.

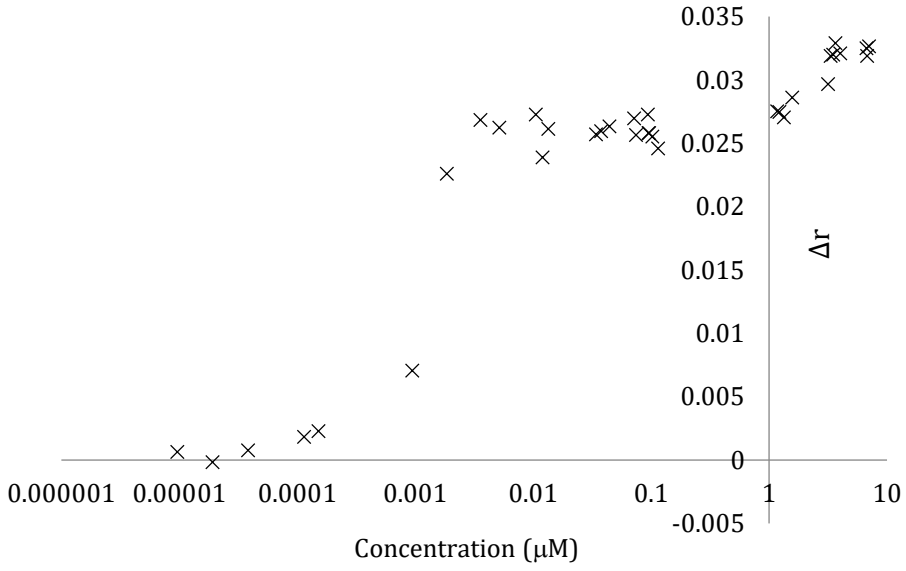


Figure 4.4.2.1.1 Fluorescence anisotropy curve showing the interaction of M protein and F peptide.

Analysis by J. M Sanderson built on the initial results from the initial FRET experiments, the reaction scheme shown again below in Figure 4.4.2.1.2.

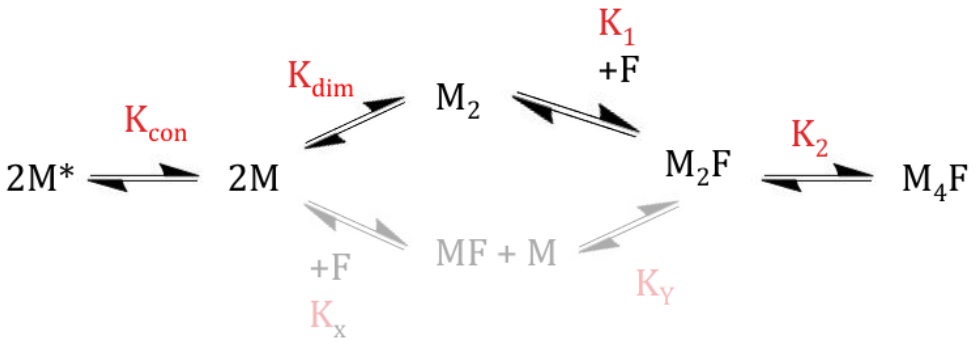


Figure 4.4.2.1.2 Reaction Scheme for FA analysis, where it presumed that the top pathway is preferred.

4.4.2.1.1 FITTING WITH M*

The M* term is used when the oligomerisation state of RSV-M protein is not defined. Previously, H. McPhee, Durham University showed that the matrix protein is a dimer, tetramer and a hexamer in solution. The 2M* term was used to fit the data using the binding isotherm as shown below and the fit of the data in the graph in Figure 4.4.2.1.1.1.

$$r_{calc} = r_{M^*}[M^*] + r_F[F] + r_1[M_2F] + r_2[M_4F]$$

The binding isotherm shows the contribution from each of the species in solution. r_{calc} is the overall signal that is seen from each species. In Figure 4.4.2.1.1.1 the term Δr is used to normalize the data.

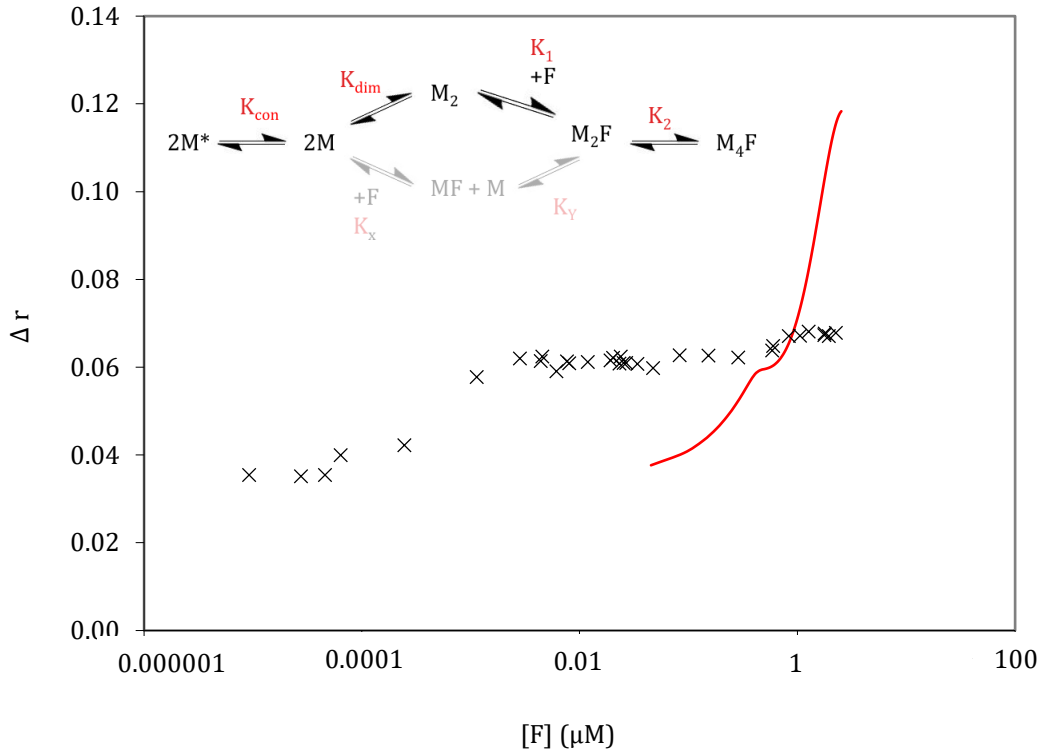


Figure 4.4.2.1.1.1 Using M* term in the binding isotherm for data fitting. [F] is the concentration of guest, in this case fusion peptide.

From the data analysis it can be seen that there is a concentration problem, where the line of best fit does not match the data. The maximum amount for a 1:1 complex would be 3 nM for 100% of F being bound, therefore 99% of the F peptide would not be bound and the signal change for the interaction would be minimal. Is there another interaction which is being seen, such as a rearrangement or a change in structure of F peptide or M? Since F peptide is only six amino acids in length, it lacks tertiary structure and could change conformation in solution which could change the signal being given in the experiment. Non-specific binding could not be ruled out either in this instance and further work would be needed to understand the interaction. The thermodynamic data from the analysis are shown below in Table 4.4.2.1.1.2.

K_2	K_1	K_{dim}	K_{con}
1.34×10^{7} M^{-1}	6.50×10^{11} M^{-1}	1.20×10^{4} M^{-1}	6.00×10^{13} M^{-1}

Table 4.4.2.1.1.2 Thermodynamic values using M* term.

4.4.2.1.2 FITTING WITHOUT THE M* TERM.

The binding isotherm, which details the total response for all species present is shown below,

$$r_{calc} = r_F \frac{[F]}{[F_{tot}]} + r_1 \frac{[M_2F]}{[F_{tot}]} + r_2 \frac{[M_4F]}{[F_{tot}]} + r_{free}$$

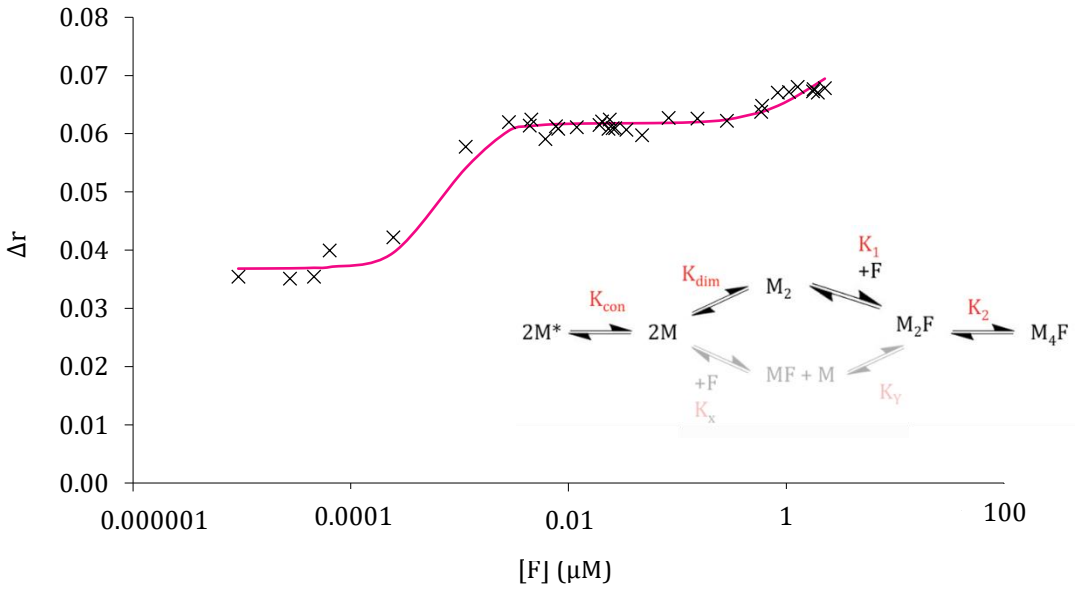


Figure 4.4.2.1.2.1 Binding curve and fit of RR-F and matrix protein interaction.

K_{dim}	K_1	K_2
4.14x10^5 M	7.43x10^{12} M	1.83x10^6 M

Table 4.4.2.1.2.2. Thermodynamic values for the dissociation of M protein to F protein.

The reaction scheme shown below, shows the two reaction mechanisms that are proposed to take place.

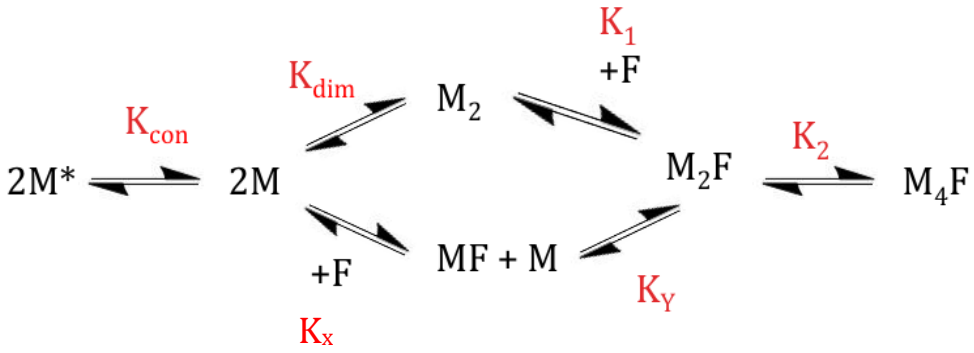


Figure 4.4.2.1.2.3. Proposed reaction equilibrium showing the two intermediate pathways proposed.

ΔG is the same irrespective of route so the formation of M_2F, from MF and M is equivalent to M_2 + F as shown below

$$K_{dim} \times K_1 = K_x \times K_y$$

The value for K_x as determined by Surface Plasmon Resonance was low (~400 μM) as shown in Chapter 4.5, in which case the value of K_y would have to be very large (~10^{16}) to align with the $K_{dim} - K_1$ pathway. The K_y value is unfeasibly large and the observed concentrations of MF are approximately zero indicating that the $K_x + K_y$ route can be discounted. The value for K_1 obtained through FA is unusually large and is up to 10^5 magnitude difference in comparison to the figures obtained in the FRET experiments, this raises concerns for the feasibility of the models used for the fitting of the data in this instance. The complex nature of the interactions could be the reason for errors in the

models. Further fluorescence experiments would need to be conducted to understand the nature of the interaction.

4.4.2.2 MUTATED FUSION PEPTIDE (NIANSN)

The same FA experiment was conducted with the Rhodamine Red mutant peptide (RR-Fmut) to understand if the phenylalanine, as described by Shaikh *et al.* [7] was key amino acid for binding to take place. The interaction was shown to be disrupted when the RR-Fmut was tested in the FRET experiments shown in chapter 4.4.1.2. The data in Figure 4.4.2.2.1 were normalised to plot the relative change in polarisation signal (Δr) against concentration.

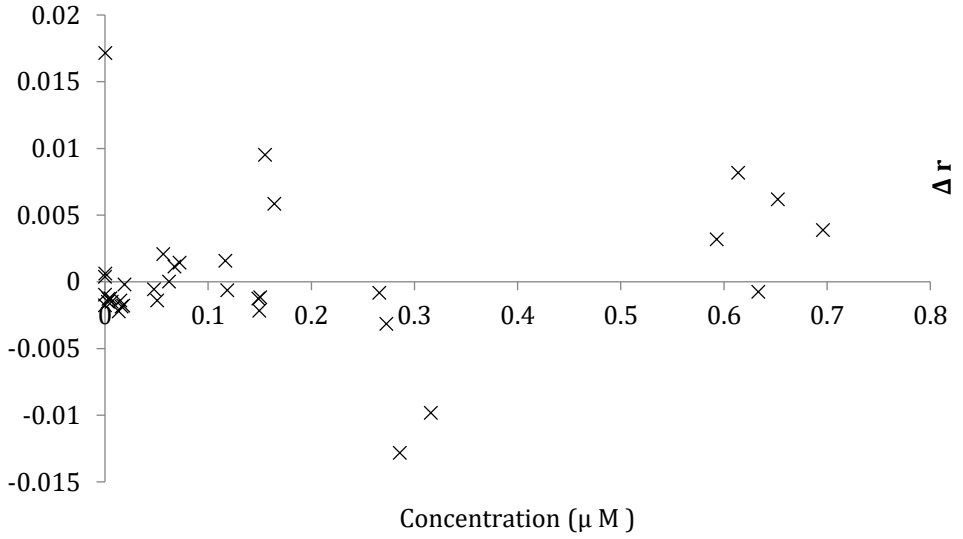


Figure 4.4.2.2.1 Change in observed polarisation plotted against concentration.

In comparison to the wild type peptide there is no large increase in polarisation signal (Δr). Concentrations were tested from 10^{-11} to 10^{-5} M where it was clear that there was no distinct pattern. The recorded changes in fluorescence shows no binding curve and confirms the non-binding result seen in FRET experiments above.

4.5. FURTHER WORK

4.5.1 ISOTHERMAL TITRATION CALORIMETRY

This technique could help corroborate the thermodynamic values obtained in chapter 4.4 where the discrepancies may be identified and the binding model confirmed. Further ITC data could help to understand the stoichiometry of the binding and if indeed as hypothesised, the value of n>1.[115] The cooperativity, as touched upon by E. Antonio can also be probed using this technique.

4.5.2 PHOTO-CROSSLINKING

The wild type F peptide could be engineered to include a photocrosslinker, where the peptide would bind the matrix protein and the binding would be analysed by mass spectrometry. A photochemical crosslinking agent is a chemically inert compound, which require energy from light to initiate and become reactive.[127] Many classes of compounds are available but aryl azides and diazirines are most commonly used for their relative low cost and ease of use.[127] The crosslinked protein and peptide would be digested using trypsin to give small manageable peptide sequences, which would be identified by MALDI mass spectrometry.[127] The crosslinked peptide and matrix protein interaction can be mapped and further mutations performed to analyse the binding to the matrix protein.

4.6 CONCLUSIONS.

In this chapter the primary aim was to confirm that the cytoplasmic tail of fusion protein binds *in vitro* to the hRSV M protein. Through previous literature[7] it has been determined that the residue F572 was key to the interaction between the cytoplasmic tail of hRSV fusion protein and matrix protein. In this chapter, the last six amino acids of the cytosolic tail have been modelled through the means of a small 6 amino acid peptide (NIAFSN). A mutated peptide was also used, where the F572 was exchanged for an asparagine (NIANSN).

Thermal shift analysis, using a SYPRO orange fluorophore was used to establish if the buffering system could stabilise hRSV M. The stabilisation was key due to hRSV M precipitating out of solution within hours once purified.

Surface Plasmon Resonance experiments were conducted to scout the concentrations needed to gain a K_d for the interaction. The technique used small amounts of protein and peptide equally so was ideal for testing concentrations to gain an initial thermodynamic understanding of the interaction. The experiment was conducted in two different ways; where the matrix protein was attached to the chip and the peptide was flowed over and when the matrix protein was attached, more M protein without a tag present and then the peptide. The two different experiments showed dramatically different results, where the binding can be accounted for by two different equilibrium pathways.

The binding was further analysed by two different fluorescent studies, FRET and Fluorescence Anisotropy. The initial FRET experiments using wild-type peptide are detailed from E. Antonio's previous work. They show a clear binding curve that when fit to a rectangular hypobola shows a nM K_d. To confirm the intial K_d the data were fitted to various statistical plots including Klotz, Scatchard and Hill plots. These showed the magnitude of binding was indeed true and further analysis was needed to understand the interaction. The mutant peptide was tested in the same way and showed no specific binding, this confirms the original hypothesis that the F572 residue identified by Shaikh *et. al* [7] is key to the interaction.

The Fluorescence Anisotropy experiments were conducted alongside the FRET experiments using both the wild-type and the mutant peptide. Again a binding curve was

seen and analysis using the same equilibrium pathway showed a binding event that was a 1000 times stronger than the FRET data. Further experiments need to be conducted to understand if the magnitude difference in binding is the technique or the interaction of the peptide and protein. The mutant peptide again showed no binding and confirmed the hypothesis and previous FRET results.

4.7 METHODS

4.7.1 EXPRESSION AND PURIFICATION OF M PROTEIN.

A pET-16-b vector containing the ORF of M protein, obtained from R. P. Yeo, was transformed into *E. coli* BL21-DE3 (see 7.2.1) and grown on ampicillin LB agar plates overnight at 37 °C. One colony was picked from the resulting plate and was incubated in 0.2 L of LB broth (Tryptone 10 g/L, sodium chloride 10 g/L, yeast extract 5 g/L; Melford Laboratories Ltd., Suffolk, UK, 25 g/L) overnight at 37 °C. 0.01 L of the starter culture was used to inoculate 1 L of LB Broth in a 2.5 L conical flask, the bacteria were grown in the presence of Ampicillin (100 mg/mL) at 37 °C for 5 hours or until Optical Density at 600 nm reached 0.6-0.8. Expression was induced using isopropyl-β-D-1-thiogalactopyranoside (IPTG, 0.8 μM; Sigma Aldrich, UK) and left to shake overnight at 28 °C at 180 rpm. The bacterial cells were harvested by centrifugation at 4000 x g for 20 minutes in a Beckmann Coulter ™ J-20 XP centrifuge, using rotor JLA 8000 and the cell pellet was stored at -80 °C.

Protein extraction was achieved through resuspending the pelleted cells in 0.02 L of Lysis Buffer (50 mM Tris,1 M NaOH, 5 mM $CaCl_2$, 5 mM $MgCl_2$ pH 7.6). Lysozyme (1 mg/mL) and protease inhibitor tablet (cOmplete Mini™, EDTA-Free, Roche Diagnostics Ltd, West Sussex, UK) were added once the pellet was resuspended. The cells were lysed by sonication for 240 seconds at 20 kHz and 3-[(3-cholamindopropyl)dimethylammonio]-1-propanesulphonate hydrate (CHAPS, final concentration 8 mM in ddH_2O) was added to the mixture and centrifuged at 50,000 × g for 30 minutes Beckmann Coulter ™ J-20 XP centrifuge, using rotor JA 25.50 to remove cellular debris.

The supernatant was then passed through a pre-equilibrated nickel affinity His-trap column (GE Healthcare Life Sciences, Buckinghamshire, UK). After the supernatant had passed through the column, Tris buffer (50 mM Tris, 300 mM NaCl pH 7.6) was run through, followed by an increasing concentration of imidazole (50 mM Tris, 300 mM NaCl, 50 mM to 1 M imidazole pH 7.6). The protein eluted at 200 – 500 mM imidazole and the fractions were analysed for M protein by SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis)

The protein was buffer exchanged overnight into Tris buffer (50 nM Tris, 200 nM NaCl, pH 7.3) or PBS (Gibco® PBS tablets, Life Technologies, Renfrew, Scotland) by dialysis

(MWCO > 12,000 Da) at 4 °C. The His-tag was removed by cleavage with Factor Xa overnight, as per the manufacturers instructions (Novagen, Merck Chemicals Ltd., Nottingham, UK). The matrix protein was then stored and used at 4 °C.

4.7.2 SDS-PAGE (SODIUM DODECYL SULPHATE POLYACRYLAMIDE GEL ELECTROPHORESIS)

The purified protein was analysed using a 15% acrylamide resolving gel: 5 mL ProtoFlow 30% acrylamide gel (Flowgen Bioscience Ltd., Nottingham, UK), 2.5 mL buffer (1.5 M Tris, 0.4 % (w/v) SDS pH 8.8), 2.3 mL water, 0.1 mL 10 % SDS (Aldrich, Poole, UK), 0.1 mL 10 % (w/v) ammonium persulphate (APS) solution in water and 0.004 mL N,N,N',N'-tetramethylethylenediamine (TEMED; Fluka). The stacking gel was made from: 1.7 mL ProtoFlow 30 % acrylamide gel (Flowgen), 2.5 mL buffer (0.5 M Tris, 0.4 % (w/v) SDS pH 6.8), 5.55 mL water, 0.1 mL 10 % SDS (Aldrich,), 0.1 mL 10 % (w/v) APS solution in water and 0.01 mL TEMED (Fluka). Protein samples were prepared by addition of protein loading buffer (2.5 mL 1 M Tris-HCl pH 6.8, 0.5 mL ddH_20, 1.0 g SDS, 0.8 mL 0.1% bromophenol blue, 4 mL 100% Glycerol, 2 mL 14.3 M β-mercaptoethanol (100% stock) and adjusted to 10mL using ddH_20, 5 μL in each 15 μL sample) and heated to 95 °C for 5 minutes then loaded onto the gel with a protein ladder molecular weight marker in the first well (10 – 250 kDa; New England Biolabs., Hertfordshire, UK). The gel was run at 200 V for 1 hour and stained using InstantBlue™ (Expedeon, Cambridge, UK)

4.7.3 DETERMINING PROTEIN CONCENTRATION

4.7.3.1 NANODROP 1000

The concentration of the protein was determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Kent, UK) by measuring absorbance at 280 nm. The protein concentration was increased using a Vivaspin 6 (MWCO 10,000; GE Healthcare Life Sciences). Samples were centrifuged at 4000 × g in a Beckman Coulter™ Allegra™ X-22R centrifuge for 20 minutes at 4 °C, tested using the NanoDrop 1000 spectrophotometer and repeated until the desired concentration was achieved.

4.7.3.2 BRADFORD ASSAY

Protein concentration was also verified by Bradford assay. Standard concentrations between 5 and 100 μM of bovine serum albumin were prepared and allowed to

equilibrate at room temperature for 30 minutes. Each sample contained the diluted protein sample using MilliQ water to the appropriate concentration at 800 μL and Bradford Reagent (Sigma Aldrich, UK) 200 μL. Each concentration was measured at 595 nm using UV-visable spectrophotometer (Multskan Go, Thermo Scientific, UK) and the absorbance was plotted against the concentration to give a linear plot. The protein was diluted and prepared in the same way and measured at 595 nm. The resulting absorbance was plotted on the standard graph and the concentration was found on the x-axis.

4.7.4 THERMAL SHIFT ASSAYS (TSA)

The interaction of various protein interactions over this work was analysed by thermal shift assay. Fluorescence data was collected on an Applied Biosystems 7500 FAST Real-Time PCR System with an excitation range of 455-485 nm. 8 μL of matrix protein (1mg/mL) in ddH_20 was pipetted into each well of a standard 96-well plate (Starlabs Semi-Skirted FAST) and 10 μL of known concentrations of F peptide were added to wells and 2 μL of SYPRO Orange (5 × final concentration) was also added to give 20 μL in each well.

The fluorescence emission signal at 567-596 nm was used for data analysis. The temperature was held for 1 min per degree from 24 to 95 °C. *NAMI*, a Python programme written by M. Grøftehauge and N. Hajizadeh, was used to analyse the data collected.

4.7.5.1 SURFACE PLASMON RESONANCE

All experiments were conducted using G. E Healthcare Buffers and chips. A Biacore 3000 machine was supplied by G. E Healthcare, UK.

4.7.5.1 MATRIX PROTEIN CONCENTRATION SCOUT.

The NTA Chip was equilibrated using 1 x HBS-EP buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20) in the machine for 20 minutes before use. Activation was achieved by injection of $NiCl_2$ (0.02M, G. E Healthcare) for 60 s at 10 µL/ min The standard NTA Biacore protocol (G. E Healthcare, UK) was then followed which detailed a 10 fold concentration range injected for 60 s at 10 µL/ min. The chip was regenerated using EDTA (0.01M, G. E Healthcare, UK) injected for 60 s at 30 µL/ min.

4.7.5.2 BLANK EXPERIMENT.

Lane 1 on NTA chip was used for blank experiments, no other experiments were conducted on this lane. Activation was achieved by injection of $NiCl_2$ (0.02M, G. E Healthcare) for 60 s at 10 µL/ min, the non-tagged protein was then injected at 60 s at 30 µL/ min at desired concentrations. The chip was regenerated using EDTA (0.01M, G. E Healthcare, UK) injected for 60 s at 30 µL/ min.

4.7.5.3 SECOND PROTEIN ADDITION.

Activation was achieved by injection of $NiCl_2$ (0.02M, G. E Healthcare) for 60 s at 10 µL/ min The standard NTA Biacore protocol (G. E Healthcare, UK) was then followed which detailed the injection of $M_{(His6)}$ protein at the desired concentration for 60 s at 10 µL/ min. The second protein (no tag) was injected for 60 s at 30 µL/ min at the desired concentration. The chip was regenerated using EDTA (0.01M, G.E Healthcare, UK) injected for 60 s at 30 µL/ min.

This was repeated with 5 different non-tagged protein concentrations, which were pre determined using the guidelines detailed in G. E. Healthcare NTA Biacore protocol.

4.7.5.4 BIAEVALUATION ANALYSIS.

A Standard protocol from G. E Healthcare, UK was followed to analyse each data set using the provided software. Each concentration was normalized and blank subtracted. The final curves were exported in a ASCII file and plotted in Microsoft Excel.

4.7.6 RHODAMINE RED STOCK SOLUTION

5 mg of 5(6)-Carboxy-rhodamine *N*-succinimidyl ester (Sigma Aldrich, UK) was dissolved in 1 mL of Dimethylsulphoxide (DMSO) to yield an 8 mM solution. 50 μL of Rhodamine stock solution was incubated with the protein (usually 1 mg/mL) at 22 °C for 4 hours. The tagged protein was separated from the untagged fluorescent molecules by PD10 column (G. E. Healthcare, UK), as per the manufacturer's instructions.

4.7.7 FLUORESCENCE ANISOTROPY

The concentration of stocks containing RR-F peptide were predetermined by E. Antonio and stored at -80 °C. RR-F peptide was added to a standard quartz cuvette and diluted to 500 μL with ddH_20. All measurements were taken on a Cary Eclipse Fluorescence Spectrophotometer (Aligent Technologies, UK) using the supplied Eclipse ADL program. The excitation and emission wavelength (570 nm/580 nm) was added to the program and the voltage was changed to reflect the sample. The voltage was changed until the reading voltage stood at approximately 700 m.A.U. The blank experiment, of only the fluorophore in buffer was taken and the G Factor noted.

Predetermined stocks of M protein were titrated in to a quartz cuvette containing the fluorophore and measured. Each measurement was plotted in Microsoft Excel and the analysis conducted in the same program.

5. NUCLEOCAPSID PROTEIN INTERACTION WITH RSV-604 AND RELATED COMPOUNDS.

The nucleocapsid protein (N) is a key drug target in hRSV as it acts as a replication template, which if disrupted could reduce the virus infectivity.[8,77] Previous work carried out by Arrow Therapeutics Ltd, identified some potential compounds, which were thought to disrupt the nucleocapsid protein structure.[128] This chapter begins with an overview of the N protein and then describes the identification of the new compounds. Further testing of these compounds used a variety of biophysical techniques, which has allowed the characterisation of the interaction in greater detail.

Firstly thermal shift assays were used to probe if the molecules would have a stabilising or destabilising effect on the protein. dynamic light scattering was utilised alongside thermal shift assays to understand if the compounds affect the size and stacking of the nucleocapsid ring formation.[9] As significant changes were seen, transmission electron microscopy was then used to visualise the difference in the protein. The results found begin to build a picture of the interaction and determine which molecules should be taken forward for further testing.

5.1. INTRODUCTION.

The nucleocapsid protein (N) has been previously crystallised as discussed in Chapter 1.[8,9] The structures have enabled drug discovery to target the protein in a way that was not obtainable before. The protein forms a helical structure, as shown below in Figure 5.1.1, which protects the viral genetic material and helps provide a scaffold for viral replication. These helical structures are usually comprised of decameric or undecameric rings inter-wound with viral RNA.[8]

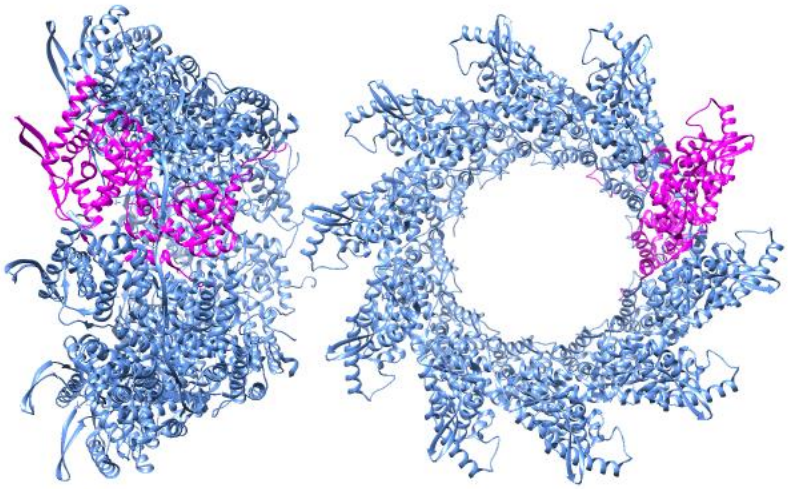


Figure 5.1.1. The crystal structure of hRSV N[8] at 3.3 Å resolution shown as a decamer in blue. A monomer unit is shown in magenta. (PDB code: 2WJ8)

5.1.1 LEAD DRUG COMPOUNDS.

Drug compounds have been previously identified through high throughput screening. 20 000 compounds from Arrow Therapeutics LtD were screened by Chapman *et al.* in a single concentration cell-based XTT (2,3-Bis-(2-Methoxy-4-Nitro-5-Sulfophenyl)-2*H*-Tetrazolium-5-Carboxanilide) assay. These assays identified active compounds which

inhibited any compound involved in protein replication[10] through the use of tetrazolium salt dyes.[10]

The compounds that have been previously identified have potential antiviral activity and were rescreened to identify their EC_{50} values by Arrow Therapeutics LtD. The EC_{50} value is defined as a molar concentration of an agonist, which produces 50% of the maximum response for that agonist. The LD_{50} is defined as the lethal dose of a drug for 50% of the population.[129] These measures are used as an indicator for drug potency and effectiveness and a number of compounds were identified with good therapeutic ratio, where the LD_{50}/EC_{50} is less than 50 and further tested against four different hRSV strains. These further identified a lead compound A-33903 as shown below.[10]

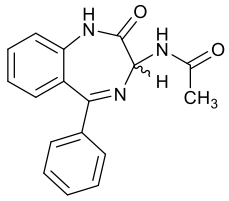


Figure 5.1.1.1. Lead compound A33903.

The lead compound, A-33903 showed an EC_{50} value in the range of 10 to 20 μM over a number of assays. The compound also showed favourable pharmacokinetic properties in a rat model with a half-life of approximately 6 hours and an oral bioavailability of 76%. This suggests the core of the benzodiazepine is stable in metabolism with good absorption properties.[10,130] The lead compound was subjected to structural modifications to increase potency for antiviral activity. The unsubstituted benzodiazepine template was shown to be optimal for antiviral activity but modification of the amide group could increase antiviral activity. An alkyl chain, branching or cyclic showed a decrease in potency but when aromatic amides were tested an increase in potency was observed. A further increase in potency was seen when electron-donating groups were included.

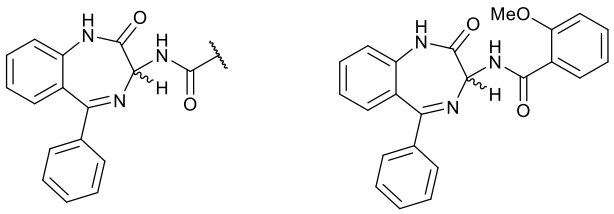


Figure 5.1.1.2. Left, lead compound A33903 with R position for electron donating group additions. Right, compound 3, R= o-methoxybenzamide.[10]

Compound 3 shown above contains electron donating constituents however, when tested in the rat model this compound showed poor bioavailabilty (4%) and a poor pharmacokinetic profile. This could be due to the methyoxy side chain forming unfavourable contacts such as electrostatic interactions, which could destabilise the pocket and ligand.[10] Further modifications were tested at position R as shown in Figure 5.1.1.3.

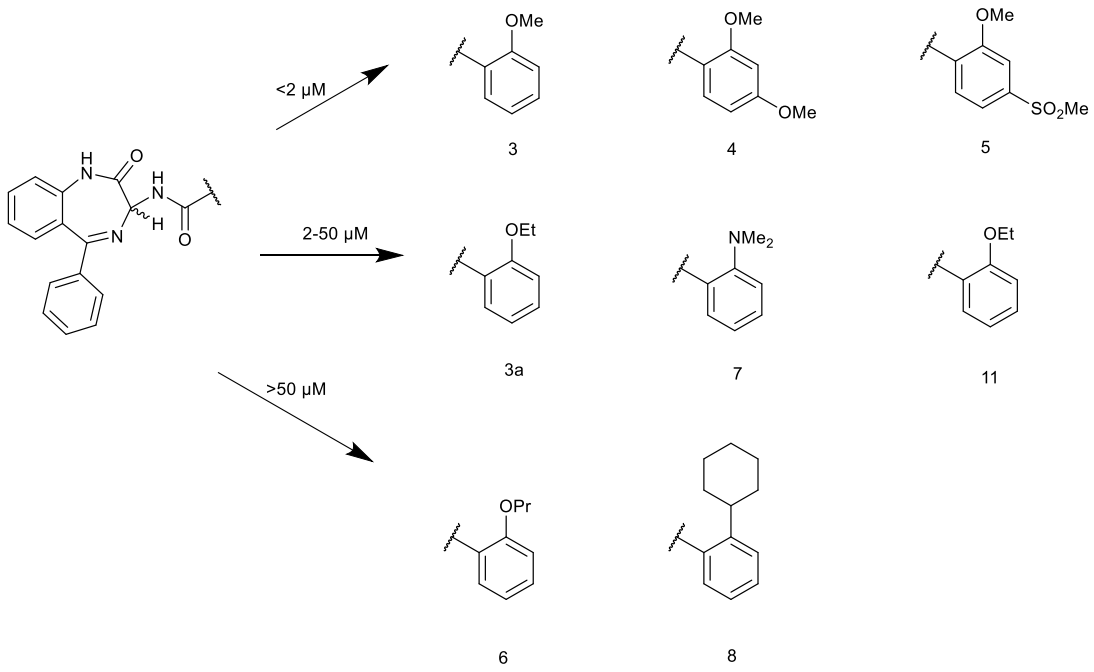


Figure 5.1.1.3. A33093 lead compound shown with primary targets.[10]

Compound 4 which contains a second methyl group retained the activity of compound 3, is in contrast to compounds 5 and 6 with lengthened alkyl chains and reduced activities. Other compounds i.e. where R = dimethylamino (compound 7) with electron-donating substituents gave reasonably potent molecules, but did not increase the potency significantly above that of the lead compound.[10,128] The combination of a lipophilic, electronically neutral substituent i.e. the cyclohexyl group (compound 8), gave a large decrease in potency.

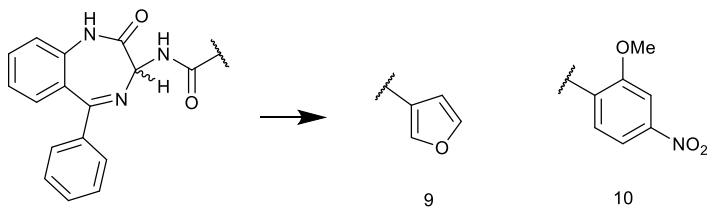


Figure 5.1.1.4. Structure activity relationship (SAR) expansion of the amide substituent in A33903.[10]

Further structure activity relationship (SAR) studies showed that heteroaromatic compounds (compound 9) gave outstanding activity, but, like the electron-rich aromatic amides, have poor pharmacokinetics when tested (compounds 3-5, Figure 5.1.1.3). The addition of nitro groups was employed to block potential aromatic hydroxylation of electron-withdrawing substituents. This gave compounds 9 and 10 as shown in Figure 5.1.1.4 above. These proved to be potent molecules, but were found to be readily metabolised. The lead was further modified to incorporate the 4-trifluoromethyl group, shown as compound 14 below in Figure 5.1.1.5. This new modification was potent in its antiviral activity, but was shown to be highly cell-toxic.[10,128] The left pendant phenyl ring with a methyl substituent group gave poor activity when replaced. This was seen when the pendant ring was replaced with a cyclohexyl, benzyl or pyridine ring.[10,128]

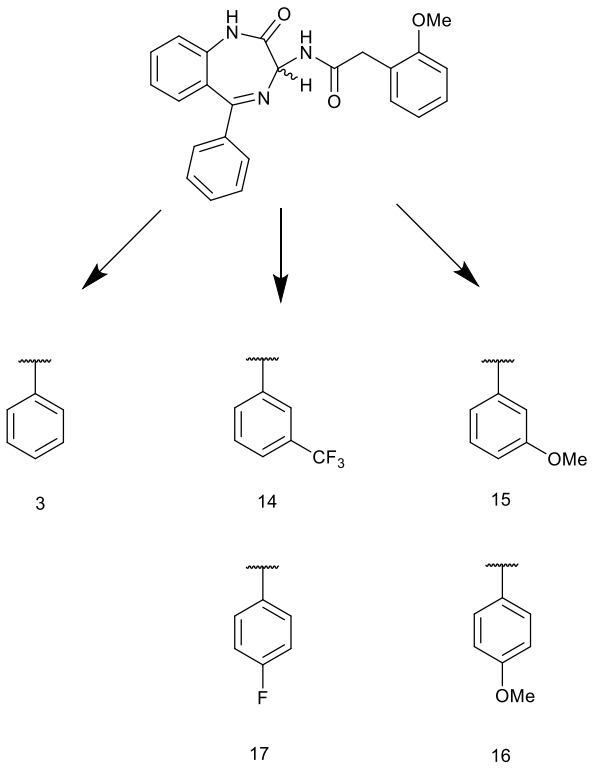


Figure 5.1.1.5 Pendant phenyl substituents[128]

Further SAR studies[130] were conducted on the pendant phenyl to see if any substituent addition could increase potency further as shown above in Figure 5.1.1.5. Any additional substituent was found to have a lowering of potency, only the substitition at the meta-position with an electron-withdrawing group (compound 14) could retain any antiviral activity. Further SAR studies added halogen substituents on to the benzodiazepine core as shown in Figure 5.1.1.6.

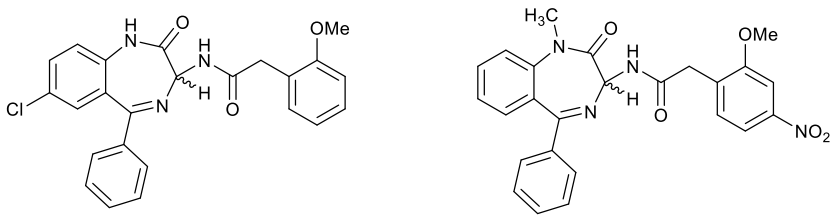


Figure 5.1.1.6. Compound 18 showing the loss of activity in the chloro-substituent. In 19 the addition of a nitro group in the *para* position to the linker region and a methyl group on 7 membered ring increases activity.[128]

Each enantiomer was tested for activity using the high throughput XTT assays used previously[10,71] and overwhelmingly the S-enantiomer showed a higher activity and potency so this was taken forward into trials. A second amine group was added at the last stage to further increase potency after the amide group in the linker region. This resulted in the final compound, RSV-604 shown in Figure 5.1.1.7, which maintained high potency and had good bioavailability in a rat model.

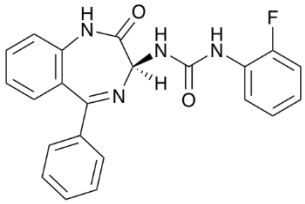


Figure 5.1.1.7 RSV-604, lead drug compound from the initial screen by Arrow Therapeutics Ltd.[10]

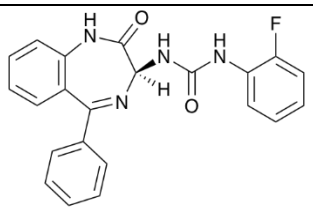
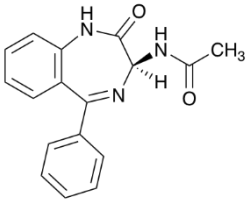
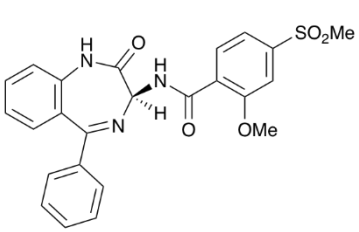
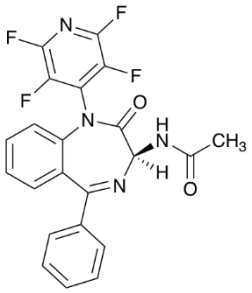
Further *in vitro* testing of the lead compounds is required from the initial screen to confirm and understand the anti-viral activity identified in the high throughput assays.

The overall aim for this part of the project involving the nucleocapsid protein was to ascertain if the ligand molecules will interact with this specific protein. Once this had been determined, the compounds that did interact were taken forward to identify binding thermodynamics. Experiments were conducted using *in vitro* biophysical techniques i.e. Surface plasmon resonance and differential scanning fluorimetry. The final question which was answered is: if the compounds do bind to the N protein, how does this binding affect the stability of the protein in solution? The N protein rings were imaged by transmission electron microscopy in order to visualise this change. Future work comprising *in vivo* studies with compounds 45 and 46 shown in Table 5.1.2.1, and

other active compounds, will also be key to understanding their potency and modes of action.

5.1.2. COMPOUNDS AVAILABLE FOR TESTING.

In 2013 S. Cockerill approached the VAM/RPY research team for further *in vitro* testing of the lead compound RSV-604 and a range of compounds made in the follow-up SAR studies. Compounds 16 and 18 were originally tested in the initial screening and modification stage by S. Cockerill and J. Chapman at Arrow Therapeutics Ltd.[130] 45 and 46 were part of a further modification round of unpublished data and not included in the initial screening, these compounds are novel and little data has been collected either *in vivo* or *in vitro*. Compound 46 is a racemic mixture. From here on the compounds will be referred to as Compound 15, 16 etc.

	RSV604 - Compound 15.
	Lead compound A33903 - Compound 16.
	Compound 18
	Compound 45.

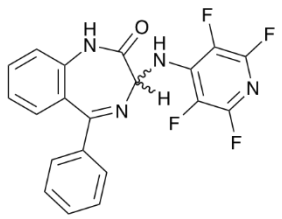
	Compound 46.

Table 5.1.2.1. Lead compounds available for testing.[10]

5.2. BIOPHYSICAL CHARACTERISATION.

5.2.1 PURIFICATION OF NUCLEOCAPSID PROTEIN.

The hRSV N gene was cloned into a pET16b vector by Dr. R. P. Yeo. All work carried out using the vector was performed by the author and R.G. Stewart, MSci student.

The purification of hRSV N was achieved by following the Tawar *et al*[8] protocol where the N protein was purified off the His-column using multiple washes of increasing concentration of imidazole from 20 mM to 500 mM. The lowest concentrations of imidazole remove any non-specific protein which have bound to the column. The higher concentrations elute the protein, which was collected in 2 mL fractions and its size and purity were confirmed by SDS-PAGE below.

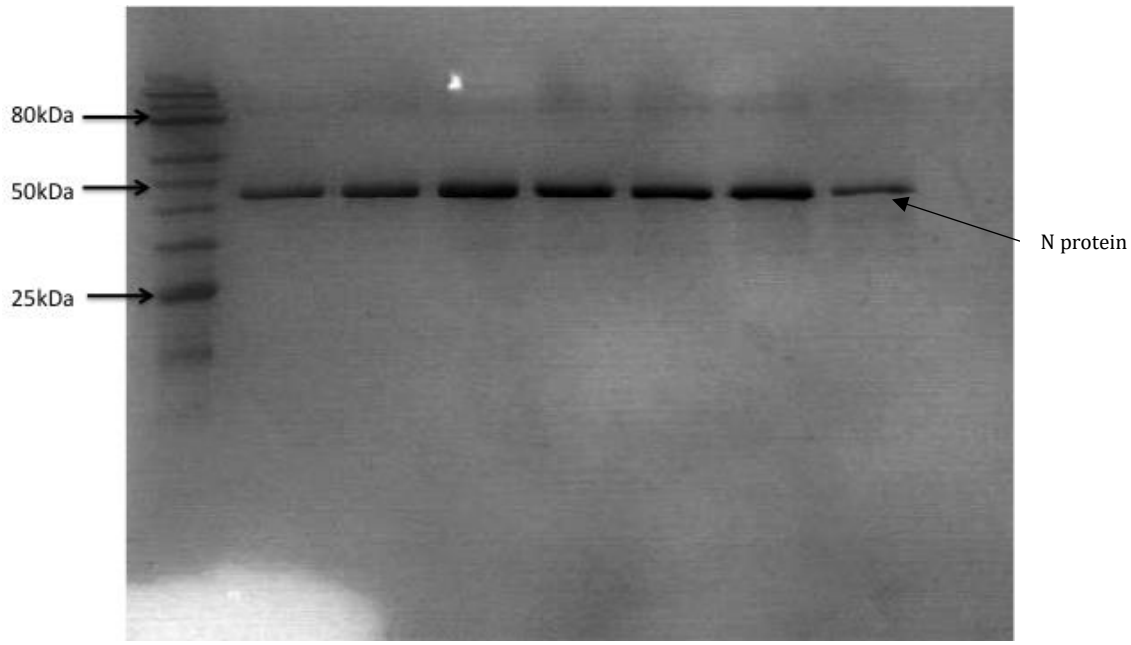
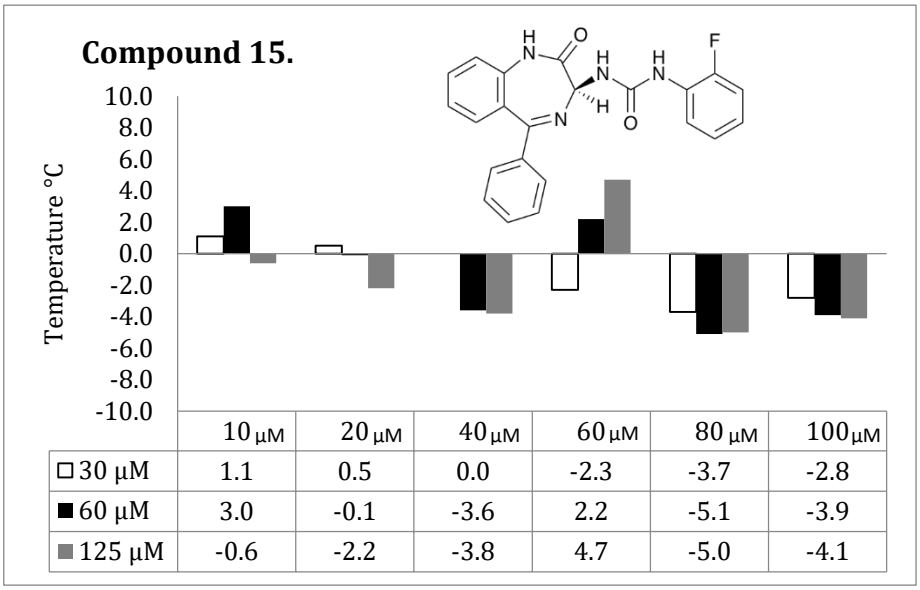


Figure 5.2.1.1 15% SDS page gel showing fractions at 300 mM Imidazole elution. Nucleocapsid protein is seen at 45 kDa from the His-Trap Column.

From Figure 5.2.1.1 it can be seen that pure, recombinant N protein was obtained after expression in *E. coli*. If a preparation of protein was found to be impure a size exclusion column was used as a final purification step.

5.2.2. THERMAL SHIFT ASSAYS (DIFFERENTIAL SCANNING FLUOROMETRY)

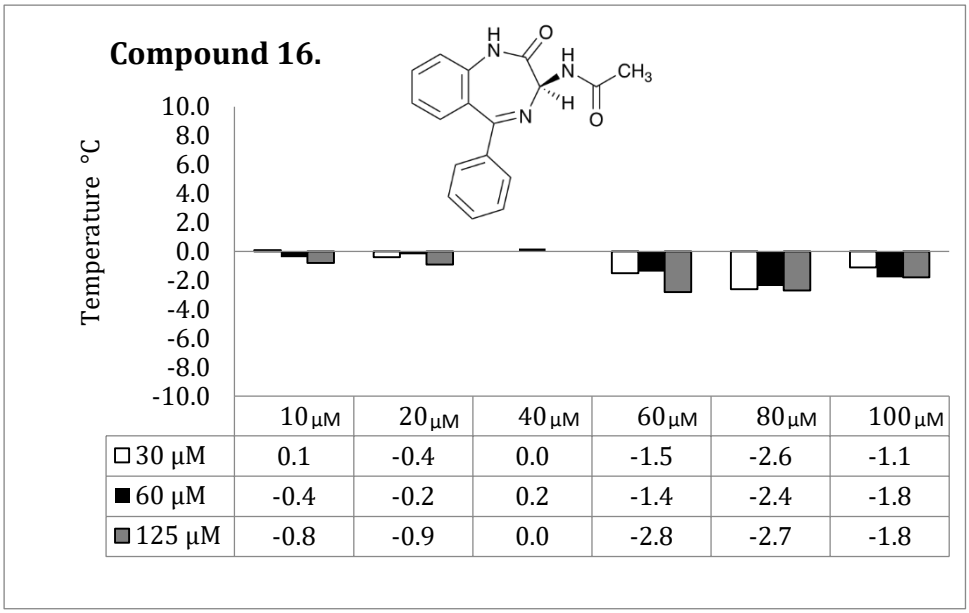
As described in Chapter 2.5, thermal shift assays can be used to determine the thermal stability of the protein in solution. This approach was used to begin to investigate the effect of the lead molecules on the nucleocapsid protein in solution. The protein was tested at varying concentrations (10 μM to 100 μM) and using different concentrations of drugs (30 μM**,** 60 μM and 125 μM). Using different concentrations allows the probing of concentration dependant binding in both the protein and the drug compound added, Figure 5.2.2.1.



	10 μM	20 μM	40 μM	60 μM	80 μM	100 μM
□ 30 μM	1.1	0.5	0.0	-2.3	-3.7	-2.8
■ 60 μM	3.0	-0.1	-3.6	2.2	-5.1	-3.9
■ 125 μM	-0.6	-2.2	-3.8	4.7	-5.0	-4.1

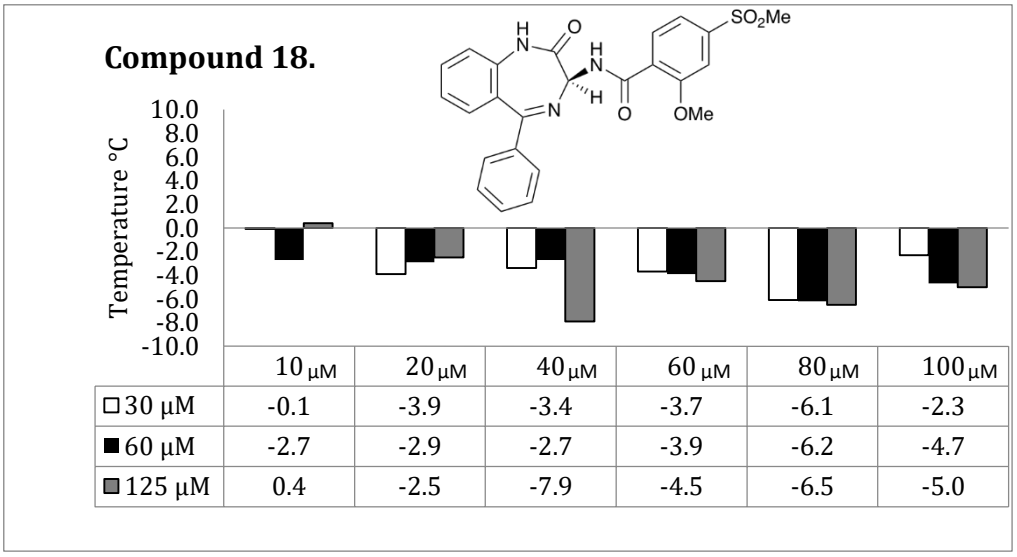
Figure 5.2.2.1. Lead compound RSV604, lead compound 15. [10] Where concentration of protein ranging from 10 μM to 100 μM **.**

RSV-604 is the lead compound to which the others are compared. The range of results shown above shows no major trend, but the ligand and the protein do interact in the assay. It is known that RSV-604, compound 15, is active with hRSV N *in vivo* but these results cast doubt on the effectiveness of the compound *in vitro* or the technique. The full method is detailed in Chapter 5.4.



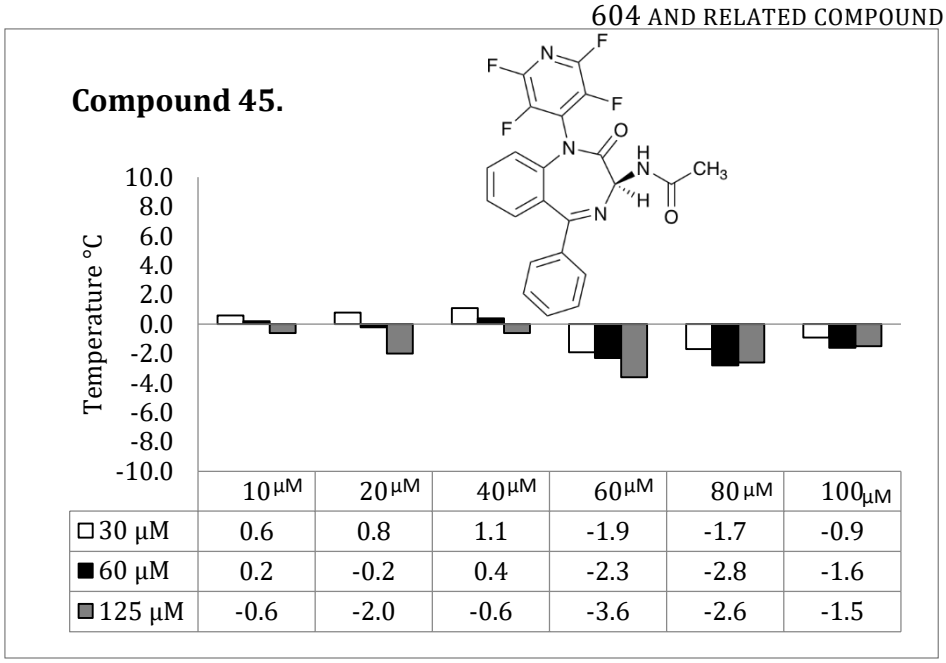
	10 μM	20 μM	40 μM	60 μM	80 μM	100 μM
30 μM	0.1	-0.4	0.0	-1.5	-2.6	-1.1
60 μM	-0.4	-0.2	0.2	-1.4	-2.4	-1.8
125 μM	-0.8	-0.9	0.0	-2.8	-2.7	-1.8

Figure 5.2.2.2. Compound 16. Where x axis is each concentration of protein ranging from 10 μM to 100 μM .



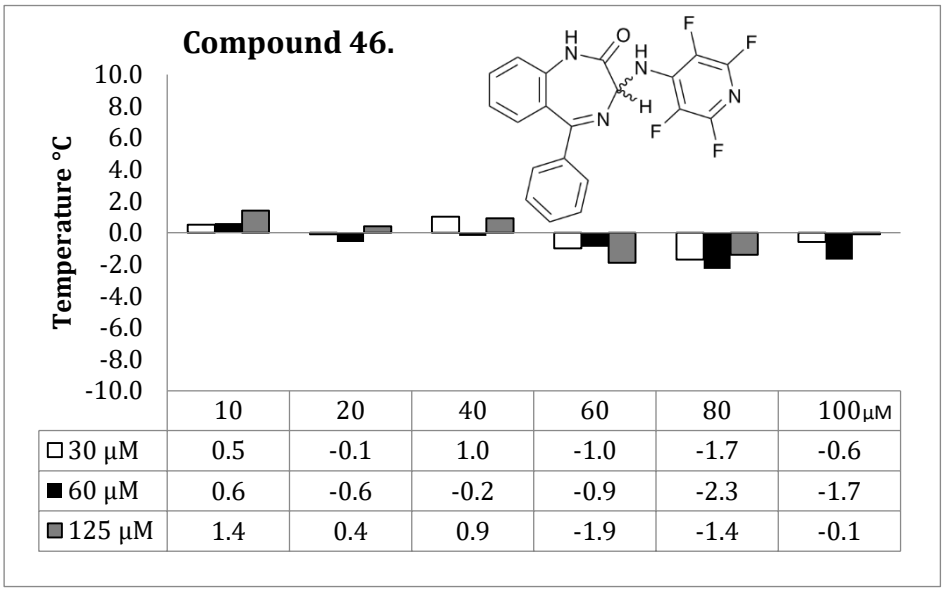
	10 μM	20 μM	40 μM	60 μM	80 μM	100 μM
30 μM	-0.1	-3.9	-3.4	-3.7	-6.1	-2.3
60 μM	-2.7	-2.9	-2.7	-3.9	-6.2	-4.7
125 μM	0.4	-2.5	-7.9	-4.5	-6.5	-5.0

Figure 5.2.2.3. Compound 18. Where x the axis is the concentration of protein ranging from 10 μM to 100 μM.



	10μM	20μM	40μM	60μM	80μM	100μM
□30 μM	0.6	0.8	1.1	-1.9	-1.7	-0.9
■60 μM	0.2	-0.2	0.4	-2.3	-2.8	-1.6
■125 μM	-0.6	-2.0	-0.6	-3.6	-2.6	-1.5

Figure 5.2.2.4. Compound 45, Where x the axis is the concentration of protein ranging from 10 μM to 100 μM.



	10	20	40	60	80	100μM
□30 μM	0.5	-0.1	1.0	-1.0	-1.7	-0.6
■60 μM	0.6	-0.6	-0.2	-0.9	-2.3	-1.7
■125 μM	1.4	0.4	0.9	-1.9	-1.4	-0.1

Figure 5.2.2.5. Compound 46. Where x the axis is the concentration of protein ranging from 10 μM to 100 μM.

All the lead compounds showed an effect on the N protein with the most marked being destabilisation of the protein in solution. The thermal shift assay, although sensitive, can only be used to determine if the drug compound is affecting the protein in solution and all compounds tested did show an effect on the N protein.

The drug compound could be weakly binding to the protein to be providing the stabilising and destabilising effects seen above. Other techniques are needed to understand the nature of the interaction so all drugs were taken forward at this point for further investigation using dynamic light scattering (DLS). DLS gives a real time size estimation of species in solution and hopefully how the addition of ligands can affect the protein in solution.[92]

5.2.3 DYNAMIC LIGHT SCATTERING.

As described in Chapter 2 dynamic light scattering can be used to obtain the hydrodynamic radius of species in solution.[92] This is particularly vital for the nucleocapsid protein as the protein self assembles into decameric rings and then stacks into a large nucleocapsid form.[8,9] Previously there have only been crude assays available to obtain interaction data between N and the various ligands, none of these indicate the oligomerisation state of the protein, if it changes conformation on ligand binding or how the drug interacts.

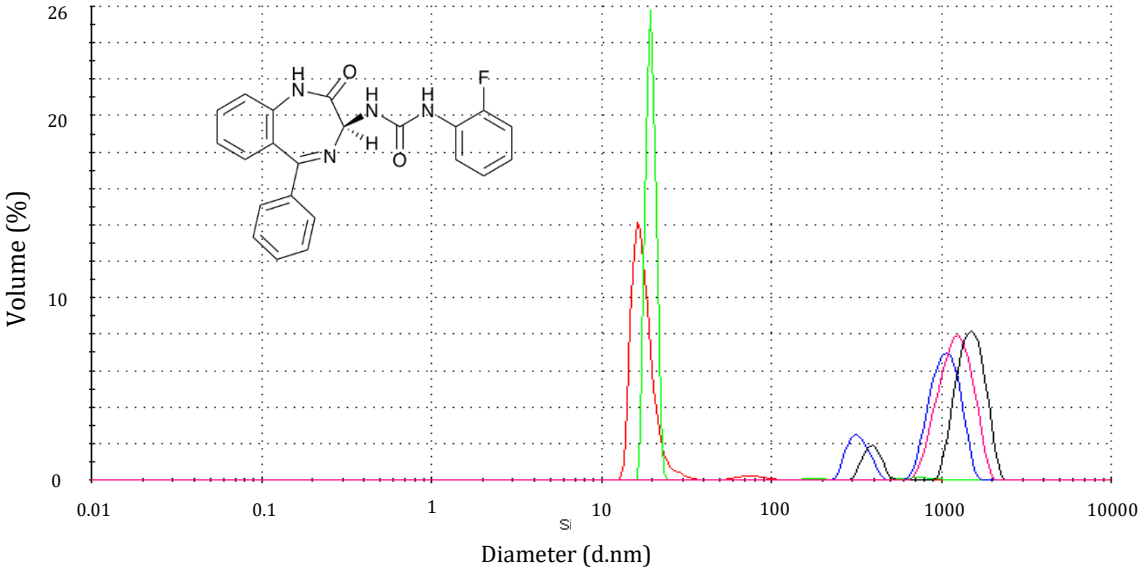


Figure 5.2.3.1. Compound 15 trace as shown on the Malvern Zetasizer software, each colour represents a different ligand to protein concentration shown below.

Colour						
Protein concentration	50 μM	50 μM	50 μM	50 μM	50 μM	50 μM
Ligand concentration	0 μM	50 μM	100 μM	150 μM	200 μM	500 μM

Legend of protein and ligand concentrations used in Figure 5.2.3.1, 5.2.3.2 – 5.2.3.5.

Compound 15 shows a shift in the protein size above a drug concentration of 100 μM. The polydispersity of the species in solution also slightly falls indicating a uniformed aggregation as shown in Table 5.2.3.7, other lead compounds were tested in the same manner.

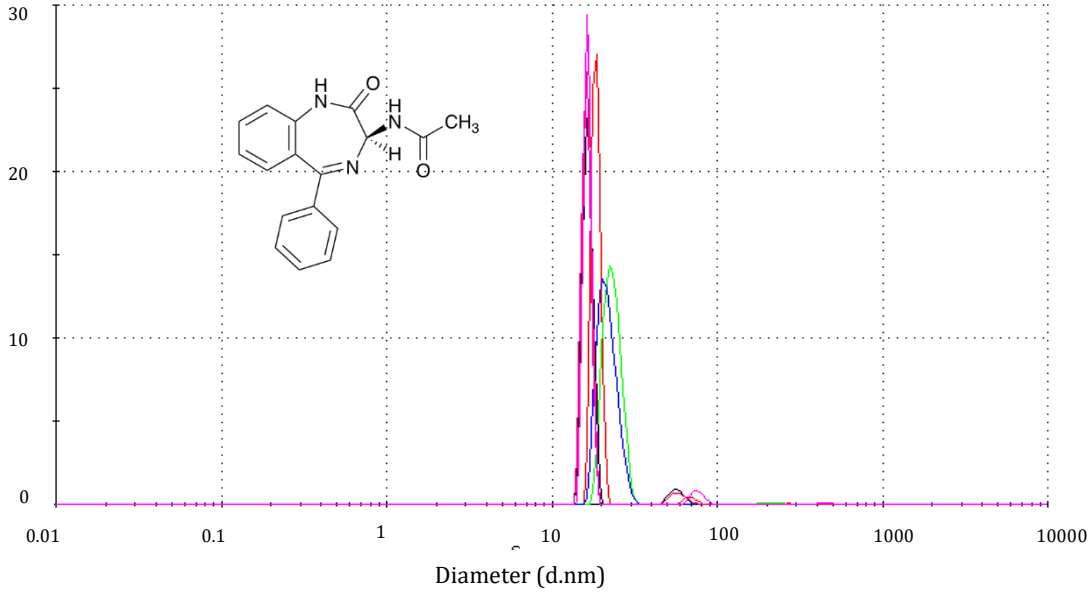


Figure 5.2.3.2. Compound 16 trace as shown on the Malvern Zetasizer software, each colour represents a different ligand to protein concentration.

Compound 16 does not seem to form a larger species by mass as the compound 15 does. As it is unknown how the compounds interact with the protein it is hard to hypothesise the role of the larger species. Again the polydispersity decreases accordingly with increasing drug concentration but not in the same magnitude as compound 15.

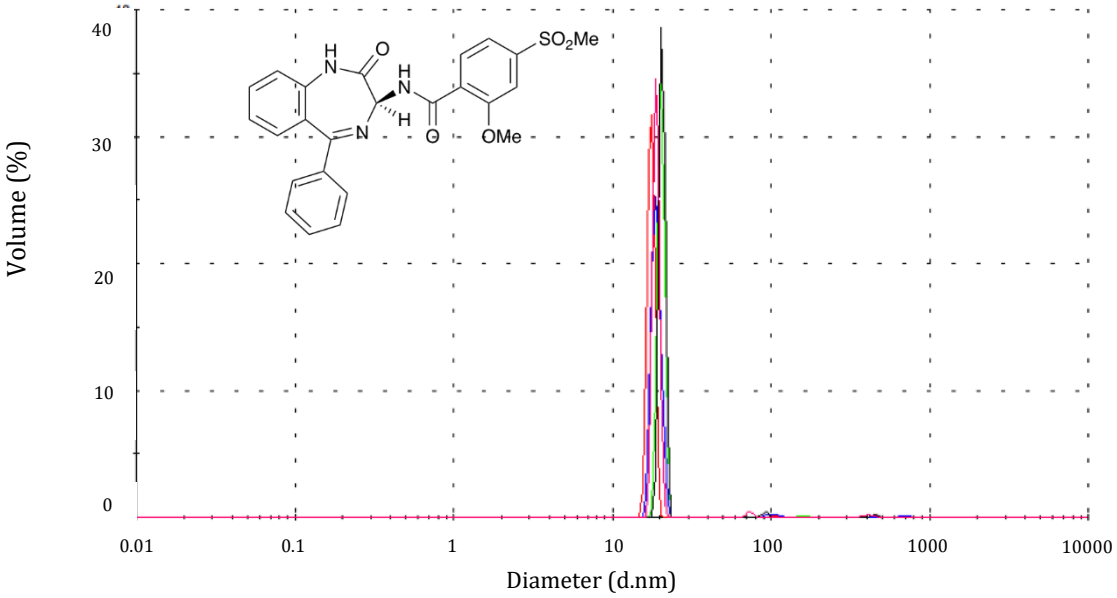


Figure 5.2.3.3 Compound 18, as shown on the Malvern Zetasizer software, each colour represents a different ligand to protein concentration.

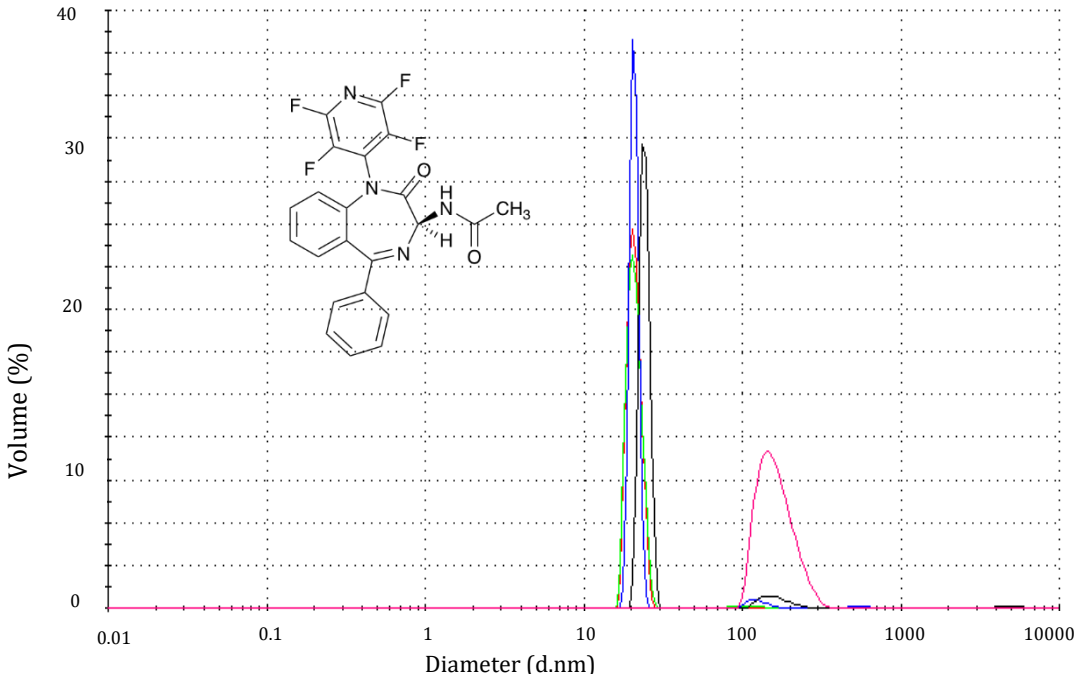


Figure 5.2.3.4 Compound 45, shown on the Malvern Zetasizer software, each colour represents a different ligand to protein concentration.[94]

Compound 45 shown above, shows a negligible increase in particle size until the addition of 200 μM where the % volume of aggregate increases significantly. In comparison to drug compound 46, compound 45 seems less potent in the formation of aggregates.

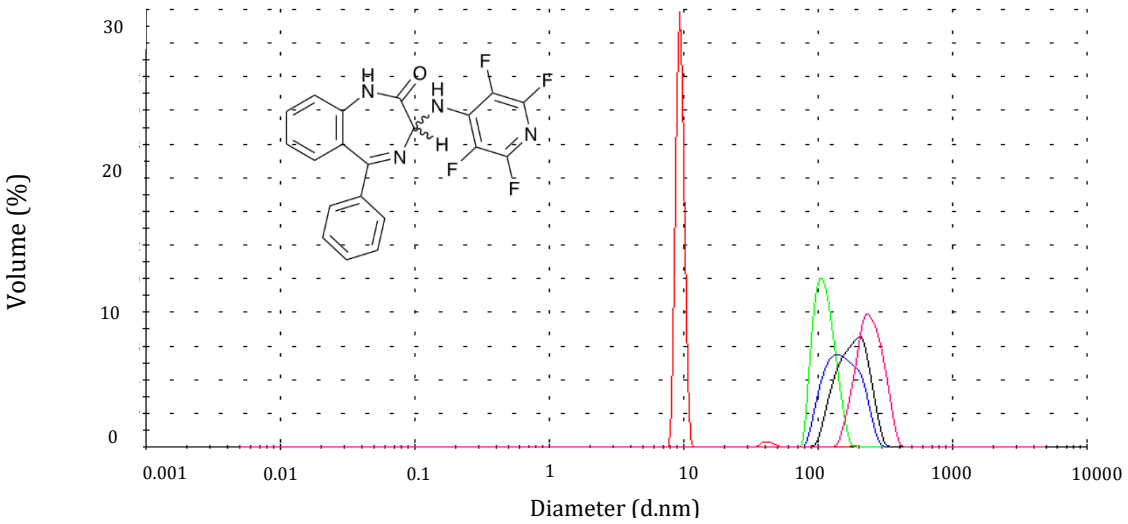


Figure 5.2.3.5 Compound 46 shown on the Malvern Zetasizer software, each colour represents a different ligand to protein concentration.[94]

	Polydispersity of the aggregates				
	15	**16**	**18**	**45**	**46**
N	0.981	0.639	0.729	0.802	0.749
N + 50 µM drug	0.861	0.986	0.444	0.671	0.149
N + 100 µM drug	0.952	0.718	0.494	0.795	0.141
N + 150 µM drug	0.620	0.695	0.535	0.372	0.099
N + 200 µM drug	0.570	0.687	0.443	0.224	0.045

Table 5.2.3.7. Polydispersity of N in the presence of the compounds. A value of 1 indicates the sample has no monodisperse species in the solution. Polydispersity in compounds 45 and 46 especially drops significantly.

Polydispersity is representative of the overall protein particle size distribution width, where the term is calculated by the width of the DLS peak at half the height.[100] Drug compounds 45 and 46 are from a further subset by Arrow Therapeutics LtD but have not been through *in vivo* testing. The results show a large decrease in the polydispersity, indicating the formation of a more monodisperse species. In the case of compound 15 the polydispersity does decrease the most out of the tested compounds but 45 and 46 seem more effective in initiating the aggregation. It is unknown if the protein is stacking in an ordered way on addition of compound, which may still allow the protein to function, or if there is random aggregation occurring. If the latter is occurring, this could be the compounds' mode of action and could affect the nucleocapsid formation.

The results above support the thermal shift assay results, mainly that the compounds have an effect on hRSV N. The formation of species with a large hydrodynamic radius in the solution indicates an aggregated species, but further experiments would be needed to characterise this.

5.2.4 TRANSMISSION ELECTRON MICROSCOPY. (TEM)

As shown in Chapter 3.2 Transmission electron microscopy is a powerful tool to image biological samples.[106] The aggregation seen in Chapter 5.2.3 provides many questions to understand which reactions are taking place and if the ligand is binding to the protein. The images were obtained with the help of A. C. Richardson, Durham University. The methods and parameters are detailed in Chapter 5.4 for clarity.

5.2.4.1 NUCLEOCAPSID TEM.

N protein was first imaged without any drug compounds present as seen in Figure 5.2.4.1.1 below. Individual rings can be seen which are consistent with the crystal structure in shape and also size.[8]

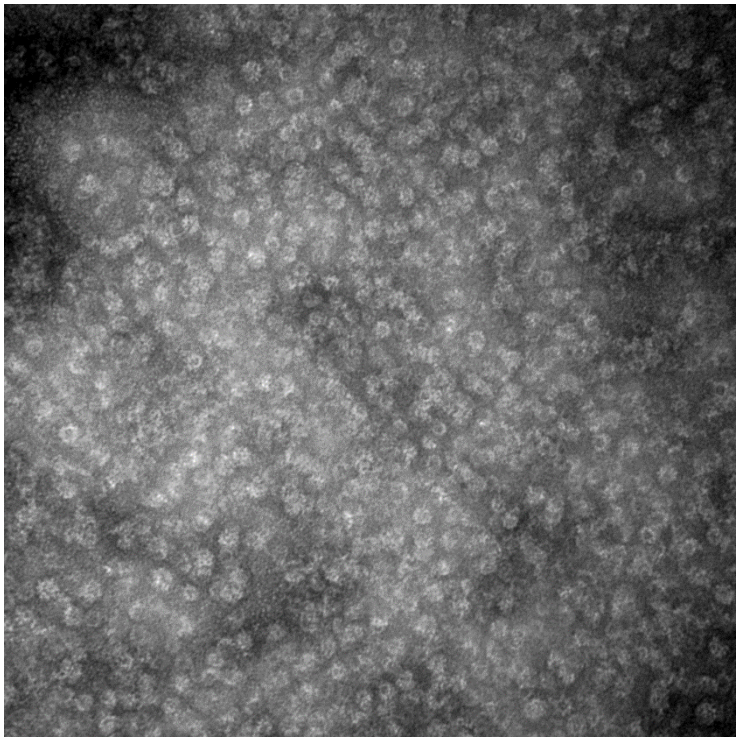


Figure 5.2.4.1.1 Nucleocapsid protein (100 μM) TEM image, stained using uranyl acetate at 200 000 × magnification

5.2.4.2. NUCLEOCAPSID PROTEIN WITH RSV604, COMPOUND 15.

As compound 15, RSV604 is a clinical candidate and is known to be the most potent of the tested compounds *in vivo* hRSV N was tested first to understand how the compound is affecting the protein.

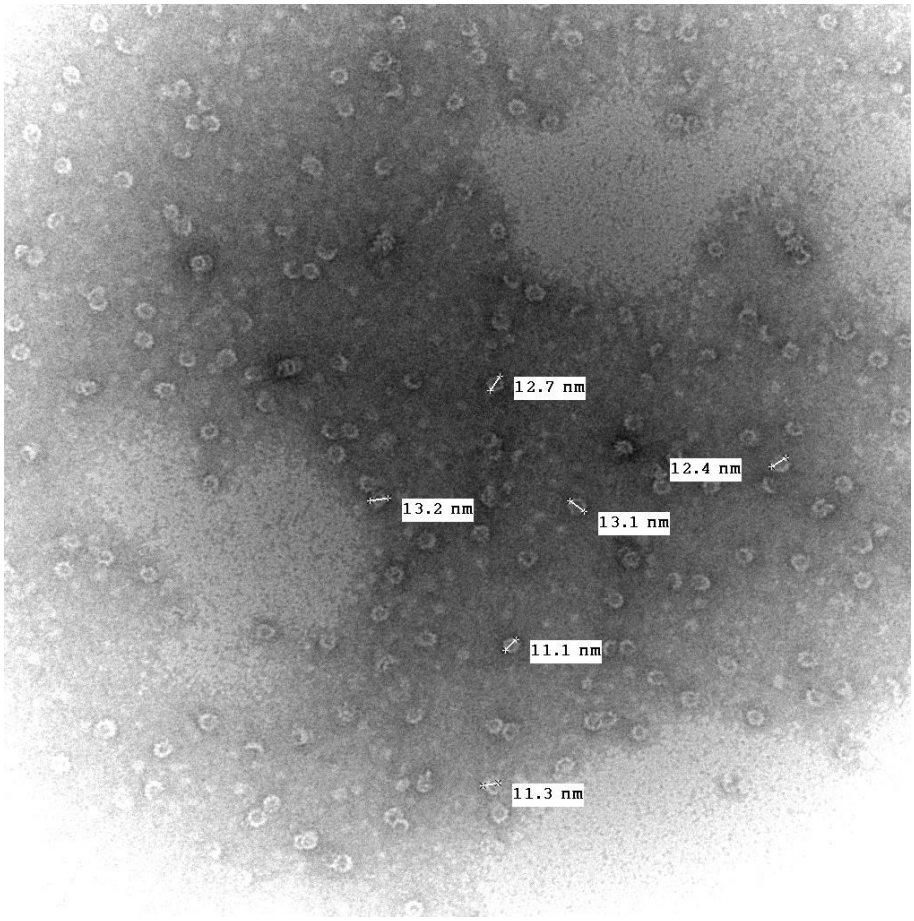


Figure 5.2.4.2.1 Nucleocapsid protein (100 μM) under TEM, shows a variation in size at 120 000 × magnification.

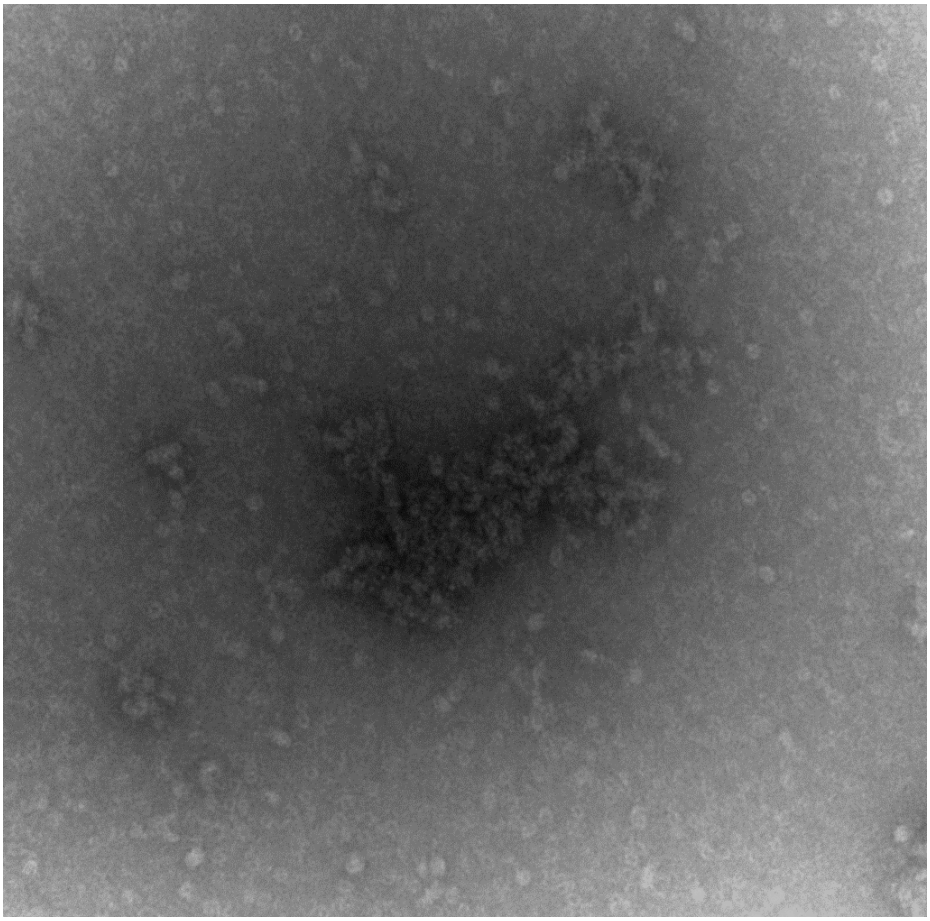


Figure 5.2.4.2.2 TEM image of Compound 15 (100 μM) and nucleocapsid protein (100 μM) aggregation at 120 000 × magnification.

A mixture of compound 15 (100 μM) and nucleocapsid (100 μM) as also imaged as shown above in Figure 5.2.4.2.2. From the above image it can be seen that the aggregates are formed from the N rings. Dynamic light scattering experiments showed that 100 μM concentration of drug induced a change in the hydrodynamic radius of the species in solution and so this would suggest that above this concentration compound 15 is active in forming aggregates.

5.2.4.3 NUCLEOCAPSID PROTEIN WITH COMPOUND 46.

Compound 46 resulted in a decrease in polydispersity and large increase in hydrodynamic radius of aggregates of N above a 200 μM concentration in dynamic light scattering experiments. The images obtained are shown below in Figures 5.2.4.3.1 and 5.2.4.3.2

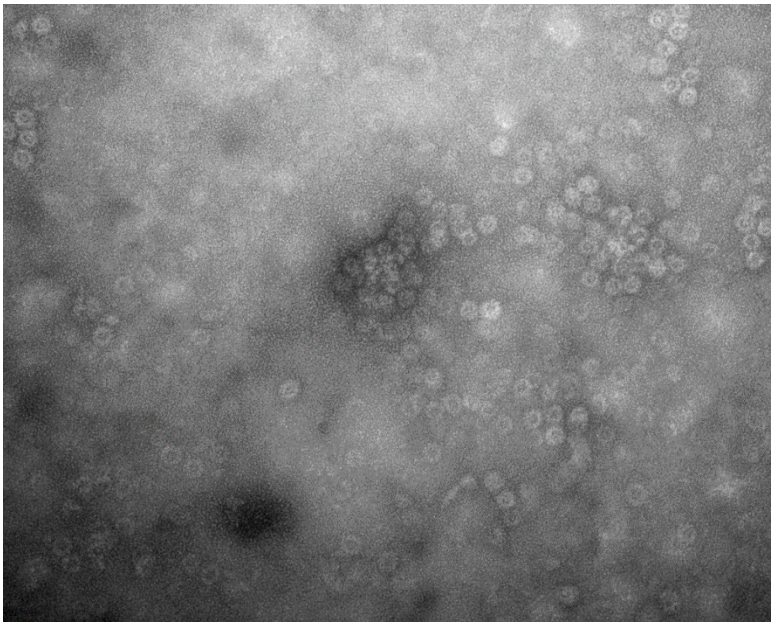


Figure 5.2.4.3.1 Initial aggregation of nucleocapsid protein (100 μM) and compound 46 (100 μM).

The protein rings can be seen in solution, but also the formation of a larger species. The aggregation that occurs seems to be random and disordered. Initially, it was thought that the drugs may create order and stack the protein to form filaments, but this shows the opposite. Further aggregation can be seen in Figure 5.2.4.3.2.

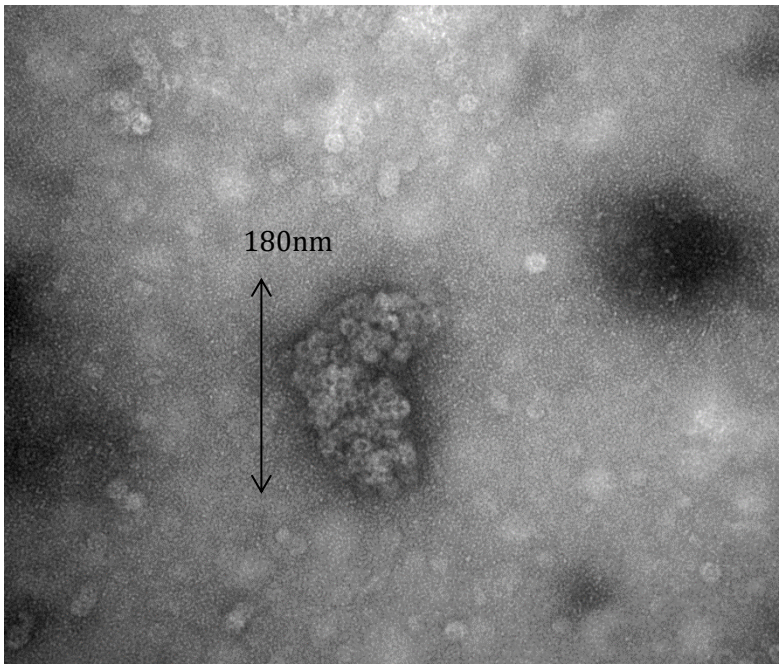


Figure 5.2.4.3.2 Further aggregation seen in TEM at the same protein and compound concentration.

The largest protein aggregate is approximately 180 nm in length and 100 nm in width, which correlates with the dynamic light scattering data reported in Figure 5.2.3.5. Compound 46 showed similar results to compound 15 in the TEM images and DLS experiments indicate a potent compound. Further analysis is needed, for example, in the form of ITC to determine the stoichiometry and thermodynamic nature of the interaction.

5.2.5 CO-CRYSTALLISATION.

Co-crystallisation is a key stage on the road to pharmaceutical research and design. The technique allows for a high-resolution representation of the protein and ligand bound in complex.[16] The structure can begin to enable us to understand the mode of action of a drug and where the binding pocket is located in the protein.[131]

There are two main methods for co-crystallisation,[16]

1. Where the two species are incubated together in a known ratio usually in hanging drop trays. This is where the mixture of protein, ligand and buffer conditions are suspended above the well and a diffusion equilibrium is allowed to be established.[16]

2. Crystals of protein alone were grown in favourable buffer conditions and the ligand was then added to the formed crystal. There can be problems with the addition of the ligand cracking the crystal if there is a structural change upon binding.

In 2009 Tawar *et al.*[8] successfully crystallised hRSV nucleocapsid protein in 10% (v/v) 2-methyl-2, 4-pentandiol (MPD), 0.05 M ammonium acetate, 0.05 M Tris pH 7.5, 0.01 M $MgCl_2$. We replicated and expanded the conditions in a sparse matrix screen and nucleocapsid protein (450 μM) was used. Various methods including sitting drop and hanging drop were tested at multiple temperatures which allowed the maximum number of variants to be tested at any one time.

hRSV N was to be crystallised alone to give a target for ligand soaking if the co-crystallisation failed to generate any crystals. This yielded mostly dark amorphous precipitate or phase separation when we employed the 2011 conditions; 10 % (w/v) PEG 400, 0.1 M KCl, 0.05M HEPES pH 7.0, 0.01 M $CaCl_2$.[9] In the 2009 conditions[8] as discussed above, phase separation was seen and other optimisation trays were set down for analysis.

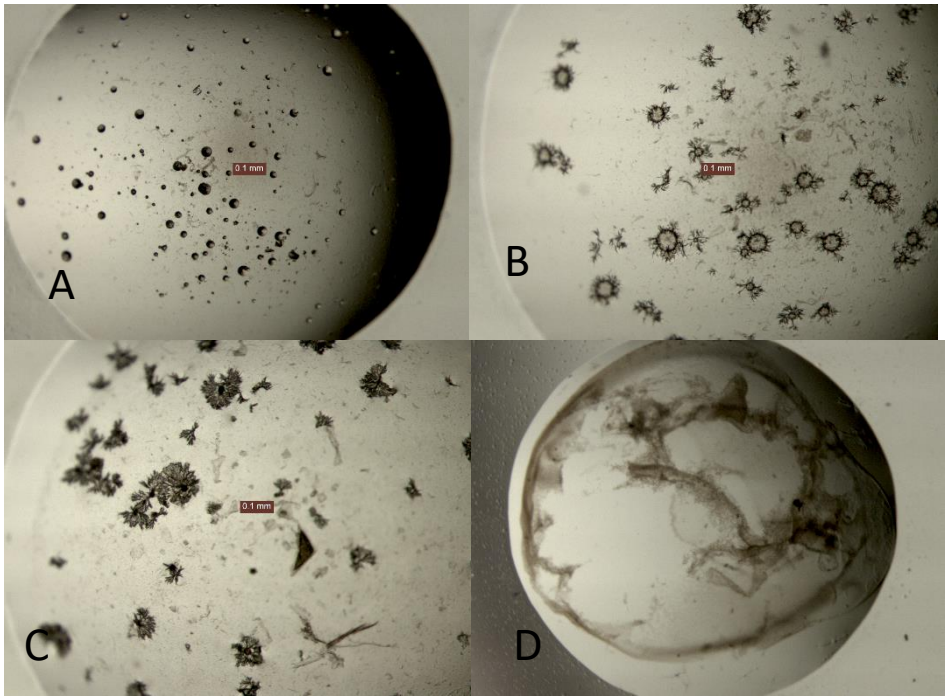


Figure 5.2.5.1 A. Phase separation, spheralites, **B.** Nucleation from spheralites, **C.** Further nucleation, **D.** amorphous precipitate

Co-crystallisation trials of hRSV N and compound 15 were more fruitful, several stages of protein crystal growth were seen in the space of 2 weeks from day 3 as shown in Figure 5.2.5.1. A clear progression is seen in the crystal tray, phase separated spherulites, were seen in the left hand side of the tray where the MPD concentration ranged from 6-8 %. Within 48 hours, the spherulites formed needles from a single nucleation site, which progressed into the needles shown in Picture C in Figure 5.2.5.1. Picture D is from the higher 12-14% MPD in the optimisation tray where the protein precipitated. Further needle-like crystals were found in the drops after seven days as shown below in Figure 5.3.5.1, these were poorly shaped and required further optimisation and streak seeding was used to optimise the crystals.

Seeding uses previously nucleated micro-crystals in new drops, which have been equilibrated at lower levels of supersaturation, this promotes crystal formation and usually produces better shaped crystals for diffraction. Conditions were again optimised to 8 % (v/v) MPD, 0.025 to 0.075 M ammonium acetate, 0.05 M Tris, pH 7.5 and 0.01 M $MgCl_2$. hRSV N was mixed with solid compound 15 and the following crystals grew, which are shown in Figure 5.2.5.2.

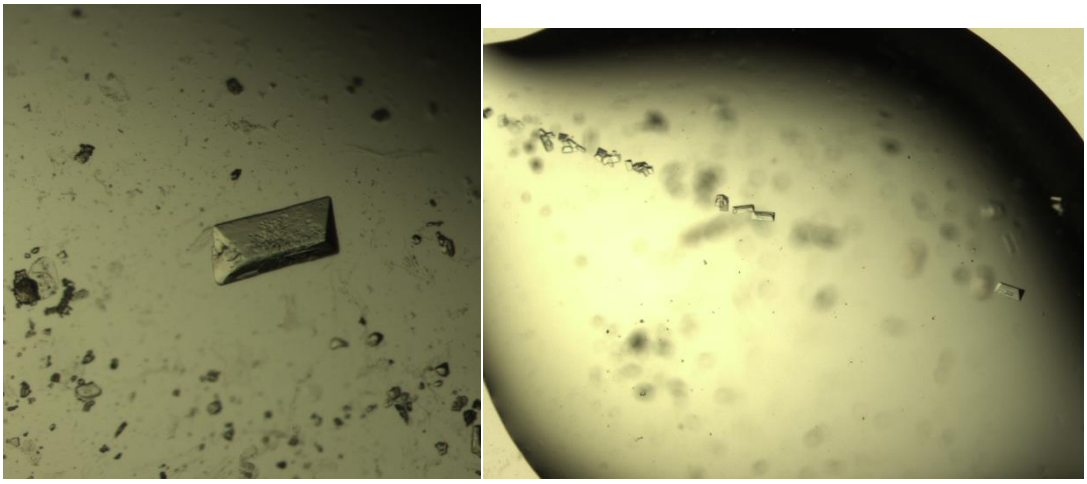


Figure 5.2.5.2 Left, a larger crystal ~200 µM in length. Right, a seeded drop showing smaller crystals ~50 µM in length.

The crystal on the left in Figure 5.2.5.2 was tested on the in-house diffractometer and did not diffract. The smaller crystals on the streak on the right were allowed to grow further before freezing down with MPD and being sent to the Diamond Light Source. Unfortunately, the crystals failed to diffract and further work is needed to investigate different crystal conditions to yield a diffracting crystal.

5.3. CONCLUSIONS

Previous to the work described in this chapter it was known that RSV-604, (compound 15) was taken through to phase two clinical trials, but it was still unknown how the compound interacted with the nucleocapsid protein.[128] The various subsets of compounds were obtained from S. Cockerhill (Durham University). The anti-viral activity of RSV-604 or compound 15 has been confirmed through *in vivo* and *in vitro* data previously.[71] In comparison compound 16 (Lead compound – A33903) and a compound 18 with an electron withdrawing substituent, which were precursors to compound 15 and have not been tested.[10] Compounds 45 and 46 were from a second round of design and had never been tested either *in vivo* or *in vitro.*

Thermal shift assays were used to assess if the compound binds the protein, each of the compounds successfully gave a response and was then analysed by dynamic light scattering. Using the technique it could be seen that at certain compound concentrations the hydrodynamic radius of the aggregates increases rapidly whereas the polydispersity decreases. Compounds 18, 16 and 45 showed less of a change in hydrodynamic radius even at high concentrations. Compound 46 showed a similar response to compound 15, where a smaller concentration of compound added to the protein induces a large increase.

These changes indicate a change in the protein structure or that a new structure forms. Transmission electron microscopy was used to identify the aggregates seen in the DLS experiments. The nucleocapsid protein alone shows small rings, similar to the crystal structure, but when the compounds were added, larger aggregates were seen just as the dynamic light scattering had shown. Large aggregates were seen with Compound 15 and 46, where it can be seen that the nucleocapsid rings are attached to each other. The rings seem to be intact and form a disordered aggregates, not the long ordered tubules that the nucleocapsid protein would form *in vivo*.

Overall this chapter shows a promising beginning for a subset of compounds, whose mode of action was unknown until now. Further work to understand how the protein binds to the ligands is key. The experiments conducted give an insight but further conclusive studies would direct experimental direction and targeted mutagenic studies.

5.4 METHODS

5.4.1. EXPRESSION AND PURIFICATION OF N PROTEIN.

A pET-16-b vector containing the ORF of N protein, obtained from R. P. Yeo, was transformed into *E. coli* BL21-DE3 and grown on ampicillin LB agar plates overnight at 37 °C. One colony was picked from the resulting plate and was incubated in 0.2 L of LB broth (Tryptone 10 g/L, sodium chloride 10 g/L, yeast extract 5 g/L; Melford Laboratories Ltd., Suffolk, UK, 25 g/L) overnight at 37 °C. 0.01 L of the starter culture was used to inoculate 1 L of LB Broth in a 2.5 L conical flask and the bacteria were grown in the presence of ampicillin (100 mg/mL) at 37 °C for 5 hours or until Optical Density at 600 nm reached 0.6-0.8. Expression was induced using isopropyl-β-D-1-thiogalactopyranoside (IPTG, 0.8 μM; Sigma Aldrich, UK) and left to shake overnight at 28 °C at 180 rpm. The bacterial cells were harvested by centrifugation at 4000 × g for 20 minutes in a Beckmann Coulter ™ J-20 XP centrifuge, using rotor JLA 8000 and the cell pellet was stored at -80 °C.

Protein extraction was achieved through resuspending the pelleted cells in 0.02 L of Lysis Buffer (50 mM Tris,1 M NaOH, 5 mM $CaCl_2$, 5 mM $MgCl_2$ pH 7.6). Lysozyme (1 mg/mL) and protease inhibitor tablet (cOmplete Mini, EDTA-Free, Roche Diagnostics Ltd, West Sussex, UK) were added once the pellet was resuspended. The cells were lysed by sonication for 240 seconds at 20 kHz and 3-[(3-cholamindopropyl)dimethylammonio]-1-propanesulphonate hydrate (CHAPS, final concentration 8 mM in ddH_2O) was added to the mixture and centrifuged at 50,000 × g for 30 minutes Beckmann Coulter ™ J-20 XP centrifuge, using rotor JA 25.50 to remove cellular debris.

The supernatant was then passed through a pre-equilibrated nickel affinity His-trap column (GE Healthcare Life Sciences, Buckinghamshire, UK). After the supernatant had passed through the column, Tris buffer (50 mM Tris, 300 mM NaCl pH 7.6) was run through, followed by an increasing concentration of imidazole (50 mM Tris, 300 mM NaCl, 50 mM to 1 M imidazole pH 7.6). The protein eluted at 200 – 500 mM imidazole and the fractions were analysed for N protein by SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis).

The protein was buffer exchanged overnight into Tris buffer (50 nM Tris, 200 nM NaCl, pH 7.3) or PBS (Gibco® PBS tablets, Life Technologies, Renfrew, Scotland) by dialysis (MWCO > 12,000 Da) at 4 °C. The His-tag was removed by cleavage with Factor Xa

overnight, as per the manufacturer's instructions (Novagen, Merck Chemicals Ltd., Nottingham, UK). The matrix protein was then stored and used at 4 °C.

5.4.2. SDS-PAGE (SODIUM DODECYL SULPHATE POLYACRYLAMIDE GEL ELECTROPHORESIS)

The purified protein was analysed using a 15 % acrylamide resolving gel: 5 mL ProtoFlow 30% acrylamide gel (Flowgen Bioscience Ltd., Nottingham, UK), 2.5 mL buffer (1.5 M Tris, 0.4 % (w/v) SDS pH 8.8), 2.3 mL water, 0.1 mL 10 % SDS (Aldrich, Poole, UK), 0.1 mL 10 % (w/v) ammonium persulphate (APS) solution in water and 0.004 mL *N,N,N',N'*-tetramethylethylenediamine (TEMED; Fluka). The stacking gel was made from: 1.7 mL ProtoFlow 30 % acrylamide gel (Flowgen), 2.5 mL buffer (0.5 M Tris, 0.4 % (w/v) SDS pH 6.8), 5.55 mL water, 0.1 mL 10 % SDS (Aldrich,), 0.1 mL 10 % (w/v) (APS) solution in water and 0.01 mL TEMED (Fluka). Protein samples were prepared by addition of protein loading buffer (2.5 mL 1 M Tris-HCl pH 6.8, 0.5 mL ddH_20, 1.0g SDS, 0.8 mL 0.1% bromophenol blue, 4 mL 100% glycerol, 2 mL 14.3 M β-mercaptoethanol (100% stock) and adjusted to 10 mL using ddH_20, 5 µL in each 15 µL sample) and heated to 95 °C for 5 minutes then loaded onto the gel with a protein ladder molecular weight marker in the first well (10 – 250 kDa; New England Biolabs., Hertfordshire, UK). The gel was run at 200 V for 1 hour and stained using InstantBlue™ (Expedeon, Cambridge, UK).

5.4.3. DETERMINING PROTEIN CONCENTRATION

5.4.3.1 NANODROP 1000

The concentration of the protein was determined using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Kent, UK) by measuring absorbance at 280 nm. The protein concentration was increased using a Vivaspin 6 (MWCO 10,000; GE Healthcare Life Sciences). Samples were centrifuged at 4000 × g in a Beckman Coulter™ Allegra™ X-22R centrifuge for 20 minutes at 4 °C, tested using the NanoDrop 1000 spectrophotometer and repeated until the desired concentration was achieved.

5.4.3.2 BRADFORD ASSAY

Protein concentration was also verified by Bradford assay. Standard concentrations between 5 and 100 μM of bovine serum albumin were prepared and allowed to equilibrate at room temperature for 30 minutes. Each sample contained the diluted protein sample at 800 μL and Bradford Reagent (Sigma Aldrich, UK) 200 μL. Each concentration was measured at 595 nm using UV-visable spectrophotometer (Multskan Go, Thermo Scientific, UK) and the absorbance was plotted against the concentration to give a linear plot. The protein was diluted and prepared in the same way and measured at 595 nm. The resulting absorbance was plotted on the standard graph with concentration was found on the x-axis.

5.4.4. THERMAL SHIFT ASSAYS (TSA)

The interaction of proteins over this work was analysed by thermal shift assays. Fluorescence data were collected on an Applied Biosystems 7500 FAST Real-Time PCR System with an excitation range of 455-485 nm. The fluorescence emission signal at 567-596 nm was used for data analysis. 8 μL of varying concentrations of RSV N protein (10 to 100 μM), 10 μl of various ligand concentrations (50 to 250 μM) and 2 μL of SYPRO Orange (5 × final concentration), were pipetted into a standard 96-well PCR plate (Starlabs Semi-Skirted FAST) to give final volumes of 20 μL in each well.

The temperature was held for 1 min per degree from 24 to 95 °C. *NAMI*, a Python programme written by M. Grøftehauge and N. Hajizadeh, was used to analyse the data that were collected. [95]

5.4.5. DYNAMIC LIGHT SCATTERING

A Malvern Zetasizer μV DLS system (Malvern Instruments Ltd., Worcestershire, UK) was used to conduct all DLS measurements in this work. The sample was held at 25 °C and allowed to equilibrate for 60 s prior to analysis. First, the DLS measurement of 100 μL of hRSV N/M was taken. Then, increasing concentrations (50 to 200 μM) of drug compound

were added, mixed and the DLS measured on each addition. Each size measurement was determined from the averaging of 13 runs of 10 s each.

5.4.6. TRANSMISSION ELECTRON MICROSCOPY

All images obtained via transmission electron microscopy using a modified negative staining protocol were with assistance of A.C. Richardson. A JEOL 2100F FEG TEM with a GATAN GIF tridiem with 4 megapixel Ultrascan™ 1000 CCD camera was used for imaging.

20 µL of a protein solution (either single or mixture) was pipetted on to a formvar-carbon coated 400-mesh copper grid, and left to dry for 30 s at room temperature. Any excess was removed with standard filter paper. The grid was negatively stained with 4% (w/v) aqueous uranyl acetate pH 4.0 for 30 s. The excess was removed from the grid, again with filter paper and allowed to dry briefly before imaging.

A standard TEM protocol for focussing and using the microscope was followed by A. C Richardson.

CHAPTER 6. IDENTIFICATION AND DOCKING OF DRUG LIKE MOLECULES INTO THE MATRIX PROTEIN OF HRSV.

This chapter will introduce the concept of drug discovery and design from the earliest stage through to biophysical testing. The techniques in this chapter were used to firstly identify potential binding sites in hRSV M using the CASTp server.[132] A suitable site was found and pocket parameters such as volume were computed. The pocket in hRSV M contains amino acid contacts in the dimer interface, if these were disrupted using ligands the structure of hRSV M would be compromised. To obtain the ligand molecules needed for docking, the ZINC server was used as a library. The ZINC server identified new potential drug-like molecules from a subset of over 1 million compounds to the final few hits. *In vitro* analysis was used to validate the drug-like molecules' activities against the target and will potentially lead the way for a new anti-viral candidate.

6.1 INTRODUCTION TO DRUG DISCOVERY AND DESIGN.

6.1.1 EARLY DEVELOPMENT.

The 20th Century dawned with a renewed interest in drugs and medical practice but it was not until the end of the Second World War and the introduction of antibiotics that drug discovery began its golden age.[133] The beginnings of rational drug design began in the 1950s where a theoretical understanding of the drug-target interaction and experimental drug testing could aid advancement in the field. The 1960s were the origin for modern antiviral therapies[134] and during this decade the first experimental antivirals were developed in the hope of treating the herpes virus. They used traditional trial and error methods where cultures of cells were infected with the target virus and the experimental compounds were introduced. This was very hit-or-miss and lacked the knowledge of how the virus worked so it was not very efficient. In the 1980s full genetic sequences of the target viruses began to emerge and virus structure and function could be examined in detail. This led to an increased understanding of the compounds needed to interrupt viral replication cycles, and its usefulness can be seen in the dozens of anti-viral treatments now available.[134]

6.1.2. MODERN PRACTICE.

New antiviral drug design follows a general idea of identifying the viral proteins, or parts of those proteins, which can be disabled by a future drug.[130,134] The proteins, and the smaller targets, should share as little homology to those found in the host as possible to lessen the potential side effects. If the target is also common across many strains of the virus or even within the same family of viruses this can be advantageous and would give the drug a wide effectiveness in its use.[130] Once the target is identified, drug-like compounds can begin to be isolated and tested from a larger subset, this process is shown in figures 6.1.2.1 and 6.1.2.2. The development pipeline below shows the stages from a target in the virus being defined to a drug being on the open market. There are two main stages in the pipeline, the pre-clinical stages coloured in red, are from molecular target discovery up to human clinical trials. The second stage is clinical, where the drug should prove its purpose, efficacy and safety before the human trial stage, and this is shown in blue in figure 6.1.2.2 [135]

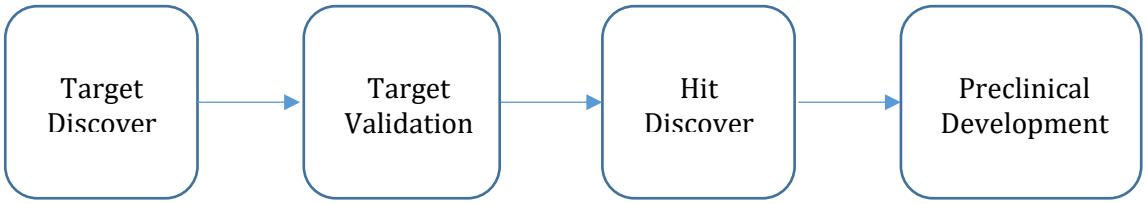


Figure 6.1.2.1 Preclinical stages of drug discovery.[136]

At the end of the preclinical development all data should be presented to the drug council of the country where the developments took place and the decision is made as to whether the drug will continue to the clinical stage. The financial investment required to trial a drug is large and only the top candidates will make it to the end. Data from testing *in vivo* using animal models and pharmacokinetic (PK) and pharmacodynamics (PD) data should also be presented at this stage and these will give some indication of the actions of the drug in the body and relevant dose. These are the precursor to clinical tests, and their results are used to justify entry into the clinic.[136]

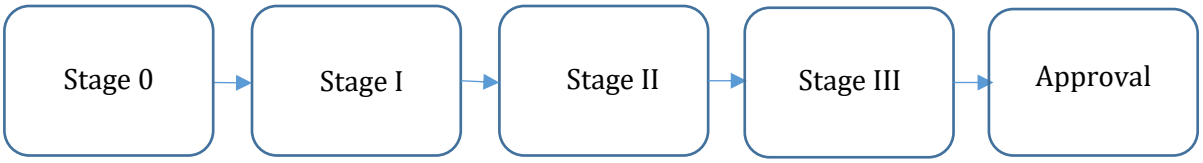


Figure 6.1.2.2 Clinical stages of drug discovery.[136]

If the drug passes pre-clinical trials it will enter stage 0, where the PK and PD data are tested in primates. Typically a small number of individuals (usually <10) are tested with sub Therapeutic doses and the oral bioavailability and half-life of the drug are determined.[134] Stage I tests the drug on healthy human volunteers to determine the dosage of the drug. The doses are often sub-therapeutic, but can rise; the patient number is larger in this stage.[130,135,136] Stage II moves the drug into patients so that efficacy and safety of the drug at a therapeutic dose can be known.[136] Until this stage the drug in question has not needed to prove its therapeutic effect on the condition. A typical sample size of ~200 patients is needed at this stage to ensure reproducibility and this is where the majority of drugs fail.[136] The reasons for failure can be as simple as the drug did not work as expected or showed toxic side effects.[137] Stage III continues on from stage II where a larger sample size (~1000 patients) is used alongside a physician to determine the full effect of the drug on the patients.[130,136] This is technically the last stage of testing and if the drug passes then it can be marketed to the public. Stage IV is often cited as post marketing surveillance where the long-term effects of the drugs can be monitored, if any problems arise at this stage the drug can be withdrawn.[136] Overall these pipelines take over 15 years and at the end the drug still may not be approved for use by government agencies.[137]

This lengthy process and the constraints involved have been a driving force for rational drug design. The technique combines X-ray crystallography along with the computational imaging and docking of a ligand. In comparison to high throughput screening and *in vivo* testing, the new technique affords many advantages which are detailed in the following chapter 6.1.3.

6.1.3 DOCKING AND VIRTUAL SCREENING

The increasing prevalence of X-ray protein crystallography during the 20th century has greatly increased our knowledge of protein structure which, coupled with increasing computer power, has paved the way for new rational structure-based drug design.[138] From the early beginnings of rigid docking through to the modern approach and improvements that can more accurately predict the biophysical state of binding, progress has been rapid and wide reaching.

6.1.3.1. EARLY DAYS – RIGID DOCKING AND EARLY TECHNIQUES.

Molecular docking is usually defined as solving the problem of identification of low-energy binding modes of a small molecule, or ligand, within the active site of a macromolecule, or receptor, whose structure is known.[139] To create a ligand the programme employed must have an accurate model of the molecular energetics to position the ligand, an efficient algorithm to search for the potential binding modes and the ability to show this in a defined way, usually on a screen through a graphical interface.[138,140]

The first definition of molecular surface was published by Lee and Richards[141], where they described how surface atoms can be distinguished from buried atoms and how the total surface of the molecule is the sum of the atomic surfaces.[141] Further work completed the theory and was published by Greer and Bush in 1978.[142] This method, which calculated the molecular surface and displayed the data as a contour map, was based on the proposition by Richards *et al*[143] in the previous year. This earlier method utilised water in the model and started to apply van der Waals interactions where water is excluded in the structure, mimicking protein dynamics in solution. The theoretical work paved the way for the field of Molecular Graphics to be developed.[144] The first automatic docking program, DOCK was developed through molecular graphics and uses a geometric approach with rigid docking. This approach explores the geometrically feasible alignments of ligands[139,145] in the known structure. Algorithms included in the

program examine binding geometries and evaluate the molecules in terms of steric overlap and thus give specific molecular conformations.[145]

Further developments through the 1980s began to include scoring functions, solvent surface calculations and flexibility into the docking model. Flexibility in docking was introduced in early docking programs such as FLOG[146] (Flexible Ligand Orientated on a Grid) and flexX[140].

6.1.3.2. DEVELOPMENT OF GOLD.

The GOLD (Genetic Optimisation for Ligand Docking) program was developed in the 1990s and uses a general algorithm[147]. It is an automated docking system which explores ligand conformational flexibility using partial flexibility of the protein and conforms to the requirement that the ligand must displace loosely bound water on binding. A general algorithm uses concatenated 8-byte strings, which represent torsions of rotatable bonds that undergo certain alterations to increase diversity.[147] The general algorithm collects data from the structure of each pose of the ligand and for each of these poses encodes a possible solution to the binding problem, these are called chromosomes. A pose of a ligand is where the same ligand is docked into the pocket, but a number of bonds are rotated or it can bind in multiple places, all within pre-selected limits.[148]

The docking uses pharmacophore points also known as fitting points which generate ligand poses. These points consist of donor-acceptor fitting points i.e. hydrogen bonds, and hydrophobic interactions.[138] The program requires the user to define an approximate binding site and location, this is pinpointed with a user-defined sphere, the radius of which is chosen to incorporate the relevant residues.[147] Each possible solution, namely, a ligand orientated within the protein binding site, is also assigned a fitness score based on the relative fit of the solution in the pocket.[147]

6.1.3.3. A GOLD RUN

Each run of GOLD has a number of operations, where GOLD uses the information of each number of rotatable bonds and the ligand flexibility to determine how many operations are required. A 100% efficiency of all possible ligands would produce a large set of poses, if there are multiple poses then each counts as an operation.[80] Search efficiency should be selected, 100% is ideal but unless the GOLD program and docking run is on a supercomputer this is unfeasible. A set search efficiency of 10% for the initial compound scout and 30% for a more in depth view, was used in this Chapter. This improves the speed of the search and although it lessens the number of answers, given the wide range, it should be sufficient to obtain a set of results which are viable. The docking run can be terminated early or run to completion, early termination will stop the run when the results reach a user defined number of solutions within a set RMSD as shown in Appendix B.1.[12,80] The general algorithm poses all possible solutions of the ligand in the binding site. A third option, diverse solution will end the docking run when a number of solutions within a certain cluster, a diverse subset, have been found.[81] These should be within a user defined RMSD and this method gives a set of poses which cover wider range of potential solutions but can take longer to reach.

6.1.3.4. SCORING FUNCTIONS.

Four fitness functions are available when docking; GoldScore, ChemScore, ASP and ChemPLP.[149] These differ in a number of ways, from what components are calculated through to which statistical model is used.[150] Firstly GoldScore which was one of the original scoring functions and designed when GOLD was first developed.[80] This fitness function calculates the ligand binding positions and derives a score. Hydrogen bond energies, van der Waals energy and torsion strain are all considered.[12] ChemScore, the second function available in this package, uses affinity to calculate fitness. This uses a subset of 82 complexes and compares the current docked ligand to known values. Hydrophobic contact areas, ligand flexibility and metal ion interactions are used in this fitness function.[12] ASP (Astex Statiscal Potential) is based on statistics. This is where the frequency of interaction between the ligand and protein atoms is calculated and compared against existing ligand-protein structure.[151] Finally, the fourth scoring function is ChemPLP. This is a combination of van der Waals forces and hydrogen bonds, which are found in ChemScore. In ChemPLP various models are also included which calculate repulsion terms and model distances.[148] There is no way of identifying which scoring function would give the most accurate answer. Comparison studies[149] have shown that scoring functions are far from being accurate but can give an overall approximation for ligand docking. Of the four scoring functions available, ChemScore is often cited as being the scoring function which gives better correlation of results.[152] In this Chapter, ChemScore will be used throughout.

6.2. MATRIX PROTEIN

As previously described in Chapter 1, the human Respiratory Syncytial Virus (hRSV) matrix protein was described as a monomer on crystallisation to a resolution of 1.6 Å.[48] As shown below in Figure 6.2.1 the protein N-terminal region is highlighted in magenta and the C-terminal in dark grey. The domains are joined by a 13-residue flexible linker region held together by hydrophobic interactions seen across related matrix proteins e.g. Ebola VP40 protein.[56]

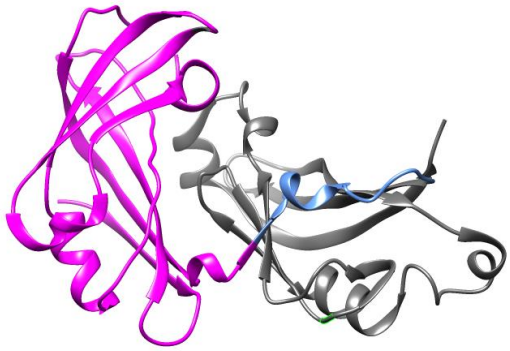


Figure 6.2.1. The crystal structure of hRSV matrix protein, colour coded N-Terminal domain in Magenta (1-126), C-Terminal domain in Dark grey (140-255) and a flexible linker region in blue (residues 127-139)[48] (PDB : 2VQP)

As discussed in Chapter 1 there are significant similarities in the proteins across the *Mononegavirales* family. Both hRSV M[48] and Ebola Virus VP40[51] were crystallised as monomers and share a similar global fold with an RMSD of 3.7 Å for C-alpha atoms. Other matrix proteins in the order of *Mononegavirales* have been crystallised but many are not similarly related to hRSV M in either sequence or structure. In 2014, the hRSV M structure was used to solve a closely related matrix protein in the same sub-family human metapnuemovirus (hMPV).[50] All work in this sub section was completed before the identification of the structure of the hRSV M dimer as discussed in Chapter 6.8.2.

6.2.1 MODELLING THE HRSV M DIMER.

Modelling of the hRSV M dimer and subsequent docking runs using HADDOCK[153] were performed in collaboration with N. J. Tatum, Durham University. The hMPV matrix crystal structure was used to guide the docking due to its proximity in structure to hRSV matrix.[48,50]

The hMPV matrix protein crystal structure was published in 2014;[50] and the phase problem was solved by molecular replacement using the hRSV M protein structure. Primary sequence identity between the two proteins is 37% when analysed through ExPASY BLAST[154]. The authors used small angle X-ray scattering (SAXS) to determine that the matrix protein of hMPV is a dimer in solution as well as in the crystal structure as shown below in 6.2.1.[50]

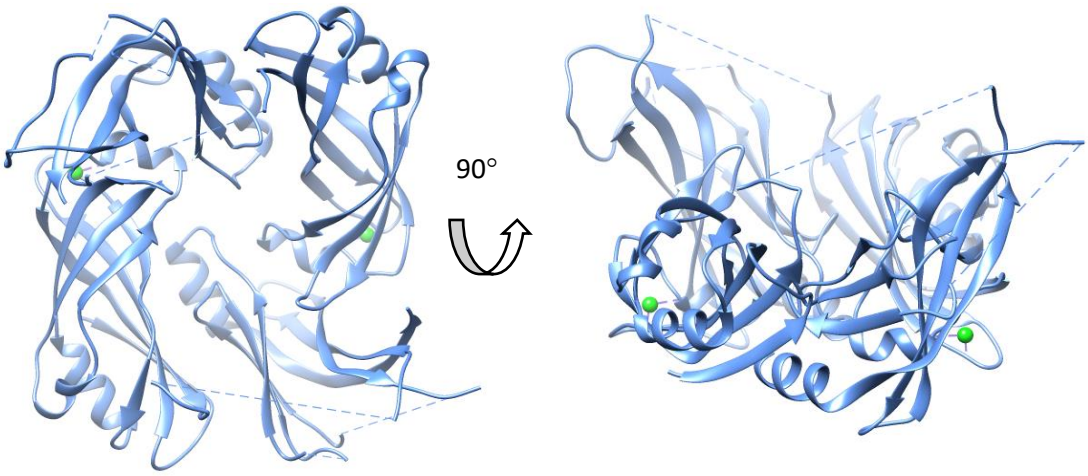


Figure 6.2.1.1. hMPV matrix protein dimer crystal structure. Calcium ions are shown in green.[50]

The dimer structure is stabilised by calcium ions, of which there is one per monomer this in contrast to the hRSV M monomer which was crystallised without any metal ions.[48] There are biophysical similarities between the proteins in solution as discussed in Chapters 3 and 4. That the hRSV M protein acts as a dimer in solution was confirmed by size exclusion chromatography and crosslinking where the protein leaves the column as a dimer of approximately 58 kDa. Also, previous work by J. Freeth[155] has shown by ICP-MS that in solution hRSV M does contain calcium ion.

6.2.2. DIMER FORMATION USING SYMMETRY-RELATED MOLECULES.

Using the hMPV matrix protein,[50] which crystallised as a dimer (pdb - 4LP7), the hRSV M protein monomer was flipped 180° and the two monomers placed together to form a dimer interface which resembles that of hMPV. PYMOL was used but it does not take into account hydrogen bonds or hydrophobic residues.[156] structure viewing software. The model was prepared by using the *symexp* command and a cut off for interactions at 3 Å and the resulting model is shown in figure 6.2.2.1. The parameters were chosen based on previous work by Arnold *et. al.*[157]This model was confirmed using PDBe PISA (v1.51) the interface of hRSV M was analysed and a similar symmetry operation [-*x*+2, *y*, -*z*] gives the same result.[158]

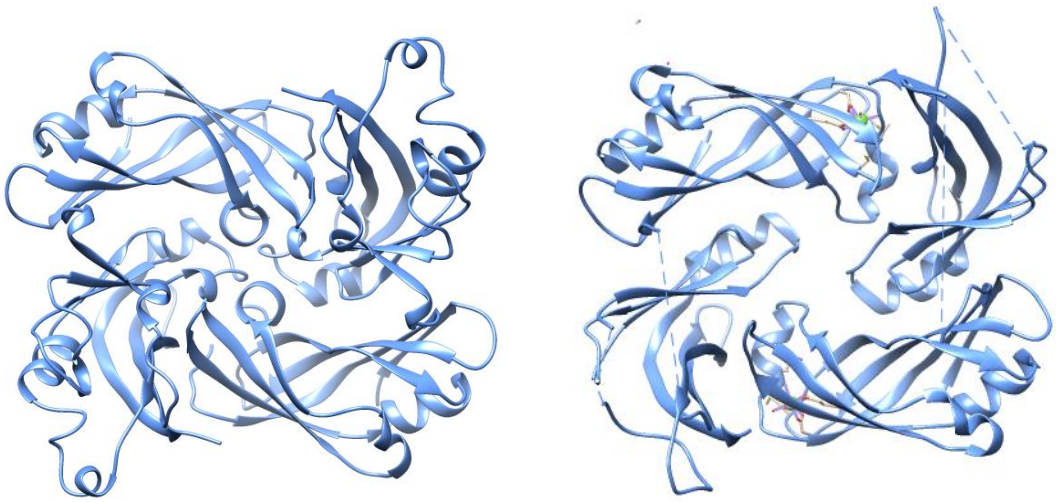


Figure 6.2.2.1. Left, RSV-M symmetry fitted dimer. Right, hMPV dimer (PDB – 4LP7)

A second method was used to form the hRSV M dimer, utilising the docking programme HADDOCK. Using the paper by Leyret *et al.*[50] where the hMPV M structure was first described, a number of residues were identified as forming the potential hRSV M dimer interface to help guide the protein-protein docking procedure in HADDOCK. To aid proper docking HADDOCK requires a number of residues to be pre-determined.[153] The molecular replacement of the hMPV M crystal structure was used to mimic the new

dimer. The resulting solution gave RMS distances of 3.157 Å for 1193 atoms in hMPV chain A, and 3.311 Å over 1287 atoms in chain C which is statistically closer so this model was used in further analysis.

It is noted in the supplementary information, that there is a hydrophobic motif *xWxPx* in the N-Terminal domain of hMPV M which is conserved across the families in the *Mononegavirales* Order including hRSV and Ebola.[49] In the hRSV M structure this corresponds to residues 34-38 (IWVPM). In addition, the hMPV dimer structure is stabilised by a large network of conserved hydrophobic interactions [50]. This gives rise to a buried surface area of 1421 $Å^2$ for each monomer and these residues were used to guide the docking. If all the residues on the buried surface area were used then this could bias the docking towards one solution, so three residues were chosen and explored independently: Trp35 from the *xWxPx* motif; Ser198; and Asp225 as highlighted in Figure 6.2.2.2. Passive residues, which are the solvent accessible surface neighbours to the active residues, were defined as within 6.5 Å of the active residues as default.

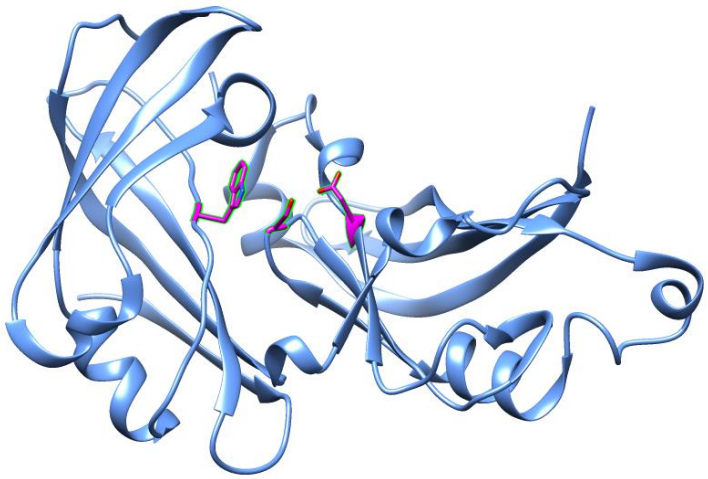


Figure 6.2.2.2 Active residues selected Trp25, Ser198 and Asp225 are shown as stick representation on the buried surface of hRSV matrix protein.[48] (PDB code 2VQP).

6.2.3. HADDOCK WEB INTERFACE

The manually fitted model of hRSV matrix protein was refined and the interface aligned to the interface of the hMPV matrix protein dimer using the HADDOCK webserver. HADDOCK (High Ambiguity Driven protein-protein DOCKing)[153] uses biochemical and/or biophysical interaction data to produce bioinformatics predictions. Firstly the protein structural data should be in a pdb format, then the residues of interest should be highlighted as shown in Figure 6.2.2.2 in Chapter 6.2.2, which will guide the docking. There should also be no duplicate residue or chain numbers as the server cannot compute the same number as different chains.[153] All water molecules and ligands were removed as HADDOCK assumes a water-mediated environment. The information given is used as Ambiguous Interaction Restraints (AIRs) to drive the docking process. The AIR is defined as an ambiguous distance between all residues shown to be involved in the reaction.[153] The server takes into account hydrogen-bond donors and acceptors, hydrophobic regions and charge properties to estimate a docking solution.

6.2.3.1. HADDOCK HRSV M DIMER MODEL

The overall result of a HADDOCK docking run can be shown as a maximum of 200 top scoring structures in clusters. The clusters were defined as a minimum of four potential dimer structures and each cluster at least 7.5 Å apart from each other by RMSD. This produces a vast range in the clusters. HADDOCK produces a graph to show the variation as shown in Figure 6.2.3.1.1 below.

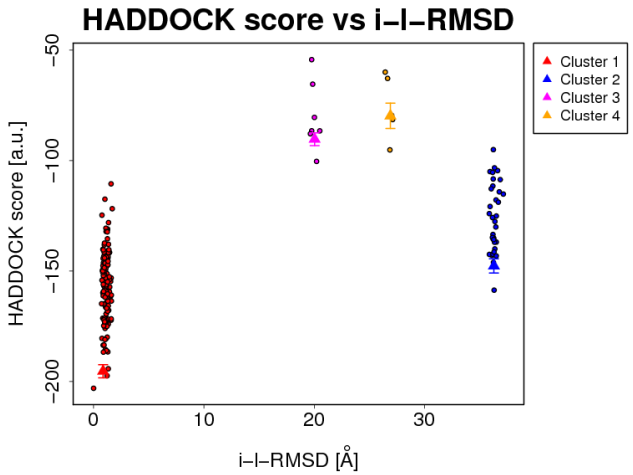


Figure 6.2.3.1.1. HADDOCK graph of score. The arbitrary units are plotted against the interface RMSD. The most negative solution is the best scoring. Interface RMSD is calculated on the backbone atoms of all residues involved in intermolecular contact using 10 Å cut off.

Clusters 2, 3 and 4 contained a total of forty six structures, all of these structures were a minimum of 10 Å different from the best scoring cluster, cluster 1. Clusters 2 and 4 did not contain any solutions, which were recognisable as similar to the hMPV structure as shown in Figure 6.2.3.1.2.

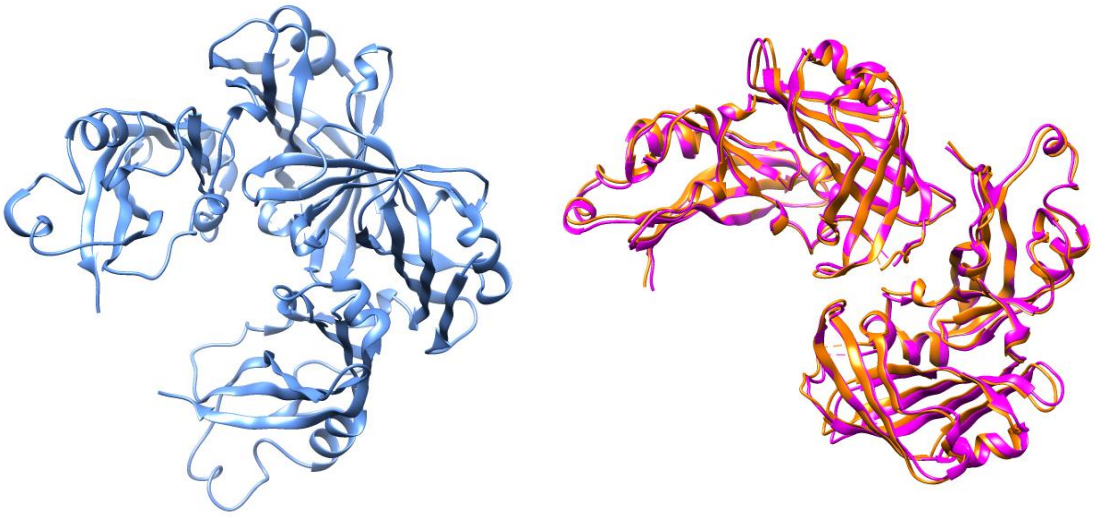


Figure 6.2.3.1.2. Top scoring models from Cluster 2 (left), Cluster 3 (orange) and Cluster 4 (magenta) (right) overlayed.

The top scoring cluster, which was cluster 1 contained 148 structures with an RMSD variation within the cluster of 0.5 ± 0.3 Å. This dimer is in the cluster 1 best scoring solution from HADDOCK and indicates strong convergence. All clusters and statistics of the solutions are shown in Appendix B.3

To verify the HADDOCK docking, ZDOCK [159] was used to independently dock the molecules, based on the active residues. The top 10 structures from a total of 500 structures, generated by ZDOCK, were analysed for comparisons to the HADDOCK results. The similarities between these 10 structures were clear to see and when structure 5 was superimposed onto the HADDOCK structure the RMSD (Root Mean Squared Deviation) difference was only ~1.3 Å. Overall the protein-protein docking from HADDOCK and ZDOCK respectively gave similar results, with cluster 1 being the highest scoring. The cluster 1 model, despite the restraint violation penalty for unfulfilled active residues, still scored above the others and when superimposed onto the best ZDOCK solution this showed a small RMSD and was taken forward.

6.2.3.2. THE HRSV M DIMER MODEL.

The most negative scoring model from HADDOCK was taken forward and is shown in Figure 6.2.3.2.1 below.

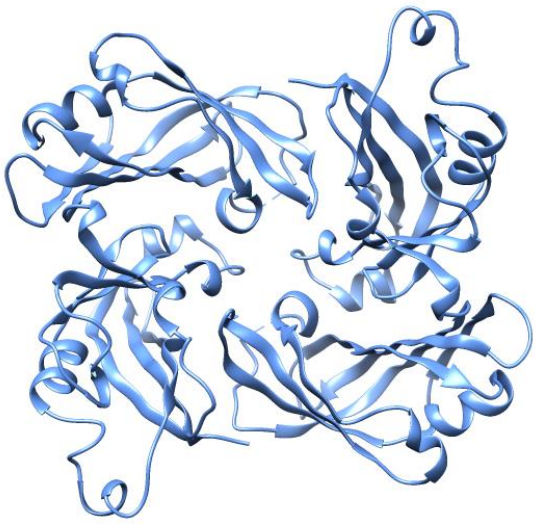


Figure 6.2.3.2.1. Dimer model of hRSV Matrix protein.

The overall fold and topology of the hRSV M and hMPV M proteins are very similar, with the N-terminal domains forming inter-domain hydrogen bonds with the C-terminal domain of the opposite chain. The overlaid structures are shown in Figure 6.2.3.2.2 below.

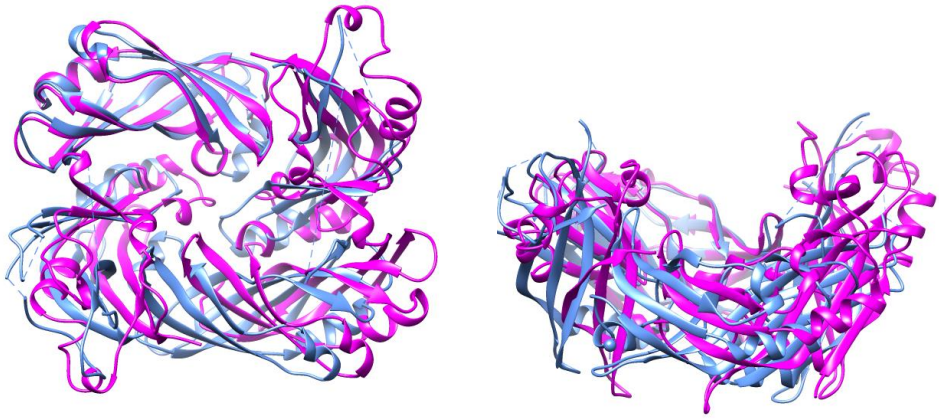


Figure 6.2.3.2.2 Overlaid structures showing the similar topology of the model hRSV M (magenta) and crystal structure of hMPV-M (blue) (pdb code: 4LP7)

The structures differ in one key region, in the loop around the hMPV-M calcium site at the N-terminus especially. This may be a consequence of hRSV M crystallising without calcium and hMPV-M crystallising with calcium and the loop being positioned to bind the metal ion rather than in hRSV M which does not bind calcium in the crystal structure. That being said the hRSV M N-terminal domain is shorter in length than the equivalent hMPV-M domain, which can be seen in the model overlay in Figure 6.2.3.2.3 below. Later comparison with the recently published hRSV M structure[79] supports this dimer model as being biologically relevant.

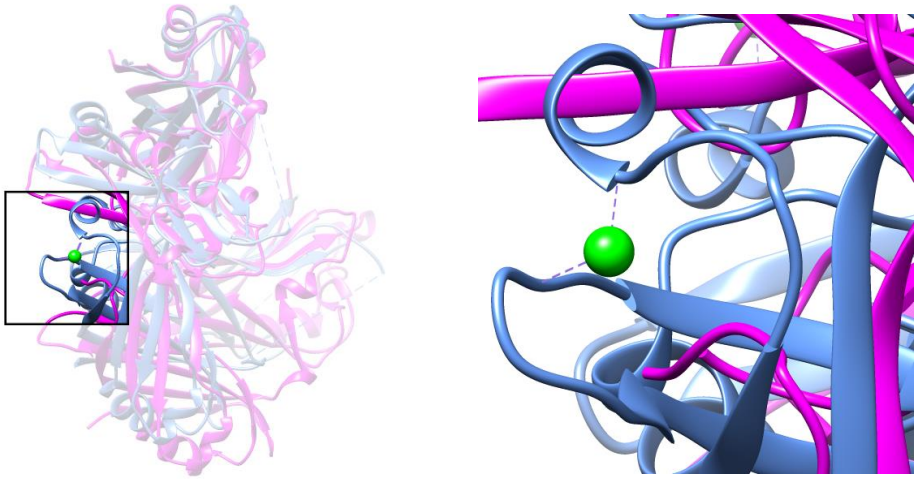


Figure 6.2.3.2.3. The low-affinity binding site in the hMPV-M structure shown in blue. (PDB code 4LP7) and the smaller loop in hRSV M in Magenta.

This is in contrast to the high affinity C-terminal calcium binding site in hMPV-M, the site is formed by residues Lys25 to Ala30. The C-terminal is solvent-exposed and is also unique across the *Paramyoxyviradae* order.[50] The residues involved in the calcium binding in the C-terminal region are conserved in both hRSV M and hMPV M but as the overlay shows the loop involved in hMPV M is in an open position and possibly active as shown in Figure 6.2.3.2.4. This is in contrast to the loop in hRSV M which is bent in the opposite direction.

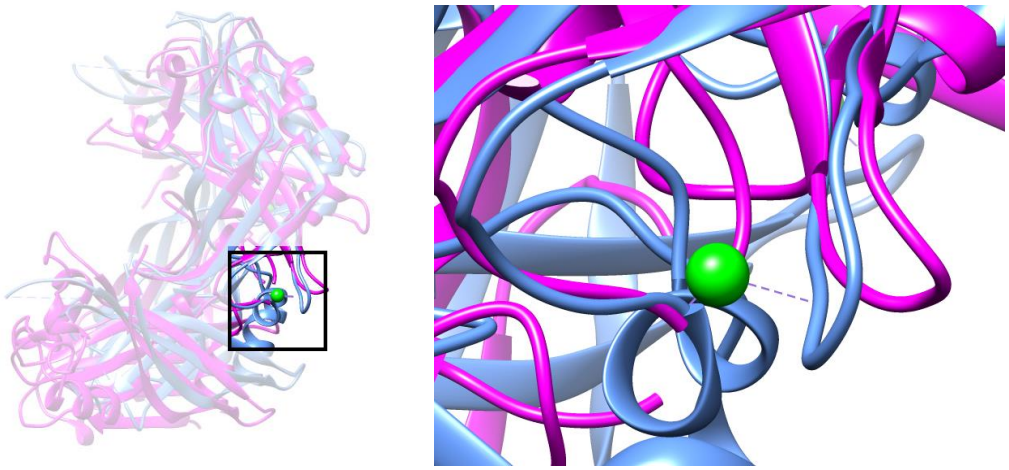


Figure 6.2.3.2.4 C-terminal calcium binding site, hMPV-M (blue) shows the loop in the active position, whereas the hRSV M (magenta) loop is angled away from the metal ion in a closed position.

Specific hydrogen bonds, which form the dimer interface can be identified using Chimera[160] shown below in Figure 6.2.3.2.5.

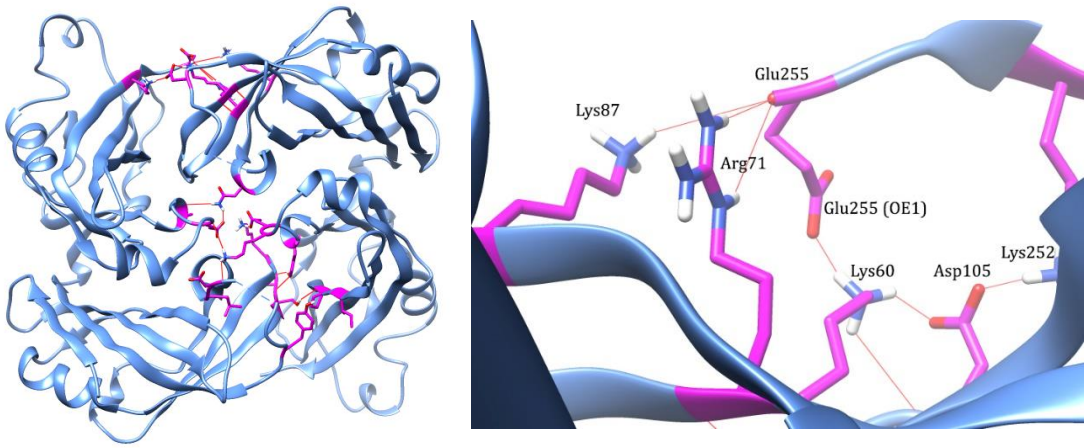


Figure 6.2.3.2.5 Interaction residues between the dimer interface of hRSV M.

Interacting residues in the dimer formation are detailed in Table 6.2.3.2.6. All values were calculated using Chimera.[160]

Chain A	**Chain B**	**Distance(Å)**
Ala 228 (Carbonyl O)	Asn 93 (ND1)	2.86
Glu231 (OE1)	Asn 93 (ND2)	2.67
Lys66 (NZ)	Glu 231 (OE1)	2.62
Glu 231 (Carbonyl O)	Lys 232 (NZ)	2.76
Asp 97(Carbonyl O)	Lys 232 (NZ)	2.94
Leu 96 (Carbonyl O)	Lys 232 (NZ)	2.93
Thr64 (Carbonyl O)	Tyr 229 (OH)	2.65
Ser 63 (OG)	Pro 161 (Carbonyl O)	2.71
Thr 104 (OH)	Ile 159 (Carbonyl O)	2.88
Lys 252 (NZ)	Asp 105 (OD2)	2.69
Glu 255 (OE1)	Lys 60 (NZ)	2.71
Glu 255 (OXT)	Arg71 (NE)	2.70
Glu 255 (OXT)	Arg 71 (NH2)	2.91
Glu 255 (Carbonyl O)	Lys 84 (NZ)	2.67

Table 6.2.3.2.6 Dimer interface residues.

6.3. FUSION PROTEIN.

The fusion protein (F) of hRSV is one of three viral surface glycoproteins.[68] The protein is essential for viral entry and is also highly conserved within the viral family.[43] F is a type I viral fusion protein, where a pre- and post-fusion state is seen. The protein is synthesised as an inactive single-chain polypeptide that assembles into active trimers which are able to bind to the target cell. Although the cytoplasmic tail of the protein is known to be necessary for filament and mature virion formation, the mechanism of this is unknown. A deletion of the terminal three residues or a mutation of the F572 residue has been found to halt filament formation, but not to affect F-protein expression or its trafficking to the cell surface. [7] The matrix protein (M) also has a critical role in viral filament formation.[11] Viruses which have been engineered without this protein, but still contain the other critical attachment proteins, have been found to produce shorter viral filaments and some did not form at all.[11]

6.3.1. FUSION PEPTIDE (NIAFSN)

The last 6 residues of the cytosolic tail of hRSV M are shown below. The peptide was previously identified through work by Shaikh *et al.* where the cytosolic tail was found to be important in the filament formation.[7] It has also been suggested that the residues interact with the matrix protein to produce mature filaments for viral assembly, budding and attachment.[79] The key residue identified by Shaikh *et. al.* was F572, a phenyalanine, which when mutated resulted in all activity being diminished.[7] The side chain of phenyalanine can now be used as a template for docking and drive the investigation into the interaction.

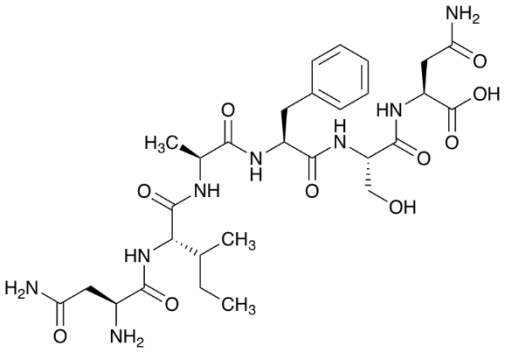


Figure 6.3.1.1. The six amino-acid peptide (NIAFSN) which represents the cytoplasmic tail of F protein in hRSV.

6.4. POCKET PICKING ON HRSV MATRIX DIMER USING CASTP.[161]

Active sites and binding sites of structural proteins are often found in pockets or cavities. The CASTp (Computed Atlas of Surface Topography of proteins) server[161] identifies multiple factors such as area and volume of the pocket and cavity. The server identifies surface accessible pockets as well as interior inaccessible cavities and measures the size of each pocket and mouth opening.[132,161] Further analysis can include solvent accessible surface calculations, mouth openings and molecular surface.

The CASTp server requires a pdb file for analysis, once it has been loaded and run through the program the user is able to manipulate the protein to identify which pocket is key, if that is not known. All the pockets available on the dimer were examined and there were three that were large enough to be ligand binding sites as shown below in Figure 6.4.1.

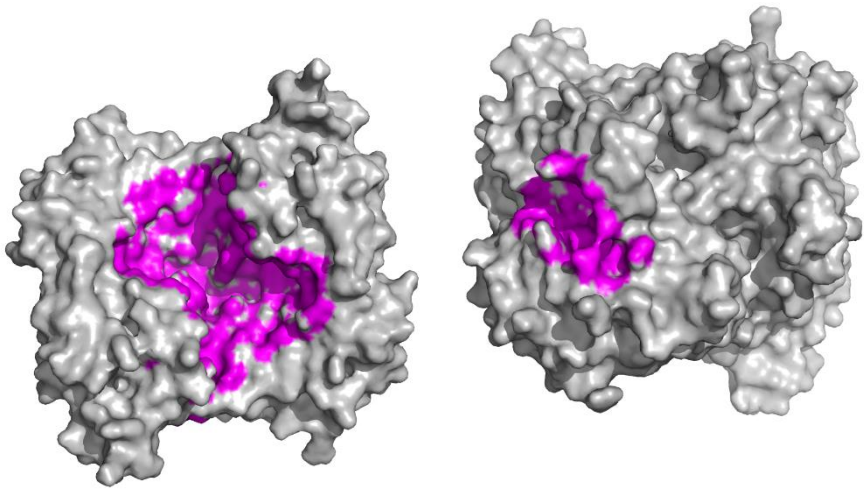


Figure 6.4.1. Major pocket 74 (left) in the cleft of the protein dimer and pocket 73 (right) shown in from the CASTp server.[161]

The major pocket shown above is approximately 8600 Å^3 in volume, which is larger than required for small ligand docking. This pocket may bind ligands, but it would be too promiscuous for our purpose of drug like compound binding.[162] The second pocket highlighted was approximately 1100 Å^3 in volume, as shown below, which was again larger than recommended for drug-like compounds under 500 Dalton molecular weight.[163] The third pocket is smaller, approximately 450 Å^3 in volume and located on

the dimer interface as shown in Figure 6.4.3. If a ligand could bind in this pocket there is the potential for disrupting the dimer interface.[49] Many of the interface residues identified in Table 6.2.3.2.6. are present in this pocket as shown in Figure 6.4.3.

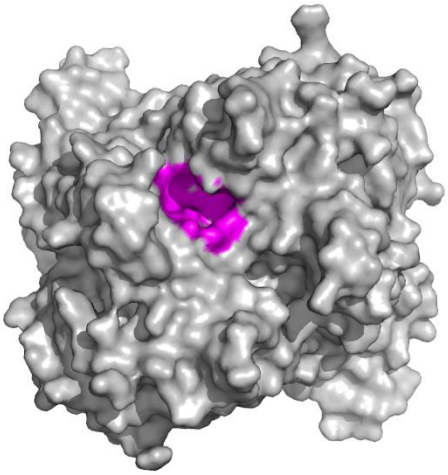


Figure 6.4.2. Pocket 72 - Dimer interface pocket.

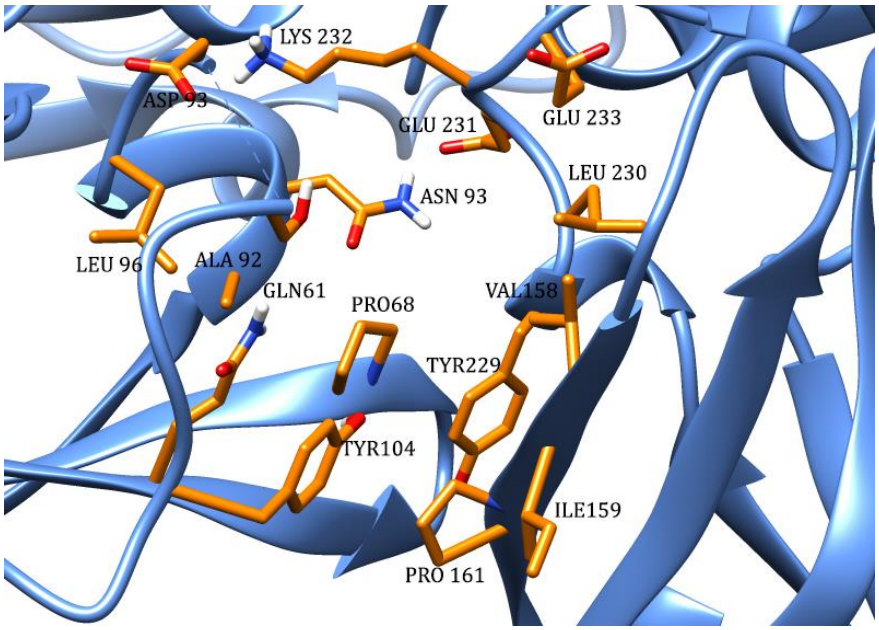


Figure 6.4.3 Pocket 72 key residues shown in Chimera.[160]

6.5 ZINC DATABASE

To determine drug-like molecules for docking studies the ZINC database[164] was used to search a subset of chemical compounds. ZINC is a commercially available subset of biologically relevant chemical compounds, which are available in a 3D format. In order to be useful for research the subsets are as diverse as possible and in biologically applicable forms.[164] The ZINC database is not unique and there is significant overlap with other well known systems i.e. ChEMBL,[165] PubChem[166] and DrugBank,[167] but ZINC is superior for our purposes due to the commercially available nature of the compounds that are included.

Millions of compounds are divided into smaller subsets; drug-like, fragment-like and lead-like are particularly important subsets, and these were tested in this work. Lead-like compounds are usually used for assays *in vivo* where binding is not directly observed and they tend to be in the higher molecular weight range.[164] The fragment-like subset are ideal for X-ray crystallography and surface plasmon resonance experiments, because they tend to be lower in molecular weight. The drug-like subset is an amalgamation of historically important fragments and compounds which obey the rule of 5,[168] but could miss out important compounds by being too similar in basic structure. The aim of the initial screen is to test a large number of diverse compounds. The fragment-like subset contained a large number of aromatic compounds, which was required to mimic the essential phenylalanine on the F peptide chain. Due to limited computing power only drug-like compounds at pH 7 were taken, which totalled 1 385 525 compounds, where these compounds were the most diverse in structure whilst still retaining the key benzene group. In later rounds, to obtain comparable data and to determine if other scaffolds could be sorted from the other subsets, the lead-like compounds were also tested.

6.6 KNIME

The aim of this program is to filter by structure, molecular weight, volume and many other parameters. It is particularly important to screen the large number of compounds in the drug-like subset obtained from the ZINC server. Knime[169] is a data-mining platform, which offers users an interface to assemble filters, known as nodes, for data reprocessing. Fragment-like and lead-like subsets were screened by the author and H. Yamada, the outcome of the subsets were investigated using the same parameters. Over 1 000 000 compounds were taken from the pH 7 fragment-like subset, to dock all of these in GOLD would only be possible on a super computer so Knime is used to process and separate the compounds into user-defined sections. From previous work[7], it is known that the F protein phenylalanine side chain (F572) is critical for binding to the matrix protein. This structure was the core of the search and a structure filter was used which constrains the subset to contain a benzene ring. A cyclohexane was decided against due to the unknown nature of the pi-pi stacking which could occur with the central tyrosine residues 104 and 229 in the pocket. Additional filters were added which were based on Lipinski's rule of 5.[168] Lipinski's rule of 5 predicts that poor absorption of the drug is likely if: there are more that 5 hydrogen bond donors; more than 10 hydrogen bond acceptors; the molecular weight is larger than 500 Da; and the calculated log P is greater than 5.[168] These factors were all used in node constraints in Knime. Other constraints that were used included the overall volume of the molecules, which was set to a maximum of 470 $Å^3$, which is equivalent to the size of the proposed pocket and also there must be an aromatic group included. This reduced the subset down to 879 173 compounds. To further reduce the number of potential compounds an additional set of more stringent rules were imposed on the subset. Further nodes were used which complied with the fragment rule of 3.[163] This took Lipinski's rule of 5 and reduced all parameters to 3 if the subset was fragment-like. The rule of three was devised by researchers[163] who found the most drug-like ligands were found under 300 Da but does not compromise the diversity of the ligands found. The parameters included: fewer than three rotatable bonds and hydrogen-bond donors and acceptors; and <300 in molecular weight. The subset was reduced to 137 498.

This was still too large to take through to GOLD docking and so was reduced further by reducing the XlogP to <3 to fit with the fragment rules. A lower limit to the volume of compounds was also set to rule out small promiscuous molecules. This dramatically reduced the number of compounds to 43 820. This was deemed a reasonable number to take through to GOLD. The full pipeline for fragment–like compounds is shown below in Appendix B.2.

6.7 GOLD

6.7.1. DEFINING THE BINDING POCKET.

As described previously in 6.1.3.3 a GOLD run can firstly be thought of as defining the binding site. In the PYMOL viewer[156] the residues, which are in the middle of the pocket and also form part of the dimer interface can be determined as Tyr 229 and Asn 93. The ball seen on screen should be placed in the centre of the binding pocket and the approximate coordinates taken, in this case these were X = 4.048 Å, Y = 6.036 Å and Z = -8.303 Å as shown below in Figure 6.7.1.1.

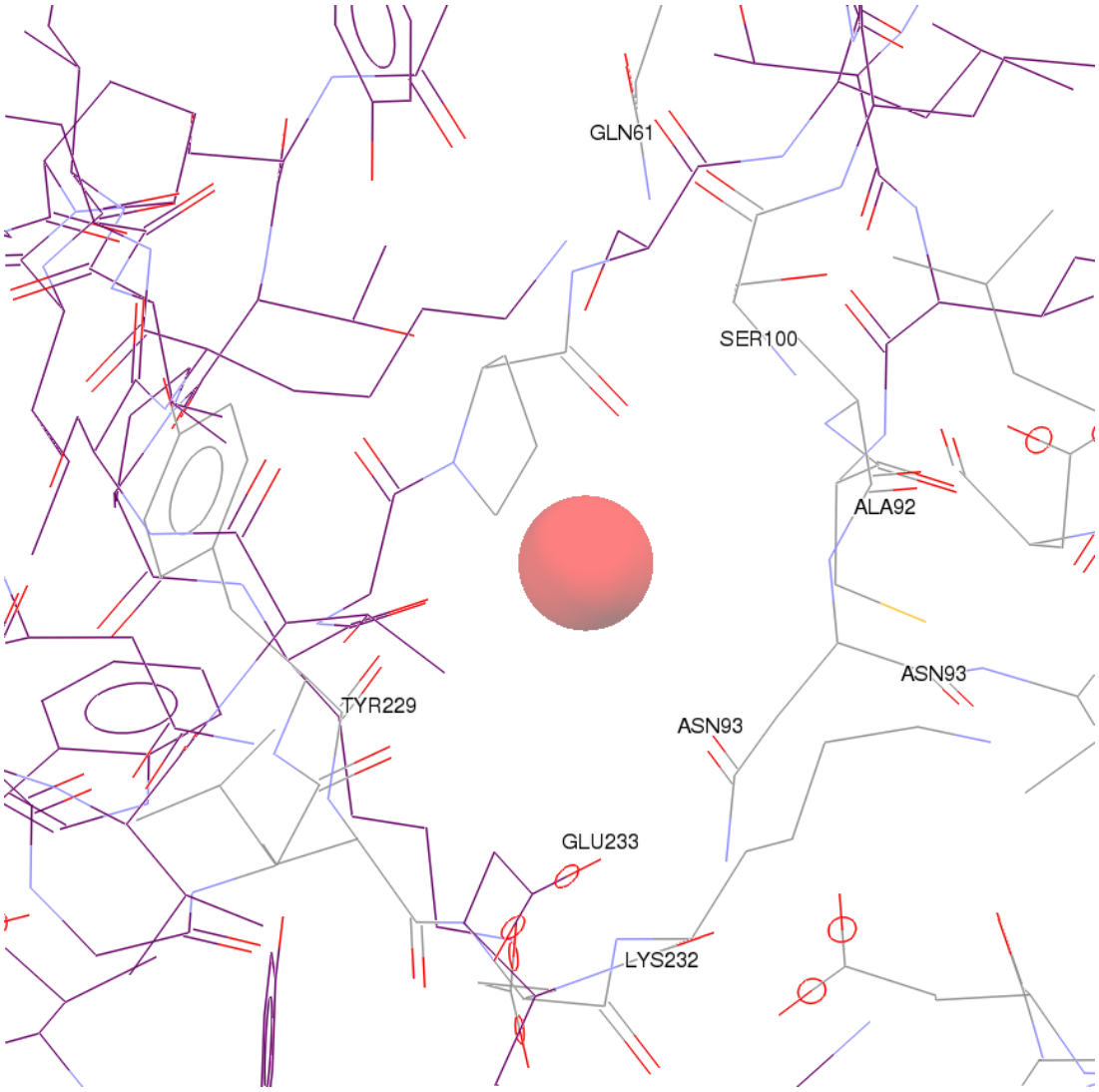


Figure 6.7.1.1 Binding pocket, highlighted with key residues and important binding residues.[150]

The GOLD run was customised so that the number of poses that each ligand was screened in was reduced to compensate for computing power. The RMSD (root mean square deviation) was also reduced to 1 Å to obtain a subset of molecules that have small differences in structure and binding, but of the same central binding pocket. Search efficiency in an ideal world would be 100% where every pose could be analysed by the program, docked into place and a fitness score awarded. 30% is considered to be a standard fragment screen and is run as a compromise to give a general screen.[12] An additional option was used to allow ligand flexibility, where amine bonds and rings can rotate and rings can twist into conformations required in the binding site. The total number of ligands tested was 131 460 (43 820 in 3 separate poses).

6.7.2. GOLDMINE

Once the run was complete the results were filtered and imaged using GOLDMine,[150] a post processing and analysis software written by the developers of GOLD. Scoring functions from GOLD alone cannot be relied upon and some docked ligands may have extreme torsion angles or form no hydrogen bonds to the protein. GOLDMine can be used to filter out undesirable ligands using histograms where outliers are shown in red and the viable data are shown in blue.[81]

Firstly hydrogen bond formation was evaluated; the ligand should form hydrogen bonds as stated by Lipinski's rule of 5. This stated that there should be no more than 5 hydrogen donors and no more than 10 hydrogen bond acceptors.[168] Hydrogen bonds, from H---X-H typically prefer to form at 180° at the lowest energy conformation. GOLDMine automatically sets the minimum angle at 90° so any angles within this constraint are shown, but this does not necessarily mean they are the best fit in the binding pocket. Using histograms the outliers can be easily identified and eliminated. Torsion and clash can also be constrained to reduce outliers. The Gold PLP Ligand Clash is another scoring function, which shows undesirable ligand-protein interactions and awards penalties, any high values were eliminated to ensure all unfavourable torsion angles and bond clashes are removed.[148]

From this set the top scoring 10% were filtered and analysed, which yielded 62 hits in 61 structures, where one ligand was posed in the pocket twice. Using CCDC Software –

Mogul[170], a geometry checker can be used to check the compounds based on a library of molecular geometry from the Cambridge Structural Database. The library contains millions of bond lengths, ring conformations and torsion angles, which can be used to validate the structure.[170]

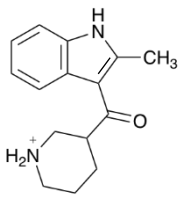
After checking manually, 22 compounds were noted as acceptable and the top 20% were chosen on the basis of fitness score were taken. The compounds were then cross-referenced back into the ZINC database for purchasing information. Even though the ZINC database is labelled as commercially available, sometimes the compounds are specially made to order or maybe out of stock at that time, this further limited the number of compounds which were available for *in vitro* testing shown in Chapter 6.8.

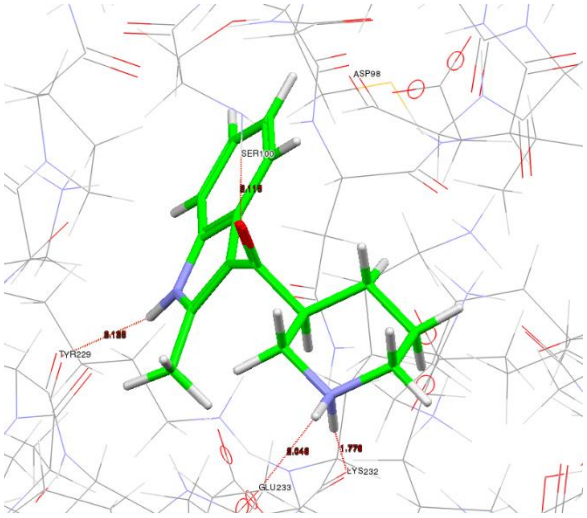
6.8 COMPOUND RESULTS

6.8.1 DOCKED COMPOUNDS USING THE MODEL.

The top 20% of the 22 compounds, in numerical order of the highest fitness score are detailed below, showing the chemical structure, binding pose in the pocket and theoretical hydrogen bonding residues in the protein.

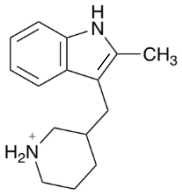
1. Compound 1. (2-methyl-1H-indol-3yl)-[(3s)-3-piperidyl] methanone

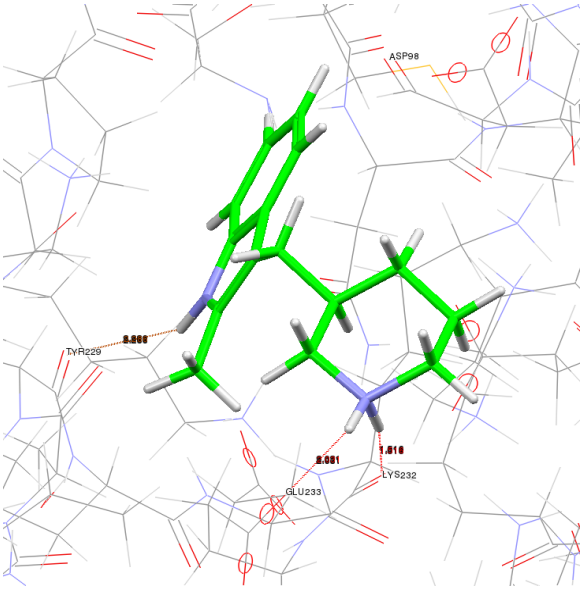




Protein Residues	Ligand group	Distance/ Å
Tyr229	NH	2.13
Ser100	C=O	2.11
Lys232	NH	1.77
Glu233	NH	2.05

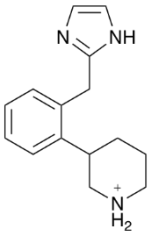
2. Compound 2. (2-methyl-1H-indol-3yl)-[(3s)-3-piperidyl] methane

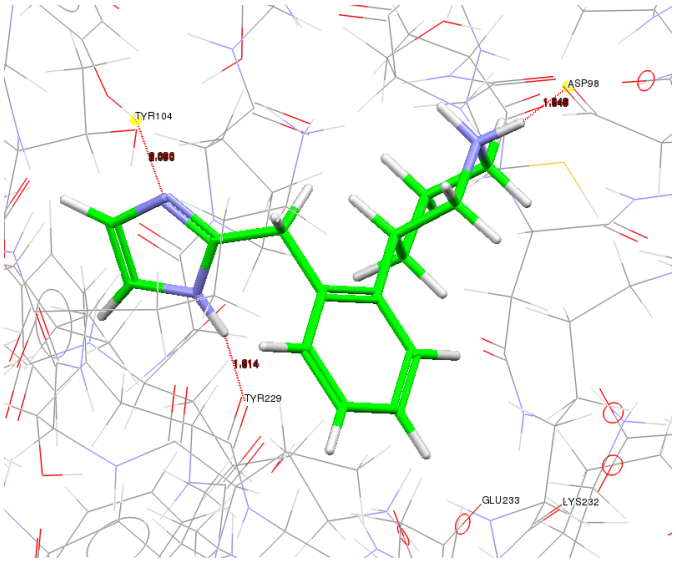




Protein Residues	Ligand group	Distance/ Å
Tyr229	NH	2.28
Lys232	NH	1.51
Glu233	NH	2.03

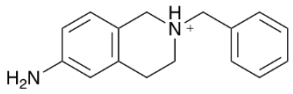
3. Compound 6. 1-(imidazole-1-yl methyl)- [(2s)-2-piperidyl]benzene

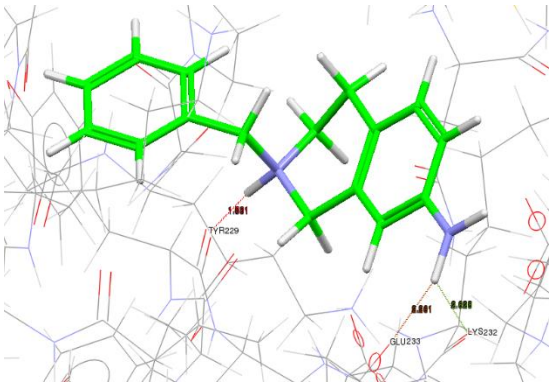




Protein Residues	Ligand group	Distance/ Å
Tyr229	NH (Imidazole)	1.81
Tyr104	N (Imidazole)	2.10
Asp98	NH	1.85

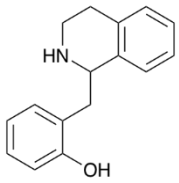
4. Compound 15. 2-benzyl-3,4-dihydro-1H-isoquinolin-7-amine

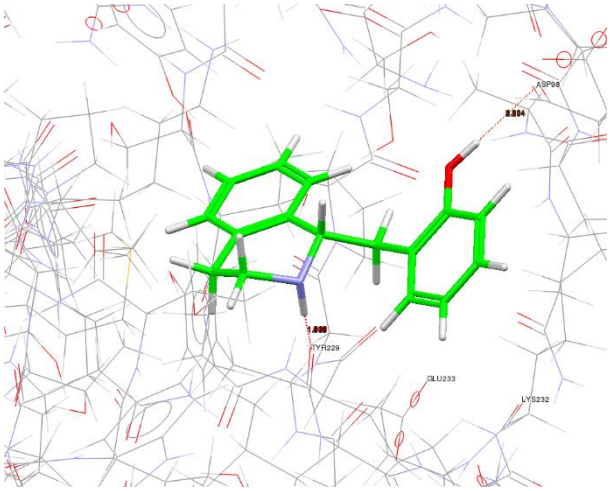




Protein Residues	Ligand group	Distance /Å
Tyr229	N-H (isoindol)	1.53
Glu233	N-H	2.26
Lys232	N-H	2.63

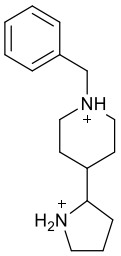
5. Compound 33. (1*S*)-1-[(2-methoxyphenyl)methyl]-1,2,3,4-tetrahydroisoquinoline.

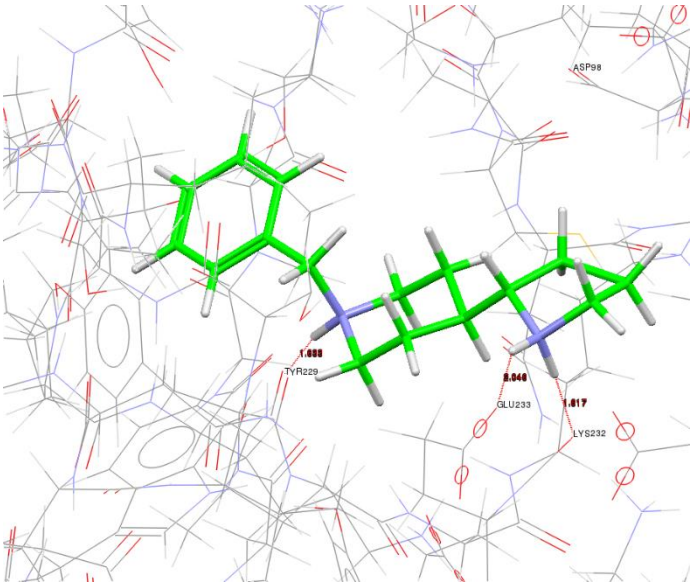




Protein residues	Ligand group	Distance /Å
Tyr229	N-H	1.99
Asp93	O-H	2.23

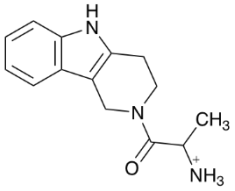
6. Compound 45. 1-benzyl-4-(pyrrdidin-zyl)piperidine

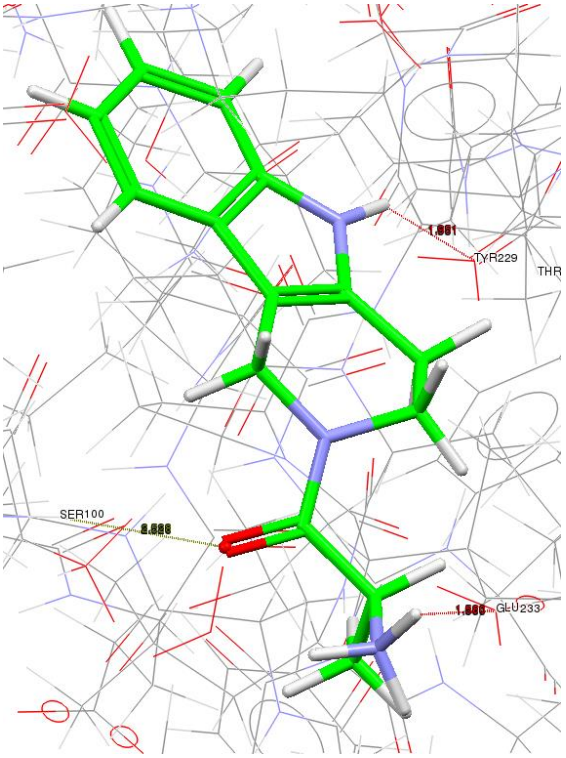




Protein Residues	Ligand group	Distance/ Å
Tyr229	NH	1.63
Lys232	NH	1.62
Glu233	NH	2.04

5. Compound 47. (2*S*-)-2-amino-1-(1,3,4,5-tetrahydropyrido[4,3-6]indol-2-yl]propan-1-one.





Protein Residues	Ligand Group	Distance Å
Tyr229	NH	1.633
Glu233	NH_3	2.04
Lys232	C=O	1.61

Many of the compounds show theoretical hydrogen bonds in the docking which are short compared to the average length from 2.8 Å to approximately 3.1 Å.[171] All bond lengths are approximate and it is believed that the flexibility of the protein and ligand may lengthen these in solution. Also the protein docking was conducted on a dimer model, this is an estimation and the values may change.

6.8.2 FURTHER DOCKING

6.8.2.1. NEW MATRIX DIMER.

In February 2015 a new paper which identifies hRSV M as a dimer was published,[79] showing the matrix protein as a dimer where no calcium ion was found in the structure as shown below. Using Pymol, the structures as shown in Figure 6.8.2.1.1. were overlaid using 1598 atoms to a RMS of 0.986. It can be seen that model was very close to the actual dimer structure, around the binding pocket and dimer interface the models are very similar.

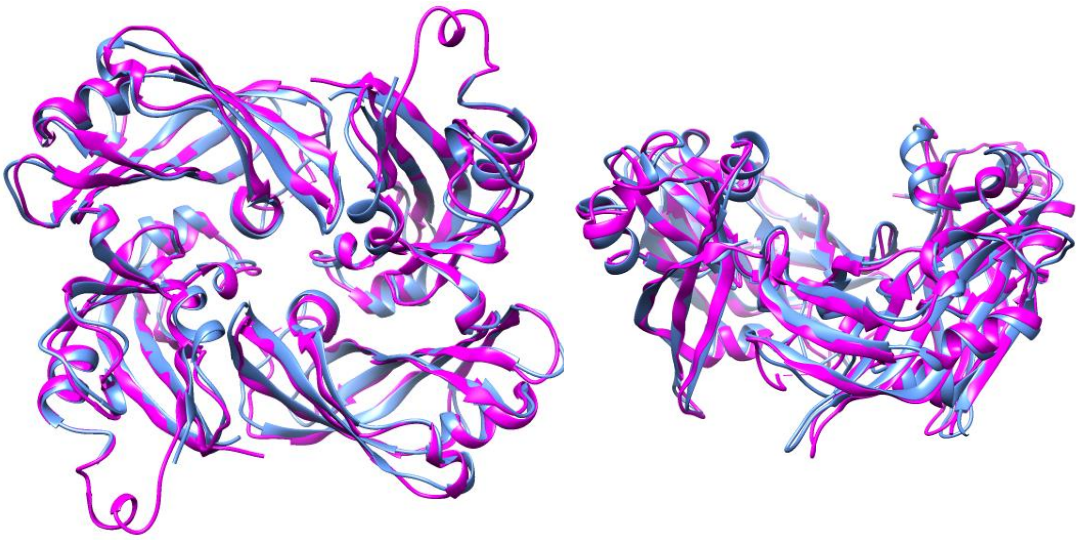


Figure 6.8.2.1.1 Left: The new hRSV matrix protein dimer (PDB code:4D4T) overlaid with the model interaction dimer M. Right: Overlaid structures rotated 90 °.

The differences seen in the loop region are where a region of electron density was undefined in the 2009 structure.[48] This was fully defined in the 2015 dimer structure, but it is far enough away to not change the pocket residues or distances. Even so, it was thought beneficial to re-dock the top 20% of the results from the 10% and 30% screen of fragment-like compounds to ensure a complete experiment. The compounds were saved in a larger sub-set file in KNIME, so each had to be filtered out by hand again, this was completed by a structure search as shown in Figure 6.8.2.1.2. This filtering was also completed for the 10% fragment-like screen, which was performed previously, and as a result 20 compounds were taken through for redocking. The file parameters are detailed in Appendix B.3.

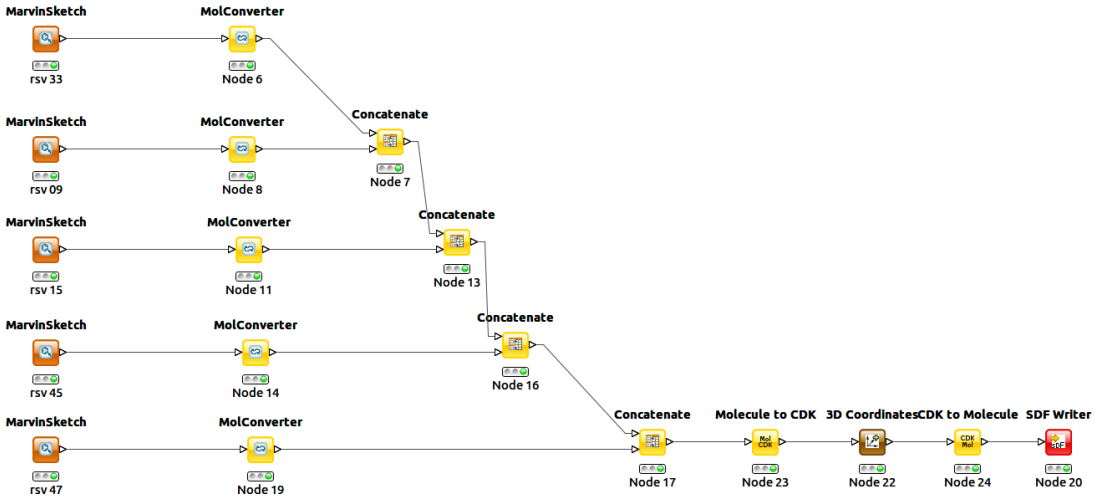
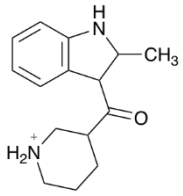


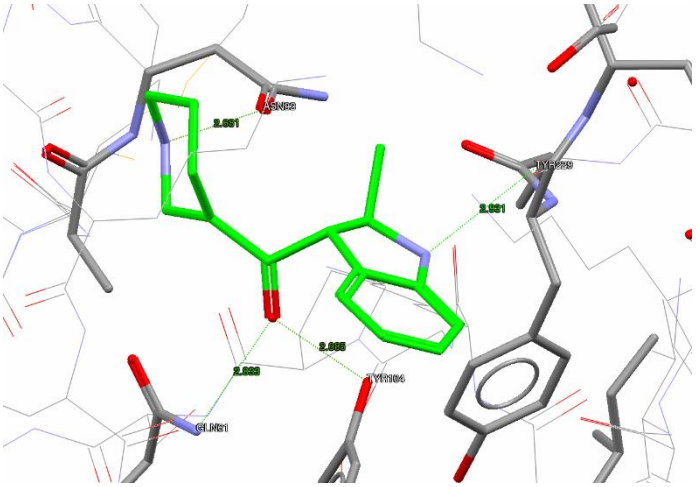
Figure 6.8.2.1.2 KNIME flow pathway for isolating specific ligands.

6.8.2.2 DOCKING COMPOUNDS USING HRSV M DIMER.

The re-docked results of the top 5 compounds from the 30% and 10% screen are shown below, ranked in order of scoring. Hydrogens are not explicitly shown but are the same as Chapter 6.8.1.

1. Compound 9. (2-methyl-1H-indol-3-yl)-[(3s)-3-piperidyl] methanone

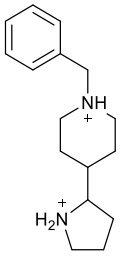


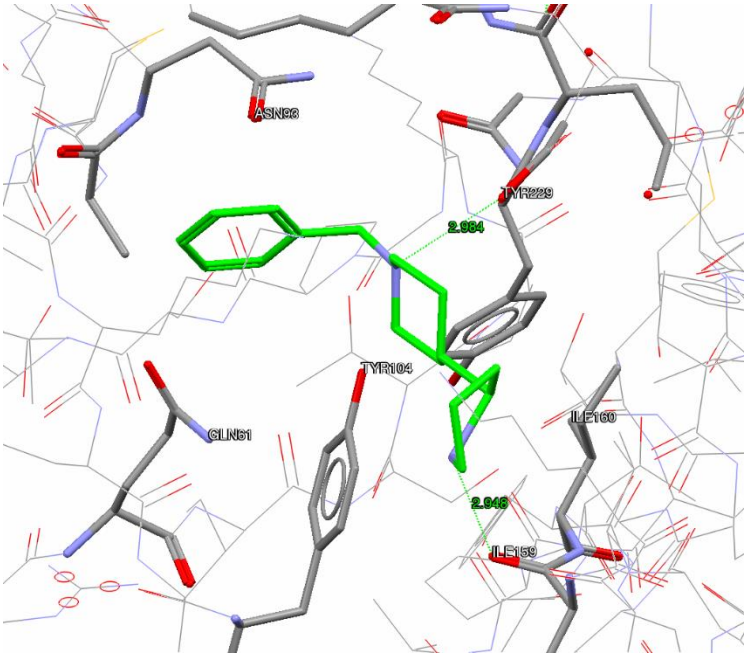


Protein Residues	Ligand group	Distance/ Å
Tyr229	Backbone NH	2.931
Gln61	C=O	2.833
Tyr104	OH	2.685

Compound 9 was found to be the highest scoring in both protein models. It was found that the same Tyr229 residue was identified in this structure as it was in our model. This residue was also found to be key in forming the dimer interface as described by Förster *et al.* and disrupting this was the aim of the compound.

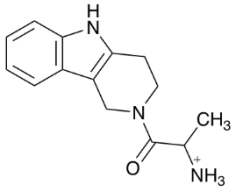
2. Compound 45. 1-benzyl-4-(pyrrdidin-zyl)piperidine

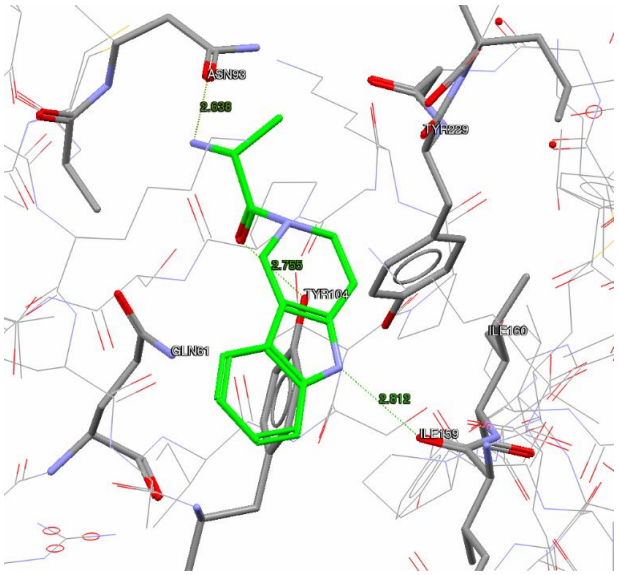




Protein Resiudues	Ligand group	Distance/ Å
Tyr229	Backbone NH	2.994
Ile259	Backbone OH	2.948

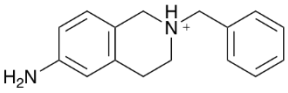
3. Compound 47a – (2S)-2-amino-1-(1,2,4,5-tetrahydropyrido[4,3-b]indol-2-yl)propan-1-one.

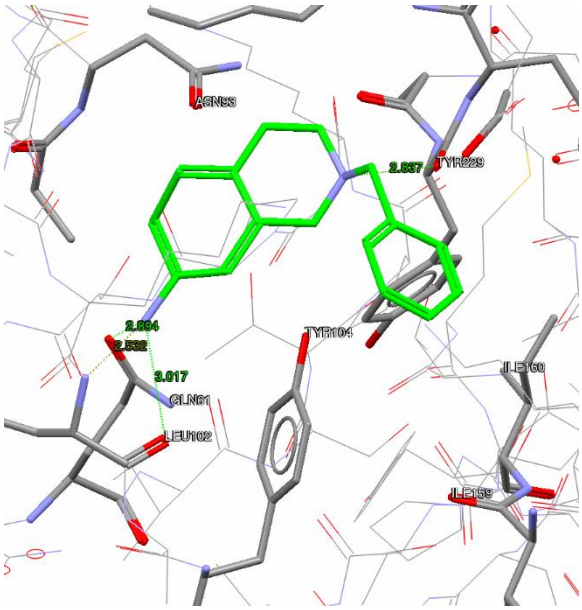




Protein Residues	Ligand group	Distance/ Å
Tyr104	Backbone NH	2.755
Ile159	Backbone OH	2.811
Asn93	NH	2.638

5. Compound 15. 2-benzyl-3,4-dihydro-1H-isoquinolin-7-amine





Protein Residues	Ligand group	Distance/ Å
Tyr229	Backbone NH	2.837
Gln61	OH	2..894
Ile102	Backbone NH	2.63

All the lengths shown are more indicative of the traditional hydrogen bond when the hRSV M dimer was used instead of the model.

6.8.2.3 MOGUL

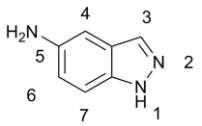
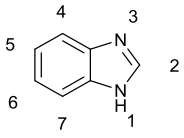
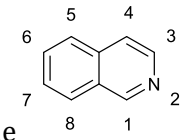
Mogul,[170] as described previously, was used to check the geometry of the bonds and angles in the compounds above. The program checks the geometry against the CSD of known compounds to identify clashes or out of range torsion angles. All compounds tested were within the recommended ranges for each bond type.

6.8.3 COMPOUNDS AVAILABLE FOR PURCHASE

Compounds 9, 15 and 47 were available for immediate purchase through the ZINC database, so these were taken forward for further testing. Only compound 9 arrived in time for initial testing.

6.8.3.1 COMPOUNDS AVAILABLE IN HOUSE

A number of compounds that contain at least 2 aromatic rings and have a similar backbone to the compounds from the KNIME subset were available from J.M. Sanderson, Durham University. The range of compounds is detailed below in Table 6.8.3.1.1

Compound group	Compound Number	Molecular weight
Indazole 	133 a) 3-aminoindazole b) 4-aminoindazole c) 5-aminoindazole d) 6-aminoindazole e) 7-aminoindazole	133.15
Benzimidazole 	133 f) 2-aminobenzimidazole h) 5-aminobendimidazole	133.15
Isoquinoline 	144 a) 3-amino isoquinoline b) 4-aminoisoquinoline c) 5-aminoisoquinoline	144.17

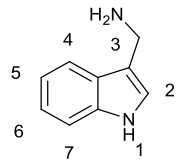
 Indole	146 d) 2-(aminomethy)indole e) (1H-indol-3-yl)methanamine f) 5-aminomethylindole g) 6-aminomethylindole	146.19

Table 6.8.3.1.1 New compounds for testing.

6.9 PRELIMINARY RESULTS

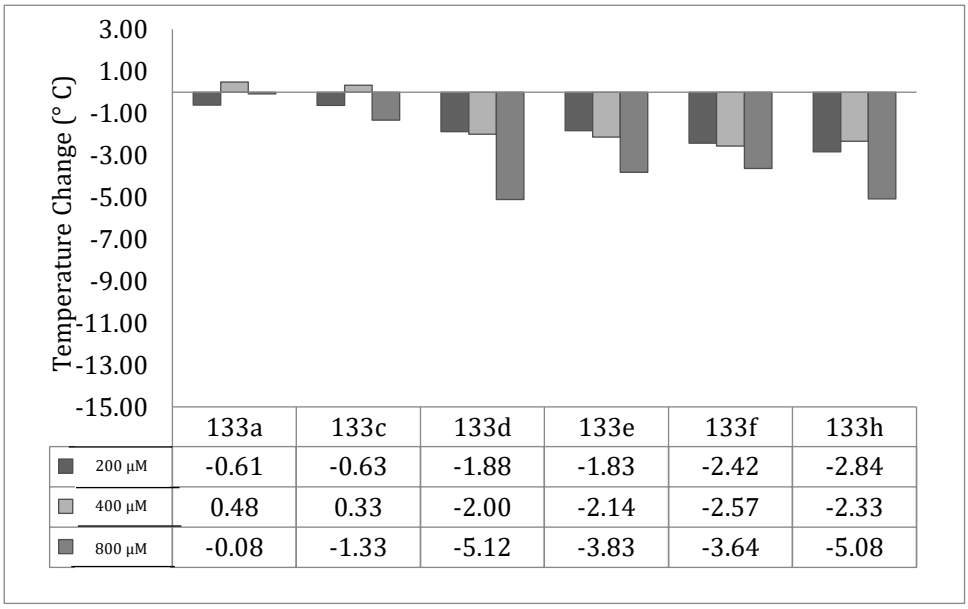
The author and H. Yamada carried out the work detailed in this section.

6.9.1 THERMAL SHIFT ASSAYS (DIFFERENTIAL SCANNING FLUORIMETRY)

As previously discussed in Chapter 5, thermal shift assays can be used to determine if the compound can induce a stabilising or destabilising change in the protein determined by a hydrophobic fluorescent dye.[95] The matrix protein was incubated at 35 μM concentration with the various compounds at three different concentrations and the results are shown below. A shift of not less than 2 °C was taken as significant due to the error in the machine and the assay. Each number in the tables below shows the melting tempertature change in comparison to the hRSV M protein in water.

6.9.1.1. INDAZOLE AND BENZIMIDAZOLE (133 a-h)

133a and 133c did not show significant changes in the assay whereas a dose dependant response was seen for the d and e subsitituents, 6 and 7-aminoindol. A similar response was seen in the imidazole compounds, in which the nitrogen atoms are moved into the 1 and 3 positions of the ring.



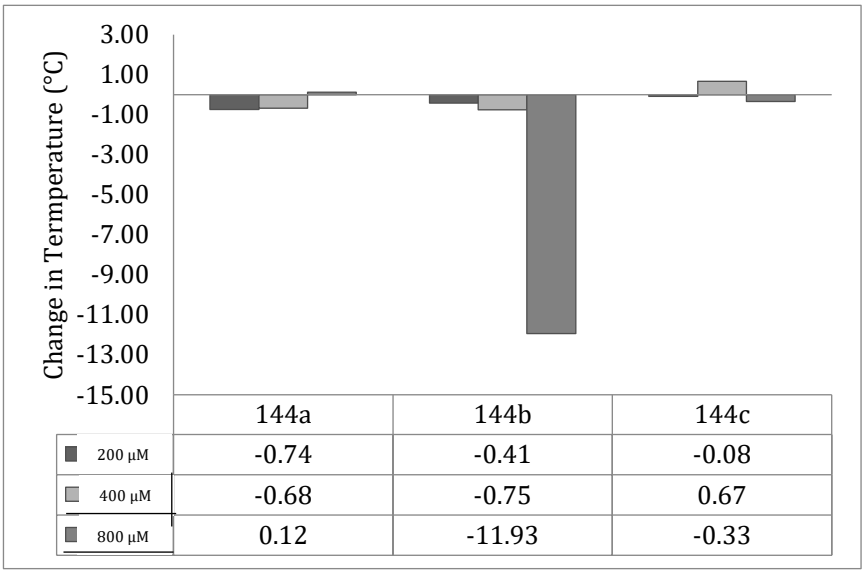
	133a	133c	133d	133e	133f	133h
200 μM	-0.61	-0.63	-1.88	-1.83	-2.42	-2.84
400 μM	0.48	0.33	-2.00	-2.14	-2.57	-2.33
800 μM	-0.08	-1.33	-5.12	-3.83	-3.64	-5.08

Figure 6.9.1.1.1 Indazole and Imidazole compound destabilisation of the matrix protein.

It was expected that the compounds, if bound to the correct pocket would destabilise the protein, which as shown above seems to be true as illustrated by the decreases in melting temperature. The protein we believe, would bind the compounds in the binding pocket on the dimer interface, destabilising the interacting dimer by forming hydrogen bonds to the compound and not the other subunit of the dimer.

6.9.1.2 ISOQUINOLINE (144 a-c)

Lesser destabilisations were seen with the isoquinoline compounds in the assay, mostly these were under the 2 °C threshold required. In sample 144b there is a large jump in the destabilisation of the protein at 400 μM, this compound was tested further at the higher concentration to determine if this was anomaly and showed that the compound is insoluble above 500 μM, rendering 144b unusable in this experiment.

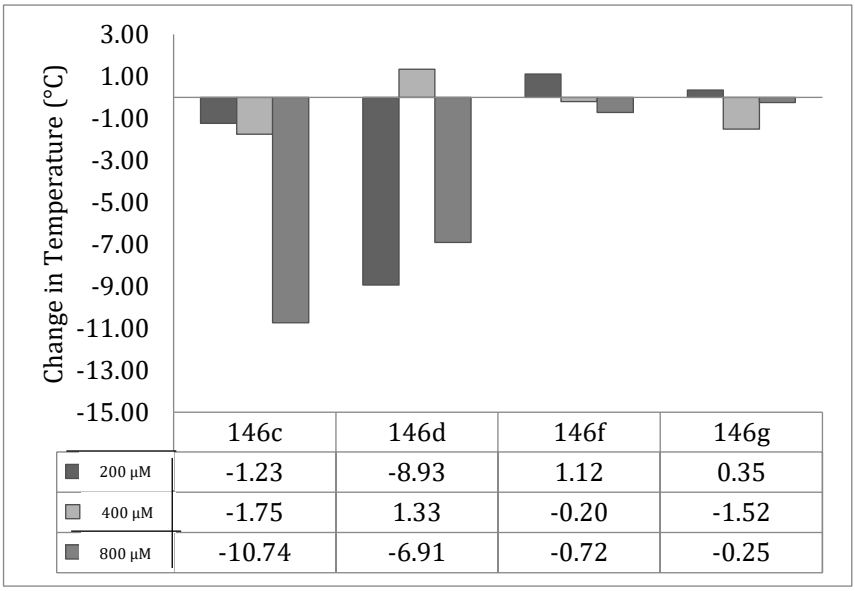


	144a	144b	144c
200 μM	-0.74	-0.41	-0.08
400 μM	-0.68	-0.75	0.67
800 μM	0.12	-11.93	-0.33

Figure 6.9.1.2.1 Isoquinoline compounds tested with matrix protein

6.9.1.3 INDOLE (146 c-g)

The indole compounds showed varied responses in assays, 146c was dose-responsive and at 800 μM showed a large destabilisation of approximately 11 °C. In contrast 146f and 146g did not show any change above 2 °C and were discounted. 146d showed a mixed result, and further investigations using the compound found that it had limited solubility and this it was ruled out of the experiment.



	146c	146d	146f	146g
200 μM	-1.23	-8.93	1.12	0.35
400 μM	-1.75	1.33	-0.20	-1.52
800 μM	-10.74	-6.91	-0.72	-0.25

Figure 6.9 1.3.1 Indole compound destabilising effect on matrix protein.

6.9.1.4 PURCHASED COMPOUNDS

Due to time constraints only one compound arrived before the deadline for testing. This was Compound 15 purchased through Mcule Inc (Budapest, Hungary) and the thermal shift data are shown below in Figure 6.9.1.4.1

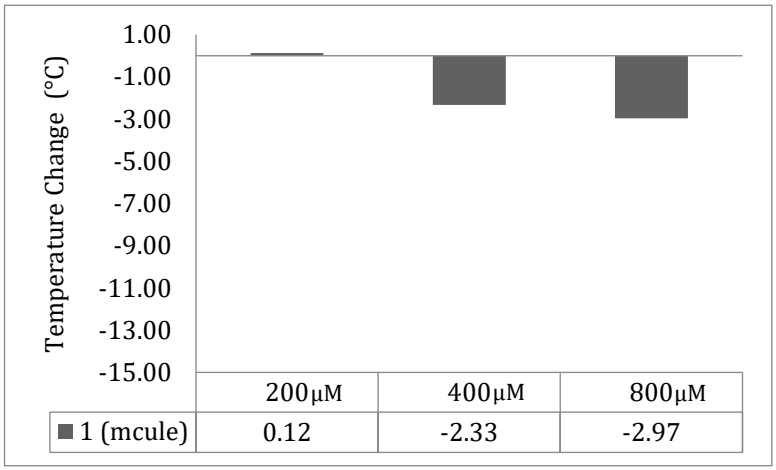


Figure 6.9.1.4.1 Compound 1 thermal shift assay result.

The compound is dose responsive and the higher concentrations go above the 2 °C threshold that was required, this is a positive sign and further experiments are needed to characterise the interaction.

6.9 FURTHER WORK

6.9.1 *IN VIVO* TESTING

Using in cell-assays could help characterise the interactions and understand how the compounds could bind and change the protein. Plaque inhibition assays to understand how effective the compounds are in hindering the spread of virus could be useful for taking the project further.[172] Plaque inhibition assays have an advantage over other *in vivo* assays due the reliable and rapid nature of determining 50% inhibitory concentrations.[172]

6.9.2 ISOTHERMAL TITRATION CALORIMETRY

This *in vitro* technique could further characterise interactions in solution and yield significant thermodynamic data to determine a binding model, dissociation constant and furthermore the changes in enthalpy, stoichiometry and entropy.[115,173] Binding of the ligands can be compared against the binding of the wild-type peptide and competitive binding studies could confirm the new ligands' effectiveness.

6.9.3 MUTATIONS OF HRSV M

Mutations within this binding pocket and across the dimer interface could be used to determine where the main contacts are forming hydrogen bonds. Residue Tyr 229 has been identified in this work but also by Förster *et al.*[49] where it was identified as a hydrogen bond contact in the dimer interface.

6.11 CONCLUSIONS.

A new understanding of viruses and their structures has helped the development of new anti-viral drug candidates. With the advent of computational docking studies, a new approach can be utilised where a potential binding target can be identified in hRSV M and small molecules can be identified and docked into place *in silico*. hRSV M has been long identified as a drug target by its function as an intergral part of the virus.[48,50] The protein's interaction with the fusion protein enables viral budding and maturation, that if disrupted results in a loss of virus maturation and a decrease in infectivity.[98] Shaikh *et. al.*[7] identified that the last 6 amino acids in the cytoplasmic tail were key to the maturation of the virus by forming a complex with hRSV M protein. Disruption of this interaction is key and it is hoped that the small molecules identified, may be able to mimic the disruption shown to be caused by mutations in the F cytoplasmic tail.

The pocket identified on the surface of hRSV M is small enough, in theory, to allow small molecules to bind without being promiscuous and an ideal candidate for docking studies. The ZINC database[164] provided a large molecule subset to identify drug-like molecules that can be taken through to the GOLD programs and subsequently to be tested *in vitro.* A fragment-like subset was taken from the database to be tested as the molecular weight and space fitting of the compounds fitted the chosen pocket. Over 1 million compounds were in the subset and these were filtered using KNIME, a data-mining platform.[139] A key residue, a phenylalanine in the fusion protein tail has been found to show binding to hRSV M[7] so an aromatic ring was used as the first constraint to filter. Further filters constrained the compounds to Lipinski's rule of 5 and subsquently the Fragment rule of 3.

The compounds that fit all constraints were taken through to GOLD to be docked into the pocket, the results have shown a key scaffold of at least 2 aromatic rings which form hydrogen bonds from an amine or alcohol group. Key residues in the pocket are Tyr229 and Asp98, which form hydrogen bonds in most of the docking results.

The recent publication of hRSV M dimer[49] only assisted in the confirmation in the dimer model, where the symmetry and pockets were identified in the same region. The new crystal structure was used to re-dock the compound, which had been identified through the initial search and the re-dockings gave positive solutions and similar hydrogen bond

contacts were found. Further validation of the compounds is needed and this could be in the form of *in vitro* testing by thermal shift assays and isothermal titration calorimetry and *in vivo* plaque inhibition assays to assess how effective the compounds are against the virus.

6.12 METHODS

6.12.1 PISA SYMMETRY DIMER

PDBe Pisa[158] was used to analyse the hRSV M structure, where the second chain was produced by the symmetry operation [-*x*+2, *y*, -*z*]. This result gave the largest buried surface area at 943.6 Å per monomer. The new model was saved as a pdb file for further modifications.

6.12.2 HADDOCK WEB INTERFACE

The new Pisa pdb file was manipulated to exclude water and all residues were labelled differently so that HADDOCK interface[153] can distinguish each individually in the docking process. Three residues were identified from the dimer interface, Ser63, Asp105 and Lys232, these were selected as active residues that the docking should be driven towards and the pdb file was run through the system.
The resulting graphs of scoring functions and energy calculations were analysed to obtain the best model going forward and the HADDOCK model was saved as a pdb file for further analysis.

6.12.3 CASTP

The HADDOCK model was uploaded to the CASTp server[161] (http://sts.bioe.uic.edu/castp/), the pockets were analysed and a total of 74 pockets were identified. Ideally the pocket would be on the dimer interface and a candidate was identified in pocket 72 at 450 $Å^3$ in volume.

6.12.4 ZINC

The Zinc database[164] was accessed via the website (http://zinc.docking.org/). Fragment-like and drug-like subsets were chosen from the subsets menu. The compounds downloaded for knime screen were from the in stock row. The files were downloaded as SDF files and saved to the computer for easy access.

6.12.5 KNIME

The Knime program was downloaded from (https://www.knime.org/).[169] The program was used to open the sdf file previously downloaded. Each node, which is the filtering station was dragged from the left menu onto the main screen. The full scheme is detailed in Appendix B. where first node used was to convert the SDF file into a KNIME readable format.

The structures were then filtered through to ensure the critical benzene from the Phenylalanine residue (F572) was maintained by the Substructure search. Lipinski's rule of 5[168] and the Fragment rule of 3[163] were used to filter down using the molecular properties and row splitter. To connect each node the arrows were connected from the output arrow to the left hand arrow on the next node. Each time the node was right clicked and 'start' was selected to start the filtering process. This took anywhere from 5 minutes to 20 hours dependent on the task and number of molecules involved.

Once the compounds were reduced, the files were converted using 'CDK to molecule' and 'Interactive table' so the remaining compounds could be viewed. Further conversion to different file types were used including 'MolConverter' especially for GOLD files.

6.12.6 GOLD

Molecular docking was carried out using GOLD (version 5.1)[80,81]. The interaction dimer and hRSV M dimer (PDB entry 4D4T) were taken from the PDB (www.rscb.org/pdb), and both models were edited in GOLD interface to include Hydrogen atoms. GOLD default settings were used throughout unless otherwise specified and shown in Appendix B.1.

A radius of 8 Å was used about a point in between Tyr229 and Asn 93 in the binding pocket (X = 4.048, Y = 6.036 and Z = -8.303) and the default scoring function CHEMPLP and ChemScore were used to score docking poses.[12] Ligands were allowed maximum available flexibility for all rotatable bonds in this run.

The output mol2 files from the filtering in KNIME were used as ligands and GOLD was run for 3 days, until all ligands were posed.

CHAPTER 7. CONCLUSIONS AND FUTURE PERSPECTIVES.

The prevalence of the Human Respiratory Syncytial Virus highlights the need for new antiviral agents. The virus and the secondary infections that follow are the second leading cause of death in elderly patients[24] and result in millions of outpatient hospital visits by children under 5.[174]

Although the virus was discovered in 1955, relatively little is known about the viral proteins which form the basis of viral infectivity. In 2001 the closely related metapneumovirus (hMPV),[26,37] was discovered and has since been found to co-infect with hRSV and lead to bronchiolitis with further complications.[26] Some infections can lead to long-term effects; where up to 70% of infants are left with respiratory problems for up to 10 years following hRSV bronchiolitis.[28,174]

In order to understand which viral proteins are key, their functions and viral replication cycle have to be determined. hRSV is a negative sense RNA virus which is contained in class V of the Baltimore classification system.[21] This class contains a number of families, of which many are important human diseases i.e. Mumps, measles and hRSV.

In the family *Paramyxoviridae* lies a subfamily *Pneumonvirinae*, of which hRSV and hMPV are members. The genome of the virus comprises of 11 genes which codes for 10 proteins.[77] Key proteins include the nucleocapsid protein (N)[46] which associates with genomic RNA to form a nucleocapsid. The fusion protein (F) is a glycoprotein on the surface and mediates cell fusion allowing the virus to enter the cell cytoplasm.[7,67] M2 is another structural protein which has two open reading frames and M2-1 is the elongation factor.[64] M2-1 is of specific importance to this work as it was one of the proteins in the virus with no published X-Ray Protein Crystallography structure in 2011. Finally the matrix protein (M) was solved by Money *et al*[48] in 2009 and Förster *et al*[49] in 2015 found to form interactions within the virus to M2-1 and F proteins. These interactions could be key for understanding and targeting a process for anti-viral agents to disrupt.

In Chapter 2, M2-1 was targeted for characterisation. The protein had no published X-Ray crystal structure, only a core solution NMR structure.[63] The protein was relatively uncharacterised with only basic information available. This ranged from the secondary structure of the protein backbone through circular dichroism[62] and the diameter from dynamic light scattering.[62] In this chapter, these parameters were confirmed and the protein was further analysed.

M2-1 proteins contain a zinc finger where a zinc ion is held by a Cys_3-His_1 motif, and this domain is key for the correct folding of the virus.[61] ICP-MS was used to quantify the zinc bound in the protein compared to the standard concentrations. The protein was found to be more stable at higher zinc concentrations indicating that the zinc finger was highly influential. From these data the expression protocol was changed to include $ZnSO_4$ at the point of expression.

Crystallisation of the full length protein was attempted through multiple trials using 96 well screening techniques. Conditions were optimised from lead conditions to yield bipyramid shaped crystals. Unfortunately the crystals did not diffract and Tanner *et al.* published a complete structure in 2014.[5]

In Chapter 3 the focus changed to protein-protein interactions. Matrix protein was first crystallised in 2009 by Money *et al*.[48] It is known to interact with M2-1 through cell inclusion assays[58,77] and this chapter confirmed that binding does take place through the use of pull down assays. Dynamic light scattering was used to understand the size of the complex formed and transmission electron microscopy was used as a tool to image the complex. The TEM found 2 distinct protein aggregates in the mixture of M2-1 and M, the sizes ranged from ~ 30-60 μM. This low resolution technique was limited for our purpose so further investigations were directed to understanding the thermodynamic nature of the binding. Further analysis through Cryo-Electron Tomography could yield a more conclusive image of the protein aggregates.

Surface Plasmon Resonance (SPR) and Fluorescence anisotropy (FA) were used to further understand the binding. The former technique gave a K_d of 3 mM where the data fitted to a 1:1 Langmuir interaction model. In FA the experiments were conducted with a fluorescent tag to monitor the binding in solution, in contrast to SPR where one is tethered to a solid surface. The results proved difficult to analyse, in the oligomerisation model, the fitted curve was found to be a close fit up to 10 μM however above this concentration the fit was poor. As matrix protein has been found to form up to a hexamer in solution[45] so the dimerized form of the protein could not be relied upon. In future work, mutagenesis of hRSV M binding region would yield potential to specifically target this interaction.

Chapter 4 introduced hRSV F and the interaction with hRSV M. This interaction is key to the viral budding and maturation. Through previous work[7] a key residue F572 found in the cytoplasmic tail of hRSV fusion protein which binds to hRSV matrix protein. A six

amino acid peptide consisting of the terminal amino acids of the cytoplasmic tail was purified and used for the following experiments.

Initially the binding was confirmed through the use of surface plasmon resonance experiments. The first method involved the matrix protein binding to the chip through a histidine tag and the fusion peptide was then added in solution. This unfortunately showed little binding so an alternative experimental procedure was employed. The tagged matrix protein was attached to the chip, as before, but another layer of matrix protein was added. This allowed the formation of possible oligomers, and when the fusion peptide was introduced, binding occurred.

Fluorescence anisotropy and FRET experiments were also conducted in Chapter 4. The initial FRET experiment was conducted by E. Antonio, which showed clear binding when analysed and a nM K_d. To confirm the value statistical plots were used to further analyse the data. A mutant peptide was also tested to confirm that the residue identified by Shaikh *et al.* [7] was the key binding residue. Fluorescence Anisotropy experiments were conducted alongside FRET to confirm the possible binding. Again, a binding curve was seen and was analysed using the same equilibrium pathway. Some discrepancies were seen in the figures and in further work these would need to be investigated to understand if the problem is an experimental or analysis issue.

In Chapter 5, the focus changes to understanding the interaction of small molecules with hRSV N protein. A small subset of compounds was made available by S Cockerhill, Durham University and were tested using various biophysical techniques. Thermal shift assays were again employed to understand the stabilising and destabilising effects of the compounds on hRSV N. This showed a varied destabilisation which prompted further investigation using Dynamic Light Scattering (DLS) and Transmission Electron Microscopy (TEM). DLS was used to understand how hRSV N changes during the interaction with the ligand compounds. It was seen that each ligand molecule formed aggregates at various concentrations as a large shift was seen in hydrodynamic radius. Transmission Electron Microscopy, which was used in Chapter 3, enables the shape of a sample to be imaged. From this a clear mode of action could be seen where the protein formed large aggregates in the presence of the ligands. Further work using mutagenesis to pin point the binding site would build a larger picture of how the ligands are interacting with the protein. Once this is known further optimisation of the structure can be obtained, to reduce potential problems with potential side effects and solubility issues.

In Chapter 6, the aim was to identify small fragment-like ligands which would bind to hRSV M. The ligands would hopefully bind in a pocket and disrupt the dimer interface, so disrupting the binding of fusion protein. A pocket was first identified on the surface of the protein using a web based server, CASTp.[161] A commercially available subset of compounds from ZINC database[164] was used to filter down from over 1 000 000 compounds through KNIME.[138] A filter was used as a template where Lipinski's rule of 5[168] and fragment rule of 3[163] and a core benzene must be in the final structure. The core benzene represented the phenylalanine side chain F572 which is key for binding.

Once a smaller number of compounds was achieved, the remaining compounds were docked in GOLD[12,80,147]. These were posed in the pocket determined in CASTp and a fitness score was attached to each. The top 20% were evaluated and ordered from the manufacturers. Unfortunately, due to time constraints only one ligand, compound 15 from the docking subset arrived for testing before the period of study ended. When tested by thermal shift assays the results showed a concentration-dependant destabilisation. This is what was expected and further studies are needed to understand how the ligands are destabilising the protein.

In 2015, a new hRSV M dimer structure was published[49] that confirmed the theory that the protein was a dimer in solution. In comparison to the model dimer structure, the published dimer was a close fit, where the shape and main Tyr229 contact were alike. The top 20% of compounds in the screen were redocked into the protein pockets using GOLD to validate the compounds. The docking reproduced the results from the model and confirmed the same hydrogen bond contacts. Further work would expand the testing of the compounds identified in the screening process. Structure-function activity relationship studies could further optimise the ligand's structure.

Overall this work begins to understand the complex interactions between viral proteins. The interactions between hRSV M2-1 and hRSV M and also hRSV F and hRSV M are key to the viruses' infectivity and maturation. Targeting these interactions with small ligand compounds begins to specifically target the binding and will hopefully lead to new trials of anti-viral agents.

REFERENCES

1. Hall, C. B. Respiratory syncytial virus in young children. *Lancet* **375,** 1500–1502 (2010).
2. Haynes, A. K. ., Prill, M., Iwane, M K. Iwane & Gerber, S. I. Respiratory Syncytial Virus — United States, July 2012–June 2014. **63,** (2014).
3. Hart, C. A. & Cuevas, L. E. Acute respiratory infections in children Infecções respiratórias agudas em crianças. **7,** 23–29 (2007).
4. Teng, M. N. & Collins, P. L. Identification of the respiratory syncytial virus proteins required for formation and passage of helper-dependent infectious particles. *J. Virol.* **72,** 5707–5716 (1998).
5. Tanner, S. J. *et al.* Crystal structure of the essential transcription antiterminator M2-1 protein of human respiratory syncytial virus and implications of its phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 1580–1585 (2014).
6. Mitra, R., Baviskar, P., Duncan-Decocq, R. R., Patel, D. & Oomens, A. G. P. The human respiratory syncytial virus matrix protein is required for maturation of viral filaments. *J. Virol.* **86,** 4432–4443 (2012).
7. Shaikh, F. Y. *et al.* A critical phenylalanine residue in the respiratory syncytial virus fusion protein cytoplasmic tail mediates assembly of internal viral proteins into viral filaments and particles. *MBio* **3,** (2012).
8. Tawar, R. G. *et al.* Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science* **326,** 1279–1283 (2009).
9. El Omari, K. *et al.* Structures of respiratory syncytial virus nucleocapsid protein from two crystal forms: details of potential packing interactions in the native helical form. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **67,** 1179–1183 (2011).
10. Chapman, J. *et al.* RSV604, a novel inhibitor of respiratory syncytial virus replication. *Antimicrob. Agents Chemother.* **51,** 3346–3353 (2007).
11. Mitra, R., Baviskar, P., Duncan-Decocq, R. R., Patel, D. & Oomens, A. G. P. The human respiratory syncytial virus matrix protein is required for maturation of viral filaments. *J. Virol.* **86,** 4432–4443 (2012).
12. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52,** 609–623 (2003).
13. S. J. Martin. *The Biochemistry of Viruses.* (Cambridge University Texts, 1978).
14. Vaney, M.-C. & Rey, F. A. Class II enveloped viruses. *Cell. Microbiol.* **13,** 1451–1459 (2011).
15. Bächi, T. & Howe, C. Morphogenesis and ultrastructure of respiratory syncytial virus. *J. Virol.* **12,** 1173–1180 (1973).
16. Dessau, M. A. & Modis, Y. Protein crystallization for X-ray crystallography. *J. Vis. Exp.* (2011). doi:10.3791/2285
17. Chirgadze, D. Protein crystallisation in action. (2001). Available at: http://www.xray.bioc.cam.ac.uk/xray_resources/whitepapers/xtal-in-action/node3.html.
18. Strauss J, S. E. *Viruses and Human Disease.* (Academic Press, 2002).

19. Simoes, E. Respiratory syncytial virus infection. *Lancet* (1999).

20. Murphy, F. A. *et al.* Virus taxonomy: classification and nomenclature of viruses. Sixth report of the International Committee on Taxonomy of Viruses. (1995).

21. Baltimore, D. Expression of animal virus genomes. *Bacteriol. Rev.* **35,** 235–241 (1971).

22. Kingsbury, D. Orthomyxoviridae and their replication. *Fields BN, Knipe DM. Virol.* (1990).

23. Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380,** 2095–2128 (2012).

24. Empey, K. M., Peebles, R. S. & Kolls, J. K. Pharmacologic advances in the treatment and prevention of respiratory syncytial virus. *Clin. Infect. Dis.* **50,** 1258–1267 (2010).

25. McNamara, P. S. & Smyth, R. L. The pathogenesis of respiratory syncytial virus disease in childhood. *Br. Med. Bull.* **61,** 13–28 (2002).

26. Hamelin, M.-E., Abed, Y. & Boivin, G. Human metapneumovirus: a new player among respiratory viruses. *Clin. Infect. Dis.* **38,** 983–990 (2004).

27. Smyth, R. L. & Openshaw, P. J. M. Bronchiolitis. *Lancet* **368,** 312–322 (2006).

28. Edell, D., Khoshoo, V., Ross, G. & Salter, K. Early ribavarin treatment of bronchiolitis: effect on long-term respiratory morbidity. *Chest* **122,** 935–939 (2002).

29. Hall, C. B. *et al.* The burden of respiratory syncytial virus infection in young children. *N. Engl. J. Med.* **360,** 588–598 (2009).

30. Stevens, W. W., Falsey, A. R. & Braciale, T. J. RSV 2007: recent advances in respiratory syncytial virus research. *Viral Immunol.* **21,** 133–140 (2008).

31. Teale, A., Deshpande, S. & Burls, A. Palivizumab and the importance of cost effectiveness. *BMJ* **338,** 1935–1940 (2009).

32. Committee on Infectious Diseases. Use of Ribavirin in the Treatment of Respiratory Syncytial Virus Infection. *Pediatrics* **92,** 501–504 (1993).

33. Roberts, S. S. *et al.* The Ribavirin Pregnancy Registry: Findings after 5 years of enrollment, 2003-2009. *Birth Defects Res. A. Clin. Mol. Teratol.* **88,** 551–559 (2010).

34. Ghildyal, R., Ho, A. & Jans, D. A. Central role of the respiratory syncytial virus matrix protein in infection. *FEMS Microbiol. Rev.* **30,** 692–705 (2006).

35. Johnson, P. R., Spriggs, M. K., Olmsted, R. A. & Collins, P. L. The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins. *Proc. Natl. Acad. Sci.* **84,** 5625–5629 (1987).

36. Liljeroos, L., Krzyzaniak, M. A., Helenius, A. & Butcher, S. J. Architecture of respiratory syncytial virus revealed by electron cryotomography. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 11133–11138 (2013).

37. Smith, E. C., Popa, A., Chang, A., Masante, C. & Dutch, R. E. Viral entry mechanisms: the increasing diversity of paramyxovirus entry. *FEBS J.* **276,** 7217–7227 (2009).

38. Cosset, F.-L. & Lavillette, D. Cell entry of enveloped viruses. *Adv. Genet.* **73,** 121–183 (2011).

39. Battisti, A. J. *et al.* Structure and assembly of a paramyxovirus matrix protein. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 13996–14000 (2012).

40. Mercer, J., Schelhaas, M. & Helenius, A. Virus entry by endocytosis. *Annu. Rev. Biochem.* **79,** 803–833 (2010).

41. Bally, M., Dimitrievski, K., Larson, G., Zhdanov, V. P. & Höök, F. Interaction of virions with membrane glycolipids. *Phys. Biol.* **9,** 26011 (2012).

42. Welsch, S., Müller, B. & Kräusslich, H.-G. More than one door - Budding of enveloped viruses through cellular membranes. *FEBS Lett.* **581,** 2089–2097 (2007).

43. Zhao, X., Singh, M., Malashkevich, V. N. & Kim, P. S. Structural characterization of the human respiratory syncytial virus fusion protein core. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 14172–14177 (2000).

44. Lee, J. E. & Saphire, E. O. Ebolavirus glycoprotein structure and mechanism of entry. *Future Virol.* **4,** 621–635 (2009).

45. McPhee, H. K. A study of the Membrane Binding Properties of the Matrix protein from Human Respiratory Syncytial Virus. (Durham University, 2009).

46. Teng, M. N. & Collins, P. L. Identification of the Respiratory Syncytial Virus Proteins Required for Formation and Passage of Helper-Dependent Infectious Particles. *J. Virol.* **72,** 5707–5716 (1998).

47. Rodriguez, L. Human respiratory syncytial virus matrix protein is an RNA-binding protein: binding properties, location and identity of the RNA contact residues. *J. Gen. Virol.* **85,** 709–719 (2004).

48. Money, V. A., McPhee, H. K., Mosely, J. A., Sanderson, J. M. & Yeo, R. P. Surface features of a Mononegavirales matrix protein indicate sites of membrane interaction. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 4441–4446 (2009).

49. Förster, A., Maertens, G. N., Farrell, P. J. & Bajorek, M. Dimerization of Matrix protein is required for budding of Respiratory Syncytial Virus. *J. Virol.* **89,** (2015).

50. Leyrat, C., Renner, M., Harlos, K., Huiskonen, J. T. & Grimes, J. M. Structure and self-assembly of the calcium binding matrix protein of human metapneumovirus. *Structure* **22,** 136–148 (2014).

51. Dessen, A., Volchkov, V., Dolnik, O., Klenk, H. D. & Weissenhorn, W. Crystal structure of the matrix protein VP40 from Ebola virus. *EMBO J.* **19,** 4228–4236 (2000).

52. Gaudier, M., Gaudin, Y. & Knossow, M. Crystal structure of vesicular stomatitis virus matrix protein. *EMBO J.* **21,** 2886–2892 (2002).

53. Timmins, J., Scianimanico, S., Schoehn, G. & Weissenhorn, W. Vesicular release of ebola virus matrix protein VP40. *Virology* **283,** 1–6 (2001).

54. Alfadhli, A., Huseby, D., Kapit, E., Colman, D. & Barklis, E. Human immunodeficiency virus type 1 matrix protein assembles on membranes as a hexamer. *J. Virol.* **81,** 1472–1478 (2007).

55. Hoenen, T. *et al.* VP40 octamers are essential for Ebola virus replication. *J. Virol.*

79, 1898–1905 (2005).

56. Bornholdt, Z. a *et al.* Structural rearrangement of ebola virus VP40 begets multiple functions in the virus life cycle. *Cell* **154,** 763–774 (2013).

57. Neumann, P. *et al.* Crystal structure of the Borna disease virus matrix protein (BDV-M) reveals ssRNA binding properties. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 3710–3715 (2009).

58. Li, D. *et al.* Association of respiratory syncytial virus M protein with viral nucleocapsids is mediated by the M2-1 protein. *J. Virol.* **82,** 8863–8870 (2008).

59. Fearns, R. & Collins, P. L. Role of the M2-1 transcription antitermination protein of respiratory syncytial virus in sequential transcription. *J. Virol.* **73,** 5852–5864 (1999).

60. Cuesta, I., Geng, X., Asenjo, A. & Villanueva, N. Structural phosphoprotein M2-1 of the human respiratory syncytial virus is an RNA binding protein. *J. Virol.* **74,** 9858–9867 (2000).

61. Esperante, S. A. *et al.* Fine modulation of the respiratory syncytial virus M2-1 protein quaternary structure by reversible zinc removal from its Cys(3)-His(1) motif. *Biochemistry* **52,** 6779–6789 (2013).

62. Tran, T.-L. *et al.* The respiratory syncytial virus M2-1 protein forms tetramers and interacts with RNA and P in a competitive manner. *J. Virol.* **83,** 6363–6374 (2009).

63. Blondot, M.-L. *et al.* Structure and functional analysis of the RNA- and viral phosphoprotein-binding domain of respiratory syncytial virus M2-1 protein. *PLoS Pathog.* **8,** (2012).

64. Cartee, T. L. & Wertz, G. W. Respiratory syncytial virus M2-1 protein requires phosphorylation for efficient function and binds viral RNA during infection. *J. Virol.* **75,** 12188–12197 (2001).

65. Mason, S. W. *et al.* Interaction between Human Respiratory Syncytial Virus (RSV) M2-1 and P Proteins Is Required for Reconstitution of M2-1-Dependent RSV Minigenome Activity. *J. Virol.* **77,** 10670–10676 (2003).

66. Hartlieb, B., Muziol, T., Weissenhorn, W. & Becker, S. Crystal structure of the C-terminal domain of Ebola virus VP30 reveals a role in transcription and nucleocapsid association. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 624–629 (2007).

67. McLellan, J. S. *et al.* Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340,** 1113–1137 (2013).

68. Chaiwatpongsakorn, S., Epand, R. F., Collins, P. L., Epand, R. M. & Peeples, M. E. Soluble respiratory syncytial virus fusion protein in the fully cleaved, pretriggered state is triggered by exposure to low-molarity buffer. *J. Virol.* **85,** 3968–3977 (2011).

69. Lee, J. E. *et al.* Structure of the Ebola virus glycoprotein bound to an antibody from a human survivor. *Nature* **454,** 177–182 (2008).

70. Wen, X. *et al.* Structure of the human metapneumovirus fusion protein with neutralizing antibody identifies a pneumovirus antigenic site. *Nat. Struct. Mol. Biol.* **19,** 461–463 (2012).

71. Challa, S. *et al.* Mechanism of action for respiratory syncytial virus inhibitor RSV604. *Antimicrob. Agents Chemother.* **59,** 1080–1087 (2015).

72. Green, T. J. *et al.* Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. *J. Virol.* **74,** 9515–9524 (2000).

73. Albertini, A. A. V *et al.* Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313,** 360–363 (2006).

74. Green, T. J., Zhang, X., Wertz, G. W. & Luo, M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313,** 357–360 (2006).

75. Green, T. J. *et al.* Access to RNA encapsidated in the nucleocapsid of vesicular stomatitis virus. *J. Virol.* **85,** 2714–2722 (2011).

76. Liu, P., Yang, J., Wu, X. & Fu, Z. F. Interactions amongst rabies virus nucleoprotein, phosphoprotein and genomic RNA in virus-infected and transfected cells. *J. Gen. Virol.* **85,** 3725–3734 (2004).

77. Ghildyal, R., Mills, J., Murray, M., Vardaxis, N. & Meanger, J. Respiratory syncytial virus matrix protein associates with nucleocapsids in infected cells. *J. Gen. Virol.* **83,** 753–757 (2002).

78. Liljeroos, L., Krzyzaniak, M. A., Helenius, A. & Butcher, S. J. Architecture of respiratory syncytial virus revealed by electron cryotomography. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 11133–11138 (2013).

79. Förster, A., Maertens, G. N., Farrell, P. J. & Bajorek, M. Dimerization of Matrix protein is required for budding of Respiratory Syncytial Virus. *J. Virol.* (2015). doi:10.1128/JVI.03500-14

80. Olsson, T., Bowden, S., Crystallographic, C. & Centre, D. An overview of protein-ligand docking using GOLD. *PROTEINS Struct. Funct. Genet.* **52,** 609–623 (2003).

81. Cambridge Crystallographic Data Centre. *GoldMine User Guide* . (2013).

82. Kaelin, W. G. *et al.* Expression cloning of a cDNA encoding a retinoblastoma-binding protein with E2F-like properties. *Cell* **70,** 351–364 (1992).

83. Harper, S. & Speicher, D. W. Purification of proteins fused to glutathione S-transferase. *Methods Mol. Biol.* **681,** 259–80 (2011).

84. Cordingleys, M. G., Callahan, P. L., Sardana, V. V, Garsky, V. M. & Colonno, R. J. THE JOURNAL OF BIOLOGICAL CHEMISTRY Substrate Requirements of Human Rhinovirus 3C Protease for Peptide Cleavage in Vitro*. **265,** 9062–9065

85. Walker, P. A. *et al.* Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Biotechnology. (N. Y).* **12,** 601–5 (1994).

86. LAEMMLI, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227,** 680–685 (1970).

87. G.E Healthcare. UNICORN ™ 7.0 software. (2005).

88. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

89. Greenfield, S. *et al.* Inductively coupled plasmas in atomic fluorescence spectrometry. A review. *J. Anal. At. Spectrom.* **9,** 565 (1994).

90. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1,** 2876–2890 (2006).

91. Sreerama, N. & Woody, R. W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR

methods with an expanded reference set. *Anal. Biochem.* **287,** 252–260 (2000).

92. Arzen, D., Co-advisor, R. P. & Ljubljana, D. K. Dynamic light scattering and application to proteins in solutions. (2010).

93. Miao, X. M., Xiong, C., Wang, W. W., Ling, L. S. & Shuai, X. T. Dynamic-light-scattering-based sequence-specific recognition of double-stranded DNA with oligonucleotide-functionalized gold nanoparticles. *Chem. - A Eur. J.* **17,** 11230–11236 (2011).

94. Instruments, M. in

95. Grøftehauge, M. K., Hajizadeh, N. R., Swann, M. J. & Pohl, E. Protein–ligand interactions investigated by thermal shift assays (TSA) and dual polarization interferometry (DPI). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71,** 36–44 (2015).

96. Lavinder, J. J., Hari, S. B., Sullivan, B. J. & Magliery, T. J. High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *J. Am. Chem. Soc.* **131,** 3794–5 (2009).

97. Oliphant, T. E. Python for scientific computing. *Comput. Sci. Eng.* **9,** 10–20 (2007).

98. Kiss, G. *et al.* Structural Analysis of Respiratory Syncytial Virus Reveals the Position of M2-1 between the Matrix Protein and the Ribonucleoprotein Complex. *J. Virol.* **88,** 7602–7617 (2014).

99. Hampton Research. Crystal scoring 101. (2009). doi:10.1139/X09-045

100. Khurshid, S., Saridakis, E., Govada, L. & Chayen, N. E. Porous nucleating agents for protein crystallization. *Nat. Protoc.* **9,** 1621–1633 (2014).

101. Newman, J. *et al.* Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Cryst* **61,** 1426–1431 (2005).

102. Wooh, J. W. *et al.* Comparison of three commercial sparse-matrix crystallization screens. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59,** 769–772 (2003).

103. Page, R. *et al.* Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the Thermotoga maritima proteome. *Acta Crystallogr. D. Biol. Crystallogr.* **59,** 1028–37 (2003).

104. Detection of protein-protein interactions using the GST fusion protein pull-down technique. *Nat. Methods* **1,** 275–276 (2004).

105. Payne, J. W. Polymerization of proteins with glutaraldehyde. Soluble molecular-weight markers. *Biochem. J.* **135,** 867–73 (1973).

106. Mirsaidov, U. M., Zheng, H., Casana, Y. & Matsudaira, P. *Imaging Protein Structure in Water at 2.7 nm Resolution by Transmission Electron Microscopy*. *Biophysical Journal* **102,** (2012).

107. Merwe, P. A. Van Der. Surface plasmon resonance GENERAL PRINCIPLES OF BIACORE EXPERIMENTS. *Physics (College. Park. Md).* **627,** 1–50 (2010).

108. Drescher, D. G., Ramakrishnan, N. A. & Drescher, M. J. Surface plasmon resonance (SPR) analysis of binding interactions of proteins in inner-ear sensory epithelia. *Methods Mol. Biol.* **493,** 323–43 (2009).

109. Healthcare, G. E. & Sciences, L. *Sensor Chip NTA*. (2000).

110. G.E Healthcare. *BIACORE BIAevaluation Version 3 Software Handbook*. (1998).

111. Thyberg, P. Fluorescence Anisotropy. (2012).

112. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. (Springer US, 2006). doi:10.1007/978-0-387-46312-4

113. Life Technologies. Analysis of FP Binding Data. 1–12

114. McPhee, H. . A Study of the Membrane Binding Properties of the Matrix Protein from Human Respiratory Syncytial Virus. *PhD Thesis.* (Durham University, 2009).

115. Pierce, M. M., Raman, C. S. & Nall, B. T. Isothermal titration calorimetry of protein-protein interactions. *Methods* **19,** 213–221 (1999).

116. Tran, T.-L. *et al.* The nine C-terminal amino acids of the respiratory syncytial virus protein P are necessary and sufficient for binding to ribonucleoprotein complexes in which six ribonucleotides are contacted per N protein protomer. *J. Gen. Virol.* **88,** 196–206 (2007).

117. Bajorek, M. *et al.* The Thr205 phosphorylation site within respiratory syncytial virus matrix (M) protein modulates M oligomerization and virus production. *J. Virol.* **88,** 6380–6393 (2014).

118. Brock, S. C., Heck, J. M., McGraw, P. A. & Crowe, J. E. The transmembrane domain of the respiratory syncytial virus F protein is an orientation-independent apical plasma membrane sorting sequence. *J. Virol.* **79,** 12528–12535 (2005).

119. Invitrogen Corporation. Theory of Binding Data Analysis. *Fluoresc. Polariz. Tech. Resour. Guid. Chapter 7* (2008).

120. Sekar, R. B. & Periasamy, A. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J. Cell Biol.* **160,** 629–33 (2003).

121. Piston, D. W. & Kremers, G.-J. Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem. Sci.* **32,** 407–414 (2007).

122. Hussain, S. . An Introduction to FRET. (2012).

123. Antonio, E. Masters Thesis. *Durham Univerisity* (2014). doi:10.1016/S0022-3913(12)00047-9

124. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. U. S. A.* **58,** 719–726 (1967).

125. Berges, J., Montagnes, D., Hurd, C. & Harrison, P. Fitting ecological and physiological data to rectangular hyperbolae: a comparison of methods using Monte Carlo simulations . *Mar. Ecol. Prog. Ser.* **114,** 175–183 (1994).

126. Prinz, H. Hill coefficients, dose-response curves and allosteric mechanisms. *J. Chem. Biol.* **3,** 37–44 (2010).

127. Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat. Methods* **11,** 1064–1070 (2014).

128. Chapman, J. & Cockerill, G. S. in *Antiviral Drugs: From Basic Discovery through clinical trials* (2011).

129. Hodgson, E. *A textbook of Modern Toxicology*. (2004).

130. Kazmierski, W. M. *Antiviral Drugs: From Basic Discovery Through Clinical Trials*. (John Wiley & Sons, 2011).

131. Steed, J. W. Durham Research Online. *Trends pharmological Sci.* **34,** 185–193 (2013).

132. Dundas, J. *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34,** 116–118 (2006).

133. Heath, G. & Colburn, W. A. An evolution of drug development and clinical pharmacology during the 20th century. *J. Clin. Pharmacol.* **40,** 918–929 (2000).

134. Littler, E. The past, present and future of antiviral drug discovery. *IDrugs* **7,** 1104–1112 (2004).

135. Roses, A. D. Pharmacogenetics in drug discovery and development: a translational perspective. *Nat. Rev. Drug Discov.* **7,** 807–817 (2008).

136. Hughes, J. P., Rees, S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **162,** 1239–1249 (2011).

137. Friedman, L. M., Furberg, C. D. & DeMets, D. L. in *Fundamentals of Clinical Trials* 445

138. Taylor, R. D., Jewsbury, P. J. & Essex, J. W. A review of protein-small molecule docking methods. 151–166 (2002).

139. Ewing, T. J. A., Makino, S., Skillman, A. G. & Kuntz, I. D. DOCK 4 . 0 : Search strategies for automated molecular docking of flexible molecule databases. 411–428 (2001).

140. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261,** 470–489 (1996).

141. Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55,** 379–400 (1971).

142. Greer, J. & Bush, B. L. Macromolecular shape and surface maps by solvent exclusion. *Proc. Natl. Acad. Sci. U. S. A.* **75,** 303–307 (1978).

143. Richards, F. M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6,** 151–176 (1977).

144. Wodak, S. J., De Crombrugghe, M. & Janin, J. Computer studies of interactions between macromolecules. *Prog. Biophys. Mol. Biol.* **49,** 29–63 (1987).

145. Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. & Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **161,** 269–288 (1982).

146. Miller, M. D., Kearsley, S. K., Underwood, D. J. & Sheridan, R. P. FLOG: A system to select ?quasi-flexible? ligands complementary to a receptor of known three-dimensional structure. *J. Comput. Aided. Mol. Des.* **8,** 153–174 (1994).

147. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267,** 727–748 (1997).

148. Korb, O., Stützle, T. & Exner, T. E. Empirical scoring functions for advanced Protein-Ligand docking with PLANTS. *J. Chem. Inf. Model.* **49,** 84–96 (2009).

149. Huang, S.-Y., Grinter, S. Z. & Zou, X. Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys* **12,** 12899–12908 (2010).

150. Cambridge Crystallographic Data Centre. GOLD User Guide. *2014*

151. Mooij, W. T. M. & Verdonk, M. L. General and targeted statistical potentials for protein-ligand interactions. *Proteins* **61,** 272–287 (2005).

152. Wang, R., Lu, Y., Fang, X. & Wang, S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J. Chem. Inf. Comput. Sci.* **44,** 2114–25

153. de Vries, S. J., van Dijk, M. & Bonvin, A. M. J. J. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc.* **5,** 883–897 (2010).

154. Gasteiger, E. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31,** 3784–3788 (2003).

155. Freeth, J. The interactions of viral matrix proteins with lipid membranes. (Durham University, 2014).

156. Schrödinger, LLC. *The {PyMOL} Molecular Graphics System, Version~1.3r1*. (2010).

157. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22,** 195–201 (2006).

158. K., K. E. H. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **43,** (2007).

159. Pierce, B. G. *et al.* ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **30,** 1771–1773 (2014).

160. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

161. Binkowski, T. A., Naghibzadeh, S. & Liang, J. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res.* **31,** 3352–3355 (2003).

162. Liang, J., Edelsbrunner, H. & Woodward ', C. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.* 71884–1897 (1998).

163. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A 'Rule of Three' for fragment-based lead discovery? *Drug Discov. Today* **8,** 876–877 (2003).

164. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **52,** 1757–1768 (2012).

165. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40,** 1100–1107 (2012).

166. Bolton, E. E., Wang, Y., Thiessen, P. a. & Bryant, S. H. PubChem: Integrated Platform of Small Molecules and Biological Activities. *Annu. Rep. Comput. Chem.* **4,** 217–241 (2008).

167. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* **42,** 1091–1097 (2014).

168. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46,** 3–26 (2001).

169. Tiwari, A. & Sekhar, A. K. T. Workflow based framework for life science informatics. *Comput. Biol. Chem.* **31,** 305–319 (2007).

170. Bruno, I. J. *et al.* Retrieval of crystallographically-derived molecular geometry information. *J. Chem. Inf. Comput. Sci.* **44,** 2133–2144 (2004).

171. Wallwork, S. C. Hydrogen-bond radii. *Acta Crystallogr.* **15,** 758–759 (1962).

172. Baer, A. & Kehn-Hall, K. Viral concentration determination through plaque assays: using traditional and novel overlay systems. *J. Vis. Exp.* e52065 (2014). doi:10.3791/52065

173. Pierce, M. M., Raman, C. S. & Nall, B. T. Isothermal titration calorimetry of protein-protein interactions. *Methods* **19,** 213–221 (1999).

174. Dawson-Caswell, M. & Muncie, H. L. Respiratory syncytial virus infection in children. *Am. Fam. Physician* **83,** 141–146 (2011).

APPENDIX A.

1. MALDI-TOF FRAGMENT ANALYSIS

The output file from the pdf report received from Dr A. P Brown showing the major fragments which were detected and the subsequent matched proteins from the BLAST database.[88]

Automated digest 2-95 spot B8 = band 20

N	Unused	Total	% Cov	Accession #	Name	Species
1	39.99	39.99	89.9	sp|P08515|GST26_SCHJA	Glutathione S-transferase class-mu 26 kDa isozyme OS=Schistosoma japonicum	
2	28.67	28.67	74.2	sp|P04545|M21_HRSVA	Matrix M2-1 OS=Human respiratory syncytial virus A (strain A2) GN=M2-1 PE=1 SV=1	HRSVA
3	8.75	8.75	25.1	sp|P00761|TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	PIG
4	4.00	4.00	19.2	sp|Q48406|BLAT_KLEOX	Beta-lactamase TEM-12 OS=Klebsiella oxytoca GN=blaT-12b PE=1 SV=1	KLEOX

Glutathione S-transferase class-mu 26 kDa isozyme OS=Schistosoma japonicum

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDF
LSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK

Matrix M2-1 OS=Human respiratory syncytial virus A (strain A2) GN=M2-1

MSRRNPCKFEIRGHCLNGKRCHFSHNYFEWPPHALLVRQNFMLNRILKSMDKSIDTLSEISGAAELDRTEEYALGVVGVLESYIGSINNITKQSACVAMSKLLTELNSDDIKKLRDNEELNS
PKIEVYNTVISYIESNRKNNKQTIHLLKRLPADVLKKTIKNTLDIHKSITINNPKESTVSDTNDHAKNNDTT

Automated digest 2-95 spot B9 = band 21

N	Unused	Total	% Cov	Accession #	Name	Species	Biological Processes	Molecular Functions	PANTHER ID
1	37.37	37.37	67.4	sp|P08515|GST26_SCHJA	Glutathione S-transferase class-mu 26 kDa isozyme OS=Schistosoma japonicum PE=1 SV=3				
2	32.96	32.96	75.8	sp|P04545|M21_HRSVA	Matrix M2-1 OS=Human respiratory syncytial virus A (strain A2) GN=M2-1 PE=1 SV=1	HRSVA			
3	7.83	7.83	17.6	sp|Q8XB34|TNAA_ECO57	Tryptophanase OS=Escherichia coli O157:H7 GN=tnaA PE=3 SV=1	ECO57			
4	6.06	6.06	10.7	sp|P04264|K2C1_HUMAN	Keratin, type II cytoskeletal 1 OS=Homo sapiens GN=KRT1 PE=1 SV=6	HUMAN			
5	5.71	5.71	27.7	sp|Q7N9B1|EFTU1_PHOLL	Elongation factor Tu 1 OS=Photorhabdus luminescens subsp. laumondii (strain TT01) GN=tuf1 PE=3				
6	4.34	4.34	20.8	sp|P00761|TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1	PIG			
7	4.01	9.71	17.9	sp|P15964|GST26_SCHMA	Glutathione S-transferase class-mu 26 kDa isozyme OS=Schistosoma mansoni PE=2 SV=1	SCHMA			
8	4.00	4.00	14.5	sp|Q6LMT1|ENO_PHOPR	Enolase OS=Photobacterium profundum GN=eno PE=3 SV=1	PHOPR			
9	3.17	20.30	57.4	sp|O42050|M21_HRSVB	Matrix M2-1 OS=Human respiratory syncytial virus B (strain B1) GN=M2-1 PE=3 SV=1	HRSVB			

Glutathione S-transferase class-mu 26 kDa isozyme OS=Schistosoma japonicum

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDF
LSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPK

2. INDUCTIVELY COUPLED PLASMA MASS SPECTROMETRY.

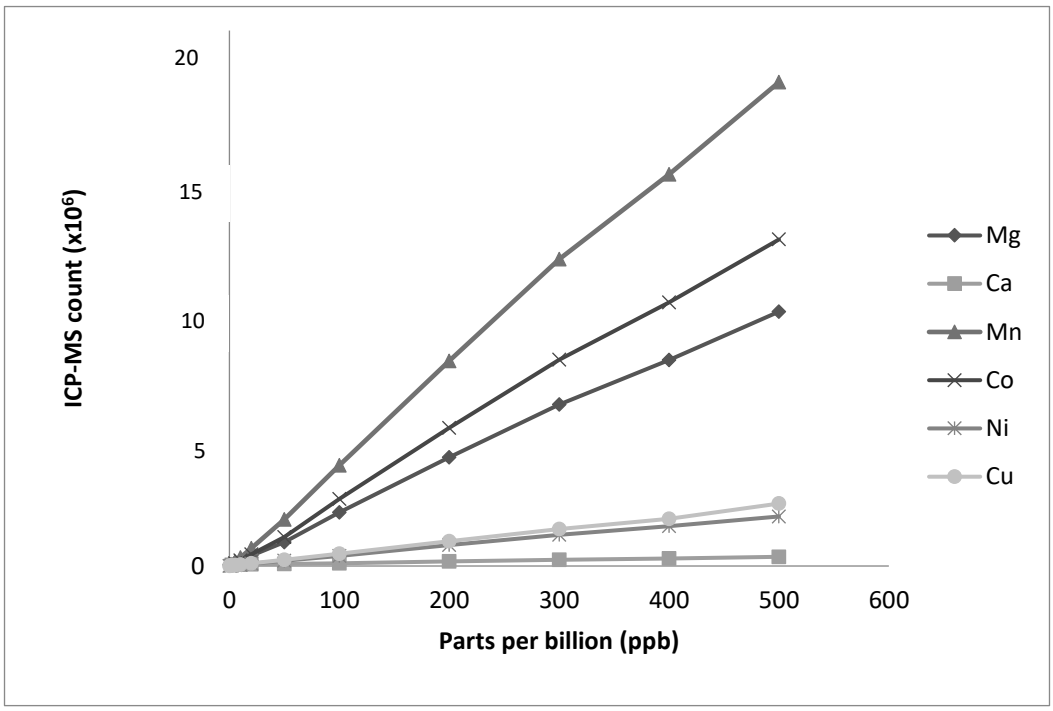


Figure A.2.1. Standard curve of 6 Standard metal ions, Mg, Ca,Mn,Co,Ni,Cu.

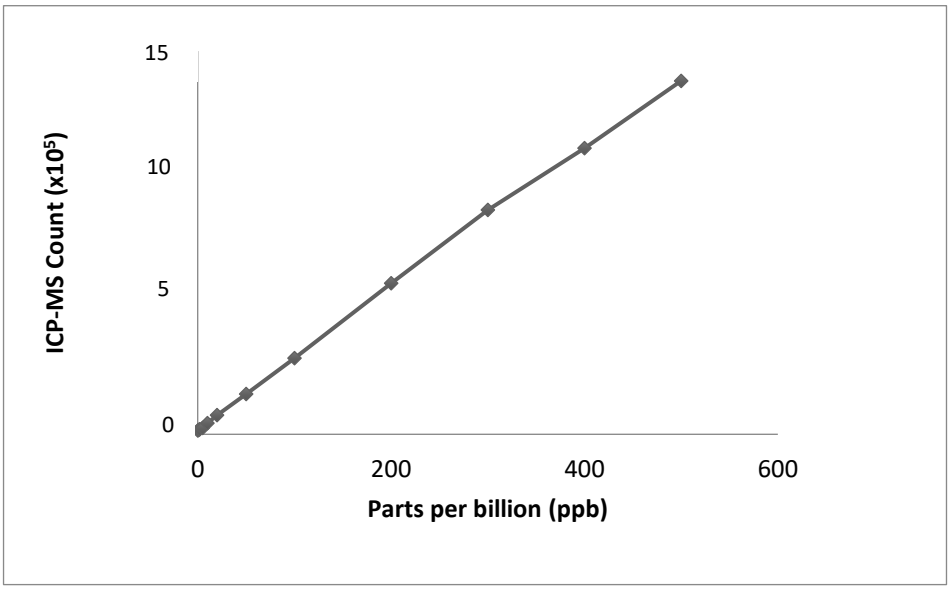


Figure 2.2.2.Appendix A.1.2. Zinc standard curve.

Figure A.2.2 Standard curves of known concentrations of 9 metal ions used in ICP-MS analysis. Zinc is shown separately due to the small scale of the ICP-MS count.

3. NAMI SALT SCREEN CONDITIONS.

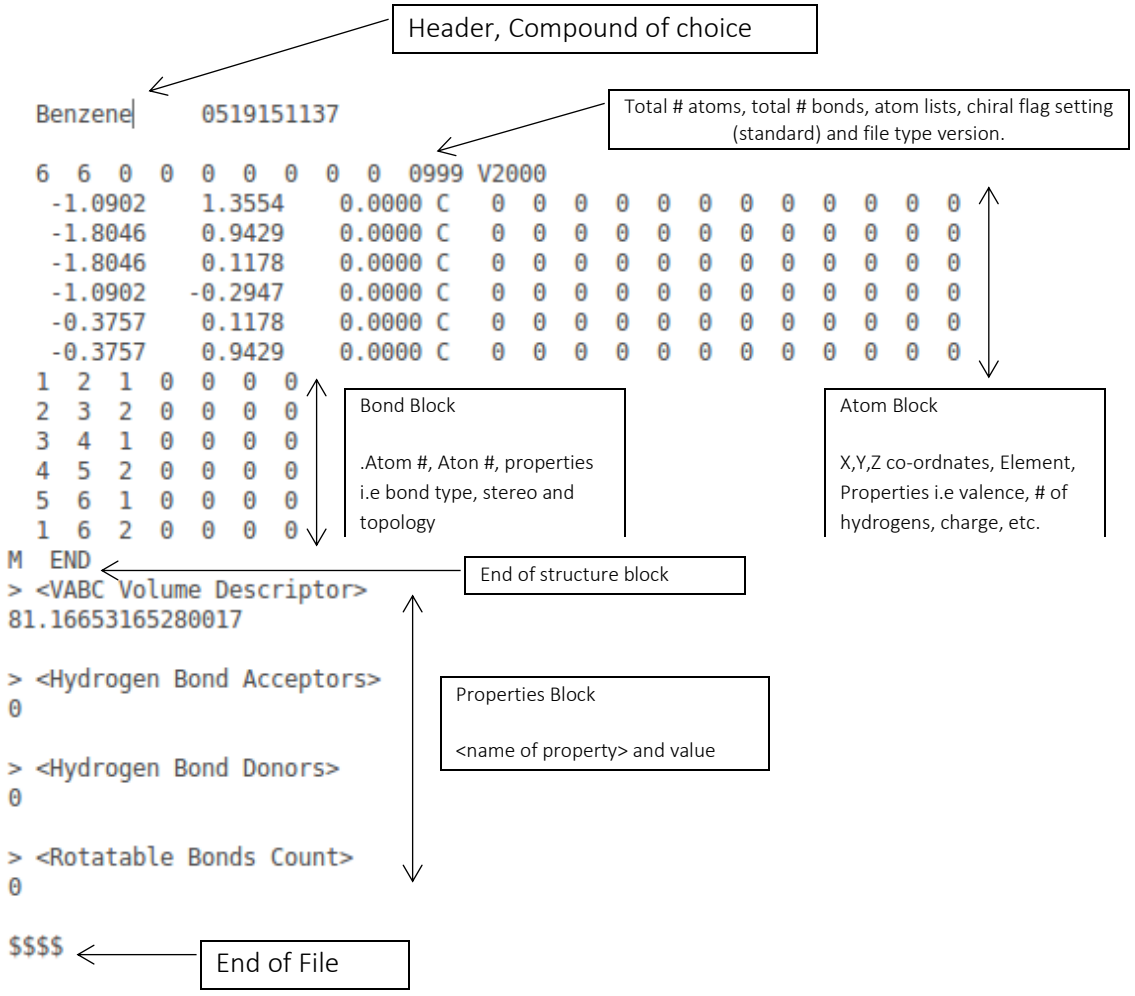
	1	2	3	4	5	6
A	Water	Water	Water	4 M Urea	0.5 M Urea	3.0 M Gu-HCl
B	1.5 M Na_2 malonate	1.0 M Na_2 malonate	0.8 M Na_2 malonate	0.6 M Na_2 malonate	0.4 M Na_2 malonate	0.2 M Na_2 malonate
C	1.5 M NaCl	1.0 M NaCl	0.8 M NaCl	0.6 M NaCl	0.4 M NaCl	0.2 M NaCl
D	1.0 M $MgSO_4$	0.8 M $MgSO_4$	0.6 M $MgSO_4$	0.4 M $MgSO_4$	0.2 M $MgSO_4$	1.0 M $NaSO_4$
E	0.5 M LiCl	0.2 M LiCl	0.5 M RbCl	0.2 M RbCl	0.5 M CsCl	0.2 M CsCl
F	0.4 M $MgCl_2$	5 mM $MgCl_2$	5 mM $CaCl_2$	5 mM $SrCl_2$	1 mM $ZnCl_2$	1 mM $NiCl_2$
G	1.5 M NaBr	magic triangle	2 mM $La(NO_3)_3$	2 mM $PrCl_3$	2 mM $NdCl_3$	2 mM $SmCl_3$
H	5 mM Na_2HPO_4	5 mM Na_3VO_4	5 mM Na_2WO_4	5 mM Na_2MoO_4	20% glycerol	10% glycerol
	7	8	9	10	11	12
A	1.0 M Gu-HCl	0.8 M Gu-HCl	0.6 M Gu-HCl	0.4 M Gu-HCl	0.2 M Gu-HCl	0.2 M Gu-HCl
B	1.5 M $(NH_4)2SO_4$	1.0 M $(NH_4)2SO_4$	0.8 M $(NH_4)2SO_4$	0.6 M $(NH_4)2SO_4$	0.4 M $(NH_4)2SO_4$	0.2 M $(NH_4)2SO_4$
C	1.5 M NH_4Cl	1.0 M NH_4Cl	0.8 M NH_4Cl	0.6 M NH_4Cl	0.4 M NH_4Cl	0.2 M NH_4Cl
D	0.8 M $NaSO_4$	0.6 M $NaSO_4$	0.4 M $NaSO_4$	0.2 M $NaSO_4$	0.5 M KCl	0.2 M KCl
E	0.4 M NaF	0.1 M NaF	0.4 M NaBr	0.1 M NaBr	0.4 M NaI	0.1 M NaI
F	5 mM $MnCl_2$	1 mM $CoCl_2$	1 mM $CuSO_4$	1 mM $CdSO_4$	5 mM EDTA pH8	5 mM EGTA pH 8.9
G	2 mM $EuCl_3$	2 mM $GdCl_3$	2 mM $DyCl_3$	2 mM $HoCl_3$	2 mM $YbCl_3$	2 mM $LuCl_3$
H	5 mM DTT	0.5 M Na_3 citrate	0.2 M Na_3 citrate	0.5 M trehalose	1 M trimethylglycine	0.5 M trimethylglycine

APPENDIX B. CHAPTER 6. COMPUTATIONAL AND INITIAL BIOPHYSICS RESULTS.

1. SDF FRAGMENT LIKE SUBSET.

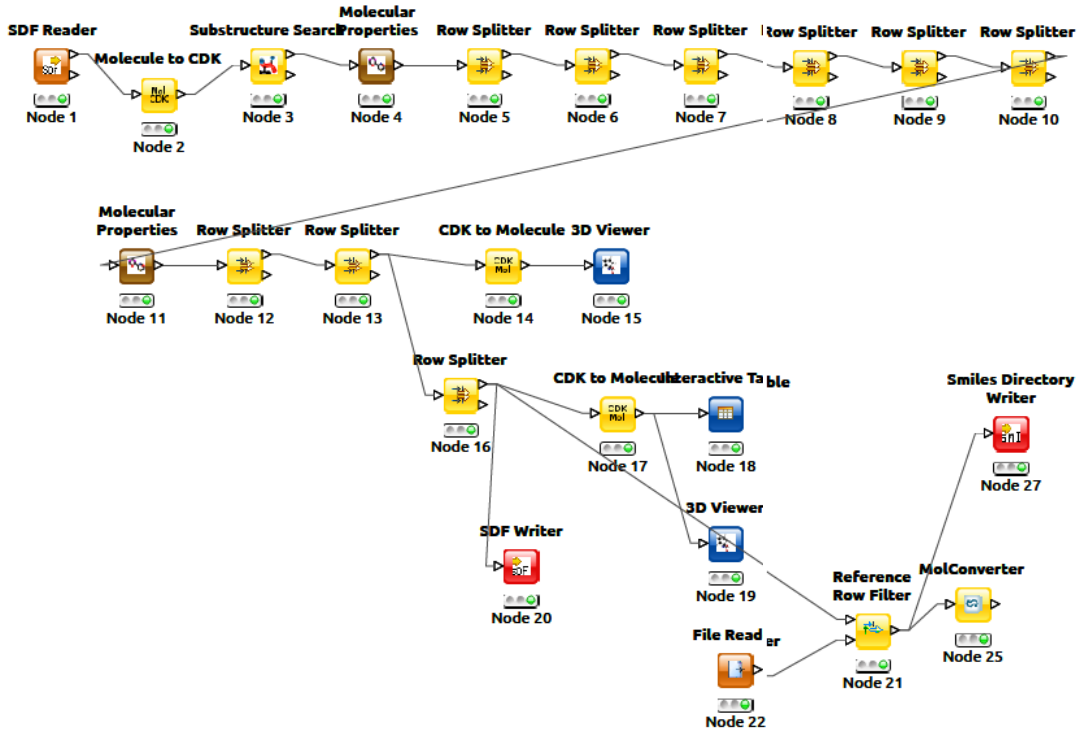
AN EXAMPLE WAS USED, THE SDF FILE CONTAINS OVER 1 000 000 COMPOUNDS AND COULD NOT BE SHOWN HERE. EACH COMPOUND HAS A DESCRIPTOR AS SHOWN BELOW.

EXAMPLE – BENZENE



2. KNIME

KNIME creates a clustered pipeline to filter from the initial SDF file obtained through ZINC. Orange nodes are SDF reader input, Red are SDF output and yellow nodes are different types of filters. Finally Brown nodes are usually structural filters and blue nodes are conversion filters. The first arrow on the left on each node accepts the information from the subset and previous nodes. The top arrow on the right of each node connects to the next node and carries the compounds that have fit the criteria in that node. The lower right arrows on the nodes are the rejected compounds, which can be checked against the criteria to ensure correct filtering.



3. HADDOCK CLUSTER TABLE

Parameter	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
HADDOCK Score	-195 ± 5.9	147.7 ± 6.5	98.3 ± 5.8	-79.7 ± 11.5
Cluster size	148	34	7	5
RMSD from Overall lowest energy structure / Å	0.5 ± 0.3	22.7 ± 0.1	11.4 ± 0.2	11.4 ± 0.0
Van der Waal Energy / kcal mol^{-1}	-48.3 ± 2.6	-81.3 ± 8.7	-46.6 ± 7.1	-33.9 ± 6.0
Electrostatic Energy / kcal mol^{-1}	-934.9 ± 76.8	-579.5 ± 16.6	-471.1 ± 14.6	-505.6 ± 19.3
Desolvation Energy / kcal mol^{-1}	25.4 ± 13.0	41.4 ± 4.3	39.5 ± 4.6	45.6 ± 7.9
Restraints Violation Energy / kcal mol^{-1}	145.1 ± 41.31	81.7 ± 26.99	110.1 ± 23.65	96.1 ± 31.69
Buried Surface Area / $Å^2$	145.1± 41.31	81.7 ± 26.99	110.1 ± 23.65	96.1 ± 31.69
Z-score	-1.4	-0.4	0.8	1.0

The Z-score shows the standard deviation from the average, so a more negative value is deemed the 'best'.

Cluster 1 also contained structures with most negative Z-scores, the lowest desolvation energy, (which indicates the best solvated models), and the lowest electrostatic energy and thus the best electrostatic contacts.

4. GOLD CONF FILE.

```
  GOLD CONFIGURATION FILE

  AUTOMATIC SETTINGS
autoscale = 0.3

  POPULATION
popsiz = auto
select_pressure = auto
n_islands = auto
maxops = auto
niche_siz = auto

  GENETIC OPERATORS
pt_crosswt = auto
allele_mutatewt = auto
migratewt = auto

  FLOOD FILL
radius = 8
origin = 4.0481 6.0364 -8.3025
do_cavity = 1
floodfill_atom_no = 0
cavity_file =
floodfill_center = point

  DATA FILES
ligand_data_file /home/haruna/Documents/KNIME/fragment like-knime.sdf 30
param_file = DEFAULT
set_ligand_atom_types = 1
set_protein_atom_types = 0
directory = /home/haruna/Documents/gold/GOLD 30% run
tordist_file = DEFAULT
make_subdirs = 0
save_lone_pairs = 1
fit_points_file = fit_pts.mol2
read_fitpts = 0

  FLAGS
internal_ligand_h_bonds = 1
flip_free_corners = 1
match_ring_templates = 1
flip_amide_bonds = 1
flip_planar_n = 1 flip_ring_NRR flip_ring_NHR
flip_pyramidal_n = 1
rotate_carboxylic_oh = flip
use_tordist = 1
postprocess_bonds = 1
rotatable_bond_override_file = DEFAULT
diverse_solutions = 1
divsol_rmsd = 1.5
divsol_cluster_size = 1
solvate_all = 1

  TERMINATION
early_termination = 1
n_top_solutions = 10
rms_tolerance = 1.5

  CONSTRAINTS
force_constraints = 0

  COVALENT BONDING
covalent = 0

  SAVE OPTIONS
save_score_in_file = 1
save_protein_torsions = 1

  FITNESS FUNCTION SETTINGS
initial_virtual_pt_match_max = 3
relative_ligand_energy = 1
gold_fitfunc_path = chemscore
score_param_file = /opt/goldsuite-5.2.2/GOLD/gold/chemscore.kinase.params

  PROTEIN DATA
protein_datafile = /home/haruna/Documents/gold/ID_protein.mol2
```

Search efficiency: 30%

General Algorithm Settings

Centre point coordinates x,y,z radius and the option to produce a cavity structure file

Source of data files

1= yes, 0 = no.

Ligand flexibility flags