Targeting the essential interactions of the M2-1 protein of human respiratory syncytial virus (HRSV) for anti-viral drug development

Rachel Linda Dods

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Molecular and Cellular Biology

August 2017

The candidate confirms that the work submitted is her own, except work which has formed part of jointly authored publications has been included. The contribution of the candidate at the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 3 is based on work contributed to a jointly authored publication: Tanner, S.J., Ariza, A., Richard, C.A., Kyle, H.F., Dods, R.L., Blondot, M.L., Wu, W., Trincão, J., Trinh, C.H., Hiscox, J.A. and Carroll, M.W., 2014. Crystal structure of the essential transcription antiterminator M2-1 protein of human respiratory syncytial virus and implications of its phosphorylation. *Proceedings of the National Academy of Sciences, 111*(4), pp.1580-1585. Mass spectrometry and negative stain EM were performed as a service by the corresponding departments at the University of Leeds. The GST pull down assay was performed by Dr Kavestri Yegambaram and based on data (personal communication) by Dr Jean-François Eléouèt. X-ray diffraction data was jointly performed with Dr Yegambaram with the best resolution data set collected by Dr Yegambaram.

Chapter 4 was part of a collaborative project with Dr Richard Foster and Dr Roger Taylor, Department of Chemistry, University of Leeds. Dr Richard Foster performed the initial computational docking. Dr Anthony Marriott and Dr Miles Carroll (Public Health England) performed initial plaque assay of compounds selected by Dr Richard Foster. Dr Roger Taylor performed all further SAR and Dr Rachel Trowbridge cherry picked all compounds assessed.

This copy has been supplied on the understanding that it is copyright material and that no quotation from this thesis may be published without proper acknowledgement.

The right of Rachel Linda Dods to be identified as author of this work has been asserted by her in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgments

Abstract

Human respiratory syncytial virus (HRSV) is a non-segmented negative stranded RNA virus classified within the *Mononegavirales* order. HRSV is the leading cause of lower respiratory tract illness in young children and the immunocompromised, with over 250,000 annual fatalities worldwide. HRSV-mediated diseases are especially prevalent in developing countries, where no financially viable treatment exists. The M2-1 protein of HRSV represents a promising potential anti-viral target for the treatment of HRSV-mediated diseases. M2-1 is a transcription anti-terminator with an essential role in viral gene expression. M2-1 binds both viral RNA and the polymerase co-factor phosphoprotein (P), and these interactions are essential for its anti-termination activity. Here, the crystal structure of M2-1 in complex with a P protein peptide was determined and used in addition to the unbound M2-1 crystal structure as a basis to identify potential anti-viral compounds. Structure based drug design (SBDD) of M2-1s interaction interface with RNA/ P produced hit compounds that were analysed through biophysical, structural and *in cellulo* methods including the use of a mini-genome and infectious HRSV assay. SBDD identified hit compounds that significantly inhibited the growth of HRSV. Subsequently, synthetic chemistry was used to generate libraries of molecules to establish structure-activity relationships (SAR) as well as assess their pharmacokinetic properties. Further, a novel facility that allowed assessment of fragment binders to HRSV M2-1 by X-ray crystallography was utilised to determine hit binders towards the entire M2-1 protein, including the previously targeted RNA/ P surface and allowed further rapid SAR. The work presented here represents an effective strategy to rationally design anti-viral compounds for the M2-1 protein of HRSV with potential applications for the related virus, bovine respiratory syncytial virus (BRSV).

Table of contents

List of Tables

List of Figures

Abbreviations

ACE	Apical recycling endosomes
ADMET	Absorption, distribution, metabolism, excretion, toxicity
AMP	Adenosine monophosphate
AMPV	Avian metapneumovirus
ARE	Apical recycling endosomes
AUC	Asymmetric unit cell
BHV-1	Bovine herpesvirus-1
BPIV-3	Bovine parainfluenza virus
BRSV	Bovine respiratory syncytial virus
BVDV-1/ 2	Bovine viral diarrhoea virus
CC	Coiled coil domain
CD	Circular dichroism
CDV	Canine distemper virus
CeMV	Cetacean morbillivirus
CK	Casein kinase
CME	Clathrin-mediated endocytosis
CTD	C-terminal domain
CV	Column volume
DEPC	Diethylpyrocarbonate
DHBNH	Dihydroxy benzoyl naphthyl hydrazine
DLS	Diamond light source
EBOV	Ebola virus
eGFP	Enhanced green fluorescent protein
eHITS	Electronic high-throughput screening
EM	Electron microscopy
ESCRT	Endosomal sorting complexes for transport
F	Fusion protein
FA	Fluorescence anisotropy
FBDD	Fragment based drug design
FDA	Food and drug administration

FeMV	Feline morbillivirus
Fl	Fluorescent
FP	Forward primer
G	Glycoprotein
GCU	Green calibrated unit
GJ	Gene junction
GST	Glutathione S transferase
HAART	Highly active antiretroviral therapy
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HMPV	Human metapneumovirus
HPIV	Human parainfluenza virus 1/ 2/ 3/ 4a/ 4b
hr	Hour
HRSV	Human respiratory syncytial virus
HSPG	Heparin sulphate proteoglycans
HTS	High throughput screening
IBs	Inclusion bodies
IGR	Intergenic region
IL	Interleukin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kd	Dissociation constant
L	Large protein
LB	Lysogeny broth
Le	Leader sequence
LRTI	Lower respiratory tract infection
M	Matrix protein
M2-1	Matrix-like protein 1
M2-2	Matrix-like protein 2
MAPV	Mapuera virus
MoRe	Molecular recognition site
MPV	Murine pneumonia virus
MuV	Mumps virus
MV	Measles virus

MW	Molecular weight
MWCO	Molecular weight cut off
N	Nucleoprotein
NMR	Nuclear magnetic resonance spectroscopy
NS1/ 2	Non structural protein 1/ 2
NTD	N- terminal domain
nts	Nucleotides
OD	Oligomerisation domain
ORF	Open reading frame
PanDDA	Pan-dataset density analysis
PCR	Polymerase chain reaction
PDB	Protein data bank
PDV	Phocine distemper virus
PHE	Public health england
pI	Isoelectric point
PIV-5	Parainfluenza virus 5
PPI	Protein- protein interaction
PPIV-1	Porcine parainfluenza virus-1
PPRV	Peste-des-petits-ruminants virus
RdRp	RNA dependent RNA polymerase
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RP	Reverse primer
RPV	Rinderpest virus
RT	Reverse transcriptase
RV	Rabies virus
SAR	Structure- activity relationship
SAXS	Small-angle X-ray scattering
SBDD	Structure based drug design
SDM	Site directed mutagenesis
SEC	Size exclusion chromatography
SENV	Sendai virus
SGC	Structural genomics consortium

SH	Small hydrophobic protein
siRNA	Small interfering RNA
SV-41	Simian virus 41
Tm	Melting temperature
TMAO	Trimethylamine N-oxide dehydrate
Tr	Trailer
TrC	Trailer complement
VHH	Variable domain of heavy chain antibodies
Vo	Void volume
VSV	Vesicular stomatitis virus
WT	Wild type
XCE	XChem explorer
ZBD	Zinc binding domain

Chapter 1 Introduction

1.1   HRSV Classification & Discovery

Human respiratory syncytial virus (HRSV) was first isolated in 1956 in chimpanzees (Chanock et al., 1957) and shortly after was isolated in children in the USA. From its initial discovery 60 years ago, HRSV is now regarded as the main cause of lower respiratory tract infections (LRTI) in infants. HRSV is classified within the *Mononegavirales* order, which was established in 1991 (Pringle, 1991) to accommodate single stranded, negative sense non-segmented RNA viruses. The *Mononegavirales* order contains many high-profile pathogens of humans, animals and plants, including Ebola virus (EBOV) (Filoviridae family), rabies virus (RV) (*Rhabdoviridae* family) and Sendai (SENV), mumps (MuV) and measles viruses (MV) of the *Paramyxovirinae* family. Recently, the *Pneumovirus* genus that includes HRSV and human metapneumovirus (HMPV) has undergone re-classification; the previous *Pneumovirus* genus has now been elevated in status to represent the family *Pneumoviridae*, and the *Pneumovirus* genus that included HRSV (Table 1) was renamed *Orthopneumovirus* to avoid confusion between family and genus taxa (Afonso et al., 2016).

Family	Genus	Species	Virus
Paramyxoviridae	*Morbillivirus*	*Canine distemper virus*	canine distemper virus (CDV)
		Cetacean morbillivirus	cetacean morbillivirus (CeMV)
		Feline morbillivirus	feline morbillivirus (FeMV)
		Measles virus	measles virus (MeV)
		Peste-des-petits-ruminants virus	peste-des-petits-ruminants virus (PPRV)
		Phocine distemper virus	phocine distemper virus (PDV)
		Rinderpest virus	rinderpest virus (RPV)

Family	Genus	Species	Virus
	Respirovirus	*Bovine parainfluenza virus 3*	bovine parainfluenza virus 3 (BPIV-3)
		Human parainfluenza virus 1	human parainfluenza virus 1 (HPIV-1)
		Human parainfluenza virus 3	human parainfluenza virus 3 (HPIV-3)
		Porcine parainfluenza virus 1	porcine parainfluenza virus 1 (PPIV-1)
		Sendai virus	Sendai virus (SENV)

Family	Genus	Species	Virus
	Rubulavirus	*Human parainfluenza virus 2*	human parainfluenza virus 2 (HPIV-2)
		Human parainfluenza virus 4	human parainfluenza virus 4a (HPIV-4a)
			human parainfluenza virus 4b (HPIV-4b)
		Mapuera virus	mapuera virus (MapV)
		Mumps virus	mumps virus (MuV)
			bat mumps virus (BMuV)
		Parainfluenza virus 5	parainfluenza virus 5 (PIV-5)
		Simian virus 41	simian virus 41 (SV-41)

Family	Genus	Species	Virus
Pneumoviridae	*Metapneumovirus*	*Avian metapneumovirus*	avian metapneumovirus (AMPV)
		Human metapneumovirus	human metapneumovirus (HMPV)

3

Family	Genus	Species	Virus
	Orthopneumovirus	*Bovine respiratory syncytial virus*	bovine respiratory syncytial virus (BRSV)
		Human respiratory syncytial virus	human respiratory syncytial virus A2 (HRSV-A2)
			human respiratory syncytial virus B1 (HRSV-B1)
			human respiratory syncytial virus S2 (HRSV-S2)
		Murine pneumonia virus	murine pneumonia virus (MPV)

Table 1 Tabulation of members of the *Mononegavirales* order, *Paramyxoviridae and Pneumoviridae* families

A full taxonomy of the Mononegavirales order can be found in Afonso et al, (2016).

1.2   HRSV epidemiology and symptoms

HRSV is extremely contagious and is found in large quantities in respiratory secretions. It is transmitted via direct close contact with respiratory droplets and contact with fomites (Hall & Douglas, 1981). HRSV epidemiology is characterised by a peak in hospitalisations in winter months and in the UK, the annual epidemic peak of infections is late December to early January (White et al., 2005). Virtually all children have been infected with HRSV by the age of 2, with non-antigen variation re-infection through the host's lifetime a common event due to evasion of IgA B cell memory (Glezen et al., 1986; reviewed by Griffiths et al., 2017) and incomplete or short-lived protective immunity (Sande et al., 2014). The highest mortality rates are associated with infants under 6 months of age, where infection of the bronchiolar and alveolar epithelium results in acute lower respiratory tract disease such as bronchiolitis and pneumonia and is associated with a 'wheezing' sound during breathing. Between 66,000-199,000 children under 5 years of age are estimated to die from HRSV associated acute lower respiratory infections yearly, with 99 % of mortality occurring in developing countries (Nair et al., 2010). In developed countries, HRSV associated mortality is low in infants and mortality is usually associated with a second underlying condition (Byington et al., 2015). HRSV mortality in the US accounts for 11,000 -17,000 deaths per year with 10 times this amount suggested to require hospitalisation or treatment, resulting in large costs to the healthcare system. These values are likely underestimations as in cases of pneumonia and LRTI hospitalisation, HRSV is non-discriminate in its symptoms from other severe respiratory viruses and is usually undetected as the cause of mortality and hospitalisation (Han et al., 1999). In the general adult population, HRSV infection re-occurs throughout life with only mild to moderate symptoms of the 'common cold' (rhinorrhoea, fever and cough), however an increase in disease severity and mortality is seen in the immunocompromised, transplant recipients and elderly cohorts.

HRSV infection and subsequent hospitalisation in infants has been linked to chronic wheezing and asthma in adulthood. IL-33 is known to signal downstream cytokines (IL-4, IL-5 and IL-13) that are implicated in asthma genesis. In a mouse model, HRSV infection induced high levels of IL-33 expression in neonates but not adult mice, with the levels of IL-33 determining the severity of asthma in adulthood (Saravia et al., 2015). In addition, bronchial epithelial cells, the site of HRSV infection that results in LRTI, express high levels

of IL-33 and poses a credible link between HRSV infection and development of asthma in adulthood (Schmitz et al., 2005).

1.3   HRSV genome

The HRSV genome (A2 strain) is composed of a single strand of negative sense RNA approximately 15.2 Kb in length (Collins et al., 1995). The HRSV genome consists of 10 genes (Collins et al., 1984), each expressing a single mRNA and together encoding a total of 11 proteins (Table 2). The *Pneumoviridae* family that includes HRSV, HMPV and BRSV is the only family of the *Mononegavirales* order to encode an M2 gene (Figure 1). The M2 gene expresses an mRNA with two open reading frames (ORFs) that encodes two proteins, M2-1 and M2-2.
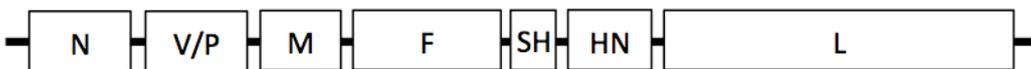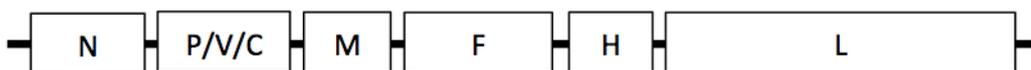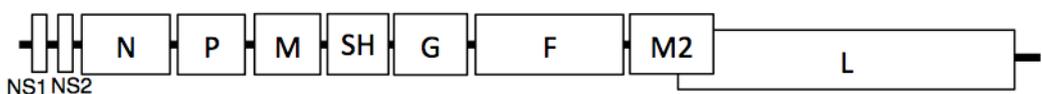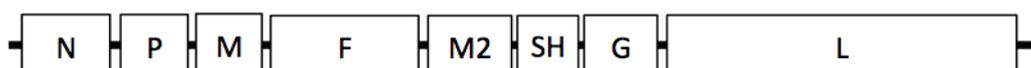
Figure 1 Genome schematic

> Schematic of the genomic organisation of members of the Paramyxovirinae and Pneumoviridae families. All genomes are represented in a 3' to 5' orientation.

Protein Acronym	Protein	Description
NS1	Non-structural protein 1	Immune evasion
NS2	Non-structural protein 2	Immune evasion
N	Nucleoprotein	Encapsidates viral RNA, prevents viral genome degradation
P	Phosphoprotein	Polymerase co-factor, interacts with L, N and M2-1 for roles in both replication and transcription
M	Matrix protein	Role in virion assembly and filamentous virion particle elongation
SH	Small hydrophobic protein	Viroporin
G	Glycoprotein	Attachment protein of virus particle to cell
F	Fusion protein	Fusion of viral and cellular membranes and syncytia formation *in cellulo*
M2-1	Matrix-like protein 1	Anti-terminator, synthesis of full mRNA transcripts and polycistronic read-though mRNAs, allowing RdRp access to distal genes
M2-2	Matrix-like protein 2	Regulation of transcription and replication
L	Large protein	RNA dependent RNA polymerase (RdRp)

Table 2 HRSV proteins

HRSV encodes 11 proteins, their full-name, acronym and function is described.

1.4 Non-structural HRSV proteins (NS1 and NS2)

NS1 and NS2 are the first 2 genes in the HRSV gene order, and their position close to the 3' end of the genome results in their corresponding mRNAs being abundantly transcribed in accordance with the transcriptional programme of HRSV (Fearns & Collins, 1999). Expression of NS proteins are defining characteristics of all viruses classified within the *Orthopneumovirus* genus, with no other members of the *Mononegavirales* Order expressing NS1 or NS2 homologs. Neither NS1 nor NS2 are essential for virus growth, as gene deleted variants can be rescued as infectious virus. However, these viruses are severely growth attenuated in both continuous and primary cell cultures (Buchholz et al., 1999) and have been trialled as live attenuated vaccine candidates (Whitehead et al., 1999).

NS1 and NS2 of HRSV are 139 and 124 amino acids in length respectively with a 69% sequence identity between NS1 of HRSV A strain and BRSV whilst NS2 had an 84% identity between HRSV A strain and BRSV (Pastey & Samal, 1995). NS1 and 2 of HRSV are multifunctional proteins (Bermingham & Collins. 1999) with roles in innate immune evasion. The NS proteins are potent antagonists of the IFN α/ β mediated host cell response and expression of both allows maximal antagonism via complementary mechanisms (Schlender et al., 2000). Other *Paramyxoviridae* family members such as MV antagonise interferon signalling via their accessory proteins C and V (reviewed in Gotoh et al., 2001). BRSV NS1 and NS2 were found to more potently antagonise the IFN response in bovine cells (MDBK) than Vero cells (African Green Monkey) (Schlender et al., 2000) whilst HRSV NS proteins were found to potently inhibit IFN in a human lung epithelia (A549) and fibroblast (MRC-5) cell lines (Atreya & Kulkarni, 1999) and suggest a species specificity of this host cell antiviral response.

NS2 is involved in proteosomal degradation of the STAT2 signal transduction factor and results in the suppression of the IFN (type I) pathway (Ramaswamy et al., 2006). NS1 is also thought to target STAT2 to the protesome using ubiquitin ligase enzymes (Elliott et al., 2007). IFN type III pathways are also inhibited in HRSV infection likely due to using the same STAT factors for downstream signalling (Spann et al., 2004). NS1 and NS2 deletion mutants illustrated that NS1 was responsible for suppressing CD103+ and CD8+ T cells activation and proliferation and supressing Th17 cells, both involved in a protective role during infection and promoted Th2 cell activation that enhance associated disease (Munir et al., 2011). NS1 has been identified in inefficient migration of human

dendritic cells to the lymphatic tissue by supressing human dendritic cell maturation and contributes to the decrease in adaptive response (Le Nouën et al., 2011). Cytoplasmic receptors that have been involved in binding the NS proteins to inhibit downstream signalling events and evade immune response pathways include; RIG-1 binding to NS2 (Ling et al., 2009), IKKε binding to NS1 and TRAF3 interactions with both NS1 and NS2 (Swedan et al., 2009).

1.5 HRSV life cycle

The primary target of HRSV are epithelial cells lining the respiratory tract. The virus enters via the apical surface, and following synthesis of viral components, the assembled progeny virus also leave the cell via the apical route, to infect neighbouring cells. All stages and viral components involved in this cycle are described below and summarised as a schematic in Figure 2.
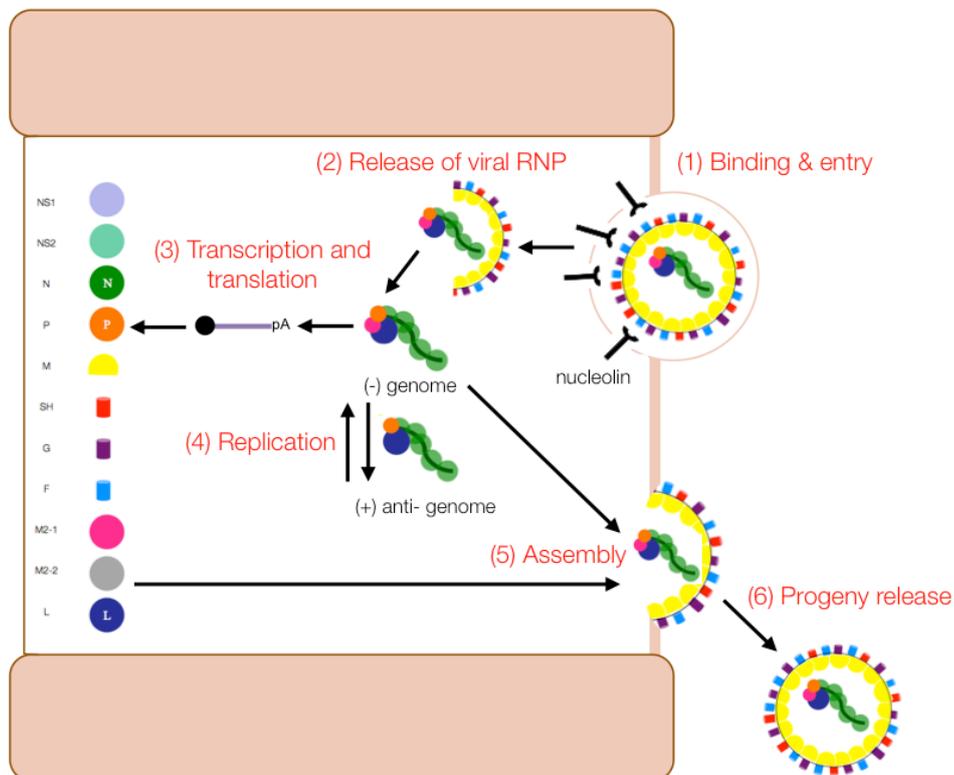
Figure 2 Schematic of the HRSV life cycle

(1) HRSV attaches (via G) to the apical cell surface of epithelial cells and binds (via F) to the nucleolin receptors, conformational change of F from pre-post fusion results in fusion of viral and cell membranes; (2) the virus is endocytosed (assumed by macropinocytosis) and the RNP is released into the cytoplasm; (3) on entering the cell, HRSV first enters a stage of transcription and translation of the 11 proteins; (4) once sufficient protein levels are acquired HRSV then begins to replicate its genome (-) via an anti-genome (+) intermediate; (5) the M protein directs viral assembly at the apical cell surface with F, G and SH inserted into the cell membrane used for the virions membrane and encases the vRNP; (6) progeny virions are released.

1.5.1 Virion

Similar to other *Paramyxoviruses*, HRSV virions are both enveloped and pleomorphic (Bächi & Howe, 1973) being visualised in both spherical (100 nm- 1 μm) and filamentous (200 nm- 2 μm; width 70- 190 nm) forms by cryo-EM (Figure 3). The role or relevance of the two infectious forms of HRSV in the virus life cycle is poorly understood, although virions with filamentous morphology appear to exhibit reduced infectivity (Liljeroos et al., 2013).

The viral envelope is derived from the host cell plasma membrane and contains three virus-encoded surface glycoproteins namely the fusion (F) protein, the attachment glycoprotein (G) and the small hydrophobic (SH) protein that do not influence the morphology of the virus. The envelope surrounds an inner layer of matrix (M) protein that also interacts with the cytoplasmic tails of the envelope glycoproteins. The matrix layer is also thought to interact with the internal ribonucleoprotein (RNP) component of the virion, which comprises the RNA genome encapsidated by the nucleocapsid protein (N). The virion also contains an RNA dependent RNA polymerase (RdRp), comprising the phosphoprotein (P) and the large (L) protein. This polymerase must be transcriptionally active when the virus enters a new cell in order to initiate viral gene expression, and so it likely also includes the M2-1 protein that acts as an essential polymerase co-factor.
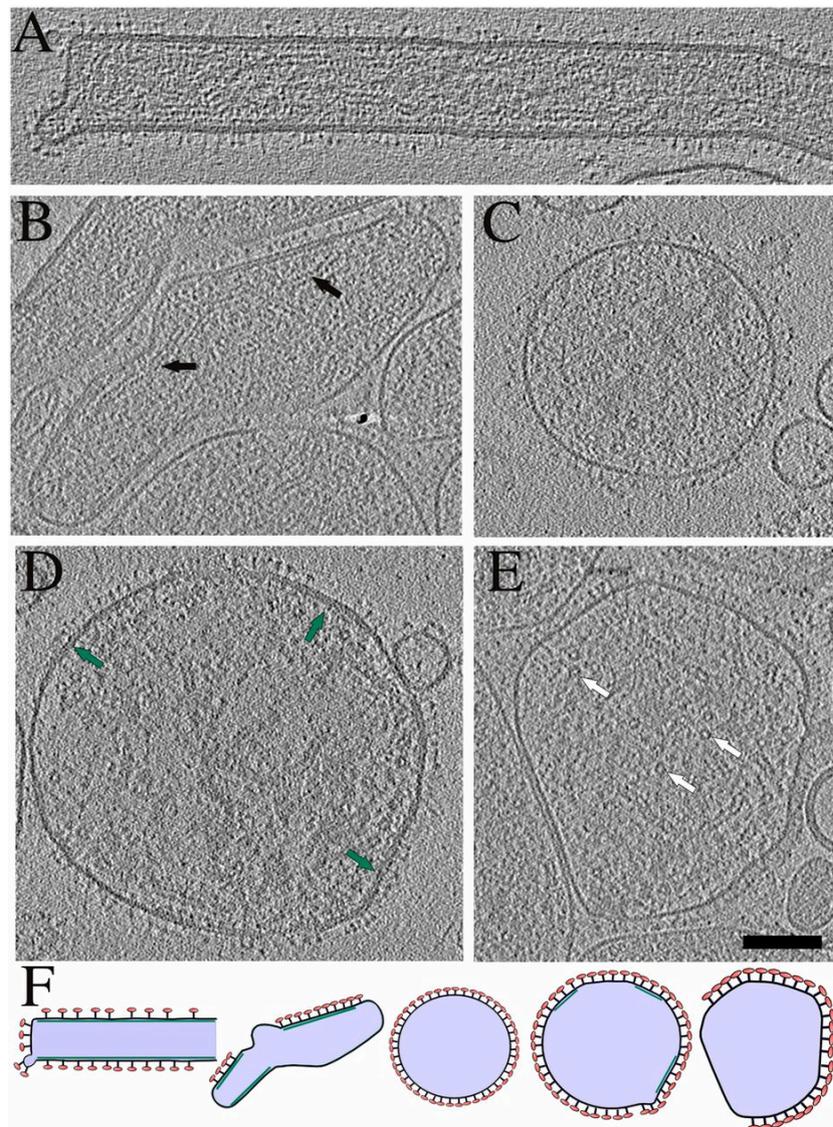
Figure 3 HRSV virion morphology

HRSV virion morphology determined by cryo-EM (Liljeroos et al., 2013). A-filamentous virion; B- intermediate virion morphology displaying both filamentous and spherical structures; C- spherical virion; D as B; E-deformed spherical virion in close proximity to other virus particles (scale bar, 100 nM). White and black arrows indicate RNPs whilst green arrows indicate a second layer of density under the viral membrane, assumed to be M2-1; F- schematic representation of virions displayed in images A-E.

1.5.2 Viral entry

The mechanism of HRSV entry is currently unclear (reviewed by Chang & Dutch, 2012). Both clathrin-mediated endocytosis (CME) and macropinocytosis have been postulated as mechanisms for HRSV entry, with evidence in the literature for both. In support of CME, a siRNA library targeting CME-associated genes such as *Pab5A* and clathrin light chain *AP1B1* resulted in decreased infection (Kolokoltsov et al., 2007) whereas inhibition of clathrin and caveolin- dependent endocytosis through the use of established small molecule inhibitors blocked the uptake of HRSV cell surface antigen-antibody complexes (Ortega et al., 2008). Furthermore, HRSV and the closely related HMPV were found to use endosomal pathways to enter cells (Kolokoltsov et al., 2007; Palgen et al., 2015) implicating CME as the entry pathway. More recently macropinocytosis has been proposed as a mechanism for HRSV entry with the virus being shown to internalise by an actin-dependent process, with actin found in large quantities within purified virions (Ulloa et al., 1999; Kallewaard et al., 2005). In further support of a macropinocytosis mechanism, the virus was endocytosed and entered Rab5 positive fluid-filled macropinosomes before fusing with the macropinosome membrane after internalisation. This mechanism of viral entry was shown to occur in polarised human epithelial cells (Krzyzaniak et al., 2013), the primary site of HRSV infection. Further, the related Nipah and EBOV viruses have been postulated to enter cells via micropinocytosis, as compounds that inhibited macropinocytosis also inhibited viral entry (Pernet et al., 2009).

The trans-membrane surface proteins implicated in HRSV entry comprise the fusion glycoprotein (F), the attachment glycoprotein (G) and the small hydrophobic (SH) protein, and are discussed below.

1.5.2.1 F Protein

The activity of a fusion protein (F) is a requirement for all enveloped viruses in order to mediate fusion between the viral envelope with a cellular membrane, and allow virus entry. Nucleolin is the only determined receptor to bind HRSV F and is located at the apical cell surface of epithelial cells, the site of HRSV entry (Tayaari et al., 2011). Nucleolin has also been identified as a host cell receptor of human parainfluenza virus type 3 (HPIV-3) (Bose et al., 2004) and Group B Coxsackieviruses (DeVerdigo et al., 1995) and thus could be an entry receptor for multiple viruses, however nucleolin was not identified as a cell receptor for

SENV raising the possibility of alternative receptors for viral entry (Fujiki et al., 2014).

After HRSV F binds to its receptor on the target cell, the protein undergoes an irreversible structural rearrangement from a pre- to post-fusion conformational state following viral-cell fusion. The pre-fusion state of F (F_0) requires independent cleavage of two furin sites into two disulphide-linked subunits (F_1 and F_2) to allow the F protein to mediate fusion (Gonzales-Reyes et al., 2001). All other fusion proteins from the *Paramyxoviridae* family only require protease cleavage at one site. For HRSV, the second furin site must be cleaved to release pep27 and results in an infectious and fusion competent virus (Krzyzaniack et al., 2013). X-ray structure determination of post-fusion F (McLellan et al., 2011; Swanson et al., 2011) (Figure 4) described a cone shaped trimer with a six-helix bundle core formed by three heptad repeats (a structural motif that forms the basis of coiled coils) with three anti-parallel heptad repeats that allow pH independent (Zhao et al., 2000) fusion by pulling cell and viral membranes to close proximity. Comparison with related viruses, PIV3 and NDV revealed high structural similarities of their respective F proteins despite low sequence identity (McLellan et al., 2011) and suggested a similar mechanism for these viruses to allow cell fusion and viral entry.
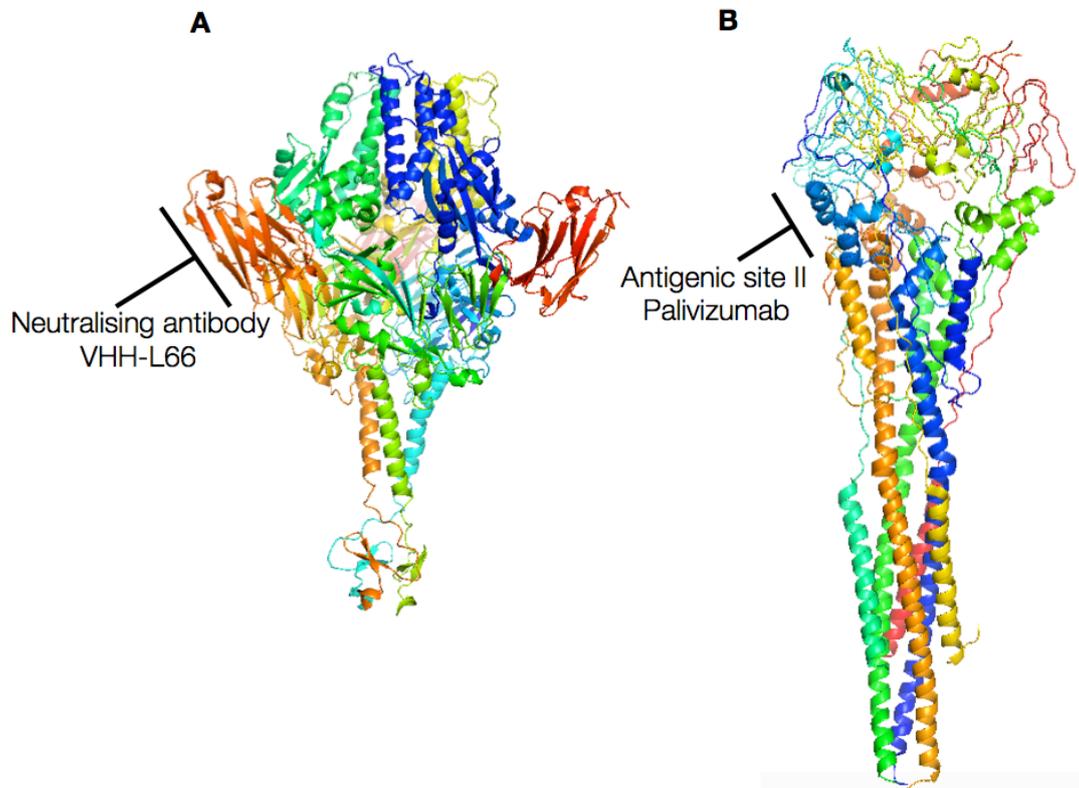
Figure 4 Structure of HRSV pre and post fusion F

> F protein structures were determined by X-ray crystallography (pre-fusion-Rossey et al., 2017; post-fusion- McLellan et al., 2011). A- a llama derived single domain antibody (VHH) neutralised HRSV by binding to F and maintaining its pre-fusion conformation, the pre-fusion form of F can be described as a 'spring-loaded' trimer; B- post-fusion F is characterised by a six helical bundle, with three heptad and antiparallel heptad repeats. The cone shaped trimer head contains antigenic site II, the binding target of the only HRSV-specific treatment palivizumab and is further discussed in section 1.9.

1.5.2.2 G protein

The HRSV attachment protein (G) is involved in attachment of HRSV to ciliated cells within the nasopharyngeal epithelium (Levine et al., 1987). HRSV G has a low level of sequence identity with other *Paramyxoviridae* family attachment proteins. HRSV G is heavily glycosylated with N- and O-linked oligosaccharides (Wertz et al., 1985) and removal of these oligosaccharides resulted in a 97% reduction in infection *in vitro* (Lambert, 1988) and suggested that G may bind to carbohydrate binding proteins on host cells.

Studies found G to bind heparin sulphate proteoglycans (HSPG) (Krusat & Sreckert, 1997) but these findings were at odds with the observation that HSPGs were not present in human epithelial cells, the primary site of HRSV infection and suggest that G does not bind HSPG for its attachment function. Annexin II has been implicated in binding the G protein, with recombinant annexin II decreasing HRSV infection in Hep2 cells (Malhorta et al., 2003). Recent studies in which virus attachment and infectivity was examined using differentiated human airway epithelial cells identified the chemokine receptor CX3CR1 as the receptor recognised by G (Johnson et al., 2015; Jeong et al., 2015; Chirkova et al., 2015).

Members of the *Paramyxoviridae* family, such as MV require G as a prerequisite to signal for fusigenic activity of the F protein (Nauaratnarajah et al., 2011) with the virus and target membrane, however HRSV G is dispensable for infection *in vitro*, revealed by the ability of recombinant HRSV lacking the G gene to infect cells in culture (Teng et al., 2001). Due to HSPGs being expressed in the aforementioned assay but not in human epithelial cells, the site of HRSV infection, the validity of this study is questionable. A recent study has shown a higher dependency on G than previously thought, determined by strain specificity of F (Meng et al., 2016).

The G protein also exists as an N-terminally truncated form that is secreted by epithelial cells (Hendricks et al 1987). Soluble forms of the attachment protein have also been recorded for VSV and showed the soluble, truncated form of G to be functional and not a by-product of virion degradation (Little & Huang, 1977). The soluble form of G is thought to play a role in immune evasion by inhibiting TLR3/4 mediated IFN beta induction (Shingai et al., 2008).

1.5.2.3 SH Protein

The role of the small hydrophobic (SH) protein (64 amino acids HRSV A2 strain) in the virus life cycle is unclear. It is non-essential for virus replication as infectious virus can be recovered following SH gene deletion, although the resulting virus shows partial growth attenuation in certain cell types and small animal models (Bukreyev et al., 1997; Jin et al., 2000; Russell et al., 2015). SH-deleted HRSV also induces altered innate immune signalling characteristics, in particular increased apoptosis (Fuentes et al, 2007). SH has been suggested to be a viroporin (Perez et al., 1997), causing membrane permeability and acting as an ion channel. Mutation of the basic His residues 22 and 51 of SH (to Ala)

abolished channel activity at low pH (5.5) suggesting pH dependence for channel activity. A requirement of His residues has also been reported for the viroporin of Influenza A, known as M2 (Schnell & Chou, 2008).

The SH protein (with attached FLAG tag) was determined by negative stain EM to display distinct 5 and 6-fold symmetry consistent with a pentamer or hexamer (Carter et al., 2010). A model was hypothesised in which SH was a hexamer in solution with a central void 'pore'. The SH protein (cleaved from His-MBP tag) was further assessed by solution NMR (Gan et al., 2012) and determined in multiple detergents to form pentamers and is most likely that natural oligomeric state of SH due to absence of expression tags during structural analysis. Each SH protomer was formed of a single transmembrane α helix, flanked by an α helix at the N-terminal domain (NTD) that forms a ring around the pore lumen and an extended beta turn at the C-terminal domain (CTD). The HRSV SH protein function is similar to that of the M2 protein of influenza A (Gonzalez & Carrasco, 2003), however M2 forms tetramers in solution as determined by NMR (Schnell & Chou, 2008) showing that these viroporins display different oligomeric states in different viruses.

1.5.3 Replication & Transcription

The HRSV RdRp is proposed to enter the genome template at promoter sequences at its 3' end for both HRSV genome replication and transcription functions (Fearns & Collins, 1999). For transcription, the HRSV polymerase comprises L, P and also M2-1 proteins. This transcriptase migrates along the genome template in the 3' to 5' direction and transcribes each gene individually to synthesise a total of 10 mRNA transcripts that are both 5' capped and 3' polyadenylated. In contrast, genome replication requires a different RdRp that comprises only L and P proteins. This replicase copies the template strand into a single contiguous complementary RNA molecule called an anti-genome that is encapsidated in N protein to form a ribonucleocapsid (RNP). The initial RNA synthesis activity performed by these polymerase-associated proteins is transcription, with viral mRNAs being detected in the cytoplasm 4-6 hours post infection (Collins & Crowe, 2007). Once HRSV proteins are in abundance the polymerase switches to HRSV genome replication that also requires a supply of N protein, although the precise events that mediate this switch in activity are poorly understood. The anti-genome RNP is then used as a template for further genome RNP synthesis.

The HRSV genome (Figure 5) contains non-coding regions at the 3' and 5' termini termed the leader (le) and trailer (tr) sequence respectively. Each HRSV gene is flanked by a conserved gene start (GS) and semi-conserved gene end (GE) sequence, which are separated by a non-conserved intergenic region (Collins et al., 1986). The *cis*-acting conserved and semi-conserved GS and GE sequences play a role in transcription initiation (Kuo et al., 1995) and transcription termination/ polyadenylation respectively (Kuo et al., 1996; Fearns & Collins, 1999). The GS and GE sequences are separated by an intergenic region (IGR), and together these three elements make up the gene junction (GJ). The diverse HRSV GJs were shown to respond differently to M2-1, however in all cases the addition of M2-1 increased the abundance of read-through transcripts with the F/M2 GJ being most sensitive to the presence of M2-1 (Hardy et al., 1999).

Figure 5 HRSV genome

> The HRSV genome consists of 10 genes with the GS shown as green, GE as red and the IGR as grey.

The IGRs vary in length from 1- 52 nucleotides and are not conserved in sequence or length except for one 'A' nucleotide at the last nucleotide position that immediately precedes the conserved GS (Collins et al., 1986; Kuo et al., 1995). The IGR of VSV plays a role in termination and initiation of transcription (Stillman & Whitt, 1997; Barr et al., 1997). However, there seems to be no defined role for the non-conserved IGRs in HRSV (Kuo et al., 1996). Transcription of the 10 HRSV genes occurs in a sequential manner from the 3' to 5' end, with termination of transcription of an upstream gene being required for initiation of transcription of a downstream gene. Transcripts generated from the 3' end of the genome are more abundant that those transcribed from the 5' end (Dickens et al., 1984), and this phenomenon is due to localised transcription attenuation that is thought to occur as the polymerase crosses the GJ. Transcripts from genes located at the 3' end such as the NS mRNA are transcribed in large quantities, probably to allow effective inhibition of innate immune response pathways whilst L at the 5' end is produced in minimal amounts due to its catalytic role in RNA polymerisation. However, sufficient levels of L exist for its polymerase function allowing for transcriptional economy

(Fearns & Collins, 1999; Barik, 2004). The rate of polymerase attenuation has been measured in the closely related VSV to be 21% per gene junction (Barr et al., 1997), and is thought to be similar in HRSV.

The 9th and 10th genes of HRSV (M2 and L) overlap such that the M2 GE sequence lies downstream of the L GS (Figure 5). Thus, in order for a polymerase to transcribe the L mRNA, it must first terminate transcription of the M2 mRNA at the M2 GE, and then track backwards to reach the L GS, located 68 nucleotides upstream (Collins et al., 1987). Once initiation has occurred at this GS, the polymerase must then ignore the proximal M2 GE by reading through this signal, and progressing downstream to the L GE. As for all other HRSV genes, the M2 GE sequence is an absolute requirement for transcription initiation of the L mRNA, whereas the sequence between the M2 GE and L GS is not conserved and does not play a role in modulating transcription efficiency (Fearns & Collins, 1999, Kuo et al 1996). Once the polymerase has recognised the M2 GE and released the nascent mRNA, the polymerase is able to move bi-directionally to find the next GS sequence in a process known as 'scanning'. Such scanning has been suggested to occur at all gene start and ends, regardless of any overlap, and there is direct evidence for this in the related VSV (Barr et al., 2008). Other viruses of the *Mononegavirales* order, such as EBOV, contain overlapping genes and thus bi-directional scanning for transcription start sites may be a conserved mechanism for this virus order (Fearns & Collins, 1999).

Unlike the divergent IGR sequences, the gene end sequences are semi-conserved in HRSV. Mutagenesis of the F and M gene ends has been performed to determine nucleotide specificity (Sutherland et al., 2001; Harmon et al., 2001) and addition of M2-1 resulted in an increase in anti-termination activity showing that M2-1 recognises the gene end sequences for efficient termination activity. Gene ends were determined to be 13 nucleotides in length with a conserved 3'-UCAAU-5' motif at nucleotides 1-5, followed by non-conserved nucleotides 6-8. The importance of nucleotides 6-8 has been disputed, with different studies focussing on different gene ends reporting inconsistent outcomes. In one study this region could not withstand a substitution to C however, for the M gene end nucleotide 6 had to be a U for efficient transcription termination (Harmon et al., 2001). In contrast, others reported that a U at position 6 was not required for transcription termination efficiency in the context of the F gene end (Sutherland et al., 2001). Gene end nucleotides 9-12/13 comprise of a conserved poly U tract of between 4 and 7

residues, with the general observation that increased U-tract length correlated with increased termination ability.

1.5.3.1 N Protein

The N protein encapsidates genomic and anti-genomic viral RNA and allows RNase resistance and escape from innate immunity (Vreede et al., 2004) in all negative sense RNA viruses. The N protein has a molecular weight of 45 kDa and the structure of recombinant decameric RNA N-RNA rings has been solved by X-ray crystallography to a resolution of 3.3 Å (Figure 6). Each N subunit consists of a core region with NTD and CTD α helical bundles linked via a hinge region. At the distal ends of the NTD and CTD lies an N and C arm that only appear folded in the quaternary structure. The N arm residues 1-35 insert into the adjacent N monomer forming the N chain whilst the C arm lies in a space determined to be for consecutive turns of the nucleoprotein complex that is not seen in the resolved ring structure (Tawar et al., 2009). The N protein is also kept in a monomeric soluble form by binding to the P protein, with the N CTD implicated in this interaction (Tran et al., 2007). The ability to maintain a pool of monomeric N is critical for subsequent genome and anti-genome encapsidation to form the RNPs. The HRSV N-RNA ring structure revealed 7 ribonucleotides of bacterial RNA bound to each N monomer (Tawar et al., 2009), whilst VSV and RV were determined to bind 9 per N monomer (Green et al., 2006; Albertini et al., 2006). A defined cavity was seen on the NTD and CTD interface (and assumed for all members of the *Monogeavirales* order) and bound three stacked bases. Despite the identification of this conserved cavity, lateral N contacts in VSV results in an opposite curvature with the RNA inside and the N oriented out. In HRSV the RNA is on the outside in relation to N and may provide a mechanism by which the *Pneumovirus* family do not need to disassemble the RNP helix to be read by the RdRp (Tawar et al., 2009; Lilijeroos et al., 2013). Earlier modelling work of N multimers bound to RNA determined the RNPs adopt right-handed helical assemblies (Tawar et al., 2009). However, analysis with the use of a helical nanoparticle with known handedness showed the HRSV RNP to be organised as a left-handed helix (Lilijeroos et al., 2013) and follows the general trend of other *Mononegavirales* order viruses, such as the MV, MuV and Marburg viruses (Bharat et al., 2011).
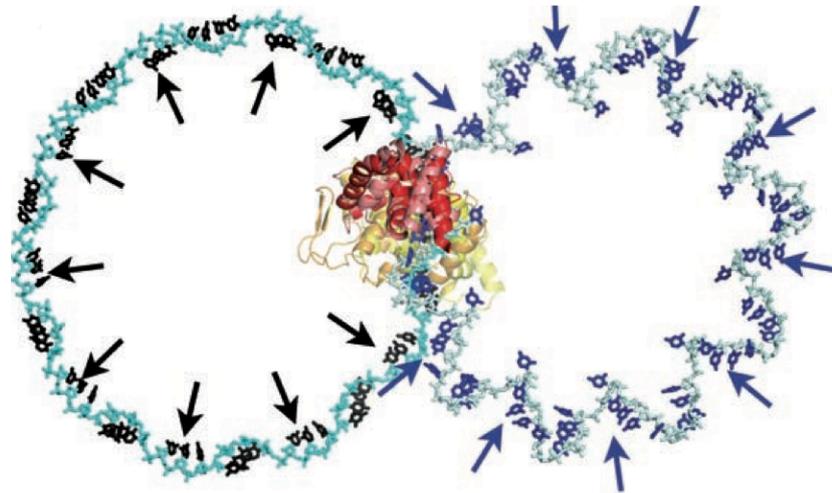
Figure 6 Comparison of HRSV and VSV N: RNA rings

HRSV decameric N: RNA ring (70 bases- left) and VSV (90 bases- right). An N monomer is super positioned for HRSV and VSV and shows the N NTD (yellow) and CTD (red). Arrows indicate the position of 3-stacked ribonucleotides that face inwards towards the N protein. For HRSV (black arrows) this results in RNA oriented on the outside of N, whilst VSV (purple arrows) results in RNA orientated on the inwards face of N. N: RNA ring structures were determined by X-ray crystallography (HRSV- Tawar et al., 2009; VSV- Green et al., 2006).

1.5.3.2 L protein

The large (L) protein is a multifunctional protein that is highly conserved within the *Mononegavirales* order and its role is as the catalytic component of the RdRp, required to polymerise ribonucleotides in order to synthesise genomic and anti-genomic replication products, and mRNA transcripts. In addition, the L protein performs co-transcriptional addition of an mRNA 5' cap and a 3' poly (A) tail (Liuzzi et al., 2005). No high resolution structure currently exists for the HRSV L protein, likely due to the proteins complexity, however the L protein of VSV bound to a fragment of P, residues 35-106, has been solved by cryo-EM to 3.8 Å (Figure 7) and is hypothesised to be in the 'initiation' state, ready to bind the 3' end of the RNA template (Liang et al., 2015). Sequence alignment showed an overall similar alignment of VSV L protein domains with HRSV L and suggest that the VSV L protein structure could be used to further understand the structure and function of HRSV L. The VSV L structure displays 5 distinct domains of the protein; 1) RdRp, 2) mRNA capping domain, 3) methyltransferase domain, 4) connector domain between the mRNA capping and methyltransferase domain and 5) a CTD.

As with all RNA polymerases, the VSV RdRp domain core consists of a right hand 'fingers-palm-thumb' structure, conserved with the RdRp domain of influenza B virus (Reich et al., 2014) and the rotavirus VP1 (Lu et al., 2008). The capping domain contains a priming loop that extended into the active site of the RdRp to allow efficient guanylation. The connector and CTD domains play an organisation role with the connector spacing the three catalytic domains whilst binding the P fragment. The CTD consists of a α helical bundle with a projecting beta hairpin. The beta hairpin extends to the RdRp and terminates at the junction and makes contacts with the mRNA capping, methyltransferase and connector domains to form the structure (Liang et al., 2015).
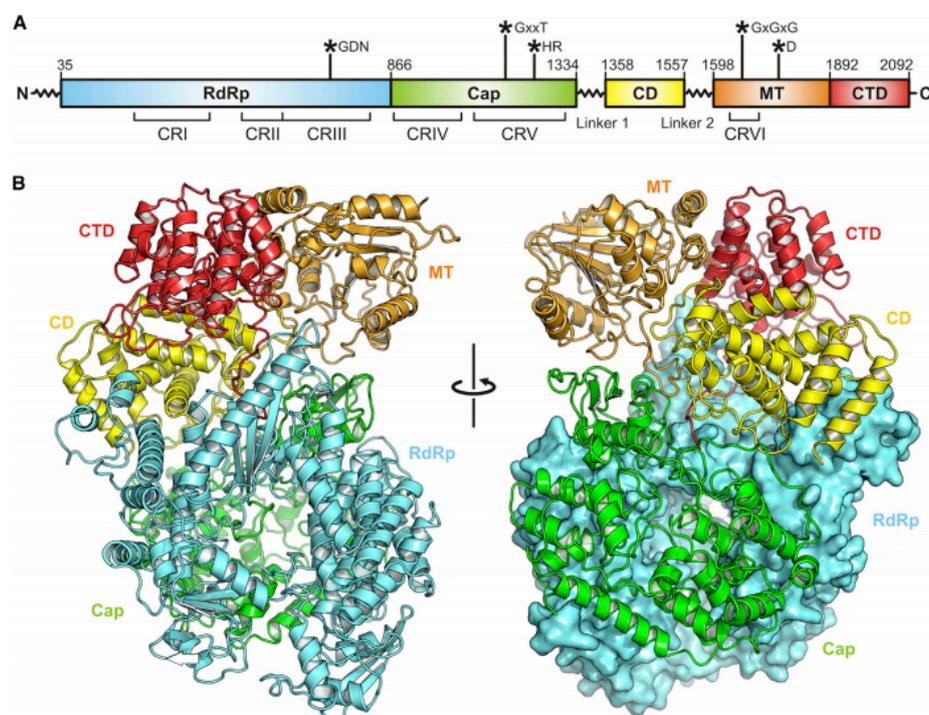
Figure 7 Structure of VSV L protein

A- schematic representation of the five domains of the VSV L protein coloured from N to C terminus: polymerase domain (RdRp)- blue, capping domain- green, connector domain- yellow, methyltransferase domain- orange and CTD- red. The labels "CR" indicate conserved regions amongst viruses in the Mononegavirales order, including HRSV; B- the cryo-EM structure of VSV- L determined to 3.8 Å with domains coloured as in A (Liang et al., 2015).

1.5.3.3 Phosphoprotein (P)

The HRSV phosphoprotein (P) is a structural component of the virion and plays critical roles in viral RNA replication and mRNA transcription through interactions with L, N and M2-1 proteins (Asenjo & Villanueva, 2000) (Figure 8). As its name implies, P is phosphorylated at multiple residues (Ser residues 116, 117, 119, 232 and 237) by cellular casein kinase (CK) II although this modification does not influence its ability to interact with L, N and M2-1 binding partners (Mason et al., 2003; Castagne et al., 2004). P phosphorylation has been implicated in regulating the transcription and replication activities of the polymerase although the mechanism by which this occurs is unknown (Barik et al., 1995).

Mononegavirales P proteins vary in length from 241 amino acid residues in HRSV (Lambden, 1985) to 709 residues for Nipah virus (Harcourt et al., 2000) with little sequence identity. However, despite low sequence identity all P proteins have 3 distinct domains: an N-terminal domain (NTD) that interacts with monomeric N, C-terminal domain (CTD) that interacts with the N-RNA complex, and a central coiled coil (CC) that represents the oligomerisation domain (OD) responsible for tetramerisation of P that is critical for its function (Figure 8). High-resolution structural information is available for the CC domain of several *Mononegavirales* members including MV (Communie et al., 2013), SENV (Tarbouriech et al., 2000), MV (Cox et al., 2013) and HMPV (Leyrat et al., 2013) that revealed a predominantly α helical structure, characteristic of a CC domain. P protomers were found to run in parallel with the exception of the MV that contains two pairs of juxtaposed protomers (Cox et al., 2014). The high sequence identity (72%) of HRSV and HMPV P proteins suggests HRSV will also adopt a similar CC structure, with protomers running in parallel.

Many *Paramyxoviridae* family viruses contain intrinsically disordered proteins for replication functions that may fold on interaction with its binding partner. This dynamic nature is proposed for the P protein of HRSV with the majority of the protein hypothesised to be unstructured, and only able to fold with ordered secondary structures on interaction with N, L and M2-1 (Communie et al., 2014). The interaction of P with M2-1 and its role in anti-termination is described further in section 1.6.2.3 whilst interactions of L and N with P are described below.

1   Monomeric N binding   **29**

1   MEKFAPEFHGEDANNRATKFLESIKGKFTS

31   PKDPKKKDSIISVNSIDIEVTKESPITSNS

61   TIINPTNETDDTAGNKPNYQRKPLVSFKED

100   M2-1 binding   **120**

91   PTPSDNPFSKLYKETIETFDNNEEESSYSY

121   Coiled-coil domain

121   EEINDQTNDNITARLDRIDEKLSEILGMLH

161

151   TLVVASAGPTSARDGIRDAMVGLREEMIEK

181   IRTEALMTNDRLEAMARLRNEESEKMAKDT

203   L binding   **233** RNP binding **241**

201   SDEVSLNPTSEKLNNLLEGNDSDNDLSLEDF

Figure 8 Binding regions of the HRSV P protein

Five functional regions have been mapped to the P protein (residues 1-241) of HRSV. (1) An N-terminal domain (NTD) region (residues 1- 29) that binds monomeric N protein; (2) a region (residues 100-120) that interacts with the M2-1 protein; (3) a coiled coil domain (CC) that forms native tetrameric P (residue 121-161); (4) a CTD region involved in binding the L protein (residues 203- 241) and (5) a CTD region involved in interactions with RNPs. The NTD and CTD of P are hypothesised to be largely disordered in solution and likely become structured on interaction with its binding partners

1.5.3.3.1 P: L interaction

The P protein interacts with L to allow RdRp-directed transcription and replication. The literature is conflicting in terms of identity of the L binding site and both the CTD and CC domain of P have been implicated in interactions with L (Khattar et al., 2001; Asenjo et al 2008; Sourimant et al., 2015). Initial experiments to determine the P residues critical for L interaction used a mini genome system alongside coimmunoprecipitation studies using [^{35}S] methionine labelled L protein and P specific serum. Khattar et al, (2001) determined the deletion of the CC domain (residues 121-160) prevented interaction with L. Whilst a second group (Asenjo et al., 2008) determined only a reduced binding (50%) when the residues 120-150 of the CC were deleted and found a much larger reduction in L: P coimmunoprecipitation when the CTD residues 203-241 were deleted. Most recently Sourimant et al, (2015) used a GST pull down approach to determine critical residues for the L: P interaction. A construct lacking CC domain residues (Δ122-160) improved binding efficiency to L and was used for all further work by the group. The improved binding seen here in contrast to the necessity of this region by Khattar et al, (2001) could be due to the GST tag having a compensatory action and allowing P oligomerisation in the absence of the CC domain. The results of Asenjo et al, (2008) are in agreement with those of Sourimant et al, (2015) in that P: L binding can occur in absence of the CC domain. However, the P OD is required to allow full function in viral synthesis. For the closely related SENV the CC of P binds to L to allow function (Tarbouriech et al., 2000), however the HRSV P CC is much shorter in length than that of other viruses and thus L may bind to a distal region of P as reported in the literature.

Deletion of CTD P residues 216-241 (Sourimant et al., 2015) and residues 203-241 (Asenjo et al., 2008) both resulted in loss of the P: L interaction. Site directed mutagenesis via alanine scanning identified three hydrophobic residues (Leu 216, Leu 233 and Leu 227) that reduced RNA polymerase activity to background levels, whilst Ala mutation of Lys 222, Asn 230 and D 231 resulted in over a 50% reduction in polymerase activity. Ala mutation of Asn 224 did not alter RNA polymerase activity but mutation to Pro at this location resulted in near background levels of activity. These data suggest the requirement of the CTD hydrophobic Leu residues for polymerase activity whilst ablation of activity on mutation of Asn 224 to Pro suggests a requirement of a helix at the CTD for interaction with its binding partner and subsequent function, potentially by

recognition of a MoRe (molecular recognition) site that folds upon binding to L (Sourimant et al., 2015).

1.5.3.3.2 P: N interaction

The P protein acts as an N protein chaperone, maintaining the solubility of newly synthesized N, to prevent self-oligomerisation and binding to cellular RNA. In addition, P binds and releases the RNP complex and allows local uncoating of the RNP for polymerase access for transcription and genome replication (Ruigrok et al., 2011). The RNP binding region domain of P lies within the P_{CTD} at residues 233- 241. In particular the terminal 4 residues of this sequence are critical for interactions with N-RNA as mutation of Leu 238 or Phe 241 resulted in complete abrogation of binding (Tran et al., 2007). Interestingly, the P_{CTD} shows little sequence similarity between the closely related HRSV and HMPV P proteins. A second RNP interaction site has been hypothesised directly after the CC domain, residues $P_{161-180}$ (Khattar et al., 2001). Removal of this site and the terminal P_{CTD} together or separately resulted in complete abrogation of RNP binding in the closely related BRSV P. An earlier study suggested an interaction of N of the RNP complex with the CC domain of HRSV $P_{132-168}$, however as the CC domain is essential for P tetramerisation, it is likely that truncation of this region resulted in destabilisation of the P protein that subsequently resulted in insufficient RNP binding (Mallipeddi et al., 1996). The P protein bound L in the absence and presence of N in a coimmunoprecipitation assay (Khattar et al., 2001) and while the most recent data illustrates an overlapping binding site of N and L at $P_{233-239}$ (Sourimant et al., 2015) and suggest competition for binding between L and N, most likely both proteins can bind simultaneously to different protomers of the tetrameric P protein.

Monomeric RNA-free N required for further encapsidation of replication products (anti-genome and genome) binds to the P_{NTD} as monomeric N bound weakly to the P_{CTD}, and was inhibited by N binding to the P_{NTD}, residues 1-29 (Mallipeddi et al., 1996; Galloux et al., 2015). The multimeric ring form of N, comprising 10 or 11 N monomers, does not bind to the P_{NTD} (Khattar et al., 2001; Galloux et al., 2015). Despite low sequence similarities, all *Paramyxoviridae* P proteins have been found to contain a short conserved motif (MoRe site) at the P_{NTD} of 11-26 amino acids in length, termed soyuz 1 and which is reported to prevent N self-binding (Karlin and Belshaw, 2012; Galloux et al., 2015). Conversely, MuV, which belongs to the *Rubaluvirus* genus of the *Paramyxoviridae* family, requires both P NTD and CTD for RNP interaction through uncoiling of the RNP and stabilising for the polymerase function respectively. However, the *Rubulavirus* P

proteins have a unique structural orientation, with the tetrameric form of P containing two pairs of juxtaposed protomers, explaining the dual functions of P NTD and CTD in RNP interactions and thus the interaction of RNP: P_{NTD} may be limited to the *Rhabdoviridae* family (Cox et al., 2013 & 2014).

1.5.4 Progeny virus assembly and exit

Following HRSV replication and transcription, progeny viruses are packaged and released from cells to allow infection of neighbouring cells or susceptible hosts. Progeny virus can be detected in cell culture 10-12 hours post infection with a peak in virus release at 24 hours (Collins, 2007). Assembly and budding of virions occurs at the apical surface of polarised epithelial cells (Roberts et al., 1995). In addition to virus-encoded proteins, virions are known to include several cellular proteins, including components of lipid raft microdomains (the organisational centre for assembly of signalling molecules), actin associated proteins and apical recycling endosomes (ACE).

Virus budding of *Paramyxoviridae* family members parainfluenza-5 (PIV5) and MV are dependent on endosomal sorting complexes for transport (ESCRT) pathways. Inhibition of a Vps4 AAA-ATPase required in this ESCRT pathway to dismantle and recycle the ESCRT complex resulted in inhibition of viral budding (Schmitt et al., 2002; Li et al., 2009). In contrast to these viruses, HRSV budding occurred in the presence of ESCRT pathway inhibitors and is thought to be ESCRT independent (Utley et al., 2008). Due to HRSV budding at the polarised apical surface, apical recycling endosomes (ARE) enriched with Rab11a are hypothesised to aid budding.

Microdomains in lipid bilayers have been hypothesised to provide nucleation points to allow the concentration of viral proteins for subsequent budding. Indeed, lipid raft markers such as caveolin-1, CD55 and CD58 have been identified in budding filaments (Brown et al., 2002; Jeffree et al., 2003). Further, PI3K, a raft associated kinase and rhoA kinase, a raft associated signalling molecule, are involved in regulating actin structure and have been implicated in the formation of the viral filament (Jeffree et al., 2007). The actin cytoskeleton has been heavily implicated in HRSV budding and has been detected in large quantities within purified virions. Disruption of actin and microtubule networks resulted in impaired HRSV entry, assembly and release (Kallewaard et al., 2005; reviewed by Harrison et al., 2010). Actin associated heat shock protein 90 (HSP90) and HSC70 proteins were detected by mass spectrometry analysis of

HRSV virus particles. Inhibition of HSP90 in a Hep2 cell line infected with the HRSV A2 strain, resulted in the inhibition of viral particle formation, likely due the inability of actin to allow viral assembly and release (Radhakrishnan et al., 2010).

1.5.4.1 M protein

The matrix protein (M) forms an internal layer of the viral membrane and plays central roles in viral particle assembly and the budding polarity of the virus (Batonick et al., 2008). In addition, the M protein can interact with the cytoplasmic tail of F and is thought to maintain the tubular shape of filamentous virions (Mitra et al., 2012; Liljeroos et al., 2013). In the filamentous virion devoid of SH and G, an extra layer of electron density below M was identified, assumed to be M2-1 that connects M and RNPs (Kiss et al., 2014). M has been reported to target RNPs to virus assembly sites for the MeV, however the M2-1 protein, not found in MeV, could provide this function for HRSV (Runkler et al., 2007).

The structure of the HRSV M protein has been solved by X-ray crystallography to a resolution of 1.6 Å, and shown to exhibit high structural similarity to the VP40 protein of EBOV. The M protein was found to be a monomer in solution however cross-linking studies of the purified protein identified higher order oligomers including dimers, tetramers and hexamers. The overall architecture of M consists of 2 domains, an NTD and CTD, separated by a 13-residue linker that is mostly unstructured (Figure 9A). The secondary structures of both NTD and CTD mainly consist of beta sheets, with the NTD comprising 2 beta sheet bundles perpendicular to each other (Figure 9A blue), whereas the CTD has a flattened beta barrel structure (Figure 9A red). Surface electrostatics of the M protein display a large positively charged surface across both the NTD and CTD, consistent with a role in interactions with the negatively charged membrane (envelope). In addition, the NTD of M contains a distinct negatively charged area (Figure 9B red) that could interact with positively charged viral proteins, required for the role of M (Money et al., 2009).

A
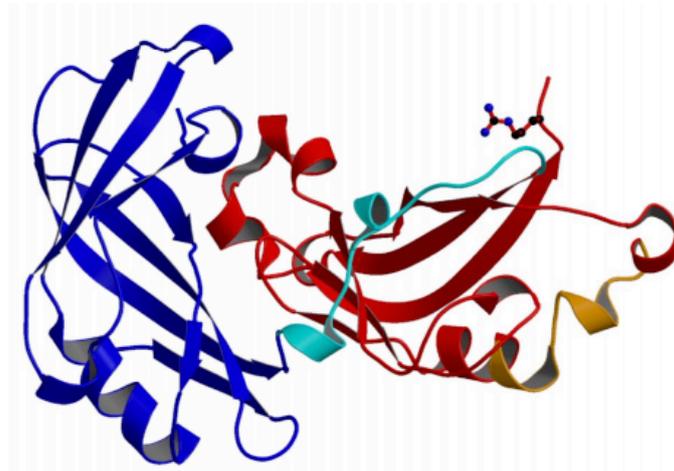
B
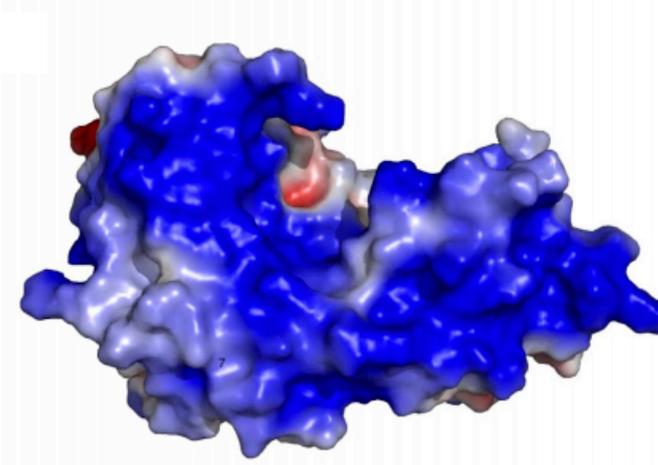
Figure 9 Structure of the HRSV M protein

The structure of the monomeric M protein was determined by X-ray crystallography to a resolution of 1.6 Å. A- ribbon diagram of the M protein displays two domains, NTD (blue) and CTD (red), separated by an unstructured linker (cyan). The NTD consists of 2 beta sheet bundles perpendicular to each other whilst the CTD consists a flattened beta barrel structure; B- surface electrostatics of the M protein displays a large positively charged surface (blue) across both the NTD and CTD and a smaller defined region of negative charge (red).

1.6 M2-1 protein

1.6.1 Function

The M2-1 gene is essential for rescue of infectious HRSV, which implies it performs a critical function in the virus life cycle. The role assigned to M2-1 is as an anti-termination transcription factor that allows full-length transcription of each viral mRNA (Collins et al., 1996) and enhances transcriptional read-through at gene-end/ gene-start junctions (Hardy & Wertz 1997). This latter read-through function has been suggested to permit access of the RdRp to downstream transcriptional units thus bypassing the transcription attenuation step that occurs at each GJ. There has been some dispute in the literature of which of these two anti-termination roles of M2-1 is correct, although it is possible that they are mechanistically related, as both involve M2-1 influencing the way in which the RdRp recognises transcription termination signals. All research to date indicates a role for M2-1 in transcription and not RNA replication (Fearns & Collins, 1999).

Analysis of M2-1 function was first performed using a mini-genome assay (Fearns & Collins, 1999); such assays are particularly powerful in that they allow the investigation of roles of essential *cis* and *trans* acting viral components in viral transcription and replication, where mutations in these components would normally be lethal to an infectious virus. Initial data from the mini-genome assay showed supply of the M2-1 protein along with N, P and L proteins increased the abundance of mRNA transcription, but had no effect on replication. Furthermore, primer extension analysis revealed that in the absence of M2-1, the abundance of full-length mRNA transcripts was dramatically reduced, and instead the most abundant transcripts were truncated such that few possessed intact 3' end sequences. This lack of full-length mRNAs was dependent on the size of the transcriptional unit, with the HRSV RdRp being capable of full-length mRNA synthesis from small genes, but not from genes in excess of more than a few hundred nucleotides. In this case, the smaller NS1 and to a lesser extent NS2 genes of HRSV could be synthesised fully in the absence of M2-1 (Collins et al., 1996; Fearns & Collins, 1999). This finding was interpreted as being due to the HRSV RdRp possessing poor intrinsic processivity during transcription, resulting in premature intragenic termination of polymerisation and thus the generation of abundant truncated mRNA transcripts. It was suggested that the

role of M2-1 was to increase polymerase processivity, and as such was described as a processivity factor. A further role of the M2-1, described as anti-termination, was assigned due to the observed increase in read-through RNA synthesis when polygenic mini-genomes containing two or more transcriptional units were used to assess M2-1 function. In these cases, M2-1 was shown to influence transcriptional termination specifically at GE sequences, resulting in synthesis of an increased abundance of read-through or polycistronic mRNAs (Hardy & Wertz, 1997).

1.6.1.1 M2-1 structure

The HRSV M2-1 protein (A2 strain) is 194 amino acids in length and has a molecular weight of 22 kDa. The first high-resolution structural information of the M2-1 protein was obtained in 2012, with the solution of the NMR structure of M2-1$_{58-177}$ (Blondot et al., 2012). In 2014, the crystal structure of the full-length native M2-1 protein was determined to a resolution of 2.5 Å (Tanner et al., 2014). M2-1 crystallised as a tetramer, and the structure revealed that contacts between individual monomers extended to all other monomers in the multimer, thus confirming previous reports that M2-1 forms a tetramer in its physiological state on the basis of chemical cross linking and negative stain EM (Tran et al., 2009; Blondot et al., 2012). Each M2-1 monomer comprises three distinct regions; the N-terminal Cys_3His_1 motif that forms the zinc binding domain (ZBD), an OD, also termed the oligomerisation helix, and the core domain. The three distinct regions in addition to phosphorylation sites, the C terminus and residues critical for interactions between M2-1 and its binding partners are highlighted in Figure 10 and Figure 11 and discussed in this section.
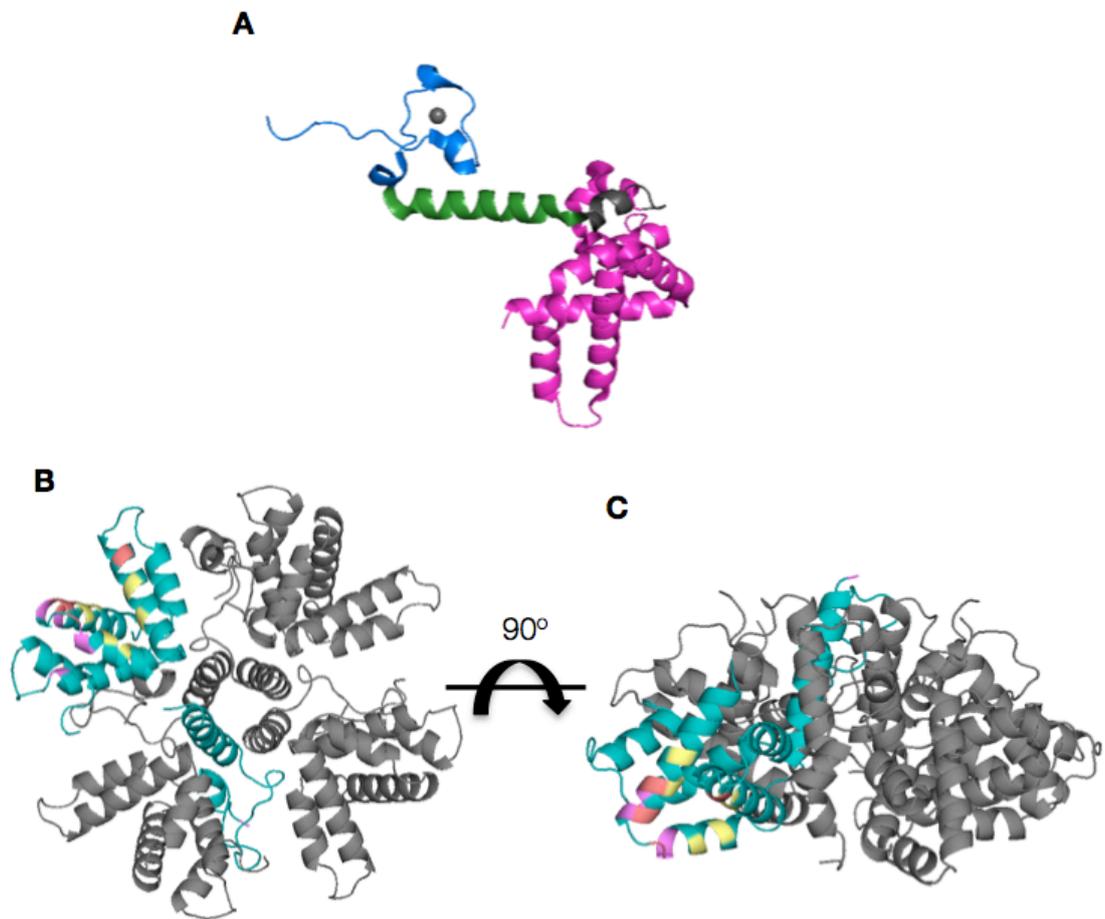
Figure 10 Structure of the HRSV M2-1 protein

A- monomeric M2-1 has 3 distinct domains, the ZBD (blue), oligomerisation helix (green) and core domain (pink); B- native M2-1 forms a tetramer (grey) with one monomer highlighted in blue. M2-1 interactions with RNA (pink), P (yellow) and both P and RNA (coral) are shown on the surface of M2-1 and further discussed in section 1.6.2. The crystal structure of HRSV M2-1 was determined to a resolution of 2.5 Å (Tanner et al., 2014; PDB: 4c3b).

```
                    7              ZBD              25
                    ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬
    1      MSRRNPCKFEIRGHCLNGKRCHFSHNYFEW
              32   Oligomerisation helix   50
              ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬
    31     PPHALLVRQNFMLNRILKSMDKSIDTLSEI
                    69            Core domain
                    ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬
    61     SGAAELDRTEEYALGVVGVLESYIGSINNI

           ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬
    91     TKQSACVAMSKLLTELNSDDIKKLRDNEEL
                         Core domain
           ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬
    121    NSPKIRVYNTVISYIESNRKNNKQTIHLLK
                                    172   174
                                    ▬▬▬▬▬▬▬▬
    151    RLPADVLKKTIKNTLDIHKSITINNPKEST
           CTD           194
           ▬▬▬▬▬▬▬▬▬▬▬▬▬▬
    181    VSDTNDHAKNNDTT
```

Figure 11 Features of the HRSV M2-1 structure

Distinct M2-1 regions are highlighted from N to C terminus; ZBD (blue), Oligomerisation helix (red), Core domain (teal) and the CTD (purple). Two M2-1 residues are phosphorylated (Ser 58 and Ser 61) and are highlighted in green. Putative interactions of M2-1 amino acids with its binding partners are highlighted RNA (pink), P (light green) and M2-1 amino acids proposed to interact with both P and RNA are highlighted in coral. The interactions of M2-1 with its binding partners are further discussed in section 1.6.2. Structural regions are based on the HRSV M2-1 crystal structure (Tanner et al., 2014; PDB: 4c3b) whilst putative M2-1 interactions with RNA and P are based on SDM studies (Blondot et al., 2012; Tanner et al., 2014).

1.6.1.2 ZBD

The Cys_3His_1 motif also known as the ZBD is responsible for zinc-coordination, and mutation of these conserved residues resulted in a loss of transcriptional anti-termination activity of M2-1 and binding to the N protein (Hardy & Wertz, 2000). Thus the zinc coordinating residues of the ZBD are critical for the functional integrity of M2-1 (Hardy & Wertz, 2000). From structure determination of the M2-1 tetramer, it was shown that residues that form the ZBD interact with residues of the adjacent M2-1 monomer, which suggests the ZBD plays a role in maintaining the tetrameric structure of M2-1 (Tanner et al., 2014). Furthermore, removal of the zinc from the ZBD by EDTA treatment resulted in dissociation of M2-1 into monomeric protein with a stable secondary structure that could still bind RNA and P (Esperante et al., 2013). Taken together these findings suggest that while the monomeric form of M2-1 is able to bind with its RNA and P ligands, formation of tetrameric M2-1 with an intact ZBD is required for its anti-termination function. The Cys_3His_1 motif has been found to co-ordinate zinc in other proteins, such as the Nup475 mammalian transcription factor and the closely related VP30 transcription factor protein of EBOV, where substitution of the zinc co-ordinating residues resulted in a reduction in its transcription activating function (Modrof et al., 2003).

1.6.1.3 Oligomerisation helix

The region encompassing the OD of M2-1 was first mapped to a putative α helix (M2-1 residues 32-63) and deletion of this region resulted in a reduction in HRSV transcription (Tran et al., 2009). The crystal structure of the full length M2-1 confirmed all previous data and revealed the domain to comprise a single α helix consisting of residues 32-50. The crystal structure showed 4 hydrophobic amino acids (Leu 36, Leu 43, Ile 46 and Met 50) that are all buried on one face of a helix and interact with the other three M2-1 monomers to form the tetramer (Figure 12). In addition to the oligomerisation helix, the crystal structure revealed that the ZBD residues were also involved in coordinating tetramerisation via interactions with the N-terminal face of adjacent monomers (Tanner et al., 2014).
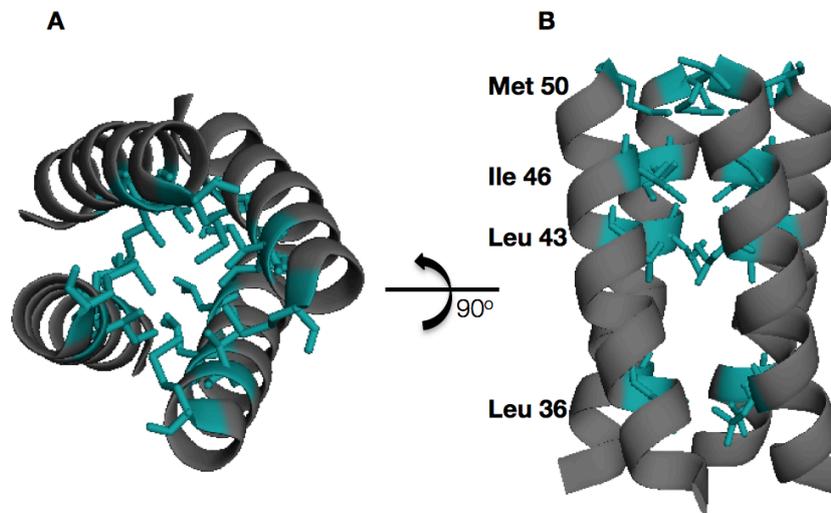
Figure 12 Structure of the HRSV M2-1 oligomerisation domain

> The structure of the M2-1 oligomerisation helix (residues 32-50) and its interactions with adjacent helices was determined by X-ray crystallography (Tanner et al., 2014). Hydrophobic residues of the helix (Leu 36, Leu 43, Ile 46 and Met 50 face inwards and interact with the same residues of adjacent protomers.

1.6.1.4 Core domain

The overall globular architecture of the core domain of M2-1 was found to be in agreement using both NMR and X-ray crystallography methods (Blondot et al., 2012; Tanner et al., 2014). The core domain is comprised of residues 58-177 and binds both RNA and P (Cuesta et al., 2000; Mason et al., 2003), with the binding of these ligands critical for the function of M2-1 as an anti-terminator, and further detail of these interactions are discussed in section 1.6.2. The M2-1 core domain compromises a single globular domain (residues Gly 75- Ile 171) and consists of 6 α helices. Helices 1, 2, 5 and 6 form a helix bundle scaffold whilst helices 3 and 4 form a hairpin and are stacked upon helix 6 (Figure 13).

The HRSV M2-1 core domain exhibits structural homology to that of M2-1 from the closely related *Pneumoviridae family member* HMPV M2-1 and also the VP30 protein of the more distantly related *Filoviridae* family member, EBOV. These relationships are further discussed in section 1.8.
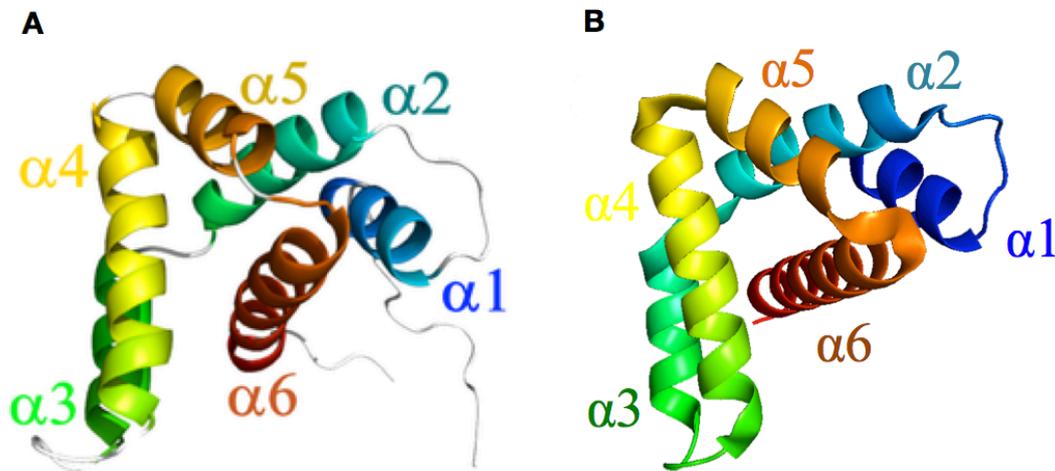
Figure 13 Comparison of HRSV M2-1 core domain structures

A- The high resolution structure of the M2-1 core domain residues 58-177 determined by solution NMR (Blondot et al., 2012); B- M2-1 core domain residues 58-177 determined by X-ray crystallography (Tanner et al., 2014). Both structures are in agreement and display a similar arrangement of α helices.

1.6.1.5 Phosphorylation sites

The OD and core domain of M2-1 are linked by a highly flexible region comprising residues 52-67, which contains two phosphorylation sites (Ser 58 and Ser 61). Phosphorylation of M2-1 was first experimentally observed in 2000 where different migrating patterns of M2-1 on SDS-PAGE gels were revealed (Hardy & Wertz, 2000). The slower migrating band was identified as a phosphorylated species and the faster migrating band to be non-phosphorylated. M2-1 phosphorylation is hypothesised to occur by cellular CK1, a serine/ threonine selective protein kinase (Cartee & Wertz, 2001). Both phosphoablatant and phosphomimetic M2-1 mutants (Ser 58/61 Ala & Ser 58/61 Asp) decreased anti-termination activity in a mini genome system (Tanner et al., 2014), whereas both phosphorylated and non-phosphorylated forms of M2-1 were detected in HRSV infected cells (Zhou et al., 2003), suggesting a requirement of dynamic phosphorylation for optimal anti-termination activity, although the mechanistic basis for this unknown (Hardy & Wertz, 2000; Cartee and Wertz, 2001; Tran et al., 2009; Tanner et al., 2014).

1.6.1.6 CTD

The CTD residues of M2-1 (174-194) were non-determinable in the crystal structure and are presumed to be unstructured in this form (Tanner et al., 2014). The requirement of the CTD for the function of M2-1 was assessed by creating deletion mutants of the M2-1 CTD (M2-1Δ127, Δ148 and Δ177) and then attempting to rescue infectious virus from cDNA. M2-1Δ177, in which the extreme 17 CTD residues were deleted resulted in the recovery of infectious virus, however a reduction in replication efficiency was seen *in vivo* and *in vitro.* Further, β galactosidase was assayed as a measure of M2-1 dependent processivity and the truncated M2-1Δ177 resulted in over a 50% reduction in mRNA synthesis compared to full-length M2-1 (Tang et al., 2001).

1.6.2 Binding partners of M2-1

1.6.2.1 RNA specificity of M2-1

Initial RNA binding studies illustrated the ability of M2-1 to bind long (700-1300 nts) RNAs with no apparent sequence specificity, whereas shorter sequences (80 nts or less) exhibited sequence specificity, with the highest affinity being determined for those RNAs that included the HRSV anti-genomic leader sequence (Cuesta et al., 2000). In contrast, other studies showed M2-1 exhibited a preference for binding viral mRNA in HRSV-infected cells, which did not contain the leader sequence (Cartee and Wertz. 2001) but were A-rich in sequence, particularly due to their 3' poly (A) tails. More recently, Blondot et al (2012) tested the binding of short (10-15 nts) negative and positive sense HRSV-specific RNAs and determined that the core M2-1$_{58-177}$ domain showed a preference for binding A-rich RNA sequences. In addition, these workers found that double stranded RNAs (representing the F GE sequence) bound to M2-1 core with higher affinity compared to single stranded RNAs of a similar size, and it was suggested that the target RNA is a duplex. The work of Tanner et al (2014) further characterised the RNA binding properties of M2-1 using fluorescent anisotropy, and the results were in broad agreement with those of Blondot et al (2012) in that positive sense GE sequences, and particularly those with high A-content were bound with highest affinities.

1.6.2.2 M2-1 residues involved in RNA binding

Functional analysis of the RNA binding ability of M2-1 was initially performed using various M2-1 mutants with residues that comprised phosphorylation sites Ser 58 and Ser 61, as well as residues that comprised the Cys_3His_1 ZBD motif being identified in RNA binding (Cuesta et al, 2000). In contrast, Tran et al (2009) defined the M2-1 RNA binding site to between residues 59-153, thus excluding the ZBD, the OD and the unstructured CTD in contributing to RNA binding ability. More recently, M2-1 core residues (Ser 58, Ser 61, Lys 92, Lys 150, Arg 151 and Lys 159) were assessed by site direct mutagenesis (mutation to Ala or Asp) and were shown to display a reduction in RNA binding ability in addition to displaying a reduction in anti-termination activity by the use of a mini genome system. This showed that the ability of M2-1 to bind RNA was critical for its anti-termination function (Tanner et al., 2014). Further core residues Arg 126, Val 127, Asn 129, Thr 130, Ser 133, Leu 152 and Val 156 were also found to be critical for RNA binding via performing NMR perturbation experiments (Blondot et al., 2012). Most recently and in contrast to earlier literature (Tran et al., 2009), the M2-1 N-terminal residues Arg 3 and Arg 4 were mutated to Ala and resulted in a loss of RNA binding and implicated the involvement of the N-terminus of M2-1 in interacting with RNA (Tanner et al., 2014). This finding was consistent with previous work showing that the M2-1$_{58-177}$ core exhibited lower affinity for RNA than full length M2-1 (Blondot et al., 2012) and thus suggested that interactions other than that with the core domain must take place for full affinity of M2-1 to RNA. The high-resolution core domain structure of M2-1$_{58-177}$ displayed a similar architecture to that of the full-length M2-1 and was determined to have two oppositely charged faces, with the face hypothesised for RNA and P binding, being a positively charged surface (Blondot et al., 2012; Tanner et al., 2014).

1.6.2.3 The interaction of M2-1 with the P protein

Mason et al (2003) determined that P binding to M2-1 *in vitro* contributed to the ability of M2-1 to act as an anti-terminator and hypothesised that this interaction was required to deliver P to the HRSV RdRp complex. Tran et al (2009) initially determined that P and RNA compete for the same binding site on M2-1 *in vitro* at the core domain, residues 59-177. Furthermore, Blondot et al (2012) used NMR to characterise two distinct binding surfaces for RNA and P that partially overlap within this core domain. Blondot et al (2012) determined that α -helices

4, 5 and 6 of the M2-1 core domain are involved in interactions with P, and contacts include residues Ser 133, Tyr 134, Val 156, Lys 159, Asn 163 and Lys 162. A loss of the M2-1: P interaction resulted in a decrease in transcription and prevented localisation of M2-1 to inclusion bodies, the site of RNA synthesis (Blondot et al., 2012). Therefore, it was hypothesised that P may play a role in transporting M2-1 to inclusion bodies (IBs) where it can bind its target RNA and perform its anti-termination function (Rincheval et al., 2017).

The region of P responsible for interaction with M2-1 was determined to involve residues 100-120, specifically residues P_{101}, P_{102} and P_{109} as determined by deletion mutations and affinity chromatography (Mason et al., 2003). Initially sections of 20 amino acid in length were deleted along the length of the P protein to determine regions involved in binding both M2-1 and N. Deletion of residues 100-120 of P resulted in a decrease in M2-1 binding (6% binding, compared to 100% for wild type P) but did not reduce binding to the N protein as determined by affinity chromatography and subsequent quantification of western blot protein band intensity. Deletion of regions encompassing residues 120-140 and 140-160 resulted in a decrease in both M2-1 and N binding, however these residues form the CC domain of P and likely a decrease in binding affinity is due to a non-native tetrameric form of P being expressed with a deletion of this region. Further assessment of the P region encompassing residues 100-120 region was performed via double and single Ala scanning mutagenesis and analysed by affinity chromatography and two mini genome systems that detected either luciferase or chloramphenicol acetyltransferase. Single Ala mutations of Leu 101 Tyr 102 and Phe 109 resulted in a decrease in M2-1 binding and reduced luciferase output in the M2-1 dependent luciferase mini genome whilst in the M2-1 independent chloramphenicol acetyltransferase mini genome no reduction in output was observed. Structural predictions using small angle X-ray scattering (SAXS) for the closely related HMPV M2-1 protein suggested no α helical propensity for residues 100- 120 of P (Leyrat et al., 2013) and may represent a MoRe site, an intrinsically disordered region that folds on binding to the M2-1 protein.

1.6.2.4 The interaction of M2-1 with the M protein

M2-1 is postulated to mediate the interaction of M and RNP for inclusion of RNPs into IBs (Li et al., 2008). M2-1 interacts directly with the NTD of M, discovered via deletion of the M NTD (residues 1-113) in a coimmunoprecipitation assay. M and M2-1 were found to co-localise in the

absence of other viral proteins, however in transfected cells, co-localisation was not seen, likely due to the transient nature of this interaction in the presence of other proteins in the viral life cycle. Viral transcription was enhanced when the M protein was blocked by an M specific antibody in a transcription reaction using purified RNPs and suggest a role for this M2-1: M interaction in inhibiting transcription and initiating viral assembly (Ghildyal et al., 2002). Despite not expressing the M2-1 protein, an interaction of M and RNPs has been seen in VSV (Kaptur et al., 1991) to inhibit transcription and allow assembly.

1.6.3 M2-1 localisation within the virus particle

It was hypothesised that in addition to its anti-termination activity, M2-1 has a role in forming the structure of mature virions. Electron cryotomography data illustrated an internal layer of electron density under the M protein that is linked to the inner leaflet of the viral envelope and adjacent to the RNP in filamentous viral particle (Liljeroos et al., 2013). This layer was later confirmed as M2-1 by laser scanning confocal microscopy, and was found in both the cytoplasm of infected cells, and assembled virus filaments, albeit in lower quantities to that of the M protein (Kiss et al., 2014). The 3D crystal structure of HRSV M2-1 in its tetrameric state does not seem consistent with the possibility of M2-1 forming a 2D layer in the virion structure. It has been suggested that M2-1 in its monomeric form, in the absence of zinc co-ordination, plays a role in assembly and provides interaction with M and RNPs (Esperante et al., 2013). Interestingly, the closely related HMPV M2-1 crystallised as a tetramer in which one or more of the protomer core domains could be 'flipped' out thus allowing the formation of a 2D layer (Leyrat et al., 2014) and provides an explanation to the proposed ability of M2-1 to form an internal layer in the virion.

1.7 M2-2 protein

The M2 gene encodes a second polypeptide from an overlapping ORF on the M2 mRNA, which is known as M2-2. This product is 90 amino acids in length, and its synthesis is mediated by a coupled translation mechanism in which the expression of M2-2 depends on the prior translation of M2-1 from the upstream ORF (Ahmadian et al., 2000). The precise function of M2-2 remains elusive, but it has been suggested to regulate the balance between mRNA transcription and RNA replication of the anti-genome (Bermingham & Collins, 1999). The M2-2 ORF can be deleted from the genome, showing it is a nonessential accessory

factor, although gene deleted viruses are growth attenuated and as such are promising vaccine candidates (Teng et al., 2000).

1.8 Comparison of HRSV M2-1 to EBOV VP30 and HMPV M2-1

The high resolution NMR and crystal structures of HRSV M2-1 have revealed clear structural homology with two other viral proteins with similar established or proposed functions relating to modulating polymerase activity. These are the M2-1 protein of HPMV and VP30 of EBOV.

1.8.1 HMPV M2-1

HMPV is the most closely related virus to the *Orthopneumovirus* genus, which includes both BRSV and HRSV. HMPV encodes an M2-1 protein that exhibits a high degree of sequence identity (38% identity, 64% similarity) with HRSV M2-1. The HMPV M2 gene does not comprise two overlapping ORFs like HRSV, and HPMV M2-1 does not have an essential role in viral transcription, as evidenced by the fact that the M2-1 gene can be deleted from the genome and infectious virus still rescued (Collins et al., 1996; Buchholz et al., 2005). The crystal structure of HMPV M2-1 has recently been determined in complex with an adenosine monophosphate (AMP) molecule and the DNA sequence 5'-AGTT- 3' (Leyrat et al., 2014). In comparison to HRSV M2-1, the structure of HMPV M2-1 was as an asymmetric tetramer, with the core of one protomer 'flipped' out (Figure 14). Molecular dynamic simulations and SAXS showed the structure of the HMPV tetramer to be dynamic with the closed symmetric tetramer conformation as seen for HRSV M2-1 being favoured on nucleotide binding. HMPV M2-1 preferentially bound AMP mononucleotides and, consistent with this, only AMP could be seen bound to HMPV M2-1 in resulting electron density maps. Despite differences in their tetrameric conformations, both *Pneumoviridae* M2-1 proteins preferentially bound A-rich RNA oligonucleotide sequences. The structurally determined nucleotide binding residues of HMPV M2-1 formed the same RNA core domain binding surface identified for HRSV M2-1 (Blondot et al., 2012; Tanner et al., 2014) with residues predicted to interact with both RNA and P highly conserved between both viruses. Furthermore, Leyrat et al. (2014) hypothesised a model of RNA binding to M2-1 in which residues within both the core domain and ZBD formed a continuous binding surface, proposed to bind a

total length of 13 nucleotides, which the authors noted coincided with the precise length of the consensus HRSV gene end sequence.
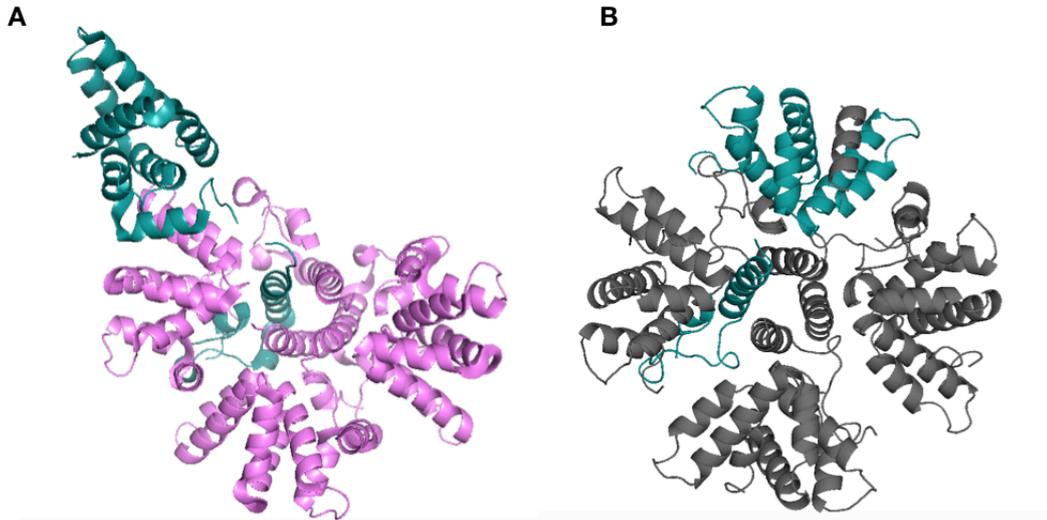
Figure 14 Comparison of HRSV and HMPV M2-1 structures

A- Tetrameric HMPV M2-1 (pink) with one monomer highlighted in blue; B-tetrameric HRSV M2-1 (grey) with one monomer highlighted in blue. Both structures are in agreement with the binding of M2-1 partners (RNA and P) hypothesised to result in the closed conformation (B) of the protein. M2-1 structures were determined by X-ray crystallography (HMPV- Leyrat et al., 2014; HRSV- Tanner et al., 2014).

1.8.2 EBOV VP30

The *Mononegavirales* member EBOV expresses an essential polymerase cofactor VP30 that shares many structural and functional similarities to HRSV M2-1. VP30 plays a critical role in EBOV mRNA transcription (Martinez et al., 2008) that is thought to specifically act during initiation of transcription, through interaction with a conserved RNA stem loop located within the N gene (Hoenen et al., 2006), although VP30 has also been shown to influence RNA replication unlike HRSV M2-1 where no defined role in replication has been observed. A role of VP30 in viral RNA synthesis initiation was suggested by the recent crystal structure of EBOV VP30 in complex with a nucleoprotein peptide. However the functional relationship between the nucleoprotein and VP30 for RNA synthesis was not identified in an EBOV mini genome assay, with mutated residues that did result in a decrease in mini genome output likely due to affecting the structural integrity of the studied protein (Xu et al., 2017). The VP30 CTD forms

hexamers within virions (Hartlieb et al., 2006) whereas full-length M2-1 forms tetramers in solution, as evidenced by cross-linking analysis and size exclusion chromatography (Tran et al., 2009). Both VP30 and M2-1 interact with RNA in order to perform their transcriptional activation functions, and phosphorylation of both proteins at defined residues is a requirement for their respective functions (Modrof et al., 2002; John et al., 2007). Furthermore, both proteins contain a zinc binding Cys_3-His motif, which has been shown to be required for function. The structure of the CTD of VP30 (Figure 15) has been solved by X-ray crystallography (Hartlieb et al., 2006), and comparison with the HRSV M2-1 core domain revealed a similar arrangement of helices. Similarities between VP30 and M2-1 allow comparisons to be drawn on their function as anti-terminator factors.
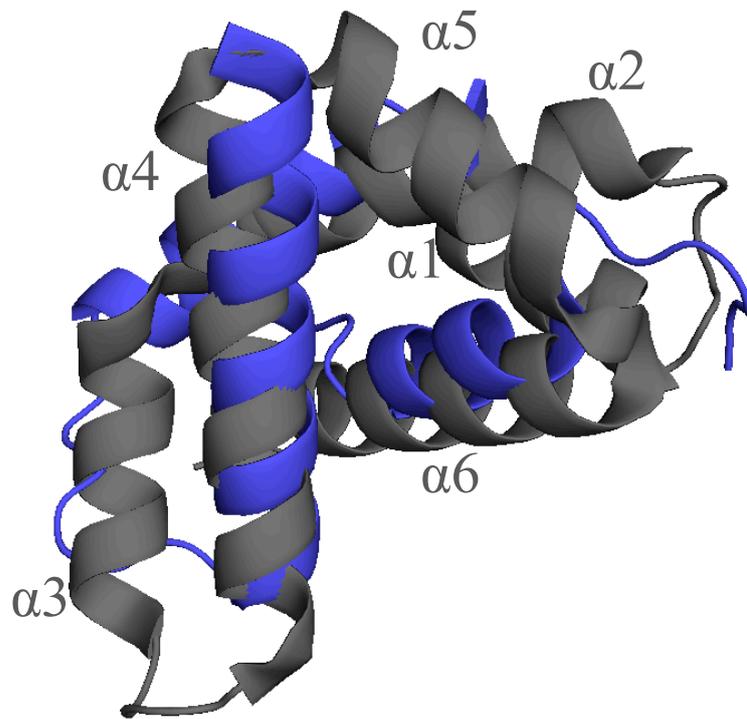
Figure 15 Similarities of the HRSV M2-1 and EBOV VP30

HRSV M2-1 core domain (grey; residues Val 76 – Asn 174; PDB: 4c3b) and EBOV VP30 CTD (purple; residues Phe 181- Pro 240; PDB: 2ieb) crystal structures were aligned and determined structural similarities for three α helices 2, 4 and 6.

1.9   Prevention and treatment of HRSV

Despite the high disease burden associated with HRSV infection, and the accumulation of a detailed understanding of the HRSV life cycle, there is still an unmet need for either a safe vaccine, or a cost-effective anti-viral therapy for HRSV.

In terms of HRSV prevention, no Food and Drug Administration (FDA) approved vaccine exists, due in part to the unsuccessful trial in the 1960s of a formalin-inactivated vaccine that exacerbated HRSV disease symptoms (Jeffocate, 1969). Recent vaccine development for HRSV has focussed on live attenuated vaccines based on viruses bearing a variety of genetic alterations including deletion of NS, SH and M2-2 genes (Karron et al., 2015), incorporation of multiple nucleotide point mutations, or more recently a genome scale codon pair de-optimization (Le Nouen et al., 2017). In addition, subunit, cDNA and nano-particle vaccines are in development (Smith et al., 2012). Despite a lack of vaccine, two treatments are currently available for the treatment of HRSV, namely the non-specific anti-viral agent ribavirin, and an anti-F monoclonal antibody known as palivizumab that is thought to prevent F protein fusagenic activity.

Anti-viral medicines are currently a large area for HRSV treatment development with a significant number of compounds described in the literature and several in clinical trials, summarised in Table 3 (although none are yet FDA approved). Recent developments in structural determination of proteins and high-throughput screening (HTS) routines by technological advancement have accelerated this developing field with many proteins being targeted for a treatment including the F, L and M2-1 proteins.

1.9.1 Current HRSV treatments

As described above, there are currently only two FDA approved drugs intended to reduce the burden of HRSV infection, namely ribavirin and palivizumab. Both treatments are limited in their use and are expensive to administer. Ribavirin is the only FDA licensed drug for therapeutic HRSV treatment (that is, not as a prophylactic). Ribavirin is a nucleoside analogue that exhibits broad range anti-viral activity although the exact mechanism of action in relation to HRSV is unknown. Treatment delivers marginal efficacy, and in addition ribavirin has been

found to be carcinogenic and teratogenic, and is not routinely used (Hall et al., 1983; Sun et al., 2013).

Palivizumab (marketed using the trade name of 'Synagis') is an FDA approved humanized anti-F protein neutralising monoclonal antibody used in the prophylactic treatment of high-risk infants. Palivizumab interacts with a conserved epitope at antigenic site II of the F protein (Figure 4) and blocks both cell-cell and virus-cell F mediated fusion (Huang et al., 2010). Palivizumab is a well-tolerated, safe and effective treatment. An HRSV IMpact trial (1998) reported a significant reduction in HRSV associated hospitalisations with palivizumab treatment. In the case of HRSV infected children, treatment with palivizumab reduced hospitalisation by 39%, whilst for premature infants, treatment resulted in a 78% reduction. Overall, palivizumab treatment for HRSV reduced the total number of hospitalisation days, reduced the number of days where extra oxygen therapy was required, lowered the overall LRI score of all cohorts (LRI score: graded from 0-5, where 0= no respiratory infection and 5= mechanical ventilation) (Feltes et al., 2003) and reduced subsequent wheezing of treated premature infants (Fenton et al., 2004; Simoes et al., 2007). However, the palivizumab treatment regime requires repeat administration and is costly, making it out of reach for many in low income countries within much of the developing world, where mortality associated with HRSV is most prevalent.

Motavizumab was derived from palivizumab by rational engineering of the F protein binding site and targets the same antigenic site on the post fusion F protein (McLellan et al., 2011). Motavizumab varies by 13 amino acids and has picomolar affinity for HRSV compared to nanomolar affinities for palivizumab. However, motavizumab did not decrease viral load or disease severity in infants compared to the already available palivizumab and development was discontinued (O'Brien et al., 2015).

1.9.2 HRSV vaccine development

The development and use of vaccines are the preferred strategy for the mitigation of many viral infections. Currently no vaccine treatment exists for HRSV however vaccine trials have been attempted and are currently in development. The formalin inactivated HRSV vaccine trialled in 1969 impeded the development of an HRSV vaccine due to the enhanced disease severity and subsequent mortality of two infants in the trial (Jeffcoate, 1969; Collins & Graham, 2008). The exact cause of the disease enhancement remains unclear,

but has been suggested to be due to T cell priming, resulting in the recruitment of abundant inflammatory infiltrate to the lung and subsequent exacerbation of disease (Delgado et al., 2009).

A multitude of growth-attenuated live vaccine candidates have been designed and trialled over a period of several decades, involving both rational and stochastic methods for growth attenuation. An intranasal live attenuated virus (HRSV MEDI ΔM2-2), lacking the M2-2 gene was trialled in seronegative children and produced significantly higher titres of HRSV neutralising serum antibodies compared to the previous leading live attenuated vaccine candidates. Furthermore, in subsequent seasons, seronegative vaccine recipients had an increase in antibodies without reported illness showing that the vaccine elicited a protective response in these naïve children (Karron et al., 2015). An alternative approach to generating a live attenuated vaccine is by introducing suboptimal nucleotide substitutions in 9 of the 11 HRSV ORFs, which has resulted in a stable vaccine candidate with improved attenuation characteristics and high immunogenicity (Le Nouen et al., 2017). HRSV F protein nanoparticle vaccines (Novavax E202 & E301), displaying antigenic site II, recognised by palivizumab, was trialled in women of childbearing age and was safe, immunogenic, reduced HRSV infection and did not have any vaccine related adverse side-effects. In further phase 2 and 3 clinical trials E303 did not show efficacy and current work is continuing with E202 to provide a vaccine for use in maternal immunisation for protection of infants against HRSV (Smith et al., 2012; Glenn et al., 2013; Glenn et al., 2015).

1.9.3 Anti-viral compounds in clinical trials

Anti-virals are currently approved (FDA) and have revolutionised the treatment of chronic viral infections such as hepatitis C virus (HCV) and human immunodeficiency virus (HIV). For HCV, anti-viral therapies have reduced the need for interferon therapy and shown a sustained virological response coupled with an improvement of the patients' quality of life. One approved anti-viral therapy is sofosbuvir, a nucleotide analogue that inhibits the NS5B protein in a broad range of HCV genotypes. Additional anti-viral compounds are approved and target the NS3 protease (simeprevir) and the NS5A protein involved in viral replication and assembly namely, ledipasvir, ombitasvir and daclatasvir (Feeney & Chung, 2014; Banerjee & Reddy, 2016). The discoveries of HIV and influenza anti-viral compounds are discussed further in chapter 4. A number of anti-viral

compounds are currently in clinical trials for the treatment of HRSV and are summarised in (Table 3). Many studies have been undertaken to provide a better understanding of the therapeutic window for anti-viral treatment and greater knowledge of the replication cycle and the availability of biomarkers to provide a higher success rate for these compounds.

Compound	Company	Target	Reference
JNJ-53718678	Janssen & Janssen	F protein	Battles et al., 2015
GS-5806	Gilead Sciences	F protein	DeVincenzo et al., 2014 Jordan et al., 2015
PC786	Pulmocide	Non-nucleoside polymerase inhibitor (L)	Personal correspondence, RSV 16
ALS-008176	Alios	Nucleoside analogue (L)	DeVincenzo et al., 2015
ALX-0171	Ablynx	F protein	Personal correspondence, RSV 16
REGN2222	Regeneron	F protein	Griffiths et al., 2017

Table 3 Anti-viral compounds targeting HRSV currently in clinical trials

The F protein has been an attractive target for small molecule anti-viral compounds due to its HRSV neutralising activity, role in viral binding and fusion, and the successful development of peptide inhibitors against the HIV-1 entry protein, gp41. Despite this, many compounds have been discontinued due to insufficient pharmacokinetic properties (reviewed by Sun et al., 2013) and as such are not discussed here, with this section focussing on the current promising therapeutics.

The Janssen & Janssen oral F protein inhibitor (JNJ-53718678) has just completed phase 2 clinical trials (Identifier: NCT02387606). Little is known on the outcome of the trials or the inhibitor but there is currently recruitment for a phase 3 clinical trial in hospitalised infants. GS5806 (Presatovir) is based on the discontinued compound, JNJ 2408068. A co-crystal structure of JNJ-2408068 bound to the F protein displayed an interaction with a cavity in the 3-fold symmetric pocket of the pre-fusion form of F. The compound stabilised the pre-fusion conformation by tethering two regions that must undergo rearrangement for membrane fusion and virus entry. However, inhibitor escape mutants were shown to have fitness cost to the virus and did not propagate well compared to the wild-type virus (Battles et al., 2016). In contrast, a single dose of GS5806 (half life - 35 hours) in a phase 2 clinical trial was sufficient to decrease viral load and symptom severity whilst having a good safety profile (DeVincenzo et al., 2014). The trial determined resistant mutants (Phe 140 Leu and Thr 400 Ile) and further to this the current trial (Identifier: NCT01797419) is looking at a dosing regiment of GS5806 to lower the mutation rate (Jordan et al., 2015).

ALX00171 (Ablynx) is a trivalent nanobody (a single domain antibody fragment that contains the structural and functional properties of heavy-chain antibodies) and binds F resulting in neutralisation of HRSV. It is delivered directly to the site of HRSV infection by nebulisation and only requires a once-a-day dose. ALX-0171 showed good efficacy in a lamb model with 87% neutralisation and a 10,000 fold decrease in viral titres (correspondence at RSV16 conference). Recruitment for further clinical trials in hospitalised infants (Identifier: NCT02979431) for ALX00171 is currently underway (Ackermann, 2014). REGN2222 (Regeneron) is an intramuscular administered human monoclonal antibody that inhibits the F protein (like palivizumab) and is presumed promising as a drug candidate as subjects are now being recruited for a phase 3 clinical trial in preterm infants (CT Identifier: NCT02325791) (reviewed in Griffiths et al., 2017).

PC786 (Pulmocide) acts by inhibiting mRNA capping and cap methylation by the L protein and is administered via inhalation. In a phase 2 clinical trial with early intervention and a once-a day dosage, the virus was completely cleared. With later intervention a higher dose of PC786 was required and a longer time of action was seen, however the virus could still be cleared. PC786 proved efficacious for up to 48 hours after dosage and could provide a once-daily treatment option. ALS008716 from Alios is a potent and selective inhibitor of the L protein via a chain termination mechanism. The active metabolite of

ALS008716 in the respiratory tract showed a long half-life (~ 29 hours). The current phase 2 clinical trial showed a decrease in disease severity coupled with an increase in viral clearance and decrease in viral load (DeVincenzo et al., 2015). Over 25 nucleoside analogues have been approved for viral infections alone and include hepatitis B & C, HIV and herpes (reviewed in Jordheim et al., 2013) and suggest a potentially promising route for HRSV anti-viral drug design.

1.9.4 Anti-viral compounds of further interest

Anti-viral compounds that inhibit HRSV but have so far not entered clinical trials are discussed here. A high-throughput screen (HTS) assessing HRSV replication found three inhibitor compounds from the AZ compound collection termed compound 1, 2 and 3 (Laganas et al., 2015). The three compounds were further assessed and each inhibited HRSV replication via different mechanisms. Escape mutations mapped compound 3 to the P protein (D231V) whilst compound 2 showed no frequent mutation of HRSV and is presumed to inhibit the host's cell cycle. Escape mutations mapped compound 1 to the L protein (Tyr 1631 Cys), and it is thought to share a similar mode of action with a compound termed AZ-27, that prevents RNA replication and transcription by inhibiting the correct structural formation of L for initiation (Tiong-Yip et al., 2014; Noton et al., 2015). Despite positive data of these three novel compounds in inhibiting HRSV replication from a HTS, they showed limited inhibition in a viral RNA synthesis assay using the isolated HRSV replication complex and further work is required to improve the potency and solubility of these compounds (Laganas et al., 2015).

The X-ray crystal structure of the NTD of the N protein was recently solved in complex with residues 1 and 2 of the P protein. Further, the co-crystal structure of a compound termed M76 was determined bound to the same site of N as the P residues 1 and 2. The structure of compound M76 bound to the N protein allowed the design of a pro-drug that displayed anti-viral activity by inhibiting HRSV replication *in cellulo*. Currently, M76 and the pro-drug only show a narrow therapeutic window and need improvement for longer lasting action (Ouizougun-Oubari et al., 2015).

1.9.5 Anti-viral compounds targeted to M2-1

The HRSV M2-1 protein is an attractive target for anti-viral compounds due to its obligatory requirement for viral function and little known homology to characterised mammalian proteins. Sequence analysis of 165 HRSV A2 strain isolates identified regions with highly conserved genome sequences that correspond to regions primarily in N, P, M, F, M2 and L genes.

Regions of high conservation, such as the M2-1 protein of the M2 gene provide a good target for anti-viral compound activity due to their lower rate of mutation and thus reduced subsequent inhibition of the compounds anti-viral function (Do et al., 2015). Recently, two groups have identified compounds that inhibit HRSV replication *in vitro*, with these compounds hypothesised to interact with M2-1.

Cyclopamine (Figure 16) is a well-known antagonist of the smoothened protein, a transcription factor involved in embryonic development, cell differentiation and tumorgenesis (Chen et al., 2002). Its analogue, jervine, was discovered from a screen of known bioactive compounds to inhibit HRSV replication *in vitro* (post adsorption infection; cyclopamine IC_{50} 116 nM; jervine IC_{50} 994 nM). Serial passage of HRSV in the presence of cyclopamine resulted in an escape variant bearing a single mutation (Arg 151 Lys) within M2-1 that maps to the established RNA and P binding site (Blondot et al., 2012; Tanner et al., 2014). Further analysis of cyclopamine showed inhibition of M2-1 and a reduction in M2-1 expression in a mini genome system whilst *in vivo* testing of cyclopamine revealed a significant reduction in mice lung viral titres (Bailly et al., 2016).
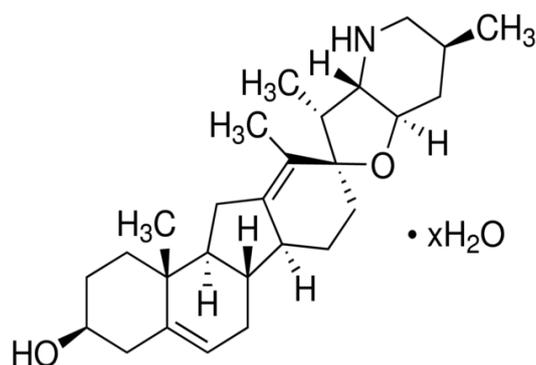
Figure 16 Chemical structure of cyclopamine

Cyclopamine resulted in a decrease in HRSV replication and transcription in a mini genome assay. Reduced M2-1 levels were visualised via an immunofluorescence assay and a reduction in viral titres using a mouse model. Escape mutations were mapped to the RNA/ P binding surface of the M2-1 protein.

A second class of M2-1 inhibitors have been developed from *in silico* docking to the ZBD of M2-1 using the crystal structure solved by Tanner et al, (2014). A zinc-ejecting compound aldrithiol (AT-2; Figure 17) that interacted with the zinc finger of retrovirus nucleocapsid proteins (Rossio et al., 1998), reducing infectivity, also inhibited HRSV infectivity through interactions with M2-1. An AT-2 inactivated HRSV vaccine induced an immunopathological response in cotton rats with no vaccine-enhanced diseases whilst an adjuvanated form of AT-2 resulted in vaccine-enhanced disease (Boukhvalova et al., 2010). Rationally designed compounds based on the AT-2 compound reduced HRSV replication *in cellulo*, with compound 10e being the most potent (IC_{50} 10e 6 µM; IC_{50} AT-2 841 µM). However, unlike AT-2, compound 10e has not been shown to be an M2-1 specific inhibitor to date (Cancielleri et al., 2015).
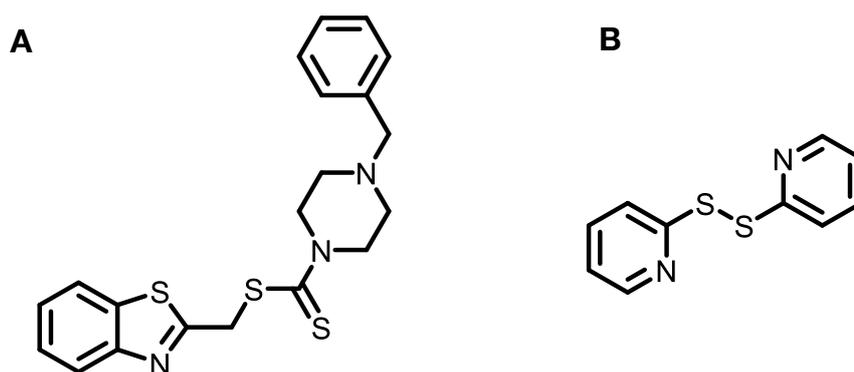
Figure 17 Structure of synthetically derived 10e compound and AT-2

A- compound 10e; B- compound AT-2. Compound 10e was synthetically derived based on the AT-2 compound, a zinc ejector known to inactive the nucleocapsid of HIV-1 by interactions with the zinc finger motif. AT-2 was shown to inactive HRSV, and was presumed to act through at the ZBD of M2-1, HRSV. Compound 10e and its derivatives that reduced HRSV replication all contained a common benzyl piperazine moiety on the scaffold.

1.10 Project Aims

The HRSV M2-1 protein is essential for virus multiplication with a primary role in synthesis of full-length viral mRNA transcripts and enhancing read-through at gene-end/ gene-start junctions. Conservation of critical residues for M2-1 function in addition to no known mammalian homologs of the protein suggest that M2-1 could be used as a target for anti-viral drug design. The recently solved crystal structure of HRSV M2-1 has provided much insight into its structure and function, however further understanding of the interactions between M2-1 and its binding partners, RNA and P, is required to determine the precise function of HRSV in the virus life cycle. The aims of this project were 3 fold; firstly, to provide further insight into the interaction of M2-1 with its binding partners RNA and P by structural and biophysical methods. Secondly, to assess the suitability of M2-1 as a drug target via consideration of the RNA/ P binding site and further potential inhibitor sites. Finally, this project aimed to characterise the bovine respiratory syncytial virus (BRSV) M2-1 protein to provide similarities with HRSV M2-1 that may allow cross-species treatment with the determined M2-1 inhibitors.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Recipes

Lysogeny broth (LB): per litre in Milli-Q H_2O (18.2 MΩcm^{-1}) 16 g tryptone 10 g yeast extract 5 g NaCl (Fisher Scientific)	RNA binding buffer: 20 mM tris-Cl pH 7.5 150 mM NaCl (Fisher Scientific) 0.01% triton X-100
Protein expression- lysis buffer: 25 mM tris-Cl pH 7.5 1 M NaCl 1 mM DTT 2% triton X-100 10 mM $MgSO_4$ 1 mM $CaCl_2$ 0.5 mg/ml lysozyme 5% glycerol 1x complete protease inhibitors EDTA-free (Roche) 1 µg/ml DNase 1 µg/ml RNase	Protein expression- wash buffer: 25 mM tris-Cl pH 7.5 150 mM NaCl (Fisher Scientific) 1 mM DTT 5% glycerol 0.1% triton X-100
His bind buffer: 150 mm NaCl 50 mM tris pH 7.4	Glutathione elution buffer: 20 mM L-glutathione (Alfa Aesar), adjusted to pH 7

Agar medium: With LB as above, supplemented with 1.5% agar and a selective antibiotic (ampicillin 100 µg/ml or kanamycin at 40 µg/ml)	Size exclusion chromatography (SEC) buffer: 25 mM tris-Cl pH 7.5 150 mM NaCl (Fisher Scientific) 1 mM DTT
Cation exchange- low salt buffer: 20 mM tris-Cl pH 7.5 50 mM NaCl 5% glycerol 1 mM DTT	Cation exchange- high salt buffer: 20 mM tris-Cl pH 7.5 1 M NaCl 5% glycerol 1 mM DTT
SDS-PAGE 15% resolving gel: 4.6 ml milli-Q H_2O (18.2 MΩcm^{-1}) 10 ml 30% acrylamide mix 5 ml tris-Cl (1.5 M pH 8.8) 200 µl SDS (10%) 200 µl 10% ammonium persulfate 10 µl TEMED 5% stacking gel: 3.4 ml milli-Q H_2O (18.2 MΩcm^{-1}) 830 µl 30% acrylamide mix 630 µl tris-Cl (1.0 M pH 6.8) 50 µl SDS (10%) 50 µl 10% ammonium persulfate 5 µl TEMED	Agarose gel electrophoresis: 50x TAE buffer (1 l) pH 8.5: 0.5 M of EDTA 2 M Tris Base (Melford) 57.1 ml of Glacetic acid Agarose gel (1.5%): 1.5 g agarose (Melford) per 100 ml of 1x TAE SYBR® safe (Invitrogen Life Technologies) at 10,000 x concentration Loading dye: 0.25% bromo B and 40% sucrose (SLS)

SDS-PAGE running buffer:	SDS-PAGE sample loading dye:
25 mM tris Cl pH 7.5 0.1% SDS 190 mM glycine pH 6.8	100 mM tris 4% SDS 20% glycerol 200 mM DTT 0.3% bromophenol blue pH 6.8
Coomassie blue de-stain: 30% methanol (Scientific Laboratory Supplies-SLS) 60% milli-Q H_2O (18.2 M′Ωcm^{-1}) 10% acetic acid	Coomassie blue stain: 45% methanol (SLS) 45% milli-Q H_2O (18.2 M′Ωcm^{-1}) 10% acetic acid (SLS) 0.25% coomassie brilliant blue R-250
CD buffer: milli-Q H_2O (18.2 M′Ωcm^{-1}) 10 mM NaPo4 150 mM NaCl	

Table 4 Consumables and recipes used for further assays

All reagents were purchased from Sigma-Aldrich, unless otherwise stated.

2.1.2 Consumables

2.1.2.1 Primers

Primers were synthesised by Sigma-Aldrich for Gibson assembly cloning (FP-forward primer, RP- reverse primer).

Name	Primer sequence (5'-3')
ΔLGS FP1	GGAAGGTCTGTTCCAGGGGCC CATGTCACGAAGGAATCCTTGC
ΔLGSMS FP1	GGAAGTTCTGTTCCAGGGGCCC CGAAGGAATCCTTGCAAATTTG
ΔLGS/MS RP1	CGATGCGGCCGCTCGAGTTAG GTAGTATCATTATTTTTGG
M2-1 $_{58-177}$ FP1	CGCGGATCCTCAGAAATAAGTGGA
M2-1 $_{58-177}$ RP1	CCGCTCGAGTTATTTTGGGTTGTTGAT

Table 5 Gibson assembly primer design

Primers were synthesised by Sigma-Aldrich for Gibson assembly cloning (FP- forward primer, RP- reverse primer).

2.1.2.2 Crystal screens

Commercial screens were utilised for the initial identification of crystals. Screens included; Salt RX, Crystal, Index (Hampton Research), Wizard 1 & 2, Wizard 3 & 4, Midas, Morpheus (Molecular Dimensions), JCSG Core Suite I-IV and PACT (Qiagen).

2.1.2.3 Cell lines

A BSRT7 cell line, derived from BHK21 cells and constitutively expressing the bacteriophage T7 RNA polymerase (described by Buchholz et al., 1999) was a kind gift from Dr Cheryl Walter, University of Leeds. A human epidermal carcinoma (Hep2) cell line was a kind gift from Dr Hussah Taqi, University of Leeds.

2.1.2.4 Un-labelled RNA synthesis

Oligoribonucleotides of various lengths (6mers-14mers) were synthesised by Dharmacon (GE Healthcare Sciences) and HPLC purified. RNA was provided

converted to the 2'-hydroxyl form and desalted ready for use. Oligoribonucleotides were re-suspended in diethylpyrocarbonate (DEPC) treated water and stored in RNase free 1.5 ml tubes at a concentration of 1 mM.

2.1.2.5 Labelled RNA synthesis

Oligoribonucleotides of 13 nucleotides in length were synthesised with a 3'-fluorescein label (Thermo Scientific). 3' Fl RNA was provided in the stable 2'-ACE® protected form to improve stability and enhance nuclease resistance. RNA in the stable 2'-ACE® protected form was de-protected with 400 μl of 2' de-protection buffer (100 mM acetic acid, adjusted to pH 3.8 with TEMED) or 800 μl for oligomers with a homopolymer A stretch of more than 12 bases. The RNA pellet was completely dissolved, vortexed, and incubated at 60°C for 30 minutes (or 2 hours for oligomers with 12 base or longer A stretches). The solution was vacuumed to dryness (SpeedVac) before re-suspension. All RNA was re-suspended in DEPC treated water and stocks stored in RNase free 1.5 ml black tubes at a concentration of 1 mM.

2.1.2.6 P 90- 110 peptide synthesis

A short region of the HRSV A2 strain P protein (residues 90-110) was synthesised by ProteoGenix SAS in an unlabelled and fluorescently labelled form. The peptide was purified by HPLC and deemed 98.66% pure with a MW of 2444.66 g/mol by mass spectrometry. The peptide was provided in a lyophilised form and stocks of 1 mM produced using DEPC treated. The synthesised amino acid peptide sequence was 5'-3': DPTPSDNPFSKLYKETIETFD.

2.1.2.7 Compounds

Compounds were computationally screened using a free database of commercially available compounds, known as Zinc 12 (Irwin & Shoichet, 2005). Hits were purchased from a variety of vendors as stated in appendix IV. Compound stocks were produced at a concentration of 10 mM with DMSO and stored at -20°C.

2.1.2.8 Fragments

Two fragment libraries were provided by the DLS (grant number: lb14607-1) for fragment based drug design (FBDD). Libraries were in a solution of DMSO and stored at 4°C. The Edelris library contained 280 fragments at a stock concentration of 250 mM. The second library, Maybridge 1000, contained 856 fragments at a stock concentration of 100 mM.

2.2   Methods

2.2.1   Molecular biology

2.2.1.1 Cloning of HRSV M2-1 cDNA into the pGEX-6P-2 vector

A cDNA representing the HRSV (A2 strain) M2-1 open reading frame (ORF) was a kind gift from Professor Gail Wertz (Hardy and Wertz, 1998). The M2-1 ORF was sub-cloned (Tanner et al., 2014) into the pGEX-6P-2 vector (GE Healthcare), that contained sequences encoding an N-terminal glutathione S-transferase (GST) tag followed by the recognition sequence for PreScission Protease. The position of this affinity tag is such that five additional non-native residues (Gly-Pro-Leu-Gly-Ser) remain at the N–terminus of M2-1 following cleavage.

2.2.1.2 Cloning of BRSV M2-1 cDNA into the pGEX-6P-2 vector

A cDNA representing the BRSV M2-1 A2 strain ORF (Accession No: P29792) was custom synthesised (Invitrogen Life Technologies) flanked by restriction sites *Xho1* and *BamHI* (NEB) that would allow for insertion by restriction digest into the pGEX-6P-2 vector used for expression of HRSV M2-1. Like HRSV, BRSV M2-1 contained 5 non-native N-terminal residues (Gly-Pro-Leu-Gly-Ser) after cleavage of the GST affinity tag.

2.2.1.3 Cloning of truncated HRSV M2-1 cDNA into the pGEX-6P-2 vector

M2-1 N and C terminus truncations were constructed using Gibson Assembly (NEB) as per the manufacturers instructions. The first truncation removed 3 non-native residues (ΔLGS) at the N-terminus of M2-1 and the second removed the 3 non-native residues in addition to native residues 1 and 2 of the M2-1 protein (ΔLGSMS). A monomeric form of M2-1 (Blondot et al., 2012) was created by removal of residues 1-57 and 178-194, with the resultant protein termed $M2\text{-}1_{58\text{-}177}$. All plasmids were cloned into the pGEX-6P-2 vector. The vector was cut using restriction enzymes *BamHI* and *Xho1* (NEB) before gel purification (Zymo Research) following the manufacturers instructions. Inserts for Gibson Assembly were amplified by a polymerase chain reaction (PCR) protocol (Table 6). Resultant cDNA plasmids were submitted to Sanger sequencing for sequence conformation (Beckmann Coulter Genomics).

	Temperature	Time	Cycles
Cycle 1	96°C	2 minutes	1 cycle
Cycle 2	94°C	30 seconds	4 cycle
	66°C	1 minute	
	72°C	1 minute	
Cycle 3	94°C	30 seconds	25 cycles
	75°C	1 minute	
	75°C	1 minute	
Cycle 4	72°C	7 minutes	1 cycle

Table 6 Thermal cycler settings for HRSV M2-1 truncation PCR reactions

2.2.1.4 Bacterial transformation

For efficient transformation 50 μl of competent *E.coli* cells were thawed on ice and 1 μl of 50 ng/μl plasmid DNA added. Samples were incubated on ice for 30 minutes, followed by heat shock at 42°C for 45 seconds. Samples were placed back on ice for a further 2 minutes before addition of 500 μl of LB and incubation at 37°C for 2 hours with shaking at 170 rpm. Samples were then plated on antibiotic-selective LB agar plates and incubated at 37°C for 16 hours. Amplification of plasmid DNA was performed by transformation into DH5α *E.coli* bacterial competent cells (Life Technologies) whilst BL21-Gold (DE3) cells (Agilent Genomics) were used for protein expression.

2.2.1.5 Plasmid DNA extraction

Prior to alkaline lysis, one colony from LB Agar plates containing the transformation mixture was picked with a pipette tip and placed in 5 ml of LB and a selective antibiotic in a vessel of 20 ml volume. The mixture was placed for 16 hours at 37°C with shaking at 220 rpm. DNA extraction was performed using a Zymo Research DNA extraction kit and performed as per manufacturers instructions.

2.2.2 Protein purification

2.2.2.1 Starter cultures

A single colony from the antibiotic-selective LB agar plates was picked and placed into 10 ml of an LB starter culture containing a selective antibiotic. Cultures were incubated at 37°C for 16 hours with shaking at 170 rpm. For large-scale protein expression 60 ml of starter culture was used.

2.2.2.2 Protein expression

Two litres of LB (1 L of LB in a 2 L conical flask) were inoculated with starter culture and incubated at 37°C, with shaking until cells reached $O.D_{600}$ 0.8. At this time point cells were induced with 340 μM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 50 nM of $ZnSO_4$ and incubated at 28°C, with shaking for 16 hours. Cells were then centrifuged at 4000 xg for 30 minutes at 4°C. The bacterial pellet was re-suspended in lysis buffer (100 ml per 1L of LB). The suspension was then freeze-thawed for four cycles before the addition of DNase and RNase. Lysates were centrifuged at 16000 xg for 30 minutes at 4°C. The supernatant was then applied to Glutathione 4B Resin (GE Healthcare) by gravity flow before washing with lysis buffer and equilibrating with wash buffer. PreScission Protease (GE Healthcare) was added to the resin and incubated with gentle rotation overnight at 4°C. The protein was eluted from the resin by gravity flow with the addition of 2 column volumes (CV) of wash buffer.

2.2.2.3 Ion exchange chromatography

SP sepharose resin was equilibrated in 1 CV of low salt (50 mM) followed by 1 CV of high salt (1mM) and finally 2 CV of low salt buffer. Eluted protein (section 1.2.2.2) was diluted from 150 mM to 50 mM NaCl and subsequently added to the resin. One column volume of a 15 mM NaCl solution was added to the resin before collection of the M2-1 protein via addition of a 90 mM NaCl solution to the resin. Ten, 5 ml elutes were collected before the addition of 1 CV of a 150 mM high salt solution. All samples were analysed by SDS-PAGE to ensure further impurities were removed.

2.2.2.4 Size exclusion chromatography (SEC)

A 320 ml HiLoad 26/600 Superdex 75 pg column (GE Healthcare) was equilibrated with SEC buffer. All SEC buffers were filtered (0.22 μm) and de-gassed (Merck Millipore). The column was kept at 4°C and used with an ÄKTA prime pump system (GE Healthcare) that measured absorbance at 280 nm. The desired protein sample was concentrated to 5 ml and filtered using a 0.22

μm filter (Sartorius) before injection into the ÄKTA using a 5 ml loop. The column was set to a flow rate of 2 ml/min and 3 ml fractions were collected after the initial void volume of 110 ml.

2.2.2.5 SDS-PAGE gels and Coomassie staining

SDS-PAGE gels were prepared with materials stated in section 2.1. Gels were run at 150 V for 60 minutes before staining in Coomassie blue for 15 minutes followed by the addition of de-stain.

2.2.2.6 Cross linking gluteraldehyde

Protein was initially buffer exchanged in to CD buffer (section 1.1.1). Protein at a concentration of 1.5 mg/ml was then mixed with varying concentration of gluteraldehyde (0- 40 mM) in a 1:1 ratio. The mixture was incubated for 30 minutes at room temperature before the addition of SDS loading dye and boiled at 95°C for 10 minutes to stop the reaction. Samples were analysed by SDS-PAGE as above.

2.2.2.7 Protein concentration and quantification

Protein was concentrated for SEC (5 ml) and for final storage (100 μl). Protein was concentrated using 20 ml, 10,000 molecular weight cut-off (MWCO) polyethersulfone membrane concentrators (Thermo Fisher) and spun at 3,800 rpm for 20 minutes. For smaller protein volumes, 0.5 ml Amicon Ultracel® low-binding regenerated cellulose membrane concentrators with a 10,000 MWCO (Merck Millipore) were used and spun at 13,000 rpm.

Protein concentration was determined by dilution of the protein to ensure the 280 nm absorbance value was in the range of 0.1- 1 using a 10 mm path length quartz cuvette (UV spectrophotometer- Astbury BioSuite). Further absorbance readings at 260, 300 and 340 nm were recorded. To determine the protein concentration (mg/ml) the absorbance value at 340 nm was subtracted from the 280 nm reading and divided by the extinction coefficient multiplied by the path length (Wilfinger et al., 1997).

2.2.3 X-ray crystallography (general)

2.2.3.1 Crystal factorials

Protein was mixed with equal amounts of mother liquor (0.2 µl of protein: 0.2 µl of mother liquor) unless otherwise stated using the NT8 (Formulatrix) or Mosquito (TTP labtech) liquid handlers. Screens were set up in 2 or 3- drop MRC crystallisation plates and sealed with ClearSeal Film™ (Hampton Research). Crystal plates were imaged using a Fibonacci series and the Rock Imager 1000 (Formulatrix) with visible and UV-TPEF imaging being recorded at 20°C. For co-crystal studies, RNA or P 90-110 was added in x1.2 molar excess.

2.2.3.2 Crystal optimisation

Crystal hits were optimised by finely altering the pH and PEG concentration of the mother liquor in addition to storage temperature of the plates. For consistency of protein formation, 96- well, 2 and 3- drop MRC crystallisation plates (Hampton Research) were used for initial screens and all optimisation.

2.2.3.3 Crystal harvesting

Crystals were picked in appropriate size nylon loops (Hampton Research) and cryo-cooled in their mother liquor substituted with 5 % glycerol, 5 % PEG 400, 5 % 2-MPD and 5 % ethylene glycol.

2.2.3.4 Data collection

Data was collected at the DLS on multiple MX beamlines. Crystals were firstly screened for diffraction at, 0°, 70° and 140° using 12,658 eV X-rays (0.9795 Å). Full data sets were obtained for crystal that diffracted below 3 Å with images recorded every 0.5°.

2.2.3.5 Molecular replacement

Data sets were integrated and auto-processed using the pre-defined pipelines available on the IsPyB at the DLS. The auto-processed mtz file (3d iii), known PDB co-ordinates (Tanner et al., 2014; PDB 4c3b and 4c3d) and protein sequence were placed into the CCP4 i2 software package for molecular replacement (Phaser). Further refinement was performed using refmac5, as part of the CCP4 i2 suite. All molecular images used in this report were created with The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.

2.2.4 X-ray crystallography (XChem facility)

This crystallography methods section refers specifically to the work performed using the XChem facility at the DLS.

2.2.4.1 Crystal condition

Crystal drops were set up using MRC 3- drop (also known as SWISS CI, Hampton Research) plates. Protein was added to drops 2 and 3 of the 3-drop plates. 0.2 μl of protein at 12 mg/ml was mixed with 0.2 μl of mother liquor using the NT8 (Formulatrix) liquid handler and sealed with ClearSeal Film ™ (Hampton Research). Plates were incubated at 20°C for 3 days prior to the DLS visit. One mother liquor condition was used to form crystals this was: 0.1 M carboxylic acid, 0.1 M BS3 pH 8.5 and 24% PEG550MME_P20K (Morpheus).

2.2.4.2 Crystal targeting

MRC 3-drop plates were imaged using the Formulatrix RockImager at the DLS and ranked using TexRankE. Crystal 'hits' were targeted by creating an x,y co-ordinate dispense site for the ECHO (Labcyte) dispenser in close proximity to the crystal.

2.2.4.3 Crystal soaking

The ECHO acoustic liquid handler used ultrasonic pulses to shoot 2.5 nl droplets at the dispense site previously stated in TeXRankE. The ECHO dispensed the fragment to the pre-specified location and was left to soak for one hour at room temperature. After this time 300 nl of the cryo-protectant mix was dispensed to the drop. The ECHO was set to dispense slower due to the viscosity of the cryo-protectant. Each condition was performed in duplicate for the optimisation screen and in singular for the full fragment screen.

2.2.4.4 Crystal harvesting

The process of crystal harvesting was streamlined by the use of the Shifter, a microscope x-y stage that automatically tracks sample positions. Crystals were harvested using 75 and 150 μM Mitegen loops. For the cryo protectant the mother liquor was substituted with 5 % glycerol, 5 % PEG 400, 5 % 2-MPD and 5 % ethylene glycol.

2.2.4.5 Data collection

Data was collected on the I04-1 beamline (Visit: lb14607-1). The Mitegen loops allow for good loop centring on the beamline and data collection was fully automated and unattended. Full data collection was performed for each crystal

with the following parameters: oscillation 0.12°, transmission 100%, exposure 0.08 s, aperture 70 μM, images collected 1500.

2.2.4.6 Data analysis and refinement

Data sets were automatically integrated using the pre-existing pipelines at the DLS. Data sets were automatically molecular replaced and electron density assessed using the PANDDA and XChemExplorer (XCE) software designed by the structural genomics consortium (University of Oxford) for the XChem fragment screening facility.

2.2.5 Biophysical assays

2.2.5.1 Fluorescence anisotropy (FA)

2.2.5.1.1 Direct binding assay

RNA binding buffer (20 μl) was added to each well of a black 384- well optiplate (Perkin Elmer). Protein was added to the first well at a final concentration of 110 μM and a 1 in 2 serial dilution of the protein was then performed. 3'-Fl labelled RNA 13mers (Thermo Scientific Ltd) were added to each well at a concentration of 40 nM. Control wells contained RNA binding buffer and protein only. The reaction was incubated at room temperature for 30 minutes. Polarisation was measured using an EnVisionTM 2103 MultiLabel plate reader (Perkin Elmer) that contained an excitation filter at 480 nm (30 nM bandwidth), polarised dichroic mirror at 505 nm and emission at 530 nm (40 nM bandwidth) S and P polarised. Anisotropy values were calculated using the raw emission values for S and P signals and the following equations:

$$intenisty = (2 \times P) + S$$

$$anisotropy = \frac{S - (P \times 1)}{intensity}$$

Anisotropy values were plotted in GraphPad Prism7 and a nonlinear regression curve fitted with the following equation:

$$y = \frac{A1 - A2}{1 + (X \div X^0)p} + A2$$

Where y is the anisotropy, X is the protein concentration, A1 is the initial value of y and A2 is the final value of y. X^0 is the dissociation constant (Kd) and p is the Hill coefficient.

2.2.5.1.2 Competition binding assay

For competition binding, 20 nM of 3'-Fl and 1 µM of protein (the amount of protein required to produce the Kd of P 90-110) (20 µl of each) was added to a black 384- well optiplate (Perkin Elmer) and incubated for 10 minutes at room temperature. Serial dilutions (1 in 2) of the unlabelled RNA or compound were performed in a separate well and 20 µl transferred (1 in 3 dilution) to the well containing protein and the 3'-Fl. The plate was incubated for 30 minutes prior to analysis using the EnVision plate reader. The concentration of un-labelled RNA added to the first well was 1 mM and for compounds the concentration was 10 mM.

2.2.5.2 Circular dichroism (CD) spectroscopy

Data was collected using the Jasco J715 spectropolarimeter (Applied Photophysics) purged with nitrogen gas. The sample temperature was controlled using a water cooled Peltier system and allowed for automated temperature ramping. Protein was buffer exchanged into CD buffer and diluted to a final concentration of 0.1 mg/ml. The data was collected as ellipticity (**θ**) in units of millidegrees (mdeg). The CD spectra was normalised to give the mean residues molar ellipticity for comparison of secondary structure between proteins. The equation used to determine the mean residues molar ellipticity was:

$$= \frac{raw\ signal}{protein\ concentration \times (Number\ of\ peptide\ bonds - 1) \times path\ length \times 10}$$

2.2.5.2.1 Collection of far UV CD spectra

UV spectra were collected in 1 mm path length quartz cuvettes. CD measurements were made in 1 nm increments, between 180 and 300 nm with a bandwidth of 2.5 nm. The temperature was maintained at 6°C.

2.2.5.2.2 Collection of thermal unfolding curves

A constant temperature ramp from 10- 90°C with an in-cell temperature probe was used to record the precise sample temperature at a wavelength of 222 nm.

2.2.5.3 Mass spectrometry analysis

All mass spectrometer analysis was performed as a service by the mass spectrometry facility, University of Leeds. For protein sequence identification and molecular mass determination, samples were provided in the form of a SDS-PAGE stained Coomassie gel.

2.2.6 *In cellulo* assays

2.2.6.1 Cell maintenance

Continuous cultures of all cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM, Sigma-Aldrich) supplemented with 10% foetal bovine serum (Sigma-Aldrich), 100 unit/ml penicillin and 100 µg/ml streptomycin (Invitrogen), herein referred to as complete media 10 (CM10). Cell lines were passaged by washing in 1x phosphate buffered saline (1x PBS, Sigma-Aldrich) followed by detachment using 1x Trypsin (Sigma-Aldrich). To maintain expression of the T7 RNA polymerase promoter, BRST7 cells were supplemented with 500 µg/ml of G418 antibiotic at alternate passages. For maintenance, cells were incubated at 37°C in a humidified atmosphere supplied with 5% CO_2. Assays measuring GFP were performed in OPTI-MEM (Gibco) reduced serum media with no phenol red.

2.2.6.2 HRSV mini genome system

The mini genome plasmids were a kind gift from the laboratory of Professor Julian Hiscox, University of Liverpool (described by Yu et al., 1995 and Grosfeld et al., 1995). A plasmid expressing a Rab5-GFP fusion protein was used as a positive control (Addgene plasmid: 31733) and was a kind gift from Professor Mark Harris, University of Leeds. One day prior to transfection cells were seeded in a 12-well dish at a confluency of 1x10^5 cells/well. On the day of transfection, 3 µl Lipofectamine 2000® Reagent (Thermo Fisher) was incubated with 250 µl (for 1 well) of OPTI-MEM (Sigma-Aldrich). In another tube 250 µl of OPTI-MEM was

incubated with the following mini-genome plasmids: 1 μg N, P, M2-1, GFP and 0.5 μ g of L. After an incubation time of 5 minutes the two tubes were mixed and further incubated for 20 minutes at room temperature. Media was then removed from wells containing cells and washed with 1x PBS before addition of the 500 μl Lipofectamine and DNA mixture. The transfected cell mixture was incubated for 4 hours before removal of the media and addition of fresh OPTI-MEM media.

Compounds were added to a final concentration of 20 μM (0.2% DMSO) at time of transfection and re-added 4 hours later following cell washing with 1x PBS. The mini genome assayed levels of green fluorescent protein (GFP) using the IncuCyte Zoom® with access kindly provided by Professor Nicola Stonehouse, University of Leeds. Transiently transfected cells were scanned and images recorded for green fluorescence and phase every hour, for 29 hours. Data was analysed using the IncuCyte ZOOM® software with the integrated algorithm and plotted as the total number of green fluorescent cells per well (described in Tulloch et al., 2014). Each compound was assayed in duplicate at a single concentration of 20 μM, to an n-2.

2.2.6.3 MTT (3-(4,5-dimethylthiazol-2-yl)- 2,5-diphenyltetrazolium bromide) assay

Clear 96-well plates were seeded with 5×10^5 BSRT7 cells/ well one day prior to transfection. Prior to the assay media was removed from the cells, washed with 1x PBS, and replaced with 100 μl of media plus relevant compound at a final concentration of 20 μM. The mixture was incubated at 37^0C with 5% CO_2 for 33 hours to mimic the conditions used for the mini genome assay. After 33 hours all media was removed from the plate and replaced with a solution of serum free media containing 1 mg/ml of MTT powder (Sigma-Aldrich). The assay plate was further incubated at 37°C for 30 minutes away from light sources. After this time, the MTT solution was discarded and replaced with DMSO. The assay plate was shaken at 60 rpm for 5 minutes to dissolve the precipitate followed by the recording of absorbance at 570 nm.

2.2.6.4 Infectious virus

Hep2 cells were seeded to be 70% confluent at time of infection. Recombinant HRSV-GFP, encoding GFP upstream of the SH gene was kindly gifted by Dr Hussah Taqi. Cells were infected with HRSV-GFP at a multiplicity of infection (MOI) of 0.1 cell infectious units (CIU). Flasks were gently rocked for 10 minutes before incubating at 37°C, 5% CO_2 for 5 hours. At this time cells were washed with 1x PBS, replaced with CM10 media and incubated for 5 days. After 5 days autoclaved glass beads were added to the flasks to assist in mechanical

disruption of the cells, after which the supernatant containing virus was decanted and spun for 3 minutes at 3,000 g before being snap frozen and stored at -80ºC.

2.2.6.5 Plaque assay

Hep2 cells were seeded at $2x10^5$ cells/well in a 24 well plate to be a confluent monolayer at time of assay. The virus suspension prepared as described above was thawed and diluted 1 in 10 in 1x PBS, serial dilutions of the virus were then performed 1 in 5. The plate was gently rocked for 20 minutes and then placed in a 37^0C 5% CO_2 incubator for 4 hours. After 4 hours the virus was removed and cells washed with 1x PBS followed by the addition of 1.6% carboxymethyl cellulose mixed in a 1:1 ratio with CM10 to each well. The mixture was further incubated for 5 days.

After 5 days the mixture was removed and cells washed 4 times with 1x PBS. Cells were then fixed with 80% cold methanol for 1 hour before being washed 3 times with 1x PBS- 0.1% Tween. Blocking buffer (5% skimmed milk in 1x PBS) was incubated with the cells for one hour at room temperature. Primary antibody (goat anti-RSV; Abcam 20745) was diluted 1 in 200 in blocking buffer, added to each well and incubated overnight at 4 degrees. Secondary antibody (horseradish peroxidase anti-goat; Abcam ab6741) was added to each well at a 1 in 1000 dilution in blocking buffer and incubated with rocking for an hour at room temperature. 4-chloro-1-napthol (Thermo Scientific) in methanol was diluted 1 in 10 in 1x PBS, supplemented with 1 in 1000 of 30% hydrogen peroxide and added to each well. The plate was further incubated at room temperature with rocking until the formation of purple plaques was visible. Titre (PFU/ml) was determined by the following equation:

$$Titre = \frac{number\ of\ plaques}{dilution\ factor\ x\ volume\ of\ diluted\ virus}$$

2.2.6.6 Live virus assay

2.2.6.6.1 Single compound dose

BSRT7 cells were seeded in a 24 well dish at a confluency of $0.05x10^6$ cells/well 24 hours prior to infection. Cells were infected with HRSV-GFP at an MOI of 4 (diluted in CM10) in addition to 20 μM of the compound (0.2% DMSO). Cells infected with HRSV-GFP were then rocked at 37ºC for 20 minutes before incubating at 37ºC, 5 % CO_2 for 5 hours. After this time cells were washed twice with 1x PBS and replaced with OPTI-MEM and 20 μM of compound.

Plates were imaged hourly and green fluorescence levels recorded using the IncuCyte Zoom® for a total of 33 hours. Optimal parameter values were pre-determined by the IncuCyte software (and used for the mini genome assay), however the green calibrated unit (GCU) value was altered for the live virus assay and was set to a minimum of 2.0, with green fluorescent objects below this threshold level classified as background. The mean number of green fluorescent cells per image was calculated for 8 non-overlapping images and used to extrapolate the mean number of green cells per well using the IncuCyteZOOM® software. Data was analysed using the integrated algorithm and plotted as the total number of green fluorescent cells per well (described in Tulloch et al., 2014). Each compound was assayed in duplicate at a single concentration of 20 µM, to an n-3.

2.2.6.6.2 Compound titration

The titration assay was performed and analysed as above for a singe compound concentration, however the 20 µM compound added to the initial well was successively diluted 2.5 times for each dilution, with the last dilution void of compound.

Chapter 3 Protein purification of the M2-1 protein of human respiratory syncytial virus (HRSV) and its interactions with viral RNA and phosphoprotein (P)

3.1   Introduction

Structure and function studies often require high quantities of homogenous and functional protein, and a common approach to generate this is the overexpression of recombinant proteins using a heterologous host such as *E. coli*. In this case, a cDNA representing the ORF of the protein of interest is inserted behind an IPTG inducible promoter within a specialised expression plasmid, and fused to an affinity tag to assist in purification. IPTG induction alleviates repression of the target gene, which allows gene expression by RNA polymerase, and subsequent high level production of the encoded fusion protein. By performing such inductions in large scale cultures, high quantities of recombinant proteins can be expressed and then purified to near homogeneity using various downstream chromatographic techniques that can, in turn, be used for structural and functional studies.

3.2   Principles of chromatography

Chromatographic techniques include affinity chromatography that purifies proteins based on ligand specificity, SEC that separates proteins based on their size and ion exchange that purifies based on the charge of the proteins. For purification of HRSV M2-1, all three of these chromatography techniques were utilised.

During affinity chromatography, the stationary phase consists of a support medium (glutathione sepharose media) on which the substrate (glutathione) is covalently attached. Proteins, in this case a GST fused M2-1 protein, bind to the glutathione substrate in the stationary phase whilst non-specific proteins are eluted. The tagged protein can then be eluted, and the media restored to an unbound form via incubation with an excess of GST. SEC media is a porous matrix of spherical particles, in which the buffer that fills the porous matrix (stationary phase) is in equilibrium with the buffer on the outside of the matrix (mobile phase). A sample containing protein is loaded and the matrix traps smaller, low molecular weight proteins whilst higher molecular weight proteins can not enter the pores and are eluted earlier. To perform ion exchange chromatography the isoelectric point (pI) of the protein is determined and a buffer

pH chosen to create a positive or negatively charged environment. Based on the charge identified for the protein, an oppositely charged resin (negative charged-cation or positively charged- anion) is chosen to bind the protein of interest, whilst contaminants are eluted in the void volume. The buffer condition to bind the protein of interest is usually 0.5-1.5 pH units above or below the pI depending on the resin type used.

3.3   Analysis of protein structure

Once a streamlined methodology for the production of pure protein has been established, the protein can be used for structural studies that include X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) and cryo-electron microscopy (EM). X-ray crystallography has been the most desirable method for atomic resolution structural information for over 60 years, however recent advancements in transmission electron microscopes and electron detectors has allowed cryo-EM to compete with X-ray crystallography with near atomic resolution now being observed for structures determined by cryo EM, in what has been termed the 'resolution revolution' (Kühlbrandt, 2014). Further, the use of cryo-EM for high resolution structural information bypasses the phase problem seen in X-ray crystallography as these can be directly recorded from the density map (Cheng, 2015).

The main requirement of X-ray crystallography that differs from other structural techniques is the need to form protein crystals for structural determination. NMR and cryo-EM require only a protein in solution for near atomic resolution structural determination and may provide a more physiologically relevant and dynamic snapshot of the protein compared to that seen in rigid protein crystals. Whilst X-ray crystallography is suitable for structural determination of any size protein (although increased difficulties are seen with larger complexes), both EM and NMR methods have limitations in the size of proteins that can be used. Proteins with a molecular weight above 100 kDa are required for cryo-EM whilst NMR requires proteins with a molecular weight below 50 kDa.

Depending on the required information for the protein of interest each structural technique has its advantages, however, with advancements in access to these structural methods multiple techniques can be performed in parallel resulting in a detailed structural knowledge of the proteins being studied. For HRSV M2-1 studied here, X-ray crystallography was the preferred method for structural determination due to the existence of an established and successful protocol (PDB: 4c3b and 4c3d), and due to this approach being the only current method

to generate structural models with the required resolution required for the research detailed here.

3.4 X-ray Crystallography

3.4.1 Crystal formation

X-ray crystallography requires the formation of crystals that contain multiple copies of the protein, regularly arranged within a crystal lattice. This arrangement allows amplification and subsequent detection of the diffracted X-rays that would not be detected using a single protein molecule. Crystallisation of a protein occurs when the protein becomes supersaturated, with a general rule of thumb being that the protein concentration must be greater than three times the solubility to enable crystallisation (Chernov, 1997). The large supersaturation is required to overcome the energy required for the process of nucleation (a small microscopic cluster of proteins that eventually forms the visible crystal) (Asherie, 2004) (Figure 18A). The phase diagram (Figure 18B) displays the three stages of supersaturation: metastable, labile and precipitation. To overcome the energy barrier and nucleate the protein must enter the labile zone, once this has occurred it can then enter the metastable zone that can sustain crystal growth. However, if supersaturation proceeds too far precipitation and aggregation formation will occur at a faster rate than crystal formation in what is known as the precipitation zone and is unfavourable to crystal formation.
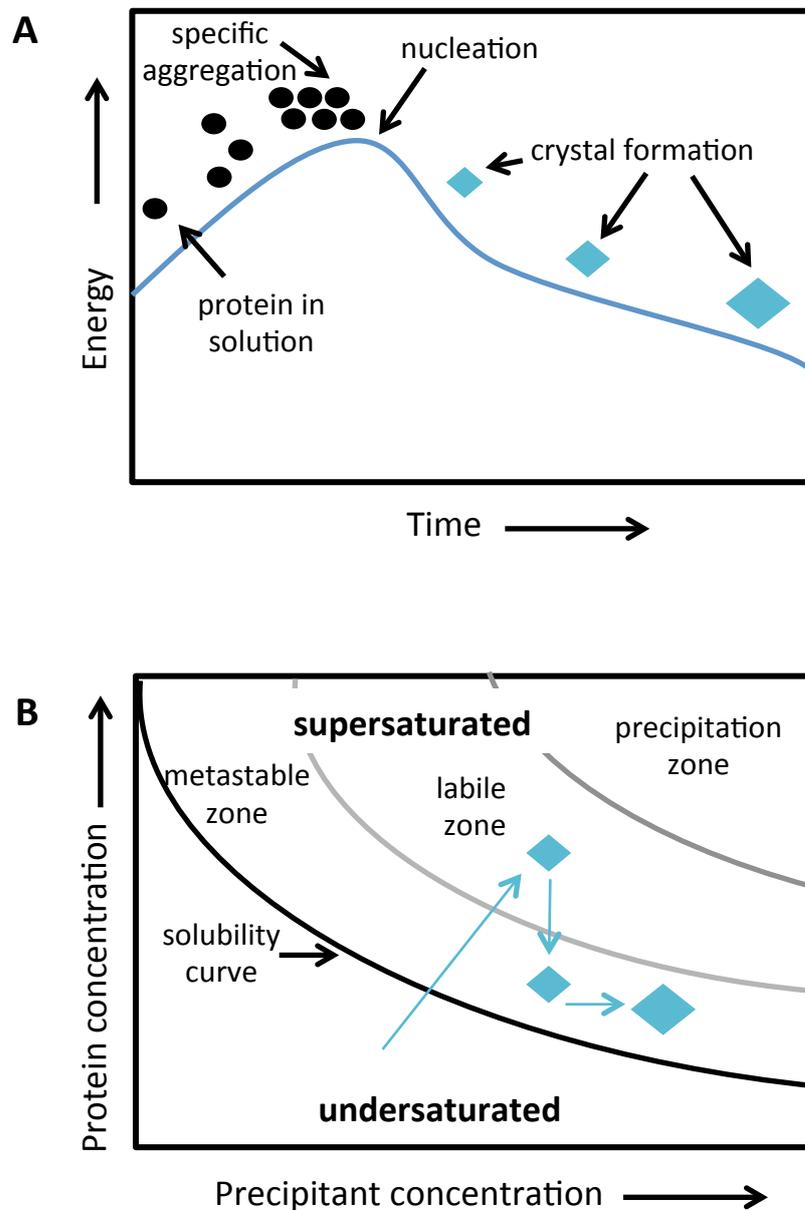
Figure 18 Energy and solubility phase diagrams for crystal formation

A- in order to crystallise the protein must overcome an energy barrier (nucleation), with a higher energy barrier resulting in a slower rate of nucleation; B- the protein must enter a supersaturated state to crystallise in which the concentration of the protein and precipitant increases. The protein must enter the labile zone in order to allow nucleation before entering the metastable zone that supports sustained crystal growth. If the protein enters the precipitation zone, aggregation is more likely and is unfavourable for crystal formation.

Crystal formation and growth is ascertained via three main methods: batch, vapour and dialysis (Figure 19). During vapour diffusion the protein: precipitant drop is suspended (via a hanging or sitting drop) and equilibrated with a reservoir of the precipitant in the well. For batch crystallisation the precipitant and protein are mixed directly under oil whilst for dialysis crystallisation the protein and precipitant are equilibrated through a dialysis membrane. Commercial screens that contain a diverse range of precipitant conditions (mother liquors) are available and aid crystal formation. Sparse matrix screens are biased towards conditions that have previously produced crystals whilst factorial screens assess each element and matrix parameter systematically.

Vapour diffusion hanging drop

Vapour diffusion sitting drop

Batch diffusion

Dialysis diffusion

Protein

Precipitant

Paraffin oil

Dialysis cassette
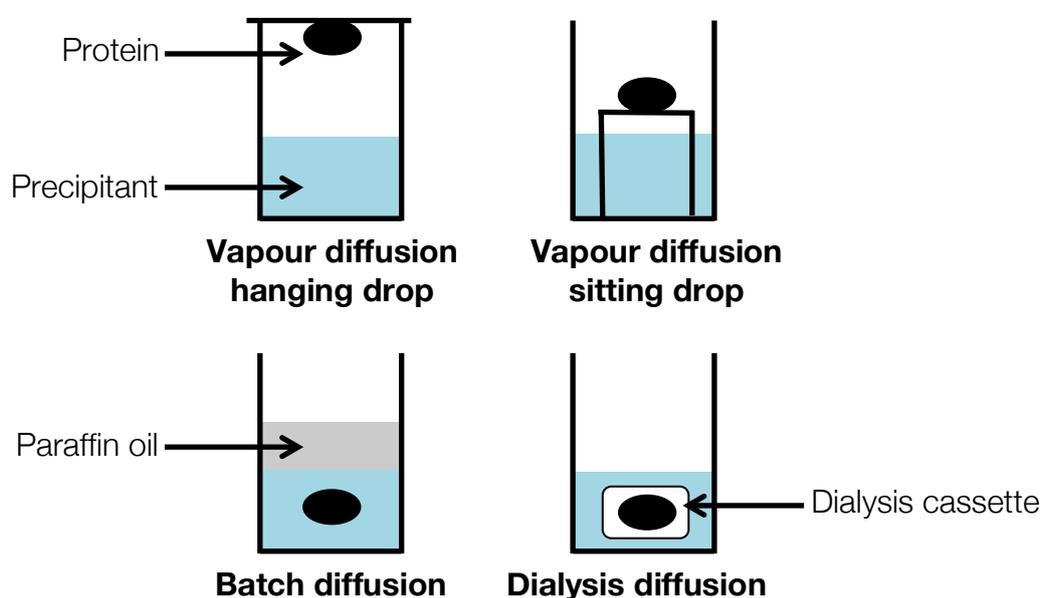
Figure 19 Protein crystal formation techniques

Schematic representation of techniques that allow crystal formation. During vapour diffusion the protein is suspended above the precipitant via either a hanging or sitting drop. For batch diffusion the protein and precipitant are mixed under a layer of oil and for dialysis a semi-permeable membrane separates the precipitant and protein.

3.4.2 Solving the crystal structure

X-rays are used to study protein crystals as the wavelengths of X-rays are the same order of magnitude (1Å= 0.1 nm) as the interatomic distances of proteins and can be used to determine the structure. Most X-rays pass through the crystal without interaction with electrons and are not recorded during the data collection process. However, some photons from the X-ray source interact with electrons of the atoms in the crystal structure and are scattered. These waves have different

scattering angles and phase and when they combine have either constructive (waves with high amplitudes) or destructive interference (waves with lower or no amplitudes). These waves hit a detector and are recorded as a pattern of discrete intestines, known as reflections and form the diffraction pattern.

In order for a diffracted X-ray to be recorded it must satisfy two conditions; Bragg's law and the Ewald sphere. Bragg's law is used to define constructive interference to produce a reflection. The law relates the angle (θ) at which there is a maximum in diffracted intensity to the wavelength of X-rays and the inter layer distances between the planes of atoms in the lattice. When scattered waves interfere constructively they remain in phase since the difference between path lengths (d) of two waves is equal to an integer multiple (n) of the wavelength. The path difference between two waves to remain in phase is 2d sinθ. Bragg's law is written as nλ= 2d sinθ and is shown in schematic form in Figure 20.
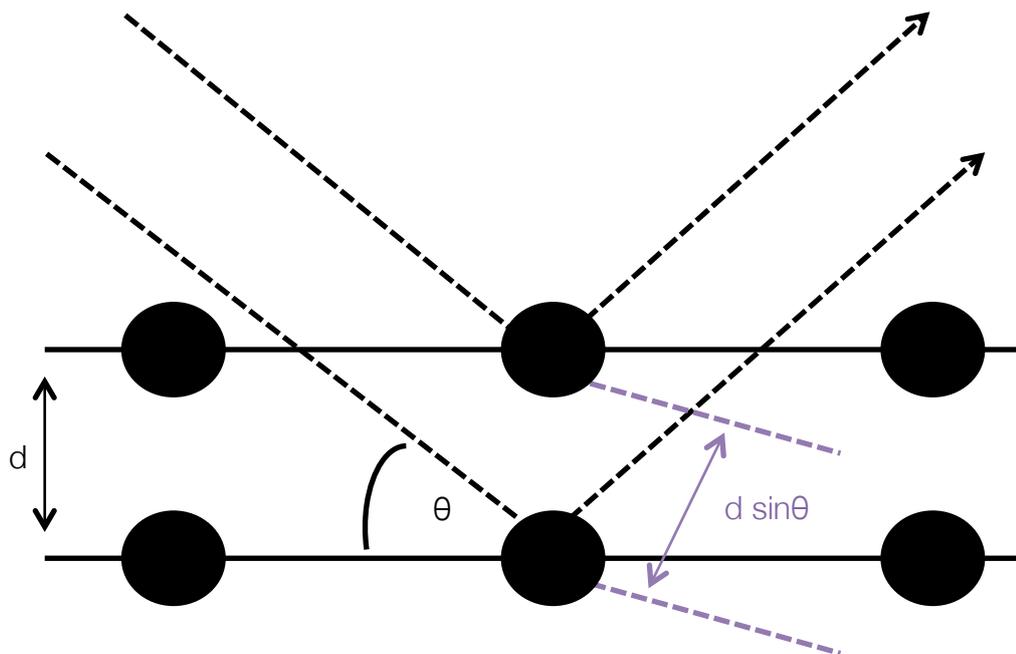
Figure 20 Bragg's Law

> If two Bragg's planes (parallel planes) are a distance (d) apart, with an angle of incidence of (θ), the two X-rays incoming waves will remain in phase if the second wave travels an addition distance (d sinθ) that is an integer (n) of the wavelength.

Bragg's law provides an explanation as to why a limited number of reflections are recorded for a given crystal. Reflections occur when Bragg's equation is satisfied and this occurs when the reciprocal lattice point of the transmitted X-ray beam lies on the Ewald sphere, a geometric construction that visualises the properties of Bragg's law. The Ewald sphere demonstrates that in order to collect more reflections, the crystal needs to be rotated to increase the likelihood of a reciprocal lattice point that lies on the Ewald sphere.

To satisfy Bragg's law waves must be in phase to result in constructive interference however destructive interference also occurs in which the scattering waves are out of phase and results in a decrease in amplitude of waves and the recorded intensity of the reflections. The intensity of a reflection provides information about the arrangement of atoms within the unit cell whilst the symmetry of the reflections provides information on the space group. The crystal system, crystal class and Bravais lattice define the space group. There are 7 crystal systems that refer to the shape of the undecorated unit cell and 14 unique Bravais lattices that are constructed from the simplest translational symmetry applied to the crystal systems. Point groups, of which there are 32 can then be associated to the Bravais lattice and represent the possible symmetries (reflection plane, rotation axis and rotary inversion axis) of the crystal.

As Bragg's planes intersect each unit cell of the crystal in the same way, each diffracting electron throughout the crystal contributes towards the resultant wave that results in a singular reflection seen. Miller indices (hkl) are assigned to each Bragg plane that intersects each unit cell and corresponding indices are given to the reflections. Further, a structure factor (Fhkl), the summation of all contributing waves for a single reflection is given to each reflection. The structure factors are based on knowing both the amplitudes and phases for each reflection. The structure factor amplitudes are directly proportional to the square root of the reflection intensity and are measured during data collection. However, unlike the amplitudes the phases cannot be directly recorded and here in lays the 'phase problem' seen in X-ray crystallography in which the phases of the resultant waves that generate the reflections are unknown. Two indirect methods can be used to determine phases; an isomorphous replacement method in which heavy atoms are soaked into the protein crystal to allow their position in the protein molecule and their phase angle to be determined and secondly, the method of molecular replacement. Molecular replacement is an appropriate method to determine phases when a related structure is available. Fitting the known model to the same orientation and position as the unknown protein by correlating intra and intermolecular vectors in the Patterson maps of both molecules allows the phases

to be extracted from the known model and used with the experimental structure factor amplitudes to apply a fourier transform (a way of describing a wave as a sum of all it components waves) and generate an electron density map for the unknown protein. If a high degree of similarity exists between the used model and the unknown structure, the subsequent electron density map will display features missing in the original model that can be built into the unknown model. The utilisation of molecular replacement formed the basis of the structural work performed in this chapter to solve the crystal structure of a P peptide bound to the previously determined HRSV M2-1 structure.

3.5   Functional studies

Previous work has suggested that M2-1 function depends on interaction with its various binding partners. In addition to structural studies, to further examine the interaction between M2-1 and its binding partners (P and RNA), fluorescence anisotropy (FA) was chosen to measure the affinity of M2-1 for its binding partners.

3.5.1 Fluorescence anisotropy (FA)

FA is a simple and robust technique that can determine the binding affinity of a fluorescently (Fl) labelled molecule to a protein of interest by measuring the rotational diffusion of the Fl-labelled molecule. FA is the method of choice for many biological studies due to its ability to be automated, requirement of low quantities of materials and ability to determine binding affinities for weak binding interactions in the µM range.

Here, the fluorescently (fluorescein) labelled molecule (RNA or P) when excited by plane polarised light emits light with a degree of polarisation inversely proportional to the tumbling rate of the molecule (Pollard, 2010). The small, unbound Fl- RNA or Fl- P molecule has a high tumbling rate in solution due to rapid Brownian motion and results in depolarisation of the emitted light. When the fluorescently labelled molecule is bound to M2-1, a large protein, the complex has a slower tumbling rate and the emitted light will be polarised to an extent. The extent of polarisation emission is described in terms of anisotropy with a larger protein having a lower rotational diffusion and a higher anisotropy reading (Figure 21).
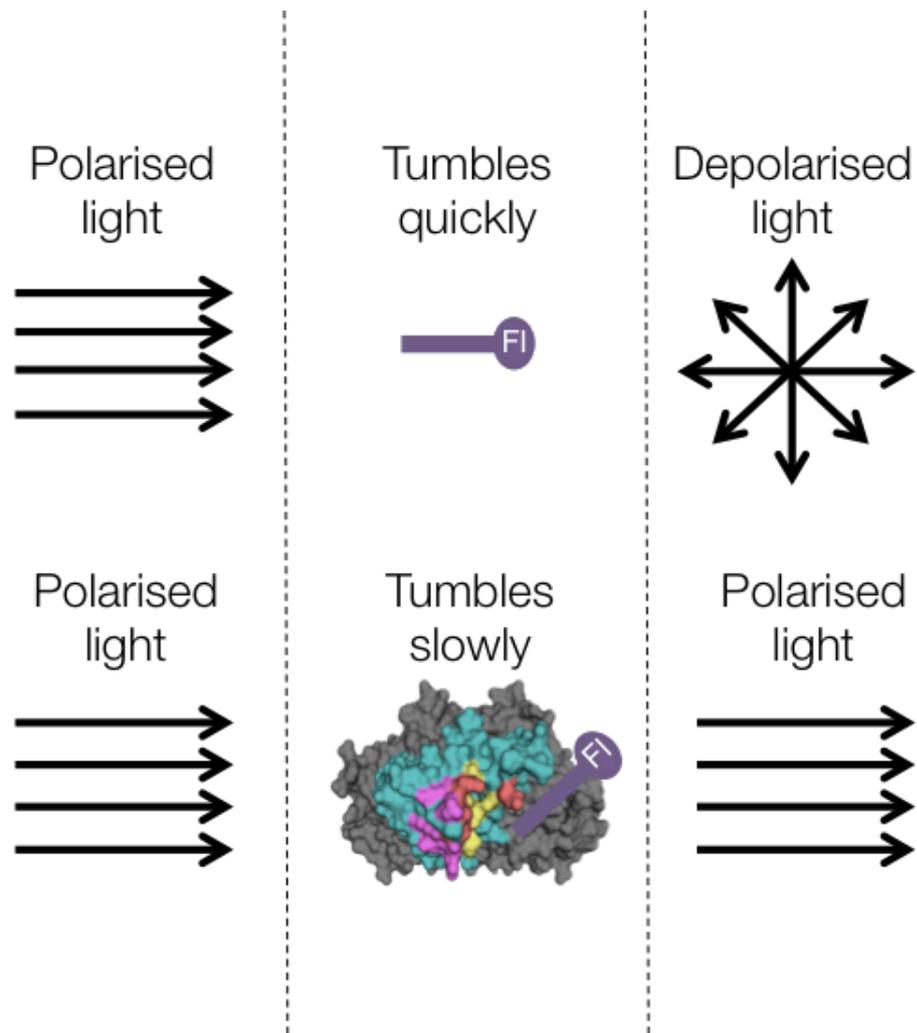
Figure 21 Fluorescence anisotropy schematic

When a fluorescently labelled molecule is excited by polarised light, the resultant light is depolarised due to the rapid motion in solution. On binding of the fluorescently labelled molecule to a large molecule the complex tumbles slower and the resultant light remains polarised to an extent.

3.5.2 Chapter aims

A greater understanding of the M2-1 HRSV protein and its interactions with its binding partners is required to allow further elucidation of M2-1's role in anti-termination and in the viral life cycle. The objective of the work described in this chapter is two-fold; (1) to optimise the heterologous expression and purification of HRSV M2-1 protein in bacteria to consistently produce high yields of homogenous protein and (2) to investigate the interactions of M2-1 with its binding partners; RNA and the phosphoprotein (P) using biophysical and structural methods.

3.6 Results

3.6.1 Optimisation of HRSV M2-1 purification

The HRSV M2-1 protein was previously expressed as a GST-fusion protein that yielded near-native M2-1, with five additional non-native amino acids (Gly-Pro-Leu-Gly-Ser) at its N-terminus (Tanner et al., 2014). However, yields of protein were inconsistent (40 g bacterial cell pellet resulted in a yield of 0-10 mg of M2-1 protein), which presented significant hurdles to perform material intensive studies such as X-ray crystallography, as well as biophysical techniques such as FA. Furthermore, the previous protocol required 5 days for purification of M2-1 resulting in significant losses through protein degradation, in addition to the requirement for high levels of starting materials and the need to concentrate the expressed protein in an impure form prior to SEC. However, perhaps the most serious issue with the existing protocol was that the low yield required the generation of multiple batches prepared on different days, which had potentially serious consequences for batch-to-batch experimental reproducibility. Previously, to circumvent these problems each batch of purified protein was assayed for its RNA binding ability using FA and the Kd determined to allow comparisons to be drawn between the functionality of each batch. However, this remedy was unsatisfactory, as batch variation was inevitable. To allow a more rigorous experimental interrogation of M2-1 structure and function, an improved protocol for expression, purification and crystallisation was required.

There are multiple benefits of having large yields of consistent, pure, functional protein. It allows a greater degree of comparison between experiments but mainly allows a vast number of techniques to be performed. In the case of the work presented here, a large quantity of protein was required for functional, biophysical and structural studies. X-ray crystallography is widely regarded as a material intensive technique. Large (milligram) quantities are required in the first instance for crystal screens with more protein being required for further crystal optimisation before a diffraction data set can be collected. In chapter 5, a new method of fragment screening by X-ray crystallography (FBDD) was assessed and required thousands of protein crystals that could be harvested with optimal diffraction for further analysis. This novel method to determine potential drug-like lead building blocks for anti-viral therapy could not be achieved without the ability to produce vast quantities of pure homogenous protein from a single preparation.

3.6.2 New purification method

Several possible stages within the existing M2-1 expression and purification pipeline were identified as being potentially problematic and responsible for poor yields and protein quality. These were the composition of the lysis buffer, sonication method, chromatography techniques and protein concentration methods, and these are summarised in Table 7. The M2-1-GST construct (Tanner et al., 2014) was not altered here, only the downstream purification as previous research (personal communication Dr Sian Tanner, University of Leeds) had assessed M2-1 expression in an insect cell expression system and bacterial expression systems, utilising a His-SUMO tagged and GST tagged expression vectors and determined the GST tagged protein optimal for M2-1 protein expression.

First, the issue of protein degradation was addressed, and one possible area this occurred was during the overnight binding to glutathione resin. To reduce degradation clarified lysate was bound to glutathione resin using gravity flow and washed with buffer immediately after (Figure 22A). Secondly, a freeze/ thaw method was employed instead of sonication to provide a more gentle method of bacterial lysis with greater temperature control and allowed a larger volume of buffer per gram of bacterial cell pellet. Both steps above in addition to cation exchange chromatography (Figure 22B) produced homogenous M2-1 protein. This improved method yielded 50 mg M2-1 protein from 10 g bacterial pellet, a ~ 10 fold improvement on the previous protocol. A final purification step was performed prior to experiments using size exclusion chromatography (Figure 23) to ensure M2-1 was free from aggregates and exchanged into the correct buffer. The M2-1 protein elutes at a larger MW than expected for a 90 kDa tetrameric protein. This discrepancy has been observed previously (Tanner et al., 2014) and is most likely due to the non-spherical shape of M2-1, which influences the SEC elution profile and thus estimation of molecular mass.

The full optimised M2-1 purification protocol is described in chapter 2; section 2.2.2. Briefly, the bacterial pellet was re-suspended in lysis buffer. The re-suspended bacteria were then lysed using four cycles of freeze-thawing. DNase and RNase was then added to degrade genomic DNA and RNA to avoid the formation of M2-1/RNA complexes. Lysates were centrifuged at 35,000 x g for 30 minutes at 4°C. The clarified lysate was applied to Glutathione 4B Resin (GE Healthcare) by gravity flow before washing with lysis buffer and followed by further washes with cleavage buffer. PreScission protease (GE Healthcare) was added to

the resin, incubated with rotation overnight at 4°C and resulted in a high efficiency of M2-1 cleavage from its N-terminal GST tag. The protein was eluted from the resin by gravity flow with the addition of 2 CV of wash buffer and resulted in the purification of the M2-1 protein with 5 non-native residues at its N-terminal (Gly-Pro-Leu-Gly-Ser.

Purification alteration	Purification improvement
Each new purification- transform plasmid and perform small-scale protein expression of bacterial pellet	Consistent and high levels of expression
Lysis buffer addition: glycerol	Increased viscosity to reduce aggregation
Lysis buffer addition: DNase/ RNase	Degrade genomic RNA and DNA that may associate with M2-1
Larger buffer volume (200 ml/ 10 g bacterial pellet	For freeze/ thaw method- less aggregation as protein not as concentrated in solution
Freeze/ thaw lysate	Greater regulation of lysis temperature resulted in less protein aggregation
Glutathione resin- bind and elute lysate immediately	Less lysate bacterial growth over time that prevents elution
Cation exchange	Protein did not have to be concentrated to remove further impurities (like seen for SEC)
Concentrator membrane	PES membrane did not result in M2-1 aggregation
SEC (when protein is pure)	Buffer exchange for consistency in future assays

Table 7 HRSV M2-1 purification alterations

Alterations from the original purification protocol described by Tanner et al., 2014 resulted in an increase in homogenous pure protein of ~ 10 fold.
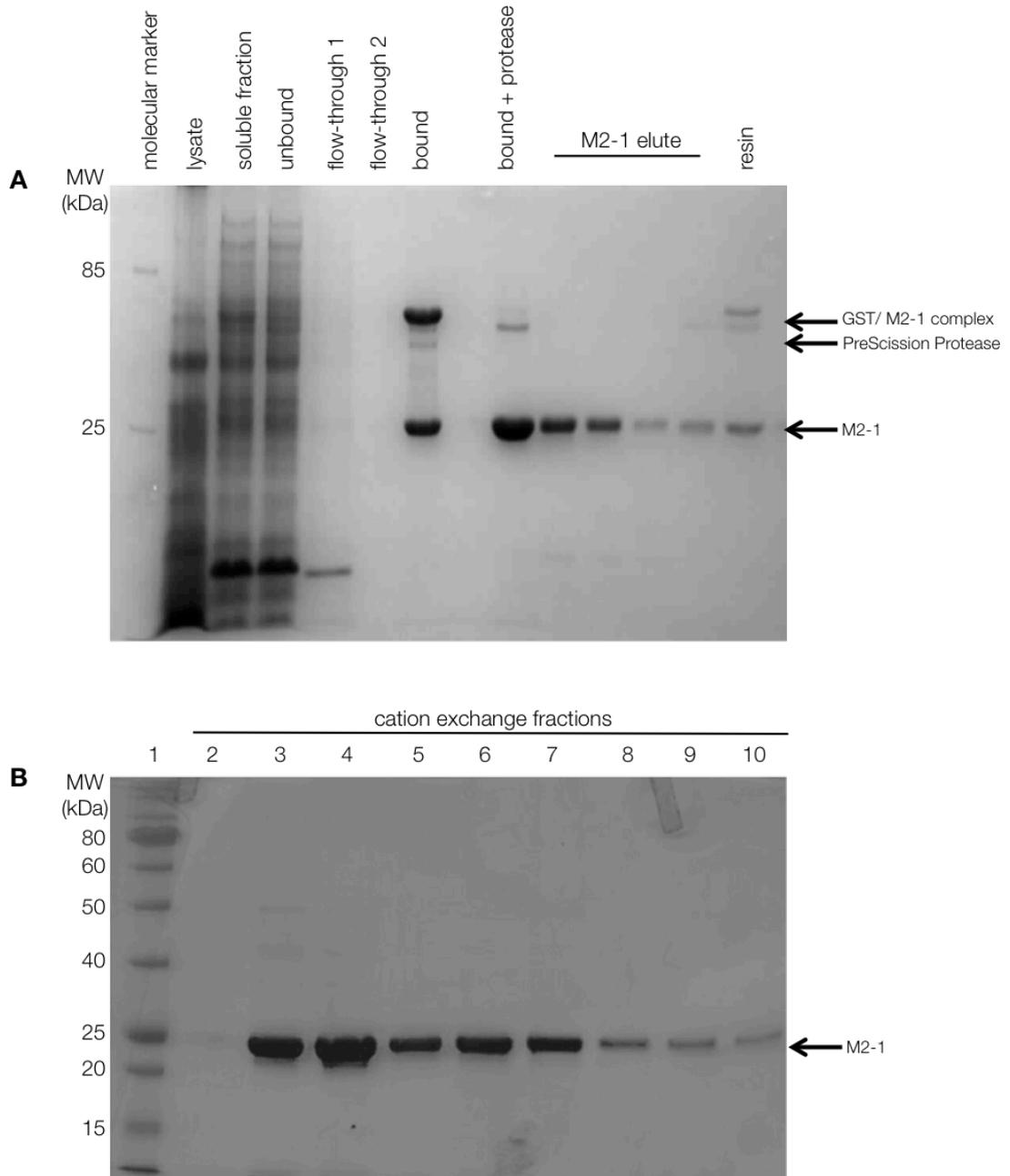
Figure 22 Purification of the M2-1 protein

SDS PAGE (15%) stained with Coomassie blue. A- purification of M2-1 by affinity chromatography. Lane 1- molecular weight marker; lane 2- clarified lysate; lane 3- clarified lysate bound to glutathione resin; lane 4- flow-through of lysate unbound to glutathione resin; lane 5- resin wash flow-through; lane 6- resin wash 2 flow-through; lane 7- glutathione resin with protein bound; lane 8- GST-M2-1 bound to glutathione resin following the addition of Precission protease; lanes 9-12 a protein, M2-1 elutes at an apparent MW of 25 kDa; lane 13- glutathione resin after M2-1 protein elution; B- M2-1 purification using cation exchange chromatography. Lane 1; molecular marker; lanes 2-9- 5 ml elutions using 60% high salt; lane 10- 100% high salt wash of the resin.
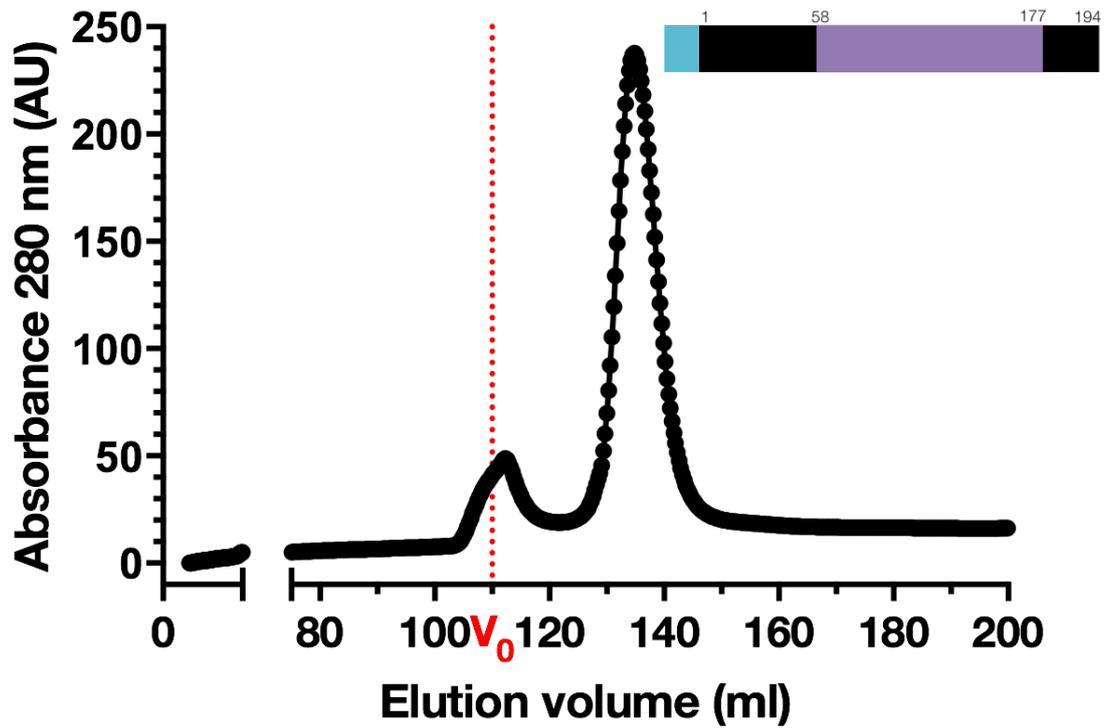
Figure 23 Final purification step for M2-1 using SEC

M2-1 protein was separated on a HiLoad 26/ 600 Superdex 75 PG column. A protein peak was seen at an elution volume between 125-140 ml corresponding to a MW ~ 80 kDa (EMBL- pepcore-chromatography calibration; appendix II).Vo=void volume.

3.7   Purification of truncated M2-1 proteins

Truncated M2-1 proteins (Table 8) were engineered to aid crystallisation of the M2-1: RNA co-complex via alterations to the unit cell packing (section 3.8.2) and were expressed as GST fusion proteins and purified using the improved M2-1 purification protocol stated above.

Protein	Schematic
M2-1	
ΔLGS	
ΔLGSMS	
M2-1 58-177	

Table 8 M2-1 protein schematic representation

> Blue indicates non native residues (Leu-Gly-Ser). Two non-native residues (Gly-Pro) remain at the N-terminal end of M2-1 in each construct due to the Precission protease cleavage. Numbers refer to the M2-1 residues.

3.7.1   N-terminal M2-1 truncations (ΔLGS and ΔLGSMS)

To aid crystallisation by allowing the formation of alternate crystal contacts we removed the non-native residues (ΔLGS) and additionally produced a construct with removal of Met1 and Ser2 of M2-1 (ΔLGSMS) (chapter 2, section 2.2.1.3). The purified protein (Figure 24A & B) was crystallised (section 3.8.2) and its crystal structure was solved using molecular replacement with the known M2-1 crystal structure (pdb: 4c3b).
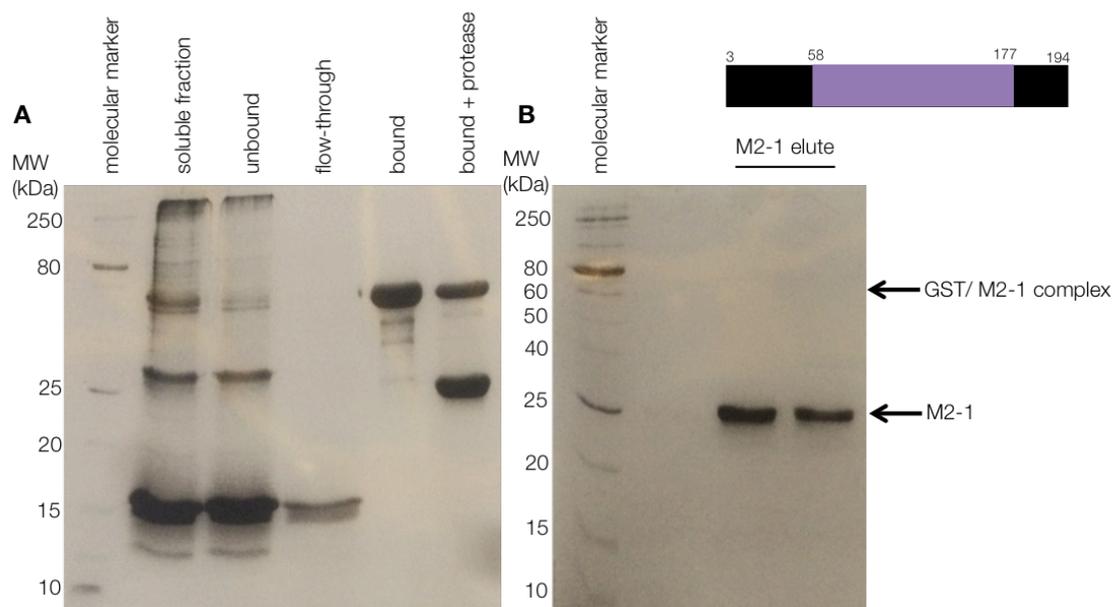
Figure 24 Purification of the ΔLGSMS M2-1 protein

SDS PAGE (15%) stained with Coomassie blue. A- purification by glutathione resin. Lane 1- molecular weight marker; lane 2- soluble fraction added to glutathione resin; lane 3- flow through of soluble lysate from glutathione resin; lane 4- resin wash 1; lane 5- glutathione resin with bound protein; lane 6- glutathione resin after cleavage of M2-1-GST site with PreScission 3C protease; B- cation exchange. Lane 7 molecular weight marker; lane 8 and 9- M2-1 elute. The protein runs at an apparent molecular weight of~ 25 kDa.

3.7.2 Monomeric M2-1 (M2-1$_{58-177}$)

Monomeric M2-1 (M2-1$_{58-177}$) was expressed and purified to aid co-crystallisation studies by preventing aggregation and production of higher order oligomers and was further assessed for its ability to bind RNA and P sequences. Monomeric M2-1 was produced by removal of the ZBD and oligomerisation domain of M2-1 (residues 1-57 deleted; see chapter 2; section 2.2.1.3). In addition the C-terminus of M2-1 was truncated to remove the disordered region not resolved in the crystal structure (residues 178-194 deleted). The resultant monomeric M2-1 protein is referred to as M2-1$_{58-177}$ (described by Blondot et al., 2012). SEC (Figure 25) of the purified protein on a 16/60 Superdex75 column revealed a protein peak at an elution volume of 170- 200 ml, corresponding to a protein of ~17 kDa, consistent with monomeric core M2-1 (residues 58-177). As discussed in the next section, the structure of this protein was not solved by X-ray crystallography. Analysis of the eluted protein by mass spectrometry determined the molecular weight to be

13864.19 ± 0.03 Da (Figure 26), which is very similar to the predicted molecular weight of 13864.8 Da.
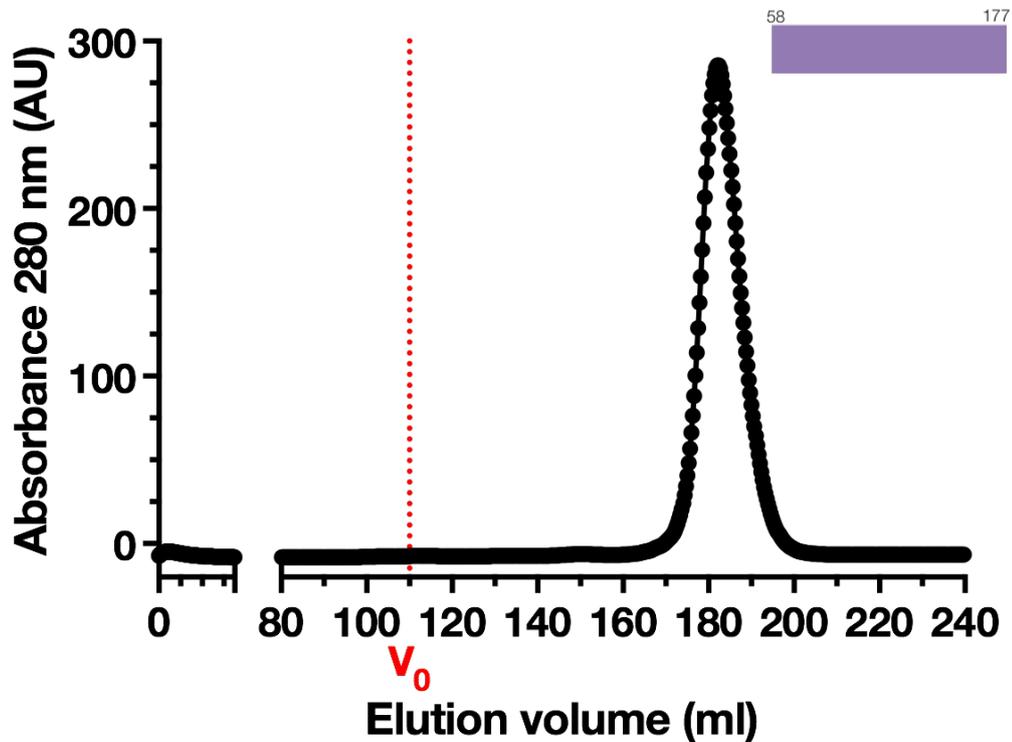
Figure 25 SEC trace of M2-1$_{58-177}$

M2-158-177 was separated on a HiLoad 26/60 Superdex 75 column. The red line indicates the column void volume at 110 ml. One peak can be seen with an absorbance at 280 nm, the protein elution volume corresponds to a protein with a molecular mass of ~17 kDa. Monomeric M2-158-177 is ~13 kDa. Discrepancies in elution volume are also seen for full length M2-1 and are assumed due to the non-spherical nature of the protein. Vo=void volume.
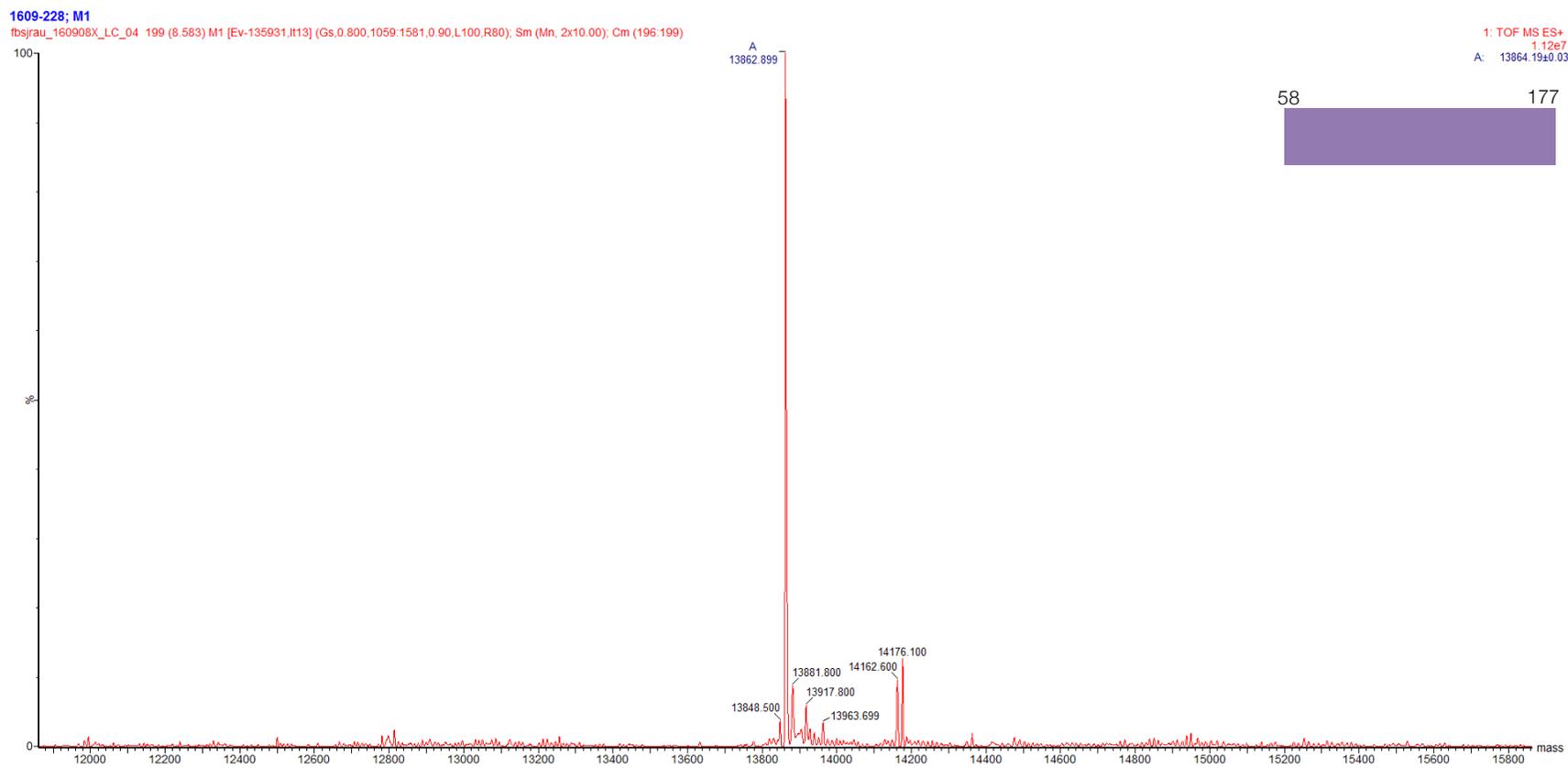
Figure 26 Mass Spectrometry analysis of M2-1$_{58-177}$.

M/Z spectrum using a electrospray ionisation mass spectrometer, performed by the Mass Spectrometry facility, University of Leeds. Dominant ions at m/z of 13864.19 ± 0.03 Da, consistent with a theoretical molecular mass of 13864.8 Da

3.7.3 Crystallisation of M2-1

The M2-1 protein, in 50 mM tris-Cl pH 7.4, 150 mM NaCl and 1 mM DTT, was concentrated to 12 mg/ ml and subjected to crystallisation trials at 18°C using commercially available screens- Morpheus, PACT, MIDAS (Molecular Dimensions) Index, Salt RX, Crystal 1 & 2 (Hampton Research) Wizard 1 & 2 and Wizard 3 & 4 (Emerald BioSystems). Using the optimised purification protocol crystals formed in more conditions than previously observed. A full list of conditions that that supported crystal growth can be found in appendix III. Despite observing crystals in varied conditions, crystals formed with one morphology (plate) for M2-1 with variation in size (Figure 27).
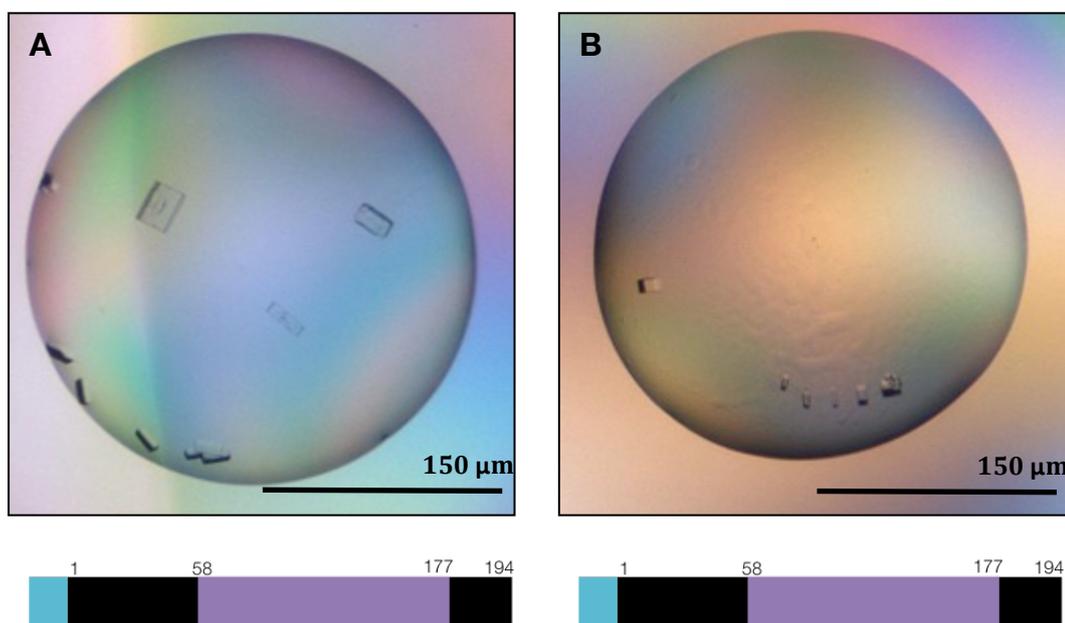
Figure 27 M2-1 crystal morphology

M2-1 crystals had a plate morphology that varied in size from 10-100 μm. A- 2.1 M DL-malic acid pH 7.0. B- 0.1 M potassium thiocyanate, 30% w/v PEG MME 2,000. Crystals were visualised using the RockMaker imager.

Crystals were assessed for protein diffraction at the DLS on the I02 beamline. The improved purification protocol resulted in more crystallisation trials to be conducted and crystals to form reproducibly. M2-1 protein crystals diffracted to a resolution of 2.2 Å (best resolution previously 2.4 Å) however, when molecular replaced these structures gave no additional information at the atomic level than previously seen in the published structure (Tanner et al., 2014). The M2-1 protein crystal structure was solved by molecular replacement (Figure 10) in its native

tetrameric conformation. One monomer is highlighted in blue with residues hypothesised from mutational studies to be involved in RNA (pink), P (yellow) and both RNA and P binding (coral) highlighted. In section 3.9.4 of this chapter the co-crystal structure of M2-1 with a P peptide bound was determined. The hypothesised RNA and P binding sites on the M2-1 protein are shown throughout to assess similarities of the hypothesised and structurally determined binding site.
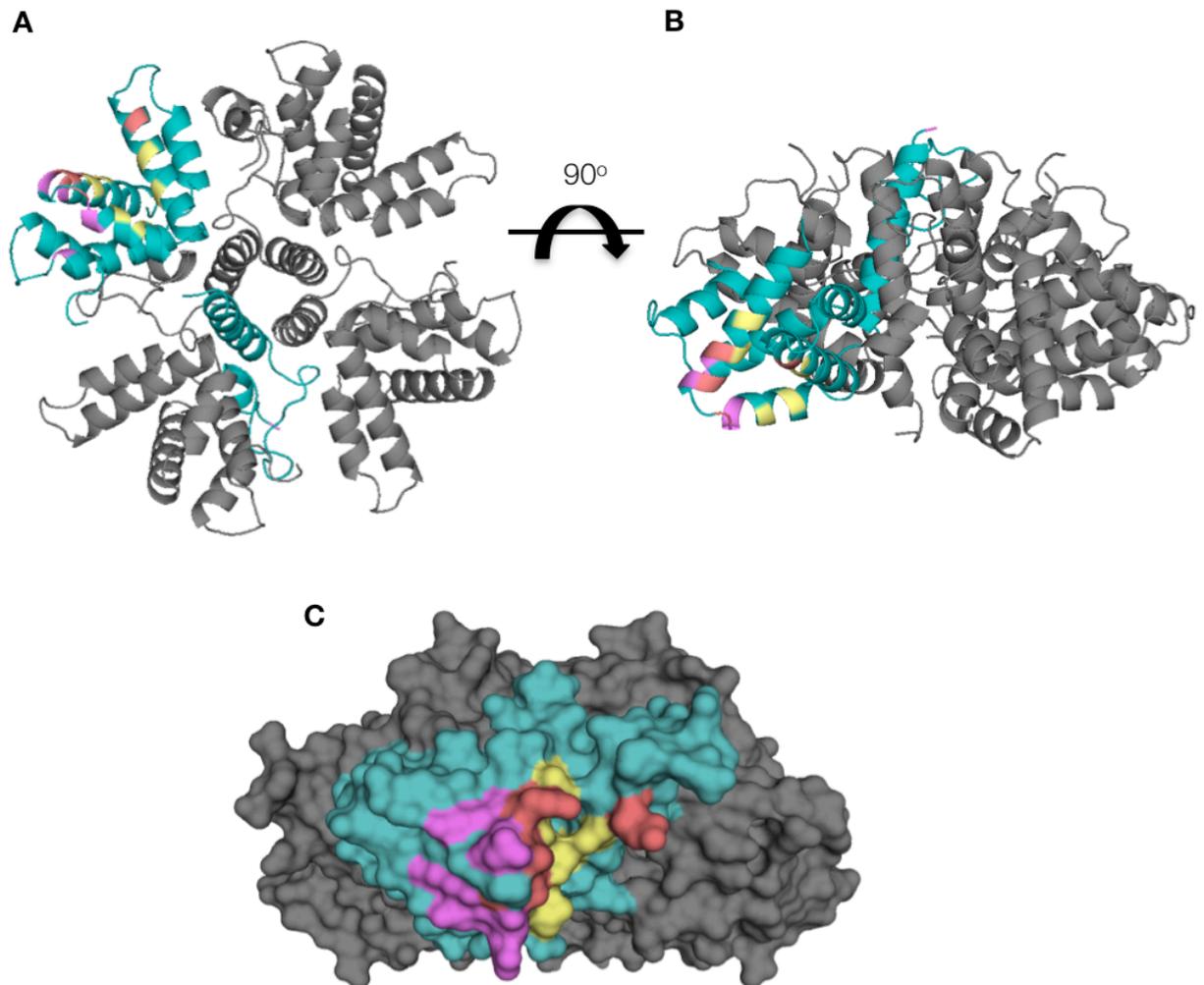
Figure 28 Structure of the HRSV M2-1 protein

> A- cartoon representation of the native tetrameric M2-1 protein (grey) with a monomer highlighted in blue. RNA (pink), P (yellow) and both RNA and P (coral) hypothesised binding residues on M2-1 are highlighted; B- same as A with rotation; C- surface representation of B. The 2.2 Å crystal structure was solved by molecular replacement based on the previously determined crystal structure (pdb 4c3b, 4c3d ; Tanner et al., 2014).

3.8 The M2-1: RNA interaction

3.8.1 Binding interaction

Anti-termination activity of M2-1 is dependent upon RNA binding with a consensus in the literature that M2-1 binds to A-rich semi-conserved gene-end (GE) signals in nascent mRNAs to function as an anti-terminator. To aid understanding of the RNA: M2-1 binding interaction and RNA sequence specificity, direct binding affinities were determined by FA for short (13 nucleotide) hRSV GE sequences of the genome sense (-) and RNA oligonucleotides referring to its positive sense complement/ mRNA (+) (affinities here formed data in Tanner et al., 2014). Binding affinities stated here were determined from the binding curves, however binding could not be saturated at the concentrations of M2-1 assayed and thus the affinities may be referred to as 'apparent' Kd values.

Higher affinity binding was always visualised for the anti-genome complement (+) oligonucleotide (Figure 29 & Table 9), with the SH GE (+) RNA sequence resulting in highest affinity of a HRSV RNA sequence (2.3 ± 0.39 µM) whilst polyA had the highest overall affinity of any sequences assayed for M2-1 (1.1 ± 0.12 µM). Further, the monomeric M2-1 construct (M2-1$_{58-177}$) that excluded the ZBD, oligomerisation domain and flexible C-termini, was assessed for binding to the RNA sequences and resulted in at least 4-fold weaker affinity than full length, tetrameric M2-1 (Table 9).
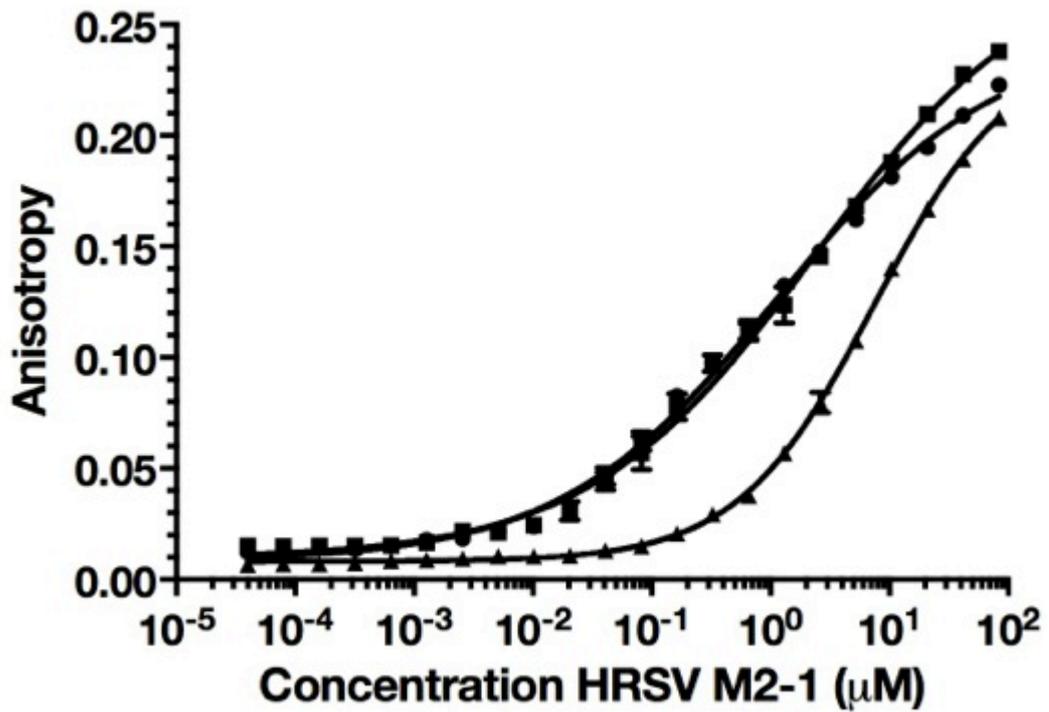
Figure 29 Direct binding of M2-1 to RNA sequences using FA

Log concentration curve of RNA 13mers, fitted to a 1:1 binding model; chapter 2; section 2.2.5.1. Data points (●) represent polyA 13mer; (n) SH GE +; (▲) SH GE -. Binding affinities (Kd) were determined from the fitted curve. Data points represent n-3 ± SEM with each n number performed in triplicate.

RNA	Sequence 5'- 3'	Apparent K_d ± SEM (µM) Tetrameric M2-1	Apparent K_d ± SEM (µM) M2-1$_{58-177}$
F GE (+)	AGUUAUAUAAAAC	5.6 ± 0.44	41 ± 2.9
F GE (-)	GUUUUAUAUAACU	17 ± 1.5	65 ± 8.8
SH GE (+)	AGUUAAUUAAAAA	2.3 ± 0.39	25 ± 2.1
SH GE (-)	UUUUUAAUUAACA	7.5 ±0.41	57 ± 16
NS1 GE (-)	UUUUAUAUUAACU	8.1 ± 0.65	65 ± 16
NS1 GE (+)	AGUUAAUAUAAAA	3.7 ± 0.35	39 ± 3.5
polyA	AAAAAAAAAAAAA	1.1 ± 0.12	11 ± 0.54
polyU	UUUUUUUUUUUUU	N.D	N.D

Table 9 Binding affinities (Kd) of RNA nucleotide sequences to the full length and truncated M2-1$_{58-177}$ protein measured by fluorescence anisotropy.

All titration curves were carried out in triplicate and the binding affinities shown with SEM. (+) indicates positive complement/ mRNA sense and (-) indicates genome sense oligonucleotide sequences. Binding affinities represent n-3 ± SEM with each n number performed in triplicate and the values displayed to 2 significant figures. N.D- non determinable.

3.8.2 Crystallisation of the M2-1: RNA complex

Various methods were trialled to co-incubate M2-1 with RNA (summarised in Table 10) for crystal formation and subsequent visualisation of the interaction via X-ray crystallography.

Co-crystallisation	Schematic	RNA	Incubation Time (hr)	Temp (°c)	Notes
Co-incubation					
M2-1		Various RNA 6-13 mers	Various 1hr-overnight	4/ 18/ 20	10 mg/ ml + SEC
M2-1		Various RNA 6-13 mers	Various 1hr-overnight	4/ 18/ 20	10 mg/ ml 20 mg/ ml 30 mg/ ml - SEC
ΔLGS		Various RNA 6-13 mers	1 hr	18	20 mg/ ml -SEC
ΔLGSMS		Various RNA 6-13 mers	1hr	18	20 mg/ ml - SEC
M2-1$_{58-177}$		Various RNA 6-13 mers	1hr	18	30 mg/ ml -SEC
Competition with P					
M2-1+ P 90-110		RNA 13 mer	1hr	18	10 mg/ ml 30 mg/ ml + SEC (off resin)
M2-1+ P 90-160		RNA 13 mer	1 hr	18	30 mg/ ml - SEC (on GST resin)

Table 10 Various methods trialled for RNA: M2-1 co-crystallisation

Initially M2-1 was incubated with various A-rich RNAs (pA 6-13 mers and SH GE +) and the co-complex either purified by SEC or entered directly to crystal trials. Further methods to form the M2-1: RNA complex exploited the M2-1: P interaction and the ability of RNA to outcompete this interaction. During the first trial of this method, unbound M2-1 was incubated with a P peptide sequence (residues 90-110) before the addition of RNA (pA13mer) and purification of the co-complex by SEC (Figure 30), however this method yielded a co-complex of M2-1: RNA that eluted in the void volume, indicative of an aggregated protein species.
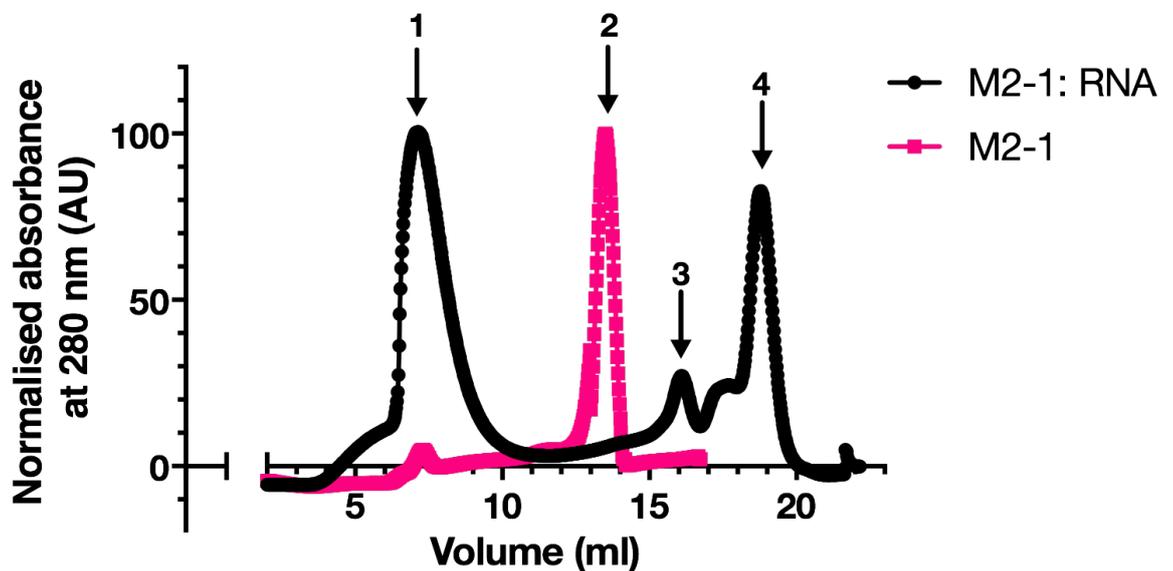
Figure 30 SEC trace of M2-1: RNA complex

M2-1 (pink) and M2-1 incubated with P 90-110 followed by pA 13mer RNA (black) were applied to a Superdex 200 10/300 gel filtration column. M2-1 (pink) eluted at a single peak (2) at ~13 ml. M2-1 (black) incubated with P 90-110, eluted at peak (4) before being outcompeted with SH GE (+) RNA. Excess RNA eluted at (3) whilst the RNA:M2-1 complex (1) eluted at ~8 ml in the void volume.

In a second trial to form an M2-1: RNA complex by exploiting the M2-1: P interaction, GST-P 90-160 was bound to glutathione resin and M2-1 was added to form the M2-1: P complex. The resin was extensively washed with buffer to remove unbound M2-1 and bound M2-1 subsequently eluted with RNA. Eluted protein was assumed to be the M2-1: RNA co-complex (Figure 31 lane 6-8) and SDS-PAGE analysis showed a protein of ~25 kDa, consistent with it being M2-1.
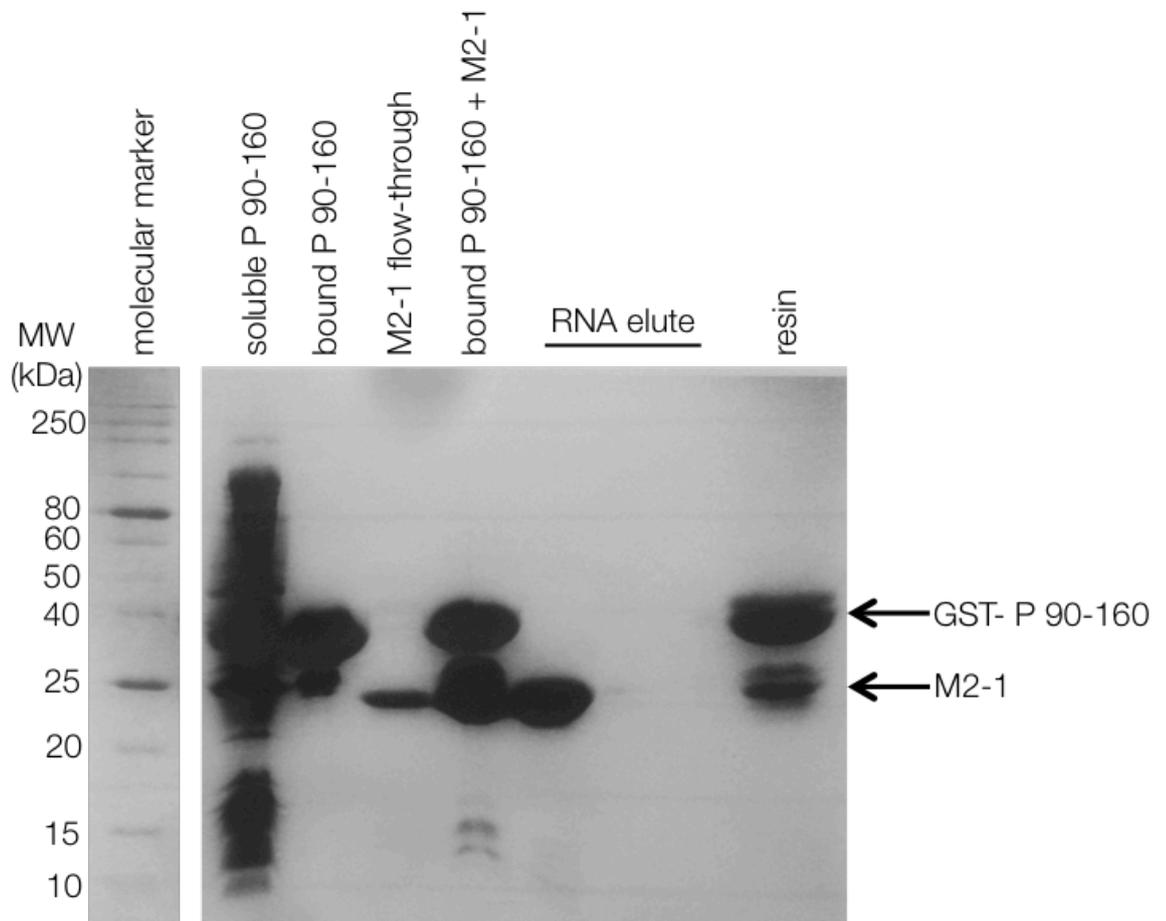
Figure 31 SDS PAGE analysis of P 90-160 and RNA competition for M2-1

SDS PAGE (15%) stained with Coomassie blue. P 90-160 fused to GST was expressed and purified on glutathione resin and competed for binding to M2-1 with an RNA13mer. Lane 1- molecular marker; lane 2- soluble fusion protein (GST-P 90-160); lane 3- GST-P 90-160 bound to glutathione resin; lane 4- flow through of excess M2-1 not bound to P 90-160 on the resin. Lane 5- P 90-160 and M2-1 bound to the glutathione resin; lanes 6-8; excess RNA13mer was added to the column and outcompeted P 90-160 and eluted M2-1 from the column. Eluent was collected in 3 fractions of 15ml however protein was only seen in the first fraction; lane 9- resin after addition of RNA and elution shows that some GST-P 90-160 and M2-1 protein still bound to the resin. Dr Kavestri Yegambaram, University of Leeds performed the expression and purification of P 90-160.

Crystal trials of the M2-1:RNA complex were performed at 4, 18 and 20°C with crystals taking an average of 12 days to form compared to 3 days for unbound M2-1 crystals. Crystals had the characteristic plate morphology (Figure 32) previously seen for the unbound M2-1 crystals. However, despite trialling various methods only unbound M2-1 crystals and not the M2-1: RNA co-complex was determined from diffraction of the crystals and subsequent analysis of the density that contained no continuous extra electron density that could account for a short RNA oligonucleotide. M2-1 crystals have only been observed in two space groups (P422 and P2$_1$) and thus it was hypothesised that the rigidity of the crystals, or existence of specific crystal contacts, may prevent the co-complex from accessing different space groups.
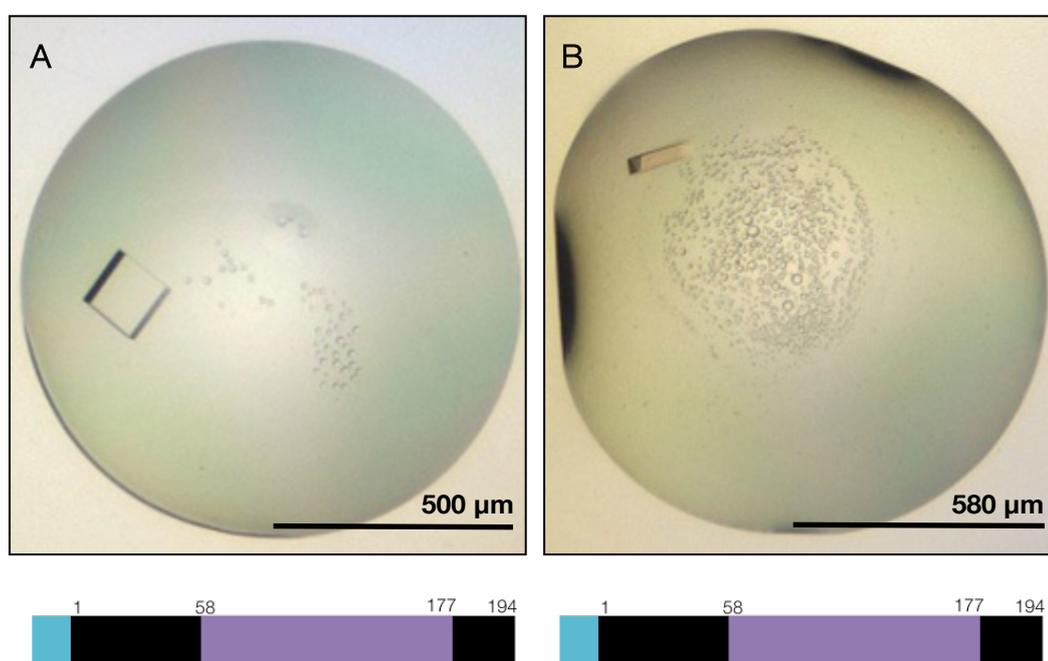
Figure 32 M2-1: RNA co-complex crystals

A- M2-1 incubated with RNA SH (+) GE 13mer in 0.15 M potassium bromide and 30% w/v PEG 2000 MME; B- M2-1 protein incubated with pA 10mer in 0.2 M sodium iodide, 20% w/v PEG 3350. Crystal formation took 34 days (A) and 13 days (B) with a plate morphology. Crystals were visualised using the RockMaker imager.

3.8.2.1 ΔLGS/MS M2-1: RNA crystallisation

In a further attempt to crystallise the M2-1: RNA complex, truncation mutants ΔLGS and ΔLGSMS (Table 8) were expressed and purified with the aim of altering crystal contacts and thus allowing solvent accessible channels for RNA to bind to the M2-1 protein in a crystal lattice. RNA was incubated in excess with ΔLGS and ΔLGSMS constructs (section 3.7.1) and crystal trials performed at a protein concentration of 20 mg/ ml and a temperature of 18°C. Crystals formed within 14 days and had a plate morphology, similar to unbound M2-1 crystals (Figure 33). Crystals were assessed for diffraction on the I02 beamline at the DLS with achieved resolutions summarised in Table 11. Diffracting crystals formed in the space group $P4_2\,2_1\,2$, with one crystal (condition 0.2 M ammonium sulphate, 0.1 M bis-tris pH 6.5, 25% w/v PEG 3350) forming in the $P1\,2_1\,1$ space group. ΔLGSMS M2-1 in complex with pA 6mer in condition 0.2 M $MgCl_2$ hexahydrate, 0.1M tris pH 8.5, 25% w/v PEG 3350 resulted in the best achievable diffraction of 2.3 Å (Figure 34). Inspection of the electron density maps after molecular replacement of the M2-1 protein showed no positive density that would account for RNA. Further, the N-terminal truncations did not alter the crystal packing seen for the previously determined M2-1 protein.
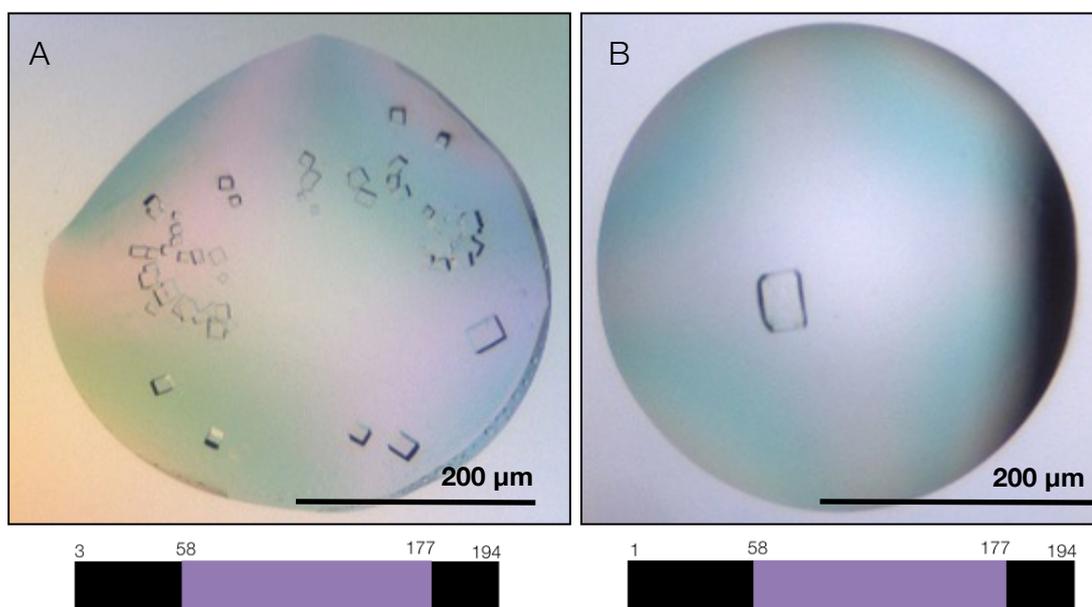
Figure 33 Images of ΔLGSMS: RNA and ΔLGS M2-1: RNA crystals

A- ΔLGSMS M2-1 incubated with pA 6mer RNA in 0.2 M MgCl2 hexahydrate, 0.1 M tris pH 8.5 and 25% w/v PEG 3350; B- ΔLGS M2-1 incubated with pA 6mer in 0.2 M potassium/ sodium tartrate, 20% w/v PEG 3350. Crystal formation took one day for both conditions with a plate morphology, and crystals were visualised using the RockMaker imager.
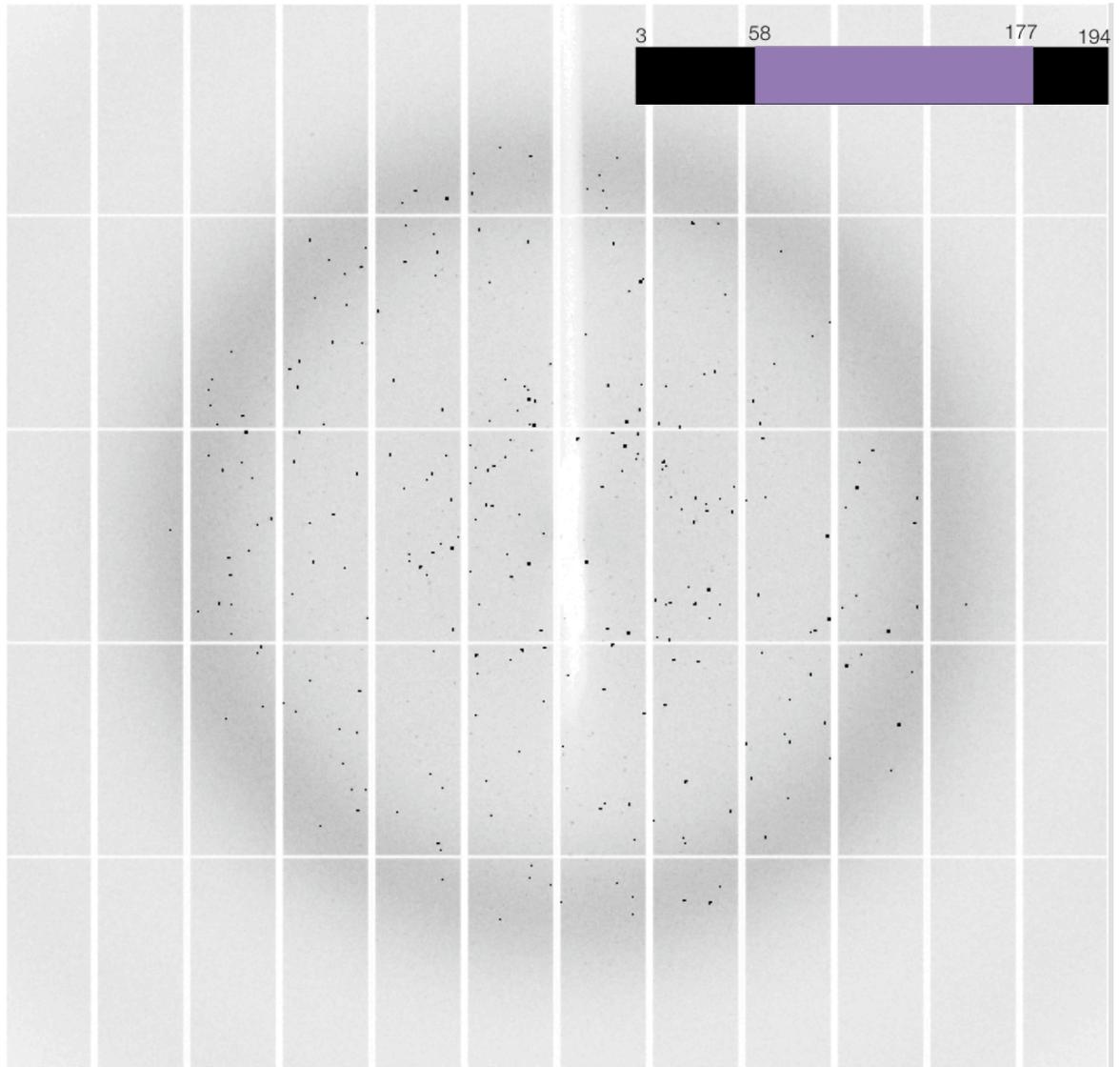
Figure 34 Diffraction image of a ΔLGSMS M2-1: pA 6mer RNA crystal

ΔLGSMS M2-1 protein incubated with pA 6mer. Condition 0.2 M MgCl2 hexahydrate, 0.1M tris pH 8.5, 25% w/v PEG 3350 . Images were collected on the I02 beamline, DLS under cryo-temperatures at 100K to a resolution of 2.3 Å.

Condition	M2-1 construct (in complex with pA 6mer)	Diffraction resolution (Å)
0.2 M MgCl$_2$ hexahydrate 0.1 M tris pH 8.5 25% w/v PEG 3350	ΔLGSMS 3 58 177 194	3.4
0.2 M MgCl$_2$ hexahydrate 0.1 M tris pH 8.5 25% w/v PEG 3350	ΔLGSMS 3 58 177 194	2.3
0.2 M potassium/ sodium tartrate 20% w/v PEG 3350	ΔLGS 1 58 177 194	3.3
0.2 M ammonium sulphate 0.1 M bis-tris pH 6.5 25% w/v PEG 3350	ΔLGS 1 58 177 194	2.7
0.2 M ammonium sulphate 0.1 M bis-tris pH 6.5 25% w/v PEG 3350	ΔLGS 1 58 177 194	3.8
0.1 M citric acid pH 5.0 20% w/v PEG 4000	ΔLGS 1 58 177 194	4.0
0.04 M potassium phosphotate (monobasic) 20% v/v glycerol 16% w/v PEG 8000	ΔLGS 1 58 177 194	4.0

Table 11 Data collection of ΔLGS: pA 6mer and ΔLGSMS: pA 6mer RNA crystal complex

Resolution of X-ray diffraction from M2-1 truncation mutants co-crystallised with RNA pA 6mer. Data collected on the I02 beamline, DLS.

3.8.2.2 M2-1: RNA complex visualisation by electron microscopy

Due to M2-1 in complex with RNA not being determined by X-ray crystallography at this time, we attempted to visualise the complex using negative stain transmission electron microscopy, reasoning that it could provide clues as to why the co-complex could not be crystallised. The negative stain images seen by electron microscopy clearly show the aggregation of M2-1 in complex with RNA, with aggregated complexes ~ 25 nm in diameter (Figure 35A). Co-incubation of M2-1 with a P peptide resulted in visibly less aggregated protein (Figure 35B). In the presence of RNA the M2-1 protein is able to aggregate and presents a heterogeneous population, not ideal for crystal formation.
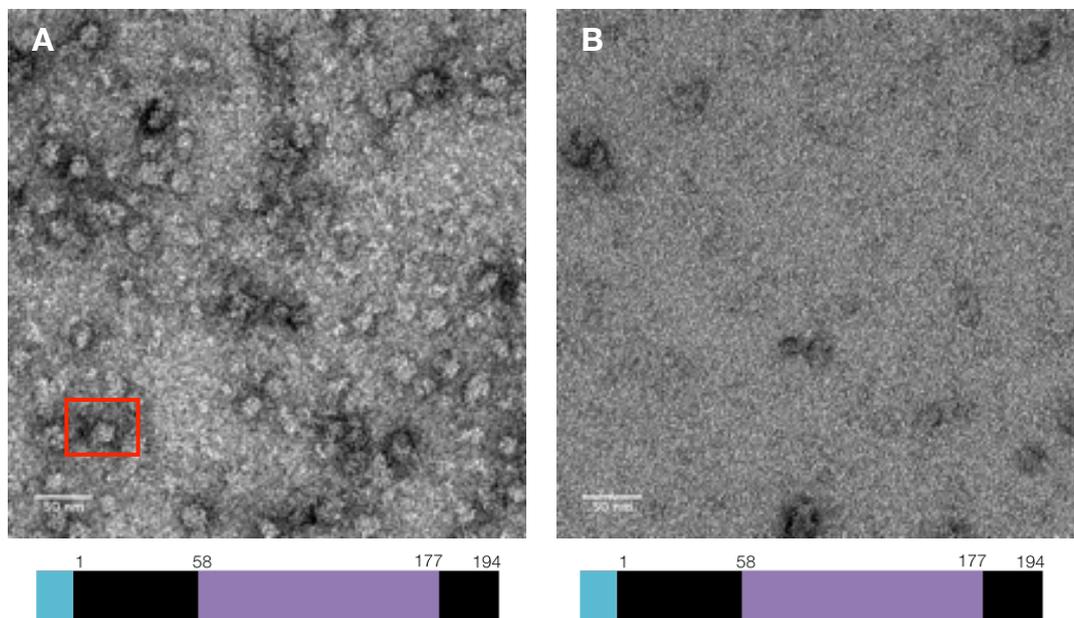
Figure 35 Negative stain electron microscopy of HRSV M2-1 in complex with RNA and P 90-110

A- RNA: M2-1; B- the M2-1: P 90-110 complex. An aggregate species is highlighted, red, in image A. Images provided by the EM facility, University of Leeds. Sidebar indicates a length of 50 nm.

3.8.2.3 M2-1$_{58-177}$: RNA crystallisation

To overcome possible interactions of RNA across more than one M2-1 tetramer, that may provide a reason for RNA induced M2-1 protein aggregation visualised in electron microscopy images (Figure 35), a monomeric M2-1 construct (M2-1$_{58-177}$) was trialled. Here, the purified M2-1$_{58-177}$ was screened for crystal formation in presence and absence of RNA with a protein concentration of 15 mg/ ml and 30 mg/ ml at 18°C. The truncated, monomeric form of M2-1 had been studied previously by NMR but not by X-ray crystallography (Blondot et al., 2012). At this time, M2-1$_{58-177}$ only formed oil-like droplets (Figure 36) in all commercial screen conditions trialled and were not suitable for further analysis by X-ray crystallography.
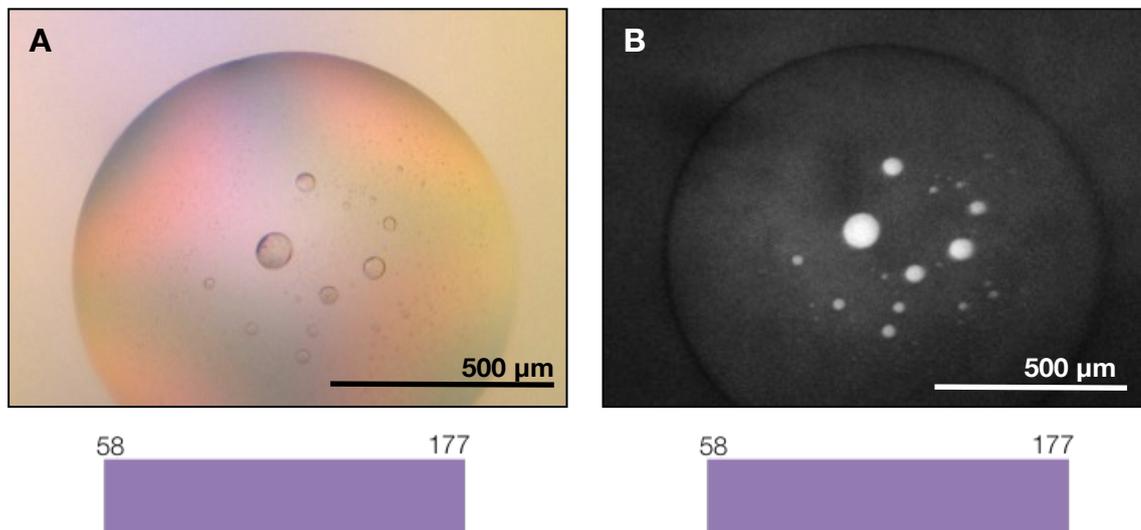
Figure 36 M2-1$_{58-177}$: RNA protein complex drop images

A- M2-1158-177 incubated with pA 6 mer in 0.2 M calcium acetate, 0.1 M tris pH 7.0, 20% w/v PEG3000 shows oil-drop like protein; B same as image A with UV imaging of the drop, fluorescence of the oil droplets suggest the presence of protein. Visualised using the RockMaker Imager.

3.9   The M2-1: P interaction

3.9.1   Binding interaction

The M2-1: P interaction, like that of M2-1: RNA, is critical for M2-1s anti-termination function with residues 90-110 of P implicated in a direct interaction with M2-1 (Figure 37) determined via a pulldown assay of GST fused to truncated P constructs. A lack of M2-1 binding at the N and C terminal regions of P indicated a single binding site with the smallest peptide that still bound to M2-1 being P 90-110 (personal communication, Dr Jean-François Eléouèt, Institut National de la Recherche Agronomique; pulldown assay Figure 37 performed by Dr Kavestri Yegambaram, University of Leeds). The HRSV P protein is punctuated by unstructured regions that have made the full length protein unsuitable for crystallography, therefore for both structural and functional studies here, the P 90-110 peptide in complex with M2-1 was further studied.
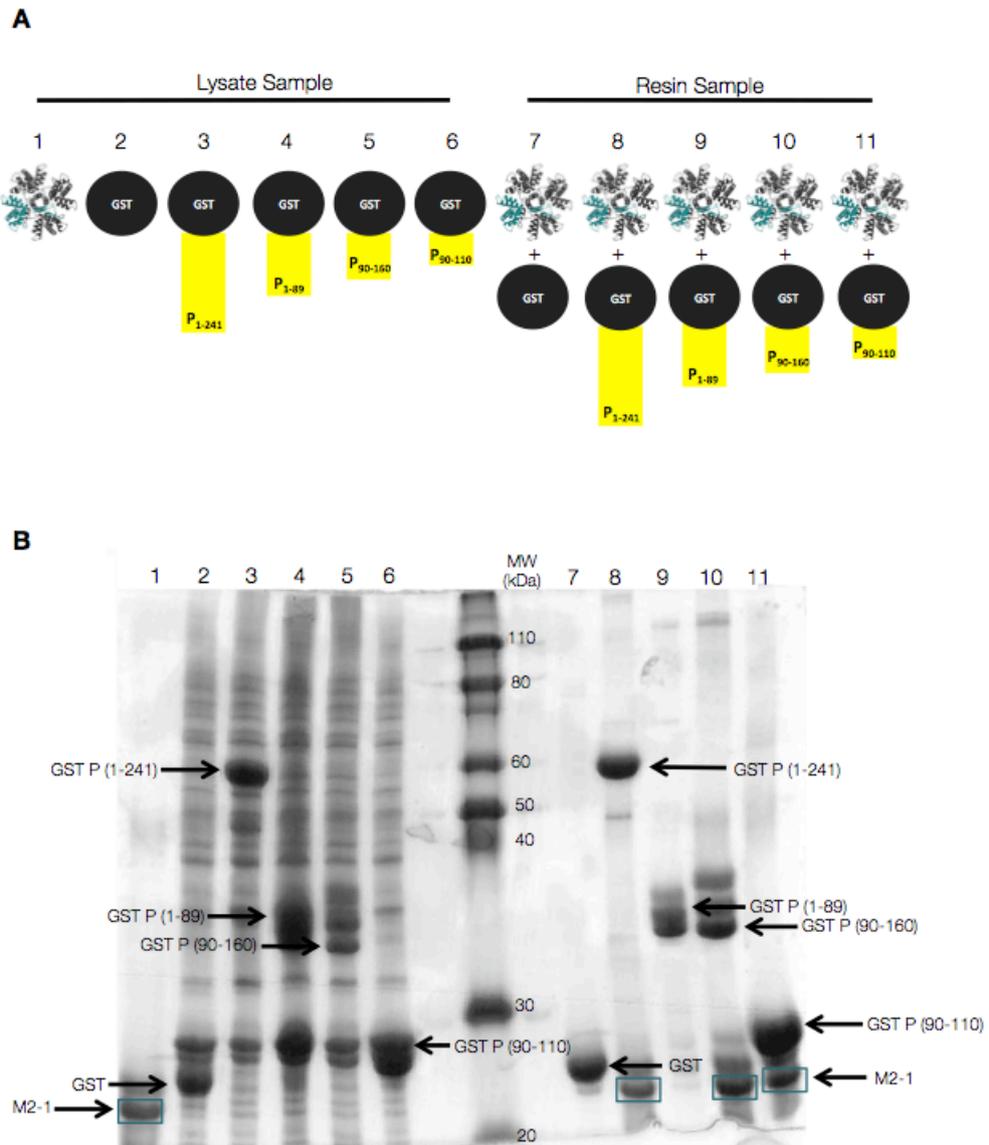
Figure 37 Binding of M2-1 to full length and truncated P constructs

Full length and truncated P constructs fused to GST that can bind glutathione resin were assessed by SDS-PAGE, stained with Coomassie blue, for the ability of the P protein to interact and pull down M2-1. A-schematic representation of GST, M2-1 or P constructs in the relative lanes in the SDS PAGE (15%) gel stained with Coomassie blue, B- blue boxes in lanes 8, 10 and 11 represent M2-1 protein that bound to the P-GST fusion. Assay constructed and Figure B performed by Dr Kavestri Yegambaram, University of Leeds.

3.9.2 P 90-110 binding interaction

With an understanding of the minimum region required for P binding to the M2-1 protein the affinity of the interaction was assessed by binding of a fluorescently labelled P 90-110 to M2-1 using FA. In the binding assay the short P peptide (residues 90-110, P 90-110) bound weakly to M2-1 with an apparent K_d of 7.5 ± 7.7 µM (Figure 38A). P 90-110 bound to M2-1 with a lower affinity than the anti-genome complement (+) GE oligonucleotides and a pA13 sequence. In a competition binding assay P 90-110 was out-competed by pA13 with an IC_{50} of 1.7 ± 0.08 µM (Figure 38B). However, RNA could not be outcompeted by P 90-110. Monomeric $M2-1_{58-177}$ was also assessed for its binding to the P 90-110 binding, as performed for RNA in section 3.8.1. The apparent affinity of P 90-110 for M2-1 did not significantly alter between the full length and truncated protein forms (P 90-110 and full length M2-1 K_d 7.5 ± 7.7 µM; P 90-110 and $M2-1_{58-177}$ K_d 12 ± 0.76).
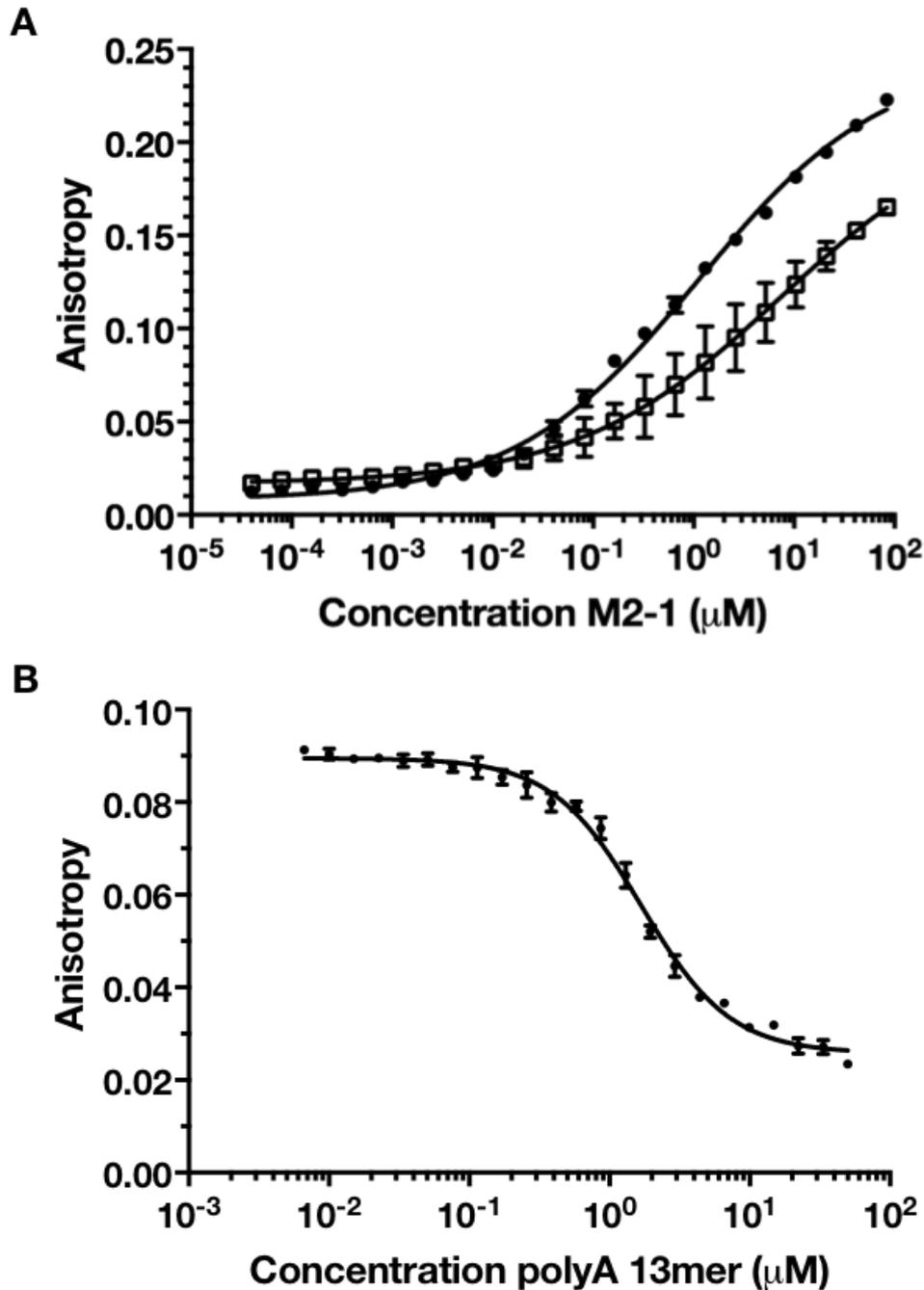
Figure 38 Binding assays of P 90-110 to M2-1.

A- direct binding of M2-1 to P 90-110 (n) and polyA 13mer RNA (●); B-competition assay. polyA13mer RNA is titrated at varying concentrations against P 90-110 for binding to M2-1 (fitted to a 1:1 model of binding). Data points represent n-3 ± SEM with each n number performed in triplicate.

3.9.3 Co-crystal trials of M2-1 and P 90-110

The results of the previous section show that M2-1 specifically binds P 90-110 (Figure 37 and Figure 38). The sequence of P 90-110 is mostly composed of hydrophobic residues and is predicted to have alpha helical propensity (Figure 39, 1D Protein Structure Prediction Server http://biomine.ece.ualberta.ca). These observations suggest that P 90-110 is an ideal peptide for co-crystal binding as the helix may bind to the surface of M2-1 in a way that is compatible with the current packing.

```
P90-110>    DPTPSDNPFSKLYKETIETFD
Secondary>  CCCCCCCHHHHHHHHHHHHHHC
```

Figure 39 Secondary structure prediction of P 90-110.

> Using the Protein Structure Prediction Server. P 90-110 was predicted to have a helical content of 64%. C-coiled, H-helical.

M2-1 was incubated with P 90-110 in a 1:1 molar ratio before further concentrating the complex to 8 mg/ ml. Crystals were seen in various conditions (see appendix III) using a sitting drop vapour diffusion method but importantly, compared to all previous crystal trials an alternative morphology was visualised. In addition to the plate morphology previously visualised, a needle-like morphology was visualised for the M2-1: P 90-110 co-crystal trials (Figure 40). These alternate morphologies were the focus of further studies as the different morphology could have been, for example, due to peptide binding and altering the space group. The needle-like morphology crystals were seen in 3 conditions incubated at 18°C:

1) 0.1 MMT buffer pH 9, 25% PEG 1,500 PACT *Premier*™
2) 0.2 M trimethylamine N-oxide dihydrate, 0.1 M tris pH 8.5, 20% w/v PEG MME 2,000
3) 0.1 M potassium thiocyanate, 30% w/v PEG 2,000

The 3 conditions that resulted in crystals with a needle-like morphology were further analysed by performing a fine screen based on the PEG concentration of the mother liquor. Drops that contained crystals with a needle-like morphology in the optimised screens were visualised for condition 1 with 25, 28 and 30% PEG 1,500 and in condition 2 with 18 and 20% PEG MME 2000. No crystals were seen for condition 3 when the PEG concentration was optimised.
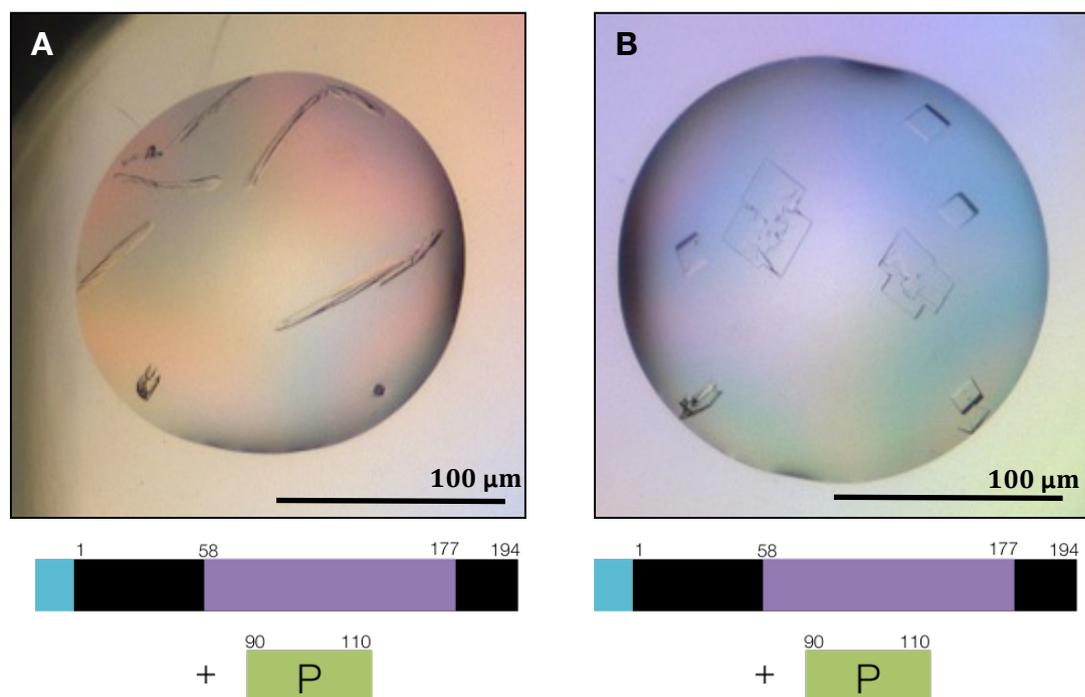
Figure 40 M2-1: P 90-110 co crystal complex morphology

A- M2-1 incubated with P 90-110 in 0.2 M trimethylamine N-oxide dihydrate, 0.1 M tris pH 8.5, 20% w/v PEG MME 2,000 and has a needle like morphology not previously seen for M2-1 crystals; B- M2-1 incubated with P 90-110 in 0.2 M sodium bromide, 0.1 M bis-tris propane pH 7.5, 20% w/v PEG 3350 has a plate morphology seen previously for the unbound M2-1 protein. Crystals were visualised using the RockMaker imager.

3.9.4 Structure determination of the M2-1: P 90-110 complex using X-ray diffraction

A subset of needle-like and plate morphology crystals were harvested and tested for diffraction on the I02 beamline, DLS. Full data sets were collected for crystals grown in condition 1 (0.1 MMT buffer pH 9, 25% PEG 1,500 PACT crystal screen) and condition 2 (0.2 M trimethylamine N-oxide dihydrate, 0.1 M tris pH 8.5, 20% w/v PEG MME 2,000), both with a needle-like morphology. The needle-like crystals had an orthorhombic (P222) crystal form and was different to the two crystal forms (P2$_1$ and P422) determined for the known crystal structure of unbound M2-1 with a plate crystal morphology. Crystals from condition 1 diffracted to a maximum resolution of 3.5 Å whereas crystals from condition 2 diffracted to a resolution of 2.6 Å (Figure 41 & Table 12) and was subsequently chosen for further X-ray structural analysis.
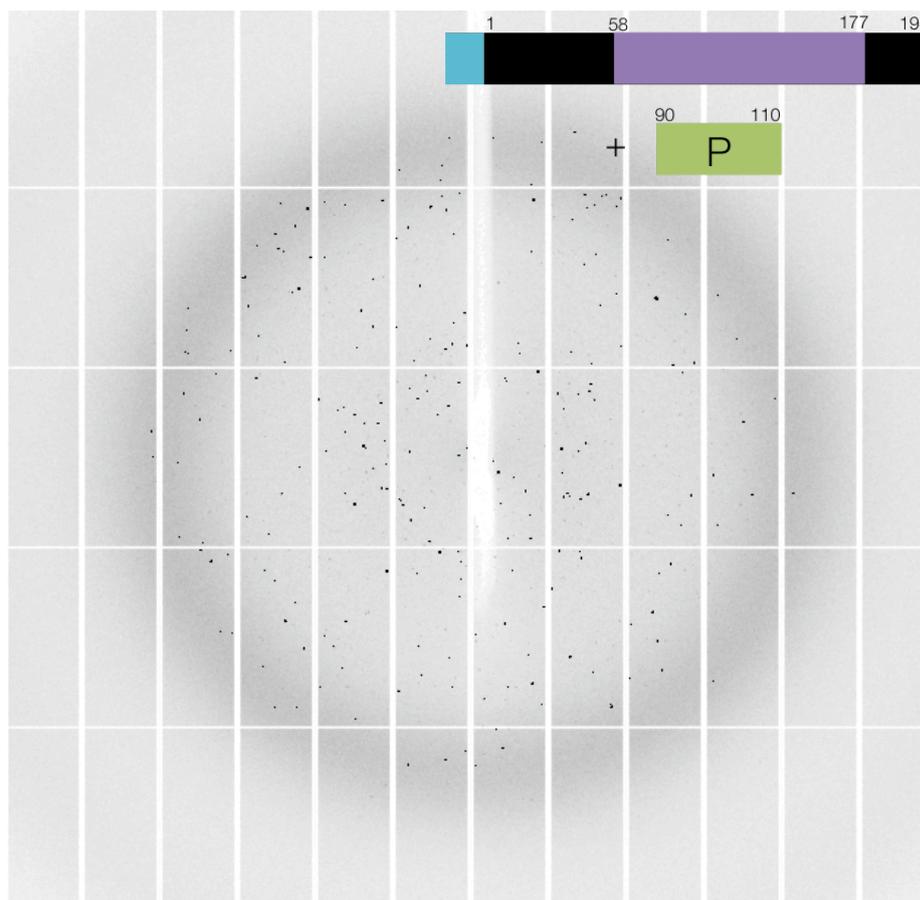
Figure 41 Diffraction image of a M2-1: P 90-110 crystal

Condition 0.2 M trimethylamine N-oxide dihydrate, 0.1 M tris pH 8.5, 20% w/v PEG MME 2,000. Images were collected on the I02 beamline, DLS under cryo-temperatures at 100K to a resolution of 2.6 Å.

The data set (Figure 41) was processed using the CCP4 suite of programmes and the structure solved using molecular replacement with M2-1 as the reference model (PDB: 4c3b). The electron density maps ($2F_{obs}$-$1F_{calc}$ and $1F_{obs}$-$1F_{calc}$ difference map) revealed a continuous area of density not accounted for by M2-1 that had obvious helical structure. The P 90-110 peptide backbone was manually built in to the density (Figure 42). Side chains were subsequently modelled after several rounds of refinement. The data collection and refinement statistics are displayed in Table 12 for this crystal. M2-1: P 90-110 crystallised as a tetramer with each M2-1 monomer bound to one P 90-110 peptide (continuous, unexplained density seen once for each M2-1 monomer) and displayed a 1:1 stoichiometry (i.e. 4 peptides per M2-1 tetramer). The co-crystal structure was solved in space group $P2_1\ 2_1\ 2$ with one tetramer per asymmetric unit (AU). Previously determined unbound M2-1 was found in two space groups, P422 with one half tetramer in the AU and $P2_1$ with four tetramers in the AU. Further, alternate crystal contacts were visualised for the co-crystal structure compared to those seen previously for the unbound M2-1 structure.
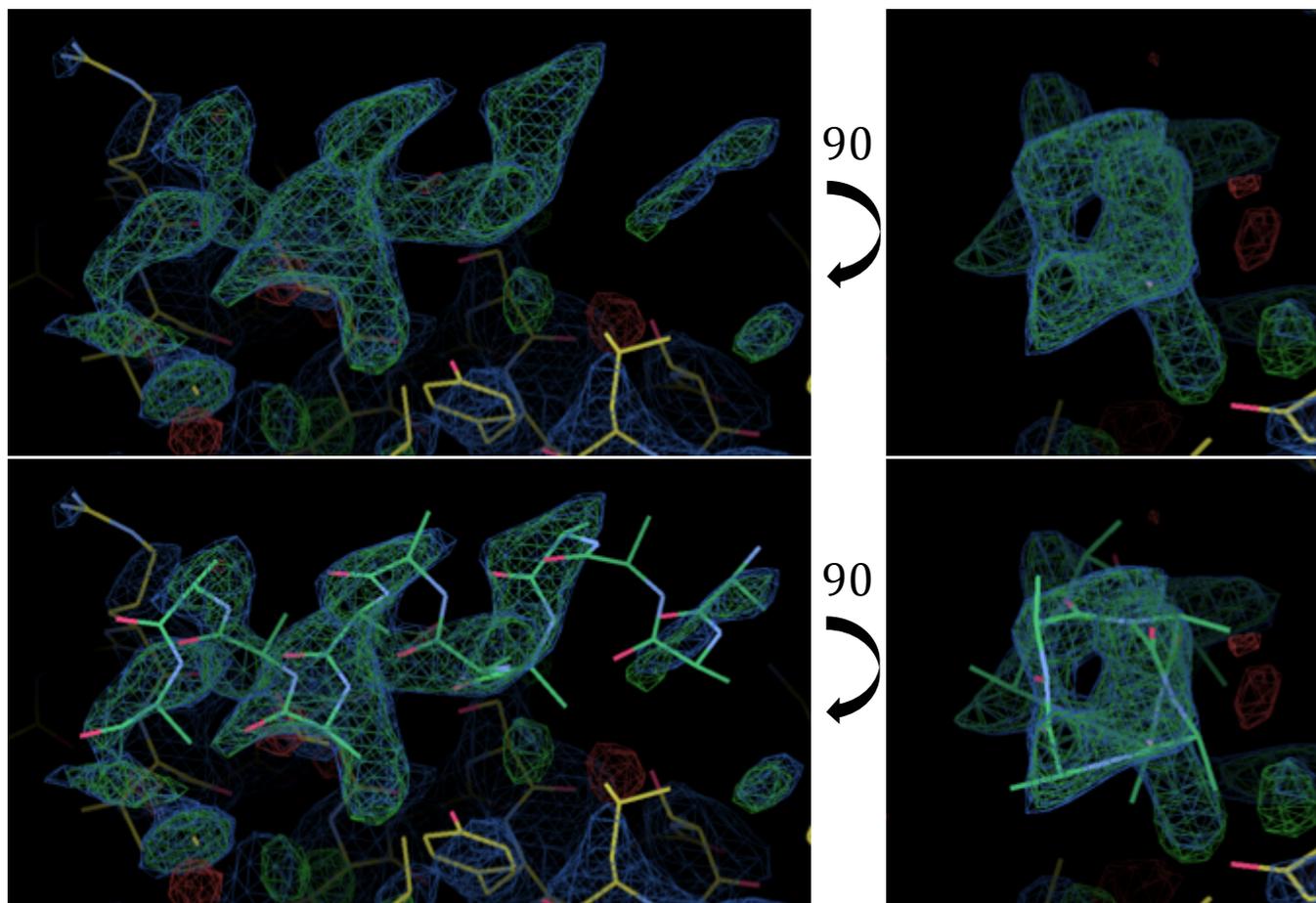
Figure 42 Electron density for the P 90-110 peptide visualised in COOT.

M2-1 residues were fitted within the density using the M2-1 model (yellow backbone). The two orientations shown are at 0o and 90o. (top panel) and show the clear density present in both the 2Fobs-1Fcalc and 1Fobs-1Fcalc difference maps before manually fitting of the peptide. The bottom panel shows auto-fitting using COOT of a helix to the density.

Property	Value
Wavelength	0.9795 Å
Space group	$P2_1\,2_1\,2$
Reflections/ unique reflections	443175/ 25813
Cell Constants a, b. c, α, β, Y	96.37 Å, 71.96 Å, 116.01 Å 90° 90° 90°
Resolution	116- 2.6 Å
% Data completeness	99.9 %
$<I/\,\sigma\,(I)>$	1.58
R, R $_{free}$	0.265, 0.292
R f$_{ree}$ test set	1289 reflections (5.26%)
Wilson B_factor (Å^2)	55.6
Anisotropy	0.744
Total number of atoms	5259
B factor ligands	75
B factor waters	67
B factor M2-1 tetramer	33
B factor P 90-110 peptide	
(1)	28
(2)	28
(3)	28
(4)	84
Ramachandran Outliers	1.7 % (11/638 residues)

Table 12 X-ray data collection and refinement statistics for the M2-1: P 90-110 co-crystal complex

Density for the P 90-110 peptide was poor in parts, which made it difficult to accurately determine the correct amino acid for the density. Similarities in amino acid side chain structure (Tyr/ Phe) and the repetitive nature of the P 90-110 sequence added to the challenge (Figure 43). The N terminus of the peptide consisted of 3 proline residues and secondary structure formation would result in a 'proline kink' thus preventing the formation of an alpha helix. The clearly visible alpha helical density would suggest the density did not contain proline residues, which helped to assign the sequence. A bulky side chain density of the un-accounted density fitted both Tyr 102 and Phe 98. Further manual fitting of the residues to the helical backbone determined a Tyr residue (102) at this position would allow successful fitting of the peptide to the density (Figure 43). Other manual fitting positions resulted in poor fitting of the side chains to the density. After sequential rounds of refinement and manual fitting and alterations, the final structure had clear density for 3 P peptides. This is further illustrated by the low B-factors which are equivalent to the B-factors for the overall protein backbone, giving confidence of reliability of the fit of the residues to the density. One peptide had poor density with a high B factor and to prevent over-fitting of the data only an all-Ala helix was modelled in the final structure. The best visible peptide density resulted in reliable fitting of P residues Pro 97 – Phe 109 (Figure 44).
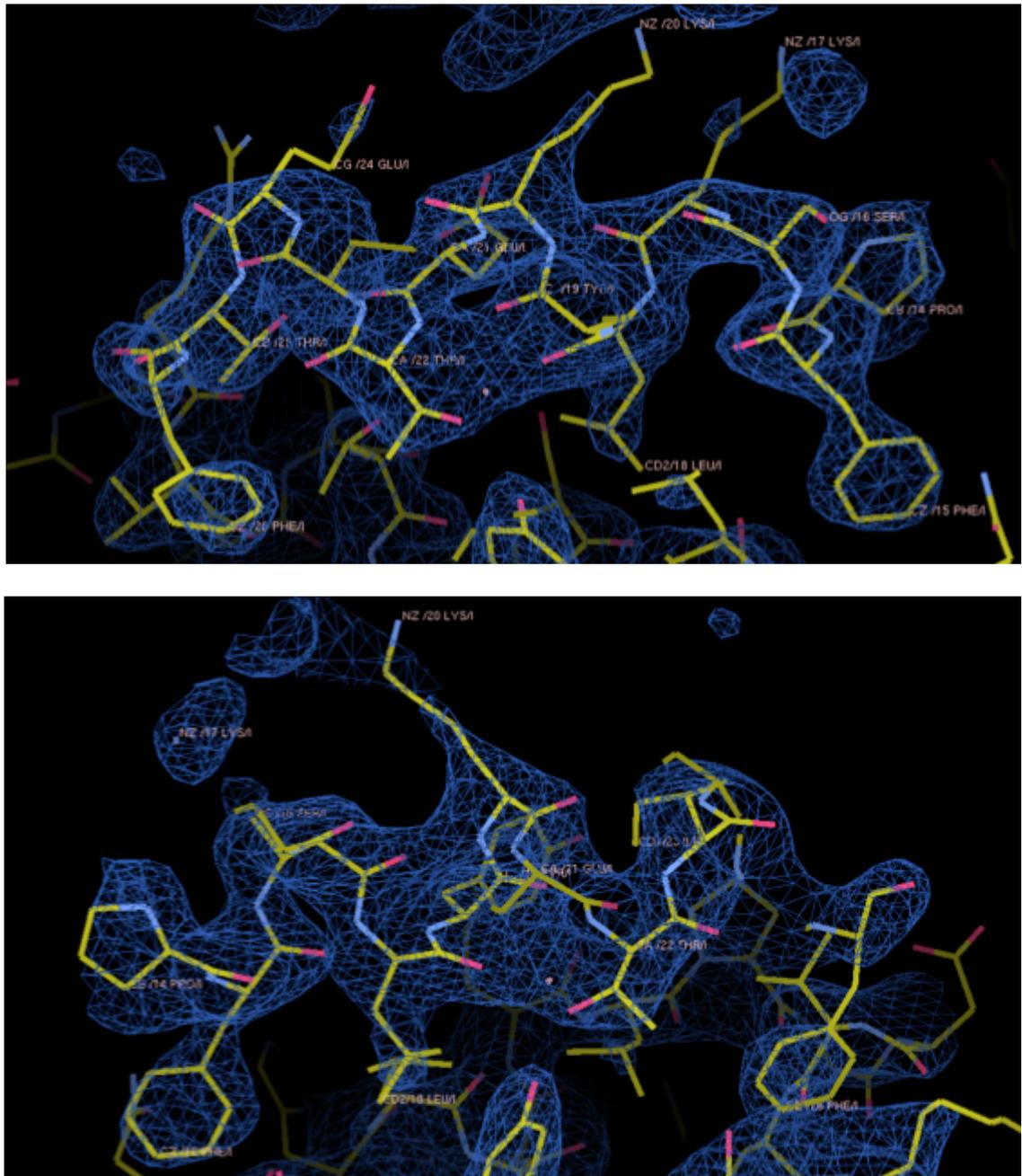
Figure 43 Model fitting of P 97-109 to the unaccounted electron density in the M2-1 crystal structure

Fitting of an alpha helix backbone to the density within the map with individual amino acids fitted into the density (yellow). Left to right P residues Pro 97 – Phe 109.
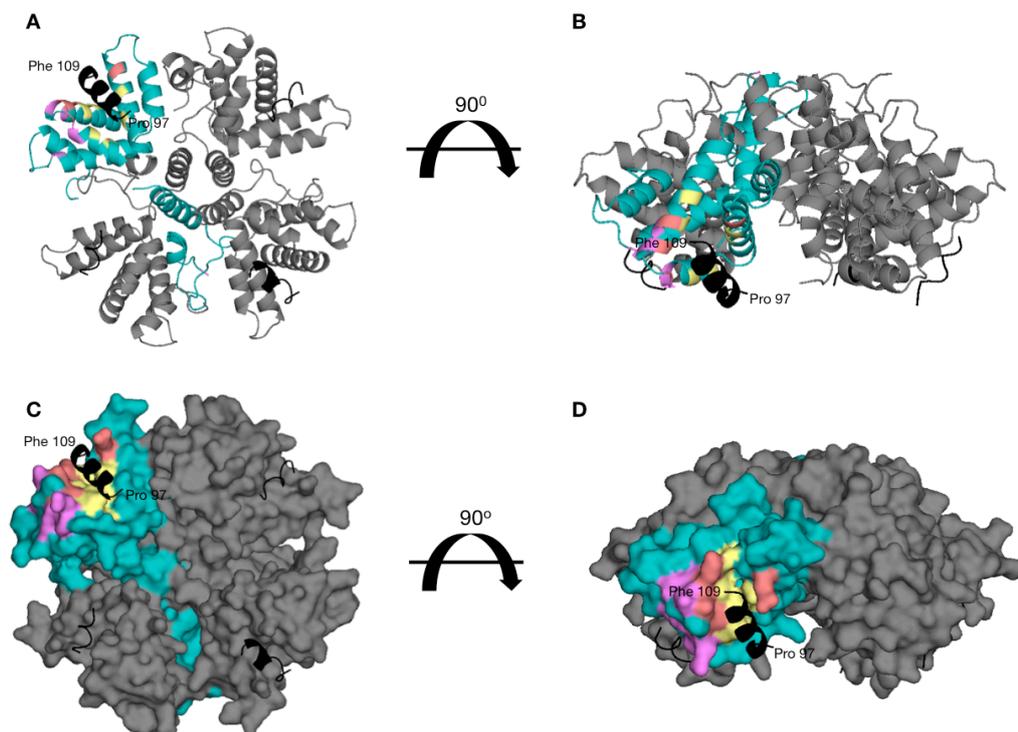
Figure 44 M2-1: P 90-110 crystal structure

The tetrameric M2-1 protein (grey) in complex with the P 90-110 peptide (black) to 2.59 Å resolution. A- the M2-1 tetramer (grey) with one monomer highlighted in blue. Residues hypothesised to be involved in M2-1 binding are highlighted as pink-RNA, yellow, P and coral- P and RNA whilst the determined P 90-110 helical structure is in black, showing an agreement of the binding region between the hypothesised P and known P binding site; B- rotation of A by 90o; C surface representation of A; D- surface representation of B.

3.9.5 Visualisation of the M2-1: P 90-110 interactions

After modelling of the P 90-110 peptide to the electron density based on the side chain positions, the interactions and orientation of P 90-110 in relation to the M2-1 protein could be visualised for the first time to determine interactions that are responsible for binding (Figure 44 and Figure 45) with the interactions of the most complete peptide in the tetramer assessed further. Four electrostatic interactions could be formed based on interaction distances (Figure 46 and Figure 47) and density for the side chains (Figure 48). From the density maps (Figure 48) all residues of M2-1 that were hypothesised to interact with P 90-110 had full occupancy and suggest a stable conformation of the side chains that would subsequently allow stable interactions with P 90-110. Bond lengths for the

interactions can be measured, however their accuracy is low due to the resolution (2. 6 Å) and relatively poor quality of the electron density map in this region with subsequent fitting of the side chains to the density. Arg 126 of M2-1 may interact with Glu 107 of P 90-110 through ionic binding. This interaction is stronger than hydrogen bonding, coupled with the low bond distance (2.8 Å) and density for the Arg 126 side chain further suggest this interaction occurs. The other three observed interactions occur through hydrogen bonding with Tyr 134, Arg 151 and Thr 130 of M2-1 interacting with Thr 105, Tyr 102 and Thr 109 of P 90-110, respectively.
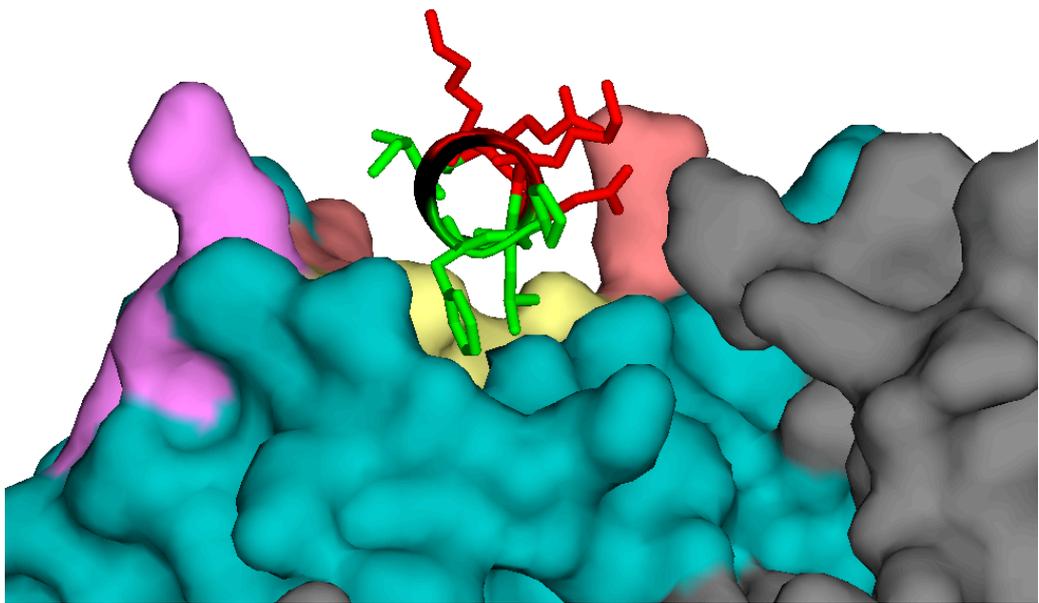
Figure 45 Orientation of the phosphoprotein side chains

The helical P 90-110 peptide in complex with M2-1. Surface representation of the M2-1 tetramer (grey) with one monomer highlighted in blue. Residues hypothesised to be involved in M2-1 binding are highlighted as pink-RNA, yellow, P and coral- P and RNA. The structurally determined P 90-110 peptide bound to M2-1 is shown as an alpha helical structure with hydrophilic charged residues facing outwards and highlighted in red with side chains shown; Lys100, Lys 103, Glu 104 and Glu 107. Hydrophobic amino acids face towards the M2-1 protein and are highlighted green with side chains shown; Pro 97, Phe 98, Leu 101 and Ile 106.
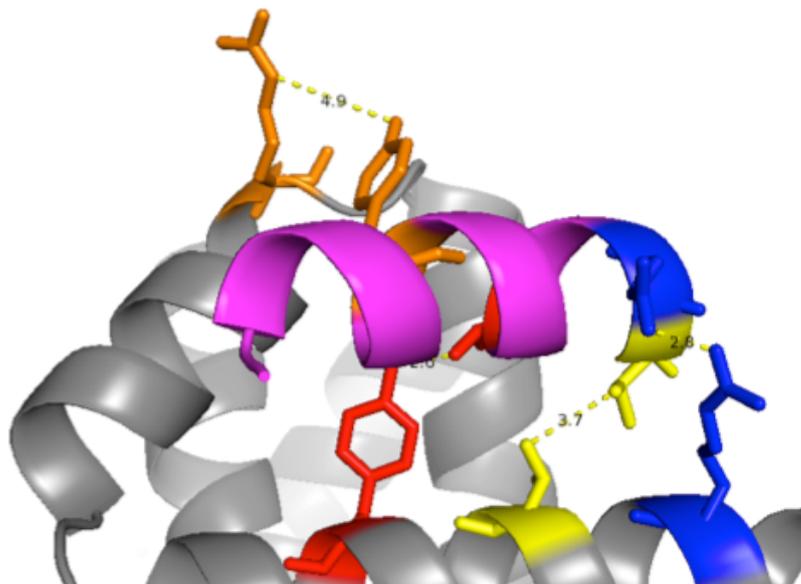
Figure 46 M2-1 and P 90-110 electrostatic interactions

Four interactions are highlighted from left to right: M2-1 Arg 151 interacts with Tyr 102 of P 90-110 and is coloured orange. M2-1 Tyr 134 interacts with P 90-110 Thr 105 and is highlighted in red. M2-1 Thr 130 interacts with Thr 108 of P 90-110 in yellow. Finally in blue M2-1 Arg 126 forms ionic interactions with Glu 107 of P 90-110.
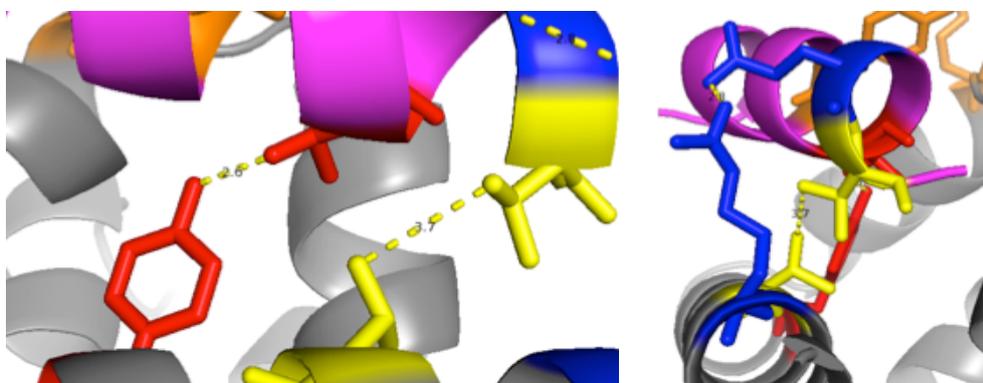
Figure 47 Further orientations of the M2-1 and P 90-110 interaction

Left panel highlights the two interactions: M2-1 Tyr 134 with P 90-110 Thr 105 (red) and M2-1 Thr 30 interacts with Thr 108 of P 90-110 (yellow). Right panel highlights the potential M2-1 Arg 126 ionic interaction with Glu 107 of P 90-110 (blue).
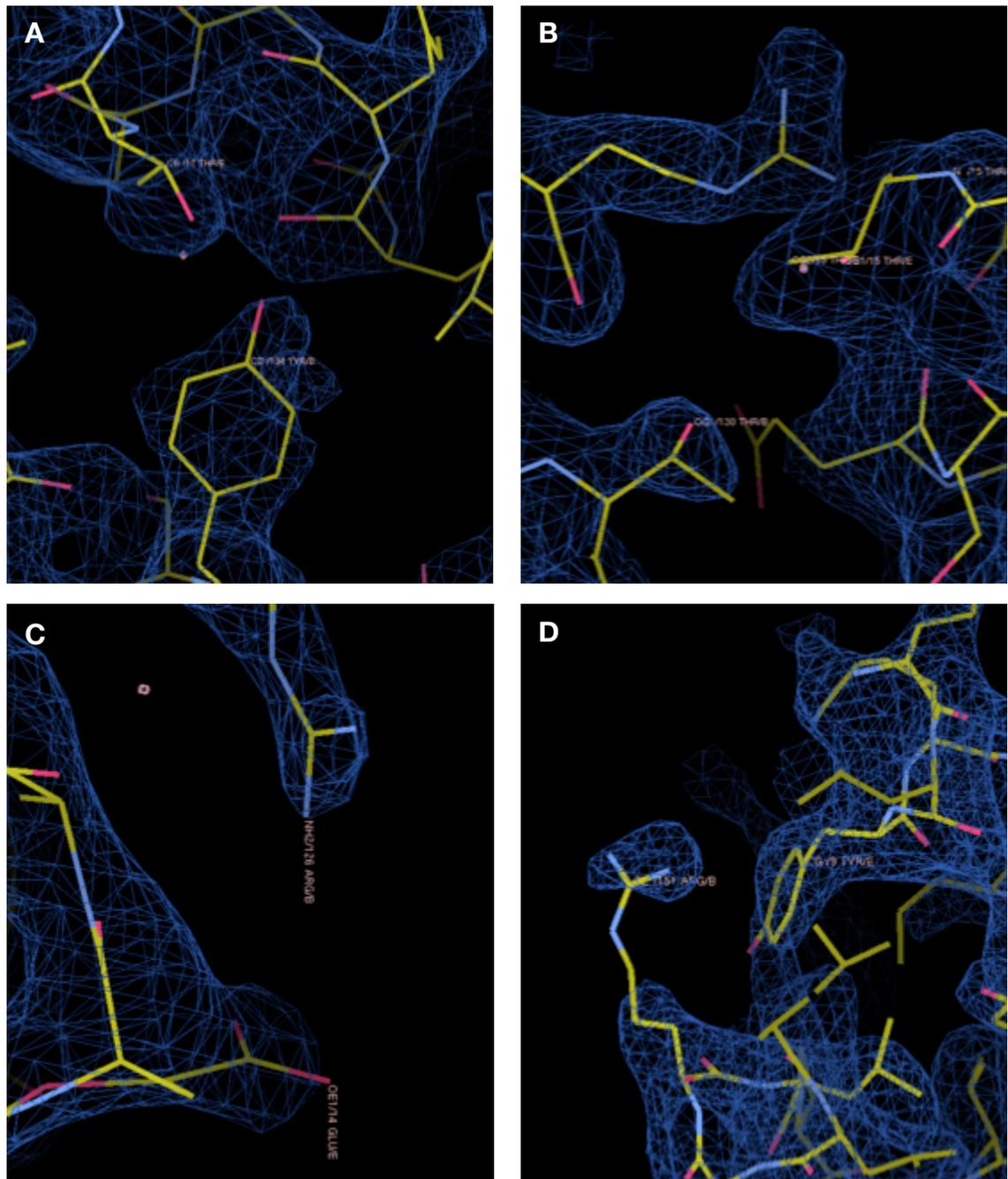
Figure 48 Electron density maps of the M2-1: P 90-110 interactions

Images of the M2-1 and P 90-110 amino acid side chains fitted to the 2Fobs-1Fcalc density map. A- M2-1 Tyr 134 and P 90-110 Thr 105; B- M2-1 Thr 130 to P 90-110 Thr 108; C- M2-1 Arg 126 to P 90-110 Glu 107 and D- M2-1 Arg 151 to P 90-110 Tyr 102.

3.10 Discussion

3.10.1 Purification improvements resulted in high protein yields

To allow material intensive methods such as X-ray crystallography to be pursued to aid our understanding of M2-1 and its binding partners there was a need for an improved M2-1 purification method that allowed consistently higher protein yields. A new purification method required less protein degradation as this resulted in a non-homogenous sample that is detrimental to crystal formation (Jancarik & Kim, 1991). The M2-1 protein degradation over time, visualised by SDS-PAGE analysis, was reduced here by performing the purification protocol in 2 days, from the previous 5 days. Further, understanding that the protein was tetrameric in nature from analytical and structural studies meant that aggregation was detrimental to our studies and any aggregation such as that from a harsh mechanical lysis methods like sonication exponentially affected the protein yield. The optimised purification method used a freeze-thaw method that allowed greater temperature regulation and higher protein yields. Improvements to the protein purification protocol resulting in ~ 50 mgs of pure M2-1 protein from 10 g of bacterial cell pellet are summarised in Table 7.

Improvements to the M2-1 purification method resulted in the ability to undertake further X-ray crystallography experiments. A vapour diffusion method with sitting drops was used initially for crystal formation with commercial factorial screens. Through optimisation, larger drop sizes, different ratios of protein to mother liquor and hanging drops using the vapour diffusion method were assessed. Interestingly, protein crystals were not seen for corresponding conditions using the hanging drop method for the M2-1 protein and a 1:1 ratio of M2-1 protein to mother liquor resulted in reproducible crystal formation for subsequent studies. Likely the 1:1 ratio and sitting drop method resulted in ideal surface tension conditions for M2-1 crystal formation. Diffraction data collected for the M2-1 protein using the new purification protocol resulted in a crystal structure of M2-1 solved by molecular replacement to 2.2 Å resolution. However, despite providing an increased resolution structure, density not resolved in the original published structure (Tanner et al., 2014) such as the C-terminus residues 175-194 and the phosphorylation loops encompassing Ser 58 and Ser 61 were not further resolved in the structure determined here and thus no further conclusions could be drawn regarding the M2-1 structure from that previously published.

3.10.2 M2-1 preferentially binds A-rich RNA sequences and competitively outcompetes P 90-110

There has been considerable speculation regarding the RNA-binding target of the M2-1 protein, and how this interaction might result in anti-termination during transcription (Kuo et al., 1996; Fearns & Collins, 1999). Assuming the RNA target is viral in origin, possibilities include either the positive sense RNA template copied during replication, or the product of transcription itself, namely the HRSV mRNAs. As M2-1 is thought to influence the transcription of all 10 HRSV genes (Hardy, Harmon & Wertz, 1999) a further prediction that could be made is that the target sequence will be present in all of these genes. Due to the heterogeneity in nucleotide sequences for the 10 HRSV genes, there are few sequences that this could comprise, but strong possibilities are the conserved GS and GE sequences, as well as the mRNA 3' poly(A) tail.

Here we analysed RNA sequences referring to HRSV specific genomic (-) GE oligonucleotides and their reverse complements (+) to determine their specificity for M2-1 based on previous work stating a necessity of GE sequences for transcription termination and polyadenylation (Kuo et al., 1996; Fearns & Collins, 1999). Using one length of genomic RNA (13 nts) and their anti-genomic sense counterparts allowed direct comparison of their Kd values and excluded issues of avidity due to nucleotide length. RNA of 13 nucleotides in length were chosen with the hypothesis that M2-1 binds to semi-conserved GE sequence to inhibit transcription and synthesise polycistronic read-through mRNAs (Hardy & Wertz, 1998; Fearns & Collins, 1999). In addition, the binding of the short P 90-110 construct that could bind to M2-1 in a pull down assay was analysed and all binding affinities compared for the full length tetrameric M2-1 and a truncated monomeric M2-1$_{58-177}$ to determine if the binding interaction was specific to the core domain.

The data presented here shows M2-1 exhibits highest affinity for A-rich RNA sequences, including those that represent the conserved HRSV GE sequences, which is in general agreement with previous data from others (Blondot et al., 2012, Tanner et al., 2014 and Leyrat et al., 2015). Leyrat et al, (2015) proposed a RNA: M2-1 binding model of 13 nucleotides in which nucleotide 5, specifically an adenosine (A5) provides the binding specificity of the RNA to M2-1, with A6 increasing this specificity. Here a U at position 5 for SH GE (-) had a lower affinity than an A at position 5 for NS1 GE (-) suggesting that an A at nucleotide 6 may aid binding specificity. The data presented here indicates that it is in fact the number of sequential adenosine nucleotides that directly contributes to the

oligonucleotides relative affinity to M2-1, with the longest A tract (polyA) followed by SH GE (+) having the highest affinities. The data illustrates no length or nucleotide position requirement for binding specificity although 13 nucleotides fits along the binding cavity and along a surface that includes ZBD residues implicated in binding Arg 3 and Arg 4 by Tanner et al., 2014.

The monomeric M2-1$_{58-177}$ form synthesised here does not contain the ZBD, oligomerisation domain or the unstructured CTD. The binding affinities demonstrate a similar patter to tetrameric M2-1 in which the length of the A tract helps binding. However, monomeric M2-1$_{58-177}$ displayed a reduced affinity for each RNA compared to tetrameric M2-1 and suggests for the higher binding affinities as seen for tetrameric M2-1 there is a requirement for the ZBD, oligomerisation or CTD region for higher affinity binding. As the binding curves determined by FA for full-length and M2-1$_{58-177}$ could not be saturated, the Kds may not be as accurate as presumed and thus have only been reported as 'apparent' Kds to 2 s.f. Despite the inability to saturate the system at the concentration of M2-1 used, the relative patterns in binding affinities (i.e. consistently higher Kds for the + counterpart) can be considered accurate and conclusions drawn that support the literature in terms of M2-1 binding to HRSV mRNA (+) transcripts.

Based on these findings, our data provides further evidence that M2-1 may recognise mRNA GE sequences rather than the genome template. The mRNA transcripts and positive sense genome complement contain a higher composition of A/ U nucleotides than the viral negative sense genome. As mRNA transcripts are not encapsidated (like the positive anti-genome) and are A rich it could be suggesting that these are the binding target of M2-1. The 3' ends of mRNAs (the GE positive complement) are A rich and thus in the literature have been described as the target of M2-1. However on looking at the A/ U composition throughout each transcript it can be seen that each transcript has an A tract of at least 4 nucleotides on average every 200 nucleotides. Research has shown that L alone can synthesis short transcripts of around 200 nucleotides hence synthesis of NS1 and partially NS2 in the absence of M2-1 (Fearns & Collins, 1999) and thus it could be suggested that in fact M2-1 recognises A rich sequences thorough the length of the transcripts to prevent intragenic anti-termination.

Gene	Gene start (%)	Gene end (%)	Gene junction (%)	Tracts of 4 or more A in each transcript	Coding transcript length (nts)	Nucleotides/no of tracks (nts)
NS1	50	92	74	2	420	210
NS2	50	92	70	1	375	375
N	50	92	100	6	1176	196
P	50	85	44	3	726	242
M	50	92	77	4	771	193
SH	50	92	73	0	195	
G	40	85	69	4	897	224
F	50	92	67	7	1725	246
M2	50	92	-	5	800	160
L	60	92	-	28	6498	232

Table 13 Composition of A/ U bases in HRSV viral mRNA transcripts

The binding affinities stated here for short RNA sequences and M2-1 are different to the published work (Tanner et al., 2014) from the same group. Protein concentration determined for previous assays used a dye binding method which can be inaccurate if standard curves are not determined and differentially interacts with different amino acids. For all purified protein used here, protein concentration was determined using UV absorbance at 280 nm which provides a more accurate determination of protein concentration from the intrinsically fluorescent aromatic rings of amino acids Trp and Tyr. In addition, the new method of M2-1 purification could have contributed to the concentration difference seen. This was ruled out by purifying M2-1 with the old method and determining binding affinities whilst measuring protein concentration by UV 280 readings. Similar concentrations to the new purification method were seen and suggests it was the method of protein concentration determination that was the issue.

Previous research has illustrated the competitive manner of RNA and P binding to M2-1 (Tran et al., 2009) via adjacent and overlapping binding sites (Figure 44). Both P and RNA have been found through functional assays to bind to the core domain and most recently Tanner et al, (2014) implicated residues 3 and 4 of the M2-1 ZBD as being involved in RNA interactions and was later supported by the findings of Leyrat et al, (2015). Here, data showed out competition of the P 90-110 peptide with short, A rich 13 nucleotide RNA sequence further suggesting an overlapping binding site for RNA and P. In a direct assay P 90-110 had a lower binding affinity than A rich RNA sequences assayed, however both bound weakly and were in the micromolar range. Interestingly, although P 90-110 could be outcompeted by a pA 13mer oligonucleotide, no RNA sequence assessed here (see Table 9 for sequences) could be outcompeted by the P 90-110 peptide.

Affinities of the RNA and P peptide sequences to both full-length tetrameric M2-1 and a truncated version that produced monomeric M2-1$_{58-177}$ were directly compared to determine the relative contributions of the full-length tetrameric protein and a monomeric protein lacking the ZBD, oligomerisation and CTD. RNA sequences bound with at least 4-fold weaker affinity to the monomeric M2-1$_{58-177}$ than full length, tetrameric M2-1 whereas the affinity of P 90-110 did not significantly alter between the full length and truncated proteins and suggests the interaction of M2-1 with RNA is not confined to the overlapping (with P) binding site at the core domain but may utilise the deleted sites or occur across M2-1 monomers of the M2-1: M2-1 interaction surface of the tetramer.

3.10.3   Crystallisation of the M2-1: RNA complex

It was seen that addition of RNA sequences resulted in protein aggregation (Figure 35) consistent with findings of Leyrat et al (2014) where RNA binding resulted in a concentration dependent aggregation determined by SAXS analysis. The M2-1: RNA protein aggregation determined here (Figure 35) by EM analysis may in part be due to the ability of RNA to interact over multiple tetramers, however aggregation, albeit to a lesser extent can be seen for M2-1: P 90-110, in which a co-crystal structure was determined here (see section 3.9.4).

During co-crystallisation efforts of the M2-1: RNA complex, Leyrat et al, (2015) published their co-crystallisation data of HMPV M2-1 to RNA and DNA. Their research showed no diffraction data for M2-1: RNA with the exception of one short DNA sequence (AGTTA) and adenine-containing di or tri-nucleotides. We synthesised the RNA nucleotide sequence (AGUUA), like the DNA counter-part crystallised by Leyrat et al, (2014) and assessed its binding to HRSV M2-1 by

SEC. No shift in elution peak was seen with the AGUUA RNA seqeunce and was not used for further co-crystal trials. It is theorised that the addition of RNA resulted in aggregation as the A-rich RNA sequences used could 'slide' along the binding surface hypothesised by Leyrat et al, (2015) allowing the RNA to bind over multiple M2-1 tetramers. Thus the M2-1$_{58-177}$ construct in which RNA could only interact with an M2-1 monomer was trialled for RNA co-crystallisation, however no crystal formation was visualised. Efforts to co-crystallise RNA with M2-1 are summarised in Table 10. Two methods not utilised here but could be trialled for future crystal formation experiments are to soak the RNA into M2-1 crystals, this method was successfully trialled for over 1000 fragments and co-crystal structures obtained (see chapter 5) and to aid stabilisation and crystallisation the RNA oligonucleotide could be UV cross-linked to M2-1.

3.10.4 The M2-1 core domain surface binds P 90-110 and induces a helical conformation of the peptide

The known inherent flexibility of large domains of the phosphoproteins across different viral clades likely prevents determination of its full length atomic resolution structure. To date only the CC domain of VSV, RV and SENV (Ding et al., 2000; Tarbouriech et al., 2000; Gérard et al., 2007) and CTD of RV (Mavrakis et al., 2004) have been structurally determined and were not in complex with its binding partners. Here, the HRSV P 90-110 peptide was structurally determined at different snapshots in its binding capacity to M2-1, evident from the differences between monomer density and is the first example of the P protein bound to its binding partner in the Mononegavirales order. The new structure provides insight into the orientation of the M2-1 in relation to its binding partner P (Figure 49) and allows further conclusions to be drawn regarding the role of this interaction in the viral life cycle. The atomic resolution structure of the M2-1: P 90-110 complex exhibits a 1:1 stoichiometry and unequivocally defines the orientation of the P protein in relation to M2-1. Here the first X-ray crystal structure of this interaction shows the C terminus of P extending towards the N terminal face of M2-1 (Figure 49).

Previous work has shown the importance of P 100-120 for M2-1 binding (Mason et al., 2003) and more recently P residues involved in M2-1 interactions (Richard et al., 2015) were finely mapped to residues P 93-110 using site directed mutagenesis. Here 13 P residues 97-109 were resolved in the M2-1 structure to a resolution of 2.6 Å. The structure determined here agrees with the importance of this P region for binding and interaction with the M2-1 protein. The P peptide

bound with similar affinity to monomeric M2-1$_{58-177}$ showing that the ZBD, oligomerisation domain or CTD are not required for binding of the P 90-110 region to M2-1. P 90-110 binds specifically to M2-1 as evident by the GST pull down assay where it did not bind to the GST control and suggest a specific interaction of this small peptide with the M2-1 protein.
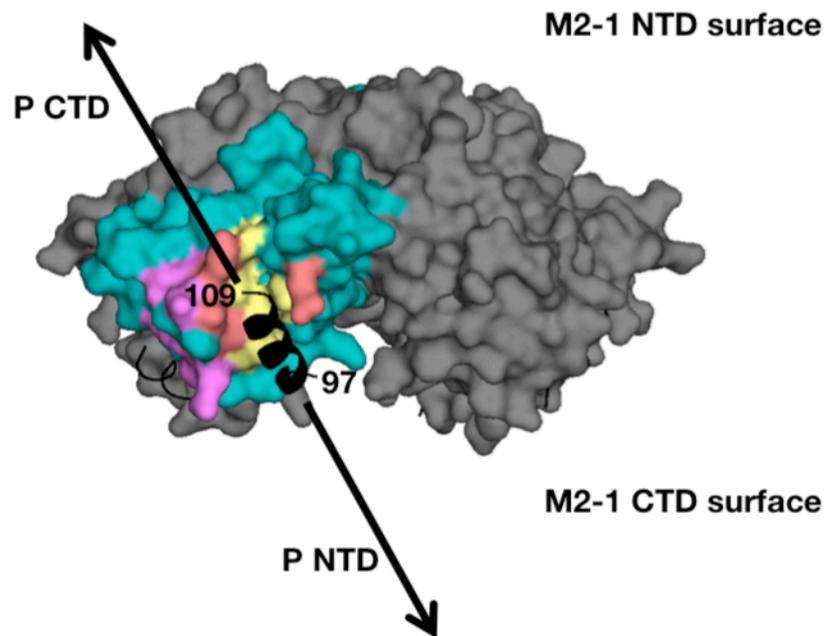
Figure 49 The orientation of M2-1 in relation to P 90-110

A schematic defining the orientation of M2-1 in relation to P 90-110 based on the X-ray crystal structure determined here.

The predicted P binding site highlighted in yellow and the overlapping P and RNA site in coral (Figure 44) was shown to support the structural determination of the P binding site determined here and shows that P 90-110 only binds to the core domain of M2-1. Both the putative and structurally determined interaction residues agree with interactions of M2-1 Arg 126 and Thr 130 with the P protein. The clear side chain density seen for Arg 126 suggests Arg 126 is stable in the structure and involved in critical interactions for phosphoprotein binding. M2-1 residues Tyr 160 and Asn 163 predicted to bind P by SDM are distal to the P 90-110 peptide in the crystal structure and do not form interactions with this peptide. These M2-1 residues decreased binding affinity to P by SDM (Blondot et al., 2012) and may not provide reasoning of a direct interaction with P but possibly of

interactions with other M2-1 residues to stabilise the binding surface. Alternatively, the hypothesised M2-1 residues may bind to a region of P distal to the region (97-109) structurally determined here.

In comparison to the M2-1 backbone, density for the P peptide was weaker and although an alpha helix could be fitted (Figure 42), similarities in functional group structures of individual amino acids proved challenging to fit side chains in parts of the density (Figure 43). Disorder/ order transitions are common in viral replication machinery and P has been shown in other *Paramyxoviridae* family members to fold on binding to its partner N (reviewed by Communie et al., 2014). The P 90-110 density does not form a perfect helical shape for all four peptides in the M2-1 tetramer and is indicative of protein folding. The co-crystal structure here demonstrates that M2-1 can bind P before helical formation and suggests that P is not ordered in solution but folds upon interaction with its binding partner, following the general trend of the *Paramyxoviridae* family. The refined structure had relatively low B factors for 3 of the 4 P peptides compared to the overall M2-1 structure which is indicative of an appropriate model in conjunction with good agreement of the electron density map. The density was too weak for one peptide and only an alanine backbone was fitted which gave a B factor value nearly 3 times larger than the other peptides (Table 12). The higher B factor may be due to partial occupancy which is expected given the weak binding affinity in the micromolar range and for full binding (and low B factor for each peptide bound to a monomer) a higher excess of P 90-110 may be required. It also needs to be noted that the crystal form that allows co-crystallisation of P 90-110 with M2-1 may only select for 3 well occupied P 90-110 bound peptides within the tetrameric structure. Crystals that contained density for the P 90-110 peptide were of a needle-like morphology and crystallised in a different point group to the previously determined protein crystal structure. For future work, crystals could be ruled out of co-crystallisation studies based on their morphology and thus streamlining the efforts of determining a co-crystal structure. Further, the helical P 90-110 peptide co-crystallisation was aided by trimethylamine N-oxide dehydrate (TMAO) in the mother liquor. TMAO is known to aid protein folding and stability and may have aided the peptides ability to form a helical structure to allow determination by X-ray crystallography (Baskakov et al., 1999).

3.10.5   Summary

Given the importance of M2-1s interactions with its bindings partners for anti-termination activity elucidated here and the requirement of an affordable,

efficacious HRVS treatment, subsequent chapters will assess the use of M2-1 as a target for anti-viral drug design through a variety of structural and functional techniques. These results were achieved in part by the development of the improved protein purification methods determined in this chapter.

Chapter 4 Anti-viral drug design- using structure based drug design (SBDD) to target the M2-1 protein of human respiratory syncytial virus (HRSV)

4.1 Introduction

Recent scientific advancements in protein structure determination, paired with new computational tools now allow for the routine use of structure based drug design (SBDD) for rational lead-drug development. This chapter covers the approaches used to identify potential anti-viral compounds effectively targeting M2-1 and discusses the use of M2-1 as a novel drug target through SBDD using *in silico*, functional and structural methods. SBDD is an iterative process that requires a collaborative effort across the diverse disciplines of structural biology, cell biology and medicinal chemistry, and an overview of the SBDD process is summarised in Figure 50.

4.1.1 Protein structure

The primary requirement for SBDD is a high quality structural model of the protein of interest, at a resolution greater than 2.5 Å to allow identification of individual atoms. Structures are primarily obtained by X-ray crystallography and to a lesser extent NMR. In the past 20 years in crystallography alone there has been over a 2000% increase in deposited structures on the RCSB PDB brought about through advances in the automation of X-ray crystallographic methods and synchrotron power. The wealth of available structural information has allowed the use of SBDD to interrogate many possible drug targets.

Protein expression and purification
⬇
Protein structure (X-ray/ NMR)
⬇
Identification of target site
⬇
Computational docking
⬇
Consensus scoring
⬇
Functional assays
⬇
Hit identification
⬇
Lead optimisation
⬇
Ligand bound structure (X-ray/ NMR)
⬇
Pharmacokinetic properties
⬇
***in vivo* testing**
⬇
Clinical trials
⬇
FDA approval
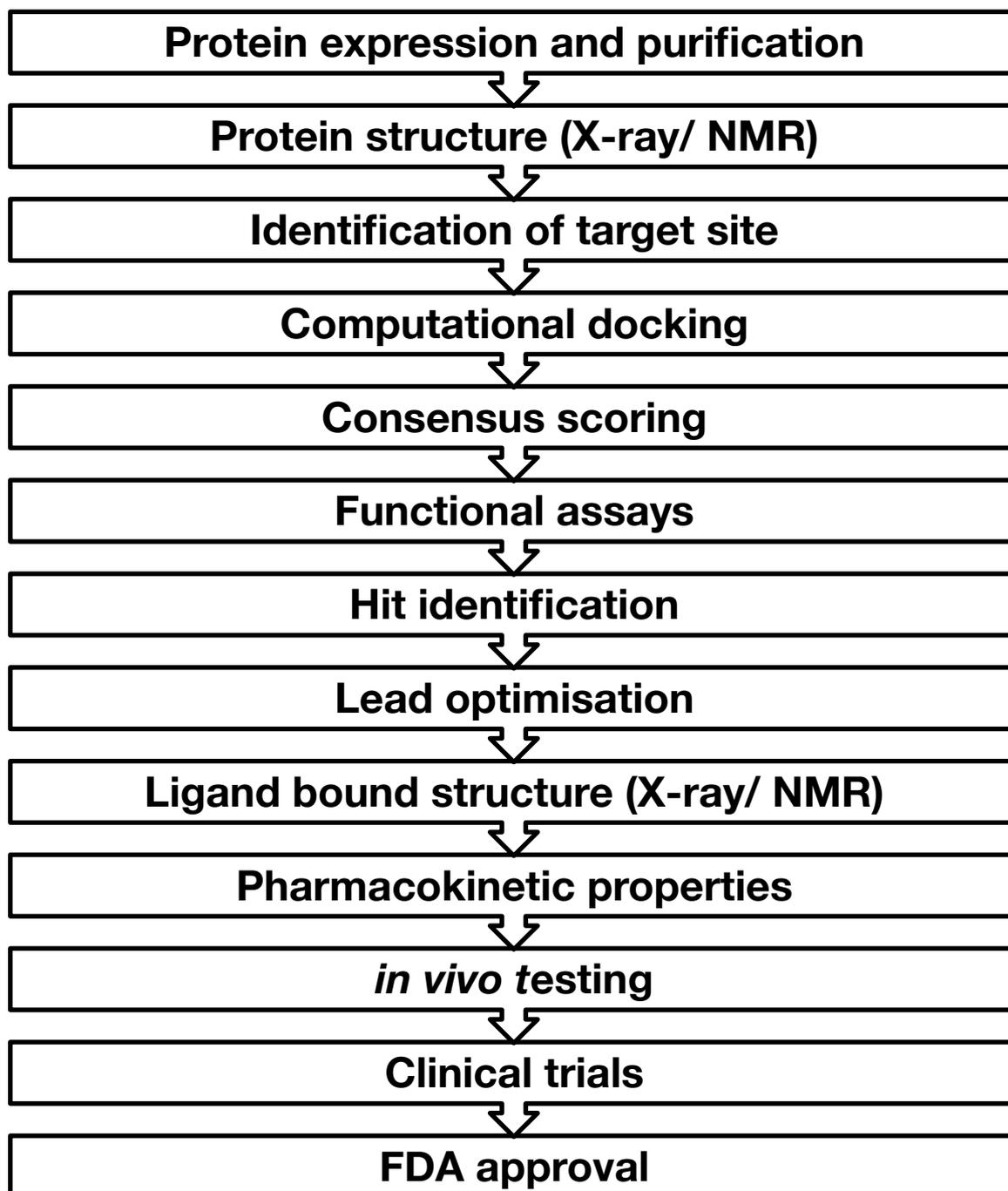
Figure 50 Flowchart of events involved in SBDD

The target protein is identified and the structure determined for computational docking. Chosen compounds are assessed and undergo multiple rounds of SAR, via functional and co-structural studies to find a lead compound with good pharmacokinetic properties for further in vivo assessment.

4.1.2 Virtual compound screening

There are three main computer-aided structural methods utilised to discover a potential pharmacologically active compound for a known protein structure, these are: (1) inspection, using a pre-determined complex-bound structure to identify a potential inhibitor, (2) virtual screening, using a library of known compound structures to dock to the protein within the region of interest, and (3) *de novo* ligand generation in which small molecule fragments are docked and linked *in silico,* this requires downstream synthesis of the 'hit' ligand (Anderson, 2003). There are many programs available for virtual screening (computer-aided method 2) to identify potential bioactive compounds such as Dekois2.0 (Bauer et al., 2013), Dud-E (Mysinger et al., 2012) and eDrug3D (Pihan et al., 2012). The e-Drug3D database uses compounds approved by the FDA and their active metabolites and currently contains 1822 structures in its database (Pihan et al., 2012). The 'Zinc15' database is the primary choice for virtual screening (Irwin & Shoichet, 2005) and the older 'Zinc12' database of compounds was used in the work described in this chapter. The database contains over 90 million commercially available compounds. Screening compounds are available in a 3D-'ready-to-dock' format and are presented in biologically relevant orientations (pose), making it ideal for future drug discovery.

4.1.3 Docking and scoring in virtual screening

The virtual docking process involves the prediction of ligand conformation and pose within a targeted binding site. Each individual pose is then evaluated by considering favourable intermolecular interactions such as hydrogen bonds and hydrophobic contacts. These are ranked in what is known as the scoring function (Kitchen et al., 2004), which predicts the binding affinity of the particular pose and is the optimal way to represent the experimental binding affinities in a high-throughput, low-cost manner. The success of the docking process is improved with multiple protein tertiary structures and in particular a co-complex structure in the targeted binding site, however docking can still be performed in the absence of a co-complex structure.

Many docking programmes are available and each utilises different algorithms to subjectively dock and rank the compounds. Gold (Jones et al., 1997; Verdonk et al., 2003) and FlexX (Rarey et al., 1996) are two of the most popular docking software's (reviewed in Zsoldos 2007), however of specific interest in this

chapter is the use of eHITS (electronic high-throughput screening) and Glide. eHITS allows each core and rotamer (an isomer that differs from other conformers due to restricted rotation around a single bond) to be treated as individual rigid fragments and allows more exhaustive positions of the ligand within the active site (Zsoldos et al., 2007; Simmons et al., 2013). Glide uses a systematic pose search where ligands with a rigid core are chosen for their low energy conformation (Friesner et al., 2004). Combining the scoring function output from multiple docking programmes through consensus scoring allows the reduction of false positives and an improvement in overall hit-rates, a method that is now routinely employed in the search for active ligands from virtual screening (Charifson et al., 1999). An overview of the computational methods involved in SBDD are summarised in Figure 51.

Progress has been limited in instances where receptor flexibility and conformational changes on binding can influence binding interactions, which has left a gap between computational and *in cellulo* binding scores. However, recent work has aimed to increase the accuracy of computationally derived scores by accounting for the receptor's flexibility using multiple conformations (Totrov & Abagyan, 2008). Despite the advancements in high resolution structures, and scoring functions for computer-aided drug design to dismiss biologically irrelevant compounds, there is still a clear need for the subjective assessment of the compound's chemistry by medicinal chemists to bridge the gap of missing dynamic information for the binding site (Ripphausen et al., 2012 & Stumpfe et al., 2012).
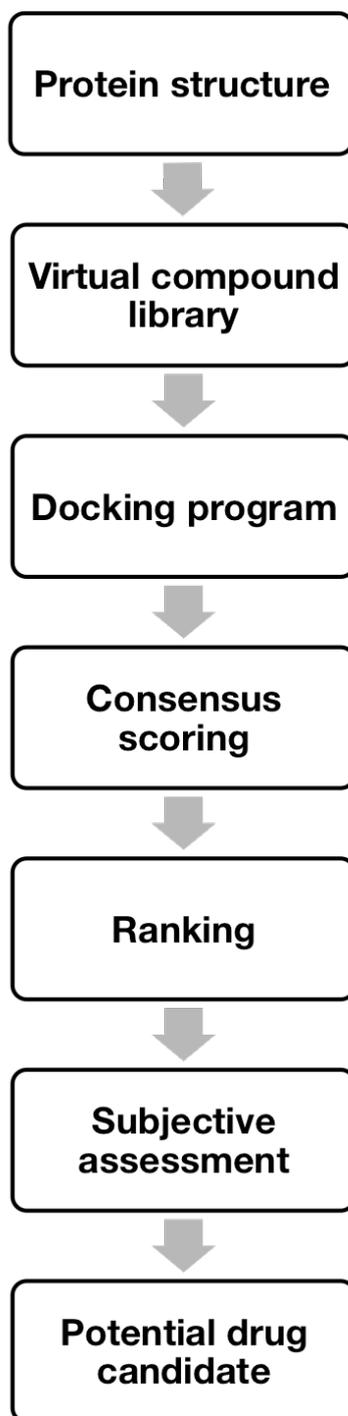
Figure 51 Flowchart of computational methods involved in SBDD

A cavity within the X-ray or NMR protein structure is highlighted for virtual docking of compounds or fragments. Multiple docking programmes allow for consensus scoring and ranking. 'Hit' compounds from the virtual screen are assessed by medicinal chemists and top 'hits' further assessed with biophysical and functional assay.

4.1.4 Beyond virtual screening

After elucidation of potential drug candidates by computational methods, further analysis is required through experimental testing. This iterative process (Figure 52) involves the design, synthesis and testing of compounds for binding or functional activity, developing structure-activity relationships to produce more potent analogues and repeating this iterative cycle until binding affinities within the nanomolar range is reached. A crystal structure of the lead-drug in complex with the protein of interest is highly desirable and allows for future development of a more potent drug candidate by allowing direct visualisation of its mode of binding.

Compounds chosen for further study satisfy Lipinksi's rule of five, which includes; a MW less than 500, a logP (a measure of compounds liphophilicity) of less than 5, less than 5 hydrogen donor atoms and less than 10 hydrogen acceptor atoms. This criteria allows the prediction of the absorption of compounds and overall drug-likeness (Leeson, 2012). Lead compounds are then further analysed computationally and experimentally for their efficacy, pharmacokinetic and ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) properties allowing a prediction of how efficacious the drug candidate will be in human candidates. Drug candidate compounds at this stage are then tested in an *in vivo* model for safety and efficacy before entering phase 1 clinical trials. The process of SBDD and drug development in general is a long, arduous process that takes many years and expertise to result in an approved drug treatment, however to date this approach is still considered to be one of the best methods of rational drug design.
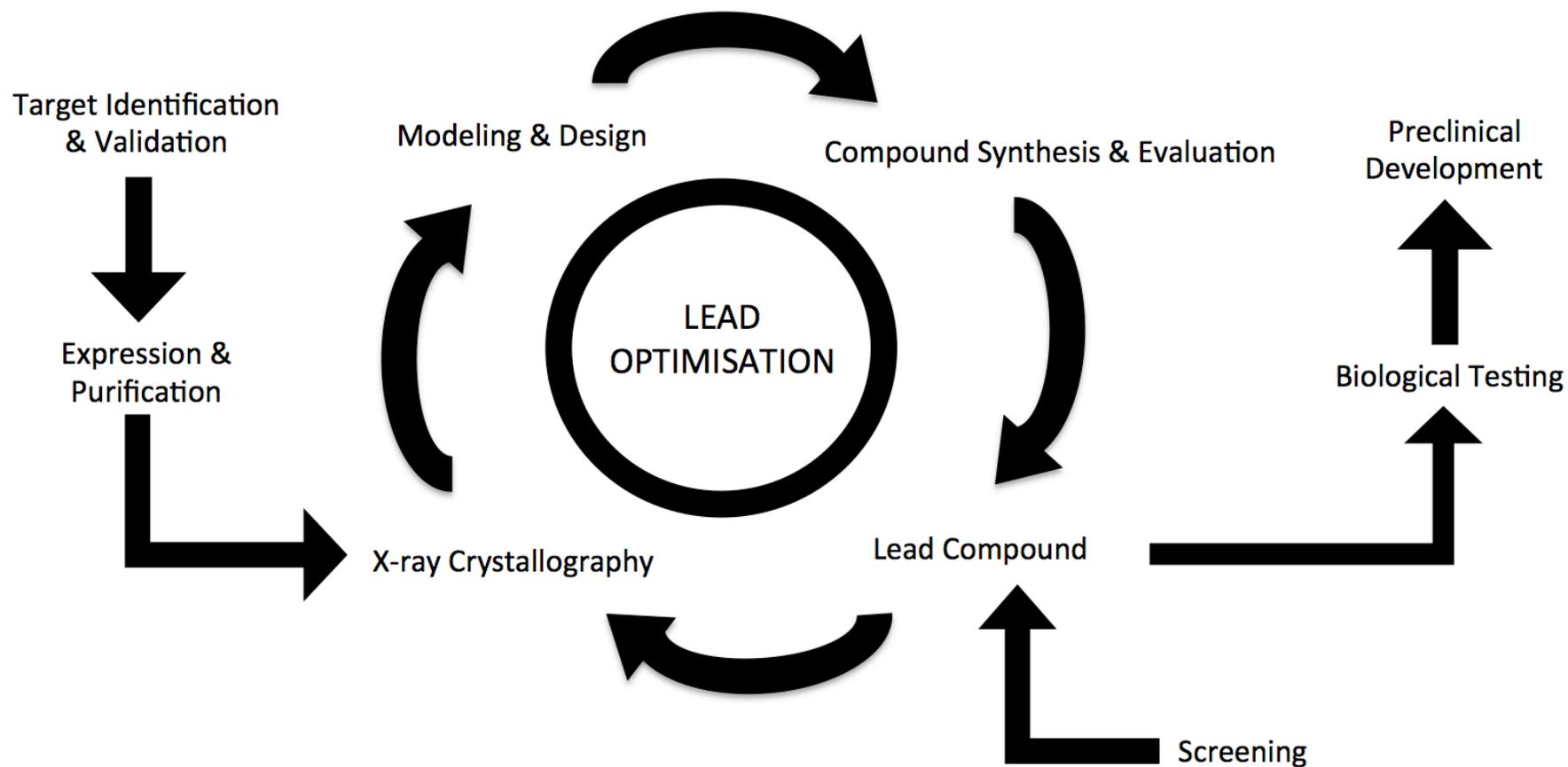
Figure 52 Lead drug compound optimisation flowchart

The protein target is identified and purified for further analysis. The target is crystallised to allow modelling of compounds into a pre-defined binding cavity with top compounds further assayed for their biological activity. Iterative rounds of SAR are performed until a lead compound with nanomolar binding affinity for the target is developed.

4.1.5 Applications of SBDD

Anti-viral compounds zanamivir (Relenza®) and oseltamivir (Tamiflu™) are classic examples of the use of SBDD, for which the rational design of neuraminidase inhibitors yielded treatments for acute illness caused by influenza A and B viruses that are still in use today. Neuraminidases (also known as sialidase) are enzymes present on the influenza virion surface that cleave sialic acid from glycoconjugates on infected cells and have an essential role in viral replication and release of progeny virus for subsequent infection (Palese & Compans, 1976; Lew et al., 2000).

The knowledge of the X-ray structure of the active site of neuraminidase with sialic acid bound (Figure 53) (Burmeister et al., 1992; Varghese et al., 1992) and later, with the inhibitor bound (Kim et al., 1997) allowed the use of SBDD and SAR (structure activity relationship) studies based on a cyclohexene scaffold isostere (a molecule with similarities in chemical properties to the original molecule being studied) of the active site's transition state and resulted in the development of a potent and specific neuraminidase inhibitor, GS 4071, of which the orally bioavailable prodrug form, oseltamivir is the first orally administered influenza treatment approved by the FDA in 1999 under the trade name Tamiflu™ (review of oseltamivir discovery Lew et al., 2000). The X-ray crystal structure of the neuraminidase active site (Burmeister et al., 1992; Varghese et al., 1992) displayed 3 major pockets, with residues involved in binding conserved across all influenza virus A and B strains. X-ray structures showed that the highly polar pocket 1 had non-polar interactions that were key to high affinity binding with the cyclohexene class of inhibitors (oseltamivir). The second pocket was not utilised in sialic acid binding nor in the rational design of a second class of inhibitor zanamivir, a sialic acid analogue and an FDA approved (1999) intranasal administered treatment for influenza (von Itzstein et al., 1993). However, pocket 2 was utilised for the potent inhibitory action of oseltamivir. The third and largest binding pocket had both hydrophobic and hydrophilic residues that have not thus far been utilised for inhibitor design but could be utilised for charged-charged interactions in future drug development. Resistance to oseltamivr has been shown, resulting from a residue mutation (His 274 Tyr) in pocket 1 of the neuraminidase protein that is the active binding site of sialic acid (Le et al., 2005). However, no such resistance mutations for zanamivir have been discovered to date, and thus it still represents a viable treatment for influenza virus. These findings for zanamivir could be due to the drug being structurally highly similar to the natural substrate for neuramindase,

sialic acid, or due to this drug being used to a lesser extent than oseltamivir (reviewed by von Itzstein, 2007) thus not providing the same opportunity for resistance to develop.
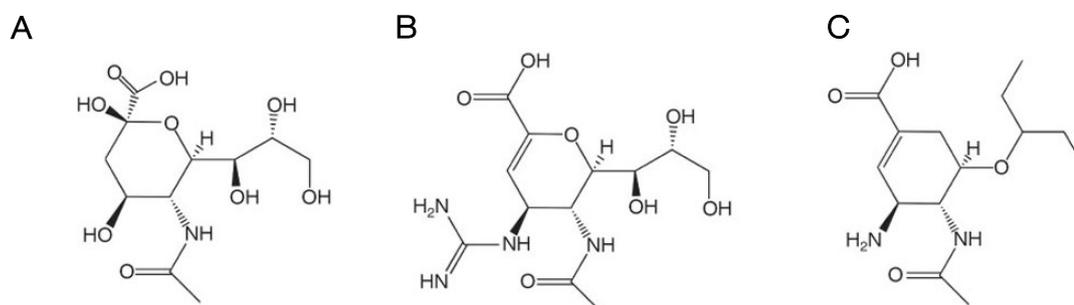
Figure 53 Influenza A and B inhibitors determined by SBDD

> The high-resolution structure of the influenza A virus neuraminidase active site was determined by X-ray crystallography in complex with its binding partner, sialic acid (A). Zanamivir (B) and oseltamivir (C) are approved treatments for influenza A and B virus infection and were both discovered through SBDD campaigns targeting the neuraminidase active site.

Anti-viral treatment for human immunodeficiency virus (HIV) has been achieved using SBDD and has resulted in the development of numerous protease and reverse transcriptase (RT) inhibitors. These inhibitors are combined in a multi-drug cocktail, and are known as highly active antiretroviral therapy (HAART). The HIV aspartyl protease cleaves precursor polyproteins (gag and gag-pol) to the respective functional HIV proteins required for viral infection and has been a successful target for rational drug design. Transition state analogue inhibitors against renin, (a aspartic protease like that of HIV; Allen et al., 1989) in addition to the X-ray structures of the HIV 1 protease (Navia et al., 1989; Lapatto et al., 1989) and the closely related Rous sarcoma virus (Miller et al., 1989) allowed rational SBDD and SAR towards the HIV-1 protease. This led to the FDA approval of saquinavir (Invirase®) in 1995 that is specific to the HIV protease and does not act on the structurally similar human aspartic protease (Roberts et al., 1990). Shortly after, in 1996 ritonavir (Norvir™) was also approved by the FDA as a HIV protease inhibitor (reviewed by Bassetto et al., 2016).

The RT of HIV is of interest as a drug target as it converts single-stranded RNA into double-stranded DNA that can be subsequently integrated into the host's chromosomes and is essential for HIV replication. The RT has 2 functions; as a

DNA polymerase and as an RNase H that selectively degrades the RNA component of a RNA-DNA heteroduplex. Interestingly, the crystal structure of an RNase H inhibitor, dihydroxy benzoyl naphthyl hydrazine (DHBNH) in complex with HIV-1 RT (Himmel et al., 2006) discovered that the inhibitor did not bind to the RNAse H active site but to a novel, allosteric site near the polymerase active site and a non-nucleoside inhibitor binding pocket. Structural analysis combined with further SBDD and SAR of DHBNH (Figure 54 left) resulted in a HIV-1 RT inhibitor (Figure 54 right) that could simultaneously inhibit both the polymerase and RNase H activity of the RT with a IC_{50} of 0.4 µM and 7 µM , respectively for both functions.
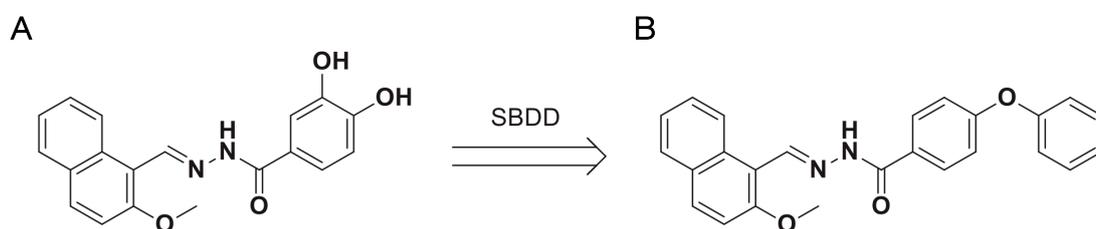
Figure 54 Dual HIV RT inhibitor designed by SBDD

A- compound DHBNH that inhibits RNase H activity of the HIV RT was found by X-ray crystallography to not bind to the RNase H active site; B-SBDD with the known X-ray structure and SAR substitutions of the benzoyl ring resulted in compound 'HIV-1' that inhibits both function of the HIV RT as a DNA polymerase and RNase H (Himmel et al., 2006 and reviewed by Bassetto et al., 2016).

4.1.6 Chapter aims

The M2-1 protein of HRSV is critical for the viral replication cycle and provides a novel target for a significant unmet medical need for an anti-viral treatment for HRSV. The results presented in this chapter are based on initial work performed by Dr Richard Foster, Department of Chemistry, University of Leeds and Public Health England (PHE). The crystal structure (PDB: 4c3b; Tanner et al., 2014) in addition to mutagenesis data for RNA and P binding (Blondot et al., 2012; Tanner et al., 2014) allowed the identification of a defined cavity for virtual docking (chapter 3, section 3.9; Figure 55) of a library of drug-like compounds (from the Zinc12 collection). Using two docking methods (eHITS & GLIDE), a total of 100,000 compounds were ranked based on their predicted Ki (inhibitory constant). A selection of these compounds (1,000) were further assessed for their drug like lead properties by Dr Richard Foster and the 500 top ranked compounds were provided to PHE who screened compound efficacy and toxicity by HRSV plaque reduction assays, in a BSC-1 cell line. In addition to the work of Dr Richard Foster, University of Leeds and PHE, the work performed here was part of collaboration with Dr Roger Taylor, Department of Chemistry, University of Leeds and Redx Pharmaceuticals and the contribution of each of these collaborators is stated. Compound 'hits' (of which there was 44) from data by PHE form the basis of the work presented in this section and illustrate the use of SBDD from hit validation to drug like lead 'hit' compounds using functional and structural methods.
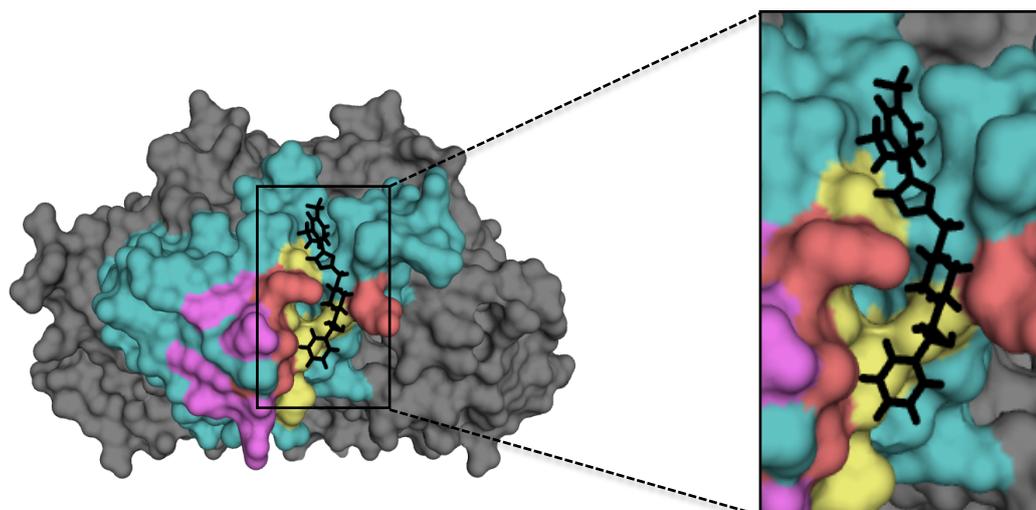
Figure 55 Consensus docking 'hit' to the predefined M2-1 cavity

Representative compound 'hit' docked to the targeted cavity using the GLIDE docking method. Residues involved in P interactions are in yellow, RNA in pink and residues involved in both interactions in coral.

4.2 Results

4.2.1 Mini genome system

The HRSV mini genome system (Figure 56) utilises the BSRT7 mammalian cell line, which are derived from BHK-21 cells but carry an episomally maintained plasmid that allows the expression of bacteriophage T7 DNA dependent RNA polymerase. The HRSV mini genome is a truncated version of the HRSV genome in which all essential genes have been deleted, but which still retains all essential *cis*-acting sequence signals required for genome transcription and replication. Replication and transcription of this mini genome requires the supply from plasmid cDNAs of the minimal HRSV components required for replication and transcription, namely N, P, L and M2-1. The mini-genome itself is expressed from a separate plasmid and consists of 2 genes flanked by an HRSV authentic gene junction and with a leader and trailer sequence. The first gene is an HRSV specific sequence (nucleotides derived from the M and SH genes), and the second gene is eGFP, which is translated to generate an easily detectable product allowing HRSV-specific RNA synthesis activities to be easily assayed.

When all five of these plasmids are transfected into BSRT7 cells, the mini genome RNP assembles, and the polymerase complex can transcribe the first and second genes of the mini genome template, allowing the translation of eGFP that can be detected and quantified using an IncuCyte ZOOM® (a live cell imaging system). The IncuCyte scanned and recorded green fluorescence and phase images every hour, for 29 hours. The recorded data was analysed using the IncuCyte ZOOM® software with the integrated algorithm and displayed as the total number of green fluorescent cells per well. All mini-genome figures displayed the total number of green fluorescent cells per well as a percentage of the total eGFP signal in the presence of all support plasmids, mini-genome reporter and the absence of compound.

In the absence of M2-1, premature termination of the first mRNA transcript occurs, preventing expression of the second gene, and thus no eGFP is detected (Figure 56). As such eGFP expression is dependent on the expression of functional M2-1. The assay was optimised as a medium-throughput assay (12-well format) for use as an initial screening method of 'hit' compounds provided by PHE to assess the ability of these compounds to block or reduce M2-1 function. Compounds (44 from the 500 assessed by PHE) were added in the presence of the transfected N, P, L, M2-1 and the reporter plasmid, with

compounds that affected M2-1 dependent RNA synthesis reducing the eGFP output. Using this system, the expression level of eGFP reporter could be utilised to rank the effectiveness of compounds to inhibit HRSV replication and transcription.
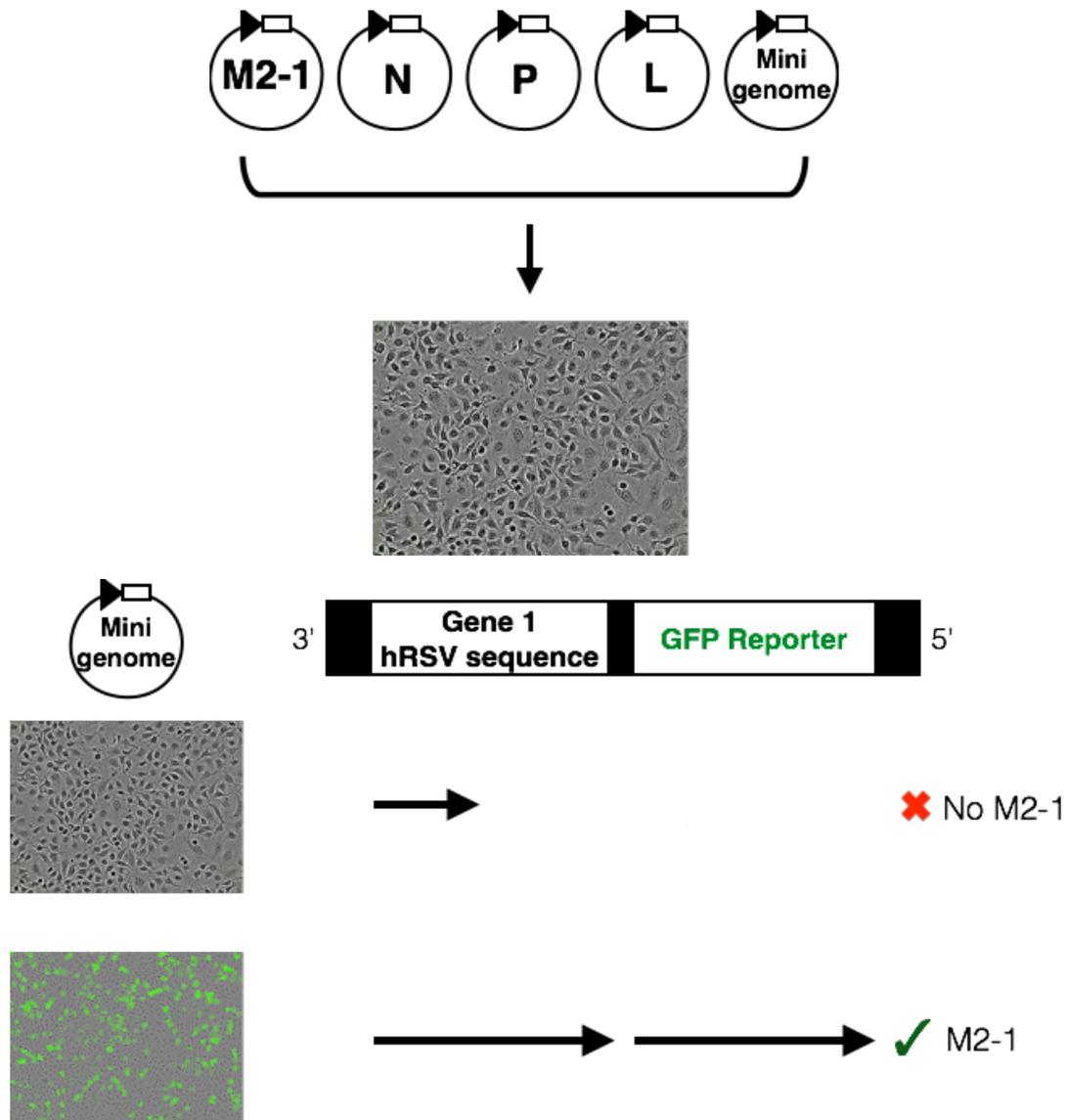
Figure 56 Overview schematic of the HRSV mini genome assay

Plasmids designed to express the minimal virus-encoded components required for HRSV replication and transcription (N, P, L and M2-1) are transfected into BSRT7 cells in addition to a plasmid that expressed a mini genome. The mini-genome 'reporter' gene consists of 2 genes flanked by an authentic HRSV gene junction and with a leader and trailer sequence. The first gene is an HRSV specific sequence (nucleotides from both M and SH genes), whereas the second gene encodes eGFP. In the absence of M2-1, premature termination of mRNA transcription from gene 1 results in no mRNA transcription from gene 2 due to the sequential nature of HRSV transcription. Consequently, no eGFP is expressed. In contrast, in the presence of M2-1, full length mRNA transcription and translation of both gene 1 and 2 results in eGFP expression, that is visualised and quantified using the IncuCyte.

4.2.2 Optimised mini genome system

Over 33 hours, in the presence of compound vehicle (0.2% DMSO) there was no reduction in reporter gene expression (100% eGFP expression= transfection of N, P, L, M2-1 and reporter plasmids) (Figure 57A). In the absence of M2-1 or L plasmids (both essential for RNA transcription) no eGFP production was visualised (Figure 57B).
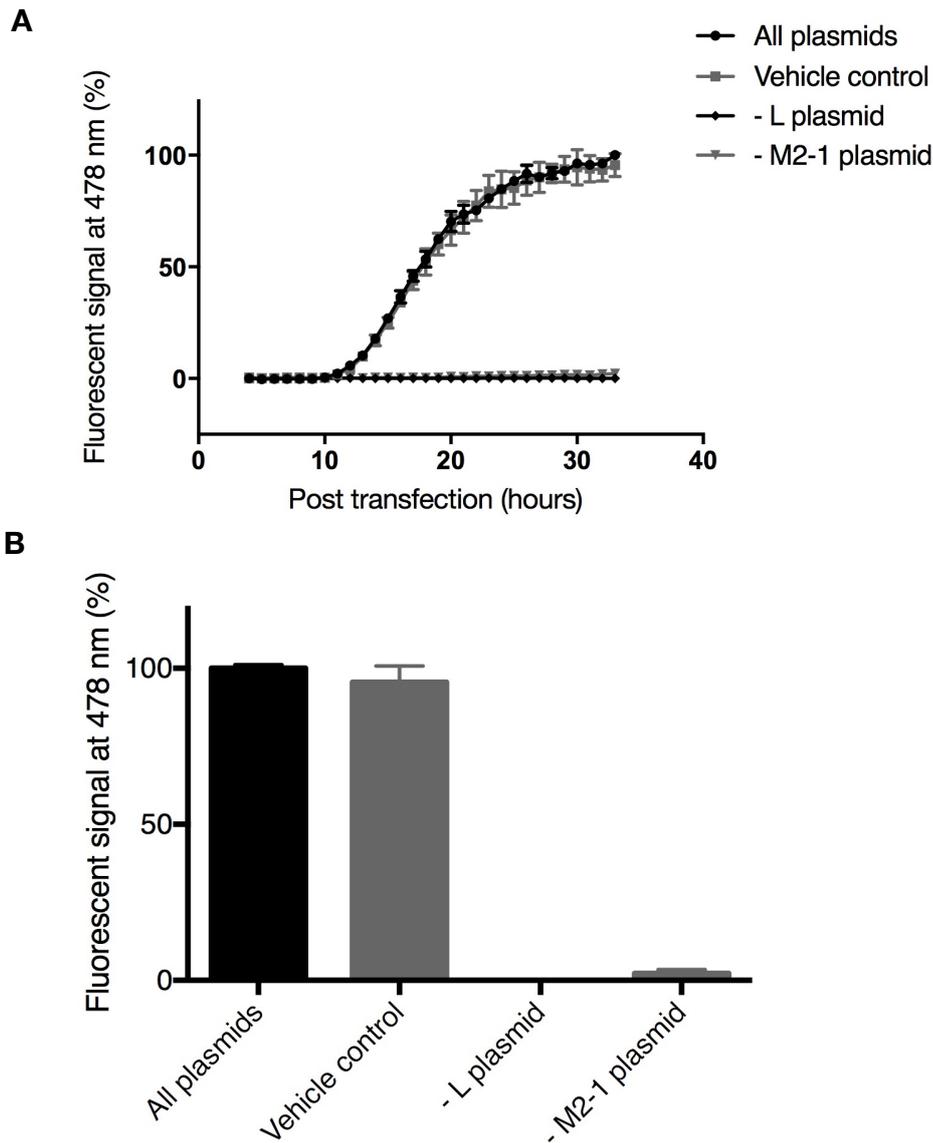
Figure 57 Mini genome assay parameters

BSRT7 cells were transfected with A- all mini genome plasmids ( ), DMSO vehicle (n), - M2-1 (▼) and -L (u); B- representation of A in bar chart form at 33 hours post-transfection. Data points represent n-2 ± SEM with each n number performed in duplicate.

4.2.3 Hit compounds from mini genome screening

In addition to testing compounds in the mini genome assay, compounds were simultaneously assayed for their effect on cell survival using an MTT assay. Compounds that resulted in a cell survival rate below 80% were eliminated from further investigation. Representative cell survival and mini genome data for compounds 9136178, LDS-016067-1 and LDS-017289-1 are shown in Figure 58. Data for all compounds tested can be found in appendix IV. All three representative compounds exhibited a sufficiently low cell toxicity level and were assayed in the mini genome for the initial compound screen. Only two (LDS-016067-1 and LDS-017289-1) of the representative compounds shown here (Figure 58) scored well (over a 50% reduction in eGFP expression) in the mini genome assay. Medicinal chemists assessed compound hits from the mini genome assay (based on over a 50% reduction in eGFP expression) and chose a subset of 13 compounds for continuation based on the mini genome data, the log$P$ values of the compounds potential solubility, a common pharmacophore group and variation in scaffolds to provide various leads for drug development. Mini genome, live virus and cell viability data for all compounds assessed here can be found in appendix IV.
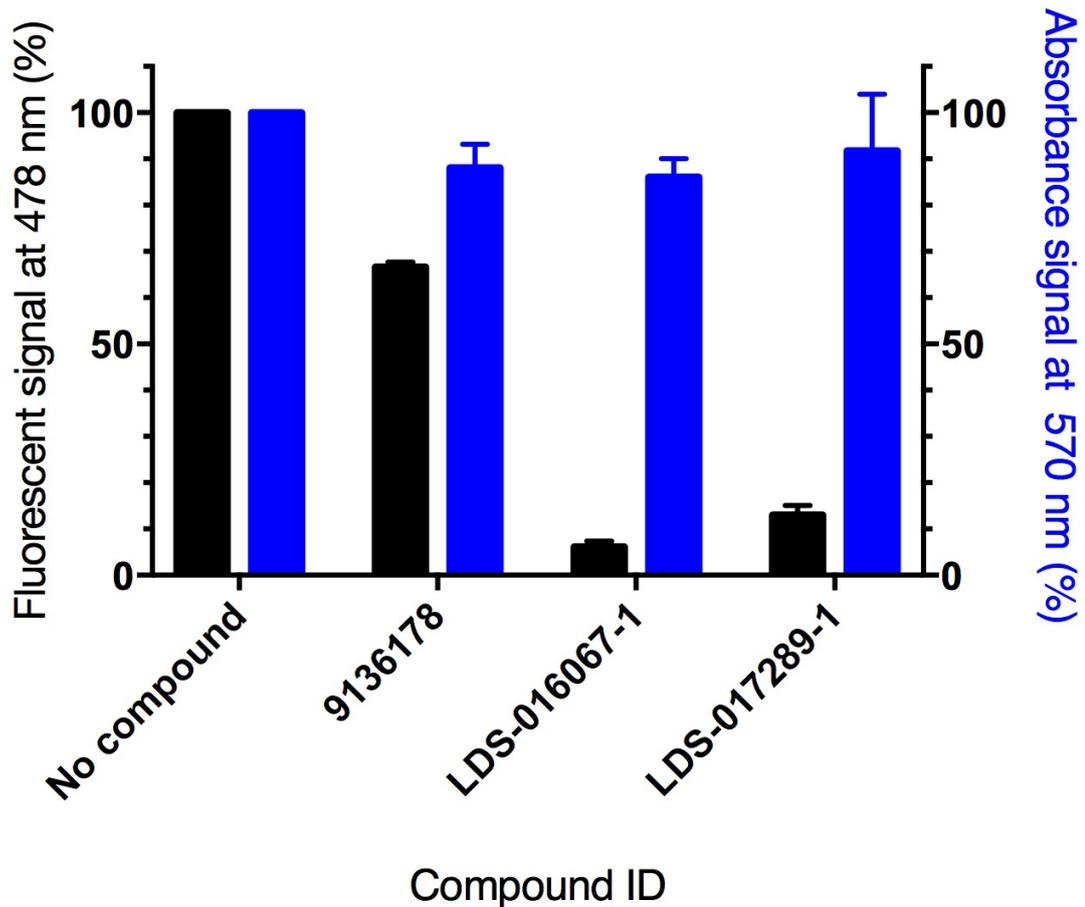
Figure 58 Representative mini genome and cell survival assay data

Representative mini genome data expressed as a percentage of eGFP expression in the absence of compound at 33 hours post-transfection (black bars). Cell survival data expressed as a percentage of absorbance at 570 nm compared to the absence of compound (blue bars). Mini genome data points represent n-2 ± SEM with each n number performed in duplicate and for the cell survival assay data points represent an n-3 ± SEM with each n number performed in duplicate. Data for all compounds assessed can be found in appendix IV.

143

4.2.4 Assessment of hit compounds using an infectious HRSV assay

To further assess compound activity, the ability of compounds to reduce virus gene expression was assayed using autonomously replicating, infectious HRSV, in which the gene for eGFP was inserted into the HRSV genome to act as a reporter for virus gene expression (Figure 59). eGFP expression levels were quantified 24 hours post infection using an IncuCyte ZOOM®. This time point allowed for sufficient levels of eGFP production to be measured before the formation of HRSV-induced cell syncytia that would stop accurate quantification of individual eGFP expressing cells.
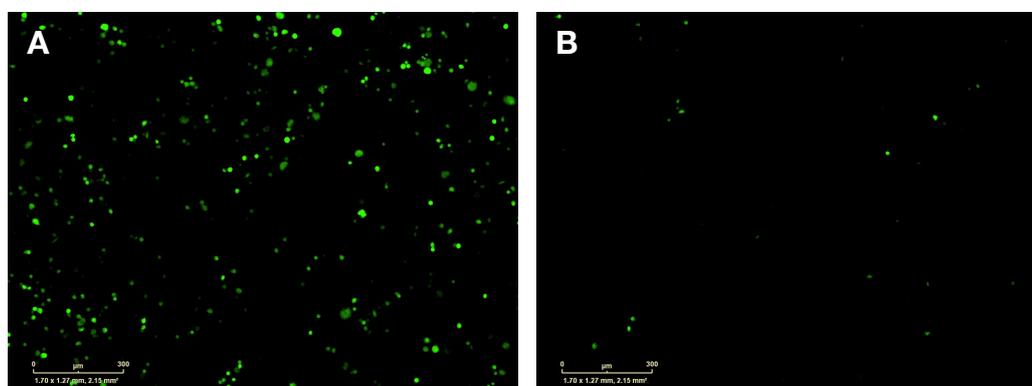
Figure 59 HRSV-eGFP infected BSRT7 cells in the presence and absence of compound LDS-01009-2

Quantification of HRSV-eGFP expressing cells 24 hours post infection. A-HRSV-eGFP infection in the absence of compound; B- in the presence of compound LDS-01009-2 a reduction in eGFP expressing cells was visualised. eGFP expression was visualised and quantified using the IncuCyte.

Compounds (LDS-017332-1, LDS-010009-2, LDS-016034-1, LDS-010149-2, LDS-016077-1 and V028-4618) showed a significant reduction ($p < 0.05$) in eGFP expression compared to the addition of no compound and are illustrated in Figure 60 as blue bars. At the time of this work, published work (Bailly et al., 2016) showed the compound cyclopamine (a smoothened receptor antagonist) to reduce HRSV viral infectivity and specifically, decrease expression levels of the M2-1 protein. To complement this data, here cyclopamine was assessed in the live virus assay at a dose of 20 µM and data showed a significant reduction

144

on viral infectivity in the presence of the compound with no significant reduction in cell survival (Figure 60).
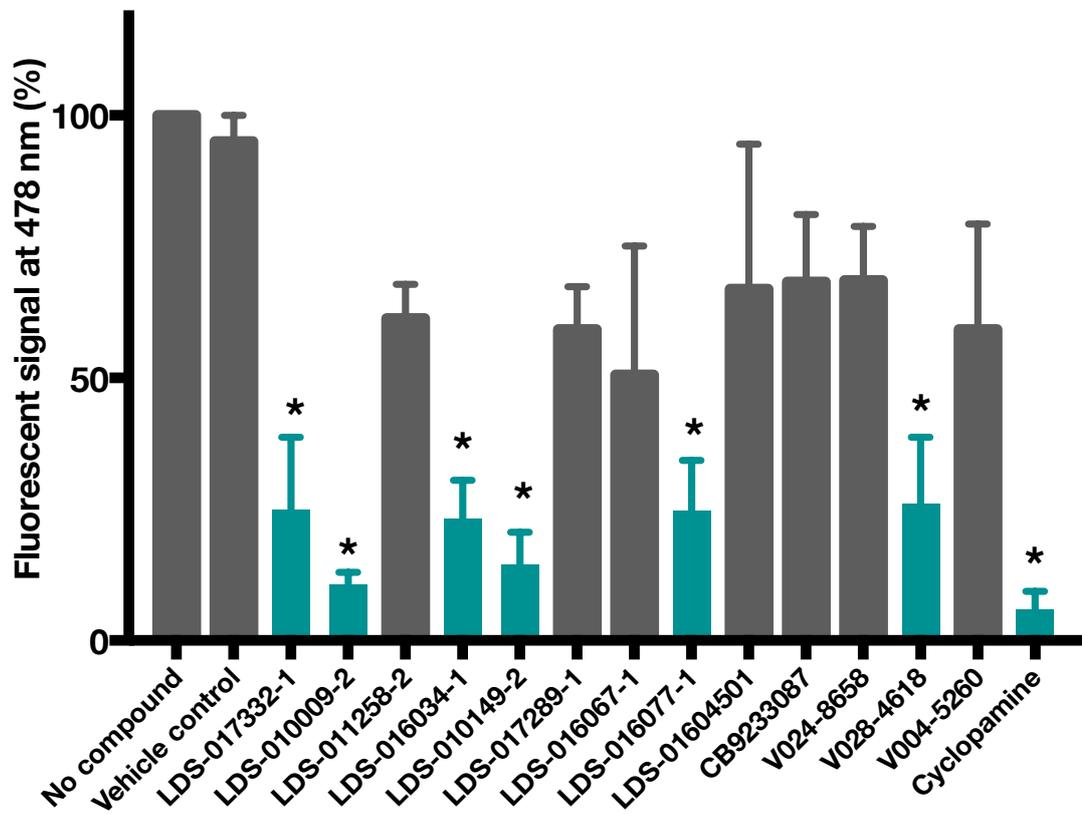
Figure 60 Compound screening with an infectious HRSV-eGFP assay

Infectious virus data expressed as a percentage of eGFP expression in the absence of compound at 24 hours post. Data points were subjected to a one-way ANOVA followed by Tukey posthoc test. A statistically significant difference is indicated by *= $p<0.05$ and bars are shown in blue. Data points represent n-3 ± SEM with each n number performed in triplicate.

4.2.5 Discovery of a drug like lead 'hit' compound

Repeated rounds of analysis were performed with compound analogues to establish a structure-activity relationship (SAR) and were further assessed using the previously determined HRSV assays (mini genome and infectious HRSV) with SAR analysis performed by Dr Roger Taylor, Department of Chemistry, University of Leeds. However, the two compounds chosen for continuation were from the original compounds assessed here (Figure 58, Figure 59, Figure 60 & appendix IV) and not from further SAR analysis. The two compounds LDS-01009-2 and V028-4618 had 92% ± 1.6 and 93% ± 4.6 cell survival, as a % of cells with no compound (Figure 61A). In the mini genome and infectious HRSV-eGFP assays, LDS-01009-2 exhibited 6.0% ± 4.5 and 11% ± 2.4 eGFP expression for the respective assays, whereas compound V028-4618 exhibited 11% ± 2.8 eGFP expression in the mini genome and 26% ± 12.6 in the infectious HRSV-eGFP assay (Figure 61B). The structures of these compounds (Figure 61C) show they are distinct but contain common pharmacophores, that when overlaid have similarly posed lipophilic and polar groups that could adopt a similar binding mode to M2-1 (Figure 62).
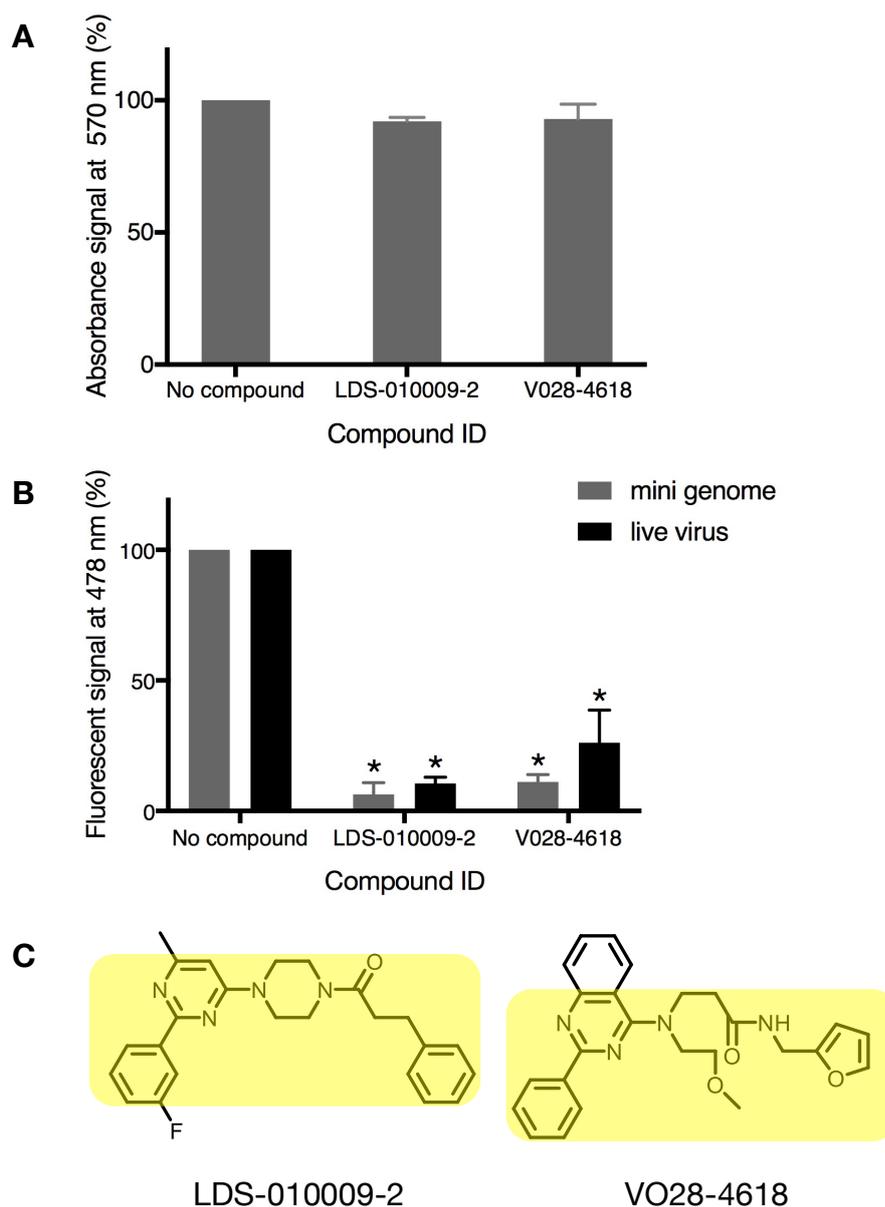
Figure 61 Data analysis of lead compounds LDS-01009-2 & V028-4618

A- cell survival of the two hit compounds in the BSRT7 cell line using an MTT assay; B- eGFP expression in the mini genome and HRSV-eGFP assay and; C- the two compound structures with pharmacophore regions highlighted in yellow. Data points were subjected to a one-way ANOVA followed by Tukey posthoc test. Statistically significant difference is indicated by *= $p < 0.05$. All data is shown ± SEM and each data point was performed in triplicate.
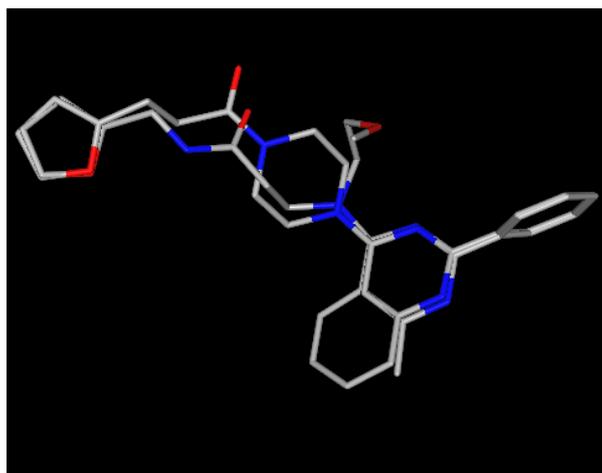
Figure 62 Superimposition of lead compounds LDS-010009-2 and V028-4618

The top two hit compounds were superimposed and showed similar lipophilic and polar groups that could adopt a similar binding mode to M2-1. Image produced by Dr Roger Taylor, Department of Chemistry, University of Leeds. The phenyl group was converted to a furyl to allow superimposition.

Dose-response curves using the infectious HRSV-eGFP assay were attempted to determine IC_{50} values for the top two compounds. eGFP levels were quantified at one time point for varying concentrations of the compounds. Unfortunately, there was too much variation in eGFP expression levels between concentrations to produce a dose-response curve. In addition syncytia formation resulted in large multi-nucleated cells in which fluorescence could not be quantified accurately. The initial assays were screened at one concentration and thus all values are relative to each other on analysis of compounds for continuation.

Compounds LDS-010009-2 and VO28-4618 were profiled (Table 14) for ADMET properties by Redx pharmaceuticals. Both hit compounds were lipophilic, with LDS-010009-2 having a much lower microsomal hepatic intrinsic clearance level. The low solubility of VO28-4618 resulted in difficulties interpreting the cytotoxicity data.

Compound ID	Hu mics Clint (µL/min/ mg)	Hu PPB (% free)	Cytotox HepG2 IC_{50} (µg/ mL)	Measured Log D	ClogD (7.4)	Kinetic solubility (µM)
LDS-010009-2	103.14	0.4	10- 100 µM	> 3.4	4.28	3.63
VO28-4618	> 277	2.39	> 100 µM	NV	4.96	NV

Table 14 Profiling data of hit compounds

Compounds LDS-010009-2 and VO28-4618 were assessed for their ADMET properties by Redx pharmaceuticals.

4.2.6 Further assessment of lead drug candidates

As the target for the selected compounds was the RNA/ P binding cavity of M2-1, we next tested whether binding of these compounds was able to disrupt these critical RNA/ P interactions, and thus suggest direct M2-1 binding as the mode of compound action. The competition assay was performed using a fluorescently labelled version of the P 90-110 peptide, which was the weakest affinity binder of all HRSV-specific ligands thus far assayed. As expected, the P 90-110 peptide was outcompeted by the RNA oligonucleotide, pA13 with an IC_{50} of 1.7 µM ± 0.08 (Figure 63A). However, lead compounds tested here did not diminish anisotropy, suggesting that none were able to outcompete the binding of peptide P 90-110 to M2-1 (Figure 63B). Compounds could not be tested for competition above a concentration of 1 mM due to some evidence of precipitation (10 mM stock concentration).
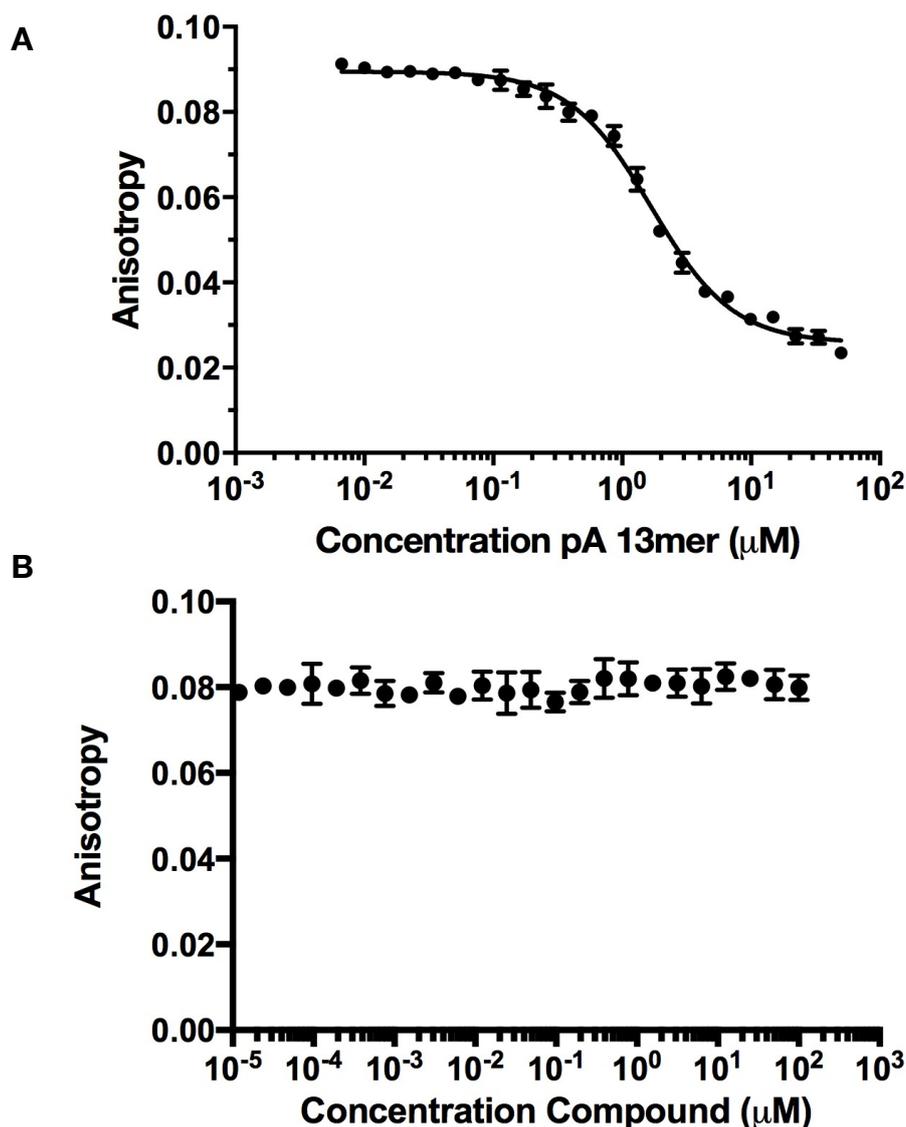
Figure 63 Competition binding assay of compounds with P 90-110 for binding to HRSV M2-1

A- the fluorescently labelled P 90-110 is out-competed by a viral specific oligonucleotide, SH GE (+); B- representative lead drug compound titrated against P 90-110 does not out-compete up to a compound concentration of 1 mM.

A co-crystal structure of a lead-drug candidate with M2-1 would allow rational design and synthesis of compounds based on the compound-protein interactions, thus creating compounds with optimal affinity and inhibitory function and greatly aiding further drug design. Compound LDS-010009-2 was incubated in excess with M2-1 (1.2: 1 ratio of compound: protein). Co-complex crystal trials were performed using commercial

screens with crystals forming after 1-3 days and visualised with a cuboidal morphology (Figure 64). Full data collection was obtained for 3 crystals using the I04-1 beamline, DLS (Table 15). The data was processed using the CCP4 suite of programs and the structure solved using the method of molecular replacement using M2-1 as the model (PDB: 4c3b). The electron density maps ($2F_{obs}-1F_{calc}$ and $1F_{obs}-1F_{calc}$ difference map) revealed no continuous density that could account for the compound. The crystal formed in the same space group as the previously determined M2-1 crystals ($P2_1$).
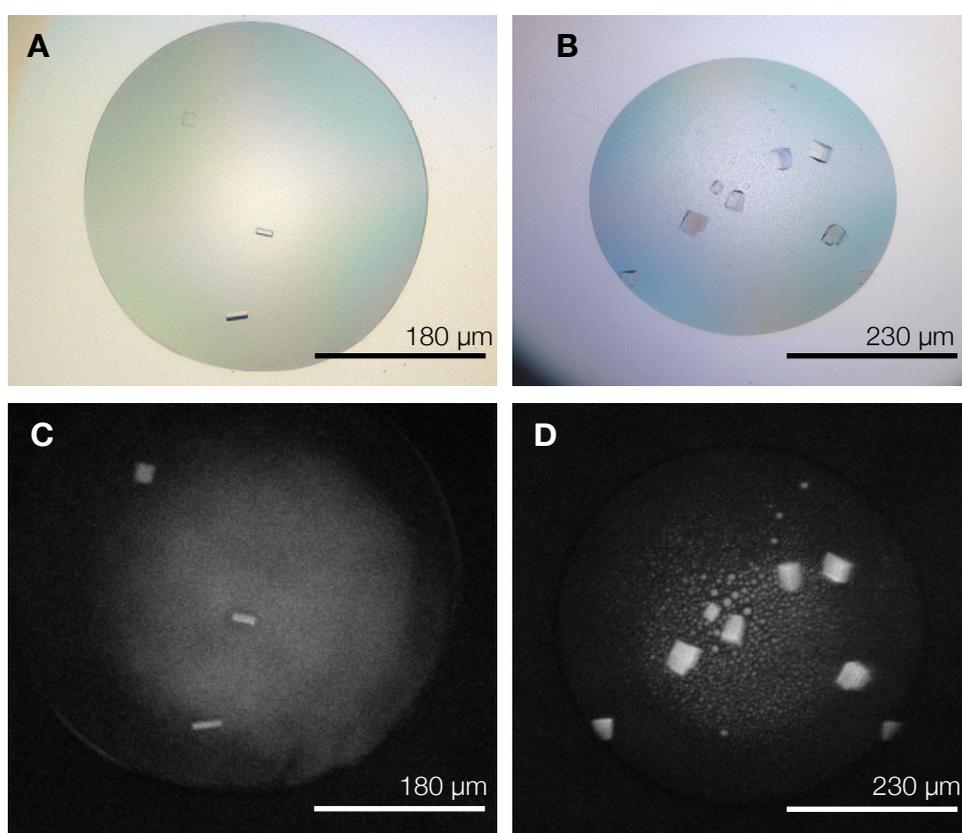
Figure 64 Crystals of the M2-1: LDS-010009-2 compound co-complex

A- condition 0.12 M monosaccharides, 0.1 M buffer system 1, pH 6.6 and 37.5 % MPD-P1K_P3350; B- condition 0.1 M Tris pH 8 and 50% v/v 5/4 PO/OH; A and B- x10 optical zoom; C and D- UV-TPEF imaging.

Data set condition	Resolution (Å)	Space group	Cell constants (a, b, c)	Cell constants (α, β ,γ)
0.1 M Tris pH 8 0.2 50% v/v 5/4 PO/OH	3.28	P1 2$_1$ 1	89.06 42.49 142.40	90 95.01 90
Morpheus Screen 0.09 M halogens 0.1 M buffer system 2 pH 7.5 37.5 % MPD_P1K_P3350	3.10	P1 2$_1$ 1	89.61 43.12 143.39	90 94.40 90
Morpheus Screen 0.12 M monosacharides 0.1 M buffer system 1, pH 6.6 37.5 % MPD-P1K_P3350	3.22	P1 2$_1$ 1	90.06 143.06 143.42	90 95.02 90

Table 15 Data collection of the M2-1: LDS-010009-2 compound complex

Resolution of co-crystal structures and space group data indicate no change in the space group to the determined unbound M2-1 protein. Data collected on the I04-1 beamline, DLS.

4.3 Discussion

4.3.1 The HRSV M2-1 docking site

M2-1 plays a critical role in the HRSV infection cycle and is described as an anti-termination transcription factor that allows full-length transcription of each viral mRNA (Collins et al., 1996), and in addition enhances transcriptional read-through at gene-end/gene-start junctions (Hardy & Wertz 1997). Site directed mutagenesis allowed the identification of residues critical for RNA and P binding (Fearns & Collins, 1999; Blondot et al., 2012; Tanner et al., 2014) and further the co-crystal structure of P90-110 and M2-1 (chapter 3) was in agreement with the putative P binding site and aided our understanding of this RNA and P binding site. The putative RNA and known P binding sites were discovered to form an adjacent and partially overlapping binding site at the M2-1 core domain, and several studies have established that disruption of interactions between M2-1 and these RNA and P binding partners leads to a reduction in its anti-termination function. Residues R126, R151 and K159 that surround the central cavity exhibit favourable electrostatic properties for further drug design. This cavity was chosen for targeting with small molecules via SBDD as these compounds could potentially disrupt RNA/ P binding and thus abolish M2-1 function. Compounds chosen for initial screens were based on their broad range of scaffolds to allow further SAR and analogue design.

Six compounds reduced the ability of the HRSV polymerase complex to perform viral replication and transcription as tested using the mini genome assay, as well as infectious virus gene expression using an infectious virus assay. Despite this encouraging evidence of compound efficacy *in cellulo,* we could not obtain any convincing evidence demonstrating a direct interaction between the compounds and M2-1 using the FA competition binding assay, which has previously been used to investigate the out-competition of P 90-110 with a polyA and HRSV specific SH GE (+) oligonucleotide sequences.

In the competition-binding assay the compounds did not out-compete the M2-1/ P 90-110 interaction, which is the weaker of the two M2-1/ ligand interactions characterised thus far. Intensity readings of the relevant fluorescence levels were taken into account and did not deviate from expected levels. In addition DMSO vehicle controls were tested and did allow competition between P and RNA. Possible reasons for why P 90-110 was not outcompeted include (1) that the compound may bind to a different area of M2-1 not affected by this binding interaction, (2) the compounds were not at a high enough concentration to

outcompete and (3) P and RNA binding to the M2-1 tetramer potentially occurs through a co-operative method (avidity) in which the compound would need to overcome to achieve binding. Further, to determine a direct interaction of the compound and M2-1, co-crystallisation was trialled. M2-1: compound co-crystals formed in 2 space groups that had been previously determined for unbound M2-1 ($P2_1$ and P422). However, no density accounting for the compound could be ascertained in these data sets. Crystals that form in the $P2_1$ and P422 space groups have a distinct cuboidal morphology and suggest a need for specific and conserved crystal contacts for regular lattice and crystal formation that may be disrupted on addition of a compound. In addition to these rationally selected compounds, a protein-protein interaction (PPI) compound library (PPI-NET) and a fragment library (Maybridge) containing over 1000 curated compounds and fragments were screened using the FA competition-binding assay, however no inhibition was seen, and for this reason the data is not shown.

4.3.2 SBDD 'hit' compounds and other determined inhibitors

Two compounds were identified (LDS-010009-2 and VO28-4618) and further analysis demonstrated that these compounds had similarly posed lipophilic and polar groups that could adopt a similar binding mode to M2-1 (Figure 62) and suggest a use of this scaffold for further SAR design with the aid of a co-crystal structure to determine nanomolar-binding affinity of a compound for M2-1. However, to further ascertain an inhibition of eGFP synthesis due to the interaction of the compounds with M2-1, other viral and cellular pathways with the ability to reduce eGFP output in the presence of compound require assessment for their ability to result in off-target effects.

At the time this work was performed, the naturally occurring alkaloid compound cyclopamine and its derivatives were described as being specific inhibitors of HRSV M2-1 (Bailly et al., 2016). The anti-viral effect of cyclopamine was analysed in a live virus replication assay and was found to inhibit HRSV growth to similar levels as the two hit compounds identified here. Similarly to the competition assay results for the hit compounds presented here, cyclopamine also did not inhibit P 90-110 binding to M2-1, and neither could it be shown to interact with the M2-1 core via NMR analysis (Dr Christina Sizun, personal communication). Evidence for a direct effect of cyclopamine on M2-1 function was by identification of a point mutation within M2-1 that provided cyclopamine resistance, and interestingly, this point mutation resided in the M2-1 core domain in the vicinity of the RNA/ P

binding groove (R151). Similarly to the hit compounds determined here, a co-crystal structure of cyclopamine with M2-1 could also not be derived at this time.

In parallel to the work presented in this chapter of determined hit compounds and the assessment of cyclopamine, a previous study identified a compound that resulted in inactivation of HRSV (Boukhvalova et al., 2010). The compound AT-2, a zinc-ejecting compound was presumed to act on M2-1 due to the requirement of each M2-1 monomer to co-ordinate a zinc ion for its function. With this knowledge, Cancellieri et al, (2015) designed and synthesised zinc reacting compounds using SBDD to target the ZBD of the HRSV M2-1 crystal structure (Tanner et al., 2014). The most active compound synthesised, 10e (Figure 65B) had an IC_{50} of 6 µM in a viral replication assay (luciferase expression driven by HRSV was a measure of viral replication) compared to the AT-2 compound (Figure 65A) that had an IC_{50} of 841 µM in the same assay. The data is promising however further work to determine the inhibitors interactions site and indeed the protein target is required for further analogue synthesis. Further, as a drug treatment, there is much debate in the literature of the practicality of using 'zinc ejectors', as zinc finger proteins account for 1% of the human genome and as such are critical for many function such as stabilising the structure of cellular components and in polynuclotide transcription (Frederickson et al., 2000).
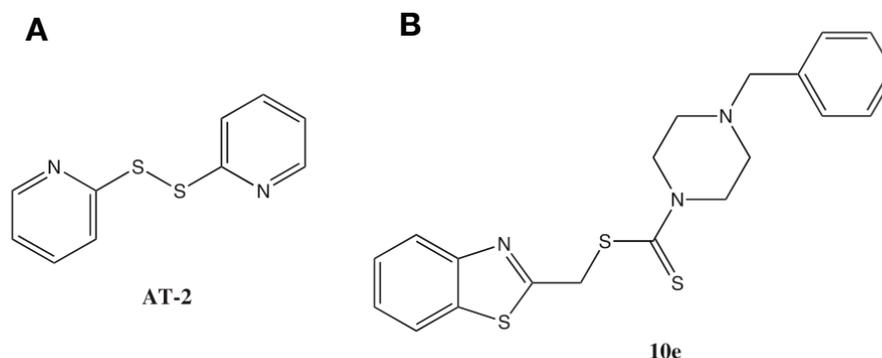
Figure 65 HRSV inhibitors determined by SBDD of the M2-1 ZBD

A- compound AT-2, a zinc-ejecting compound reduced HRSV infectivity and was assumed to inhibit via M2-1 binding, as this protein required zinc co-ordination for its activity; B- SBDD focussed on the ZBD of M2-1 using the crystals structure (PDB: 4c3b) and 'hit' compound 10e was determined to potently inhibit HRSV replication (Boukhvalova et al., 2010; Cancielleri et al., 2015).

4.3.3 Summary

This chapter has identified compounds targeted to the RNA/ P binding cavity of M2-1 using SBDD. These compounds have been shown to reduce viral replication/ transcription (mini genome) and HRSV infectivity (infectious virus assay). However, further work must ascertain the mode of binding for further structure-activity relationship work. As a co-crystal structure of the hit compounds bound to M2-1 could not be ascertained at this time to aid SAR and hit-lead development, a technique was trialled that allowed direct visualisation of fragments (the building blocks of compounds) bound to M2-1 using X-ray crystallography as the method. The new facility, known as XChem, allowed rapid screening of 1000s fragments to determine fragment hits bound to the M2-1 protein in a high-throughput manner and is discussed in chapter 5 as an alternative method to determine anti-viral compounds targeted to the M2-1 protein of HRSV.

Chapter 5 Anti-viral drug design- using fragment based drug design (FBDD) to target the M2-1 protein of human respiratory syncytial virus (HRSV)

5.1 Introduction

In terms of drug discovery a fragment can be described as a building block for the synthesis of compounds. The method of fragment-based drug discovery by X-ray crystallography has been utilised from the early 2000s but its broader application has been severely limited due to the length of time required to produce the many high quality crystals required for screening (often several thousand), as well as practical limitations in the collection of X-ray diffraction data (Nienaber et al., 2000). Recent advances in synchrotron power and automation of the crystallography process has resulted in the development of a new facility at the DLS, termed XChem that allows routine fragment screening for previously un-targetable proteins. The high-throughput XChem facility allows 1000s of fragments to be screened in combination with a potential protein target for structure determination, and allows hit-to-lead compound optimisation to be done in a matter of days. The XChem facility has been operational since April 2015 and is still being reviewed for its effectiveness as a method to aid the drug discovery process.

5.1.1 FBDD by X-ray crystallography using the XChem facility

Automation of protein purification, crystal trials and imaging has streamlined the crystallography process and made the technique accessible to a wider audience. Fragment screening by X-ray crystallography has seen a number of notable software and equipment advances that have been utilised for the XChem facility and a selection of these are listed here:

(1) TexRankE- is a program that uses previously defined images and allows rapid identification of crystals for ranking and to define a target x, y location for the fragment soak (Ng et al., 2014).

 (2) ECHO- the acoustic liquid handler uses ultrasonic pulses to eject rapidly and accurately fragment and cryo-protectant droplets. The ECHO allows 100s of crystals to be soaked with fragments in a few minutes (Collins et al., 2017).

(3) Shifter- was designed for the XChem facility by the Structural Genomics Consortium (SGC), University of Oxford and automatically tracks fragment soaked crystals to be harvested using data from the TexRankE software whilst keeping the rest of the plate sealed to avoid evaporation. In house, to harvest crystals, one puck (16 crystals) takes on average one hour, the shifter allows sub-10 second harvesting times per crystal, routinely allowing one puck to be filled within 10 minutes.

(4) BART robot on the I04-1 beamline- allows data collection in an automated unattended mode due to the large capacity of the BART robot and the automated centring of loops. A crystal can be mounted and a data set collected in under 2 minutes.

(5) Pan-Dataset Density Analysis (PanDDA) and XChemExplorer (XCE) (Krojer et al., 2016) software packages- were also produced by the SGC to amalgamate 1000s of data sets for their rapid evaluation and analysis. The recently created PanDDA software models fragments with partial occupancy using an ensemble of bound and unbound states for enhanced refinement and to allow better density for the ligand to be fitted (Pearce et al., 2017).

In combination, the software and equipment mentioned above allows identification, ranking, targeting, soaking, harvesting and data collection of 1000s of crystals per day. A process that was once only a sought-after vision is now a reality that can be achieved in a short time frame to routinely find fragment binders for lead-drug optimisation.

5.1.2 Advantages of FBDD by X-ray crystallography

The use of FBDD using X-ray crystallography as a primary hit identification method allows the collection of high resolution structural information of the fragment: protein interaction, which can then be used as the basis of further synthetic chemistry to develop fragments into potential lead compounds. In addition, the poised fragment library used within the XChem screening platform is commercially available and can be readily re-purchased for further analysis of hit binders. Previous FBDD by X-ray crystallography used fragment cocktails to allow a relative level of high-throughput (Nienaber et al., 2000; Hartshorn et al., 2004). In contrast, the XChem facility uses single fragment soaks per crystal, resulting in identification of the correct ligand density and a reduced number of damaged crystals from the toxicity of one fragment in a cocktail mix. The low affinity binding of fragments make them difficult to detect in bioassay based

screening methods (Hartshorn et al., 2004) but are easily detected using X-ray crystallography with few false positives and allows for potent lead compounds development due to knowledge of the binding site. Protein-protein interactions (PPI) as seen between M2-1 and the P protein of HRSV are renowned for their difficulty as drug targets. PPI interactions generally have shallow binding pockets with a low binding affinity and as such result in difficulty in detecting hit binders in a bioassay-based screen. X-ray crystallography bypasses this requirement and the XChem facility may now become routine in assaying protein-fragment interactions.

5.1.3 Applications of FBDD using structural methods

Both fragment and compound libraries have been utilised in drug discovery for decades. However, libraries are usually subjected to an initial screening protocol based on a functional assay, with high resolution structural information of the fragment: protein complex only being determined for lead hits due to the demanding requirements for appropriate crystals, and the time intensive nature of structural methods that must be employed.

Fragment screening in combination with structural methods provides an initial screen to determine fragment binders that may not be detected in functional assays, and this protocol has been successfully utilised for inhibitor development of the BACE-1 protein. The BACE-1 protein cleaves the amyloid precursor protein to form the distinct amyloid beta plaques present in the central nervous system of patients suffering from Alzheimer's disease (Luo et al., 2001). While previous interacting compounds have been identified, these have failed to inhibit BACE-1 *in vivo* due to their inability to cross the blood-brain barrier. A custom fragment library was screened against the BACE-1 protein and chemical shifts representing binding were determined using NMR. Analogues of these hit fragments were further screened using a functional assays and a 36-fold increase in binding compared to initial NMR identified hits was obtained (Wang et al., 2010). Two distinct classes of inhibitors were identified by this method with μM and nM affinity respectively (Wang et al., 2010; Zhu et al., 2010). These were subsequently developed into lead compounds that are currently undergoing clinical trials (Stamford & Strickland, 2013).

Of the 7 proteins analysed using the XChem facility since it became operational in April 2015, only one protein yielded no fragment hits (Dr Patrick Collins, personal communication). Due to the infancy of the XChem facility there has

been a lag in publication of experimental findings and as such there is only one paper describing the effectiveness of this screening facility. Here, an atypical non-asparagine bromodomain with no identified inhibitors was assayed using the XChem facility and the resulting work illustrates the first reported inhibitors (binding within the mM affinity range) of this atypical protein family linked to disease progression in diabetes and cancers (Cox et al., 2016).

5.1.4 Chapter aims

Due to the lack of high resolution structural information of M2-1 in complex with hit compounds identified in chapter 4 (SBDD), an alternative method was examined that would allow structural determination of protein: fragment complexes with downstream evaluation of the activity of hit fragments using the multiple functional assays that are currently available for HRSV (chapter 4; mini genome, infectious virus, FA assay). Firstly, this chapter discusses the identification of a crystal condition that allowed the production of 1000s of crystals with re-producible diffraction of at least 3 Å resolution. Once a suitable condition was determined the work described in this chapter focused on the identification of novel fragment binders to the HRSV M2-1 protein by screening fragment libraries using X-ray crystallography at the recently constructed XChem facility at the DLS. Due to the infancy of the XChem facility in providing large-scale fragment screening the optimisation required for efficient running of the facility is also described to allow efficient and routine screening for further users.

5.2 Results

5.2.1 Optimisation of crystals for the XChem facility

5.2.1.1 In house crystal optimisation

M2-1 protein crystals have previously been obtained in various chemical conditions (see appendix III), however reproducibility of the X-ray diffraction has not been assessed. Rarely in crystallography is there a requirement to determine a condition that reliably and reproducibly diffracts however, this was a pre-requisite for the facility to perform the fragment screen by X-ray crystallography. For the XChem facility, 1000s of M2-1 protein crystals were required that diffracted to below 3 Å resolution. Crystal conditions that had previously resulted in collection of diffraction data with resolution of at least 3 Å were tested for reproducibility in SWISS CI plates at 18°C and at room temperature, mimicking conditions of the XChem facility (Figure 66, Table 16). The M2-1 protein was purified using the improved purification method discussed in chapter 2 section 2.2.2 and chapter 3, section 3.6.2 for reproducibility.

Conditions were chosen for continuation if they reproducibly produced crystals along that row. Drop 1 of the SWISS CI plate, on average crystallised less readily thus to save reagents, only drops 2 and 3 were used of the 3 drop SWISS CI plates. These specific plates were used for in-house optimisation as they were consistent with equipment at the XChem. A fine screen of 12 conditions, altering pH and PEG was performed to determine protein crystals that may be more amenable for crystal harvesting. Representations of the optimisation conditions are displayed in Table 17.
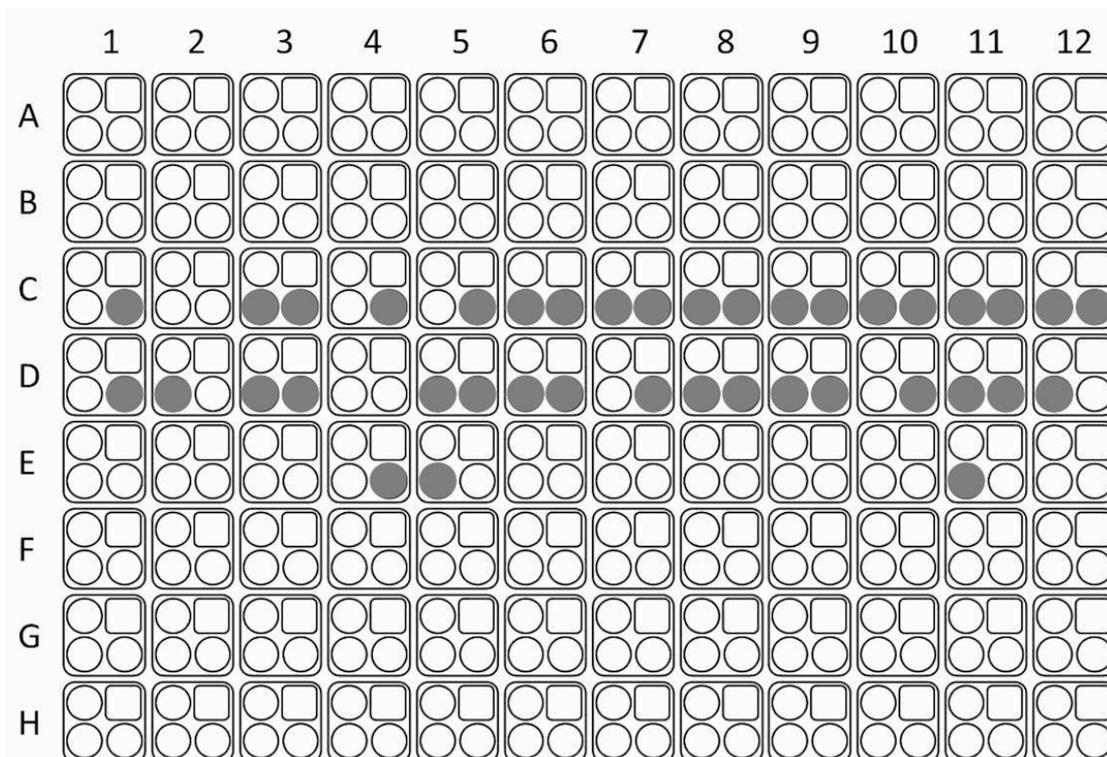
Figure 66 Representative plate studying reproducible protein crystal formation

3 drop SWISS CI plate; square= mother liquor well, circle= drops 1, 2 and 3. Each crystal condition was set up across a row 1-12 in drops 1-3 of the SWISS CI plates and placed in the RockMaker Imager. Drops were scored (grey circle) if they contained protein crystals that could be harvested. Rows A-E contained protein. Mother liquors are stated in Table 16 with screen and well stated here; A- Morpheus F5; B- Morpheus G5; C- Morpheus G9; D- Morpheus H1; E- PACT F12. Reproducible conditions were further fine screened and tested for diffraction in at least 3 crystals from one condition.

Screen & Well	Salt	Buffer System	Precipitant
PACT B1	0.1 M MIB pH 4		25% PEG 1500
PACT B10	0.2 M MgCl$_2$	0.1 M MES pH 6	20% PEG 6000
PACT D1	0.1 M MMT pH 4		25% PEG 1500
PACT D6	0.1 M MMT pH 9		25% PEG 1500
PACT D11	0.2 M CaCl$_2$	0.1 M tris pH 8	20% PEG 6000
PACT E4	0.2 M potassium		20% PEG 3350
PACT F1	0.2 M sodium fluoride	0.1 M bis-tris	20% PEG 3350
PACT F11	0.2 M sodium citrate	0.1 M bis-tris	20% PEG 3350
PACT F12	0.2 M sodium	0.1 M bis-tris	20% PEG 3350
Morpheus A1	Divalents	BS1 pH 6.5	P550MME_P20K
Morpheus A5	Divalents	BS2 pH 7.5	P550MME_P20K
Morpheus B4	Halogens	BS1 pH 6.5	MPD_P1K_P3350
Morpheus B5	Halogens	BS2 pH 7.5	P550MME_P20K
Morpheus C1	NPS	BS1 pH 6.5	P550MME_P20K
Morpheus D1	Alcohols	BS1 pH 6.5	P550MME_P20K
Morpheus E1	Ethylene Glycols	BS1 pH 6.5	P550MME_P20K
Morpheus F1	Monosaccharides	BS1 pH 6.5	P550MME_P20K
Morpheus F5	Monosaccharides	BS2 pH 7.5	P550MME_P20K
Morpheus G5	Carboxylic acids	BS2 pH 7.5	P550MME_P20K
Morpheus G9	Carboxylic acids	BS3 pH 8.5	P550MME_P20K
Morpheus H1	Amino acids	BS1 pH 6.5	P550MME_P20K

Table 16 Commercial screen conditions to assess crystal reproducibility

Each condition was set up in one row of a 3-drop SWISS CI plate and protein crystal reproducibility determined as started in Figure 66.

Morpheus	1	2	3	4	5	6	7	8	9	10	11	12
0.1 M Carboxylic acids												
0.1 ml BS3 (pH)	8.1	8.5	8.9	8.1	8.5	8.9	8.1	8.5	8.9	8.1	8.5	8.9
P550MME_P20K (ml)	0.4	0.4	0.4	0.5	0.5	0.5	0.6	0.6	0.6	0.7	0.7	0.7
H_2O (ml)	0.4	0.4	0.4	0.3	0.3	0.3	0.2	0.2	0.2	0.1	0.1	0.1

PACT	1	2	3	4	5	6	7	8	9	10	11	12
0.1 M MMT (pH)	4	4.5	5	4	4.5	5	4	4.5	5	4	4.5	5
PEG 1500 (%)	20	20	20	25	25	25	27	27	27	30	30	30
H_2O (ml)	0.7	0.7	0.7	0.65	0.65	0.65	0.63	0.63	0.63	0.6	0.6	0.6

Table 17 Fine screen optimisation conditions to assess crystal reproducibility

Optimisation of pH and PEG concentration for a representative Morpheus and PACT commercial screen condition.

Fine screen hits from one well (i.e. formation of crystals in both assessed drops) were then further screened across one row (24 drops) of a SWISS CI plate to further assess reproducibility. Crystals took 0-3 days to form and conditions that reproducibly formed crystals at this stage were selected for assessment of their X-ray diffraction at the DLS. Crystals that could not easily be harvested were not chosen for continuation due to the requirement to harvest over 1000 crystals when using the XChem facility. Only one condition highlighted in Table 18 (green) contained crystals that formed reproducibly and diffracted to 3 Å or better.

Diffraction assessment conditions
0.1 M Divalents, BS1 pH 6.9 0.4 ml, P550MME_P20K
0.1 M Divalents, BS1 pH 6.1 0.5 ml, P550MME_P20K
0.1 M Carboxylic acids, BS3 pH 8.5, 0.4 ml PEG550MME_PEG20K
MIB pH 4, 20% PEG 1500
MIB pH 4.5, 20% PEG 1500
MIB pH 4.5, 27% PEG 1500
MIB pH 5, 27% PEG 1500

Table 18 Fine screen optimisation conditions to assess diffraction of the protein crystals

Optimised pH and PEG conditions were assessed across one row of a SWISS CI plate. Conditions above resulted in reproducible crystal formation and were subsequently harvested and cryo-cooled for diffraction analysis. Green highlights the condition with < 3 Å resolution. Three diffraction images were recorder per crystal at 0, 90 and 180O, and high-resolution data collected for at least 3 crystals per condition.

5.2.1.2 XChem facility- crystal screening optimisation

Events required for the XChem fragment screening to allow identification of fragments bound to a protein of interest are summarised in Figure 67. Further optimisation was required at the DLS to determine the concentration of DMSO (fragment vehicle) and cryo-protectant required for the XChem screening process that still allowed the collection of < 3 Å diffraction for all the crystals (1000s) screened. To assess diffraction resolution at varying DMSO and cryo-protectant concentrations, the ECHO acoustic liquid handling dispenser was used to dispense varying concentrations of DMSO (0- 20%) and cryo-protectant mix (0- 20%) to pre-defined (x, y) plate co-ordinates (TexRankE).
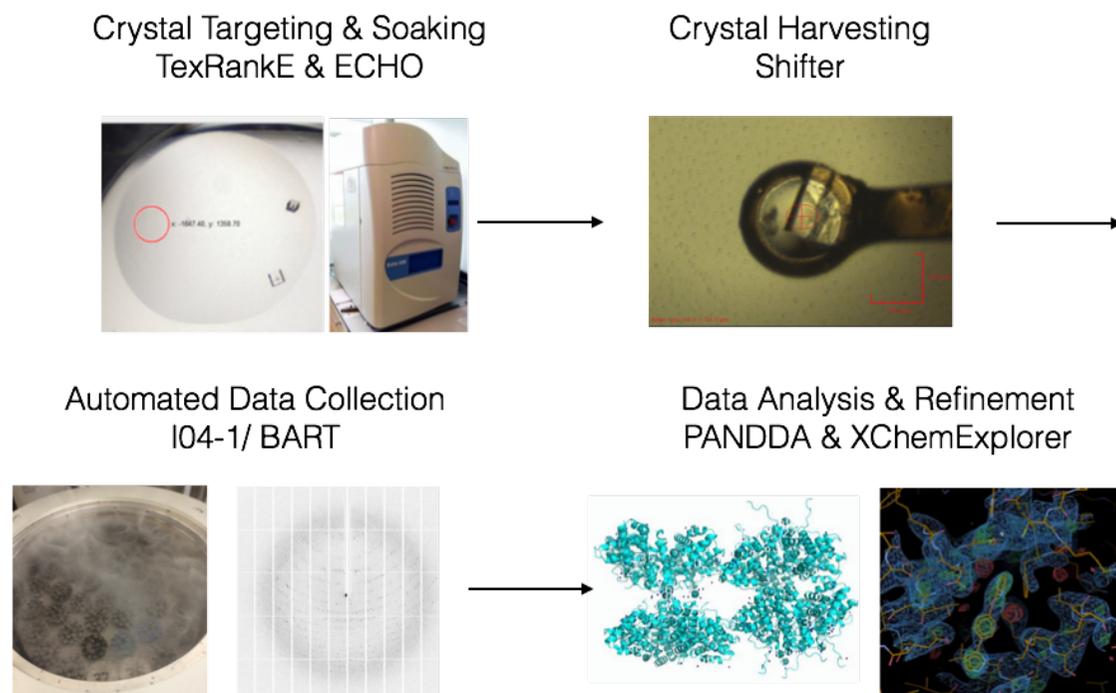
Figure 67 XChem fragment screening methodology

Crystals were imaged using the RockMaker Imager and targeted with an (x, y) co-ordinate position in the drop, away from the crystal using TexRankE. Crystals were soaked using the ECHO acoustic liquid dispenser to the previously marked (x, y) co-ordinates. Crystals were harvested and cryo-cooled using auto-centring loops and using a motorised plate mover known as 'the Shifter' to allow crystal harvesting in sub-10 seconds. Data collection was fully automated on the I04-1 beamline with use of automated centring loops and the BART robot, allowing 34 pucks containing 16 crystals each to be analysed in one visit. Data collections were automatically processed using the pipelines currently available at the DLS. Software packages PanDDA and XCE designed for the XChem facility merged all data sets and displayed the fo/fc, 2 fo/fc and an event map of amplified intensity of extra density not accounted for in the unbound M2-1 structure that may be the soaked fragment. Density is visualised in XCE and the fragment fitted manually to the density.

Soaking with DMSO and cryo-protectant was performed for 1 and 3 hour time points before crystal harvesting and cryo-cooling. Two methods of cryo-protectant addition at time of harvesting were explored to streamline the process for the full fragment screen. For the first method trialled, a 10% solution of cryo-protectant was placed in a vacant well and this subset of crystals placed in the mixture after harvesting. A second method used the ECHO acoustic liquid handler to add varying concentrations of the cryo-protectant mix (0- 20%) directly to the pre-defined x, y co-ordinates of the drop that contained the protein crystal and fragment (DMSO control).

Diffraction data sets were automatically collected on the I04-1 beamline and processed using the existing pipelines at the DLS. Upper resolution limits defined by the Xia3dii DLS pipeline are stated for each concentration of DMSO and each cryo-protectant addition method in Table 19 and Table 20. Unexpectedly, a great variation in resolution was seen between duplicates. Further analysis of crystal morphology (Figure 68) revealed clear differences between the data sets. Crystals with a cracked appearance were seen to diffract with resolution between 3-10 Å, whilst crystals with no cracks appeared to be able to diffract consistently to 2.3 Å, which represents a higher resolution than previously determined for M2-1.
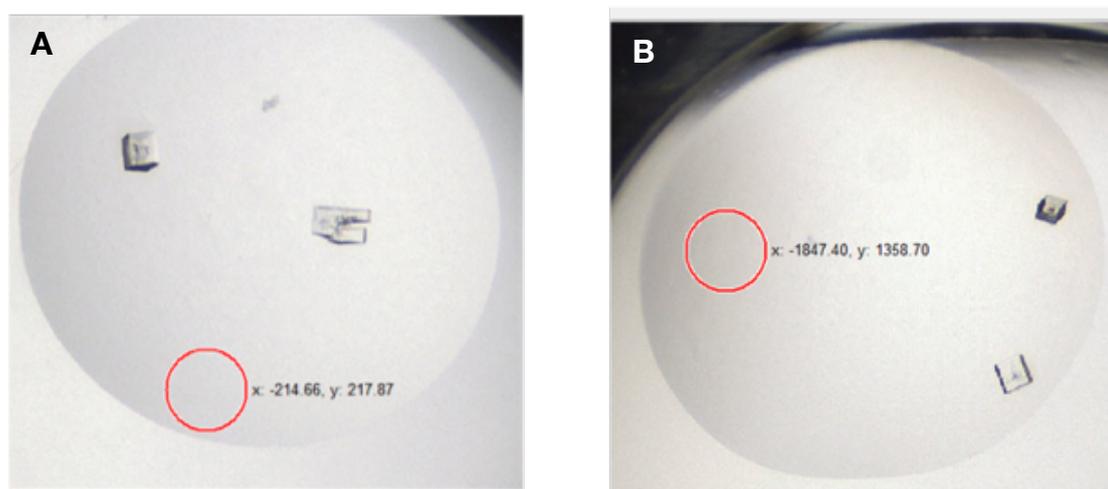
Figure 68 HRSV M2-1 crystal morphologies

> A- crystal morphology appears cracked, and diffracted to 3.23 Å; B-crystal had a cuboidal, non-cracked morphology and diffracted to 2.36 Å. Both crystals were soaked in 10% DMSO with a 5% cryo-protectant mix for 1 hour before harvesting. This pattern of resolution in relation to morphology occurred at all concentrations tested.

Solvent DMSO (%)	Cryo-protectant (%)	Resolution Å	Resolution Å
0	10	2.91	2.58
0	10	2.79	3.02
10	10	LD	2.52
10	10	LD	LD
20	10	2.4	MF
20	10	MF	MF

Table 19 XChem optimisation of DMSO concentration using a transfer method to cryo-protectant

Crystals soaked with the fragment vehicle, DMSO were harvested at 1 and 3 hours post soaking before transfer to a pre-defined well containing 10% cryo-protectant. Each condition was performed in duplicate. Data set highlighted in green displays the best resolution condition. LD- low-resolution diffraction; MF- crystal could no longer be harvested due to DMSO/ cryo-protectant addition.

5.2.1.3 Optimised condition for XChem fragment screen

During the optimisation process, crystals diffracted successfully to 2.36 Å and 2.33 Å in the absence of DMSO (Table 20). However, during the screening process DMSO is required as the vehicle for the fragments and thus a condition that could withstand DMSO was required. Due to differences in resolution because of crystal morphology all crystal images were re-analysed through TexRankE and only larger crystals (~150 µM) with a non-cracked morphology were chosen for continuation. Optimal resolution (2.3 Å) was seen for a crystal 1 hour post-soaking with 10% DMSO and 5% cryo-protectant (Table 20).

To summarise, M2-1 protein (10 mg/ ml) was set up in drops 2 and 3 of the 3 drop SWISS CI sitting drop plates in a 1:1 ratio with the mother liquor (0.1 M Carboxylic acid, 0.1 ml BS3 pH 8.5, 24% PEG550MME_PEG20K- Morpheus screen condition). Plates were incubated for 3 days to form crystals. Crystals of ~ 150 µM with no cracked appearance were marked with (x, y) co-ordinates by TexRankE for soaking, Crystal drops were soaked with 10% DMSO (containing the fragment) for 1 hour before addition of a 5% cryo-protectant and harvested and cryo-cooled prior to automated data collection.

Solvent DMSO (%)	Cryo-protectant (%)	Resolution Å	Resolution Å
0	0	MF	NA
0	0	MF	NA
10	0	3.58	NA
10	0	2.85	NA
20	0	3.02	NA
20	0	3.26	NA
0	5	LD	3.14
0	5	2.36	3.05
10	5	2.32	LD
10	5	3.23	LD
20	5	2.55	MF
20	5	2.54	MF
0	10	2.84	3.21
0	10	2.83	LD
10	10	LD	MF
10	10	3.16	3.4
20	10	3.47	LD
20	10	LD	LD
0	20	2.33	3.02
0	20	LD	LD
10	20	3.36	LD
10	20	MF	3.42
20	20	LD	LD
20	20	LD	MF

Table 20 XChem DMSO and cryo-protectant concentration optimisation

Cryo-protectant was added directly using the ECHO liquid handling apparatus to the soaked crystal drop using the previously determined TexRankE (x, y) co-ordinates. Crystals were harvested at 1 and 3 hour time points, in duplicate and resolution limits determined. Data sets highlighted in green show the best diffracting conditions. LD- low-resolution diffraction; MF- crystal could no longer be harvested due to DMSO/ cryo-protectant addition.

5.2.2 Fragment screening data

5.2.2.1 Hit fragments bound to two distinct regions of HRSV M2-1

Automated data collection (Figure 69), and subsequent processing using automated pipelines available at the DLS, resulted in 1,062 data sets for inspection. Software packages PanDDA and XCE have been designed specifically for the XChem facility to provide high-throughput identification of potential weak binding ligands with low occupancy by subtraction of a proportion of the averaged ground (non fragment bound) state to leave clear density for only the changed (fragment bound) state, and is displayed as an event map in addition to fo/fc and 2fo/fc maps. The software eliminated strong electron density with no significance (i.e. water molecules) and only identified regions of unique density (i.e. fragment) for each data set. XCE in addition to COOT summarised all unique densities and allowed manual fitting of the fragment to the density.
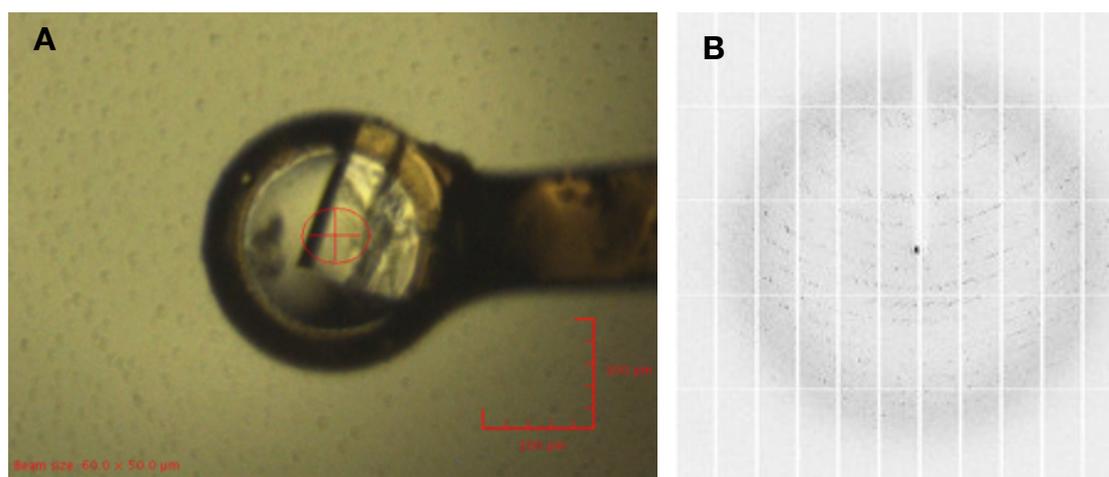
Figure 69 XChem crystal harvesting and data collection

A- M2-1 HRSV crystal incubated with fragment M2-1-x0037; B- diffraction image of crystal shown in A. The crystal diffracted to 2.4 Å, representative of the average resolution seen for all data sets.

Initial analysis and fitting of the fragment structure to the density by visual inspection was achieved for 21/ 1062 fragments, providing a hit-rate of 2%. Due to the low occupancy nature of FBDD and solvent accessible channels, full occupancy of each of the 16 M2-1 monomers in the asymmetric unit cell (AUC) was not expected and modelling of the fragments in the available density was achieved for between 1 and 5 M2-1 monomers. Fragments were fitted to the

density using XCE and COOT, where they were refined once and collated to reveal the location of each fragment in relation to M2-1. Fragments involved in crystal contacts and not binding to M2-1 were not further assessed. Further refinement of the fragments and visualisation of potential interaction with M2-1 revealed seven fragments that were modelled to the density with high confidence and bound to two distinct regions (Table 21 & Table 22), suitable for further analysis as inhibitors. Four fragments bound near the RNA/ P interaction surface (x0079, x0373, x0770, x0816) and three near the ZBD (x0108, x0869, x0937) and these were further refined and assessed.

Fragment	Resolution (Å)	$R_{Crystal}$	R_{free}
M2-1x0079	2.6	0.221	0.279
M2-1x0108	2.6	0.233	0.289
M2-1x0373	2.4	0.234	0.284
M2-1x0770	2.4	0.240	0.294
M2-1x0816	2.4	0.238	0.289
M2-1x0869	2.4	0.239	0.286
M2-1x0937	2.4	0.227	0.276

Table 21 XChem hit fragment refinement values

Only the electron density representing the fragment was further manually refined and not the M2-1 backbone. Values were determined using the XCE software.

Fragment	Fragment Structure	M2-1 binding area	Density (/16)
M2-1x0079		RNA/ P	1
M2-1x0108		ZBD	4
M2-1x0373		RNA/ P	1
M2-1x0770		RNA/ P	1
M2-1x0816		RNA/ P	5
M2-1x0869		ZBD	3
M2-1x0937		ZBD	1
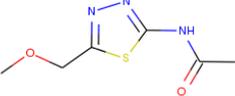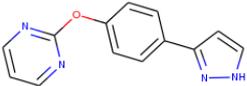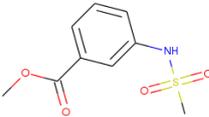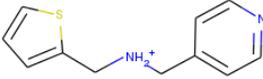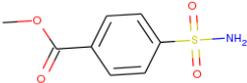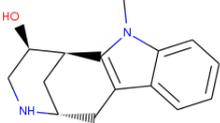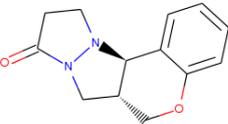
Table 22 XChem hit fragment structures

Extra electron density that was not accounted for by the M2-1 protein and allowed fitting of the fragment were determined 'hits' and were further assessed. The corresponding location of the observed fragment was recorded as the RNA/ P site or ZBD. Density out of a possible 16 refers to density fitted with the fragment at the same binding region of the 16 M2-1 monomers within the AUC.

5.2.2.1.1 Zinc binding domain hit fragments

Fragments M2-1x0108, x0869 and x0937 were refined and interacted near the zinc co-ordination site. M2-1x0937 (Figure 70A pink) and M2-1x0869 (Figure 70A yellow) interacted with M2-1 at the same binding site and may bind through similar interactions. Multiple structures at this solvent accessible binding site will greatly aid further drug development targeted to this region. M2-1-x0108 (Figure 70B green) binds at a distinct site at the zinc co-ordinating area, away from the other 2 fragments.
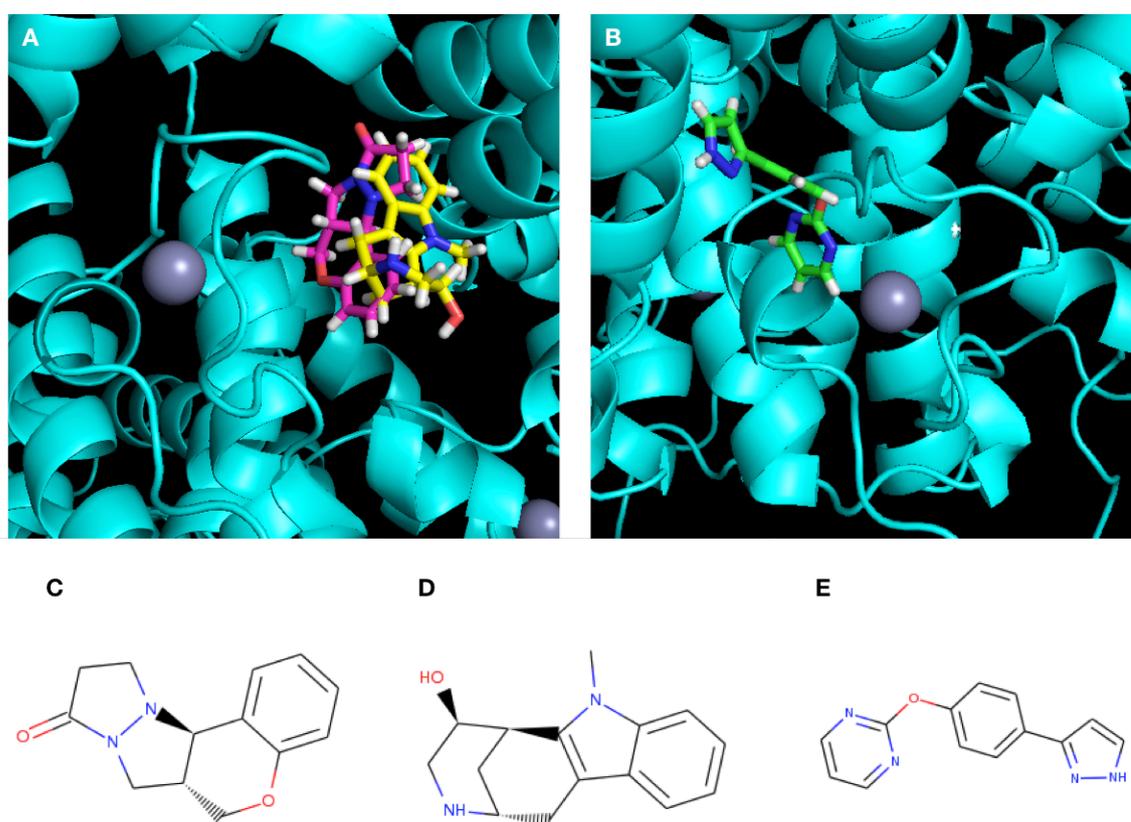
Figure 70 Fragment structures determined at the ZBD by X-ray crystallography

A- fragments M2-1x0869 (yellow) and M2-1x0937 (pink) bind to M2-1 at the same site; B- M2-1x0108 binds to a distinct area at the ZBD; C- 2D structure of M2-1x0937; D- M2-1x0869 and E-M2-1x0108.

5.2.2.1.2 RNA/ P binding surface hit fragments

Four fragments bound near the RNA/ P binding surface with fragment M2-1x0079 (Figure 71B & D green) binding directly to the cavity previously targeted for structure based drug design (chapter 4) that is involved in both RNA and P interactions. M2-1x0373 (Figure 72A & C pink) interacts with residues that comprise the RNA/ P binding surface on the M2-1 protein. Fragments M2-1x0770 (Figure 71A & C orange) and M2-1x0816 (Figure 72B & D yellow) interact near the RNA/ P surface and have a shallow binding surface.
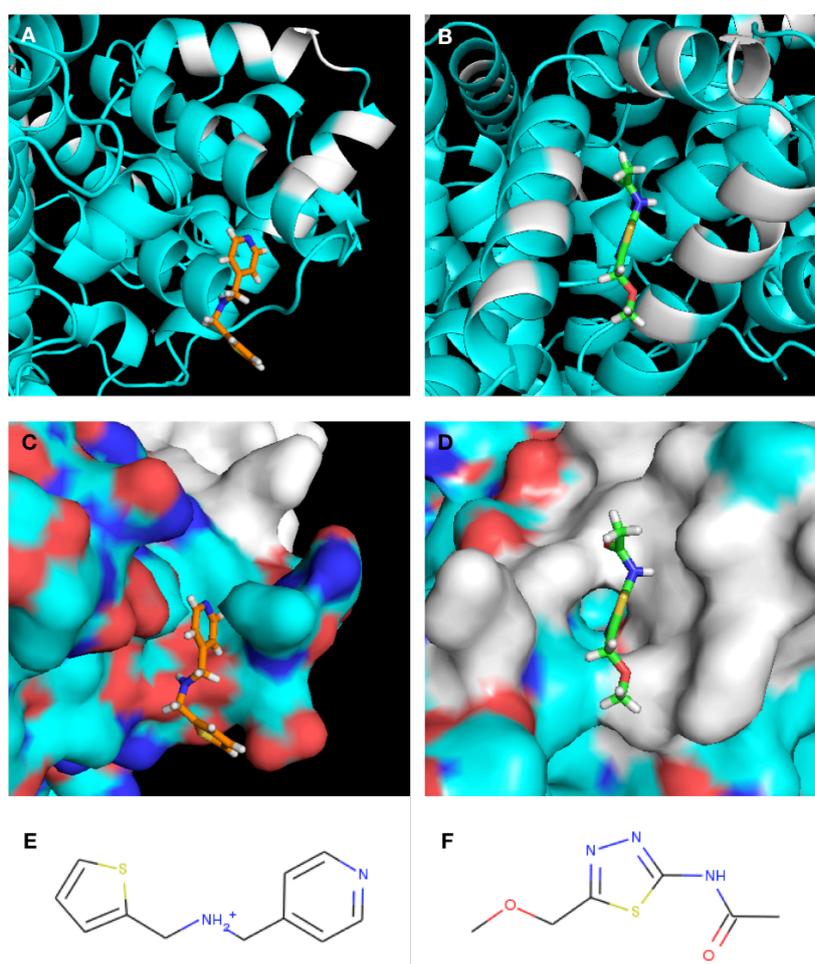
Figure 71 Fragment structures determined near the RNA/ P interaction surface by X-ray crystallography

A- cartoon representation of M2-1 with fragment M2-1x0770 bound; B- cartoon representation of M2-1 with fragment M2-1x0079; C; surface representation of interaction A; D; surface representation of interaction B; E- 2D structure of M2-1x0770; F- 2D structure of M2-1x0079. Residues involved in interactions with the M2-1 binding partners RNA and P are highlighted in white.
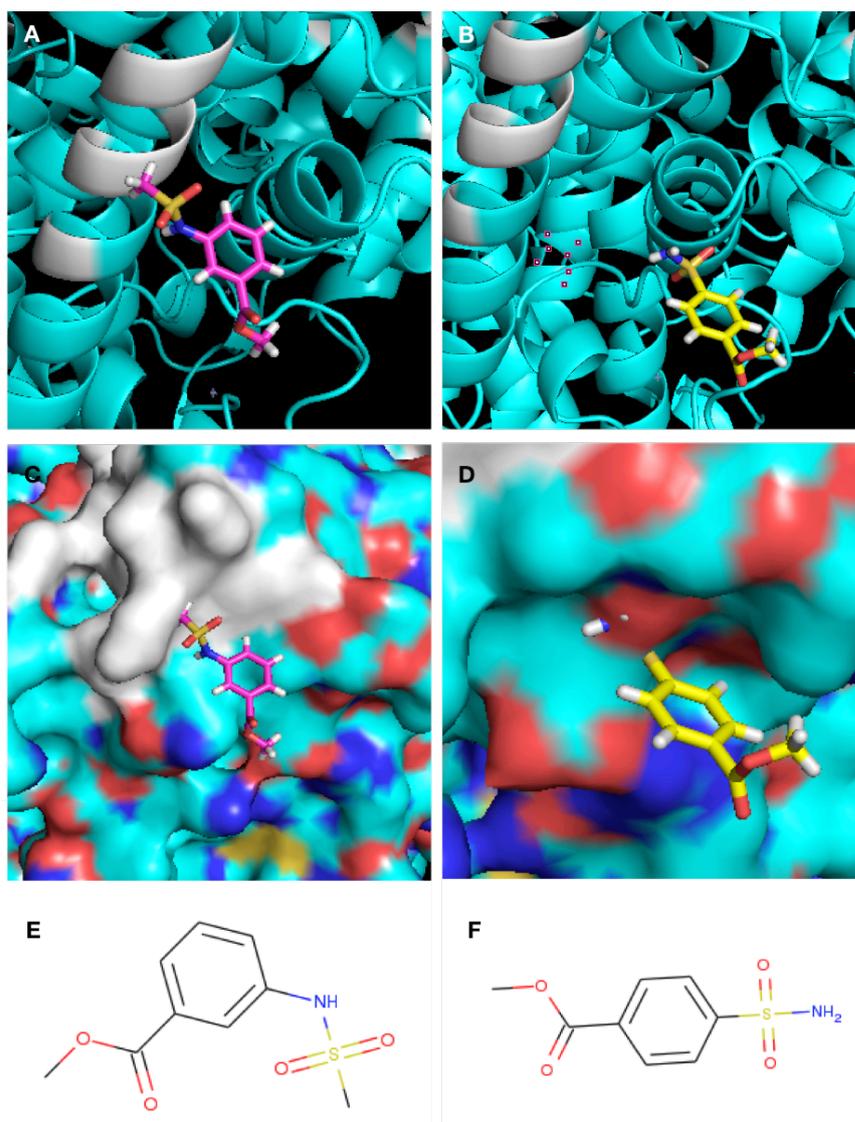
Figure 72 Further fragment structures determined near the RNA/ P interaction surface by X-ray crystallography

A- cartoon representation of M2-1 bound to fragment M2-1x0373; B- cartoon representation of M2-1 bound to M2-1x0816; C- surface representation of interaction A; D- surface representation of interaction; E- 2D structure of M2-1x0373 bound in A and C; F- 2D structure of M2-1-x0816 bound in B and D. Residues involved in interactions with M2-1s binding partners RNA and P are highlighted in white.

5.2.2.2 M2-1: fragment interactions

The possible interactions between the hit fragments and M2-1 were determined based on the electron density maps and distances between promising interactions and are discussed below.

5.2.2.2.1 M2-1x0770

Fragment M2-1x0770 was modelled with high confidence due to the high-resolution density available (Figure 73). The negatively charged side chain of Asp 116 could form an ionic bond with the positively charged NH group of the fragment, whilst Thr 69 of M2-1 could hydrogen bond with the nitrogen atom of the pyridyl ring.
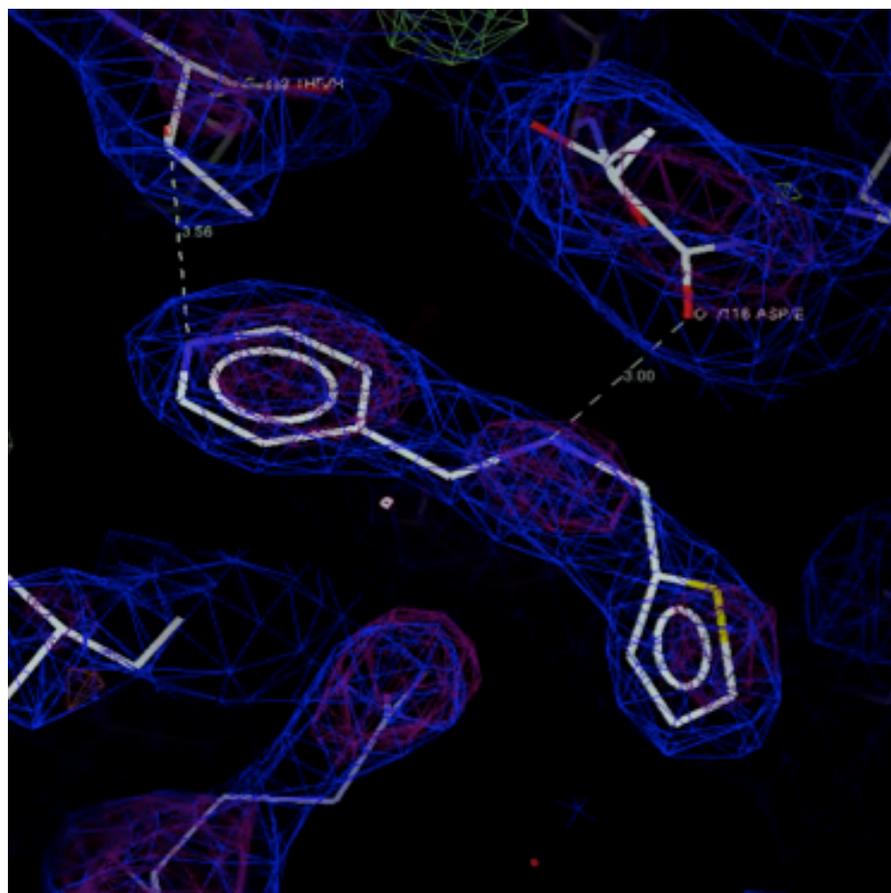
Figure 73 Electron density of fragment M2-1x0770 bound to HRSV M2-1

> View of the fragment fitted to the electron density 2 fo/fc and event maps (PanDDA). The negative charged side chain of Asp 116 could form an ionic bond with the positively charged NH group of the fragment, whilst Thr 69 of M2-1 could hydrogen bond with the nitrogen atom of the pyridyl ring.

5.2.2.2.2 M2-1x0079

The fragment could only be modelled to the density in 1 of the 16 M2-1 monomers (Figure 74). Interactions of fragment M2-1x0079 with M2-1 could result from hydrogen bonding of the carbonyl of the fragments amide group to the side chain of Lys 159 of M2- 1 and via an oxygen atom of the fragment with Thr 130 of M2-1.
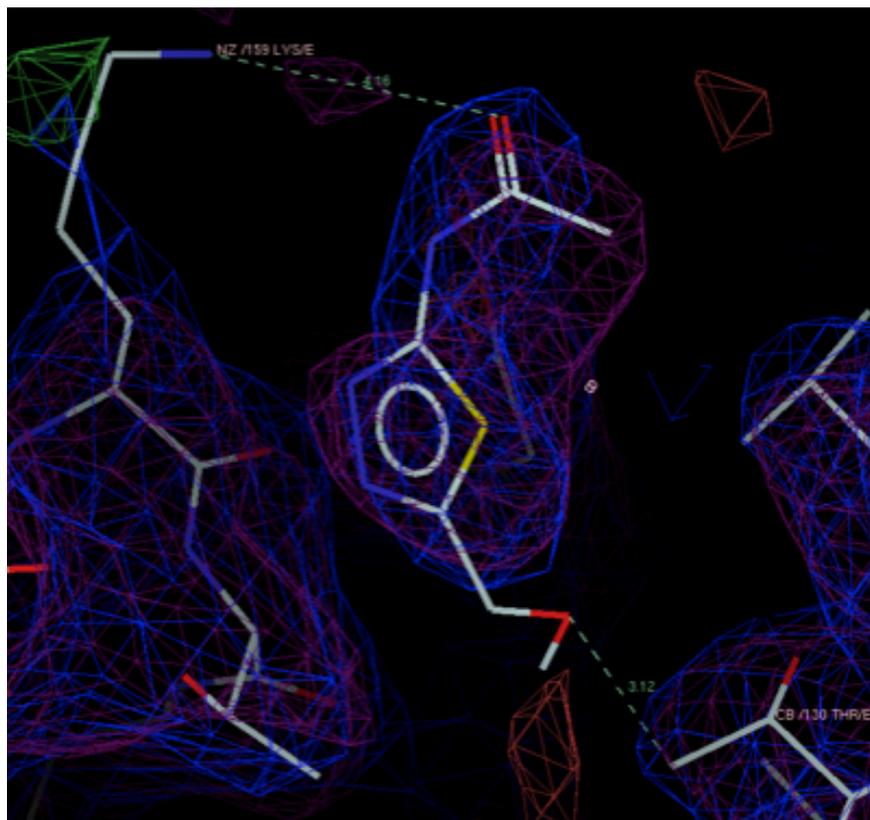
Figure 74 Electron density of fragment M2-1x0079 bound to HRSV M2-1

View of the fragment fitted to the electron density 2 fo/fc and event maps (PanDDA). An interaction between the carbonyl of the fragments amide group to the side chain of Lys 159 of M2- 1 and via an oxygen atom of the fragment with Thr 130 of M2-1 is visualised.

5.2.2.2.3 M2-1x0108

The M2-1x0108 fragment binds near the ZBD of M2-1 (Figure 75). M2-1x0108 could be fitted accurately to the electron density of 4 monomers of M2-1 in the AUC. The fragment displayed potential hydrogen bonding between the side chain of Glu 81 and the NH of the fragment pyrazole ring. Further, a pi-stack interaction is visualised between the fragment pyrimidine ring and Phe 9 of M2-1.
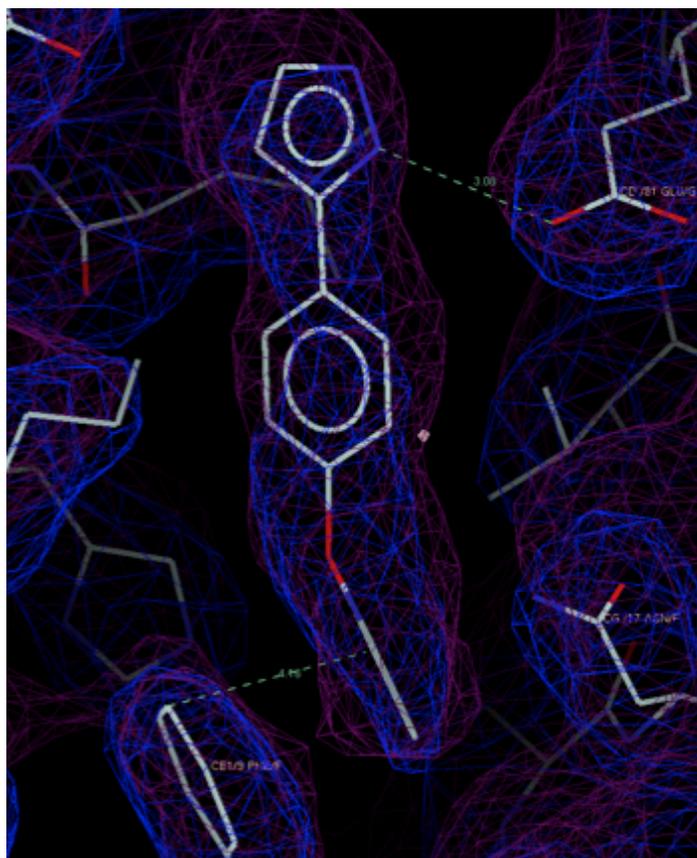
Figure 75 Electron density of fragment M2-1x0108 bound to HRSV M2-1

View of the fragment fitted to the electron density 2 fo/fc and event maps (PanDDA). The fragment displayed potential hydrogen bonding between the side chain of Glu 81 and the NH of the fragments pyrazole ring and a pi-stack interaction is visualised between the fragments pyrimidine ring and Phe 9 of M2-1.

5.2.2.2.4 M2-1x0373

The fragment M2-1x0373 could be fitted with high confidence to 2 of the M2-1 monomers (Figure 76) of the AUC. The side chain of Lys 162 could form hydrogen bond interactions with the oxygen atom of the ester group of the fragment. The benzene ring of the fragment could result in hydrophobic interactions with Ile 84 of M2-1 whilst the side chain of Lys 159 may exhibit hydrogen bonding with the sulphur termini of the fragment.
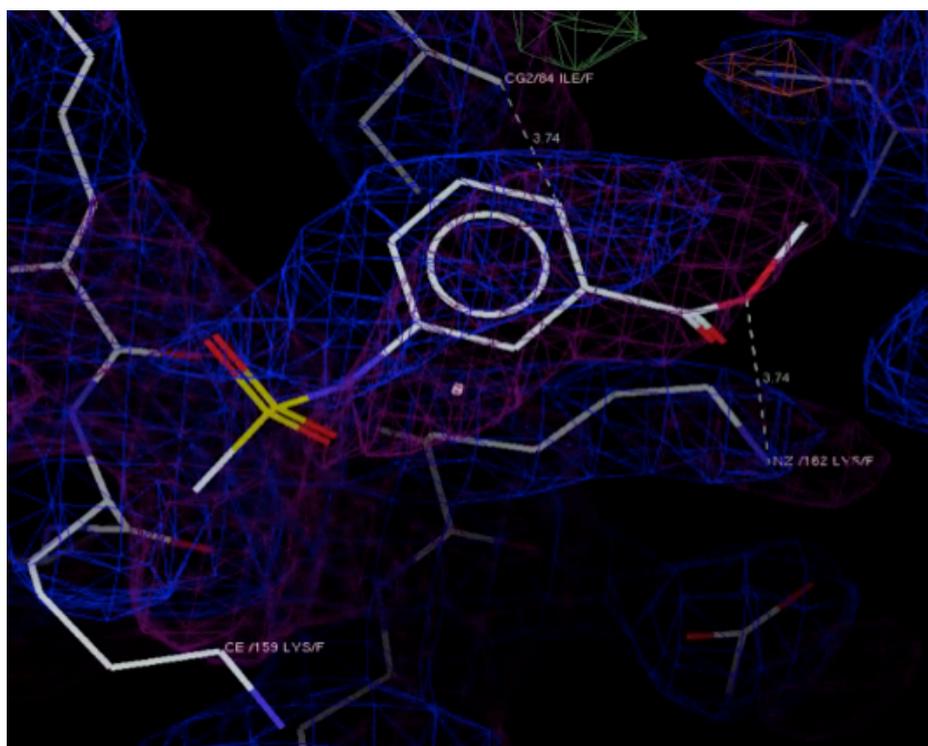
Figure 76 Electron density of fragment M2-1x0373 bound to HRSV M2-1

View of the fragment fitted to the electron density 2 fo/fc and event maps (PanDDA). The side chain of Lys 162 could form hydrogen bond interactions with the oxygen atom on the ester group of the fragment and the benzene ring of the fragment could result in hydrophobic interactions with Ile 84 of M2-1.

5.2.2.2.5 M2-1x0816

Fragment M2-1x0816 (Figure 77) was fitted with high confidence to the additional electron density in 5 of the 16 M2-1 monomers in the AUC. Density was visible but weaker in other monomers at the same interaction site and the fragment could not be fitted accurately. Fragment M2-1x0816 could form potential hydrogen bond interactions with M2-1; through a hydrogen donor group (NH_2) at the fragments sulphonamide to Glu 81 of M2-1 (Figure 77) and via an oxygen atom at the fragments sulphonamide with the hydrogen donor group (NH2) of the side chain of Arg 20 of M2-1. Both hydrogen bond interactions formed within suitable molecular distances of 2.3 Å and 2.6 Å respectively.
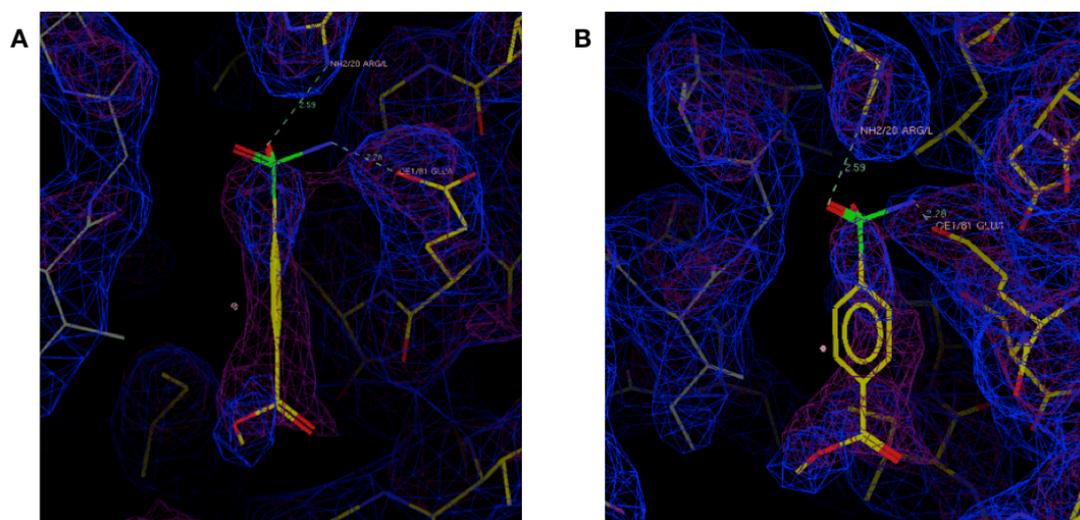
Figure 77 Electron density of fragment M2-1x0816 bound to HRSV M2-1

A- fragment fitted to the electron density 2 fo/fc and event maps (PanDDA); B- as A at a different angle to further visualise the interaction. Hydrogen bond interactions with M2-1 are seen via a hydrogen donor group (NH2) at the fragments sulphonamide to Glu 81 of M2-1 and via an oxygen atom at the fragments sulphonamide with the side chain of Arg 20 of M2-1.

5.2.2.2.6 Further fragment interactions

For two of the three fragments shown to bind to the ZBD of M2-1 (Figure 70), the electron density was too weak to correctly predict the fragment orientation and thus its interactions with the M2-1 protein.

Fragment M2-1x0869 was fitted to the density of 5 of the 16 M2-1 monomers. Electrostatic interactions of M2-1 residues Lys 19 and Asp 116 may form the binding site for fragment M2-1x0869, however further visualisation of the fragments interactions with M2-1 are limited due to the poor density. Density for fragment M2-1x0937 (Figure 78) was seen in 6 of the M2-1 monomers within the AUC, however weak density resulted in little confidence in the fitting of the fragment. Only one fragment (Figure 78) was fitted to the density with some confidence, here clear side chain density can be seen for Lys 19 of Chain B M2-1 and Asp 16 Chain D M2-1 suggesting an interaction with the fragment or most likely a stabilisation of the binding site. Further, the side chain of Tyr 72 could form a hydrogen bond with the carbonyl oxygen atom of the fragment.
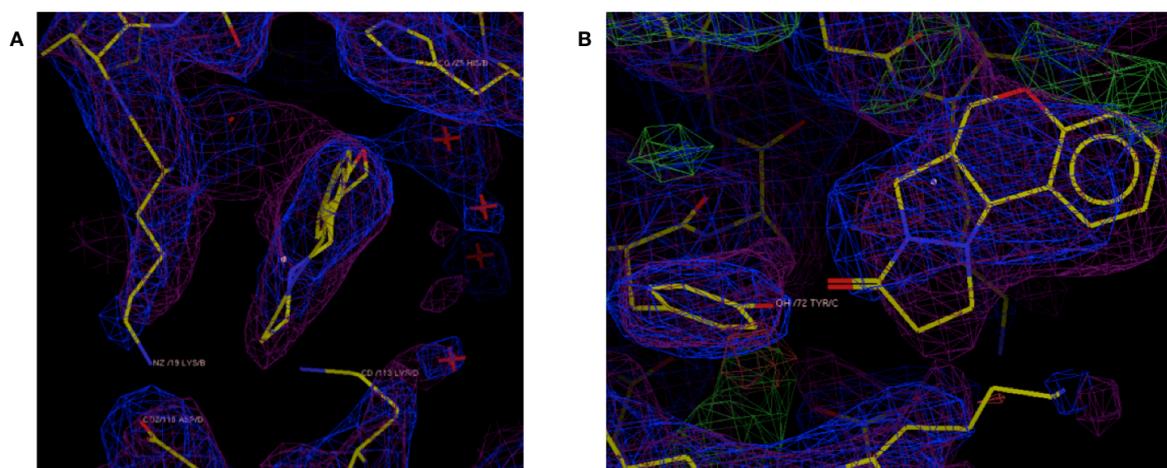
Figure 78 Electron density of fragment M2-1x0937 bound to HRSV M2-1

A- view of the fragment fitted to the electron density 2 fo/fc and event maps (PanDDA); B- as image A at a different angle to further visualise the interaction. Clear side chain density can be seen for Lys 19 of Chain B M2-1 and Asp 16 Chain D M2-1 suggesting an interaction with the fragment or most likely a stabilisation of the binding site. Further, the side chain of Tyr 72 could form a hydrogen bond with the carbonyl oxygen atom of the fragment.

5.3 Discussion

5.3.1 Crystal optimisation

Fragment screening by X-ray crystallography is a material intensive method and would not have been possible without the improved expression and purification of M2-1, described in chapter 3. Each protein that is assessed using the XChem facility requires optimisation for the production of reproducible, high resolution diffracting crystals that maintain high-resolution diffraction following addition of the fragment in its DMSO vehicle. Despite the novel expression and purification method for M2-1 that resulted in protein crystals in many conditions of a commercial screen (appendix III), the batch-batch reproducibility of crystal conditions has never been examined. Here, a single condition was identified and resulted in reproducible diffraction to a resolution of at least 3 Å. The identification of this optimal condition could be used for further co-crystal soaking experiments beyond the XChem facility.

Further optimisation of crystals required for the XChem screening gave additional insight into the M2-1 crystals used. Crystals with a non-cracked morphology consistently gave reliable diffraction data between 2.2- 2.3 Å and were reliably easier to harvest with sub-10 second harvesting times and proved ideal for the XChem facility pipeline. M2-1 protein crystals were found to weakly diffract upon soaking in DMSO for 3 hours and could not be soaked for more than 1 hour before cryo-cooling. Despite DMSO being a sufficient cryoprotectant for certain proteins, M2-1 required additional cryoprotectant for optimal diffraction. The maximum DMSO concentration the M2-1 crystals could withstand was 10%, whilst other users of the XChem facility found suitable crystal diffraction with up to 40% DMSO (Dr Patrick Collins, personal communication). The higher (40%) DMSO concentration resulted in over 40 hit fragments due to visualisation of weak binders at a high concentration of the fragment. However, due to the high DMSO concentration, many false positives were observed that would likely be excluded at the lower DMSO concentration as used here. Thus, there is a balance between detecting more fragment binders without attracting too many false positives. It is to be noted that all values of DMSO and cryoprotectant presented here are values provided to the ECHO liquid handler to dispense and are not precise volumes, due to the viscosity of solutions. In addition, the protein crystal drop had most likely dehydrated in size from the initial 0.4 µl volume and again indicates that all values are relative.

5.3.2 Fragments bound to two distinct regions of M2-1

Fragments were soaked into M2-1 protein crystals, and using X-ray crystallography were determined to bind to two distinct regions of the protein. Extra density not accounted for by M2-1 was highlighted by the XChem software at other regions (not the ZBD or RNA/ P site) but these were either crystal contacts or had poor density that may not have been the fragment and would not aid further SAR and drug discovery. These data show 7 hit fragments that bind two distinct regions of the M2-1 protein; the RNA/ P binding surface that has been previously targeted as a drug target site (see chapter 4) and the ZBD that has been suggested for anti-viral drug design (Esperante et al., 2013; Cancielleri et al., 2015). Large sigma values due to zinc co-ordination have been accounted for with the previously described PanDDA software, allowing fragment density to be seen independently of the zinc density and resulted in high confidence of the fitted fragment to the remaining density.

Three fragments bound to the ZBD that was previously targeted by Cancielleri et al, (2015) using SBDD of the HRSV M2-1 crystal structure (Tanner et al., 2014). In this study a compound (compound 10e) inhibited HRSV replication, however a direct interaction with M2-1 was not discovered (Cancielleri et al., 2015). Here, the structures of 3 fragments bound to 2 distinct sites at the ZBD could aid further SAR studies. Two fragments (M2-1x0937 and M2-1x0869) bound to the same region of M2-1, however as density was poor for M2-1x0937 the correct orientation of the fragment to the density could not be determined and possible interactions could not be hypothesised.

Four fragments were identified as binding near the RNA/ P binding surface that was previously targeted for SBDD via computational methods, described in chapter 4. Further, inhibition of RNA and P binding to M2-1 has been identified to result in a loss of transcriptional function (Cuesta et al., 2000; Mason et al., 2003) and provides a good target site for further SAR. Fragments M2-1x0079 and M2-1x0373 interact directly to the region implicated in binding RNA and P, whilst fragments M2-1x0770 and M2-1x0816 bind near the RNA/ P binding surface and may require analogue synthesis that extend towards the RNA/ P binding surface to inhibit this function.

5.3.3 Further assessment of hit fragments bound to M2-1

Although the XChem facility is useful for identification of initial weak binders to the protein of interest there is no streamlined method to follow up fragment hits with functional assays at the XChem facility. Hit fragments identified here could be assayed for their ability to disrupt the RNA/ P binding surface with the optimised FA assay, however the sensitivity of this functional assay is in question and likely would not compete with P 90-110 for M2-1 binding. Fragments could be further assessed using the mini genome assay to survey replication and transcription or in the infectious HRSV assay to identify infectivity, both of which have been optimised for drug design work (see chapter 3). Due to the low molecular weight of the fragments (appendix VI) and limited sensitivity of the assays for HRSV and M2-1 inhibition it is possible that these fragments may not be identified as inhibitors and thus further analogue synthesis of the hit fragment may be required. Analogues may be synthesised based on the initial bound fragment structure and extended at the fragments termini to increase MW and potential affinity for HRSV M2-1 to allow detection in a functional assay. Previous methods of co-crystallisation of hit compounds (chapter 4) have proved unsuccessful for M2-1. The XChem facility has resulted in the optimisation of M2-1 crystal conditions to better understand the parameters required for efficient diffraction on soaking. Further, the XChem facility has resulted in the visualisation of fragments bound to the M2-1 protein allowing direct analysis of the interaction between the protein and fragment and reducing the time required for further SAR and drug development.

5.3.4 Summary

The data presented here and in chapter 4 has resulted in lead drug discovery compounds (SBDD) and fragments (as the building blocks of these lead drug compounds) targeted specifically to HRSV M2-1. Given the need for an affordable efficacious treatment for both HRSV and BRSV in addition to the high sequence identity between the two proteins, data in the final chapter here aims to identity further similarities between BRSV and HRSV M2-1 proteins using structural and biophysical methods with an overall aim to determine M2-1 inhibitors that can inhibit both human and bovine species of RSV.

Chapter 6 Characterisation of the M2-1 protein of bovine respiratory syncytial virus (BRSV)

6.1 Introduction

Bovine respiratory syncytial virus (BRSV) was first isolated in 1970 in Europe from cattle with respiratory diseases (Paccaud and Jacquier, 1970) and has now been detected in cattle herds worldwide (Inaba et al., 1970; Wellemans et al., 1970; Edington & Jacobs, 1970). BRSV accounts for over 60% of lower respiratory tract infection and bronchopneumonia in cattle (Gershwin, 2007). Economic loss from bovine respiratory diseases, including BRSV, is estimated at $1 billion worldwide annually due to mortality of both beef and dairy cattle and a decrease in milk yield in dairy cows, with a field study presenting a direct link between BRSV infection and decreased milk production (Paccaud & Jacquier, 1969; Inaba et al., 1972; Smith et al., 1974; Ferguson et al., 1997; Sacco et al., 2013). HRSV and BRSV belong to the *Pneumoviridae* family that also includes the important human pathogen human metapneumovirus (HMPV). HRSV and BRSV are further classified within the *Orthopneumovirus* genus and have a high level of genetic similarity. Both viruses have 10 genes, in the same order, with similar conserved *cis*-acting sequence signals and genome length. HRSV and BRSV encode 11 highly conserved proteins and their similarities are highlighted in Table 23 and Table 24. BRSV provides a good model for the study of HRSV due to their similar immunological response in their respective hosts. The pathological effect of BRSV infection in bovines is a viable option to study novel HRSV vaccines and anti-viral compounds with fewer ethical concerns (Guzman & Taylor, 2015).

6.1.1 BRSV disease symptoms

Infection with BRSV is predominantly fatal in calves of 1- 9 months, similar to the fatality incidence in human infants infected with HRSV (Stott et al., 1980). Evidence of respiratory distress caused by bronchopneumonia can be observed in calves by an open mouth with a grunt, a stretched neck and an extended tongue with pouring saliva. BRSV infected cattle are susceptible to secondary infections with bacterial pathogens such as *M. haemolytica, P multocida, H. somni and M. bovis*. Secondary infections due to BRSV result in mortality and for the surviving calves, a lag in weight gain and growth (Belknap, 1993; Valarcher & Taylor, 2007; Sacco et al., 2013). In addition, the massive overuse of antibiotics

to treat such secondary infections poses a public health concern through the risk of emergence of pathogenic bacterial strains with acquired antibiotic resistance (reviewed in Seiffert et al., 2013).

6.1.2 Comparison of BRSV and HRSV

Identification of BRSV and subsequent comparison with HRSV revealed a closely related antigenic structure through cross-neutralisation experiments with guinea pig anti-sera. In addition, BRSV and HRSV both induced the formation of characteristic syncytia in cultured cells, and inclusion bodies were visualised in the infected cell cytoplasm (Inaba et al., 1972). For virus entry and exit, HRSV and BRSV employ the same apical surface of the species-specific polarised airway epithelial cells (Alansari & Potgieter, 1994). The majority (95%) of adult cattle displayed sera positive results for BRSV antibodies (Patton et al., 1998) and similar to the situation of HRSV and humans, re-infection of cattle with BRSV is common.

HRSV and BRSV both inhibit IFN-induced cellular response through expression of their respective NS1 and NS2 proteins. However, HRSV NS proteins modulate type I IFN by inhibition of Stat2 expression (Lo et al., 2005) whilst BRSV NS proteins block activation of IRF3 (Bossert et al., 2003). Further, in BRSV, NS2 appears to have a greater inhibitory effect on type 1 interferon induction than NS1 and the converse of this is seen for HRSV (Valarcher et al., 2003; Spann et al., 2004). The pep27 peptide produced by cleavage of the F protein at the two furin sites contains a tachykinin motif in BRSV, not seen in HRSV. This could have an immunological function in cattle or be a naturally occurring species-specific sequence (Zimmer et al., 2002; Zimmer et al., 2003). As previously described, HRSV is divided into two subgroups (A and B) based on antigenic and genetic differences. In contrast, classification of BRSV based on monoclonal antibody reactivity to the G protein reveals limited divergence, and consequently BRSV isolates have been postulated to exist as one antigenic group (Furze et al., 1997).

M2-1 has an essential role in the HRSV infection cycle (Collins et al., 1996), evidenced by the inability to rescue infectious HRSV lacking the M2-1 coding region. In common with this, to produce recombinant BRSV, a plasmid encoding the M2-1 must be transfected into cells in addition to those expressing N, P and L proteins (Buchholz et al., 1999) and suggests that the BRSV M2-1 protein also plays an essential role in the virus life cycle. BRSV and HRSV M2-1 proteins have an amino acid sequence identity of 80.1% and a similarity of 90.3%, making it highly conserved between the two virus species. The majority of sequence

differences are located at the CTD of the protein, as seen by the alignment of BRSV and HRSV M2-1 in Figure 79. These CTD residues (174-194) in HRSV M2-1 were not visible in electron density obtained by X-ray crystallography (Tanner et al., 2014), and are presumed disordered. Like HRSV M2-1, the BRSV M2-1 C terminal residues 174-194 are presumed disordered (Drozdetskiy et al., 2015).

```
HRSV MSRRNPCKFEIRGHCLNGKRCHFSHNYFEWPPHALLVRQNFMLNRILKSM
     ||||||||*|||||||||||*||||||||||||||||||||||*|||||
BRSV MSRRNPCKYEIRGHCLNGKKCHFSHNYFEWPPHALLVRQNFMLNKILKSM

HRSV DKSIDTLSEISGAAELDRTEEYALGVVGVLESYIGSINNITKQSACVAMS
     |** ||||||||||||||||||||||*||||||* |||||||||||||||
BRSV DRNNDTLSEISGAAELDRTEEYALGVIGVLESYLSSINNITKQSACVAMS

HRSV KLLTELNSDDIKKLRDNEELNSPKIRVYNTVISYIESNRKNNKQTIHLLK
     |||  |*|*||||*||* |    |||||*|||||||||*||**| ||||||||
BRSV KLLAEINNDDIKRLRNKEVPTSPKIRIYNTVISYIDSNKRNTKQTIHLLK

HRSV RLPADVLKKTIKNTLDIHKSITINNPKESTVSDTNDHAKNNDTT
     |||||||||||||*|||  |   ||  *  | * | *  |*
BRSV RLPADVLKKTIKNTIDIHNEINGNNQGDINVDEQNE
```

Figure 79 Pairwise sequence alignment of HRSV and BRSV M2-1

Local alignment between HRSV and BRSV M2-1 open reading frames was performed using the Smith-Waterman algorithm (EMBL server). The symbol '|' indicates an identical residue, '*' represents a substituted residue with similar functional characteristics. Where no symbol is shown, this indicates a substituted residue with no similarities.

Properties	HRSV	BRSV
First isolated	1956	1970
Genome Length	15,128 bp (A strains NCBI: KF530268.1) 15,119 bp (B strain NCBI: KJ723485.2)	15,140 bp (NCBI: AF092942.1)
Proteins	11	11
Approved Vaccine	No	Yes
Transmission	Large droplets Aerosol	Large droplets Aerosol
Peak infectivity (UK)	Winter months	Winter months

Table 23 Comparison of HRSV and BRSV properties

187

BRSV protein	BRSV protein (amino acids)	Identity with HRSV (%)
NS1	136	69
NS2	124	84
N	391	93
M	256	89
SH	73	38-44
G	257	30
F	574	81
P	241	81
M2-1	186	80
M2-2	95	43
L	2161	84

Table 24 Protein sequence identity

HRSV and BRSV genomes encode 11 highly conserved proteins. Their amino acid sequence identity is compared here (adapted from Valarcher et al., 2007).

6.1.3 BRSV vaccine treatments

Unlike HRSV for which no vaccine treatment exists, live attenuated vaccines (Delforge and Zygraich 1978; Bohlender, 1984) and inactivated vaccines (Howard et al., 1987) have been approved for BRSV treatment since the late 1970s in Europe and the North America with varying levels of efficacy, discussed here (and reviewed in Ellis, 2016).

A commercially available modified live combination BRSV parenteral vaccine (BRSV, bovine parainfluenza virus-3 [BPIV-3], bovine herpesvirus-1 [BHV-1] and bovine viral diarrhea virus [BVDV]) was used to vaccinate 2-4 week old BRSV antibody negative calves before challenge 3 weeks later with 10^5 plaque forming units of BRSV. Vaccines were provided in non-adjuvanted (Bovishield-4 SmithKline Beecham Animal health BRSV 375 isolate) and adjuvanted (Pyramid-4, Fort Dodge/ Ayerst) forms. Vaccinated calves had significantly less lung lesions and nasal shedding of BRSV compared to unvaccinated control calves (West et al., 2000). In a second study using a commercially available modified live

combination parenteral vaccine (BRSV, BHV-1, BPIV-3, BVDV-1 and 2; Vista 5 SQ, Merck Animal Health, BRSV 375 isolate) and a control vaccine as above, minus BRSV (Vista 3 SQ, Merck Animal Health) calves were fed BRSV high antibody positive colostrum prior to vaccination at 3-9 days of age. Calves were challenged 11 weeks after vaccination with 10^7 plaque forming units of BRSV and all calves developed moderate to severe respiratory disease. Calves had a lack of cytokine response (IL-4 and IFN-gamma) consistent with the idea that the calves were not immunologically primed and that the maternal antibodies inhibited the response of parenteral delivery of the vaccine (Ellis et al., 2014). As BRSV infection is more severe in these young calves with varying levels of maternal antibodies, modified live vaccines were further assessed via a mucosal (intranasal) administration method.

A combination modified live vaccine (BRSV, BPIV-3 and BHV-1; INFORCE 3, Pfizer Animal Health, BRSV 375 isolate) significantly decreased (p< 0.05) compared to control vaccinated calves, clinical disease associated with BRSV in seropositive and to a greater extent in seronegative calves when challenged with 10^6 $TCID_{50}$ of nebulised BRSV 9 and 7 weeks respectively after vaccination. The most critical finding of this study resulted from a longer period of time between vaccination and challenge (16 week challenge). In these calves, no difference was seen in clinical disease severity between vaccinated and control vaccinated calves. These findings suggest a use of intranasal MVL vaccines for BRSV protection, however immunity provided by the vaccines is relatively short (Ellis et al., 2013; Taylor et al., 2014).

Inactivated BRSV parenteral vaccines are also commercially available for the treatment of BRSV. A field trial of an inactivated BRSV vaccine (betapropriolactone-inactivated, alum and saponin adjuvanted), later withdrawn from market, resulted in increased cattle mortality on re-infection compared to non-vaccinated calves (Schreiber et al., 2000). These observations mirrored the issues seen for HRSV with the formalin inactivated vaccine that was trialled in the late 1960s and which led to exacerbated diseases (Kim et al., 1969). Further, enhanced disease severity in calves has also been reported for a modified live vaccine (Kimman et al., 1989). These findings suggest that similar issues still remain for BRSV and HRSV vaccine development.

The recent resurgence in development of second generation, more efficacious BRSV vaccines has yielded potential candidates. A chimpanzee adenovirus vectored vaccine expressing secreted forms of the HRSV proteins F, N and M2-1 induced neutralising antibodies in calves, protected against upper and lower RTI and showed no enhanced respiratory disease on re-infection (Taylor et al., 2015).

In addition to providing a model for BRSV infection the Valarcher group have designed a recombinant attenuated vaccine with deletion of the SH gene, that prevents reversion to virulence and an inactivated subunit vaccine containing HRSV P, M2-1 and N with BRSV F and G surface epitopes adjuvanted by oil emulsion (Hagglund et al., 2014; Blodorn et al., 2014). Both vaccines have been shown to induce separate immune responses in calves with maternally derived BRSV antibodies and current research is looking at the use of both vaccine in combination to offer long lasting protection against BRSV.

6.1.4 BRSV anti-viral treatment

Both HRSV and BRSV require further research into the underlying immunological events that result in vaccine exacerbated disease and illustrate the need for anti-viral development that may bypass the vaccine-induced immunological events and provide an effective anti-BRSV treatment option. No anti-viral treatment exists for BRSV and compounds in clinical trials (see chapter 1; section 1.9.3) have been designed to target HRSV. The complementarity of the BRSV and HRSV fields has previously been exploited for drug development (Blodorn et al., 2014) in which findings for HRSV are often mirrored for BRSV (Jordan et al., 2015) and can be extended to provide much needed anti-viral development research.

6.1.5 Chapter aims

The M2-1 protein is perhaps the least studied of the 6 essential HRSV structural proteins with even less information available on the bovine homologue. Given the current resurgence in interest for developing a more efficacious BRSV treatment, the M2-1 protein may provide a novel avenue for anti-viral treatment. As with HRSV, BRSV M2-1 is a good candidate for anti-viral drug design due to its essential function as a transcription anti-terminator and limited resemblance to mammalian proteins. This chapter describes the purification of BRSV M2-1, as well as the biochemical, structural and functional analysis of the protein, with relevance to the use of BRSV M2-1 as an anti-viral target.

6.2 Results

6.2.1 Protein purification

Expression and purification of the BRSV M2-1 protein was achieved using a protocol based on that optimised for the HRSV M2-1 protein, as described in chapter 3. For optimal BRSV M2-1 expression and purification this established method was modified in two ways; firstly bacterial cells (BL21 Gold) were induced with less IPTG (0.1 M) than for HRSV M2-1 (0.34 M) and secondly, after elution from GST resin the ion exchange chromatography step used for HRSV M2-1 purification was omitted. BRSV M2-1 purified with few contaminants when eluted from the GST resin (Figure 80- lane 6) and had an apparent molecular weight of ~ 24 kDa. Once purified, a 15 g bacterial cell pellet yielded ~150 mg high purity BRSV M2-1 protein.
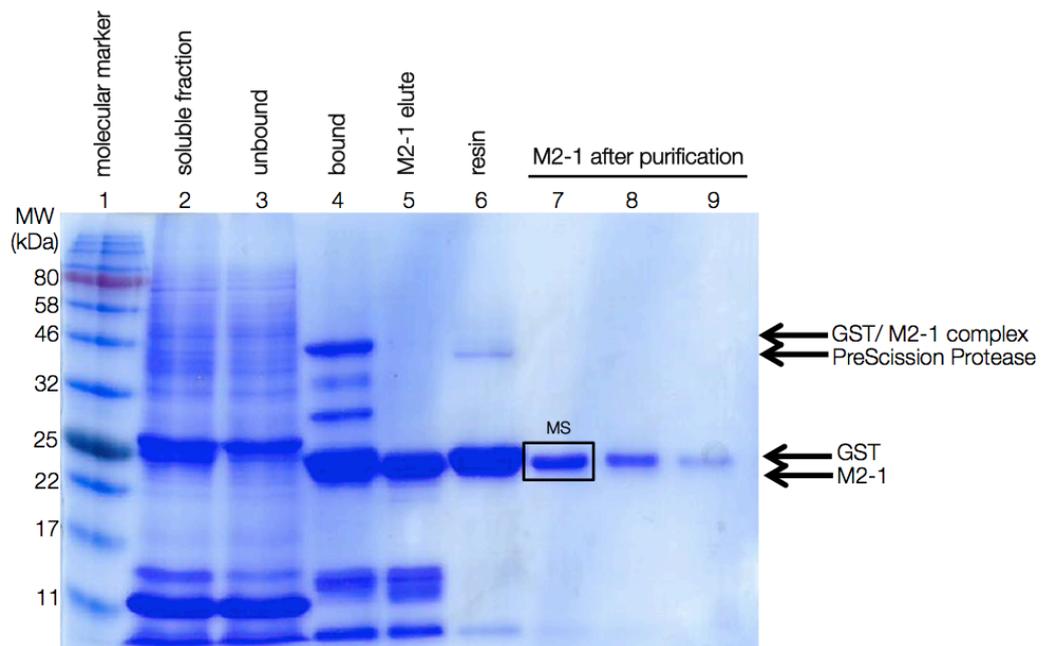
Figure 80 BRSV fraction analysis by denaturing SDS-PAGE

SDS PAGE (15%) stained with Coomassie blue. Lane 1, molecular weight markers; lane 2, 200 ml of clarified lysate applied to GST resin; lane 3, unbound protein flow through; lane 4, protein bound to GST resin; lane 5, M2-1 elute flow-through after proteolytic cleavage; lane 6, resin with remaining GST tag bound after proteolytic cleavage; lanes 7-9- M2-1 after size exclusion chromatography purification. M2-1 has an apparent molecular weight of 24 kDa by SDS-PAGE analysis. Black box labelled MS indicates gel sample used for mass spectrometry analysis to confirm protein identity (see section 6.2.3).

6.2.2 Size exclusion chromatography

Size exclusion chromatography (SEC) was performed to separate the BRSV M2-1 species from unwanted by-products of purification based on their relative differences in molecular mass. A calibration curve using proteins of known molecular masses allowed the elution volume to be used as an indication of the molecular mass of BRSV M2-1. BRSV M2-1 eluted at a single peak (120-135 ml) on a Superdex 75 column (Figure 81A). The protein peak correlated with the peak seen for HRSV M2-1 with an apparent molecular weight of around 85 kDa. The peak most likely represented tetrameric M2-1 (EMBL-pepcore- chromatography calibration). No other peaks with absorbance at 280 nm were seen suggesting no major contaminant proteins were co-eluted with M2-1. This was confirmed by SDS-PAGE (Figure 81B), which revealed a major eluted protein, with an apparent molecular weight of 24 kDa, assumed as M2-1, with low abundance of additional contaminating proteins (Figure 81B lanes 3-9). The absence of absorbance in the void volume (V_0) indicates the protein did not aggregate during the purification process.

6.2.3 Molecular mass and sequence identification

The identity and molecular mass of BRSV M2-1 was confirmed by mass spectrometry. A Coomassie stained gel slice containing a protein band (Figure 80) believed to represent M2-1 was submitted to the mass spectrometry department, University of Leeds for analysis. The molecular mass of the protein was determined to be 21762.12 ± 0.03 Da (Figure 82A). This is within 0.68 Da of the molecular mass predicted by ExPasy (21762.80 Da). Further to molecular mass analysis the protein was trypsin digested and submitted to positive ionisation mass spectrometry to allow sequence analysis of the digested peptides. The protein was identified as BRSV M2-1 and peptides determined by mass spectrometry are indicated by a blue line under the sequence of Figure 82B. Peptides were identified across the full-length protein with sequencing of both the NTD and CTD demonstrating no degradation occurred during the purification process.
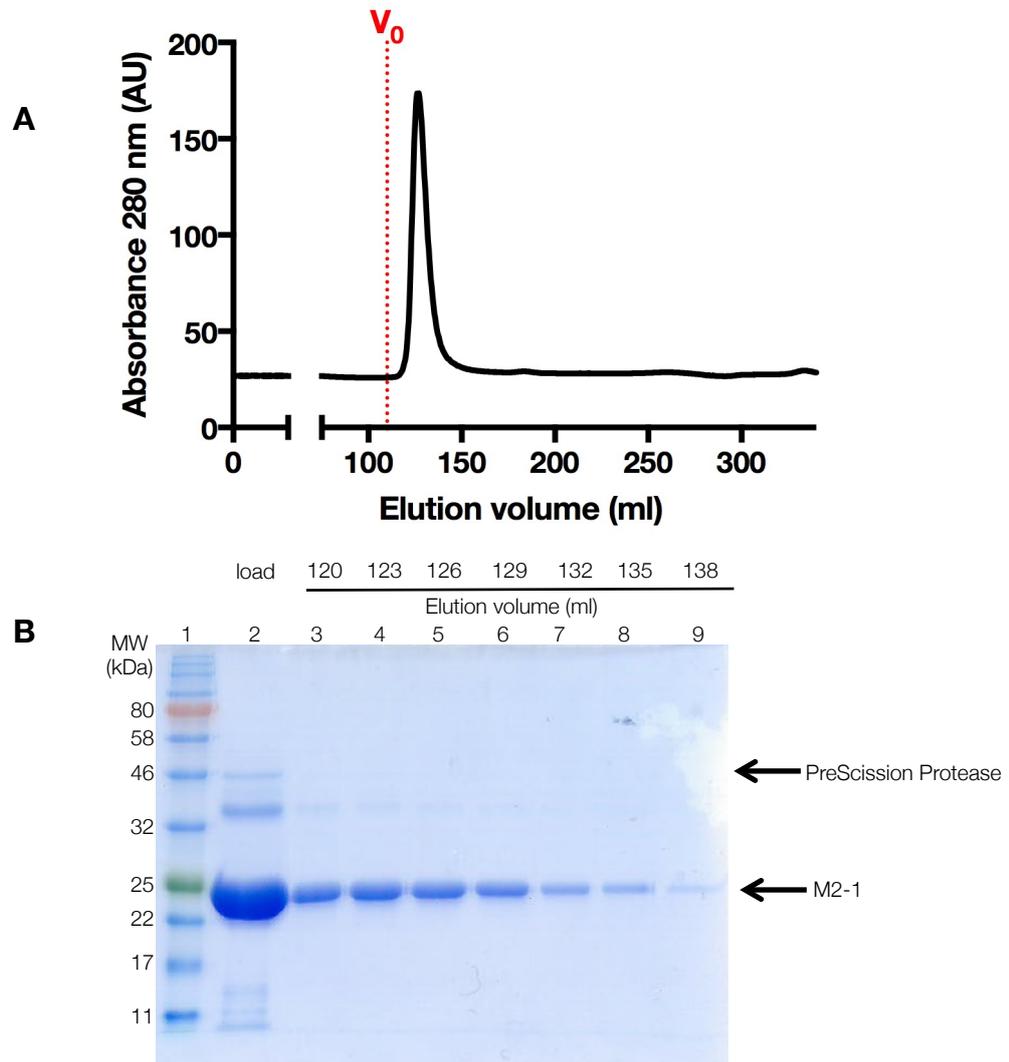
Figure 81 Size exclusion chromatography of BRSV M2-1

A- size exclusion chromatography was performed using a HiLoad 26/60 Superdex 75 pg column. V0 indicated the void volume. One peak is seen at 280 nm for BRSV M2-1 at an elution volume of 120-135 ml; B- SDS PAGE (15%) stained with Coomassie blue. Lane 1- protein marker; lane 2- protein loaded on the SEC column; lanes 3-9 are the collected fractions from SEC throughout the major peak, with elution volumes of each fraction shown above the corresponding lanes.
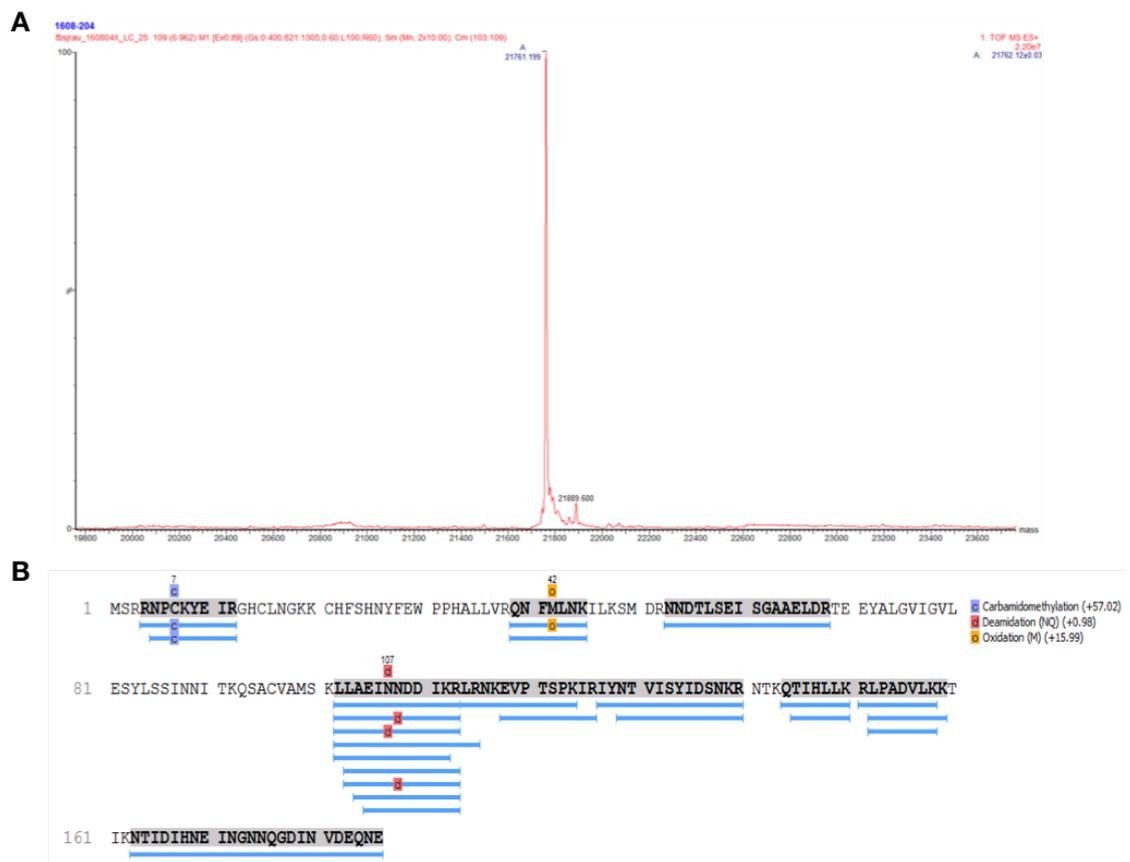
Figure 82 Mass Spectrometry analysis of BRSV M2-1

A- molecular mass of BRSV M2-1; B- the digested protein was submitted to MS/ MS analysis to determine the peptide sequence and confirm the protein was BRSV M2-1 by comparing with a protein database.

6.2.4 BRSV M2-1 forms a tetramer in solution

In addition to SEC that determined the apparent molecular weight of BRSV M2-1 in solution to be ~ 85 kDa indicative of a tetramer (Figure 81) a gluteraldehyde cross linking assay was performed to further examine the oligomeric state of M2-1 protein species by indiscriminate covalent attachment of adjacent protein molecules. BRSV M2-1 was incubated with varying concentrations of gluteraldehyde and analysed on a 12% SDS-PAGE gel (Figure 83). After gluteraldehyde addition, four distinct protein bands can be visualised with the highest protein band representing non specific aggregates. The ability of M2-1 to multimerise into its native tetrameric form under these conditions suggests the protein is correctly folded. At higher concentrations of gluteraldehyde above 10 mM (lane 7) the stained bands become less intense, likely due to the increased mass of cross-linked protein species that were too large to enter the resolving gel

matrix, and which remained at the stacking/ resolving gel boundary. By comparison with protein standards, the molecular mass of the four most predominant M2-1 protein bands were estimated to be approximately 24, 40, 65 and 150 kDa. The size discrepancies of these oligomeric species compared to their predicted masses may be due to the non-globular shape of HRSV M2-1, or inaccuracies in mass estimation using the protein standards. The four distinct protein bands seen by gluteraldehyde cross linking further support the suggestion that BRSV M2-1 is a tetrameric protein species in solution.
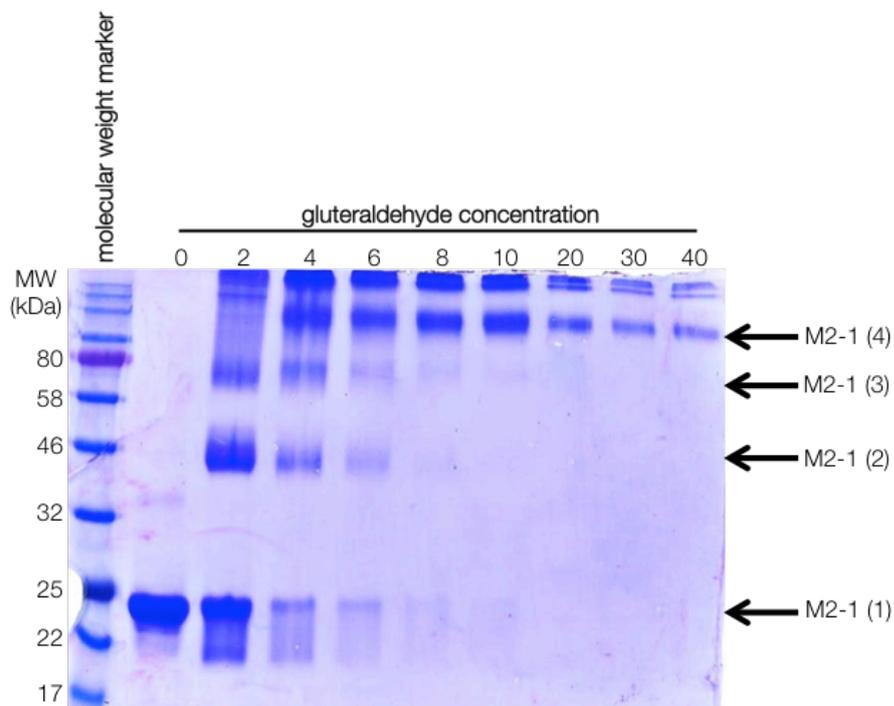
Figure 83 Cross linking of BRSV M2-1

SDS PAGE (15%) stained with Coomassie blue. BRSV M2-1 protein was incubated with gluteraldehyde of varying concentrations from 0- 40 mM and incubated for 30 minutes before stopping the reaction by incubation at 95oC. Samples were run on a 12% SDS-PAGE gel and stained with Coomassie. Four distinct protein bands were visualised, labelled M2-1 (1) – (4). Lane 1- protein molecular marker; lanes 2-10 had increasing concentrations of gluteraldehyde of 0, 2, 4, 6, 8, 10, 20, 30 and 40 mM respectively.

6.2.5 Secondary structure analysis of BRSV M2-1

Circular dichroism (CD) was used to assess the secondary structure characteristics and thermal stability of BRSV M2-1 in sodium phosphate buffer, at pH 7.0. A CD spectrum at 5°C shows minima at 208 and 222 nm (Figure 84A black) characteristic of proteins comprising predominantly alpha-helical secondary structure, and confirms the protein is folded in solution. CD spectra was collected for HRSV M2-1 and showed the same minima at 208 and 222 nm (Figure 84A purple). The flat baseline signal indicates good buffer matching. The CD signal was recorded at a single wavelength of 222 nm to monitor the dissipation of secondary structure as a function of temperature. BRSV M2-1 had an approximate melting temperature (Tm), where 50% of the protein is assumed unfolded, of 50°C (Figure 84B). HRSV M2-1 (not shown) had an approximate melting temperature of 56°C at 222 nm. Grey data points (Figure 84B) at high temperatures were not fitted to deduce the melting temperature as they did not plateau, and is indicative of protein aggregation.

A

B
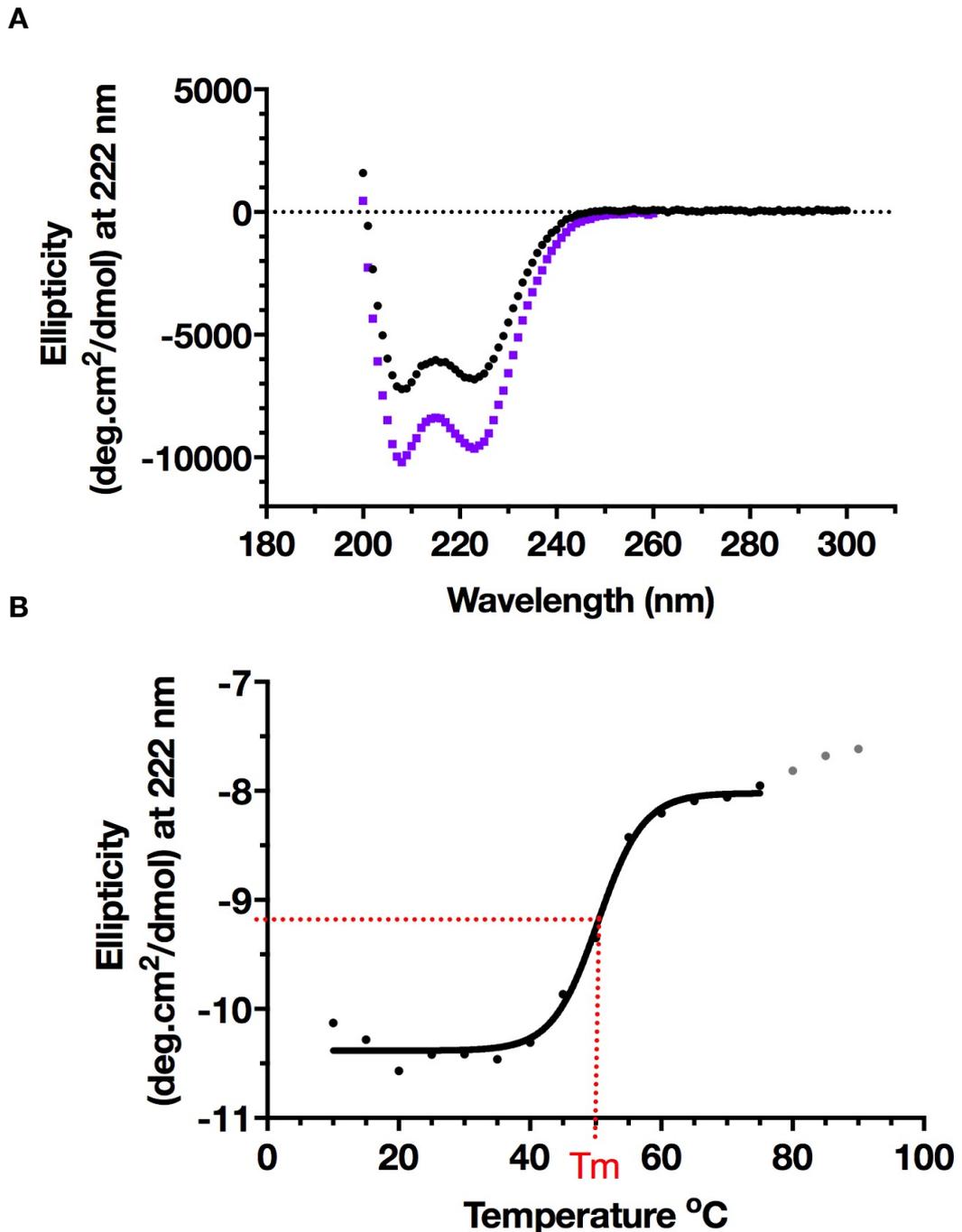
Figure 84 Circular dichroism analysis of BRSV M2-1

Spectra collected on purified M2-1 protein in phosphate buffer, pH 7.0. Triplicate reading were recorded and the spectra normalised using the equation stated in chapter 2; section 2.2.5.2. A- CD spectrum (180 nm- 300 nm) collected at 5oC for BRSV M2-1 ( ) and HRSV M2-1 (n) reveals minima at 208 and 222 nm; B- spectra was collected at 222 nm as a function of temperature and an arbitrary sigmoidal curve fitted to the black data points to determine an approximate protein melting temperature (Tm-red line) for BRSV M2-1 (50oC).

6.2.6 RNA and Phosphoprotein (P) binding to BRSV M2-1

The HRSV M2-1 protein has been shown to interact with various RNA sequences (Tanner et al., 2014; Chapter 3) as well as a short peptide representing residues 90-110 of the HRSV P protein. To assess whether BRSV M2-1 also has the ability to interact with RNA and P, we used fluorescence anisotropy to examine its ability to interact with the same RNA sequences and P peptide bound by HRSV M2-1 (Figure 85). Binding affinities stated here, as in chapter 3, were determined from the binding curves, however the binding could not be saturated at the concentrations of M2-1 used and thus the affinities may be referred to as 'apparent' Kd values. This analysis showed that BRSV M2-1 was able to bind RNA sequences representing both viral and non-viral sequences with binding affinities (Kd values) that were similar to those determined for HRSV M2-1 (Table 25). Furthermore, the relative apparent binding affinities of the panel of RNAs derived for BRSV M2-1 closely matched that determined for HRSV M2-1, with highest affinities consistently displayed for RNAs with a high A-content (HRSV M2-1 affinities are discussed in more detail in the results section of chapter 3). The HRSV P peptide (residues 90-110) bound to both HRSV and BRSV M2-1 with Kds of 7.5 µM and 19 µM, respectively. The lower affinity exhibited by BRSV M2-1 is likely due to the P peptide sequence being HRSV-specific, with only 15 out of 21 residues identical to the corresponding peptide from BRSV M2-1.
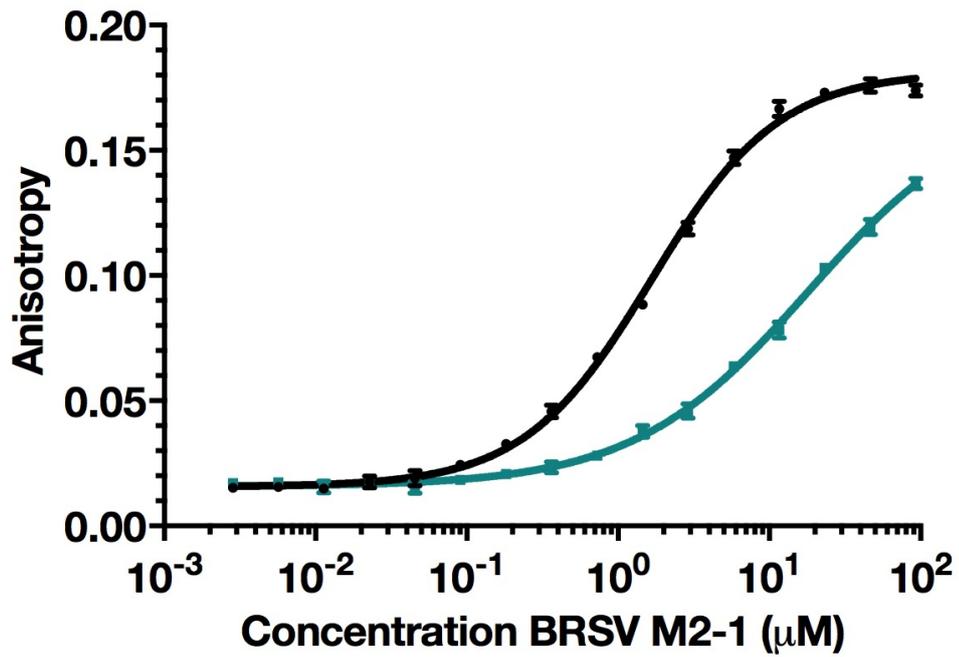
Figure 85 Binding interactions of BRSV M2-1 with RNA and P

A representative curve illustrating titration of BRSV M2-1 against polyA13mer ( ) and P90-110 (n). All vRNA and P sequence binding affinities are reported in Table 25. Data points represent n-3 ± SEM with each n number performed in triplicate.

RNA	Sequence	BRSV apparent Kd ± SEM (µM)	HRSV apparent Kd ± SEM (µM)
SH GE (+)	AGUUAAUUAAAAA	2.3 ± 0.11	2.3 ± 0.39
SH GE (-)	UUUUUAAUUAACA	4.3 ± 0.25	7.5 ± 0.41
F GE (+)	AGUUAUAUAAAAC	3.7 ± 0.20	5.6 ± 0.44
F GE (-)	GUUUUAUAUAACU	4.7 ± 0.48	17 ± 1.5
NS1 GE (+)	AGUUAAUAUAAAA	2.3 ± 0.12	3.7 ± 0.35
NS1 GE (-)	UUUUUAUAUUAACU	18 ± 5.1	8.1 ± 0.65
F GS (-)	UGUUAUUUGCCCC	7.2 ± 1.3	16 ± 1.6
polyA	AAAAAAAAAAAAA	1.7 ± 0.06	1.1 ± 0.12
Peptide			
P90-110	DPTPSDNPFSKLYKETIETFD	18 ± 2.0	7.5 ± 7.7

Table 25 Binding affinities (Kd) of BRSV M2-1 and HRSV M2-1 with RNA sequences and P90-110

Binding affinities were determined by fluorescent anisotropy for each RNA 13mer and HRSV P90-110 peptide. Values represent an n-3 ± SEM with each n number performed in triplicate. RNA nucleotide sequences are stated, as is the amino acid sequence of P90-110. (+) indicates an mRNA sense and (-) indicates a genome sense oligonucleotide.

6.2.7 Crystal trials of BRSV M2-1

The results of the previous sections showed that the BRSV M2-1 protein was purified to a high level of homogeneity, as well as being correctly folded and functional in ligand binding assays. These data raised the strong possibility that the protein may crystallise and so BRSV M2-1 crystal trials were performed using the JCGS Core Suite commercial screens. Crystals were only seen in 3 conditions: (1) 0.1 M MES pH 6.5, 12% PEG 20,000, (2) 0.1 M MES pH 5.0, 10% PEG 6,000 and (3) 0.1M Imidazole pH 8, 10% PEG 8,000 at a protein concentration of 10 mg/ ml. The crystals (Figure 86) had a thin plate morphology and were difficult to harvest. To facilitate the growth of sturdier crystals for X-ray diffraction a fine-screen of pH and PEG concentrations was performed (Table 26) for the three crystal conditions. In addition an additive screen (Molecular Dimensions) of 96 conditions was added to the three crystal conditions. Further, the mother liquor: protein ratios and speed of mother liquor: protein drop formation were altered between the Mosquito Liquid handler and Formulatrix NT8 liquid handler for the three initial conditions. Optimisation of crystal conditions did not yield further crystals for harvesting and the initial generated crystals (Figure 86) were tested for their ability to diffract X-rays.
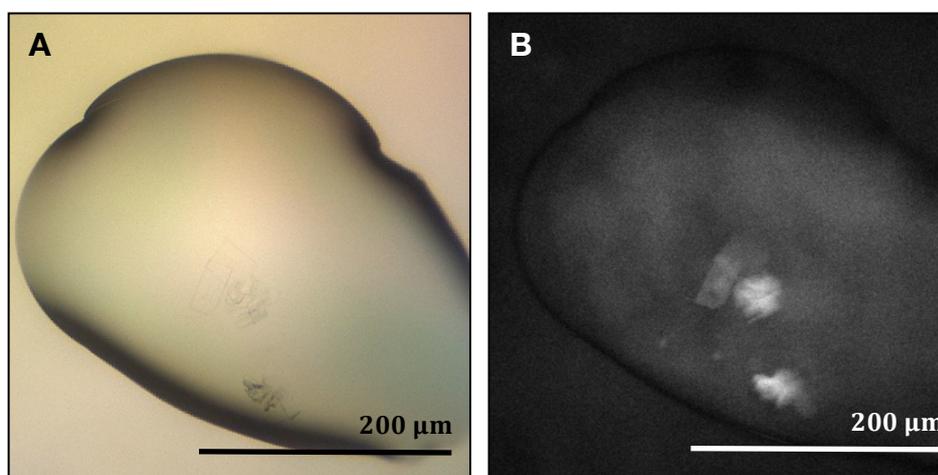
Figure 86 BRSV M2-1 crystals

The Mosquito liquid handler dispensed a 1:1 ratio of mother liquor to M2-1 protein. Crystal shown formed in 0.1M imidazole pH 8, 10% PEG 8,000. A- Liquid drop with 10x magnification; B- UV-TPEF imaging. Crystals were visualised using the RockMaker imager.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.1M MES pH 6 8% PEG 2,000	0.1M MES pH 6 10% PEG 2,000	0.1M MES pH 6 12% PEG 2,000	0.1M MES pH 6 15% PEG 2,000	0.1M MES pH 6 18% PEG 2,000		0.1M imidazole pH 7.5 5% PEG 8,000	0.1M imidazole pH 7.5 10% PEG 8,000	0.1M imidazole pH 7.5 15% PEG 8,000	0.1M imidazole pH 7.5 20% PEG 8,000	0.1M imidazole pH 7.5 25% PEG 8,000	
B	0.1M MES pH 6.5 8% PEG 2,000	0.1M MES pH 6.5 10% PEG 2,000	0.1M MES pH 6.5 12% PEG 2,000	0.1M MES pH 6.5 15% PEG 2,000	0.1M MES pH 6.5 18% PEG 2,000		0.1M imidazole pH 8 5% PEG 8,000	0.1M imidazole pH 8 10% PEG 8,000	0.1M imidazole pH 8 15% PEG 8,000	0.1M imidazole pH 8 20% PEG 8,000	0.1M imidazole pH 8 25% PEG 8,000	
C	0.1M MES pH 7 8% PEG 2,000	0.1M MES pH 7 10% PEG 2,000	0.1M MES pH 7 12% PEG 2,000	0.1M MES pH 7 15% PEG 2,000	0.1M MES pH 7 18% PEG 2,000		0.1M imidazole pH 8.5 5% PEG 8,000	0.1M imidazole pH 8.5 10% PEG 8,000	0.1M imidazole pH 8.5 15% PEG 8,000	0.1M imidazole pH 8.5 20% PEG 8,000	0.1M imidazole pH 8.5 25% PEG 8,000	
D	0.1M MES pH 4.5 5% PEG 6,000	0.1M MES pH 4.5 10% PEG 6,000	0.1M MES pH 4.5 15% PEG 6,000	0.1M MES pH 4.5 20% PEG 6,000	0.1M MES pH 4.5 25% PEG 6,000							
E	0.1M MES pH 5 5% PEG 6,000	0.1M MES pH 5 10% PEG 6,000	0.1M MES pH 5 15% PEG 6,000	0.1M MES pH 5 20% PEG 6,000	0.1M MES pH 5 25% PEG 6,000							
F	0.1M MES pH 5.5 5% PEG 6,000	0.1M MES pH 5.5 10% PEG 6,000	0.1M MES pH 5.5 15% PEG 6,000	0.1M MES pH 5.5 20% PEG 6,000	0.1M MES pH 5.5 25% PEG 6,000							

Table 26 BRSV M2-1 fine crystal screen conditions

Each crystal condition was set up using various concentrations of PEG and pH using 15 wells of a 96 well sitting drop plate. Across a row PEG was increased and down each column pH was increased at 0.5 unit intervals. Plates were set up using the Formulatrix NT8 liquid handler at 0.2: 0.2 µl drops.

6.2.8 Data collection

Crystals harvested from the initial three conditions were cryo-protected and screened on the I03 beamline at the DLS (chapter 2; section 2.2.3). Three diffraction images were collected for each sample to assess the initial diffraction quality of the crystal. Diffraction images contained relatively few diffraction spots (best resolution spot ~ 6 Å for crystal condition 0.1M MES pH 6.5 and 12% PEG 20,000 with 5% glycerol, 5% MPD, 5% glycerol and 5% PEG 400 as cryo-protectant) and therefore the crystals were not suitable for high resolution structure determination of BRSV M2-1, however the cryo condition used was seen to prevent ice formation.

6.2.9 Computational modelling of BRSV M2-1

At this time, the atomic resolution structure of BRSV M2-1 by X-ray crystallography could not be determined. As an alternative, the 3D structure of BRSV M2-1 was modelled using the I-TASSER server (Zhang, 2008; Yang et al., 2015), which gave a likely structural model based on the HRSV M2-1 crystal structure. The modelled protein is structurally similar to HRSV M2-1 (Figure 87), which was predicted from similarities in amino acid identity, biochemical and functional analysis. Non-conserved residues between HRSV and BRSV from the pairwise alignment in Figure 79 were mapped on to the HRSV M2-1 structure (Figure 88A –non-conserved residues in black). Residues not conserved between the two proteins are primarily on the external surface of the tetramer involved in crystal contacts (Figure 88B). These contacts are required for the formation of a regular lattice to produce crystals and could be one of many reasons that BRSV M2-1 did not crystallise.
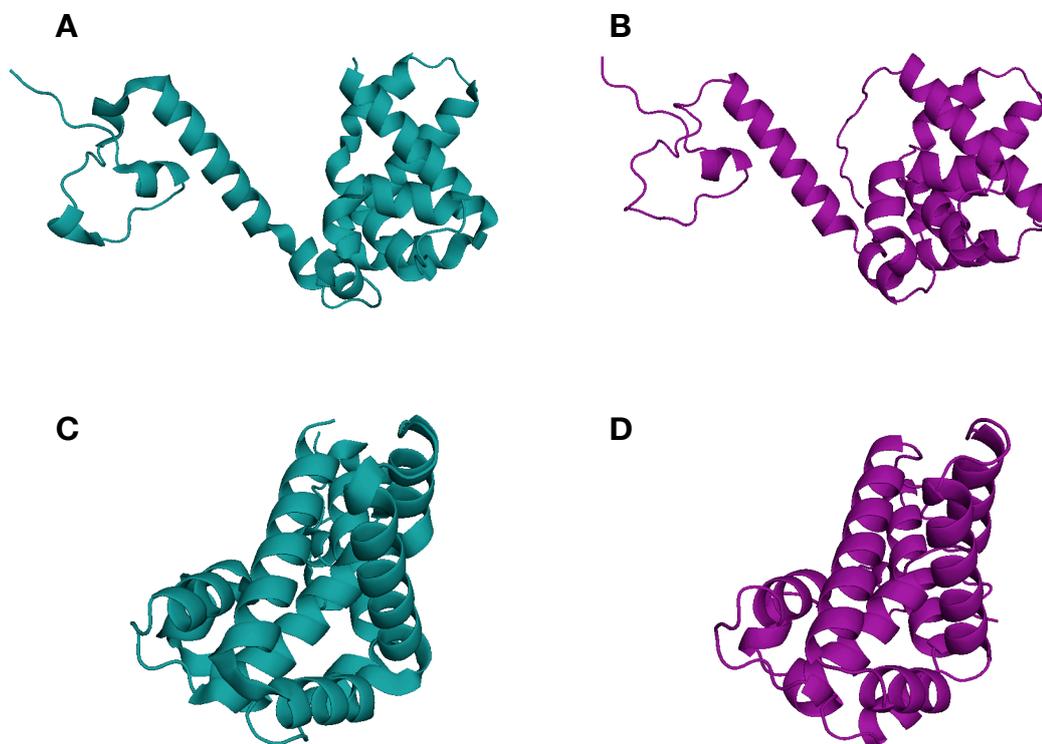
Figure 87 Comparison of the HRSV M2-1 crystal structure and the BRSV M2-1 modelled structure

A- crystal structure of HRSV M2-1 monomer (pdb: 4c3b); B- BRSV M2-1 monomer modelled by the I-TASSER server and based on the HRSV protein; C- HRSV M2-1 and D- BRSV M2-1 are views of A and B, turned 90°.

6.2.10 The RNA/ P interaction site of BRSV M2-1

With over 80% sequence identity, similarities in secondary structure (CD) and binding affinities, the modelled BRSV and crystal HRSV M2-1 RNA/ P binding sites were further analysed. Amino acid residues involved in RNA and P binding in HRSV M2-1 are conserved in BRSV M2-1 (Figure 89A) and exhibit a similar, but not identical arrangement around a central cavity in the BRSV M2-1 modelled tertiary structure (Figure 89B), however the modelled cavity for BRSV appears shallower than the HRSV crystal structure cavity.
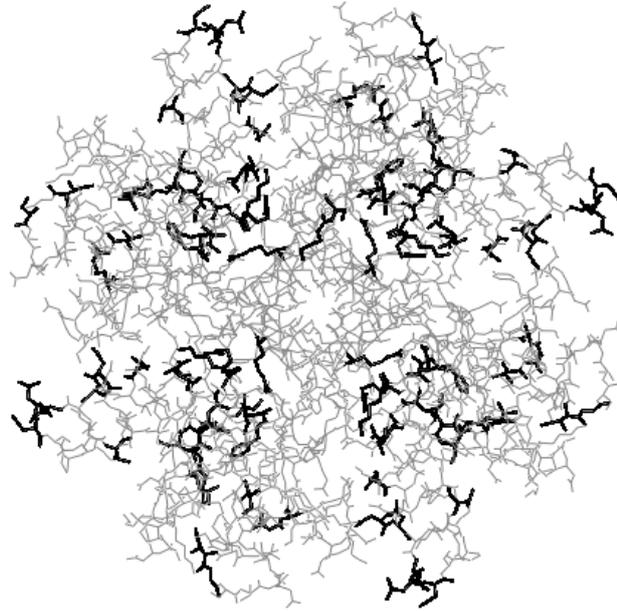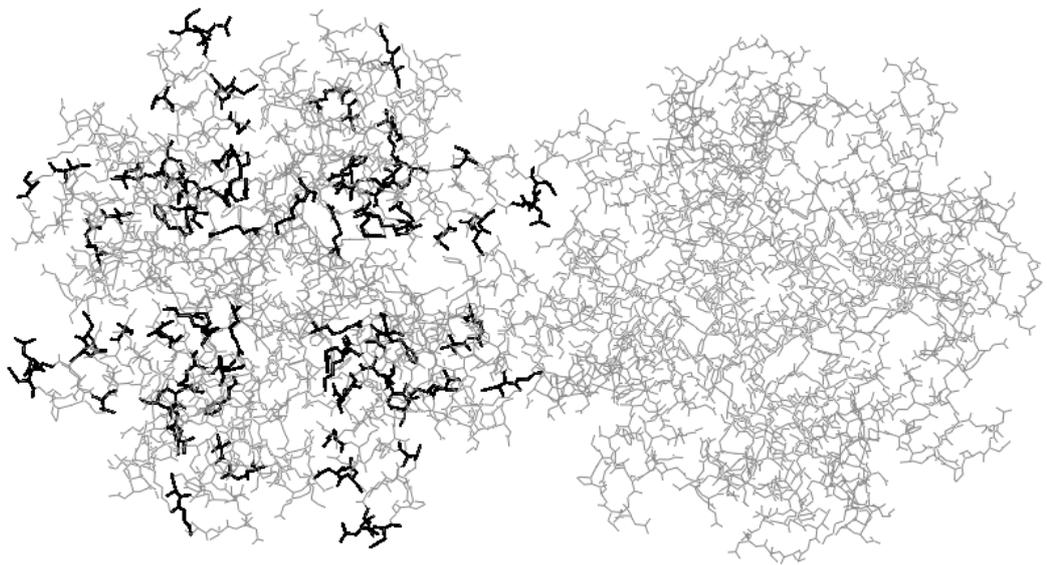
Figure 88 Non-conserved HRSV and BRSV M2-1 residues

A- differences in HRSV and BRSV amino acid identity are highlighted on HRSV tetrameric M2-1 in black on all four monomers; B two HRSV tetramers displaying the non conserved residues involved in crystal contacts

A

```
HRSV MSRRNPCKFEIRGHCLNGKRCHFSHNYFEWPPHALLVRQNFMLNRILKSM
     ||||||||*|||||||||||*|||||||||||||||||||||||*|||||
BRSV MSRRNPCKYEIRGHCLNGKKCHFSHNYFEWPPHALLVRQNFMLNKILKSM

HRSV DKSIDTLSEISGAAELDRTEEYALGVVGVLESYIGSINNITKQSACVAMS
     |**||||||||||||||||||||||||*||||||*|||||||||||||||
BRSV DRNNDTLSEISGAAELDRTEEYALGVIGVLESYLSSINNITKQSACVAMS

HRSV KLLTELNSDDIKKLRDNEELNSPKIRVYNTVISYIESNRKNNKQTIHLLK
     |||||*|*||||*||*||||||*|||||||||*||**||||||||
BRSV KLLAEINNDDIKRLRNKEVPTSPKIRIYNTVISYIDSNKRNTKQTIHLLK

HRSV RLPADVLKKTIKNTLDIHKSITINNPKESTVSDTNDHAKNNDTT
     ||||||||||||||*||||||||*||*||*|*
BRSV RLPADVLKKTIKNTIDIHNEINGNNQGDINVDEQNE
```

Key: RNA binding residues, P binding residues and RNA and P

B **HRSV M2-1**
Crystal Structure **BRSV M2-1**
Modelled Structure

Figure 89 RNA and P binding residues conserved in HRSV and BRSV M2-1

RNA binding residues (pink), phosphoprotein binding residues (yellow) and residues involved in both RNA and P binding are shown in coral. A- local alignment was performed using the Smith-Waterman algorithm. RNA and P binding residues are conserved between BRSV and HRSV M2-1 denoted by a '|', '*' represent a substituted residue with similar functional characteristics. Where no symbol is shown, this indicates a substituted residue with no similarities; B- I-TASSER modelling of the BRSV M2-1 (purple monomer) protein based on the HRSV M2-1 crystal structure (PDB: 4c3b- blue monomer).

6.3 Discussion

6.3.1 Expression and purification of BRSV M2-1 protein

Our first attempt at expression and purification of BRSV M2-1 utilised the previously optimised method used for HRSV M2-1, on account of the high sequence identity of these proteins, which suggested a similar purification strategy would be effective. However, this method yielded protein that was highly aggregated and poorly active, possibly as a consequence of higher overexpression compared to HRSV M2-1. Reducing the concentration of IPTG used to induce protein expression resulted in reduced over-expression, but higher yields of non-aggregated and active protein (~150 mg / 15 g bacterial cell pellet). BRSV M2-1 was not present in the void volume (V_0) during SEC and showed that the buffers and purification procedure used resulted in stable, non-aggregated protein. High purity was seen for both HRSV and BRSV M2-1 proteins, determined by SEC UV 280 nm reading and mass spectrometry following SDS-PAGE analysis. However, both proteins are susceptible to degradation over time due to proteases encoded by the expression host. Performing cation exchange on BRSV M2-1 resulted in the identification of multiple bands by SDS-PAGE that were of a lower molecular weight than 24 kDa, the molecular weight of M2-1. Mass spectrometry protein sequence analysis determine these bands to be cleaved BRSV M2-1 sequences. Purifying BRSV M2-1 by SEC only, resulted in purification of a protein with little degradation, as determined by SDS-PAGE analysis. Performing SEC for BRSV M2-1 resulted in a protein that was buffer exchanged as achieved for HRSV M2-1 and allowed direct comparisons to be made in future assays.

6.3.2 BRSV M2-1 preferentially binds A-rich RNA sequences

BRSV M2-1 displayed an affinity for A-rich tracts in RNA oligonucleotides and had the highest binding affinity for the pA13 mer and A-rich SH (+) GE RNA, that refers to the genome sense sequence. The preferential affinity for A-rich sequences has been observed for both HRSV M2-1 (Blondot et al., 2014, Tanner et al., 2014) and HMPV M2-1 (Leyrat et al., 2014). The RNA and P sequence assayed here were derived from HRSV and not BRSV. The HRSV P90-110 peptide sequence had a lower affinity for BRSV M2-1 than HRSV M2-1, and this is not surprising given that the HRSV and BRSV P protein sequences are not identical in the region represented by the P 90-110 peptide, with differences between these two species located entirely at the peptide N-terminus (Figure 90).

A natural extension of this work to allow functional comparisons of the M2-1: P interaction would be to design a BRSV specific P 90-110 peptide and determine its relative binding affinities, however this was not possible here due to time restraints. Previous work (chapter 3; section 3.9.3) revealed for the first time the co-crystal structure of HRSV M2-1 bound to P 90-110. Residues P 97-109 were fitted reliably to the density with C-terminal P residues (Tyr 102, Thr 105, Thr 108 and Glu 107) seen to directly interact in the crystal structure with M2-1. Density was poor for the P 90-110 peptide N-terminus and residues could not be fitted, thus no direct conclusion could be drawn of the importance of these residues for interactions with HRSV M2-1.

Extending the P 90-110 peptide from the available density would result in the possibility of interaction of the N-terminus P90-110 residues with the M2-1 protein thus it could be suggested that BRSV M2-1 binds with lower affinity to HRSV P 90-110 than HRSV M2-1 due to species specific residues in 90-100 of the P protein. Alternatively, the SEM for the Kd of HRSV M2-1 is much higher (7.7 µM) than for BRSV M2-1 (2.0 µM) and affinities may be largely similar with the same conserved residues in P (Tyr 102, Thr 105, Thr 108 and Glu 107) critical for binding to HRSV and BRSV M2-1.

```
BRSV  P  90-110   ELPSSENPFTKLYKETIETFD
                    *     |*|||  |||||||||||
HRSV  P  90-110   DPTPSDNPFSKLYKETIETFD
```

Figure 90 Sequence identity of HRSV and BRSV P 90-110

Alignment of the phosphoprotein sequences of HRSV and BRSV represented by synthetic peptide P90-110. The symbol '|' indicates identical residues and '*' indicates residue similarity. Where no symbol is shown, this indicates a substituted residue with no similarities.

6.3.3 Missing structural information

With the biophysical and functional data suggesting BRSV M2-1 forms a stable tetrameric protein in solution with similar binding characteristics for RNA and P as previously described for HRSV M2-1, it was reasoned that the protein may crystallise in a similar manner to HRSV M2-1 and that the HRSV M2-1 model could be used for molecular replacement even with low resolution data. However, BRSV M2-1 crystals did not yield any useful diffraction data for structure determination. Figure 88 shows the non-conserved residues of BRSV, displayed

on the HRSV M2-1 crystal structure. The non-conserved residues are on the surface of the protein and are primarily involved in forming crystal contacts between neighbouring tetramers. The alteration in the residues involved in forming crystal contacts could result in the protein not being able to form a regular crystal lattice and may explain why very few and only thin fragile crystals were visualised. However, to aid RNA crystallisation in complex with M2-1, residues hypothesised to be involved in crystal contacts were removed ΔLGSMS and ΔLGS (chapter 3; section 3.8.3), or point mutations made (M2-1 Asn 89 Ala and Leu 120 Ala; data not shown, personal communication from Dr Kavestri Yegambaram). Although M2-1 was not co-crystallised along with RNA, the mutated HRSV M2-1 constructs still crystallised despite altering the crystal contacts and suggests that differences in BRSV M2-1 crystal contacts may not be the only contribution to the lack of crystals seen. The CD spectrum showed both HRSV and BRSV M2-1 have a high alpha helical content and the oligomeric state of the protein suggests a correctly folded protein. However, the signal intensities and melting temperature are lower for BRSV, characteristic of a more flexible protein. This instability could also contribute towards lack of crystal formation.

Despite using numerous commercial screens and optimisation methods, high-resolution diffracting crystals were not observed. Additive screens and optimisation of PEG and pH conditions did not produce crystals sturdier than the initial commercial screens. Crystals formed at 18 and 20℃, and one possibility to improve crystal quality and yield could be to further slow the rate of crystal formation at a lower temperature to provide sturdier crystals.

6.3.4 Comparison of BRSV and HRSV M2-1 proteins.

For the HRSV and BRSV M2-1 proteins, at the biochemical level we see differences in structure and stability of the proteins by CD spectrum analysis but conservation in function and functional residues through local sequence alignment and fluorescence anisotropy assays. HRSV and BRSV M2-1 share a high sequence identity particularly of hydrophobic residues that are critical for the formation of the protein core structure. In addition, the crystal structure of HMPV M2-1 (Leyrat et al., 2015) is structurally similar to that of HRSV M2-1 (Tanner et al., 2014). In comparison to HMPV M2-1, BRSV M2-1 is more closely related with HRSV M2-1 and the predicted BRSV M2-1 structure supports the data seen here to allow more confident conclusions to be drawn of the similarities between the two M2-1 proteins. RNA and P bind to BRSV M2-1 with similar affinities to HRSV and the conservation of RNA and P binding residues in addition to the predicted

structure suggest a similar binding and function of HRSV and BRSV M2-1 with RNA and P.

6.3.5 The use of BRSV M2-1 for anti-viral drug design

Unlike HRSV, modified live vaccines (Delforge and Zygraich 1978; Bohlender, 1984) and inactivated vaccines (Howard et al., 1987) have been available for the treatment of BRSV since the late 1970s. However field studies (reviewed in Ellis et al., 2016) displayed varying levels of efficacy of these vaccine with an upper limit of 4 months immunity against BRSV in calves (Ellis et al., 2013; Taylor et al., 2014). With this low level of protection and the large costs to the beef and cattle industry in mortality and milk yields (in dairy cows) there is an urgent unmet need for a BRSV treatment (Ferguson et al., 1997; Sacco et al., 2013).

Ideally, and as seen for the prevention of seasonal influenza, an efficacious vaccine would be routinely used, however in non-vaccinated individuals, or following infection with a new strain, anti-viral compounds could be administered when symptoms are evident, prior to the development of a more specific vaccine. Currently the BRSV vaccines are not efficacious enough and vaccine enhanced disease severity remains the critical issue in BRSV (and HRSV) vaccine development with issues being seen for both modified live BRSV vaccines (Kimman et al., 1989) and inactivated BRSV vaccines (Schreiber et al., 2000). To overcome the enhanced disease severity associated with BRSV vaccination and to allow the treatment of calves already displaying respiratory symptoms characteristic of BRSV, there is an urgent and unmet need for an anti-viral treatment. The M2-1 protein of BRSV (and HRSV) is a good target for anti-viral compounds due to its essential function as a transcription anti-terminator and limited resemblance to mammalian proteins.

The primary aim of this research was to determine similarities between the two viral protein species to allow previously determined HRSV M2-1 drug candidates (chapters 4 & 5) to be assayed with BRSV M2-1 and identify possible cross-species activity. Chapters 4 and 5 of this thesis have focussed on the use of the HRSV M2-1 protein for anti-viral drug design using a structure-guided method that yielded hits with a common scaffold and mode of action and a fragment screening method that has shown a direct interaction of fragments and the HRSV M2-1 protein by X-ray crystallography. It was hoped that the crystal structure of BRSV M2-1 would give us further insight into the similarities between the two protein structures and subsequent use for HRSV and BRSV anti-viral design. Despite no X-ray structure being obtained, SEC, FA and modelling data has

allowed comparisons to be made between the two M2-1 proteins. Both proteins have shown high similarities in sequence (Figure 79), oligomeric state (Figure 81) and binding affinities (Figure 85) in addition to the highly similar modelled BRSV M2-1 structure with the crystal structure of HRSV M2-1 (Figure 87). Residues critical for P (co-crystal structure- chapter 3; section 3.9.3) and RNA binding in HRSV M2-1 (SDM- Blondot et al., 2012; Tanner et al., 2014) are conserved in BRSV M2-1, further adding to the high similarity of the proteins and the potential of inhibitors targeted to the RNA/ P binding site to inhibit the anti-termination function of HRSV and BRSV M2-1 proteins. Further, conserved RNA and P residues in the modelled BRSV M2-1 structure formed around a similar central cavity (Figure 89) to that seen for HRSV M2-1 and would suggest a similar mode of binding and inhibition could be achieved for hit compounds targeted towards this cavity by SBDD in HRSV M2-1 (chapter 4).

It is anticipated that the assays discussed in this thesis for SBDD and FBDD (infectious HRSV and X-ray crystallography) and for the study of BRSV (FA and SEC) could be used as an initial starting point for BRSV to assay hit compounds identified for HRSV M2-1. FA displayed (Figure 85) the weakest affinity for HRSV P 90-110 with BRSV M2-1, this weaker interaction than seen for HRSV P 90-110 with HRSV M2-1 may allow out-competition by compounds that could not be determined previously with higher affinity binders. In addition, the BRSV infection model (Hagglund et al., 2014; Blodorn et al., 2014) could be used to assay viral infectivity and test hit compounds like achieved with the HRSV-eGFP infection model (chapter 4; section 4.2.4).

Chapter 7 Discussion

HRSV poses a major global disease burden resulting in LRTI, pneumonia and death of premature infants, the elderly and immunocompromised. In addition, HRSV infection early in life is associated with the development of asthma in adulthood. Mortality due to HRSV largely occurs within developing countries where access to the only HRSV specific approved treatment is limited due to cost. The M2-1 protein is an essential transcription anti terminator that is only expressed by HRSV and other members of the *Pneumoviridae* family. However, only in the *Orthopneumovirus* genus is M2-1 expression required for virus viability. Overall, there is a need for basic research into the function of M2-1 in the viral life cycle, as well as research to determine whether M2-1 represents a valid target for antiviral therapies.

7.1 Further understanding of the interaction between M2-1 and its binding partners, RNA and P

By studying the interactions between the M2-1 protein and RNA we determined M2-1 possessed greatest affinity and thus specificity for A-rich tracts. The central question that we next wanted to ask was what might be the identity of this A-rich RNA sequence in the context of the HRSV life cycle. We reasoned that this information might aid in elucidating the mechanism of M2-1 function. The gene-end sequences of HRSV in the genome complementary polarity (i.e. plus sense anti-genome; cRNA) and also mRNA transcripts are A rich, with the HRSV mRNAs in particular possessing high A content at their 3' ends on account of their extensive polyadenylation (Hardy & Wertz, 1998). However, on analysis of the A/ U composition of the mRNA transcripts, A-rich tracts were seen not only at the GE sequence but also throughout the transcript length. This analysis identified A-rich tracts on average, every 200 nucleotides of the transcribed gene and it is thus suggested here that M2-1 may recognise these A-rich tracts along the length of each transcript to prevent premature intragenic termination and permit full-length mRNA transcript synthesis. It has been demonstrated that short transcripts such as NS1 can be synthesised in the absence of M2-1 and this may be because at around 200 nucleotides in length, NS1 can be fully transcribed without the requirement of A-rich tracts that are required to transcribe lengths above 200 nucleotides (Fearns & Collins, 1999). To assess the hypothesis that A-rich tracts are required on average every 200 nucleotides, an assay that removed the A-rich tracts from the mRNAs whilst preserving the coding sequences could

be performed to potentially identify a decrease in transcriptional activity. Both the anti-genome and mRNA transcripts have A-rich tracts, however it is hypothesised that M2-1 interacts with mRNA transcripts during transcription and not with the anti-genome during replication as during replication the RNA is encapsidated with N protein to form the RNPs and is not as readily accessible.

Direct and competition-binding assays were performed to assess binding of RNA and P to their binding partner, M2-1. A-rich RNA sequences were seen to outcompete the P 90-110 peptide, mirroring findings by Blondot et al (2012). In contrast P 90-110 could not outcompete RNA and suggested that in the viral life cycle M2-1 may preferentially bind to RNA. However, without assessing the full-length P protein and the precise oligonucleotide sequence that binds M2-1 for its anti-termination activity this cannot be concluded with absolute certainty. It is hypothesised here that rather than M2-1 switch between distinct RNA and P binding modes that are mutually exclusive, instead the M2-1 protein may simultaneously bind both RNA and P to alternate M2-1 monomers within the tetramer.

The relative binding affinities of RNA sequences for M2-1 were decreased on binding to a monomeric form of M2-1 that excluded the ZBD, oligomerisation helix and CTD termed, M2-1$_{58-177}$. These findings support those by Tanner et al, (2014) in which SDM work implicated the ZBD in binding RNA in addition to the core domain for full affinity binding. Negative stain EM studies (chapter 3, section 3.8.4) of M2-1 in complex with A-rich RNA oligonucleotides displayed aggregation over multiple tetramers whilst Leyrat et al, (2014) reported M2-1 aggregation on binding to RNA via an increase in radius of gyration via SAXS studies. Taken together, these data suggest interactions of RNA with both the core domain and ZBD of M2-1 and suggest, via visualisation of aggregation, that the interaction of a single RNA oligonucleotide may not be to the ZBD and core domain of an individual M2-1 monomer but occur over multiple tetramers. In contrast to RNA, binding affinities for P 90-110 did not alter between the full length and truncated form of M2-1. In the crystal structure determined here of M2-1 in complex with P 90-110 a 1:1 stoichiometry was observed. However, these studies were in the absence of RNA and one peptide of four in the tetrameric AUC had a high B factor. This may suggest that in fact four P peptides do not bind to tetrameric M2-1 and that in the viral life cycle both RNA and P bind to alternate monomers of the M2-1 tetramer, with RNA requiring additional contacts with M2-1 for full affinity binding that, in the presence of P binding to other M2-1 monomers, prevents the RNA binding over multiple tetramers that

would result in M2-1 aggregation in the viral life cycle that would impede virus function.

The crystal structure of the co-complex displayed P 90-110 bound to the surface of M2-1 with FA analysis illustrating this is a weak interaction. The weak, surface binding of P 90-110 presumably allows an intermittent interaction that would also allow RNA to bind. The surface binding (and not a strong, irreversible binding interaction) is relevant for RNA as the RNA transcript, as it is synthesised would be required to move along M2-1 presumably and this data illustrates that this could be achieved with the surface binding model. The M2-1: P 90-10 crystal structure determined a snap shot of the density at 4 different stages of binding the peptide (4 time point from 4 monomers in the AUC). The highest resolution density allowed the fitting of a helical peptide to the density whilst other density appeared less helical illustrating that M2-1 can bind P 90-110 before it is helical and thus it is suggested here that binding to M2-1 induces the helical conformation of P 90-110 identified here via recognition of a MoRE site. This unstructured to structured hypothesis is consistent with the replication machinery of other *Paramyxoviridae* family members (Communie et al., 2014).

7.2   The function of M2-1 in the viral life cycle

Given that M2-1 has not been identified in any related virus (except a non-essential function in HMPV) and the P protein is considerably shorter in length than all other related viruses, it could be suggested that in HRSV, M2-1 is performing a processivity role (in addition to anti-termination activity) that in other viruses within the *Mononegavirales* order is provided by P. Kiss et al (2014) directly visualized an additional layer of electron density below M, assumed to be M2-1 and implicated M2-1 in virion assembly. Whilst most data points to a role of M2-1 in transcription a role in assembly cannot be discounted. To further assess the hypothesis of a role in assembly, functional regions of M2-1 could be removed by truncations and residue ablations and assessed for the ability of the mutant proteins to interact with RNA/ P or M.

The crystal structure of M2-1 in complex with P 90-110 unequivocally reveals the orientation of both proteins in the M2-1: P complex, and this information increases our understanding of the HRSV transcriptase machinery. Together with previous reports an updated model that describes how M2-1 might function within the entire HRSV transcriptase complex is proposed (Figure 94). In this model M2-1 interacts with the viral transcripts as they are synthesised and would recognise A-rich sequences when they are reached on both intra and intergenic

sequences resulting in full-length mRNA transcripts and polycistronic read-though mRNAs and thus polymerase access to distal genes.
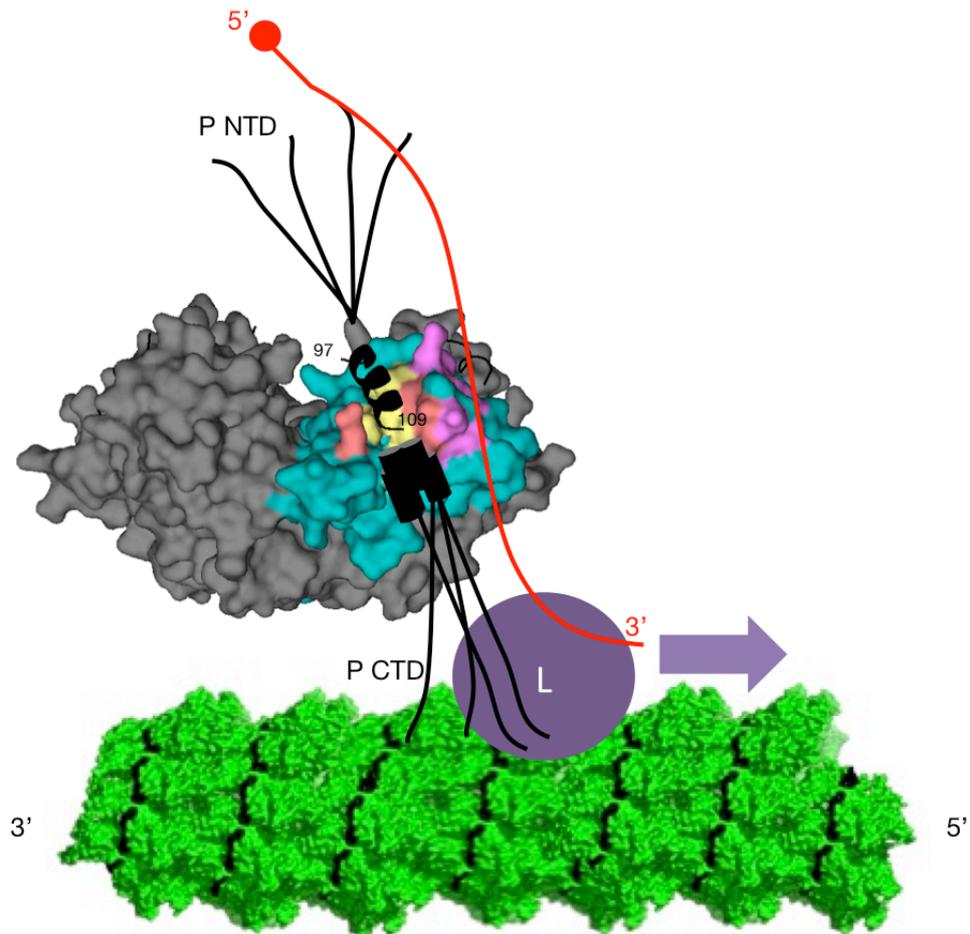
Figure 91 Schematic of viral transcription

P and RNA can bind simultaneously to one M2-1 tetramer. Here a P monomer (of the native tetramer- black) is shown bound to M2-1 from the co-crystal structure whilst an mRNA transcript (red) is shown bound to the adjacent M2-1 monomer. Both M2-1 and P binding to M2-1 is required for efficient anti-termination activity of M2-1 in the viral life cycle.

7.3 Identification of HRSV M2-1 inhibitors

There is a critical need for HRSV inhibitors that can be routinely provided at low costs with no resistance. Palivizumab is the only HRSV-specific FDA approved treatment and despite being efficacious, requires repeat administration and is not amenable due to high costs in developing countries where HRSV associated mortality is high. Approved vaccines exist for BRSV, however they have shown limited efficacy (Taylor et al., 2014) whilst vaccine development for HRSV is on-going but impeded due to previous issues of vaccine enhanced disease severity, which have resulted in poor patience compliance. There is need for an affordable HRSV treatment that does not result in an enhanced immunological response and that could be administered once symptoms are presented. Small molecule anti-viral compounds are viable options for affordable and efficacious treatments that have been developed for other viral infection such as influenza, HCV and HIV. For HRSV specifically, a better understanding of the HRSV life cycle, the window of opportunity for treatment and improved biomarkers have all aided the drug development field.

Here we targeted the M2-1 protein of HRSV for anti-viral drug design as M2-1 exhibits limited sequence homology to mammalian proteins, is highly conserved and could provide cross-species reactivity for the treatment of BRSV. At the time of carrying out this research 2 papers determined compounds that reduce HRSV infection and are presumed to act via M2-1. Compound 10e is assumed to interact with the ZBD that was the target of the study and was derived from known zinc ejector compounds, whilst cyclopamine developed resistance at the RNA/ P interaction site of the M2-1 core domain (Cancielleri et al., 2015; Bailly et al., 2016). Both sites are valid for further drug design; inhibition of M2-1 interactions with RNA and P prevent its anti-termination function (Cuesta et al., 2000) and ejecting the zinc of M2-1 results in non–functional M2-1 protein (Esperante et al., 2013); however, the latter result has not been reproduced by other groups and a direct assay to assess the M2-1 ZBD interaction is currently unavailable. Cellular assays that have been used successfully to identify HRSV inhibitors here could have been used to analyse compounds targeted to the ZBD. However such compounds may not show a direct interaction, and more importantly, the zinc ejector compounds could negatively affect the cell lines used for the assays that require zinc for cell growth and survival, resulting in false positive results.

Due to having an assay (competition FA) that could directly look at inhibition of M2-1 with its binding partners RNA and P we continued with designing inhibitors to a cavity involved in both RNA and P interactions. Unfortunately, the competition assay has only shown competition of P peptides with A-rich RNA oligonucleotides. Fragment and compound libraries in addition to top hits from chapter 4 were assessed in the assay but no out-competition with the P 90-110 peptide was seen. Although the point must be raised that these inhibitors designed to target the RNA/ P surface may not bind to their intended site, a singular hit would have been expected for the libraries that were screened with the competition assay. It is likely that the assay was not sensitive enough for the interaction studied due to the low molecular weight of the compounds in addition to a limit of inhibitory concentrations assessed, due to precipitation. A sensitive assay to determine interactions of inhibitors with M2-1 specifically is paramount for further drug studies.

Chapter 4 resulted in the identification of two compounds that reduced viral infectivity and transcription/ translation using *in cellulo* methods for lead drug discovery. However, a direct interaction with M2-1 could not be ascertained using a competition binding assay or using X-ray crystallography. Chapter 5 took a different approach to drug design and aimed to determine co-crystal structures as a first route to the identification of hits that could then be assayed as inhibitors. This would allow rapid SAR and drug discovery based on the co-crystal structure. Seven hits were identified bound to M2-1 with suitable density to determine the fragments pose. It was interesting to see that hits were visualised for two regions of M2-1 that have both been implicated in drug design, further suggesting a positive use of these regions for future drug design work. Further assessment of the fragment hits as inhibitors and analogue synthesis to determine a lead compound are now required. Due to a lack of knowledge of how the two hit compounds from SBDD interacted with M2-1, limited conclusions can be drawn about their similarities in M2-1 inhibition.

7.4 Similarities between BRSV and HRSV M2-1 proteins

To complement the HRSV M2-1 drug design work, we aimed to determine similarities of the structure and function of HRSV M2-1 with that of BRSV M2-1. BRSV M2-1 formed a tetramer in solution with similarities in secondary structure and binding affinities to both P and RNA. At this time, the BRSV M2-1 atomic structure could not be solved, however all residues involved in RNA/ P interactions are conserved between the two proteins and modelling of the BRSV M2-1 structure suggests a similar arrangement of residues around a central cavity

of residues involved in RNA/ P binding and used for SBDD (chapter 4). The work presented here for BRSV M2-1 ascertained many similarities between the two proteins and it is high likely that HRSV M2-1 inhibitors designed here for the RNA/ P interface could inhibit BRSV infection.

7.5 Future directions

The lack of a sensitive assay to directly determine M2-1 inhibition is currently halting further drug development. An assay is needed to measure the specific interaction affinity, the ability to inhibit and produce IC_{50} curves for further SAR. It has been previously discussed that the fluorescence anisotropy assay may not be sensitive enough for out-competition with compounds or fragments especially if each additional RNA nucleotide or P amino acid results in the combined avidity to M2-1. Further assessment could utilise ITC or SPR to ascertain if these methods are sensitive enough to assess compound and fragment hits determined here. It could still be reasoned that the compound hits (chapter 4) may not have an effect on HRSV infection through interactions with M2-1. Mutational passaging in which the virus can be continuously passaged in the presence of an inhibitor compound until virus levels return to wild type and then sequenced would potentially reveal the site of inhibitor interaction. This approach may tell us whether the negative results by FA and X-ray crystallography are due to the compound not binding to M2-1 or the sensitivity of the FA assay and ability to co-crystallise. With an understanding of the similarities between HRSV and BRSV M2-1 proteins, and the possibility of cross species reactivity of these anti-viral compounds it would be of interest to test hit compounds/ fragments against BRSV M2-1. With a sensitive direct assay this could be done simply, as the protein can be purified in high quantities (chapter 5) whilst a live virus encoding eGFP (as used here for HRSV) has been used by other research groups and could be used here to identity the ability of these compounds/ fragments to inhibit BRSV infectivity (Hagglund et al., 2014).

7.6 Final comments

The results presented here have contributed to understanding the structure and function of HRSV and BRSV M2-1 proteins, their interactions with binding partners RNA and P, and the feasibility of targeting M2-1 for anti-viral drug development. The improved HRSV M2-1 expression purification and crystallisation protocol has resulted in large quantities of pure homogenous protein for material-intensive research. In addition, for further X-ray crystallography

studies we have determined a single crystal condition and morphology that allows repeated sub 3 Å resolution. The co-crystal structure of M2-1: P 90-110 is the first co-structure of a P protein for any *Mononegavirales* order virus and has allowed the direct visualisation of the P interaction site whilst additional functional studies of RNA and P interactions with M2-1 allowed conclusions to be drawn regarding the mechanism of M2-1s action in transcription. Finally, we have identified using various structural and functional methods, two lead-like compounds that are predicted to bind M2-1 through a similar mechanism, and seven fragments that bind directly to the M2-1 protein at two distinct regions; the RNA/ P surface and ZBD. Both compounds and fragments require further assessment for HRSV and it is hoped these experiments can be performed in parallel with assessment of the compounds and fragments for BRSV M2-1.

References

Ackermann, M.R. 2014. Lamb model of respiratory syncytial virus-associated lung disease: Insights to pathogenesis and novel treatments. *ILAR Journal*. **55**(1),pp.4–15.

Ahmadian, G., Randhawa, J.S. and Easton, a J. 2000. Expression of the ORF-2 protein of the human respiratory syncytial virus M2 gene is initiated by a ribosomal termination-dependent reinitiation mechanism. *The EMBO Journal*. **19**(11),pp.2681–9.

Alansari, H. and Potgieter, L.N.D. 1994. Molecular cloning and sequence analysis of the phosphoprotein, nucleocapsid protein, matrix protein and 22K (M2) protein of the ovine respiratory syncytial virus. *Journal of General Virology*. **75**(12),pp.3597–3601.

Albertini, A.A. V., Wernimont, A.K., Muziol, T., Ravelli, R.B.G., Clapier, C.R., Schoehn, G., Weissenhorn, W. and Ruigrok, R.W.H. 2006. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science (New York, N.Y.)*. **313**(5785),pp.360–3.

Allen, M.C., Fuhrer, W., Tuck, B., Wade, R. and Wood, J.M. 1989. Renin inhibitors. Synthesis of transition-state analogue inhibitors containing phosphorus acid derivatives at the scissile bond. *Journal of medicinal chemistry*. **32**,pp.1652–1661.

Amarasinghe, G.K., Bào, Y., Basler, C.F., Bavari, S., Beer, M., Bejerman, N., Blasdell, K.R., Bochnowski, A., Briese, T., Bukreyev, A., Calisher, C.H., Chandran, K., Collins, P.L., Dietzgen, R.G., Dolnik, O., Dürrwald, R., Dye, J.M., Easton, A.J., Ebihara, H., Fang, Q., Formenty, P., Fouchier, R.A.M., Ghedin, E., Harding, R.M., Hewson, R., Higgins, C.M., Hong, J., Horie, M., James, A.P., Jiāng, D., Kobinger, G.P., Kondo, H., Kurath, G., Lamb, R.A., Lee, B., Leroy, E.M., Li, M., Maisner, A., Mühlberger, E., Netesov, S. V., Nowotny, N., Patterson, J.L., Payne, S.L., Paweska, J.T., Pearson, M.N., Randall, R.E., Revill, P.A., Rima, B.K., Rota, P., Rubbenstroth, D., Schwemmle, M., Smither, S.J., Song, Q., Stone, D.M., Takada, A., Terregino, C., Tesh, R.B., Tomonaga, K., Tordo, N., Towner, J.S., Vasilakis, N., Volchkov, V.E., Wahl-Jensen, V., Walker, P.J., Wang, B., Wang, D., Wang, F., Wang, L.F., Werren, J.H., Whitfield, A.E., Yan, Z., Ye, G. and Kuhn, J.H. 2017. Taxonomy of the order Mononegavirales: update 2017. *Archives of Virology*. **161**,pp.1–12.

Anderson, A. 2003. The Process of Structure-Based Drug Design. *Chemistry and Biology*. **10**,pp.787–797.

Ariza, A., Tanner, S.J., Walter, C.T., Dent, K.C., Shepherd, D.A., Wu, W., Matthews, S. V, Hiscox, J.A., Green, T.J., Luo, M., Elliott, R.M., Fooks, A.R., Ashcroft, A.E., Stonehouse, N.J., Ranson, N.A., Barr, J.N. and Edwards, T.A. 2013. Nucleocapsid protein structures from orthobunyaviruses reveal insight into ribonucleoprotein architecture and RNA polymerization. *Nucleic Acids Research*. **41**(11),pp.5912–5926.

Asenjo, A., Mendieta, J., Gómez-Puertas, P. and Villanueva, N. 2008. Residues in human respiratory syncytial virus P protein that are essential for its activity on RNA viral synthesis. *Virus Research*. **132**,pp.160–173.

Asenjo, A. and Villanueva, N. 2000. Regulated but not constitutive human respiratory syncytial virus (HRSV) P protein phosphorylation is essential for oligomerization. *FEBS Letters*. **467**(2–3),pp.279–284.

Asherie, N. 2004. Protein crystallization and phase diagrams. *Methods*. **34**(3),pp.266–272.

Atreya, P.L. and Kulkarni, S. 1999. Respiratory syncytial virus strain A2 is resistant to the antiviral effects of type I interferons and human MxA. *Virology*. **261**(2),pp.227–241.

Bächi, T. and Howe, C. 1973. Morphogenesis and ultrastructure of respiratory syncytial virus. *Journal of Virology*. **12**(5),pp.1173–80.

Bailly, B., Richard, C.-A., Sharma, G., Wang, L., Johansen, L., Cao, J., Pendharkar, V., Sharma, D.-C., Galloux, M., Wang, Y., Cui, R., Zou, G., Guillon, P., Von Itzstein, M., Eléouët, J.-F. and Altmeyer, R. 2016. Targeting human respiratory syncytial virus transcription anti- termination factor M2-1 to inhibit in vivo viral replication. *Scientific Reports*. **6**(25806),pp.1–11.

Banerjee, D. and Reddy, K.R. 2016. Review article: Safety and tolerability of direct-acting anti-viral agents in the new era of hepatitis C therapy. *Alimentary Pharmacology and Therapeutics*. **43**(6),pp.674–696.

Barik, S. 2004. Control of nonsegmented negative-strand RNA virus replication by siRNA. *Virus Research*. **102**(1),pp.27–35.

Barik, S., McLean, T. and Dupuy, L.C. 1995. Phosphorylation of Ser232 directly regulates the transcriptional activity of the P protein of human respiratory syncytial virus: phosphorylation of Ser237 may play an accessory role. *Virology*. **213**(2),pp.405–412.

Barr, J.N., Whelan, S.P.J. and Wertz, G.W. 1997. Role of the Intergenic Dinucleotide in Vesicular Stomatitis Virus RNA Transcription. *Journal of Virology*. **71**(3),pp.1794–1801.

Baskakov, I. V, Kumar, R., Srinivasan, G., Ji, Y.-S., Bolen, D.W. and Thompson, E.B. 1999. Trimethylamine N -Oxide-induced Cooperative Folding of an Fragment of Human Glucocorticoid Receptor. *The Journal of Biological Chemistry*. **274**(16),pp.10693–10696.

Bassetto, M., Massarotti, A., Coluccia, A. and Brancale, A. 2016. Structural biology in antiviral drug discovery. *Current Opinion in Pharmacology*. **30**,pp.116–130.

Battisti, A.J., Meng, G., Winkler, D.C., McGinnes, L.W., Plevka, P., Steven, A.C., Morrison, T.G. and Rossmann, M.G. 2012. Structure and assembly of a paramyxovirus matrix protein. *Proceedings of the National Academy of Sciences of the United States of America*. **109**(35),pp.13996–4000.

Battles, M.B., Langedijk, J.P., Furmanova-Hollenstein, P., Chaiwatpongsakorn, S., Costello, H.M., Kwanten, L., Vranckx, L., Vink, P., Jaensch, S., Jonckers,

T.H.M., Koul, A., Arnoult, E., Peeples, M.E., Roymans, D. and McLellan, J.S. 2015. Molecular mechanism of respiratory syncytial virus fusion inhibitors. *Nature Chemical Biology*. **12**(2),pp.87-93

Baybutt, H.N. and Pringle, C.R. 1987. Molecular cloning and sequencing of the F and 22K membrane protein genes of the RSS-2 strain of respiratory syncytial virus. *The Journal of General Virology*. **68**,pp.2789–96.

Beddell, C.R., Goodford, P.J., Stammer, D.K. and Wootton, R. 1979. Species differences in the binding of compounds designed to fit a site of known structure in adult human haemoglobin. *British Journal of Pharmacology*. **65**(3),pp.535–543.

Bermingham, a and Collins, P.L. 1999. The M2-2 protein of human respiratory syncytial virus is a regulatory factor involved in the balance between RNA replication and transcription. *Proceedings of the National Academy of Sciences of the United States of America*. **96**(20),pp.11259–64.

Biedenkopf, N., Hartlieb, B., Hoenen, T. and Becker, S. 2013. Phosphorylation of Ebola virus VP30 influences the composition of the viral nucleocapsid complex: impact on viral transcription and replication. *The Journal of Biological Chemistry*. **288**(16),pp.11165–74.

Blodörn, K., Hägglund, S., Fix, J., Dubuquoy, C., Makabi-Panzu, B., Thom, M., Karlsson, P., Roque, J.L., Karlstam, E., Pringle, J., Eléouët, J.F., Riffault, S., Taylor, G. and Valarcher, J.F. 2014. Vaccine safety and efficacy evaluation of a recombinant Bovine Respiratory Syncytial Virus (BRSV) with deletion of the SH gene and subunit vaccines based on recombinant human RSV proteins: N-nanorings, P and M2-1, in calves with maternal antibodies. *PLoS ONE*. **9**(6),pp.e100392.

Blondot, M.-L., Dubosclard, V., Fix, J., Lassoued, S., Aumont-Nicaise, M., Bontems, F., Eléouët, J.-F. and Sizun, C. 2012. Structure and functional analysis of the RNA- and viral phosphoprotein-binding domain of respiratory syncytial virus M2-1 protein. *PLoS pathogens*. **8**(5),p.e1002734.

Blundell, T.L., Jhoti, H. and Abell, C. 2002. High-Throughput Crystallography for Lead Discovery in Drug Design. *Nature Reviews Drug Discovery*. **1**,pp.45–54.

Bornholdt, Z.A., Noda, T., Abelson, D.M., Halfmann, P., Wood, M.R., Kawaoka, Y. and Saphire, E.O. 2013. Structural rearrangement of ebola virus vp40 begets multiple functions in the virus life cycle. *Cell*. **154**(4),pp.763–774.

Bose, S., Basu, M. and Banerjee, A.K. 2004. Role of Nucleolin in Human Parainfluenza Virus Type 3 Infection of Human Lung Epithelial Cells. *Society*. **78**(15),pp.8146–8158.

Bossert, B., Marozin, S. and Conzelmann, K.-K. 2003. Nonstructural proteins NS1 and NS2 of bovine respiratory syncytial virus block activation of interferon regulatory factor 3. *Journal of Virology*. **77**(16),pp.8661–8.

Boukhvalova, M.S., Prince, G.A. and Blanco, J.C.G. 2010. Inactivation of respiratory syncytial virus by zinc finger reactive compounds. **7**(20),pp.1–10.

Brown, G., Aitkein, J., Rixon, H.W.M. and Sugrue, R.J. 2002. Caveolin-1 is incorporated into mature respiratory syncytial virus particles during virus assembly

on the surface of virus-infected cells. *Journal of General Virology*. **83**(3),pp.611–621.

Bruhn, J.F., Barnett, K.C., Bibby, J., Thomas, J.M.H., Keegan, R.M., Rigden, D.J., Bornholdt, Z.A. and Saphire, E.O. 2014. Crystal structure of the nipah virus phosphoprotein tetramerization domain. *Journal of Virology*. **88**(1),pp.758–62.

Buchholz, U.J., Finke, S. and Conzelmann, K.-K. 1999. Generation of Bovine Respiratory Syncytial Virus (BRSV) from cDNA: BRSV NS2 Is Not Essential for Virus Replication in Tissue Culture, and the Human RSV Leader Region Acts as a Functional BRSV Genome Promoter. *Journal of Virology.* **73**(1),pp.251–259.

Bukreyev, A., Whitehead, S.S., Murphy, B.R. and Collins, P.L. 1997. Recombinant respiratory syncytial virus from which the entire SH gene has been deleted grows efficiently in cell culture and exhibits site-specific attenuation in the respiratory tract of the mouse. *Journal of Virology*. **71**(12),pp.8973–8982.

Burmeister, W.P., Ruigrok, R.W.H. and Cusack, S. 1992. The 2.2 A resolution crystal structure of influenza B neuraminidase and its complex with sialic acid. *The EMBO Journal*. **11**(1),pp.49–56.

Cancellieri, M., Bassetto, M., Widjaja, I., Van Kuppeveld, F., De Haan, C.A.M. and Brancale, A. 2015. In silico structure-based design and synthesis of novel anti-RSV compounds. *Antiviral Research*. **122**,pp.46-50.

Cartee, T.L., Megaw, A.G., Oomens, A.G.P. and Wertz, G.W. 2003. Identification of a Single Amino Acid Change in the Human Respiratory Syncytial Virus L Protein That Affects Transcriptional Termination. *Journal of Virology*. **77**(13),pp.7352–7360.

Cartee, T.L. and Wertz, G.W. 2001. Respiratory Syncytial Virus M2-1 Protein Requires Phosphorylation for Efficient Function and Binds Viral RNA during Infection. *Americal Society for Microbiology*. **75**(24),pp.12188–12197.

Carter, S.D., Dent, K.C., Atkins, E., Foster, T.L., Verow, M., Gorny, P., Harris, M., Hiscox, J.A., Ranson, N.A., Griffin, S. and Barr, J.N. 2010. Direct visualization of the small hydrophobic protein of human respiratory syncytial virus reveals the structural basis for membrane permeability. *FEBS letters*. **584**(13),pp.2786–90.

Castagné, N., Barbier, A., Bernard, J., Rezaei, H., Huet, J.C., Henry, C., Da Costa, B. and Eléouët, J.F. 2004. Biochemical characterization of the respiratory syncytial virus P-P and P-N protein complexes and localization of the P protein oligomerization domain. *Journal of General Virology*. **85**(6),pp.1643–1653.

Cavasotto, C.N., Orry, A.J.W., Murgolo, N.J., Czarniecki, M.F., Kocsi, S.A., Hawes, B.E., O'Neill, K.A., Hine, H., Burton, M.S., Voigt, J.H., Abagyan, R.A., Bayne, M.L. and Monsma, F.J. 2008. Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *Journal of Medicinal Chemistry*. **581**,pp.581-588.

Chang, A. and Dutch, R.E. 2012. Paramyxovirus fusion and entry: Multiple Paths to a common end. *Viruses*. **4**(4),pp.613–636.

Chapman, J., Abbott, E., Alber, D.G., Baxter, R.C., Bithell, S.K., Henderson, E.A., Carter, M.C., Chambers, P., Chubb, A., Cockerill, G.S., Collins, P.L., Dowdell, V.C.L., Keegan, S.J., Kelsey, R.D., Lockyer, M.J., Luongo, C., Najarro, P., Pickles, R.J., Simmonds, M., Taylor, D., Tyms, S., Wilson, L.J. and Powell, K.L. 2007. RSV604, a novel inhibitor of respiratory syncytial virus replication. *Antimicrobial Agents and Chemotherapy*. **51**(9),pp.3346–3353.

Charifson, P., Corkery, J., Murcko, M. and Walters, P. 1999. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins Paul. *J. Med. Chem*. **42**,pp.5100–5109.

Chen, J.K., Taipale, J., Cooper, M.K. and Beachy, P. a 2002. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened service Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes & Development*. (410),pp.2743–2748.

Cheng, Y. 2015. Single-particle Cryo-EM at crystallographic resolution. *Cell*. **161**(3),pp.450–457.

Chirkova, T., Lin, S., Oomens, A.G.P., Gaston, K.A., Boyoglu-Barnum, S., Meng, J., Stobart, C.C., Cotton, C.U., Hartert, T. V., Moore, M.L., Ziady, A.G. and Anderson, L.J. 2015. CX3CR1 is an important surface molecule for respiratory syncytial virus infection in human airway epithelial cells. *The Journal of general Virology*. **96**(9),pp.2543–56.

Chung, D.-H., Moore, B.P., Matharu, D.S., Golden, J.E., Maddox, C., Rasmussen, L., Sosa, M.I., Ananthan, S., White, E.L., Jia, F., Jonsson, C.B. and Severson, W.E. 2013. A cell based high-throughput screening approach for the discovery of new inhibitors of respiratory syncytial virus. *Virology Journal*. **10**,p.19.

Cianci, C., Langley, D.R., Dischino, D.D., Sun, Y., Yu, K.L., Stanley, A., Roach, J., Li, Z., Dalterio, R., Colonno, R., Meanwell, N.A. and Krystal, M. 2004. Targeting a binding pocket within the trimer-of-hairpins: small-molecule inhibition of viral fusion. *Proc Natl Acad Sci U S A*. **101**(42),pp.15046–15051.

Cianci, C., Meanwell, N. and Krystal, M. 2005. Antiviral activity and molecular mechanism of an orally active respiratory syncytial virus fusion inhibitor. *The Journal of antimicrobial chemotherapy*. **55**(3),pp.289–92.

Coates, M., Brookes, D., Allen, H., Fordyce, E., Colley, T., Hunt, F., Onion, S., Yim, K., Knowles, I., Murray, P.J., Ito, K., Strong, P. and Rapeport, G. Preclinical characterization of a novel anti-respiratory syncytial virus agent, PC786.

Collins, P.L., Hill, M.G., Camargo, E., Grosfeld, H., Chanock, R.M. and Murphy, B.R. 1995. Production of infectious human respiratory syncytial virus from cloned cDNA confirms an essential role for the transcription elongation factor from the 5' proximal open reading frame of the M2 mRNA in gene expression and provides a capability for vaccine. *Proceedings of the National Academy of Sciences of the United States of America*. **92**(25),pp.11563–11567.

Collins, P.L., Hill, M.G., Cristina, J. and Grosfeld, H. 1996. Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand

RNA virus. *Proceedings of the National Academy of Sciences of the United States of America*. **93**(1),pp.81–5.

Collins, P.L. and Wertz, G.W. 1985. The envelope-associated 22K protein of human respiratory syncytial virus: nucleotide sequence of the mRNA and a related polytranscript. *Journal of Virology*. **54**(1),pp.65–71.

Collins, P.M., Ng, J.T., Talon, R., Nekrosiute, K., Krojer, T., Douangamath, A., Brandao-Neto, J., Wright, N., Pearce, N.M. and Von Delft, F. 2017. Gentle, fast and effective crystal soaking by acoustic dispensing. *Acta Crystallographica Section D: Structural Biology*. **73**,pp.246–255.

Collins, P.L., Dickens, L.E., Buckler-White, A., Olmsted, R.A., Spriggs, M.K., Camargo, E. and Coelingh, K. V 1986. Nucleotide sequences for the gene junctions of human respiratory syncytial virus reveal distinctive features of intergenic structure and gene order. *Proceedings of the National Academy of Sciences of the United States of America*. **83**(13),pp.4594–8.

Collins, P.L. and Graham, B.S. 2008. Viral and host factors in human respiratory syncytial virus pathogenesis. *Journal of Virology*. **82**(5),pp.2040–55.

Collins, P.L., Huang, Y.T. and Wertz, G.W. 1984. Identification of a Tenth mRNA of Respiratory Syncytial Virus and Assignment of Polypeptides to the 10 Viral Genes. *Journal of Virology*. **49**(2),pp.572–578.

Collins, P.L., Olmsted, R.A., Spriggs, M.K., Johnson, P.R. and Buckler-White, A.J. 1987. Gene overlap and site-specific attenuation of transcription of the viral polymerase L gene of human respiratory syncytial virus (negative-strand RNA viruses/transcriptive signals/mode of transcription/alternative transcripts). *Biochemistry*. **84**,pp.5134–5138.

Communie, G., Crépin, T., Maurin, D., Jensen, M.R., Blackledge, M. and Ruigrok, R.W.H. 2013. Structure of the tetramerization domain of measles virus phosphoprotein. *Journal of Virology*. **87**(12),pp.7166–7169.

Communie, G., Ruigrok, R.W., Jensen, M.R. and Blackledge, M. 2014. Intrinsically disordered proteins implicated in paramyxoviral replication machinery. *Current Opinion in Virology*. **5**(1),pp.72–81.

Cox, O.B. and et al. 2016. A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem. Sci.* **7**(3),pp.2322–2330.

Cox, R., Green, T.J., Purushotham, S., Deivanayagam, C., Bedwell, G.J., Prevelige, P.E. and Luo, M. 2013. Structural and Functional Characterization of the Mumps Virus Phosphoprotein. *Journal of Virology*. **87**(13),pp.7558–7568.

Cox, R., Pickar, A., Qiu, S., Tsao, J., Rodenburg, C., Dokland, T., Elson, A., He, B. and Luo, M. 2014. Structural studies on the authentic mumps virus nucleocapsid showing uncoiling by the phosphoprotein. *Proceedings of the National Academy of Sciences of the United States of America*. **111**(42),pp.15208–15213.

Cuesta, I., Geng, X., Asenjo, A. and Villanueva, N. 2000. Structural phosphoprotein M2-1 of the human respiratory syncytial virus is an RNA binding protein. *Journal of Virology*. **74**(21),pp.9858–67.

de Ruyck, J., Brysbaert, G., Blossey, R. and Lensink, M.F. 2016. Molecular docking as a popular tool in drug design, an in silico travel. *Advances and Applications in Bioinformatics and Chemistry*. **9**,pp.1–11.

Debnath, A., Radigan, L. and Jiang, S. 1999. Strucutre-Based Identification of Small Molecule Antiviral Compounds Targeted to the gp41 Core Structure of the Human Immunodeficiency Virus Type 1. *J. Med. Chem*. (42),pp.3203–3209.

Delgado, M.F., Coviello, S., Monsalvo, A.C., Melendi, G.A., Hernandez, J.Z., Batalle, J.P., Diaz, L., Trento, A., Chang, H.Y., Mitzner, W., Ravetch, J., Melero, J.A., Irusta, P.M. and Polack, F.P. 2009. Lack of antibody affinity maturation due to poor Toll-like receptor stimulation leads to enhanced respiratory syncytial virus disease. *Nature medicine*. **15**(1),pp.34–41.

DeVincenzo, J.P., McClure, M.W., Symons, J.A., Fathi, H., Westland, C., Chanda, S., Lambkin-Williams, R., Smith, P., Zhang, Q., Beigelman, L., Blatt, L.M. and Fry, J. 2015. Activity of Oral ALS-008176 in a Respiratory Syncytial Virus Challenge Study. *The New England Journal of medicine*. **373**(21),pp.2048–58.

DeVincenzo, J.P., Whitley, R.J., Mackman, R.L., Scaglioni-Weinlich, C., Harrison, L., Farrell, E., McBride, S., Lambkin-Williams, R., Jordan, R., Xin, Y., Ramanathan, S., O'Riordan, T., Lewis, S.A., Li, X., Toback, S.L., Lin, S.-L. and Chien, J.W. 2014. Oral GS-5806 activity in a respiratory syncytial virus challenge study. The New England *Journal of medicine*. **371**(8),pp.711–22.

Do, L.A., Wilm, A., Van Doorn, H.R., Lam, H.M., Sim, S., Sukumaran, R., Tran, A.T., Nguyen, B.H., Tran, T.T.L., Tran, Q.H., Vo, Q.B., Tran Dac, N.A., Trinh, H.N., Nguyen, T.T.H., Le Binh, B.T., Le, K., Nguyen, M.T., Thai, Q.T., Vo, T.V., Minh Ngo, N.Q., Dang, T.K.H., Cao, N.H., Van Tran, T., Ho, L.V., Farrar, J., De Jong, M., Chen, S., Nagarajan, N., Bryant, J.E. and Hibberd, M.L. 2015. Direct whole-genome deep-sequencing of human respiratory syncytial virus A and B from Vietnamese children identifies distinct patterns of inter- and intra-host evolution. *Journal of General Virology*. **96**(12),pp.3470–3483.

Douglas, J.L., Panis, M.L., Ho, E., Krawczyk, S.H., Grant, D.M., Cai, R., Cihlar, T., Lin, K. and Swaminathan, S. 2003. Inhibition of Respiratory Syncytial Virus Fusion by the Small Molecule VP-14637 via Specific Interactions with F Protein Inhibition of Respiratory Syncytial Virus Fusion by the Small Molecule VP-14637 via Specific Interactions with F Protein. *Journal of Virology*. **77**(9),pp.5054–5064.

Drozdetskiy, A., Cole, C., Procter, J. and Barton, G.J. 2015. JPred4: A protein secondary structure prediction server. *Nucleic Acids Research*. **43**,pp.389–394.

Esperante, S.A., Noval, M.G., Altieri, T.A., De Oliveira, G.A.P., Silva, J.L. and De Prat-Gay, G. 2013. Fine modulation of the respiratory syncytial virus M2-1 protein quaternary structure by reversible zinc removal from its Cys3- His1 motif. *Biochemistry*. **52**(39),pp.6779–6789.

Fearns, R., Collins, P.L. and Peeples, M.E. 2000. Functional analysis of the genomic and antigenomic promoters of human respiratory syncytial virus. *Journal of Virology*. **74**(13),pp.6006–14.

Fearns, R. and Collins, P.L. 1999. Role of the M2-1 Transcription Antitermination Protein of Respiratory Syncytial Virus in Sequential Transcription Role of the M2-1 Transcription Antitermination Protein of Respiratory Syncytial Virus in Sequential Transcription. *Journal of Virology*. **73**(7),pp.5852–5864.

Fearns, R. and Collins, P.L. 1999. Role of the M2-1 transcription antitermination protein of respiratory syncytial virus in sequential transcription. *Journal of Virology*. **73**(7),pp.5852–64.

Fearns, R. and Collins, P.L. 1999. Model for polymerase access to the overlapped L gene of respiratory syncytial virus. *Journal of Virology*. **73**(1),pp.388–97.

Fearns, R., Peeples, M.E. and Collins, P.L. 2002. Mapping the transcription and replication promoters of respiratory syncytial virus. *Journal of Virology*. **76**(4),pp.1663–72.

Feeney, E., Chung, R. 2014. Antiviral treatment of hepatitis C. *The British Medical Journal*. **349**,pp.g3308

Feltes, T.F., Cabalka, A.K., Meissner, H.C., Piazza, F.M., Carlin, D. a, Top, F.H., Connor, E.M. and Sondheimer, H.M. 2003. Palivizumab prophylaxis reduces hospitalization due to respiratory syncytial virus in young children with hemodynamically significant congenital heart disease. *The Journal of pediatrics*. **143**(4),pp.532–40.

Fenton, C., Scott, L.J. and Plosker, G.L. 2004. Palivizumab. *Pediatric Drugs*. **6**(3),pp.177–197.

Frederickson, C.J., Suh, S.W., Silva, D., Frederickson, C.J. and Thompson, R.B. 2000. Zinc and Health : Current Status and Future Directions Importance of Zinc in the Central Nervous System : The Zinc-Containing. *The Journal of Nutrition*. **130**,pp.1471–1483.

Freisner, R., Banks, J., Murphy, R., Halgren, R., Klicic, J., Mainz, D., Repasky, M., Knoll, E., Jason, M., Perry, J., Shaw, D., Francis, P. and Shenkin, P. 2004. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy Richard. *J. Med. Chem*. **47**,pp.1739–1749.

Frey, G., Rits-Volloch, S., Zhang, X.Q., Schooley, R.T., Chen, B. and Harrison, S.C. 2006. Small molecules that bind the inner core of gp41 and inhibit HIV envelope-mediated fusion. *Proc Natl Acad Sci U S A*. **103**(38),pp.13938–13943.

Fujiki, H., Watanabe, T. and Suganuma, M. 2014. Cell-surface nucleolin acts as a central mediator for carcinogenic, anti-carcinogenic, and disease-related ligands. *Journal of Cancer Research and Clinical Oncology*. **140**(5),pp.689–699.

Furze, J.M., Roberts, S.R., Wertz, G.W. and Taylor, G. 1997. Antigenically distinct G glycoproteins of BRSV strains share a high degree of genetic homogeneity. *Virology*. **231**(1),pp.48–58.

Galloux, M., Gabiane, G., Sourimant, J., Richard, C.-A., England, P., Moudjou, M., Aumont-Nicaise, M., Fix, J., Rameix-Welti, M.-A. and Eléouët, J.-F. 2015. Identification and Characterization of the Binding Site of the Respiratory Syncytial Virus Phosphoprotein to RNA-Free Nucleoprotein. *Journal of Virology*. **89**(7),pp.3484–3496.

Gan, S.W., Tan, E., Lin, X., Yu, D., Wang, J., Tan, G.M.Y., Vararattanavech, A., Yeo, C.Y., Soon, C.H., Soong, T.W., Pervushin, K. and Torres, J. 2012. The small hydrophobic protein of the human respiratory syncytial virus forms pentameric ion channels. *Journal of Biological Chemistry*. **287**(29),pp.24671–24689.

Gao, Y., Dickerson, J.B., Guo, F., Zheng, J. and Zheng, Y. 2004. Rational design and characterization of a Rac GTPase-specific small molecule inhibitor. *Proceedings of the National Academy of Sciences*. **101**(20),pp.7618–7623.

García-Barreno, B., Delgado, T. and Melero, J.A. 1996. Identification of protein regions involved in the interaction of human respiratory syncytial virus phosphoprotein and nucleoprotein: significance for nucleocapsid assembly and formation of cytoplasmic inclusions. *Journal of Virology*. **70**(2),pp.801–8.

Garlapati, S., Garg, R., Brownlie, R., Latimer, L., Simko, E., Hancock, R.E.W., Babiuk, L.A., Gerdts, V., Potter, A. and van Drunen Littel-van den Hurk, S. 2012. Enhanced immune responses and protection by vaccination with respiratory syncytial virus fusion protein formulated with CpG oligodeoxynucleotide and innate defense regulator peptide in polyphosphazene microparticles. *Vaccine*. **30**(35),pp.5206–5214.

Gershwin, L. 2007. Bovine Respiratory Syncytial Virus Infection: Immunopathogenic Mechanisms. *Animal Health Research Reviews*. **8**(2),pp.207-213.

Ghildyal, R., Mills, J., Murray, M., Vardaxis, N. and Meanger, J. 2002. Respiratory syncytial virus matrix protein associates with nucleocapsids in infected cells. *J Gen Virol*. **83**(Pt 4),pp.753–757.

Glenn, G.M., Fries, L.F., Thomas, D.N., Smith, G., Kpamegan, E., Lu, H., Flyer, D., Jani, D., Hickman, S.P. and Piedra, P.A. 2016. A randomized, blinded, controlled, dose-ranging study of a respiratory syncytial virus recombinant Fusion (F) nanoparticle vaccine in healthy women of childbearing age. *Journal of Infectious Diseases*. **213**(3),pp.411–422.

Glenn, G.M., Smith, G., Fries, L., Raghunandan, R., Lu, H., Zhou, B., Thomas, D.N., Hickman, S.P., Kpamegan, E., Boddapati, S. and Piedra, P.A. 2013. Safety and immunogenicity of a Sf9 insect cell-derived respiratory syncytial virus fusion protein nanoparticle vaccine. *Vaccine*. **31**(3),pp.524–532.

Glezen, W.P., Taber, L.H., Frank, A.L. and Kasel, J.A. 1986. Risk of Primary Infection and Reinfection With Respiratory Syncytial Virus. *AJDC*. **140**,pp.543–546.

González-Reyes, L., Ruiz-Argüello, M.B., García-Barreno, B., Calder, L., López, J.A., Albar, J.P., Skehel, J.J., Wiley, D.C. and Melero, J.A. 2001. Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for

activation of membrane fusion. *Proceedings of the National Academy of Sciences of the United States of America*. **98**(17),pp.9859–64.

Gonzalez, M.E. and Carrasco, L. 2003. Viroporins. *FEBS Letters*. **552**(1),pp.28–34.

Gotoh, B., Komatsu, T., Takeuchi, K. and Yokoo, J. 2001. Paramyxovirus accessory proteins as interferon antagonists. *Microbiology and immunology*. **45**(12),pp.787–800.

Green, T.J., Zhang, X., Wertz, G.W. and Luo, M. 2006. Structure of the Vesicular Stomatitis Virus Nucleoprotein-RNA Complex. *Science*. **313**(5785),pp.357–360.

Greer, J., Erickson, J.W., Baldwin, J.J. and Varneyl, M.D. 1994. Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design. *Journal of Medicinal Chemistry*. **37**(8),pp.1036–1054.

Griffiths, C., Drews, S.J. and Marchant, D.J. 2017. Respiratory syncytial virus: Infection, detection, and new options for prevention and treatment. *Clinical Microbiology Reviews*. **30**(1),pp.277–319.

Grosfeld, H., M. G. Hill, and P. L. Collins. 1995. RNA replication by respiratory syncytial virus (RSV) is directed by the N, P, and L proteins: transcription also occurs under these conditions but required RSV super infection for efficient synthesis of full-length mRNA. Journal of Virology. **69**,pp. 5677–5686.

Ha Do, L.A., Wilm, A., Van Doorn, H.R., Lam, H.M., Sim, S., Sukumaran, R., Tran, A.T., Nguyen, B.H., Tran, T.T.L., Tran, Q.H., Vo, Q.B., Tran Dac, N.A., Trinh, H.N., Nguyen, T.T.H., Le Binh, B.T., Le, K., Nguyen, M.T., Thai, Q.T., Vo, T.V., Minh Ngo, N.Q., Dang, T.K.H., Cao, N.H., Van Tran, T., Ho, L.V., Farrar, J., De Jong, M., Chen, S., Nagarajan, N., Bryant, J.E. and Hibberd, M.L. 2015. Direct whole-genome deep-sequencing of human respiratory syncytial virus A and B from Vietnamese children identifies distinct patterns of inter- and intra-host evolution. *Journal of General Virology*. **96**(12),pp.3470–3483.

Hägglund, S., Hu, K., Blodörn, K., Makabi-Panzu, B., Gaillard, A.-L.L., Ellencrona, K., Chevret, D., Hellman, L., Lövgren Bengtsson, K., Riffault, S., Taylor, G., Valarcher, J.-F.F., Eléouët, J.-F.F, Panzu, B., Gaillard, A.-L.L., Ellencrona, K., Chevret, D., Hellman, L., Bengtsson, K.L., Riffault, S., Taylor, G., Valarcher, J.-F.F. and Eléouët, J.-F.F. 2014. Characterization of an experimental vaccine for bovine respiratory syncytial virus. *Clinical and vaccine immunology : CVI*. **21**(7),pp.997-1004.

Hall, C., McBride, J., Walsh, E., Bell, D. and Gala, C. 1983. Aerosolized ribavirin treatment of infants with respiratory syncytial viral infection. *The New England Journal of Medicine*. **308**(24),pp.1443–1447.

Harcourt, B.H., Tamin, A., Ksiazek, T.G., Rollin, P.E., Anderson, L.J., Bellini, W.J. and Rota, P.A. 2000. Molecular characterization of Nipah virus, a newly emergent paramyxovirus. *Virology*. **271**,pp.334–349.

Hardy, R.W., Harmon, S.B. and Wertz, G.W. 1999. Diverse gene junctions of respiratory syncytial virus modulate the efficiency of transcription termination and

respond differently to M2-mediated antitermination. *Journal of Virology*. **73**(1),pp.170–6.

Hardy, R.W. and Wertz, G.W. 2000. The Cys 3 -His 1 Motif of the Respiratory Syncytial Virus M2-1 Protein Is Essential for Protein Function The Cys 3 -His 1 Motif of the Respiratory Syncytial Virus M2-1 Protein Is Essential for Protein Function. *Journal of Virology*. **74**(13),pp.5880–5885.

Hardy, R.W. and Wertz, G.W. 1998. The product of the respiratory syncytial virus M2 gene ORF1 enhances readthrough of intergenic junctions during viral transcription. *Journal of Virology*. **72**(1),pp.520–6.

Hardy, R.W. and Wertz, G.W. 2000. The Cys 3 -His 1 Motif of the Respiratory Syncytial Virus M2-1 Protein Is Essential for Protein Function. *Journal of Virology*. **74**(13),pp.5880–5885.

Harris, J. and Werling, D. 2003. Binding and entry of respiratory syncytial virus into host cells and initiation of the innate immune response. *Cellular Microbiology*. **5**(10),pp.671-680.

Hartlieb, B., Muziol, T., Weissenhorn, W. and Becker, S. 2007. Crystal structure of the C-terminal domain of Ebola virus VP30 reveals a role in transcription and nucleocapsid association. *Proceedings of the National Academy of Sciences of the United States of America*. **104**(2),pp.624–9.

Hartshorn, M.J., Murray, C.W., Cleasby, A., Frederickson, M., Tickle, I.J. and Jhoti, H. 2005. Fragment-based lead discovery using X-ray crystallography. *Journal of Medicinal Chemistry*. **48**(2),pp.403–413.

Hendricks, D.A., Baradaran, K., McIntosh, K. and Patterson, J.L. 1987. Appearance of a soluble form of the G protein of respiratory syncytial virus in fluids of infected cells. *Journal of General Virology*. **68**(6),pp.1705–1714.

Heylen, E., Neyts, J. and Jochmans, D. 2017. Drug candidates and model systems in respiratory syncytial virus antiviral drug discovery. *Biochemical Pharmacology*. **127**,pp.1–12.

Himmel, D.M., Sarafianos, S.G., Dharmasena, S., Hossain, M.M., McCoy-Simandle, K., Ilina, T., Clark, A.D., Knight, J.L., Julias, J.G., Clark, P.K., Krogh-Jespersen, K., Levy, R.M., Hughes, S.H., Parniak, M.A. and Arnold, E. 2006. HIV-1 reverse transcriptase structure with RNase H inhibitor dihydroxy benzoyl naphthyl hydrazone bound at a novel site. *ACS chemical biology*. **1**(11),pp.702–712.

Hoenen, T., Groseth, A., Kolesnikova, L., Theriault, S., Ebihara, H., Hartlieb, B., Bamberg, S., Feldmann, H., Ströher, U. and Becker, S. 2006. Infection of naive target cells with virus-like particles: implications for the function of ebola virus VP24. *Journal of Virology*. **80**(14),pp.7260–4.

Hoetelmans, R., Meenhorts, R., Mulder, J., Burger, D., Koks, C. and Beijnen, J. 1997. Clinical pharmacology of HIV protease inhibitors: focus on saquinavir, indinavir, and ritonavir. *Pharm World Sci*. **19**(4),pp.159–175.

Hong, J., Cheng, X., Zhou, H., Li, S and Seddiqui, A. 2000. Respiratory Syncytial Virus That Lacks Open Reading Frame 2 of the M2 Gene (M2-2) Has Altered

Growth Characteristics and Is Attenuated in Rodents. *Journal of Virology.* **74**(1),pp.74-82.

Huang, K., Incognito, L., Cheng, X., Ulbrandt, N.D. and Wu, H. 2010. Respiratory Syncytial Virus-Neutralizing Monoclonal Antibodies Motavizumab and Palivizumab Inhibit Fusion. *Journal of Virology*. **84**(16),pp.8132–8140.

Inaba Y, Tanaka Y, Omori T. 1970. Isolation of Bovine Respiratory Syncytial Virus. *Japan Journal of Exp Medicine.***40**,pp.473–474.

Inaba, Y., Tanaka, Y., Sato, K., Omori, T. and Matumoto, M. 1972. Bovine respiratory syncytial virus. *Japan. J. Microbiol*. **16**(5),pp.373–383.

Irwin, J.J. and Shoichet, B.K. 2005. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model*. **45**,pp.177–182.

Irwin, J.J. and Shoichet, B.K. 2005. ZINC - A free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*. **45**(1),pp.177–182.

Iseni, F., Garcin, D., Nishio, M., Kedersha, N., Anderson, P. and Kolakofsky, D. 2002. Sendai virus trailer RNA binds TIAR, a cellular protein involved in virus-induced apoptosis. *EMBO Journal*. **21**(19),pp.5141–5150.

Jacobs, J.W.; Edington, N. 1971. Isolation of respiratory syncytial virus from cattle in Britain. *Vet. Rec*. **88**,pp.694

Jancarik, J. and Kim, S.H. 1991. Sparse matrix sampling: A screening method for crystallisation of proteins. *J. Appl. Cryst*. **24**(63),p.409.

Jeffree, C.E., Brown, G., Aitken, J., Su-Yin, D.Y., Tan, B.H. and Sugrue, R.J. 2007. Ultrastructural analysis of the interaction between F-actin and respiratory syncytial virus during virus assembly. *Virology*. **369**(2),pp.309–323.

Jeong, K. Il, Piepenhagen, P.A., Kishko, M., DiNapoli, J.M., Groppo, R.P., Zhang, L., Almond, J., Kleanthous, H., Delagrave, S. and Parrington, M. 2015. CX3CR1 is expressed in differentiated human ciliated airway cells and co-localizes with respiratory syncytial virus on cilia in a G protein-dependent manner. *PLoS ONE*. **10**(6),p.e0130517.

Jin, H., Cheng, X., Zhou, H.Z.Y., Li, S. and Seddiqui, A. 2000. Respiratory Syncytial Virus That Lacks Open Reading Frame 2 of the M2 Gene (M2-2) Has Altered Growth Characteristics and Is Attenuated in Rodents Respiratory Syncytial Virus That Lacks Open Reading Frame 2 of the M2 Gene (M2-2) Has Altered Growth Char. *Journal of Virology*. **74**(1),pp.74–82.

John, S.P., Wang, T., Steffen, S., Longhi, S., Schmaljohn, C.S. and Jonsson, C.B. 2007. Ebola Virus VP30 Is an RNA Binding Protein. *Journal of Virology*. **81**(17),pp.8967–8976.

Johnson, P.R., Spriggs, M.K., Olmsted, R.A. and Collins, P.L. 1987. The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins.

Proceedings of the National Academy of Sciences of the United States of America. **84**(16),pp.5625–9..

Johnson, S.M., McNally, B.A., Ioannidis, I., Flano, E., Teng, M.N., Oomens, A.G., Walsh, E.E. and Peeples, M.E. 2015. Respiratory Syncytial Virus Uses CX3CR1 as a Receptor on Primary Human Airway Epithelial Cultures. *PLoS Pathogens*. **11**(12),p.e1005318.

Jordan, R., Shao, M., Mackman, R.L., Perron, M., Cihlar, T., Lewis, S.A., Eisenberg, E.J., Carey, A., Strickley, R.G., Chien, J.W., Anderson, M.L., McEligot, H.A., Behrens, N.E. and Gershwin, L.J. 2015. Antiviral efficacy of a respiratory syncytial virus (RSV) fusion inhibitor in a bovine model of RSV infection. *Antimicrobial Agents and Chemotherapy*. **59**(8),pp.4889–4900.

Jordheim, L.P., Durantel, D., Zoulim, F. and Dumontet, C. 2013. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nature Reviews Drug Discovery*. **12**,pp.447–464.

Kallewaard, N.L., Bowen, A.L. and Crowe, J.E. 2005. Cooperativity of actin and microtubule elements during replication of respiratory syncytial virus. *Virology*. **331**(1),pp.73–81.

Kaptur, P.E., Rhodes, R.B. and Lyles, D.S. 1991. Sequences of the vesicular stomatitis virus matrix protein involved in binding to nucleocapsids. *J Virol*. **65**(3),pp.1057–1065.

Kargel, A., Schmidt, U. and Buchholz, U.J. 2001. Recombinant bovine respiratory syncytial virus with deletions of the G or SH genes: G and F proteins binds heparin. *Journal of General Virology*. **82**(3),pp.631–640.

Karlin, D. and Belshaw, R. 2012. Detecting remote sequence homology in disordered proteins: Discovery of conserved motifs in the N-termini of mononegavirales phosphoproteins. *PLoS ONE*. **7**(3),p.e31719.

Karron, R. a, Buonagurio, D. a, Georgiu, A.F., Whitehead, S.S., Adamus, J.E., Clements-Mann, M. Lou, Harris, D.O., Randolph, V.B., Udem, S.A., Murphy, B.R. and Sidhu, M.S. 1997. Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication in vitro: clinical evaluation and molecular characterization of a cold-passaged, attenuated RSV subgroup B mutant. *Proceedings of the National Academy of Sciences of the United States of America*. **94**(25),pp.13961–6.

Khattar, S.K., Yunus, A.S. and Samal, S.K. 2001. Mapping the domains on the phosphoprotein of bovine respiratory syncytial virus required for N-P and P-L interactions using a minigenome system. *Journal of General Virology*. **82**(4),pp.775–779.

Kim, C.U., Lew, W., Williams, M.A., Liu, H., Zhang, L., Swaminathan, S., Bischofberger, N., Chen, M.S., Mendel, D.B., Tai, C.Y., Laver, W.G. and Stevens, R.C. 1997. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: Design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *Journal of the American Chemical Society*. **119**(4),pp.681–690.

Kimman, T., Sol, J., Westenbrink, P. and Strayer, P. 1989. A severe outbreak of respiratory tract disease associated with bovine respiratory syncytial virus probably enhanced by vaccination with modified live vaccine. *Veterinary Quarterly*. **11**(4),pp.250–253.

Kimura, S.R., Tebben, A.J. and Langley, D.R. 2008. Expanding GPCR homology model binding sites via a balloon potential: A molecular dynamics refinement approach. *Proteins: Structure, Function and Genetics*.

Kiss, G., Holl, J.M., Williams, G.M., Alonas, E., Vanover, D., Lifland, A.W., Gudheti, M., Guerrero-Ferreira, R.C., Nair, V., Yi, H., Graham, B.S., Santangelo, P.J. and Wright, E.R. 2014. Structural Analysis of Respiratory Syncytial Virus Reveals the Position of M2-1 Between the Matrix Protein and the Ribonucleoprotein Complex. *Journal of Virology*. **88**(13),pp.7602–7617.

Kitchen, D.B., Decornez, H., Furr, J.R. and Bajorath, J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*. **3**(11),pp.935–949.

Kolokoltsov, A.A., Deniger, D., Fleming, E.H., Roberts, N.J., Karpilow, J.M. and Davey, R.A. 2007. Small interfering RNA profiling reveals key role of clathrin-mediated endocytosis and early endosome formation for infection by respiratory syncytial virus. *Journal of Virology*. **81**(14),pp.7786–800.

Krojer, T., Talon, R., Pearce, N., Collins, P., Douangamath, A., Brandao-Neto, J., Dias, A., Marsden, B. and Von Delft, F. 2017. The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination. *Acta Crystallographica Section D: Structural Biology*. **73**,pp.267–278.

Krusat, T and Steckert, H.-J. 1997. Heparin-Dependent Attachment of Respiratory Syncytial Virus (RSV) to Host Cells. *Archives of Virology*. **142**(6),pp.1247-1254.

Kuhlbrandt, W. 2014. The Resolution Revolution. *Science*. **343** (6178),pp.1443-1444.

Kuo, L., Fearns, R. and Collins, P.L. 1996. The structurally diverse intergenic regions of respiratory syncytial virus do not modulate sequential transcription by a dicistronic minigenome. *Journal of Virology*. **70**(9),pp.6143–6150.

Kuo, L., Grosfeld, H., Cristina, J., Hill, M.G. and Collins, P.L. 1996. Effects of mutations in the gene-start and gene-end sequence motifs on transcription of monocistronic and dicistronic minigenomes of respiratory syncytial virus. *Journal of Virology*. **70**(10),pp.6892–901.

Laganas, V.A., Dunn, E.F., McLaughlin, R.E., Tiong-Yip, C.L., Yuzhakov, O., Isabella, V.M., Hill, P. and Yu, Q. 2015. Characterization of novel respiratory syncytial virus inhibitors identified by high throughput screen. *Antiviral Research*. **115**,pp.71–74.

Lambden, P.R. 1985. Nucleotide sequence of the respiratory syncytial virus phosphoprotein gene. *Journal of General Virology*. **66**(7),pp.1607–1612.

Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J., Whittle, P., Danley, D., Geoghegan, K., Hawrylik, S., Lee, E., Scheld,

K. and Hobart, P. 1989. X-ray analysis of HIV-1 proteinase at 2.7 A resolution confirms structural homolgy among retroviral enzymes. *Nature*. **342**,pp.299–302.

Lapatto, R., Blundell, T., Hemmings, A., Overington, J., Wilderspin, A., Wood, S., Merson, J.R., Whittle, P.J., Danley, D.E. and Geoghegan, K.F. 1989. X-ray analysis of HIV-1 proteinase at 2.7 A resolution confirms structural homology among retroviral enzymes. *Nature*. **342**(6247),pp.299–302.

Le Nouën, C., Hillyer, P., Winter, C.C., McCarty, T., Rabin, R.L., Collins, P.L. and Buchholz, U.J. 2011. Low CCR7-mediated migration of human monocyte derived dendritic cells in response to human respiratory syncytial virus and human metapneumovirus. *PLoS Pathogens*. **7**(6),p.e1002105.

Le Nouën, C., McCarty, T., Brown, M., Smith, M.L., Lleras, R., Dolan, M.A., Mehedi, M., Yang, L., Luongo, C., Liang, B., Munir, S., DiNapoli, J.M., Mueller, S., Wimmer, E., Collins, P.L. and Buchholz, U.J. 2017. Genetic stability of genome-scale deoptimized RNA virus vaccine candidates under selective pressure. *Proceedings of the National Academy of Sciences*. pp.e386–e395.

Le, Q.M., Kiso, M., Someya, K., Sakai, Y.T., Nguyen, T.H., Nguyen, K.H.L., Pham, N.D., Ngyen, H.H., Yamada, S., Muramoto, Y., Horimoto, T., Takada, A., Goto, H., Suzuki, T., Suzuki, Y. and Kawaoka, Y. 2005. Avian flu: Isolation of drug-resistant H5N1 virus. *Nature*. **437**,p.1108.

Lee, Y.M., Duh, Y., Wang, S.T., Lai, M.M.C., Yuan, H.S. and Lim, C. 2016. Using an Old Drug to Target a New Drug Site: Application of Disulfiram to Target the Zn-Site in HCV NS5A Protein. *Journal of the American Chemical Society*. **138**(11),pp.3856–3862.

Leeson, P. 2012. Drug discovery: Chemical beauty contest. *Nature*. **481**(7382),pp.455–456.

Levine, S., Kaliaber-Franco, R. and Paradiso, P.R. 1987. Demonstration that glycoprotein G is the attachment protein of respiratory syncytial virus. *Journal of General Virology*. **68**(9),pp.2521–2524.

Lew, W., Chen, X. and Kim, C.U. 2000. Medicinal Chemistry Feature Molecule Discovery and Development of GS 4104 (oseltamivir): An Orally Active Influenza Neuraminidase Inhibitor. *Current Medicinal Chemistry*. **7**,pp.663-672.

Leyrat, C., Renner, M., Harlos, K. and Grimes, J.M. 2013. Solution and crystallographic structures of the central region of the phosphoprotein from human metapneumovirus. *PLoS ONE*. **8**(11),p.e80371.

Leyrat, C., Renner, M., Harlos, K., Huiskonen, J.T. and Grimes, J.M. 2014. Drastic changes in conformational dynamics of the antiterminator M2-1 regulate transcription efficiency in Pneumovirinae. *eLife*. p.e02674.

Li, D., Jans, D.A., Bardin, P.G., Meanger, J., Mills, J. and Ghildyal, R. 2008. Association of respiratory syncytial virus M protein with viral nucleocapsids is mediated by the M2-1 protein. *Journal of Virology*. **82**(17),pp.8863–70.

Li, J., Rahmeh, A., Morelli, M. and Whelan, S.P.J. 2008. A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping. *Journal of Virology*. **82**(2),pp.775–84.

Li, Y., To, J., Verdià-Baguena, C., Dossena, S., Surya, W., Huang, M., Paulmichl, M., Liu, D.X., Aguilella, V.M. and Torres, J. 2014. Inhibition of the human respiratory syncytial virus small hydrophobic protein and structural variations in a bicelle environment. *Journal of Virology*. **88**(20),pp.11899–914.

Liang, B., Li, Z., Jenni, S., Rahmeh, A.A., Morin, B.M., Grant, T., Grigorieff, N., Harrison, S.C. and Whelan, S.P.J. 2015. Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. *Cell*. **162**,pp.314–327.

Liesman, R.M., Buchholz, U.J., Luongo, C.L., Yang, L., Proia, A.D., Devincenzo, J.P., Collins, P.L. and Pickles, R.J. 2014. RSV-encoded NS2 promotes epithelial cell shedding and distal airway obstruction. *The Journal of Clinical Investigation*. **124**(5),pp.2219–2233.

Liljeroos, L., Anna, M., Helenius, A., Jane, S., Krzyzaniak, M.A., Helenius, A. and Butcher, S.J. 2013. Architecture of respiratory syncytial virus revealed by electron cryotomography. *Proceedings of the National Academy of Sciences of the United States of America*. **110**(27),pp.11133–8.

Lim, S.S., Vos, T., Flaxman, A.D., Danaei, G., Shibuya, K., Adair-Rohani, H., Amann, M., Anderson, H.R., Andrews, K.G., Aryee, M., Atkinson, C., Bacchus, L.J., Bahalim, A.N., Balakrishnan, K., Balmes, J., Barker-Collo, S., Baxter, A., Bell, M.L., Blore, J.D., Blyth, F., Bonner, C., Borges, G., Bourne, R., Boussinesq, M., Brauer, M., Brooks, P., Bruce, N.G., Brunekreef, B., Bryan-Hancock, C., Bucello, C., Buchbinder, R., Bull, F., Burnett, R.T., Byers, T.E., Calabria, B., Carapetis, J., Carnahan, E., Chafe, Z., Charlson, F., Chen, H., Chen, J.S., Cheng, A.T.-A., Child, J.C., Cohen, A., Colson, K.E., Cowie, B.C., Darby, S., Darling, S., Davis, A., Degenhardt, L., Dentener, F., Des Jarlais, D.C., Devries, K., Dherani, M., Ding, E.L., Dorsey, E.R., Driscoll, T., Edmond, K., Ali, S.E., Engell, R.E., Erwin, P.J., Fahimi, S., Falder, G., Farzadfar, F., Ferrari, A., Finucane, M.M., Flaxman, S., Fowkes, F.G.R., Freedman, G., Freeman, M.K., Gakidou, E., Ghosh, S., Giovannucci, E., Gmel, G., Graham, K., Grainger, R., Grant, B., Gunnell, D., Gutierrez, H.R., Hall, W., Hoek, H.W., Hogan, A., Hosgood, H.D., Hoy, D., Hu, H., Hubbell, B.J., Hutchings, S.J., Ibeanusi, S.E., Jacklyn, G.L., Jasrasaria, R., Jonas, J.B., Kan, H., Kanis, J. a, Kassebaum, N., Kawakami, N., Khang, Y.-H., Khatibzadeh, S., Khoo, J.-P., Kok, C., Laden, F., Lalloo, R., Lan, Q., Lathlean, T., Leasher, J.L., Leigh, J., Li, Y., Lin, J.K., Lipshultz, S.E., London, S., Lozano, R., Lu, Y., Mak, J., Malekzadeh, R., Mallinger, L., Marcenes, W., March, L., Marks, R., Martin, R., McGale, P., McGrath, J., Mehta, S., Mensah, G. a, Merriman, T.R., Micha, R., Michaud, C., Mishra, V., Mohd Hanafiah, K., Mokdad, A. a, Morawska, L., Mozaffarian, D., Murphy, T., Naghavi, M., Neal, B., Nelson, P.K., Nolla, J.M., Norman, R., Olives, C., Omer, S.B., Orchard, J., Osborne, R., Ostro, B., Page, A., Pandey, K.D., Parry, C.D.H., Passmore, E., Patra, J., Pearce, N., Pelizzari, P.M., Petzold, M., Phillips, M.R., Pope, D., Pope, C.A., Powles, J., Rao, M., Razavi, H., Rehfuess, E. a, Rehm, J.T., Ritz, B., Rivara, F.P., Roberts, T., Robinson, C., Rodriguez-Portales, J. a, Romieu, I., Room, R., Rosenfeld, L.C., Roy, A., Rushton, L., Salomon, J. a, Sampson, U., Sanchez-Riera, L., Sanman, E., Sapkota, A., Seedat, S., Shi, P., Shield, K., Shivakoti, R., Singh, G.M., Sleet, D. a, Smith, E., Smith, K.R., Stapelberg, N.J.C., Steenland, K., Stöckl, H., Stovner, L.J., Straif, K., Straney, L., Thurston, G.D., Tran, J.H., Van Dingenen, R., van

Donkelaar, A., Veerman, J.L., Vijayakumar, L., Weintraub, R., Weissman, M.M., White, R. a, Whiteford, H., Wiersma, S.T., Wilkinson, J.D., Williams, H.C., Williams, W., Wilson, N., Woolf, A.D., Yip, P., Zielinski, J.M., Lopez, A.D., Murray, C.J.L., Ezzati, M., AlMazroa, M. a and Memish, Z. a 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. **380**(9859),pp.2224–60.

Lionta, E., Spyrou, G., Vassilatis, D.K. and Cournia, Z. 2014. Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Current Topics in Medicinal Chemistry*. **14**(16),pp.1923–1938.

Liuzzi, M., Mason, S.W., Cartier, M., Lawetz, C., McCollum, R.S., Dansereau, N., Bolger, G., Lapeyre, N., Gaudette, Y., Lagacé, L., Massariol, M.-J., Dô, F., Whitehead, P., Lamarre, L., Scouten, E., Bordeleau, J., Landry, S., Rancourt, J., Fazal, G. and Simoneau, B. 2005. Inhibitors of respiratory syncytial virus replication target cotranscriptional mRNA guanylylation by viral RNA-dependent RNA polymerase. *Journal of Virology*. **79**(20),pp.13105–15.

Lo, M.S., Brazas, R.M. and Holtzman, M.J. 2005. Respiratory Syncytial Virus Nonstructural Proteins NS1 and NS2 Mediate Inhibition of Stat2 Expression and Alpha/Beta Interferon Responsiveness. *Journal of Virology*. 79(14),pp.9315-9319.

Lounnas, V., Ritschel, T., Kelder, J., McGuire, R., Bywater, R.P. and Foloppe, N. 2013. Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Computational and structural biotechnology Journal*. **5**(6),p.e201302011.

Lu, X., McDonald, S.M., Tortorici, M.A., Tao, Y.J., Vasquez-Del Carpio, R., Nibert, M.L., Patton, J.T. and Harrison, S.C. 2008. Mechanism for Coordinated RNA Packaging and Genome Replication by Rotavirus Polymerase VP1. *Structure*. **16**(11),pp.1678–1688.

Luo, Y., Bolon, B., Kahn, S., Bennett, B. D., Babu-Khan, S., Denis, P., Fan, W., Kha, H Zhang, J., Gong, Y., Martin, L., Louis, J. C., Yan, Q., Richards, W. G., Citron, M Vassar, R. 2001. Mice deficient in BACE1, the Alzheimer's beta-secretase, have normal phenotype and abolished beta-amyloid generation. Nat. Neuroscience. **4**, pp. 231–232.

Lyne, P. 2002. Structure-based virtual screening: an overview. *Drug Discovery Today*. **7**(20),pp.1047–1055.

Malhotra, R., Ward, M., Bright, H., Priest, R., Foster, M.R., Hurle, M., Blair, E. and Bird, M. 2003. Isolation and characterisation of potential respiratory syncytial virus receptor(s) on epithelial cells. *Microbes and Infection*. **5**(2),pp.123–133.

Mallipeddi, S.K., Lupiani, B. and Samal, S.K. 1996. Mapping the domains on the phosphoprotein of bovine respiratory syncytial virus required for N-P interaction using a two-hybrid system. *Journal of General Virology*. **77**(5),pp.1019–1023.

Martínez, M.J., Biedenkopf, N., Volchkova, V., Hartlieb, B., Alazard-Dany, N., Reynard, O., Becker, S. and Volchkov, V. 2008. Role of Ebola virus VP30 in transcription reinitiation. *Journal of Virology*. **82**(24),pp.12569–73.

Mason, S.W., Aberg, E., Lawetz, C., DeLong, R., Whitehead, P. and Liuzzi, M. 2003. Interaction between human respiratory syncytial virus (RSV) M2-1 and P proteins is required for reconstitution of M2-1-dependent RSV minigenome activity. *Journal of Virology*. **77**(19),pp.10670–10676.

McGivern, D.R., Collins, P.L. and Fearns, R. 2005. Identification of internal sequences in the 3' leader region of human respiratory syncytial virus that enhance transcription and confer replication processivity. *Journal of Virology*. **79**(4),pp.2449–2460.

McLellan, J.S., Yang, Y., Graham, B.S. and Kwong, P.D. 2011. Structure of Respiratory Syncytial Virus Fusion Glycoprotein in the Postfusion Conformation Reveals Preservation of Neutralizing Epitopes. *Journal of Virology*. **85**(15),pp.7788-7796.

McLellan, J.S., Chen, M., Leung, S., Graepel, K.W., Du, X., Yang, Y., Zhou, T., Baxa, U., Yasuda, E., Beaumont, T., Kumar, A., Modjarrad, K., Zheng, Z., Zhao, M., Xia, N., Kwong, P.D. and Graham, B.S. 2013. Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science.* **340**(6136),pp.1113–7.

Meng, J., Hotard, A.L., Currier, M.G., Lee, S., Stobart, C.C. and Moore, M.L. 2015. The Respiratory Syncytial Virus Attachment Glycoprotein Contribution to Infection Depends on the Specific Fusion Protein. *Journal of Virology*. **90**, pp.245–253.

Miller, M., Jaskólski, M., Rao, J.K.M., Leis, J. and Wlodawer, A. 1989. Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature*. **337**(9),pp.576–579.

Mitra, R., Baviskar, P., Duncan-Decocq, R.R., Patel, D. and Oomens, A.G.P. 2012. The human respiratory syncytial virus matrix protein is required for maturation of viral filaments. *Journal of Virology*. **86**(8),pp.4432–43.

Modrof, J., Becker, S. and Mühlberger, E. 2003. Ebola Virus Transcription Activator VP30 Is a Zinc-Binding Protein. *Journal of Virology*. **77**(5),pp.3334–3338.

Modrof, J., Mühlberger, E., Klenk, H.D. and Becker, S. 2002. Phosphorylation of VP30 impairs Ebola virus transcription. *Journal of Biological Chemistry*. **277**(36),pp.33099–33104.

Money, V.A., McPhee, H.K., Mosely, J.A., Sanderson, J.M. and Yeo, R.P. 2009. Surface features of a Mononegavirales matrix protein indicate sites of membrane interaction. *Proceedings of the National Academy of Sciences of the United States of America*. **106**(11),pp.4441–6.

Munir, S., Hillyer, P., Le Nouen, C., Buchholz, U.J., Rabin, R.L., Collins, P.L. and Bukreyev, A. 2011. Respiratory syncytial virus interferon antagonist NS1 protein suppresses and skews the human T lymphocyte response. *PLoS Pathog*. **7**(4),p.e1001336.

Murata, Y. and Falsey, A.R. 2007. Respiratory syncytial virus infection in adults. *Antiviral Therapy*. **12**(4 B),pp.659–670.

Mysinger, M.M., Carchia, M., Irwin, J.J. and Shoichet, B.K. 2012. Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking. *Journal of Medicinal Chemistry*. **55**(14),pp.6582–6594.

Nair, H., Nokes, D.J., Gessner, B.D., Dherani, M., Madhi, S.A., Singleton, R.J. and O'Brien, K.L. 2010. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *Lancet*. **376**(9746),pp.1074–1084.

Navaratnarajah, C.K., Oezguen, N., Rupp, L., Kay, L., Leonard, V.H.J., Braun, W. and Cattaneo, R. 2011. The heads of the measles virus attachment protein move to transmit the fusion-triggering signal. *Nature Structural & Molecular Biology*. **18**(2),pp.128–134.

Navia, M., Fitzgerald, P., McKeever, B., Leu, C.-T., Heimbach, J., Herber, W., Sigal, I., Darke, P. and Springer, J. 1989. Three-dimenional structrure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature*. **337**(16),pp.615–620.

Ni, L., Zhao, L., Qian, Y., Zhu, J., Jin, Z., Chen, Y.W., Tien, P. and Gao, G.F. 2005. Design and characterization of human respiratory syncytial virus entry inhibitors. *Antiviral Therapy*. **10**,pp.833–840.

Nienaber, V.L., Richardson, P.L., Klighofer, V., Bouska, J.J., Giranda, V.L. and Greer, J. 2000. Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nature biotechnology*. **18**(10),pp.1105–1108.

Nieva, J.L., Madan, V. and Carrasco, L. 2012. Viroporins: structure and biological functions. *Nature reviews. Microbiology*. **10**(8),pp.563–74.

Nisole, S., Krust, B., Callebaut, C., Guichard, G., Muller, S., Briand, J.P. and Hovanessian, A.G. 1999. The anti-HIV pseudopeptide HE-19 forms a complex with the cell-surface-expressed nucleolin independent of heparan sulfate proteoglycans. *Journal of Biological Chemistry*. **274**(39),pp.27875–27884.

Noton, S.L., Nagendra, K., Dunn, E.F., Mawhorter, M.E., Yu, Q. and Fearns, R. 2015. Respiratory Syncytial Virus Inhibitor AZ-27 Differentially Inhibits Different Polymerase Activities at the Promoter. *Journal of Virology*. **89**(15),pp.7786–7798.

O 'brien, K.L., Chandran, A., Weatherholtz, R., Jafri, H.S., Griffi, M.P., Bellamy, T., Millar, E. V, Jensen, K.M., Harris, B.S., Reid, R., Moulton, L.H., Losonsky, G.A., Karron, R.A. and Santosham, M. 2015. Effi cacy of motavizumab for the prevention of respiratory syncytial virus disease in healthy Native American infants: a phase 3 randomised double-blind placebo-controlled trial. *The Lancet Infectious Diseases*. **15**,pp.1398–1408.

Oltersdorf, T., Elmore, S.W., Shoemaker, A.R., Armstrong, R.C., Augeri, D.J., Belli, B.A., Bruncko, M., Deckwerth, T.L., Dinges, J., Hajduk, P.J., Joseph, M.K., Kitada, S., Korsmeyer, S.J., Kunzer, A.R., Letai, A., Li, C., Mitten, M.J., Nettesheim, D.G., Ng, S., Nimmer, P.M., O'Connor, J.M., Oleksijew, A., Petros, A.M., Reed, J.C., Shen, W., Tahir, S.K., Thompson, C.B., Tomaselli, K.J., Wang, B., Wendt, M.D., Zhang, H., Fesik, S.W. and Rosenberg, S.H. 2005. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature*. **435**(7042),pp.677–681.

Ouizougun-Oubari, M., Pereira, N., Tarus, B., Galloux, M., Lassoued, S., Fix, J., Tortorici, M.A., Hoos, S., Baron, B., England, P., Desmaële, D., Couvreur, P., Bontems, F., Rey, F.A., Eléouët, J.-F., Sizun, C., Slama-Schwok, A. and Duquerroy, S. 2015. A Druggable Pocket at the Nucleocapsid/Phosphoprotein Interaction Site of Human Respiratory Syncytial Virus. *Journal of Virology*. **89**,pp.11129–11143.

Paccaud M.F., Jacquier C.L. 1970. A respiratory syncytial virus of bovine origin, *Arch. Gesamte Virusforsch.* **30**,pp.327–342.

Pastey, M.K. and Samal, S.K. 1995. Nucleotide sequence analysis of the non-structural NS1 (1C) and NS2 (1B) protein genes of bovine respiratory syncytial virus. Journal of General Virology. **76**(1),pp.193–197.

Paton, D.J., Christiansen, K.H., Alenius, S., Cranwell, M.P., Pritchard, G.C. and Drew, T.W. 1998. Prevalence of antibodies to bovine virus diarrhoea virus and other viruses in bulk tank milk in England and Wales. The Veterinary record. 142(15),p.385 — 391.

Pearce, N.M., Bradley, A.R., Krojer, T., Marsden, B.D., Deane, C.M. and Delft, F. Von 2017. Partial-occupancy binders identified by the Pan-Dataset Density Analysis method offer new chemical opportunities and reveal cryptic binding sites. *Structural Dynamics*. **4**(3),pp.32104–1.

Pearce, N.M., Krojer, T., Bradley, A.R., Collins, P., Nowak, R.P., Talon, R., Marsden, B.D., Kelm, S., Shi, J., Deane, C.M. and von Delft, F. 2017. A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nature communications*. **8**,p.15123.

Pearce, N.M., Krojer, T. and Von Delft, F. 2017. Proper modelling of ligand binding requires an ensemble of bound and unbound states. *Acta Crystallographica Section D: Structural Biology*. **73**,pp.256–266.

Perez, M., Garcia-Barreno, B., Melero, J.A., Carrasco, L. and Guinea, R. 1997. Membrane permeability changes induced in Escherichia coli by the SH protein of human respiratory syncytial virus. *Virology*. **235**(2),pp.342–351.

Pernet, O., Pohl, C., Ainouze, M., Kweder, H. and Buckland, R. 2009. Nipah virus entry can occur by macropinocytosis. *Virology*. **395**(2),pp.298–311.

Pihan, E., Colliandre, L., Guichou, J.F. and Douguet, D. 2012. E-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design. *Bioinformatics*. **28**(11),pp.1540–1541.

Pollard, T.D. 2010. A guide to simple and informative binding assays. *Molecular biology of the cell*. **21**,pp.4061–4067.

Pringle, C.R. 1991. The order Mononegavirales. *Archives of Virology*. **117**(1),pp.137–140.

Qiu, S., Ogino, M., Luo, M., Ogino, T. and Green, T.J. 2016. Structure and function of the N-terminal domain of the vesicular stomatitis virus RNA polymerase. *J Virol*. **90**,pp.715–724.

Raab De Verdugo, U., Selinka, H.-C., Huber, M., Kramer, B., Kellermann, J., Hofschneider, P.H. and Kandolf, R. 1995. Characterization of a 100-Kilodalton

Binding Protein for the Six Serotypes of Coxsackie B Viruses. *JOURNAL OF VIROLOGY*. **69**(11),pp.6751–6757.

Radhakrishnan, A., Yeo, D., Brown, G., Myaing, M.Z., Iyer, L.R., Fleck, R., Tan, B.-H., Aitken, J., Sanmun, D., Tang, K., Yarwood, A., Brink, J. and Sugrue, R.J. 2010. Protein Analysis of Purified Respiratory Syncytial Virus Particles Reveals an Important Role for Heat Shock Protein 90 in Virus Particle Assembly. *Molecular & Cellular Proteomics*.,pp.1829–1848.

Rahaman, A., Srinivasan, N., Shamala, N., Shaila, M.S., Castagne, N., Barbier, A., Bernard, J., Rezaei, H., Huet, J.-C., Henry, C., Da Costa, B., Eleouet, J.-F., Llorente, M.T., García-Barreno, B., Calero, M., Camafeita, E., López, J.A., Longhi, S., Ferrón, F., Varela, P.F., Melero, J.A., Choudhary, S.K., Malur, A.G., Huo, Y., De, B.P., Banerjee, A.K., CURRAN, J., BOECK, R., LIN-MARQ, N., LUPAS, A., KOLAKOFSKY, D., Bousse, T., Takimoto, T., Matrosovich, T., Portner, A., Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W.P., Ruigrok, R.W.H. and Marion, D. 2004. Paramyxovirus Phosphoproteins Form Homotrimers as Determined by an Epitope Dilution Assay, via Predicted Coiled Coils. *Virology*. **279**(22),pp.23606–23614.

Ramaswamy, M., Shi, L., Varga, S.M., Barik, S., Behlke, M.A. and Look, D.C. 2006. Respiratory syncytial virus nonstructural protein 2 specifically inhibits type I interferon signal transduction. *Virology*. **344**(2),pp.328–339.

Rarey, M., Kramer, B., Lengauer, T. and Klebe, G. 1996. A Fast Flexible Docking Method using an Incremental Construction Algorithm. *J. Mol. Biol*. **261**,pp.470–489.

Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crépin, T., Hart, D., Lunardi, T., Nanao, M., Ruigrok, R.W.H. and Cusack, S. 2014. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature*. **516**(7531),pp.361–366.

Rincheval, V., Lelek, M., Gault, E., Bouillier, C., Sitterlin, D., Blouquit-Laye, S., Galloux, M., Zimmer, C., Eleouet, J.F. and Rameix-Welti, M.A. 2017. Functional organization of cytoplasmic inclusion bodies in cells infected by respiratory syncytial virus. *Nature Communications*. **8**(1),pp.1–11.

Ripphausen, P., Stumpfe, D and Bajorath, J. 2012. Analysis of structure-based virtual screening and characterisation of identified active compounds. *Future Science. **4**(5),pp.603-613.*

Roberts, S.R., Compans, R.W. and Wertz, G.W. 1995. Respiratory syncytial virus matures at the apical surfaces of polarized epithelial cells. *Journal of Virology*. **69**(4),pp.2667–73.

Rossio JL, Esser MT, Suryanarayana K, Schneider DK, Bess JW, Jr, Vasquez GM, Wiltrout TA, Chertova E, Grimes MK, Sattentau Q, Arthur LO, Henderson LE, Lifson JD. 1998.Inactivation of human immunodeficiency virus type 1 infectivity with preservation of conformational and functional integrity of virion surface proteins. Journal of Virology. **72**,pp.7992–8001.

Roymans, D., De Bondt, H.L., Arnoult, E., Geluykens, P., Gevers, T., Van Ginderen, M., Verheyen, N., Kim, H., Willebrords, R., Bonfanti, J.-F., Bruinzeel,

W., Cummings, M.D., van Vlijmen, H. and Andries, K. 2010. Binding of a potent small-molecule inhibitor of six-helix bundle formation requires interactions with both heptad-repeats of the RSV fusion protein. *Proceedings of the National Academy of Sciences of the United States of America*. **107**(1),pp.308–13.

Ruigrok, R.W.H., Crépin, T. and Kolakofsky, D. 2011. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Current Opinion in Microbiology*. **14**(4),pp.504–510.

Russell, R.J., Haire, L.F., Stevens, D.J., Collins, P.J., Lin, Y.P., Blackburn, G.M., Hay, A.J., Gamblin, S.J. and Skehel, J.J. 2006. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature*. **443**(7107),pp.45–49.

Russell, R.F., McDonald, U., Ivanova, M., Zhong, Z., Burkreyev, A., Tregoning, J. 2015. Partial Attenuation of Respiratory Syncytial Virus with a Deletion of a Small Hydrophobic Gene is Associated with Elevated Interluekin-IB Responses. *Journal of Virology.* **89** (17),pp. 8974-8981.

Sande, C.J., Cane, P.A. and Nokes, D.J. 2014. The association between age and the development of respiratory syncytial virus neutralising antibody responses following natural infection in infants. *Vaccine*. **32**,pp.4726–4729.

Saravia, J., You, D., Shrestha, B., Jaligama, S., Siefker, D., Lee, G.I., Harding, J.N., Jones, T.L., Rovnaghi, C., Bagga, B., DeVincenzo, J.P. and Cormier, S.A. 2015. Respiratory Syncytial Virus Disease Is Mediated by Age-Variable IL-33. *PLoS Pathogens*. **11**(10),p.e1005217.

Sarmiento-Silva, R.E., Nakamura-Lopez, Y. and Vaughan, G. 2012. Epidemiology, molecular epidemiology and evolution of bovine respiratory syncytial virus. *Viruses*. **4**(12),pp.3452–3467.

Schlender, J., Bossert, B., Buchholz, U. and Conzelmann, K.-K. 2000. Bovine Respiratory Syncytial Virus Nonstructural Proteins NS1 and NS2 Cooperatively Antagonize Alpha/Beta Interferon-Induced Antiviral Response. *Journal of Virology*. **74**(18),pp.8234–8242.

Schmidt, U., Beyer, J., Polster, U., Gershwin, L.J. and Buchholz, U.J. 2002. Mucosal immunization with live recombinant bovine respiratory syncytial virus (BRSV) and recombinant BRSV lacking the envelope glycoprotein G protects against challenge with wild-type BRSV. *J Virol*. **76**(23),pp.12355–12359.

Schmitz, J., Owyang, A., Oldham, E., Song, Y., Murphy, E., McClanahan, T.K., Zurawski, G., Moshrefi, M., Qin, J., Li, X., Gorman, D.M., Bazan, J.F. and Kastelein, R.A. 2005. IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. *Immunity*. **23**(5),pp.479–490.

Schnell, J.R. and Chou, J.J. 2008. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature*. **451**(31),pp.591–596.

Schobel, S.A., Stucker, K.M., Moore, M.L., Anderson, L.J., Larkin, E.K., Shankar, J., Bera, J., Puri, V., Shilts, M.H., Rosas-Salazar, C., Halpin, R.A., Fedorova, N., Shrivastava, S., Stockwell, T.B., Peebles, R.S., Hartert, T. V. and Das, S.R. 2016.

Respiratory Syncytial Virus whole-genome sequencing identifies convergent evolution of sequence duplication in the C-terminus of the G gene. *Scientific Reports*. **6**,p.26311.

Schreiber, P., Matheise, J.P., Dessy, F., Heimann, M., Letesson, J.J., Coppe, P. and Collard, A. 2000. High mortality rate associated with Bovine Respiratory Syncytial Vires (BRSV) infection in Belgian White Blue calves previously vaccinated with an inactivated BRSV vaccine. *Journal of Veterinary Medicine, Series B*. **47**(7),pp.535–550.

Scior, T., Bender, A., Tresadern, G., Medina-Franco, J.L., Martínez-Mayorga, K., Langer, T., Cuanalo-Contreras, K. and Agrafiotis, D.K. 2012. Recognizing pitfalls in virtual screening: A critical review. *Journal of Chemical Information and Modeling*. **52**(4),pp.867–881.

Simmons, K.J., Chopra, I. and Fishwick, C.W.G. 2010. Structure-based discovery of antibacterial drugs. *Nature reviews. Microbiology*. **8**(7),pp.501–510.

Simmons, K.J., Gotfryd, K., Billesbølle, C.B., Loland, C.J., Gether, U., Fishwick, C.W.G. and Johnson, A.P. 2013. A virtual high-throughput screening approach to the discovery of novel inhibitors of the bacterial leucine transporter, LeuT. *Molecular Membrane Biology*. **30**(2),pp.184–194.

Simoes, E.A.F., Groothuis, J.R., Carbonell-Estrany, X., Rieger, C.H.L., Mitchell, I., Fredrick, L.M. and Kimpen, J.L.L. 2007. Palivizumab Prophylaxis, Respiratory Syncytial Virus, and Subsequent Recurrent Wheezing. *Journal of Pediatrics*.,pp.34–42.

Smerdon, S.J., Jäger, J., Wang, J., Kohlstaedt, L.A., Chirino, A.J., Friedman, J.M., Rice, P. a and Steitz, T.A. 1994. Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proceedings of the National Academy of Sciences of the United States of America*. **91**(9),pp.3911–5.

Smith, M-H., Frey, M-L and Derks, R-E. 1975. Isolation, characterization and pathogenicity studies of a bovine respiratory syncytial virus. *Archives of Virology*. **47**(3),pp.237-247

Smith, G., Raghunandan, R., Wu, Y., Liu, Y., Massare, M., Nathan, M., Zhou, B., Lu, H., Boddapati, S., Li, J., Flyer, D. and Glenn, G. 2012. Respiratory Syncytial Virus Fusion Glycoprotein Expressed in Insect Cells Form Protein Nanoparticles That Induce Protective Immunity in Cotton Rats. *PLoS ONE*. **7**(11),p.e50852.

Sourimant, J., Rameix-Welti, M.-A., Gaillard, A.-L., Chevret, D., Galloux, M., Gault, E. and Eléouët, J.-F. 2015. Fine Mapping and Characterization of the L-Polymerase-Binding Domain of the Respiratory Syncytial Virus Phosphoprotein. *Journal of Virology*. **89**,pp.4421–4433.

Spann, K.M., Tran, K.-C., Chi, B., Rabin, R.L. and Collins, P.L. 2004. Suppression of the Induction of Alpha, Beta, and Gamma Interferons by the NS1 and NS2 Proteins of Human Respiratory Syncytial Virus in Human Epithelial Cells and Macrophages. *Journal of Virology*. **78**(8),pp.4363–4369.

Stamford, A. and Strickland, C. 2013. Inhibitors of BACE for treating alzheimer's disease: A fragment-based drug discovery story. *Current Opinion in Chemical Biology*. **17**,pp.320–328.

Steven W.J. Lamberts, Hajo A.Bruining, F.H.D.J. 1997. Review Article Drug Therapy. *The New England Journal of medicine*. **337**(18),pp.1285–1292.

Stillman, E.A. and Whitt, M.A. 1997. Mutational analyses of the intergenic dinucleotide and the transcriptional start sequence of vesicular stomatitis virus (VSV) define sequences required for efficient termination and initiation of VSV transcripts. *Journal of Virology*. **71**(3),pp.2127–37.

Stott, B.E., Thomas, L.H., Collins, A.P., Crouch, S., Jebbett, J., Smith, G.S., Luther, P.D. and Caswell, R. 1980. A survey of virus infections of the respiratory tract of cattle and their association with disease. *J. Hyg., Camb*. **85**,pp.257–2.

Sun, Z., Pan, Y., Jiang, S. and Lu, L. 2013. Respiratory syncytial virus entry inhibitors targeting the F protein. *Viruses*. **5**(1),pp.211–25.

Tang, R.S., Nguyen, N., Cheng, X. and Jin, H. 2001. Requirement of Cysteines and Length of the Human Respiratory Syncytial Virus M2-1 Protein for Protein Function and Virus Viability. *Society*. **75**(23),pp.11328–11335.

Tanner, S.J., Ariza, A., Richard, C.-A., Kyle, H.F., Dods, R.L., Blondot, M.-L., Wu, W., Trincão, J., Trinh, C.H., Hiscox, J.A., Carroll, M.W., Silman, N.J., Eléouët, J.-F., Edwards, T.A. and Barr, J.N. 2014. Crystal structure of the essential transcription antiterminator M2-1 protein of human respiratory syncytial virus and implications of its phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*. **111**(4),pp.1580–1585.

Tarbouriech, N., Curran, J., Ruigrok, R.W.H. and Burmeister, W.P. 2000. Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nat Struct Biol*. **7**(9),pp.777–781..

Tawar, R.G., Duquerroy, S., Vonrhein, C., Varela, P.F., Damier-Piolle, L., Castagne, N., MacLellan, K., Bedouelle, H., Bricogne, G., Bhella, D., Eleouet, J.-F. and Rey, F.A. 2009. Crystal Structure of a Nucleocapsid-Like Nucleoprotein-RNA Complex of Respiratory Syncytial Virus. *Science*. **326**(5957),pp.1279–1283.

Taylor, G., Wyld, S., Valarcher, J.F., Guzman, E., Thom, M., Widdison, S. and Buchholz, U.J. 2014. Recombinant bovine respiratory syncytial virus with deletion of the SH gene induces increased apoptosis and pro-inflammatory cytokines in vitro, and is attenuated and induces protective immunity in calves. *Journal of General Virology*.**95**,pp.1244-1254.

Taylor, G., Thom, M., Capone, S., Pierantoni, A., Guzman, E., Herbert, R., Scarselli, E., Mapolitano, F., Giuliani, A., Folgori, A., Colloca, S., Cortese, R., Nicosia, A. and Vitelli, A. 2015. Efficacy of a virus vectored vaccine against human and bovine respiratory syncytial virus infections. *Sci Transl Med*. **7**(300),pp.300ra127.

Tayyari, F., Marchant, D., Moraes, T.J., Duan, W., Mastrangelo, P. and Hegele, R.G. 2011. Identification of nucleolin as a cellular receptor for human respiratory syncytial virus. *Nature medicine*. **17**(9),pp.1132–5.

Techaarpornkul, S., Barretto, N. and Peeples, M.E. 2001. Functional analysis of recombinant respiratory syncytial virus deletion mutants lacking the small hydrophobic and/or attachment glycoprotein gene. *Journal of Virology*. **75**(15),pp.6825–34.

Teng, M.N., Whitehead, S.S., Bermingham, A., St Claire, M., Elkins, W.R., Murphy, B.R. and Collins, P.L. 2000. Recombinant respiratory syncytial virus that does not express the NS1 or M2-2 protein is highly attenuated and immunogenic in chimpanzees. *Journal of Virology*. **74**(19),pp.9317–9321.

Teng, M.N., Whitehead, S.S. and Collins, P.L. 2001. Contribution of the respiratory syncytial virus G glycoprotein and its secreted and membrane-bound forms to virus replication in vitro and in vivo. *Virology*. **289**(2),pp.283–96.

The IMpact-RSV Study Group. 1998. Palivizumab, a Humanized Respiratory Syncytial Virus Monoclonal Antibody, Reduces Hospitalization From Respiratory Syncytial Virus Infection in High-risk Infants. *Pediatrics.* **102**(3).

Tiong-Yip, C.L., Aschenbrenner, L., Johnson, K.D., McLaughlin, R.E., Fan, J., Challa, S., Xiong, H. and Yu, Q. 2014. Characterization of a respiratory syncytial virus L protein inhibitor. *Antimicrobial Agents and Chemotherapy*. **58**(7),pp.3867–3873.

Totrov, M. and Abagyan, R. 2008. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Current Opinion in Structural Biology*. **18**(2),pp.178–184.

Tran, D.N., Pham, T.M.H., Ha, M.T., Tran, T.T.L., Dang, T.K.H., Yoshida, L.-M., Okitsu, S., Hayakawa, S., Mizuguchi, M. and Ushijima, H. 2013. Molecular epidemiology and disease severity of human respiratory syncytial virus in Vietnam. *PloS one*. **8**(1),p.e45436.

Tran, T.L., Castagné, N., Bhella, D., Varela, P.F., Bernard, J., Chilmonczyk, S., Berkenkamp, S., Benhamo, V., Grznarova, K., Grosclaude, J., Nespoulos, C., Rey, F.A. and Eléouët, J.F. 2007. 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 promoter. *Journal of General Virology*. **88**(1),pp.196–206.

Tran, T.-L.T.-L., Castagne, N., Dubosclard, V., Noinville, S., Koch, E., Moudjou, M., Henry, C., Bernard, J., Yeo, R.P., Eleouet, J.-F., Castagné, N., Dubosclard, V., Noinville, S., Koch, E., Moudjou, M., Henry, C., Bernard, J., Yeo, R.P. and Eléouët, J.-F. 2009. The respiratory syncytial virus M2-1 protein forms tetramers and interacts with RNA and P in a competitive manner. *Journal of Virology*. **83**(13),pp.6363–74.

Tremaglio, C.Z., Noton, S.L., Deflubé, L.R. and Fearns, R. 2013. Respiratory syncytial virus polymerase can initiate transcription from position 3 of the leader promoter. *Journal of Virology*. **87**(6),pp.3196–207.

Tulloch, F., Pathania, U., Luke, G.A., Nicholson, J., Stonehouse, N.J., Rowlands, D.J., Jackson, T., Tuthill, T., Haas, J., Lamond, A.I. and Ryan, M.D. 2014. FMDV replicons encoding green fluorescent protein are replication competent. *Journal of Virological Methods*. **209**,pp.35–40.

Valarcher, J-F and Taylor, G. 2007. Bovine Respiratory Syncytial Virus Infection, *Veterinary Research*. **38**,pp.153-180

Valarcher, J.-F., Furze, J., Wyld, S., Cook, R., Conzelmann, K.-K. and Taylor, G. 2003. Role of Alpha/beta interferons in the attenuation and immunogenicity of recombinant bovine respiratory syncytial viruses lacking NS proteins. *Journal of Virology*. **77**(15),pp.8426–8439.

Van Der Poel, W.H.M., Brand, A., Kramps, J.A. and Oirschott, J.T. Van 1994. Respiratory syncytial virus infections in human beings and in cattle. *Journal of Infection*. **29**,pp.215–228.

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W. and Taylor, R.D. 2003. Improved protein-ligand docking using GOLD. *Proteins: Structure, Function and Genetics*. **52**(4),pp.609–623.

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W. and Taylor, R.D. 2003. Improved protein-ligand docking using GOLD. *Proteins: Structure, Function and Genetics*. **52**(4),pp.609–623.

von Itzstein, M., Wu, W., Kok, G., M, P., J, D., Jin, B., Phan, T., SMythe, M., White, H., Oliver, S., Colman, P., Varghese, J., Ryan, M., Woods, J., Bethell, R., Hotham, V., Cameron, J. and Penn, C. 1993. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature*. **363**,pp.418–423.

von Itzstein, M. 2007. The war against influenza: discovery and development of sialidase inhibitors. *Nature Reviews Drug Discovery*. **6**(12),pp.967–974.

Vreede, F.T., Jung, T.E. and Brownlee, G.G. 2004. Model suggesting that replication of influenza virus is regulated by stabilization of replicative intermediates. *Journal of Virology*. **78**(17),pp.9568–72.

Wang, Y. Sen, Strickland, C., Voigt, J.H., Kennedy, M.E., Beyer, B.M., Senior, M.M., Smith, E.M., Nechuta, T.L., Madison, V.S., Czarniecki, M., McKittrick, B.A., Stamford, A.W., Parker, E.M., Hunter, J.C., Greenlee, W.J. and Wyss, D.F. 2010. Application of fragment-based NMR screening, X-ray crystallography, structure-based design, and focused chemical library design to identify novel uM leads for the development of nM BACE-1 inhibitors. *Journal of Medicinal Chemistry*. **53**(3),pp.942–950.

Wellemans, G.; Leunen, J.; Luchsinger, E. 1970. Respiratory ailments of cattle: isolation of a virus (220/69) with serologic resemblance to the human respiratory syncytial virus. *Ann. Med. Vet*. **114**,pp.89–93.3.

White, L., Waris, M., Cane, P., Nokes, J. and Medley, G. 2004. The transmission dynamics of groups A and B human respiratory syncytial virus (hRSV) in England & Wales and Finland: seasonality and cross-protection. *Epidemiol Infect*. **133**,pp.279–289.

Whitehead, S.S., Bukreyev, A., Teng, M.N., Firestone, C.Y., St Claire, M., Elkins, W.R., Collins, P.L. and Murphy, B.R. 1999. Recombinant respiratory syncytial virus bearing a deletion of either the NS2 or SH gene is attenuated in chimpanzees. *Journal of Virology*. **73**(4),pp.3438–3442.

Wilfinger, W., Mackey, K and Chomczynski, P. 1997. Effect of pH and Ionic Strength on the Spectrophotometric Assessment of Nucleic Acid Purity. *BioTechniques*. **22**,pp.474-481

Worthington, M.T., Amann, B.T., Nathans, D. and Berg, J.M. 1996. Metal binding properties and secondary structure of the zinc-binding domain of Nup475. *Proceedings of the National Academy of Sciences of the United States of America*. **93**(24),pp.13754–9.

Wu, P. and Hartert, T. 2011. Evidence for a causal relationship between respiratory syncytial virus infection and asthma. *National Institute of Health*. **9**(9),pp.731–745.

Xiong, H., Foulk, M., Aschenbrenner, L., Fan, J., Tiong-Yip, C.-L., Johnson, K.D., Moustakas, D., Fleming, P.R., Brown, D.G., Zhang, M., Ferguson, D., Wu, D. and Yu, Q. 2013. Discovery of a potent respiratory syncytial virus RNA polymerase inhibitor. *Bioorganic & medicinal chemistry letters*. **23**(24),pp.6789–6793.

Xu, W., Luthra, P., Wu, C., Batra, J., Leung, D.W., Basler, C.F. and Amarasinghe, G.K. 2017. Ebola virus VP30 and nucleoprotein interactions modulate viral RNA synthesis. *Nature Communications*. **8**,p.15576.

York, A., Hengrung, N., Vreede, F.T., Huiskonen, J.T. and Fodor, E. 2013. Isolation and characterization of the positive-sense replicative intermediate of a negative strand RNA virus. *Proc Natl Acad Sci USA*. **110**,pp.E4238–E4245.

Yu, Q., Hardy, R.W. and Wertz, G.W. 1995. Functional cDNA clones of the human respiratory syncytial (RS) virus N, P, and L proteins support replication of RS virus genomic RNA analogs and define minimal trans-acting requirements for RNA replication. *Journal of Virology*. **69**(4),pp.2412–9.

Zhao, X., Singh, M., Malashkevich, V.N. and Kim, P.S. 2000. Structural characterization of the human respiratory syncytial virus fusion protein core. *Proceedings of the National Academy of Sciences of the United States of America*. . **97**(26),pp.14172–7.

Zhou, H., Cheng, X. and Jin, H. 2003. Identification of Amino Acids That Are Critical to the Processivity Function of Respiratory Syncytial Virus M2-1 Protein. *Journal of Virology*. **77**(9),pp.5046–5053.

Zhu, Z., Sun, Z.Y., Ye, Y., Voigt, J., Strickland, C., Smith, E.M., Cumming, J., Wang, L., Wong, J., Wang, Y. Sen, Wyss, D.F., Chen, X., Kuvelkar, R., Kennedy, M.E., Favreau, L., Parker, E., McKittrick, B.A., Stamford, A., Czarniecki, M., Greenlee, W. and Hunter, J.C. 2010. Discovery of cyclic acylguanidines as highly potent and selective ??-site amyloid cleaving enzyme (BACE) inhibitors: Part I - Inhibitor design and validation. *Journal of Medicinal Chemistry*. **53**(3),pp.951–965.

Zimmer, G., Conzelmann, K.-K. and Herrler, G. 2002. Cleavage at the Furin Consensus Sequence RAR/KR 109 and Presence of the Intervening Peptide of the Respiratory Syncytial Virus Fusion Protein Are Dispensable for Virus Replication in Cell Culture. *Journal of Virology* **76**(18),pp.9218–9224.

Zimmer, G., Rohn, M., McGregor, G.P., Schemann, M., Conzelmann, K.K. and Herrler, G. 2003. Virokinin, a Bioactive Peptide of the Tachykinin Family, Is

Released from the Fusion Protein of Bovine Respiratory Syncytial Virus. *Journal of Biological Chemistry*. **278**(47),pp.46854–46861.

Zsoldos, Z., Reid, D., Simon, A., Sadjad, S.B. and Johnson, A.P. 2007. eHiTS: A new fast, exhaustive flexible ligand docking system. *Journal of Molecular Graphics and Modelling*. **26**(1),pp.198–212.

Appendix I

HRSV M2-1, BRSV M2-1 proteins were cloned and expressed in the pGEX-6P-2 vector. All clones were sequenced via sanger sequencing prior to further analysis. Below are the nucleotide sequences obtained and the inferred amino acid sequence.

Key: non-native residues

HRSV M2-1 (length-194 amino acids)

```
gggcccctgggatccatgtcacgaaggaatccttgcaaatttgaaattcgaggtcattgc
 G   P   L   G   S   M   S   R   R   N   P   C   K   F   E   I   R   G   H   C
ttaaatggtaagaggtgtcattttagtcataattattttgaatggccaccccatgcactg
 L   N   G   K   R   C   H   F   S   H   N   Y   F   E   W   P   P   H   A   L
cttgtaagacaaaactttatgttaaacagaatacttaagtctatggataaaagtatagat
 L   V   R   Q   N   F   M   L   N   R   I   L   K   S   M   D   K   S   I   D
accttatcagaaataagtggagctgcagagttggacagaacagaagagtatgctcttggt
 T   L   S   E   I   S   G   A   A   E   L   D   R   T   E   E   Y   A   L   G
gtagttggagtgctagagagttatataggatcaataaacaatataactaaacaatcagca
 V   V   G   V   L   E   S   Y   I   G   S   I   N   N   I   T   K   Q   S   A
tgtgttgccatgagcaaactcctcactgaactcaatagtgatgatatcaaaaagctgagg
 C   V   A   M   S   K   L   L   T   E   L   N   S   D   D   I   K   K   L   R
gacaatgaagagctaaattcacccaagataagagtgtacaatactgtcatatcatatatt
 D   N   E   E   L   N   S   P   K   I   R   V   Y   N   T   V   I   S   Y   I
gaaagcaacaggaaaaacaataaacaaactatccatctgttaaaaagattgccagcagac
 E   S   N   R   K   N   N   K   Q   T   I   H   L   L   K   R   L   P   A   D
gtattgaagaaaaccatcaaaaacacattggatatccataagagcataaccatcaacaac
 V   L   K   K   T   I   K   N   T   L   D   I   H   K   S   I   T   I   N   N
ccaaaagaatcaactgttagtgatacaaatgaccatgccaaaaataatgatactacctaa
 P   K   E   S   T   V   S   D   T   N   D   H   A   K   N   N   D   T   T   –
```

BRSV M2-1 (length- 186 amino acids)

```
gggcccctgggatccatgagccgtcgtaatccgtgtaaatatgaaattcgtggtcattgc
 G   P   L   G   S   M   S   R   R   N   P   C   K   Y   E   I   R   G   H   C
ctgaacggcaaaaaatgtcattttagccacaactattttgagtggcctccgcatgcactg
 L   N   G   K   K   C   H   F   S   H   N   Y   F   E   W   P   P   H   A   L
ctggttcgtcagaattttatgctgaacaaaatcctgaaaagcatggatcgcaataatgat
 L   V   R   Q   N   F   M   L   N   K   I   L   K   S   M   D   R   N   N   D
accctgagcgaaattagcggtgcagcagaactggatcgtaccgaagaatatgcactgggt
 T   L   S   E   I   S   G   A   A   E   L   D   R   T   E   E   Y   A   L   G
gttattggtgttctggaaagctatctgagcagcattaacaacattaccaaacagagcgca
 V   I   G   V   L   E   S   Y   L   S   S   I   N   N   I   T   K   Q   S   A
tgtgttgcaatgagcaaactgctggcagaaattaacaacgatgatattaaacgcctgcgc
 C   V   A   M   S   K   L   L   A   E   I   N   N   D   D   I   K   R   L   R
aataaagaagttccgaccagcccgaaaattcgcatttataacaccgtgatcagctacatc
 N   K   E   V   P   T   S   P   K   I   R   I   Y   N   T   V   I   S   Y   I
gatagcaataaacgtaatacgaaacagaccatccatctgctgaaacgtctgcctgcagat
 D   S   N   K   R   N   T   K   Q   T   I   H   L   L   K   R   L   P   A   D
gttctgaaaaaaaccattaaaaacaccatcgacatccacaacgaaatcaatggtaataac
 V   L   K   K   T   I   K   N   T   I   D   I   H   N   E   I   N   G   N   N
cagggcgatatcaatgtggatgaacagaacgaataa
 Q   G   D   I   N   V   D   E   Q   N   E   –
```

Appendix II

SEC column elution profile and calibration curve as taken from the EMBL Protein purification website

(https://www.embl.de/pepcore/pepcore_services/protein_purification/chromatography/hiload26-60_superdex75/index.html)

Calibration of 26/60 HiLoad Superdex 75 prep grade

$$V_{elution} = 631.3 - 104.3 \times \log(M_r)$$
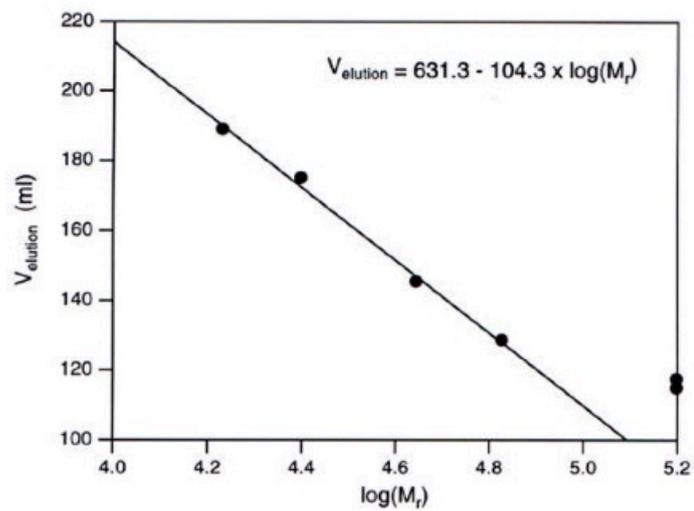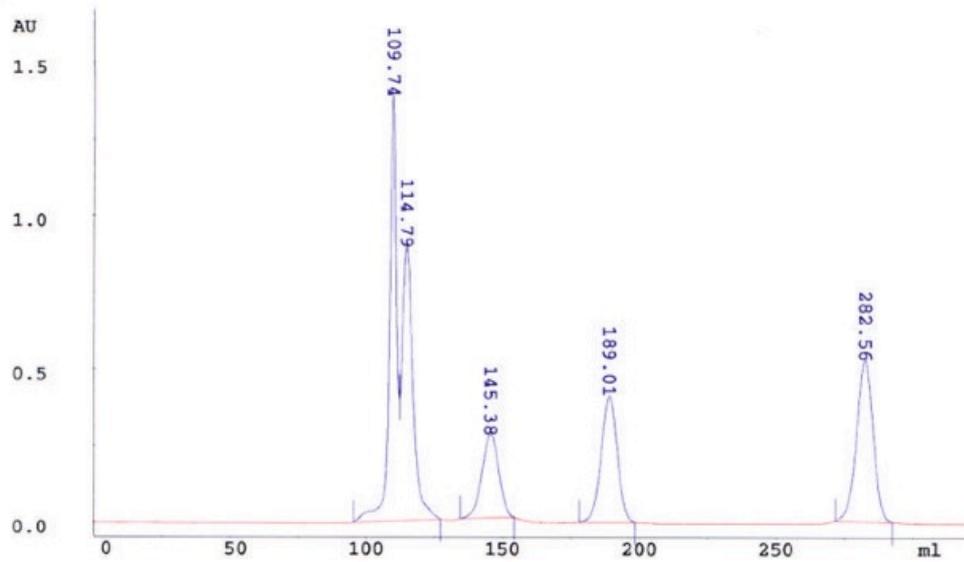
Appendix III

Commercial factorial crystal screens were assessed for their ability to sustain crystal formation. Each screen with 96 conditions is shown below, with a purple box indicating a condition that resulted in crystal growth. Well numbers refer to those of the commercial screens to allow the correct condition to be referenced. The following screens were used:

For M2-1 protein

- Morpheus
- Midas
- PACT
- Wizard 1 & 2/ 3 & 4
- Salt
- Index
- Crystal

For M2-1 protein in complex with P 90-110:

- Morpheus
- Midas
- PACT
- Wizard 1 & 2 / 3 & 4
- Index

Commercial Screen: Index (Hampton Research) Protein: M2-1 HRSV (12 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A					■	■		■	■	■	■	■
B	■	■	■	■	■	■				■	■	■
C		■		■	■	■	■	■	■	■	■	■
D	■	■										■
E	■					■	■			■	■	
F	■				■		■	■	■	■	■	■
G	■	■	■	■	■	■	■	■		■	■	■
H	■	■	■	■	■		■	■		■	■	■

Commercial Screen: Wizard 1&2 (Emerald BioSystems) Protein: M2-1 HRSV (12 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A		■	■						■		■	■
B	■			■	■	■			■	■	■	■
C	■		■		■					■		
D	■			■					■	■	■	
E			■	■	■	■		■		■		
F	■	■	■		■		■				■	
G	■		■	■					■	■		
H							■		■	■		■

Commercial Screen: Index (Hampton Research) Protein: M2-1 HRSV (12 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A					■	■		■	■	■	■	■
B	■	■	■	■	■	■				■	■	■
C		■		■	■	■	■	■	■	■	■	■
D	■	■										■
E	■					■	■			■	■	
F	■				■		■	■	■	■	■	■
G	■	■	■	■	■	■	■	■		■	■	■
H	■	■	■	■	■		■	■		■	■	■

Commercial Screen: Wizard 1&2 (Emerald BioSystems) Protein: M2-1 HRSV (12 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A		■	■						■		■	■
B	■			■	■	■			■	■	■	■
C	■		■		■					■		
D	■			■					■	■		
E			■	■	■	■		■		■		
F	■	■	■		■		■				■	
G	■		■	■					■	■		
H							■		■	■		■

Commercial Screen: Wizard 3 & 4 (Emerald BioSystems) Protein: M2-1 HRSV (12 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A			■		■				■	■		■
B		■		■	■		■		■		■	
C	■					■	■	■	■	■	■	
D		■	■	■	■		■		■			
E		■		■		■				■	■	■
F	■						■		■		■	
G	■			■	■	■			■	■	■	■
H		■			■						■	■

Commercial Screen: SaltRX™ (Hampton Research) Protein: M2-1 HRSV (12 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C												
D												
E												
F												
G												
H												

Commercial Screen: MIDAS™ (Molecular Dimensions) Protein: M2-1 HRSV (12 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A		■	■		■	■	■		■		■	
B	■	■	■			■	■	■				■
C	■				■				■		■	
D		■							■			
E	■			■	■			■	■	■	■	■
F	■	■		■	■		■	■	■	■	■	■
G	■				■	■	■	■		■		
H	■			■	■		■		■		■	

Commercial Screen: Morpheus™ (Molecular Dimensions) Protein: M2-1 HRSV (12 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A		■	■		■	■	■		■		■	
B	■	■	■			■	■	■				■
C	■				■				■		■	
D		■							■			
E	■			■	■			■	■	■	■	■
F	■	■		■	■		■	■	■	■	■	■
G	■				■	■	■	■		■		
H	■			■	■		■		■		■	

Commercial Screen: PACT *Premier™* (Molecular Dimensions) Protein: M2-1 HRSV (12 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A									█			
B	█	█					█	█		█	█	
C	█	█	█	█			█	█	█	█	█	
D	█	█	█				█	█	█	█	█	
E	█	█	█	█	█	█	█	█		█		█
F	█	█		█	█	█	█	█	█	█	█	█
G	█	█	█	█	█	█	█	█	█	█	█	█
H	█	█	█	█	█	█	█	█	█	█	█	█

Commercial Screen: Crystal 1& 2 Screen ™ (Hampton Research) Protein: M2-1 HRSV (12 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A		█	█				█	█			█	█
B		█		█		█			█		█	█
C	█		█		█				█	█		█
D				█		█	█	█				
E		█				█	█	█	█		█	
F			█	█		█	█	█				
G							█	█	█		█	
H			█	█	█			█	█			█

Commercial Screen: Index (Hampton Research) Protein: M2-1 HRSV + P 90-110 (8 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A			■	■		■	■	■	■	■	■	■
B	■	■	■	■	■	■	■		■	■	■	■
C		■			■	■	■	■			■	■
D		■										
E	■	■	■	■	■						■	
F												
G												
H												

Commercial Screen: Wizard 3 & 4 (Emerals BioSystems) Protein: M2-1 HRSV + P 90-110 (8 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A			■	■		■	■	■	■	■	■	■
B	■	■	■	■	■	■	■		■	■	■	■
C		■			■	■	■	■			■	■
D		■										
E	■	■	■	■	■						■	
F		■										
G												
H												

Commercial Screen: Morpheus™ (Molecular Dimensions) Protein: M2-1 HRSV + P 90-110 (8 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C												
D												
E												
F												
G										�		▒
H	▒					▒			▒			

Commercial Screen: MIDAS ™ (Molecular Dimensions) Protein: M2-1 HRSV + P 90-110 (8 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A					▒		▒	▒				
B								▒	▒	▒	▒	
C			▒	▒				▒	▒	▒		▒
D				▒	▒							
E	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒	▒
F	▒	▒		▒	▒			▒	▒	▒	▒	▒
G										▒	▒	
H	▒			▒	▒	▒						

Commercial Screen: Wizard 1 & 2 (Emerals BioSystems) Protein: M2-1 HRSV + P 90-110 (8 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A							■	■	■			
B					■	■				■	■	■
C				■	■		■		■			■
D			■	■						■		
E	■	■		■		■	■					
F	■											
G						■			■			
H	■			■		■	■					

Commercial Screen: PACT™ Premier Protein: M2-1 HRSV + P 90-110 (8 mg/ ml) Buffer: 25 mM Tris 7.5/ 150 mM NaCl and 1 mM DTT

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C						■	■	■	■	■	■	■
D				■	■	■	■	■		■	■	■
E								■	■	■	■	■
F					■							
G		■			■		■	■	■	■	■	■
H												

Appendix IV

Chapter 4 SBDD functional data:

- Mini genome assay- GFP fluorescent signal (expressed as a % of total GFP, in the absence of compound). Data values were taken 33 hours post infection in the presence of 20 μM compound. All values are expressed ± SEM n-2, to 2 significant figures.

- MTT assay- Signal at 570 nm as a % of viable BSRT7 cells in the absence of compound. All values are expressed ± SEM n-3, to 2 significant figures.

- Infectious virus assay- GFP fluorescent signal (expressed as a % of total GFP, in the absence of compound). All values are expressed ± SEM n-3 to 2 significant figures.

Vendor	Vendor ID	LDS ID	Mini genome assay (%)	MTT assay (%)	Infectious virus assay (%)	clogP	MW (Da)	Smiles
Asinex	AEM14637001	LDS-017332-1	6 ± 1.1	93 ± 2.3	51 ± 14	3.33	445.44	O=C(C1=CC(C2=CC=CO2)=NN1)N(CC3CCC4)CC3C4(O)C5=CC=CC(C(F)(F)F)=C5
ChemDiv	6873-1886	LDS-015832-1	1 ± 0.10	107 ± 4.7		4.89	296.39	CC1=C(C2=CC=CC=C2)N=C(NCC3=CC=CC=C3O)S1
ChemDiv	7399-0218	LDS-015883-1	7 ± 2.2	88 ± 2.0		3.89	333.2	O=C1C2(NC(C=CC=C3)=C3S2)C4=C

							C(Br)=CC=C4N1
AMRI	ALB-H09459938	LDS-010009-2	6 ± 4.5	92 ± 1.6	11 ± 2.4	4.94	404.49	O=C(CCC1=CC=CC=C1)N2CCN(CC2)C3=NC(C4=CC(F)=CC=C4)=NC(C)=C3
AMRI	ALB-H01555172	LDS-010041-2	8 ± 2.0	86 ± 0.44		1.63	487.53	O=C(N1C(C2=C(F)C=CC=C2)CC(C3=C(OC)C=C(OC)C=C3)=N1)CN(C(COC)=O)CCOC

AMRI	ALB-H01149618	LDS-010051-2	6 ± 0.16	79 ± 4.4		4.44	430.51	O=C(CCN(C1=NC(C2=CC=CC=C2)=NC3=CC=CC=C31)CCOC)NCC4=CC=CO4
AMRI	ALB-H11583697	LDS-010883-2	33 ± 5.9	89 ± 5.0		4.41	346.43	O=C(C1=NN(C2=CC(C)=CC=C2)C(CCC3=CC=CC=C3)=N1)NC4CC4
AMRI	ALB-H09272499	LDS-011258-2	18 ± 1.1	98 ± 0.71	61 ± 6.5	3.65	434.94	O=S(C1=CC=C(Cl)C=C1)(NC(C=C2CN(CC3CC3)C4=O)=CC=C2

								OC4CC)=O
ChemDiv	C226-4009	LDS-016034-1	3 ± 0.042	109 ± 7.0	23 ± 7.4	3.29	326.4	O=C(N1CCCCCC1)C2=NOC3=C2CCC4=C3C=CC(OC)=C4
ChemDiv	C714-0122	LDS-016090-1	5 ± 0.46	110 ± 12.9		3.64	394.41	O=C1N(CC(NC2=CC=CC(C(F)(F)F)=C2)=O)C(C=CC=C3)=C3SC(C)C1
ChemDiv	C737-1885	LDS-016044-1	33 ± 9.4	82 ± 13		4.41	423.53	O=S(C1=C2N=CC=CC2=CC=C1)(NC3(CCCCC3)C(NC(C=C4)=CC=C4C)=O)

								=O
ChemDiv	C768-0209	LDS-016079-1	N/A	82 ± 6.3			3.51	390.44	O=C(C(C=C1)=CC=C1C)C2=C(N3CCOCC3)C(C=C4OCCOC4=C5)=C5N=C2
Asinex	LMK19563975	LDS-018052-1	3 ± 0.30	60 ± 2.8			4	439.48	FC(F)(F)C1=CC(C2=CC=CS2)=NC(CCCC3)CN3C(C4=CSC(N)=N4)=O)=N1
AMRI	ALB-H09258955	LDS-010149-2	12 ± 7.5	90 ± 1.4	15 ± 6.2	4.93	468.48	O=C(CC1=CC=CC=C1)NC(=C2CN(CC3=CC(C(F)(F)F)=CC=C

							3)C4=O)=CC=C2OC4C
AMRI	ALB-H05321448	LDS-010902-2	1 ± 0.21	73 ± 1.6		4.01	461.44	O=C(OC)C1=C(C)N(CC2=CC=CC(C(OC)=O)=C2)C(CC1C3=CC(C(F)(F)F)=CC=C3)=O
ChemDiv	C743-0014	LDS-016067-1	6 ± 1.3	86 ± 4.1	54 ± 21	3.49	335.23	O=C(CC1=CC(Cl)=C(Cl)C=C1)N2CC3=CC=CC=C3NCC2
ChemDiv	C696-0163	LDS-	15 ± 2.7	84 ± 3.2	25 ± 9.6	2.93	370.41	C(C1=CC=C

		016077-1						C=C12)(CC3=CC=CN=C3)=NN=C2NCC4=CC=C5OCOC5=C4
Chembridge	9143187	LDS-026845-1	24 ± 1.8	77 ± 0.76		4.16	324.42	O=C(C1=CC(NC(C2=CC=C(C(C)C)C=C2)=O)=CC=C1)NC(C)C
AMRI	ALB-H10724892	LDS-029768-1	25 ± 9.6	85 ± 2.3		3.85	418.54	O=C(NC(C)COC)C1=CC(N2CCN(CC3=CC=CC=C3)CC2)=NC4=CC=CC=C41

AMRI	ALB-H01151177	LDS-013300-2	25 ± 9.6	93 ± 1.5		4.08	436.53	O=C(CCN(C1=NC(C2=CC=CS2)=NC3=CC=CC=C31)CC4=CC=CO4)NCCOC
AMRI	ALB-H04784528	LDS-015434-2	31 ± 0.22	97 ± 0.71		4.21	419.46	O=C(CCC(C)=N1)=C(N1C2=CC=CC=C2)OC3=CC(F)=CC=C3)NCC4=CC=CO4
AMRI	ALB-	LDS-	1 ± 0.30	84 ± 0.65		5.49	450.53	O=C(CCN(C1=NC(C2=C

	H01151501	013281-2						C=CS2)=NC3=CC=CC=C31)CCOC)NC4=CC=CC=C4F
AMRI	ALB-H05647691	LDS-015378-2	51 ± 22	95 ± 6.3		4.35	459.54	O=C(CC1=CC=CC(N(C(C2=CC=CS2)=O)CC3=CC=C(C=C3)F)=C1)NCC4=NC=CC=C4
Chembridge	7928320	LDS-020048-1	N/A	71 ± 9.4		3.91	387.44	O=C(C1=CC(NC(C(C=C2)=CC=C2C)=O)=CC=C1)NC(C=C3)=CC=C3NC(C)=O

ChemDiv	C700-1881	LDS-016101-1	54 ± 9.7	95 ± 0.34		2.61	407.49	O=C1N(CC(NCC(C=CC=C2)=C2OC)=O)C(C=CC=C3)=C3N4C=CC=C4SC1
ChemDiv	C700-1866	LDS-016089-1	73 ± 6.9	90 ± 3.4		2.53	328.34	O=C(NCC1=CC=C(F)C=C1)C2=CC=C(CC3=C(C)ON=C3C)O2
Asinex	AEM14662337	LDS-017289-1	13 ± 2.2	92 ± 12.5	59 ± 8.2	4.12	432.49	O=C(C1=CC=C(N(C)C)C=C1)N(CC2CCC3)CC2C3(O)C4=CC(C(F)(

								F)F)=CC=C4
ChemDiv	C239-0807	LDS-016046-1	35 ± 4.0	83 ± 4.8		4.12	321.4	CSC1=CC=C(C=C1)C2=C(NC3=CC=CC=C3)N4C(NC=N4)=N2
ChemDiv	C206-0335	LDS-016045-1	8 ± 0.10	83 ±3.1	67 ± 28	3.41	365.42	O=S(N1C2=CC(F)=CC=C2CCC1C)(C3=CC(OC)=CC=C3OC)=O
AMRI	ALB-H15023047	LDS-013192-2	69 ± 14	89 ± 2.4		3.15	493.51	O=C(CCC1=CC=CC=C1)N2CCC(NC(C3=C(NC(C4=CC(F)=C(F)C=C4)=O)N=CC=N3)=O)C

							C2
Asinex	AEM12409173	LDS-018011-1	41 ± 1.2	93 ± 4.3		1.98	407.52	O=C(N1CCN2C(NC(C)(C)C)=C(C3=CC=C(N(C)C)C=C3)N=C2C1)C4=CC=NN4
Chembridge		CB9073422	4 ± 2.1	98 ± 9.0		3.84	333.2	O=C(NC1=C2C=C(Br)C=C1)C32SC4=CC=CC=C4N3
Chembridge		CB7356502	2 ± 0.44	88 ± 2.1		4.89	296.39	CC1=C(C2=CC=CC=C2)

							N=C(NCC3=CC=CC=C3O)S1
Chembridge		CB9233087	14 ± 1.6	122 ± 1.3	68 ± 13	3.24	326.4	O=C(N1CCCCCC1)C2=NOC3=C2CCC4=CC(OC)=CC=C43
ChemDiv		V024-8658	38 ± 4.2	105 ± 9.6	69 ± 10	5.25	346.43	CC1=CC(N2C(CCC3=CC=CC=C3)=NC(C(NC4CC4)=O)=N2)=CC=C1
ChemDiv		V028-4618	11 ± 2.8	93 ± 4.6	26 ± 13	4.42	430.51	COCCN(CCC(NCC1=CC=CO1)=O)C2=NC(C3=CC

							=CC=C3)=NC4=C2C=CC=C4
ChemDiv		V028-4596	6 ± 2.3	97 ± 1.6		5.81	404.49	CC1=CC(N2CCN(C(CCC3=CC=CC=C3)=O)CC2)=NC(C4=CC(F)=CC=C4)=N1
ChemDiv		6873-1886	5 ± 0.80	107 ± 9.9		5.03	296.39	CC1=C(C2=CC=CC=C2)N=C(NCC3=CC=CC=C3O)S1
ChemDiv		V004-5260	36 ± 13	91 ± 6.2	59 ± 20	4.96	433.48	CC1=NN(C2=CC=CC=C2C)C(OC3=C

							C=CC(F)=C3)=C1CCC(NCC4=CC=CO4)=O
Asinex		LMK195639 75	7 ± 0.34	80 ± 3.0		3.23	439.48	O=C(C1=CSC(N)=N1)N(C2)CCC[C@@]2([H])C3=NC(C4=CC=CS4)=CC(C(F)(F)F)=N3
Chembridge		9083816	64 ± 5.6	89 ± 2.3		-0.66	314.36	CC(C(NC)=O)N(S(=O)(C)=O)C1=CC(OCCO2)=C2C=C1
Chembridge		9136178	67 ± 1.1	88 ± 5.4		1.69	330.4	O=C(C1=CC=C(NS(=O)(C

							)=O)C=C1)N2CCCC3=C2C=CC=C3
Chembridge		9136061	89 ± 13	88 ± 2.9		1.62	318.39	CC1=CC(CNC(C2=CC=C(NS(=O)(C)=O)C=C2)=O)=CC=C1
Chembridge		9143187	7 ± 2.7	78 ± 1.2		3.87	324.42	CC(C)C1=CC=C(C(NC2=CC=CC(NC(C)C)=O)=C2)=O)C=C1

Appendix V

Chapter 4 SBDD compound structures; see above table for functional data.

LDS ID	Structure
LDS-017332-1	
LDS-015832-1	
LDS-015883-1	
LDS-010009-2	
LDS-010041-2	
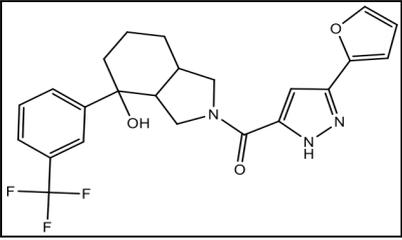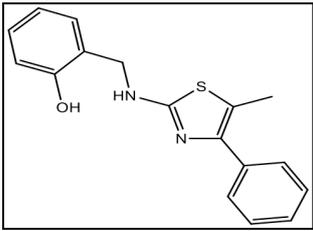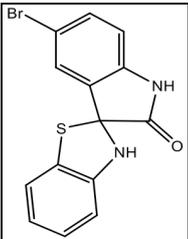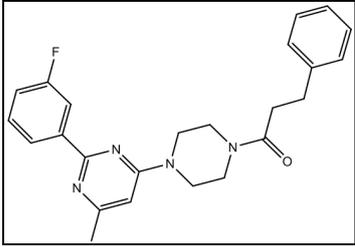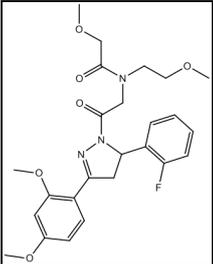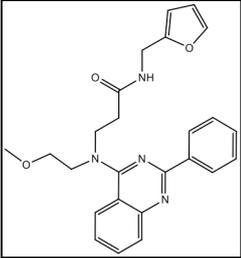
LDS-010051-2	
LDS-010883-2	
LDS-011258-2	
LDS-016034-1	
LDS-016090-1	
LDS-016044-1	
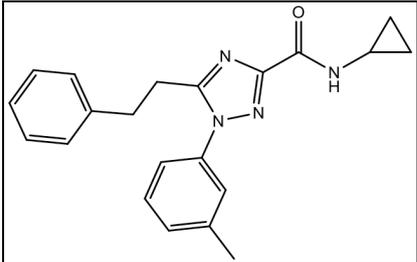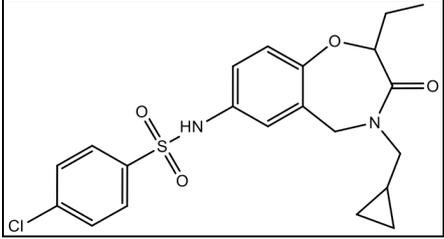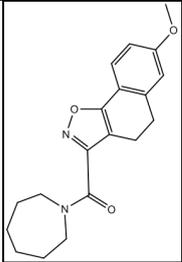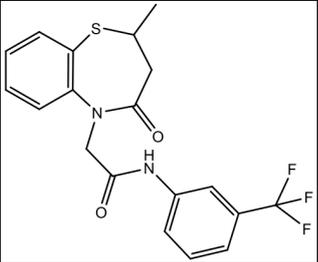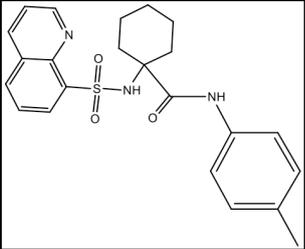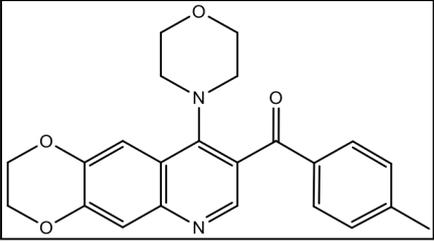
LDS-016079-1	
LDS-018052-1	
LDS-010149-2	
LDS-010902-2	
LDS-016067-1	
LDS-016077-1	
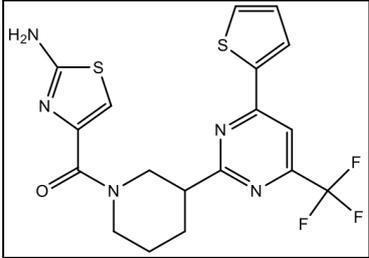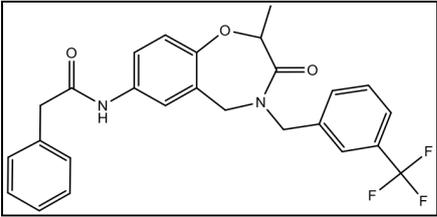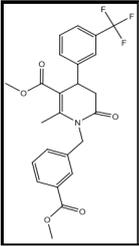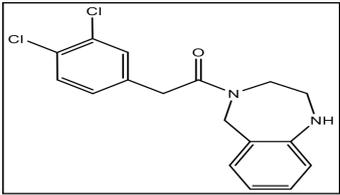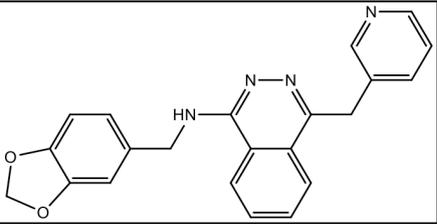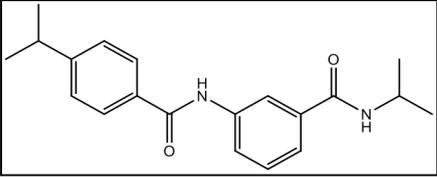
LDS-026845-1	
LDS-029768-1	
LDS-013300-2	
LDS-015434-2	
LDS-013281-2	
LDS-015378-2	
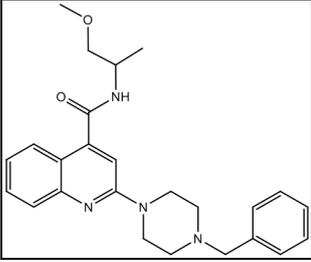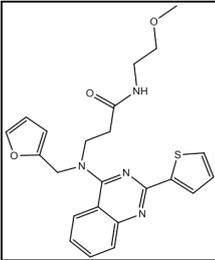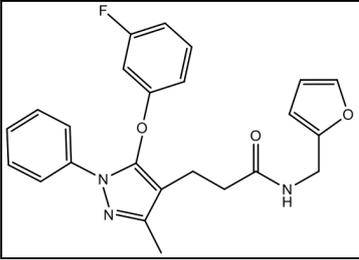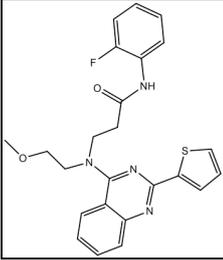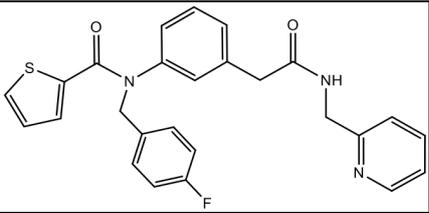
LDS-020048-1	
LDS-016101-1	
LDS-016089-1	
LDS-017289-1	
LDS-016046-1	
LDS-016045-1	

LDS-013192-2	
LDS-018011-1	
CB9073422	
CB7356502	
CB9233087	
V024-8658	

V028-4618	
V028-4596	
6873-1886	
V004-5260	
LMK19563975	
9083816	

9136178	
9136061	
9143187	

Appendix VI

Chapter 5- FBDD hits using the XChem facility to determine the structure of fragments bound to HRSV M2-1 by X-ray crystallography.

Fragment	Xchem ID	Vendor	Vendor ID	SMILES	MW
M2-1-x0079	N13517a	Asinex	BAS05277195	CC(=O)Nc1nnc(s1)COC	187.04
M2-1-x0108	N13426a	Key Organics Building Blocks	8N-020	c1cnc(nc1)Oc2ccc(cc2)c3cc[nH]n3	238.09
M2-1-x0373	N13441a	Maybridge	DFP00321	COC(=O)c1cccc(c1)NS(=O)(=O)C	229.04
M2-1-x0770	N13491a	Specs	AN-465/42246748	c1cc(sc1)C[NH2+]Cc2ccncc2	205.08
M2-1-x0816	N13698a	ChemBridge	7962225	COC(=O)c1ccc(cc1)S(=O)(=O)N	215.23
M2-1-x0869	N14177a	Edelris	FMOOA000509a	Cn1c2[C@H]3C[C@H](Cc2c4ccccc14)NC[C@H]3O	242.32
M2-1-x0937	N14247a	Edelris	FMOOA000579a	O=C1CCN2[C@@H]3[C@H](COc4ccccc34)CN12	230.26