Structural studies of the human Metapneumovirus (HMPV) nucleocapsid and the associated polymerase cofactors M2-1 and P



Max Renner

Medical Sciences Doctoral Training Centre
Linacre College
University of Oxford

A thesis submitted in fulfilment of the requirements for the degree of *Doctor of Philosophy* at the University of Oxford

Trinity Term 2016

Structural studies of the human Metapneumovirus (HMPV) nucleocapsid and the associated polymerase cofactors M2-1 and P

Max Renner, Linacre College, University of Oxford
DPhil, Medical Sciences Doctoral Training Centre
Trinity Term 2016

Abstract

Non-segmented, negative-strand RNA viruses (*Mononegavirales*) cause serious human disease and include major pathogens such as respiratory syncytial virus (RSV), measles, or Ebola virus. These infectious agents share replication and transcription strategies, and possess a similar genome organization. Human metapneumovirus (HMPV, belonging to the *Pneumovirinae* subfamily of the *Paramyxoviridae*), a recently described member of this order of viruses is a major cause of respiratory tract illness in children worldwide. As in other mononegaviruses, the replication/transcription machinery of HMPV consists of a set of conserved protein factors.

The genome serves as the template for the synthesis of viral RNA by the L polymerase. It is never found naked in the cytoplasm of infected cells, but is instead encapsidated in a sheath of nucleoproteins N, forming the nucleocapsid. The L polymerase associates with the nucleocapsid template via the assistance of a cofactor, termed P protein. Lastly, in HMPV, the antitermination protein M2-1 is required for the processive transcription of viral mRNAs. The spatially and temporally regulated interplay between nucleocapsid, L polymerase, P protein, and M2-1 lies at the heart of the controlled synthesis of progeny genomes and viral mRNAs in HMPV and related viruses. The structural elucidation of how these components act in concert could spur the development of novel antiviral strategies.

In this *DPhil* project I have carried out biophysical investigations of the HMPV replication/transcription components N, P, and M2-1. Here, I present the crystallographic structures of HMPV N in its oligomeric RNA-bound form (N-RNA) and in a naive RNA-free state, bound to the polymerase cofactor P (N^0-P). Direct comparison of these structures sheds light on the conformational transition between the N-RNA and N^0-P states and illustrates how P can inhibit the premature self-association of N in *Pneumovirinae*. Furthermore, in this work I describe structures of the HMPV antitermination protein M2-1 bound to nucleotides. In conjunction with small-angle X-ray scattering analyses, these data provide the basis of how M2-1 may recognise nascent viral mRNAs, an essential feature of this highly dynamic processivity factor. Finally, the structure of a central region of HMPV P reveals how P tetramerises via a short Leucine-rich coiled-coil domain. Due to their relatedness, the results provided here are broadly relevant for other pneumoviruses, like the major pathogen RSV.

ACKNOWLEDGEMENTS

I would like to thank Jon and Juha for being the best PIs anyone could wish for. It also was a mind-altering experience to work with my friend Cedric who, most of all, taught me how to be a happy scientist and who is always able to see a silver lining. Equally, I thank Guido and Mattia who patiently listened to my (pseudo-)philosophical (and likely nonsensical) ramblings and, puzzlingly, were still willing to do experiments with me.

Strubi has become a bit like an extended family to me and will be difficult to leave, eventually. It is hard to imagine another place with so many nice and like-minded people whom I am happy to call friends. I would like to especially mention Christina, Nayab, Rhys, Liz, Alice, Joy, Dimple, Simonas, Abhay, Helen, Steinar, Pascale and, last but not least, Serban. We're all still waiting for that guided tour of Romania! I hope I can keep in touch with you all, even while vigilantly boycotting social media. Cheers to David B for identifying a specific kit on the way to Hellfest. I would like to thank Ilona who, I'm convinced, must be a long-lost relative of mine. And also Luigi, who not only showed me what I was doing wrong in crystallography, but also how to do it right (after having a bit of a laugh first). I would also like to acknowledge Jona, who sparked my interest in high-pressure NMR. Finally, thanks to Sam for having a (semi-)good taste in music and Tony who was spontaneously willing to store all of our worldly possessions in his flat when our landlord had a sudden change of heart.

My experiments would not have been possible without the fantastic support provided by Margaret, Ollie, Karl, and Tom who kept everything running like clockwork. Kamel and Tao were invaluable when Sina and I were trying to solve her structure with S-SAD. Many thanks also to the admin team, especially Mel and Agata for helping me with my proposal and showing real interest in what I am doing.

Thanks go to my family, Carol, PeRe, Paul, Susie and FuFu, who always had my back, no matter what, and demonstrate that a conventional upbringing is totally overrated. Thanks, Sina, for putting up with me for so many years and tolerating watching sheer endless amounts of cartoon series with me.

DECLARATION OF WORK

All the described work herein is my own, except that detailed immediately below:

Chapter 2 (HMPV N Protein):

- The chapter is based (including figures) on an article published in eLife which was prepared by Jonathan Grimes and myself (Renner et al. 2016). Some sections of the chapter are quoted verbatim.
- In the course of this project I supervised Master's student Mattia Bertinelli who provided valuable assistance with purifications of HMPV N^0-P. Mattia also helped me with some of the crystallographic screening and accompanied me to the small-angle x-ray scattering beamline for data collection and analysis.
- Jonathan Grimes, Karl Harlos, Guido Paesen and myself were all involved in mounting and cryo-cooling of the numerous crystals of HMPV N-RNA or N^0-P screened during the project.
- Juha Huiskonen assisted with plunge-freezing of grids and prepared the FEI Tecnai Polara for cryo-EM single-particle data collection of N-RNA rings.
- Guido Paesen produced the recombinant baculovirus stock for the expression of N-RNA in insect cells

Chapter 3 (HMPV M2-1):

- The data presented in this chapter (including figures) is based on an article (Leyrat et al. 2014b) prepared by Jonathan Grimes, Cedric Leyrat, and myself, and parts of it are quoted verbatim.

- Data collection for SAXS was jointly carried out by Cedric Leyrat and myself. SAXS data were analysed by Cedric Leyrat.
- Crystals were mounted and cryo-cooled by Karl Harlos.
- Collection of crystal diffraction data and structure refinement were mainly performed by me, under the supervision of Cedric Leyrat.

Chapter 4 (HMPV P Protein):

- The results described in this chapter (including figures) are based on an article (Leyrat et al. 2013) prepared by Jonathan Grimes, Cedric Leyrat, and myself. Sections of it are quoted verbatim.
- SAXS data were collected jointly by Cedric Leyrat and myself.
- Crystals were mounted and cryo-cooled by Karl Harlos. Diffraction data were collected jointly by Karl Harlos, Cedric Leyrat and myself.
- The first crystal of the HMPV P tetramerisation domain was phased and refined by Cedric Leyrat, the second by myself.

Chapter 5 (Concluding Remarks)

- Parts of this chapter draw upon ideas put forth in the Discussions of the three corresponding publications noted above which have been co-authored by myself (Renner et al. 2016; Leyrat et al. 2014b, 2013).

TABLE OF CONTENTS

LIST OF FIGURES.....VII

LIST OF TABLES.....IX

LIST OF NON-STANDARD ABBREVIATIONS.....X

1 INTRODUCTION.....1

1.1 The *Mononegavirales* – A Brief Overview.....2

1.2 Human Metapneumovirus.....7

1.2.1 Epidemiology and Clinical Features.....8

1.2.2 Structure and Components of the Virion.....10

1.2.3 Genome Organization and Gene Expression.....17

1.2.4 Lifecycle and Replication.....19

1.3 The Transcriptase/Replicase Machinery.....25

1.3.1 The Nucleoprotein N.....25

1.3.2 The P Protein.....30

1.3.3 The L Polymerase.....36

1.3.4 The M2-1 Transcription Antiterminator.....41

1.4 Project Aim.....44

2 HMPV N.....45

2.1 RNA-Free N^0-P Complex.....45

2.1.1 Construct Design Rationale.....45

2.1.2 Expression and Purification of the N^0-P Chimera.....48

2.1.3 Crystallisation, Structure Solution and Refinement.....50

2.1.4 Structure of the HMPV N^0-P Complex.....54

2.2 Assembled N-RNA.....60

2.2.1 Initial Characterization of N-RNA from Different Expression Hosts.....60

2.2.2 Crystallisation, Structure Solution and Refinement.....65

2.2.3 Structure of HMPV N-RNA.....69

2.3 Comparison of N^0-P and N-RNA Crystal Structures.....72

2.4 MD Simulations of N^0-P.....79

2.5 Solution Characterization by SAXS.....82

2.6 Discussion.....86

3 HMPV M2-1.....93

3.1 Background.....93

3.2 Purification of Recombinant Sample.....97

3.3 M2-1 in Complex with Nucleotides....98
3.3.1 Crystal Production....98
3.3.2 Nucleotide Soaking Experiments and Structure Refinement....100
3.3.3 Comparison of the Apo-Forms....106
3.3.4 Nucleotide-Bound Structures....108
3.4 Analysis of RNA-binding in Solution....114
3.5 Discussion....118
4 HMPV P....123
4.1 Background....123
4.2 Revisiting HMPV P Domain Organization....123
4.3 The M2-1 · P Complex....127
4.3.1 Reconstitution of the Complex....127
4.3.2 SAXS Characterization of M2-1 · $P_{135-294}$....128
4.3.3 Crystallisation Trials....131
4.4 P Tetramerisation Domain....131
4.4.1 Purification of $P_{(158-237)}$....131
4.4.2 SAXS Characterization of $P_{(158-237)}$....133
4.4.3 Crystallisation, Structure Solution, and Refinement....134
4.4.4 Structure of the HMPV P Tetramerisation Domain....138
4.5 Discussion....141
5 Concluding Remarks and Future Work....145
5.1 The Numerous Interactions of Pneumovirus P....145
5.2 Similarities of Filo- and Pneumoviruses....151
5.3 Future Directions in Nucleocapsid Research....156
6 Materials and Methods....158
6.1 Molecular Biology and Microbiology Methods....158
6.1.1 Preparation of Solutions and Media....158
6.1.2 Cultivation and Storage of E. coli....158
6.1.3 Polymerase Chain Reaction....159
6.1.4 Agarose Gel Electrophoresis....159
6.1.5 In-Fusion Ligation Independent Cloning....160
6.1.6 Transformation of Chemically-Competent E. coli....161
6.1.7 Preparation of Plasmid-DNA from E. coli....161
6.2 Protein Production....162
6.2.1 Recombinant Expression in E. coli....162
6.2.2 Purification of Recombinant Proteins from E. coli....162
6.2.3 Production and Purification of N-RNA Filaments in Insect Cells....165

6.2.4 SDS-PAGE.....*166*
6.3 Structural Biology Methods.....167
6.3.1 Crystallisation.....*167*
6.3.2 Cryo-Protection.....*167*
6.3.3 Crystal Ligand Soaking.....*168*
6.3.4 Data Collection and Processing.....*168*
6.3.5 Small-Angle X-ray Scattering (SAXS).....*168*
6.3.6 Electron Microscopy.....*169*
6.3.7 Molecular Dynamics Simulations.....*170*
6.4 Amino Acid Sequences of Studied Proteins.....172
7 REFERENCES.....176
8 APPENDIX.....201
8.1 List of Publications.....201

LIST OF FIGURES

Figure 1.1: Phylogenetic tree of *Mononegavirales*....3
Figure 1.2: Schematic of an HMPV virion....11
Figure 1.3: Pre- and Post-Fusion F on Intact RSV Virions....12
Figure 1.4: Quaternary Structures of Matrix Proteins....14
Figure 1.5: Tomography of RSV particles....16
Figure 1.6: Genome Organisation of *Pneumovirinae*....18
Figure 1.7: Lifecyle of HMPV....20
Figure 1.8: Model of Transcription and Replication in RSV....23
Figure 1.9: Architecture of Mononegavirus Nucleoproteins....27
Figure 1.10: RNA-Binding Groove of RSV N....28
Figure 1.11: Recombinant Nucleocapsids of RSV....29
Figure 1.12: Architecture of *Paramyxovirinae* P....30
Figure 1.13: P Oligomerisation Domains....31
Figure 1.14: Interaction between N and P in *Paramyxovirinae*....32
Figure 1.15: Structural Phylogeny of P....34
Figure 1.16: VSV N bound to P and RNA....36
Figure 1.17: Structure of VSV-L by Cryo-EM Single Particle Reconstruction....40
Figure 1.18: The Antitermination Factor M2-1 from HMPV....43
Figure 2.1: Disorder and Domain Analysis of HMPV N and P....47
Figure 2.2: Purification of HMPV N^0-P....49
Figure 2.3: Crystal of HMPV N^0-P....51
Figure 2.4: SDS-PAGE Analysis of Trypsin-Digested N^0-P....54
Figure 2.5: The RNA-free N^0-P complex from HMPV....55
Figure 2.6: Mononegavirus Nucleoprotein-Phosphoprotein Complexes....59
Figure 2.7: Recombinant Nucleocapsid-Like Filaments in Low-Salt Buffer....61
Figure 2.8: Recombinant Nucleocapsid-Like Filaments in High-Salt Buffer....62
Figure 2.9: HMPV N Recombinantly Expressed in Bacteria....64
Figure 2.10: Crystal of HMPV N-RNA....65
Figure 2.11: HMPV N-RNA Electron Density Maps and Model Validation....69
Figure 2.12: HMPV N-RNA....71
Figure 2.13: Comparison of HMPV N-RNA and N^0-P....73
Figure 2.14: A Conserved Aromatic Residue Within the Hinge....75
Figure 2.15: Conformational Changes of the CTD-Arm....78
Figure 2.16: Molecular Dynamics Simulation of N^0-P....81

Figure 2.17: Characterization of the N^0-P Chimera in Solution....83
Figure 2.18: Solution Comparison of N-RNA and N^0-P via SAXS....86
Figure 2.19: Examples for Disorder Metapredictions of P Proteins....88
Figure 2.20: Hypothetical Model of Nucleocapsid-Filament Growth....92
Figure 3.1: Comparison of RSV and HMPV M2-1 Proteins....95
Figure 3.2: Purification of HMPV M2-1....97
Figure 3.3: Section of the Crystallographic Packing of HMPV M2-1....98
Figure 3.4: Representative Examples of M2-1 Crystals....99
Figure 3.5: Sample Difference Map Density for Soaking Experiments....101
Figure 3.6: Overview of AMP-Soaked M2-1....103
Figure 3.7: Electron Density of the Ligands in the Refined Structures....104
Figure 3.8: Overview of AGTTA DNA-Soaked M2-1....105
Figure 3.9: Interdomain motions within M2-1 Crystals....107
Figure 3.10: Structures of M2-1 bound to Nucleotides at the Zn-finger....110
Figure 3.11: Comparison of HMPV M2-1 and Nab2 Zn-fingers....111
Figure 3.12: Core-domain Interactions with Crystallographically Related M2-1 Copies....113
Figure 3.13: Solution Characterization of RNA-binding by SAXS....115
Figure 3.14: Model of Gene-End Recognition by M2-1....120
Figure 3.15: Schematic Model of the Pneumoviral Transcriptase....122
Figure 4.1: Construct-Design to Study P-Oligomerisation and M2-1 Recruitment....125
Figure 4.2: Purification of M2-1·$P_{(135-294)}$....128
Figure 4.3: SAXS Analysis of the M2-1·$P_{(135-294)}$ Complex....130
Figure 4.4: Purification of $P_{(158-237)}$....132
Figure 4.5: SAXS Characterization of the Central Domain of HMPV P....133
Figure 4.6: Crystal of P_{core-1}....134
Figure 4.7: Crystal of P_{core-2}....135
Figure 4.8: Structure of P_{core-1}....139
Figure 4.9: Structure of P_{core-2}....141
Figure 4.10: Phylogenetic Tree of P Multimerisation Domains....144
Figure 5.1: Schematic models of Pneumoviral RNA-Synthesis Machinery....146
Figure 5.2: MSA of P Proteins from *Pneumovirinae* Family Members....147
Figure 5.3: Schematic of HMPV P Domain Organization....148
Figure 5.4: Similarity Between VP30 and M2-1....151
Figure 5.5: Nucleoprotein-Based Phylogenetic Tree of *Mononegavirales*....155

LIST OF TABLES

Table 1: Data Collection and Refinement Statistics for HMPV N^0-P.....53
Table 2: Sequence Identities of *Pneumovirinae* N Proteins.....58
Table 3: Data Collection and Refinement Statistics for HMPV N-RNA.....68
Table 4: Isoelectric points of N protein CTD-arms in *Pneumovirinae*.....90
Table 5: Data Collection and Refinement Statistics for HMPV M2-1.....106
Table 6: A-Content of HMPV mRNAs (Strain NL1-00, A1).....118
Table 7: Data Collection and Refinement Statistics for HMPV P_{core}.....137
Table 8: Antibiotics used in this work.....158
Table 9: Representative PCR Protocol.....159

LIST OF NON-STANDARD ABBREVIATIONS

Å	Ångström (10^{-10} m)
AMP	adenosine monophosphate
AMPV	avian metapneumovirus
ASU	asymmetric unit
BDV	Borna disease virus
CC	cross-correlation
CCCH zinc finger	$(Cys)_3(His)_1$ zinc finger
CTD	C-terminal domain
DLS	Diamond Light Source
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
ER	endoplasmatic reticulum
ET	electron tomography
F protein	(viral) fusion protein
G protein	(viral) attachment glycoprotein
Gdn-HCl	guanidinium hydrochloride
GTase	guanylyltransferase
HMPV	human metapneumovirus
HPIV	human parainfluenza virus
ICTV	International Committee on Taxonomy of Viruses
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl β-D-1-thiogalactopyranoside
kDa	kilo Dalton
L protein	large viral polymerase protein
LACV	La Crosse orthobunyavirus
LB	Luria-broth or lysogeny-broth
LSSR	Local Structure Similarity Restraints
M protein	(viral) matrix protein
M2-1	(viral) matrix protein 2-1
MBP	maltose binding protein
MD	molecular dynamics
MoRE	molecular recognition element
MPD	methyl-2 4-pentanediol
MR	molecular replacement
mRNA	messenger ribonucleic acid
MSA	multiple sequence alignment
N protein	(viral) nucleoprotein

NCS	non-crystallographic symmetry
NDSB	non-detergent sulphobetaine
NDV	Newcastle disease virus
nt	nucleotides
NTA	nitrilotriacetic acid
NTD	N-terminal domain
OD_{600}	optical density at 600 nm
ORF	open reading frame
P protein	(viral) phosphoprotein
PAGE	polyacrylamide gel-electrophoresis
P_{ced}	P central domain
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEG-MME	polyethylene glycol monomethyl ether
pI	isoelectric point
PME	particle-mesh Ewald
P_{NBD}	P nucleocapsid binding domain
P_{OD}	P oligomerisation domain
PRNTase	polyribonucleotide transferase
RdRP	RNA-dependent RNA polymerase
RIGI	retinoic acid-inducible gene I
RMSD	root mean square deviation
RMSF	root mean square fluctuation
RNA	ribonucleic acid
rpm	revolutions per minute
RSV	respiratory syncytial virus
rVSV	recombinant vesicular stomatitis virus
SAXS	small angle x-ray scattering
SDS	sodium dodecyl sulphate
SEC	size-exclusion chromatography
SLAM	signalling lymphocytic activation molecule
SOC	Super Optimal broth with catabolite repression
TB	terrific broth
TAE	Tris-acetate-EDTA
TE	Tris-EDTA
TEM	transmission electron microscopy
TM	transmembrane
VSV	vesicular stomatitis virus

1 INTRODUCTION

The work presented herein is concerned with the biophysical and structural characterization of a recently discovered pathogen, human metapneumovirus (HMPV). Infection with HMPV is often asymptomatic or, in most cases when healthy adults are concerned, causes only mild respiratory symptoms. However, in individuals with pre-existing underlying conditions, the immunocompromised (e.g. transplant patients), or the elderly, HMPV may cause life-threatening pneumonia. HMPV belongs to a biomedically important family of viruses, the *Paramyxviridae,* which includes other major human pathogens such as measles and mumps virus. The paramyxoviruses are part of a wider phylogenetic group termed *Mononegavirales*, the members of which share a set of common features. The viral machinery involved in storing, expressing and replicating the genetic material of HMPV is the focus of this work and I have utilised a biophysical and structural approach to characterise a number of the involved components. In the following sections I will introduce key aspects of the biology of HMPV and will place the virus into the context of the viral order *Mononegavirales*. I will address which viral factors are necessary for transcription and replication and what is already known about them. Although my focus will be on the biology of HMPV, many characteristics are generally true for all mononegaviruses, while others differ. I will thus contrast and compare, wherever appropriate, and explain how HMPV fits into the scheme of the viral order. Finally, I will describe the aims of this project.

1.1 The *Mononegavirales* – A Brief Overview

The *Mononegavirales* are a large and diverse order of viruses which include both plant and animal pathogens with an enormous impact on global human well-being (Acheson 2006; Knipe and Howley 2007). In all mononegaviruses the genome is a non-segmented, negative-sense RNA molecule, i.e. it cannot be directly translated by the host-cell machinery. The genome is encapsidated in the form of a helical ribonucleoprotein complex termed nucleocapsid in which the RNA is sequestered in a sheath of oligomerised viral nucleoproteins (Ruigrok et al. 2011a). The viral particle, which contains the nucleocapsid, possesses a lipid envelope stemming from the host cell plasma membrane. The order is subdivided into 5 families[1]: the*, Rhabdoviridae, Filoviridae, Bornaviridae, Nyamiviridae* and *Paramyxoviridae* (**Fig. 1.1**).

The *Rhabdoviridae* family contains viruses that can infect a variety of hosts, including plants, arthropods, marine vertebrates and terrestrial vertebrates (Mann and Dietzgen 2014). Rhabdoviruses are named for their characteristic rod-shaped morphology (from Greek *rhabdos* rod). Plant-specific rhabdoviruses are often carried by arthropods and penetrate the plant host cells via insect bites. From a human health perspective, the most noteworthy member is the Rabies virus which infects carnivores and can be transmitted to humans through the bite of infected animals. In humans rabies causes illness with neurological symptoms which is, if left untreated via post-exposure vaccination, nearly always lethal (Modrow et al. 2010). Vesicular stomatitis virus (VSV), also a member of the Rhabdoviridae, is a valuable model organism and

1 The mononegavirus order was very recently emended by the addition of the *Mymonaviridae* and *Sunviridae* families through the International Committee on Taxonomy of Viruses (ICTV) (Afonso et al. 2016). Furthermore, the *Pneumovirinae* subfamily was elevated to the status of family (*Pneumoviridae*). This is further discussed in chapter 5 (Concluding Remarks and Future Work).

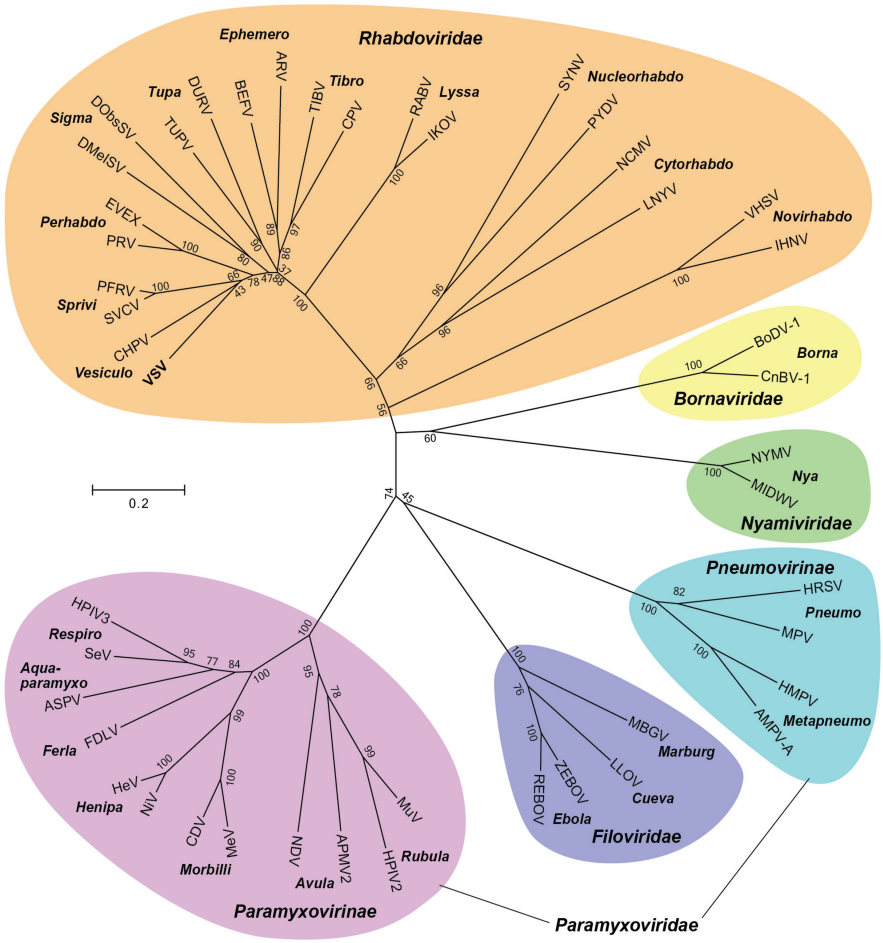


Figure 1.1: Phylogenetic tree of *Mononegavirales*.
The tree was generated using MEGA 6.0 (Tamura et al. 2013). Viral families and subfamilies are labelled in bold. Figure adapted from Neubauer et al. 2016. Viruses are abbreviated as follows: *AMPV-A*, avian metapneumovirus-A, *APMV2*, avian paramyxovirus 2, *ARV*, Adelaide River virus, *ASPV*, Atlantic salmon paramyxovirus, *BEFV*, Bovine ephemeral fever virus, *BoDV-1*, Borna disease virus 1, *CDV*, canine distemper virus, *CHPV*, Chandipura virus, *CnBV-1*, canary bornavirus 1, *CPV*, Coastal Plains virus, *DmelSV*, Drosophila melanogaster sigma virus, *DObsSV*, Drosophila obscura sigma virus, *DURV*, Durham virus, *EVEX*, eel virus European X, *FDLV*, Fer-de-Lance virus, *HeV*, Hendra virus, *HMPV*, human metapneumovirus, *HPIV2*, human parainfluenzavirus 2, *HPIV3*, human parainfluenzavirus 3, *HRSV*, human respiratory syncytial virus, *IHNV*, infectious haematopoietic necrosis virus, *IKOV*, Ikoma lyssavirus, *LLOV*, Lloviu virus, *LNYV*, lettuce necrotic yellows virus, *MBGV*, Marburgvirus, *MeV*, measles virus, *MIDWV*, Midway virus, *MPV*, murine pneumonia virus, *MuV*, mumps virus, *NCMV*, northern cereal mosaic virus, *NDV*, Newcastle disease virus, *NiV*, Nipah virus, *NYMV*, Nyamanini virus, *PFRV*, pike fry rhabdovirus, *PRV*, perch rhabdovirus, *PYDV*, potato yellow dwarf virus, *RABV*, rabies virus, *REBOV*, Reston ebolavirus, *SeV*, Sendai virus, *SVCV*, spring viremia of carp virus, *SYNV*, Sonchus yellow net virus, *TIBV*, Tibrogargan virus, *TUPV*, Tupaia rhabdovirus, *VHSV*, viral hemorrhagic septicemia virus, *VSV*, vesicular stomatitis virus, *ZEBOV*, Zaire ebolavirus

has been thoroughly studied, revealing many mechanistic details of genome replication and transcription in rhabdoviruses and non-segmented, negative-sense RNA viruses in general (Liang et al. 2015). VSV has also been proposed as a potential therapeutic tool as an oncolytic virus (Buijs et al. 2015) and as a vaccine vector (Geisbert and Feldmann 2011). The use of viruses as oncolytic therapeutics exploits the property of some viruses to more efficiently destroy cancer cells than healthy cells (Hastie and Grdzelishvili 2012). In the case of VSV, this increased susceptibility has been attributed to defects in the antiviral type I interferon pathway, which are very common in cancer cells (Hastie and Grdzelishvili 2012). VSV has further advantages, such as its well-studied biology (allowing engineering to improve oncolytic properties), no pre-existing immunity in humans, and no risk of genomic integration (Buijs et al. 2015). For vaccination efforts, recombinant, G-glycoprotein-deleted VSV (rVSVΔG) is used as a delivery backbone for other viral antigens (Martins et al. 2016; Geisbert and Feldmann 2011). A specifically successful example would be the development of a rVSV-based Ebola virus vaccine in the course of the recent 2013 Ebola virus outbreak in Western Africa, in which the VSV G-glycoprotein was replaced by Ebola virus $GP_{1,2}$ (Martins et al. 2016).

The *Filoviridae* include deadly zoonotic viruses that cause plasma leakage and haemorrhagic fevers (Acheson 2006). The name filovirus is derived from the often filamentous or thread-like morphology observed by electron microscopy (EM). Transmission from animal reservoirs (bats in the case of Ebola virus) can lead to sudden and sporadic outbreaks with potentially devastating consequences as observed in the West African Ebola outbreak from 2014 to 2016 (Henao-Restrepo et al. 2016). Lethality of infections with Marburg or Ebola virus can range from 30% to 90%

(Acheson 2006). The West African epidemic is thought to have caused more than 28 000 cases and 11 000 deaths, although these numbers may be in fact underestimated (Moll et al. 2016).

The *Bornaviridae* and *Nyamiviridae* differ from most other mononegaviruses as they replicate in the nucleus of their host, rather than in the cytoplasm (Herrel et al. 2012; Briese et al. 1992). A further example of nuclear mononegaviruses are members of the *Nucleorhabdovirus* genus from the Rhabdoviridae (**Fig. 1.1**). Borna disease virus (BDV) is a neurotropic agent which possesses a remarkably wide host range, infecting birds and a variety of mammals. BDV can cause persistent infections of neurons and has been linked to neurological disease and behavioural changes in horses and sheep (Dürrwald and Ludwig 1997). In humans, infection with BDV has been associated with the pathogenesis of neuropsychiatric diseases. Specifically, high BDV marker prevalence has been observed in psychiatric patients (reviewed in Kinnunen et al. 2013). However, this link is controversial and a conclusive causal relationship has yet to be established (Kinnunen et al. 2013). Strikingly, the nucleoprotein N gene of Borna virus is endogenised (integrated) into a wide range of eukaryotic genomes, including humans, an occurrence that was thought to be restricted mainly to retroviral sequences (Belyi et al. 2010; Horie et al. 2010). Endogenization events may be driven by high concentrations of viral RNA in the nucleus and facilitated by the host transposition machinery (Feschotte 2010). Remarkably, chromosomally integrated N copies have been suggested to be transcribed into mRNAs in human tissues, perhaps serving an antiviral purpose by offering protection against exogenous infection (Horie et al. 2010). Furthermore,

experimental infections with BDV have resulted in integration events of N DNA (Horie et al. 2010; Kinnunen et al. 2011).

The *Nyamiviridae* are a novel and recently proposed viral family that have been associated with infections of ticks and birds (Mihindukulasuriya et al. 2009). Although no infections of humans with nyamiviruses have been reported, the family member Nyamanini virus has been shown to productively infect human cells in tissue culture (Herrel et al. 2012). The Nyamiviridae remain poorly characterised to date with no structural and little functional data available.

The *Paramyxoviridae* family is composed of two subfamilies, the *Paramyxovirinae* and *Pneumovirinae*. The *Paramyxovirinae* subfamily includes major human pathogens like measles and mumps virus. Measles virus is closely related to rinderpest virus and is thought to have zoonotically spilled over to humans via infected cattle. Measles is a highly contagious airborne disease which causes serious illness in humans. Indeed, when the virus was introduced to the Americas by European conquerors it had devastating consequences for native populations which lacked herd-immunity (Acheson 2006). Today, a very successful and long-lasting vaccine is available which has ultimately led to a significant reduction of mortality from 562,400 in the year 2000 to 114,900 in 2014 (Holzmann et al. 2016). Other notable members of the subfamily are mumps virus, human parainfluenza virus 2 (HPIV2), and Sendai virus which is an important model organism. The *Paramyxovirinae* also include the zoonotic viruses Hendra and Nipah virus which have high mortality rates and circulate in fruit bat reservoirs in Asia and Australia. The *Pneumovirinae* subfamily harbours the human pathogens respiratory syncytial virus (RSV) and HMPV. Both viruses cause seasonal outbreaks of lower respiratory

tract illness. The impact of RSV is especially severe with nearly 100% of children being infected in the first two years of life and high infant mortality, especially in developing countries (Saso and Kampmann 2016). HMPV was only recently discovered in 2001 and causes symptoms similar to those of RSV (van den Hoogen et al. 2001, 2002). It constitutes an important human pathogen and there is a great paucity of information concerning the molecular biology of HMPV.

1.2 Human Metapneumovirus

Acute respiratory tract infections are a leading cause of mortality in children under five years of age, with ~1.5 million deaths reported in the year 2008 (Black et al. 2010). HMPV is an important cause of viral respiratory illness worldwide, being a major factor in pneumonia of young children (Schildgen et al. 2011). The virus was initially characterised in the Netherlands in 2001, but consecutive analysis of serum samples led to the detection of HMPV in samples going back as far as 1958 (van den Hoogen et al. 2001). HMPV may have remained undetected until recently because, contrary to most respiratory viruses, the *in vitro* propagation was found to be trypsin-dependent and the replication kinetics are slow (van den Hoogen et al. 2001). Furthermore, standard serological assays lacked cross-reactivity with HMPV. Sequence analyses were used to estimate that HMPV evolved from an avian-metapneumovirus (AMPV)-like ancestor around 200 years ago (de Graaf et al. 2008b). It has been suggested that a cross-species transmission event from birds to humans occurred around this time (de Graaf et al. 2008b). This is supported by the close relatedness between AMPV and HMPV subgroups, the high mutation rates of

RNA viruses, and the interchangeability of AMPV-C and HMPV proteins in chimeric viruses and minireplicon systems (de Graaf et al. 2008a; Govindarajan et al. 2006; Pham et al. 2005).

HMPV is subdivided into two genotypes, A and B, which are further subdivided into subgroups A1, A2 (consisting of A2a and A2b), B1 and B2 (van den Hoogen et al. 2004; Williams et al. 2004). Sequence analysis of 84 HMPV fusion protein (F) genes revealed 97% – 99% amino-acid sequence identity between members of different subgroups and 94% – 97% amino-acid sequence identity between members of different genotypes (van den Hoogen et al. 2004).

1.2.1 Epidemiology and Clinical Features

HMPV is a ubiquitous respiratory pathogen with worldwide distribution (Panda et al. 2014) with the majority of infections occurring in the winter and spring months, similar to influenza and RSV (van den Hoogen et al. 2003; Williams et al. 2006). Most children are infected within the first 5 years of life (seroprevalence is close to 100% by the fifth year of life) but reinfection of adults is common, despite the high seroprevalence (Haas et al. 2013). The annual infection rates of adults has been reported to vary between 3% – 7.1% (Walsh et al. 2008). A significant fraction (5-20%) of acute lower respiratory tract infections can be attributed to HMPV and the virus has been suggested to be the second most common cause of lower respiratory tract illness after RSV (Céspedes et al. 2016; Haas et al. 2013).

In healthy adults infection with HMPV is often asymptomatic or accompanied by mild cold-like symptoms. Symptomatic infections with the virus show features which are typical for lower respiratory tract illness and which usually last no longer

than one week. These include cough, fever, rhinorrhea (discharge of nasal fluids) and wheezing due to bronchiolitis and pneumonitis (Boivin et al. 2002, 2003; Céspedes et al. 2016). Clinical features caused by HMPV are usually indistinguishable from those caused by RSV. The virus is rarely detected in the serum of patients, indicating it is restricted to the respiratory tract (Schildgen et al. 2011). HMPV was also shown to be a lesser cause of upper respiratory tract infection in otherwise healthy children, causing rhinitis (Williams et al. 2006). In addition, HMPV has been linked to exacerbations of asthma (Williams et al. 2005a), although there is conflicting evidence on the matter (Schildgen et al. 2011). While one Australian study did not show a link between HMPV and asthma exacerbations (Rawlinson et al. 2003), a related study identified a significant one (Williams et al. 2004).

Infection with HMPV is often severe in cases where patients are immunocompromised, very young or elderly. In adults over 65 and frail subjects HMPV is reported to be especially significant (Schildgen et al. 2011). Elderly patients show a higher rate of hospitalization and exacerbated symptoms like congestive heart failure and severe pneumonia (Falsey et al. 2014; Liao et al. 2012; Te Wierik et al. 2012), which can even result in death (Boivin et al. 2002; Chiu et al. 2007; Liao et al. 2012). The presence of underlying conditions can additionally increase fatality rates (Schildgen et al. 2011). Similarly, very young age (under 5 months) is associated with increased hospitalization rates and more severe disease (Williams et al. 2010). Severe and fatal HMPV infections have also been reported in immunocompromised individuals, such as transplant recipients, or cancer patients (Boeckh et al. 2005; Cane et al. 2003; Williams et al. 2005b). Although rare, severe infection with HMPV has

even been shown to occur in otherwise healthy adults (Haas et al. 2012; Hasvold et al. 2016). The reasons behind severe infection in healthy adults are still unclear.

1.2.2 Structure and Components of the Virion

Members of the *Paramyxoviridae* form pleomorphic particles and are thus not amenable to cryo-EM single particle averaging approaches as is for instance the case with icosahedral viruses. Nonetheless much can be learned of the structures and functions of these viruses through X-ray crystallographic analyses of isolated components of the virion. In addition, cryo-EM tomographic studies can elucidate the overall architectures and provide a more holistic view of these viruses.

Much of what is known about the HMPV virion is inferred by homology to other pneumoviruses, in particular the extensively studied RSV. Virions of HMPV are highly variable with sizes ranging between 150 and 600 nm (Schildgen et al. 2011). These size variations may be accompanied with varying abundancies of viral components inside the virion (e.g. nucleocapsids, see further below). The viral particles possess a lipid envelope which harbour the G and F glycoproteins and the small hydrophobic SH protein (**Fig. 1.2**).

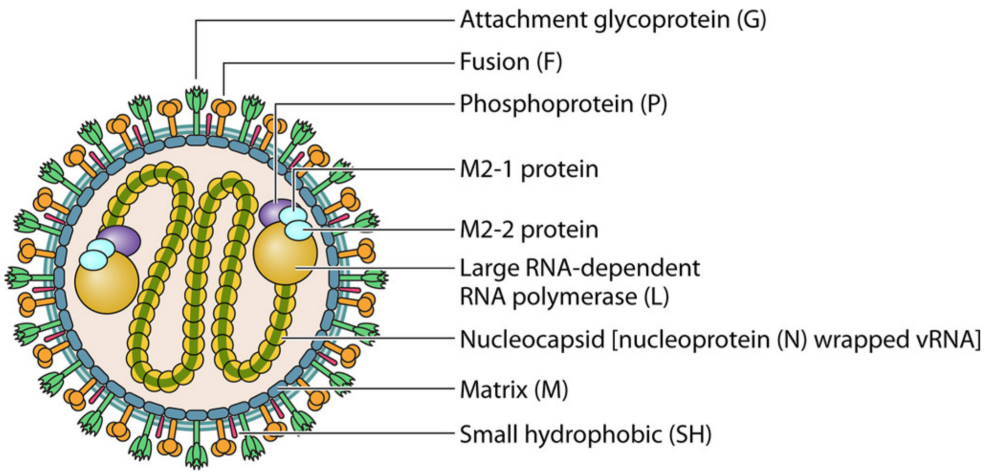


Figure 1.2: Schematic of an HMPV virion.
On the inside of the particle the nonsegmented (-)RNA genome is encased in a sheath of nucleoprotein N. This nucleocapsid complex is associated with the phosphoprotein P, the M2-1 and M2-2 proteins and the RdRP L. Just below the viral envelope the matrix protein M interacts with the cytoplasmic extensions of the glycoproteins F and G and forms the matrix layer. Three membrane proteins are inserted into the viral envelope: F, G and SH. Figure adapted from (Céspedes et al. 2016).

The F protein is a type I viral fusion protein which forms trimers on the viral surface and constitutes the main antigenic target (Cox et al. 2012; Wen et al. 2012). It contains the hydrophobic fusion loop which inserts into the target membrane of the host cell (see also 1.2.4). Crystal structures of the F-protein in its pre- and post-fusion conformations have been solved for RSV (McLellan et al. 2011, 2013). Interestingly, in cryo-EM tomographic analyses, both pre- and post-fusion F proteins have been observed on the surface of intact RSV (**Fig. 1.3**) and parainfluenza virions (Ludwig et al. 2008; Liljeroos et al. 2013). The heavily glycosylated and intrinsically disordered G protein has been suggested to contribute to adhesion to target cells and immune-evasion by a steric masking mechanism (Leyrat et al. 2014a). Many other paramyxoviruses (e.g. respiro- and rubulaviruses, see **Fig. 1.1**) possess a protein termed HN (for hemagglutination-neuraminidase activity) instead of G (El Najjar et

al. 2014). The neuraminidase activity assists the virus during release by cleaving the sialic acid groups exposed on the host cell surface which are exploited as attachment factors (Acheson 2006). In morbilliviruses such as measles this attachment protein is termed H as it lacks the neuraminidase activity. In contrast to viruses with HN, measles does not use sialic acid groups for attachment (instead it uses receptor proteins such as signalling lymphocytic activation molecule, SLAM, Schneider et al. 2002) and thus does not require the neuraminidase activity. It should be noted, that the G proteins of *Pneumovirinae* are not absolutely required for replication in primates and in tissue culture (Biacchesi et al. 2005) which is in contrast to the *Paramyxovirinae*.

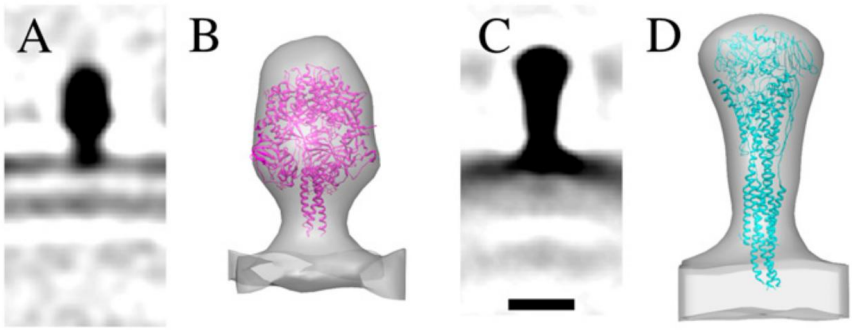


Figure 1.3: Pre- and Post-Fusion F on Intact RSV Virions.
Two differing conformations of RSV F spikes (amino-acid sequence identity of 33% vs. HMPV, van den Hoogen et al. 2002) have been observed on the surface of the virion. (A) and (C) depict central sections of subvolume averages of the surface spikes. Crystal structures of F in its pre-fusion (B) and postfusion (D) conformations could be fit into the corresponding densities. Figure taken from (Liljeroos et al. 2013).

A further protein that is embedded into the membrane envelope of HMPV is the SH protein. The protein is significantly larger in HMPV than its counterparts in related viruses (179 residues in HMPV vs. 65 in RSV) and is predicted to possess a single transmembrane (TM) helix (Masante et al. 2014). SH has been shown to be important for viral replication in animal models and in tissue culture in the case of the closely related AMPV (Ling et al. 2008). The same study also demonstrated that deletion of SH increased syncytium formation of infected cells. In primates, however, replication of HMPV was only slightly reduced after deletion of SH (Biacchesi et al. 2005) and SH seems to be dispensable in tissue culture (Biacchesi et al. 2004). Recent work strongly suggests that the SH protein is an oligomeric, cation-selective viroporin in both RSV and HMPV. RSV SH has been shown to change membrane permeability upon expression (Perez et al. 1997) and form pentameric pores in lipid bilayers (Gan et al. 2008). Similarly, HMPV SH forms higher order oligomers, alters membrane permeability and can regulate the activity of the fusion protein (Masante et al. 2014). Further work is necessary to clearly define the role of SH in the HMPV virion and during the replication cycle.

Matrix proteins of *Mononegavirales* are important coordinators of assembly and budding of nascent virions at the host plasma membrane (reviewed in Liljeroos and Butcher 2013). A layer of matrix proteins M is located just beneath the viral envelope of HMPV where M subunits are thought to interact with cytoplasmic extensions of F and G (Amarasinghe and Dutch 2014; Loo et al. 2013). The structures of both RSV and HMPV M proteins have been solved by x-ray crystallography and have revealed an arrangement of two consecutive beta-sandwich domains (**Fig. 1.4**) (Förster et al. 2015; Leyrat et al. 2014c; Money et al. 2009). Both RSV and HMPV M

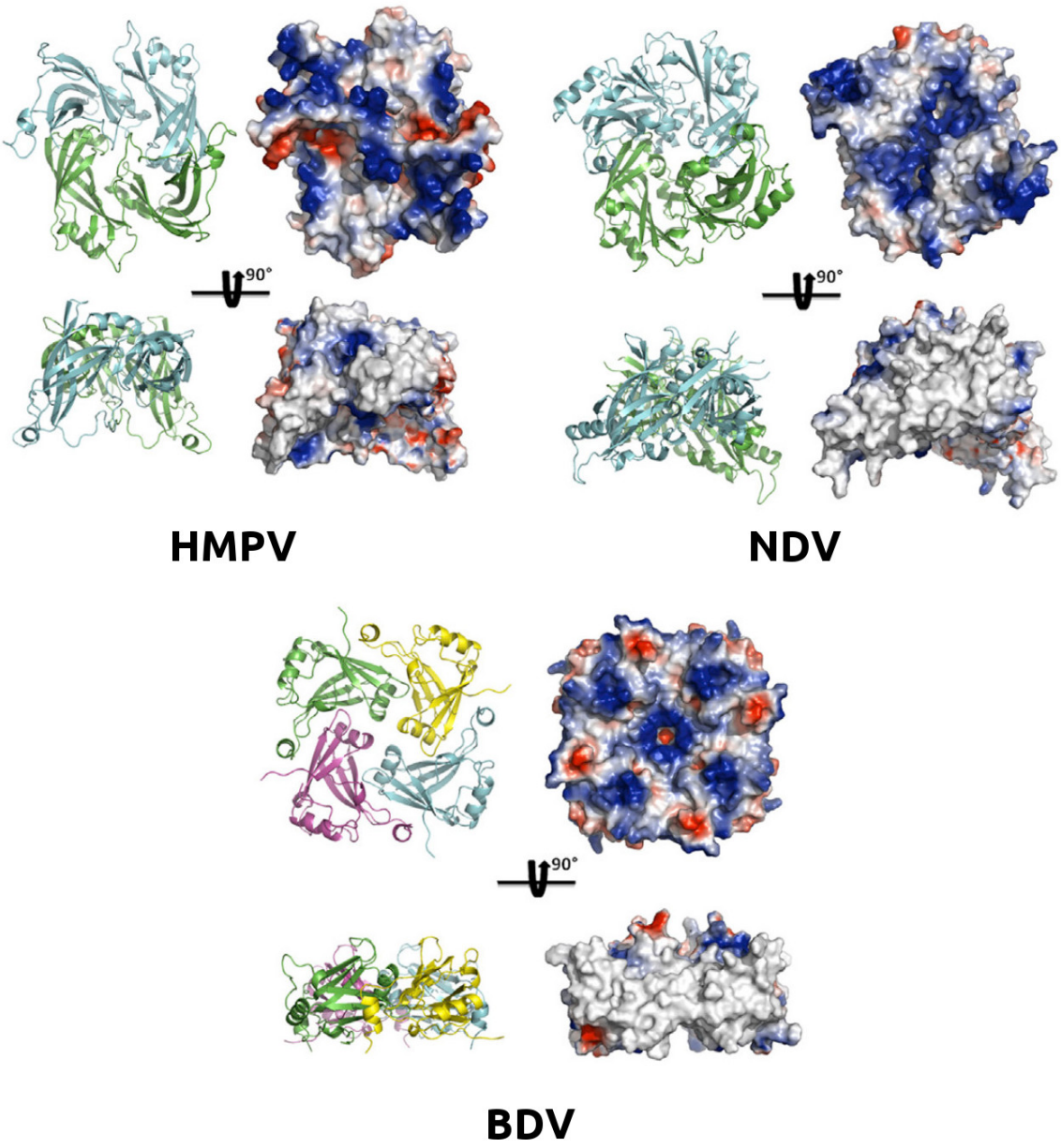


Figure 1.4: Quaternary Structures of Matrix Proteins.
Matrix proteins of human metapneumovirus (HMPV), Newcastle disease virus (NDV), and Borna disease virus (BDV) form oligomers with a similar diamond-shaped quaternary structure. Higher order M-arrays are thought to assemble from the diamond-shaped subunits and are associated with lipid bilayers. The lateral surfaces (see side-views) possess a distinctly hydrophobic character and have been reported to facilitate higher-order assembly. Figure adapted from (Leyrat et al. 2014c).

form dimers in solution exhibiting a diamond-like quaternary structure, similar to what has been observed for Newcastle disease virus (NDV) and BDV (**Fig. 1.4**) (Battisti et al. 2012; Neumann et al. 2009). Matrix proteins are known to interact with membranes and addition of lipids has been shown to result in spontaneous self-assembly of HMPV and RSV M into higher-order filamentous structures (Leyrat et al. 2014c; McPhee et al. 2011). These matrix protein arrays are formed through lateral interactions of the diamond-shaped dimers. Strikingly, HMPV M is also a calcium-binding protein and addition of calcium leads to a dramatic increase of thermal stability of M (Leyrat et al. 2014c). It is tempting to speculate that the cation-selective viroporin SH collaborates in the regulation of cytoplasmic calcium levels and thus controls functions of the matrix protein. However, this might be specific to metapneumoviruses as the calcium-binding activity of M is not conserved in RSV (Förster et al. 2015).

Within the virion the genetic information of the virus is stored as a nucleocapsid complex, of which multiple copies (exact amounts have not been quantified for RSV yet) can be found in each virus particle (Liljeroos et al. 2013; Loney et al. 2009; Kiss et al. 2014). The nucleocapsid is also associated with the phosphoprotein P, the antitermination factor M2-1 and the RNA-dependent RNA polymerase (RdRP). These are components of the transcription/replication machinery of HMPV and the focus of this work. As such I will devote later sections in this chapter to a detailed description (section 1.3).

A recent study has utilised cryo-EM tomography to render, in beautiful detail, the structure of virions from RSV, a close relative to HMPV (Liljeroos et al. 2013). The study revealed diverse morphologies, including filamentous and spherical

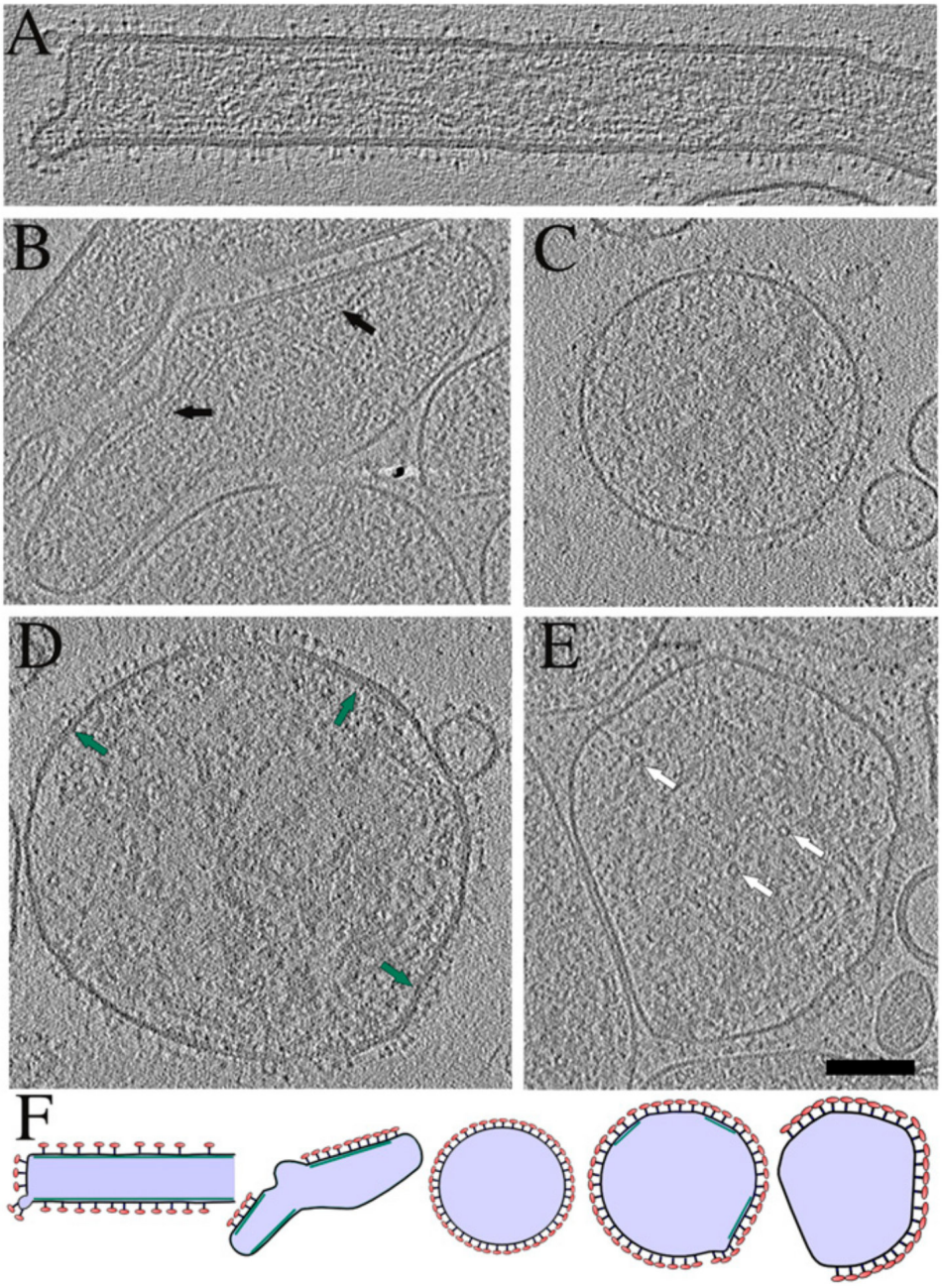


Figure 1.5: Tomography of RSV particles.
Panels A to E illustrate the morphological diversity and deformability of the virions. A schematic representation is shown in F. Both filamentous (A) and spherical (D) virions were observed. Density protruding from the viral surface has been attributed to the F protein. The black arrows in B indicate nucleocapsids viewed from the side, while the white arrows point to nucleocapsids that are viewed from the top. The green arrows indicate density matrix layer density in spherical particles. The figure was taken from Liljeroos et al. 2013.

particles (**Fig. 1.5**). A layer of density below the viral envelope was attributed to the M protein, although it seemed to be observable mainly in the filamentous forms of the virus and only rarely in spherical forms (Liljeroos et al. 2013). This further demonstrates the structure-giving properties of M. Multiple nucleocapsid filaments could be observed within the virions (black arrows, **Fig. 1.5B)** which possessed left-handed helical symmetry

1.2.3 Genome Organization and Gene Expression

The genome of HMPV is a single RNA molecule of roughly 13.3 kb length and contains sequences encoding 8 genes (**Fig. 1.6**). The overall order of encoded genes is conserved among nearly all *Mononegavirales* (Pfaller et al. 2015). The prototypical mononegavirus genome begins with the nucleoprotein N, which is found at the 3'-end, followed by a gene encoding a polymerase cofactor (P in paramyxoviruses), the matrix protein M, the surface glycoproteins and, finally the L RdRP at the 5'-end. Both pneumoviruses (e.g. RSV) and metapneumoviruses also possess the M2 and SH genes (**Fig. 1.6B**). By utilising multiple open reading frames (ORFs) the M2 gene codes for both the M2-1 and M2-2 proteins (Schildgen et al. 2011). Interestingly, the order of genes differs between pneumoviruses and metapneumoviruses, with the F and M2 genes being located further towards the 3'-end in pneumoviruses (**Fig. 1.6B**). It should also be noted that RSV encodes two additional non-structural proteins, NS1 and NS2, which have been associated with immune subversion but are absent in HMPV (Lo et al. 2005).

In contrast to positive-sense RNA viruses, the negative-sense genomes of *Mononegavirales* cannot be directly translated by the host cell machinery. Instead,

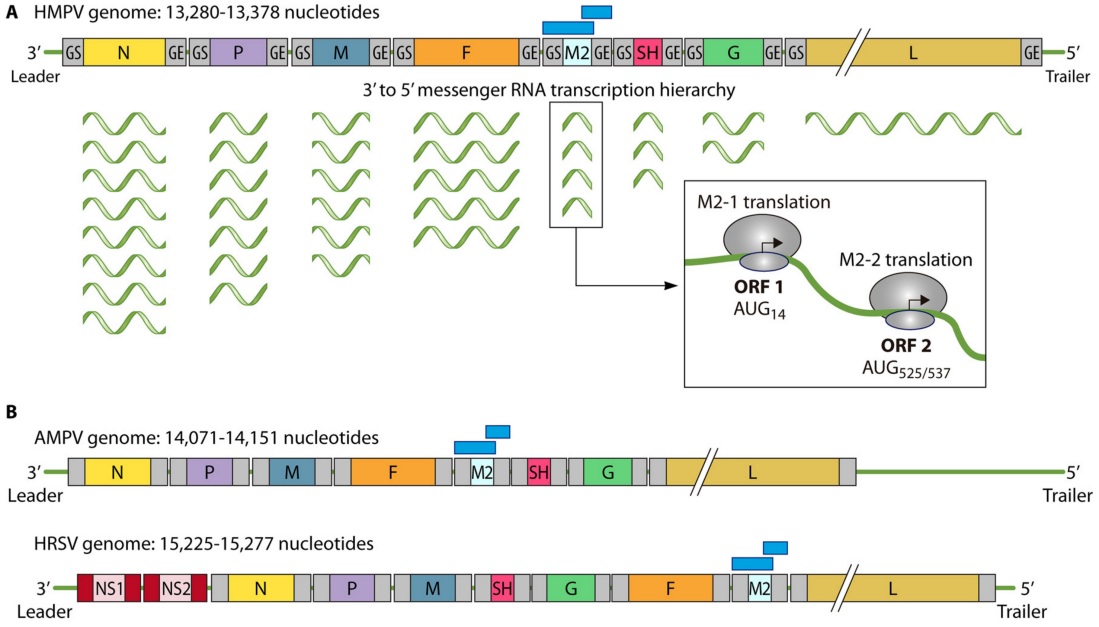


Figure 1.6: Genome Organisation of *Pneumovirinae*.
Genome organisation of HMPV (A), AMPV (B, upper schematic) and HRSV (B, lower schematic). The genomes begin with a short leader sequence at their 3'-ends and terminate with a trailer sequence at their 5'-ends. The genes are flanked by gene start (GS) and gene end (GE) sequences, which regulate transcription initiation and termination of subgenomic mRNAs (depicted as green strands). Short intergenic sequences are located between GE and GS. Because the transcriptase has a chance to disengage the template at each gene junction a gradient of mRNAs develops, with high abundance of N mRNAs and low abundance of L mRNAs. The M2 mRNA can utilise two ORFs and codes for the two proteins M2-1 and M2-2 (see inset). RSV codes for two additional nonstructural proteins, NS1 and NS2 (B), which are absent in HMPV and AMPV. Figure taken from (Céspedes et al. 2016).

mononegaviruses must package a multi-component transcription machine (L, P, N and M2-1 in HMPV) in the virion, allowing them to generate mRNAs. Extensive studies of transcription in VSV have elucidated common principles of gene expression in mononegaviruses (Knipe and Howley 2007). The transcription process initiates at the 3'-end of the genome, where a leader sequence is located which acts as promoter (Herfst et al. 2004). In general, transcription proceeds linearly and sequentially along the genome in a "start-stop-restart" process and can only initiate at the 3'-end and not within the genome. Intergenic regions at gene junctions contain

conserved gene start (GS) and gene end (GE) sequences which direct termination and reinitiation of the transcriptase (**Fig. 1.6A**). When the polymerase reaches the GE region of a gene junction it is thought to stutter at a poly-U tract, generating a poly-A tail and consecutively releasing the transcript (Whelan et al. 2004). The polymerase then continues moving along the template until it restarts transcription at the next GS site. Importantly, at each gene junction the transcriptase has a chance of dissociating from the template instead of reinitiating, which leads to a gradient of decreasing subgenomic mRNA abundance from the N gene to the L gene (**Fig. 1.6A**). The described mechanism directly controls the amount of gene product being translated by the host cell. However, there are virus-dependent variations of this theme, for instance read-through synthesis of consecutive genes (yielding poly-cistronic mRNAs) via the pneumoviral M2-1 protein. This will be discussed in more detail further below.

1.2.4 Lifecycle and Replication

Much of the HMPV replication cycle is inferred by homology from other paramyxoviruses, in particular RSV. An overview of the HMPV lifecycle is shown in **Fig. 1.7**. The intricate details of the attachment and fusion steps of the lifecycle (reviewed in Cox and Williams 2013) are not the focus of this work and, as such, will be outlined briefly in this section. As described above, mature virions possess two surface glycoproteins which are thought to be important in the early stages of infection, the F and G protein. In analogy to RSV, the glycosaminoglycan-binding G protein has been suggested to facilitate adhesion to target cells. However, conservation between HMPV and RSV G is low (amino-acid sequence identity of

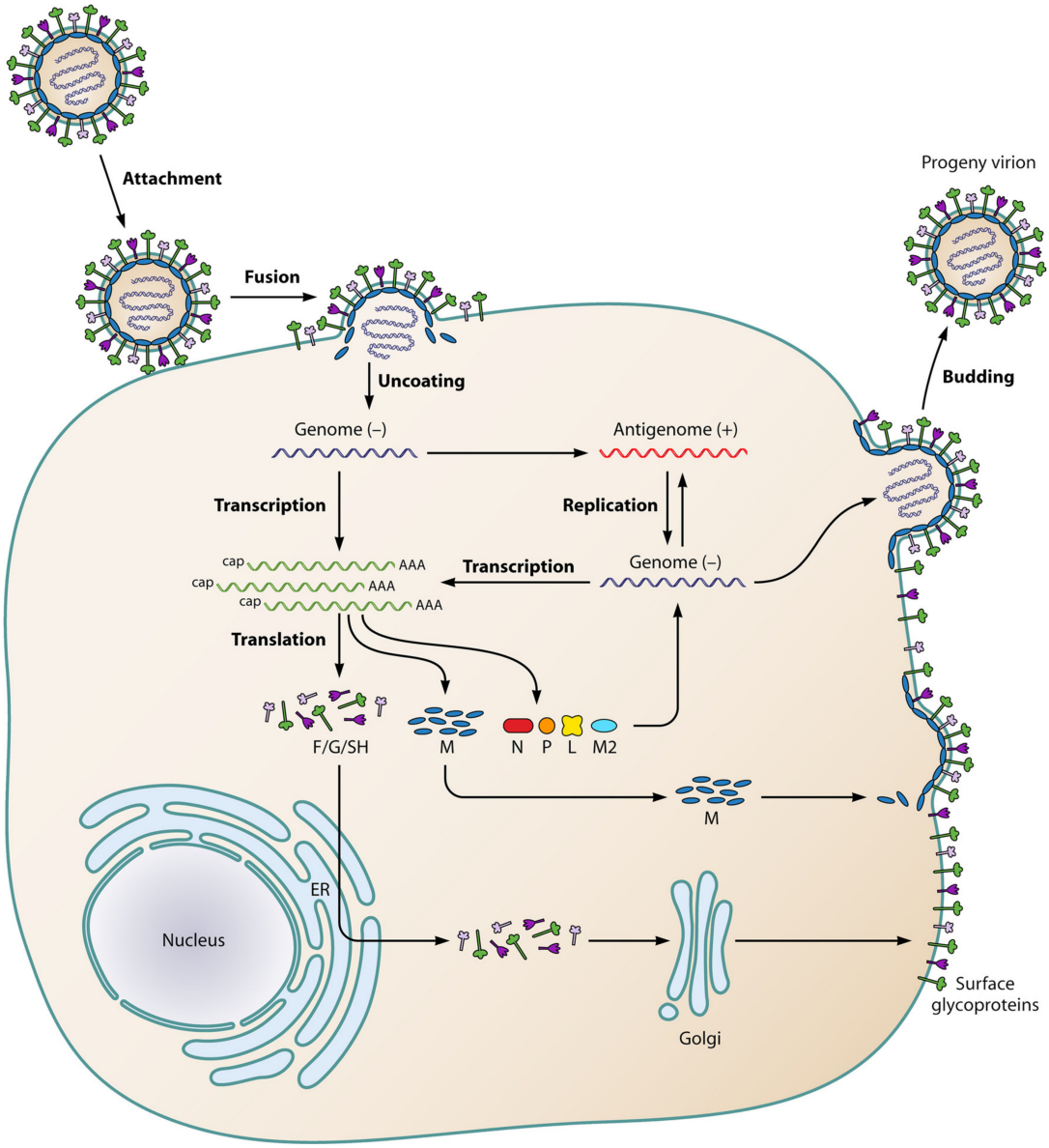


Figure 1.7: Lifecyle of HMPV.
The infectious particle attaches to the target host cell membrane via the F and G glycoproteins. Interactions between F and RGD-binding integrins and G/F with glycosaminoglycans have been suggested to mediate attachment (see accompanying text). Some evidence suggests that HMPV does not fuse with the plasma membrane like other paramyxoviruses (as depicted) but instead is internalised via clathrin-coated endosomes. After uncoating the nucleocapsid and transcription components are released into the cytosol. The viral transcription machinery produces subgenomic mRNAs which are translated into the corresponding viral proteins. The F, G and SH proteins proceed along the secretory pathway through the ER and Golgi apparatus. When sufficient N has accumulated antigenomes (+) are synthesised which then serve as templates for genome synthesis (-). Budding and assembly of progeny virions occurs at the plasma membrane. There is evidence that in pneumoviruses pre-assembly complexes consisting of nucleocapsids, transcriptase/replicase components, and M protein form in cytoplasmic inclusions before being transported to the glycoproteins at the plasma membrane. Figure taken from Schildgen et al. 2011.

~20%) (Biacchesi et al. 2003; van den Hoogen et al. 2002) and G-deficient HMPV can actively replicate in primates and tissue culture (Biacchesi et al. 2004, 2005). Instead, there is evidence that F mediates both fusion and attachment (Cox et al. 2012). Recent studies strongly suggest that HMPV F alone is sufficient for attachment by binding to multiple RGD-binding integrins (Cox et al. 2012; Cseke et al. 2009; Wei et al. 2014). Indeed, in 79 F sequences from HMPV lineages A1, A2, B1, and B2, an RGD motif (residues 329-331) was found to be strictly conserved (Cseke et al. 2009). In addition, F has been shown to bind to glycosaminoglycans (Chang et al. 2012). A picture is emerging in which both the F and G proteins mediate initial low-affinity adhesion to extracellular carbohydrate structures, followed by more specific integrin-mediated binding solely by F.

All enveloped viruses need to fuse with a cellular membrane to transport their genetic material into the host. In general, members of the *Paramyxoviridae* family, like HMPV, fuse with the plasma membrane of the target cell at neutral pH (Russell and Luque 2006). This is in contrast to many viruses which are first internalised into endosomes (e.g. dengue, VSV, Semliki Forest virus, etc.) and subsequently initiate the fusion reaction through a low-pH environmental trigger in the course of the endocytic pathway (reviewed in White and Whittaker 2016).

Interestingly, in some genotype A strains of HMPV low pH seems to promote fusion events, whilst for genotype B strains this is not the case (Herfst et al. 2008; Mas et al. 2011; Schowalter et al. 2006, 2009). The pH-dependent phenotype has been linked to a tetrad of residues within the F protein (amino acids ENWP at positions 294, 296, 396, and 404). Notably, one study shows that HMPV may be internalised into human bronchial epithelial cells via a clathrin-dependent pathway,

although the consecutive fusion process was not dependent upon pH (Cox et al.2 2015). Clathrin-dependent internalization has also been described for influenza and VSV (Cureton et al. 2009). In the case of VSV, there is evidence that the viral cargo enters the cells via unconventional, only partially clathrin-coated vesicles which can accommodate much larger cargo (such as pleomorphic viruses) (Cureton et al. 2009). It may be that HMPV utilises multiple pathways to gain access to the host cell and these may be strain-specific. More work is necessary to clearly understand the process of attachment and fusion in HMPV.

Following uncoating of the virion, the nucleocapsids and replication/transcription components are released into the cytoplasm. The genome serves as the template for both transcription (synthesis of viral mRNAs) and replication (synthesis of viral progeny genomes). As described above, transcription yields a gradient of capped and polyadenylated mRNAs which are translated by host ribosomes to produce viral proteins. In replication mode, however, the polymerase reads through the gene-junctions and new full-length RNA antigenomes (positive polarity) are synthesised and co-transcriptionally bound by N protein. The resulting antigenomes then serve as templates for the synthesis of progeny genomes (negative polarity), which are also encapsidated by N co-transcriptionally (Acheson 2006).

The switch between transcription and replication has been thoroughly studied in VSV and RSV and a model of how this is achieved in *Pneumovirinae* has recently been proposed (Noton and Fearns 2015). Transcription by the RSV polymerase usually initiates opposite nucleotide 3 of the leader at the beginning of the template (Noton and Fearns 2011; Tremaglio et al. 2013). A short leader is then transcribed and released, followed by recognition of an adjacent GS signal at the beginning of a gene (**Fig. 1.8A**). However, it is thought that the polymerase can also (although very rarely) ratchet backwards from position 3 and utilise a self-generated 5'-AC primer, giving the impression that transcription starts at nucleotide 1 of the leader (**Fig. 1.8B**). The presence of this 5'-AC seems to be the signal for co-transcriptional encapsidation and thus replication (Noton and Fearns 2015). It is important to note that transcripts

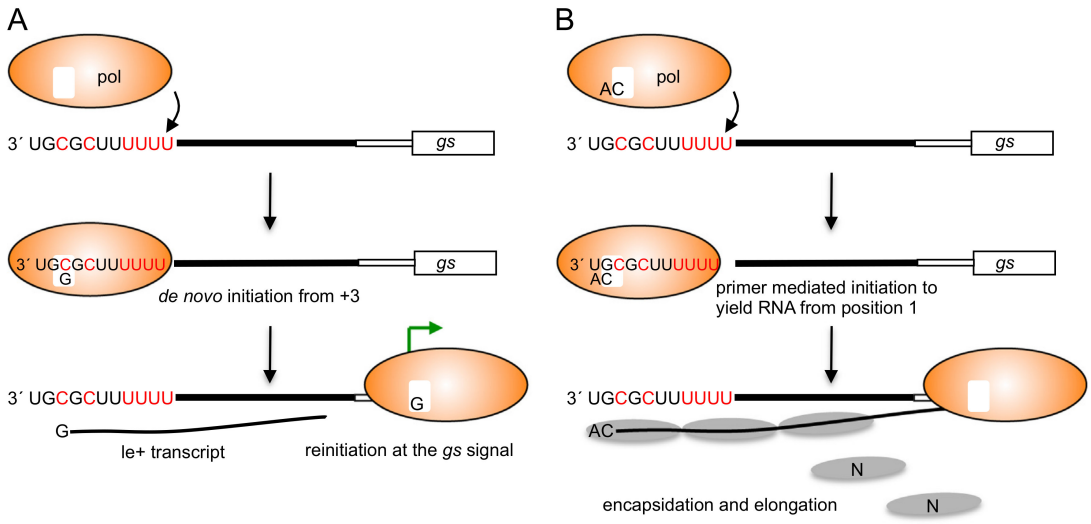


Figure 1.8: Model of Transcription and Replication in RSV.
(A) In the majority of cases transcription starts opposite nucleotide 3 of the leader at the 3'-end of the genome. A short abortive leader transcript is synthesised (le+) and then released. The polymerase continues along the template until it recognises the first GS of the N protein and then reinitiates transcription. (B) In a small number of cases the polymerase can ratchet backwards from nucleotide 3, base-paring a self-generated 5'-AC primer with the first two nucleotides of the leader. Transcripts initiating in this way appear to have started at nucleotide 1 of the template. The 5'-AC is the signal for co-transcriptional encapsidation of the nascent transcript. If N protein is not available transcription proceeds as in (A). However, if sufficient N (grey ovals) is available, the 5'-AC transcript is immediately encapsidated and the polymerase goes into a super-processive mode in which a whole antigenome is transcribed, ignoring GE and GS signals. Figure taken from Noton and Fearns 2015.

initiating opposite nucleotide 3 cannot become replication products. Co-transcriptional encapsidation of the 5'-AC initiated transcript (and thus commitment to replication) can only occur if sufficient amounts of N have accumulated in the cell in the late stages of the lifecycle. To summarise, while transcription is constitutive and starts opposite nucleotide 3 of the leader, replication is switched on for the small sub-population of transcripts that initiate with a 5'-AC primer when intracellular levels of N are high enough.

Generally, for budding of infectious *Mononegavirales* particles to occur, the surface glycoproteins, the matrix protein and the nucleocapsid components need to assemble at the plasma membrane of the host cell (El Najjar et al. 2014; Schmitt and Lamb 2004). The matrix protein is considered the principal organiser of the budding process, serving as a molecular bridge between the different virion components and promoting membrane curvature and scission (Liljeroos and Butcher 2013). M proteins from several viruses can bind to biological membranes and to the cytoplasmic tails of the glycoproteins, thus recruiting M to budding sites (El Najjar et al. 2014). Binding of M to cytoplasmic tails of G has been shown for HMPV and RSV (Ghildyal et al. 2005; Loo et al. 2013). In addition, the cytoplasmic tails of F have been demonstrated to be important for the assembly of the virion and budding of viral filaments (Batonick and Wertz 2011; Shaikh et al. 2012). While the F, G and SH proteins reach the plasma membrane by proceeding through the secretory pathway, the nucleocapsids have to be recruited to the budding sites through other means. It has been demonstrated for a number of *Paramyxovirinae* that the M protein binds to nucleocapsids and mediates their incorporation into progeny virions (Coronel et al. 2001; Iwasaki et al. 2009; Schmitt et al. 2010). In pneumoviruses the situation seems

to be more complicated, however. The available evidence suggests that here the M protein is linked to the nucleocapsid components via M2-1 which all together form pre-assembly complexes in cytoplasmic inclusions prior to transport to the assembly site (El Najjar et al. 2014; Kiss et al. 2014; Li et al. 2008a).

1.3 The Transcriptase/Replicase Machinery

The RNA synthesis machinery of mononegaviruses utilises the nucleocapsid as its assembly platform and transcription/replication occurs in the vicinity of viral template which is contained in a sheath of nucleoprotein (Ruigrok et al. 2011a). In pneumoviruses this machine consists of the following parts: (1) the L (large) protein which possesses RdRP, polyadenylation, methylation and capping activities. (2) the polymerase co-factor P (or phosphoprotein) which is attributed with a variety of functions, including positioning of L on the template, chaperoning N and recruitment of M2-1. (3) The antitermination factor M2-1 which is important for synthesis of full-length transcripts in transcription mode (it can be seen as an essential processivity factor). (4) The nucleocapsid which constitutes the template.

1.3.1 The Nucleoprotein N

The N protein is the sole protein responsible for forming the protective sheath around the genomes of members of *Mononegavirales*. In these viruses the genome is never found naked in the cytoplasm (Ruigrok et al. 2011a). Instead, it is sequestered in homo-oligomerised subunits of N (forming the nucleocapsid) to protect the genetic material from digestion by host nucleases and prevent activation of antiviral

signalling in host cells. Nucleocapsids of mononegaviruses are known to vary in morphology and are often flexible, with heterogeneous pitch and helical symmetry parameters (Bhella et al. 2002; Cox and Plemper 2015). A direct consequence of the encapsidation is that stretches of RNA need to be transiently released so that the elongating polymerase L may access the genetic information (Liang et al. 2015). The P protein is thought to position L onto the nucleocapsid template, but is also important in chaperoning naive, RNA-free nucleoprotein (N^0) to replication sites, thus facilitating assembly of progeny nucleocapsids (section 1.3.2). It is noteworthy that N proteins are interesting targets for therapeutic intervention, potentially offering low rates of resistance development (Cox and Plemper 2015). One example is the RSV inhibitor RSV604 which has been shown to bind directly to N (Challa et al. 2015; Chapman et al. 2007). RSV604 has been shown to inhibit synthesis of viral RNAs in HeLa cells by an as-of-yet undetermined mechanism (Challa et al. 2015).

Our structural knowledge of the nucleocapsid building block N is primarily derived from x-ray crystallography. An interesting phenomenon observed with the recombinant expression of N proteins in bacteria is that these spontaneously assemble into RNA-bound sub-nucleocapsid rings, which c an be interpreted to represent a single turn of the nucleocapsid in absence of helical rise (Albertini et al. 2006; Tawar et al. 2009). Although these ring-like structures are likely to be non-physiological they present an opportunity for crystallographic analysis of nucleoprotein oligomerisation and RNA-uptake. Crystal structures of N proteins from mononegaviruses reveal a common core architecture consisting of a C-terminal domain (CTD) and a N-terminal domain (NTD) that are connected via a hinge-like helix (**Fig. 1.9**). The viral RNA is inserted into a positively charged cleft formed

between the CTD and NTD (**Fig. 1.10**) (Albertini et al. 2006; Green et al. 2006; Tawar et al. 2009).

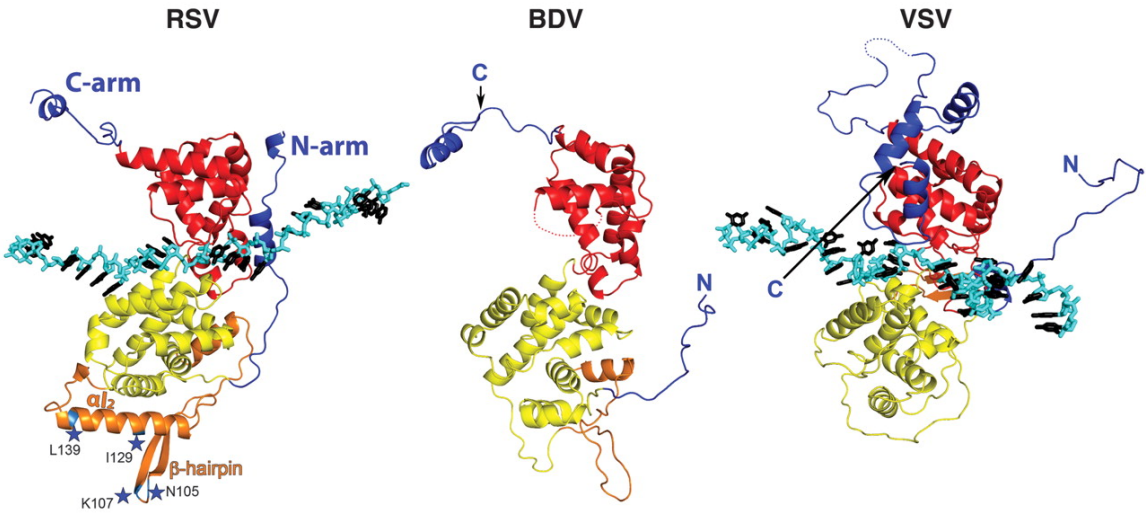


Figure 1.9: Architecture of Mononegavirus Nucleoproteins.
The crystallographic structures of respiratory syncytial virus (RSV), Borna disease virus (BDV), and vesicular stomatitis virus (VSV) N are shown. N proteins share a similar architecture with a conserved C-terminal domain (CTD, coloured in red) and N-terminal domain (NTD, coloured in yellow). In addition, the region coloured in orange is most variable within *Mononegavirales*. The N proteins oligomerise laterally via C-terminal and N-terminal extensions (coloured in blue) that latch onto neighbouring subunits. Mutations which confer resistance to the N-inhibitor RSV604 are indicated with stars. Figure taken from Tawar et al. 2009.

Nucleoproteins oligomerise linearly along the viral genome with each protomer of N contacting six nucleotides in the case of *Paramyxovirinae* (known as the rule-of-six, Calain and Roux 1993), seven nucleotides in *Pneumovirinae* and nine nucleotides in *Rhabdoviridae* (Ruigrok et al. 2011a). The lateral interactions between the protomers are dominated by C-terminal and N-terminal exchange subdomains (CTD-arm and NTD-arm, see **Fig. 1.9**) that latch onto the neighbouring N subunits, thus facilitating oligomerisation. In rhabdoviruses the CTD-arm is longer and, after

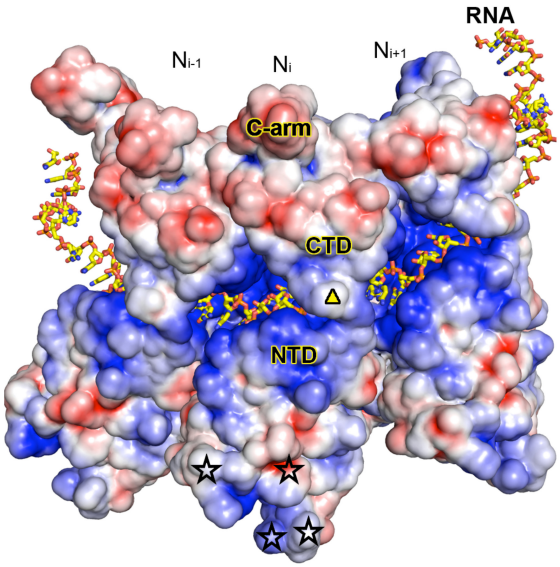


Figure 1.10: RNA-Binding Groove of RSV N. Three neighbouring protomers (indicated) of RSV N in surface representation. The surface is coloured according to charge from red (negative) to blue (positive). The RNA is shown in stick representation. The CTD, NTD and CTD-arm (C-arm) are indicated. The stars indicate positions of mutations conferring resistance to a anti-RSV compound, RSV604 (see original publication, Tawar et al. 2009).

contacting a neighbouring N, wraps back onto the nucleoprotein from which it originated (see VSV N in **Fig. 1.9**). For paramyxoviruses the strand of RNA follows a continuous outer groove along the outside of the oligomerised nucleocapsid (**Fig. 1.10**). Interestingly, while this groove is on the exterior in RSV, it is found on the interior of the ring in the Rhabdoviruses VSV and Rabies virus (Albertini et al. 2006; Green et al. 2006). Helical reconstruction from cryo-EM data of VSV virions confirmed the positioning of the RNA: in the rod-shaped VSV particles the nucleocapsid is just below the matrix layer and the RNA faces inwards, towards the particle interior (Ge et al. 2010).

The nucleoprotein most closely related to HMPV, of which the structure has been determined by crystallography, is that of RSV (sequence identity ~40%). In addition, previous cryo-negative stain EM studies have imaged nucleocapsid-like filaments which were produced by recombinant expression of RSV N in insect cells (Bakker et al. 2013; Tawar et al. 2009). The studies found that the RSV nucleocapsids possessed a considerable amount of flexibility (**Fig. 1.11A**), complicating high-resolution reconstruction. Nonetheless, combining the crystallographic data with a

map reconstructed from the EM data, a full atomistic model of RSV N proteins assembled into a left-handed helical complex could be obtained (**Fig. 1.11B**). In contrast to *Pneumovirinae*, *Paramyxovirinae* N proteins have an additional stretch of ~120 residues which are predicted to be intrinsically disordered. This so-called "N-tail" has been shown to be cleavable by trypsin digestion in recombinant nucleocapsids which leads to a considerable rigidification of the nucleocapsid helix (Ruigrok et al. 2011a).

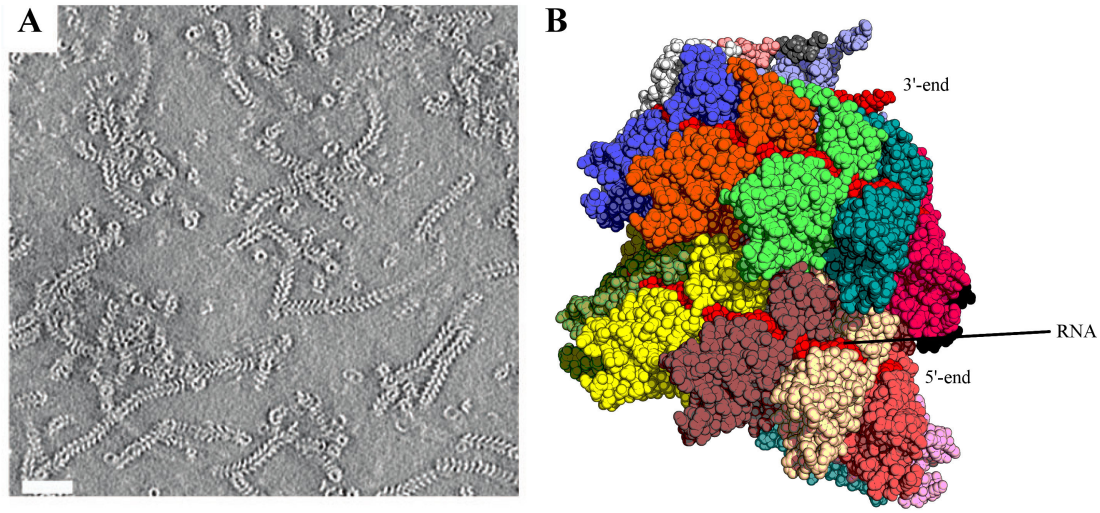


Figure 1.11: Recombinant Nucleocapsids of RSV.
(A) Section of a tomogram of recombinant RSV nucleocapsids obtained via cryo-negative stain EM. The filamentous nucleocapids reveal a typical "herringbone" morphology. (B) Model of oligomerised RSV N subunits forming a left-handed helical superstructure. Each copy of N is coloured individually, while the strand of RNA (indicated) is coloured in red. The RNA follows a continuous, positively charged groove along the exterior of the assembly. Panel A taken from Bakker et al. 2013. Panel B was rendered from PDB ID: 4bkk.

1.3.2 The P Protein

The polymerase cofactor P is often considered to be the central organiser of the transcription complex. P brings together the viral components of transcription and replication in a spatially and temporally coordinated manner (Cox and Plemper 2015). Of the *Mononegavirales* all family members except the *Filoviridae* possess a P protein (Pfaller et al. 2015). Instead, filovirus genomes encode the VP35 protein which shares key similarities to P (Leung et al. 2010).

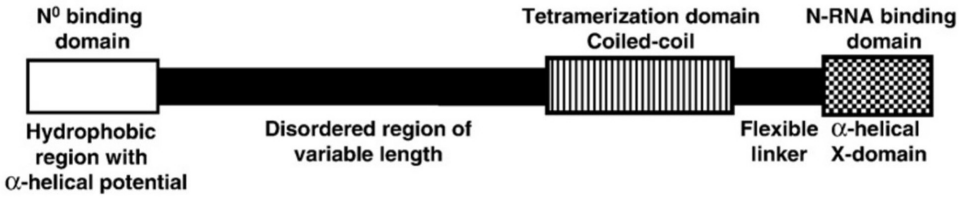


Figure 1.12: Architecture of *Paramyxovirinae* P.
Schematic of the organization of P proteins in Paramyxovirinae (Karlin et al. 2003).

In general, P proteins are organised in a modular fashion (see **Fig. 1.12** for the P organization in *Paramyxovirinae*), with distinct functional domains separated by large intrinsically disordered regions (Habchi and Longhi 2012; Karlin et al. 2003). The common architecture of P consists of an oligomerisation domain in the centre of the protein (P_{OD}), a nucleocapsid binding domain (P_{NBD}) at the C-terminus, and a partially ordered N-chaperoning region at the N-terminus (Karlin et al. 2003). Despite these similarities P proteins vary greatly in length between viruses and often have no detectable sequence conservation between viral families (Karlin et al. 2003; Pfaller et al. 2015). The P proteins are attributed with a variety of functions in transcription, replication and immune evasion. Some of these properties of P are conserved

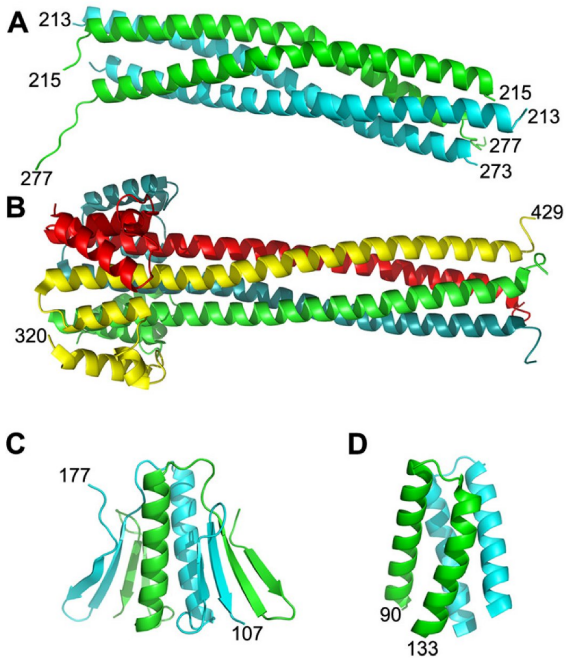


Figure 1.13: P Oligomerisation Domains.
The P proteins of Paramyxoviruses oligomerise via tetrameric coiled coils. Examples are shown for mumps virus (A) and Sendai virus (B). In rhabdoviruses P proteins form dimers instead. Structures of the dimerisation domains of VSV (C) and rabies virus (D) are shown. Figure adapted from Cox et al. 2013.

throughout all mononegaviruses, while others are virus-specific. In the following sections I will provide an overview of the most important features of P, starting with its oligomerisation domain.

A hallmark of the P protein is that it self-assembles into a distinct oligomeric state, a tetramer in the case of *Paramyxoviridae* (Pfaller et al. 2015). The oligomerisation domains have been solved for a variety of viruses, including Sendai virus (Tarbouriech et al. 2000), VSV (Ding et al. 2006), rabies virus (Ivanov et al. 2010), measles virus (Communie et al. 2013a), mumps virus (Cox et al. 2013), and Nipah virus (Bruhn et al. 2014). The tetramerisation domains of the paramyxoviruses and pneumoviruses (Llorente et al. 2006, 2008) are composed of extended helices which form coiled coils (**Fig. 1.13A** and **B**). Surprisingly, while the helices of the coiled coil are arranged in a parallel fashion for Sendai, Nipah, and measles virus (all of the N- and C-termini of the helices being located at the same end of the bundle), in the case of mumps virus the tetramerisation domain is formed by two dimers in antiparallel orientation (**Fig. 1.13A**). This places the N-termini and C-termini of the dimers at opposite ends of the tetramer. The oligomerisation domains of

Rhabdoviridae facilitate the formation of P dimers instead of tetramers. In VSV the dimer is formed through domain-swapping of a β-hairpin (Ding et al. 2006), while in rabies virus each monomer contributes a helix-loop-helix structure (Ivanov et al. 2010). The importance of P multimerisation for the proper functioning of the transcription and replication machinery has been demonstrated for bornaviruses, paramyxoviruses and rhabdoviruses (Chen et al. 2006; Kolakofsky et al. 2004; Schneider et al. 2004).

An important function of P is to ensure that the L polymerase is positioned onto the nucleocapsid template by acting as a molecular bridge, binding both L and N. Previous studies utilising x-ray crystallography have shown that the VSV P_{NBD} binds two adjacent subunits of assembled and RNA-bound nucleoprotein (Green and Luo 2009). P_{NBD} structures have also been determined for rabies virus and Mokola virus, revealing a well-ordered, mostly helical domain resembling a sliced pear (Assenberg et al. 2010; Mavrakis et al. 2004).

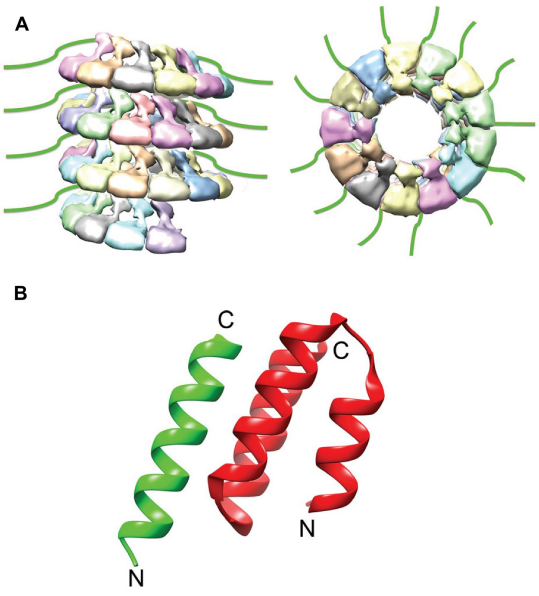


Figure 1.14: Interaction between N and P in *Paramyxovirinae*.
(A) Schematic of a helical nucleocapsid with the intrinsically disordered N-tail (green) extending from its subunits. (B) Structure of the measles virus P_{NBD} (red) in complex with a helix of the N-tail (green). Figure adapted from Cox and Plemper 2015.

The P_{NBD} is highly conserved within the *Paramyxovirinae*, however, the fold is different to that of rhabdovirus Ps. Several P_{NBD} structures have been solved for *paramyxovirinae*, including measles virus (Johansson et al. 2003; Kingston et al. 2004), mumps

virus (Kingston et al. 2008), and Hendra virus (Communie et al. 2013b), revealing a compact three-helix bundle. These domains have been shown to attach to the nucleocapsid by binding a region of the intrinsically disordered N-tail of N, thereby inducing folding of a transient helix (Erales et al. 2015) (**Fig. 1.14**). The affinity of P to N-tail has been reported to be low (μM range) consistent with the notion that the P-L complex needs to progress rapidly along the template, with the P_{NBD} binding and unbinding N-tails along the way.

In RSV from the *Pneumovirinae* subfamily it was shown that nine residues at the C-terminus of P facilitate binding to the nucleocapsid with the C-terminal Phe being especially important (Ouizougun-Oubari et al. 2015; Tran et al. 2007). Even though the P_{NBD} sequences are not conserved between *Mononegavirales* families it has been suggested that there is a degree of structural homology. A recent study proposed that all P_{NBD} structures may have a shared minimum topology and arrangement of secondary structure elements, justifying the construction of a structural phylogenetic tree (**Fig. 1.15**) and providing further evidence of the relatedness of mononegaviruses (Martinez et al. 2013).

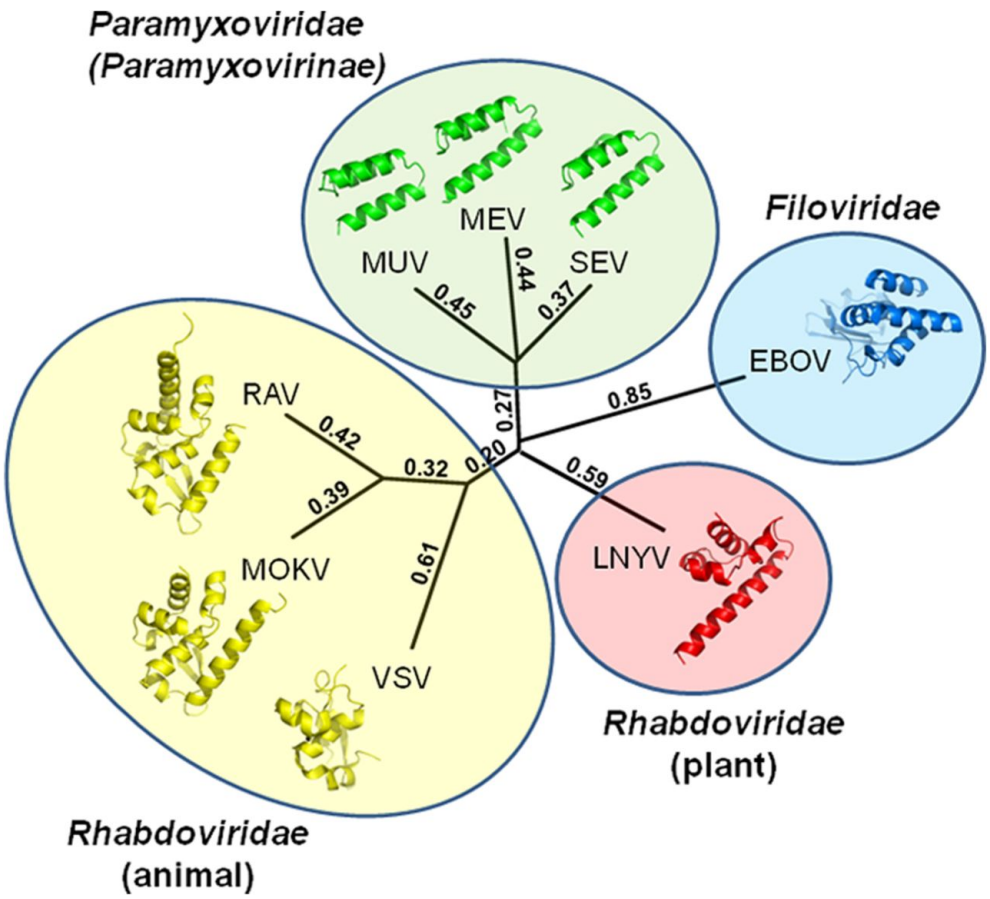


Figure 1.15: Structural Phylogeny of P.
The ordered C-terminal regions of the polymerase cofactors from different mononegaviruses were used to generate a structural phylogenetic tree. The viral families are indicated. Viruses are abbreviated as follows: *EBOV*, Ebola virus, *LNYV*, lettuce necrotic yellows virus, *MEV*, measles virus, *MOKV*, mokola virus, *MUV*, mumps virus, *RAV*, rabies virus, *SEV*, Sendai virus, *VSV*, vesicular stomatitis virus. Figure taken from Martinez et al. 2013.

To date there is only sparse data on how P proteins interact with the L-polymerase. For VSV it has been demonstrated that P acts as a processivity factor and binding of P to L induces structural rearrangements in the polymerase (Morin et al. 2012; Rahmeh et al. 2010). The region of VSV P which is responsible for this interaction and facilitates the conformational changes has been mapped to residues 41-106 (Rahmeh et al. 2012). The L-interacting domain has also been mapped for

RSV (Sourimant et al. 2015). In this study, the region consisting of P residues 216-239 was shown to bind full-length L and its importance for RNA synthesis was validated by introducing point mutations in a minigenome assay. The region includes a transient alpha-helix which was suggested to fold-upon-binding to the L-protein. Besides binding to L, P from *Pneumovirinae* also recruits the antiterminator M2-1 (see section 1.3.4) within the transcriptive complex (Mason et al. 2003; Tran et al. 2009). This interaction has been shown to be dependent on a region of P just upstream of the oligomerisation domain with residues 100-120 of RSV P being especially important (Asenjo et al. 2006; Mason et al. 2003). In addition, a truncated fragment of RSV P encompassing residues 100-166 was shown to be sufficient to bind to M2-1.

Finally, the P protein also acts as a molecular chaperone for N, inhibiting premature binding of RNA and oligomerisation (Chen et al. 2007; Curran et al. 1995; Masters and Banerjee 1988; Mavrakis et al. 2006). As this supplies the replicative polymerase with fresh N proteins needed to build a new nucleocapsid it constitutes a key function of P. A region in the N-terminal portion of P acts as a molecular recognition element and binds nascent RNA-free N (N^0). This is necessary because N would otherwise take up unspecific RNA of the host cell, leading to a dead-end state. By keeping N^0 free of RNA and delivering it directly to sites of viral replication, P ensures that N only encapsidates viral genomic RNA (Ortín and Martín-Benito 2015). At the time my project was initiated our structural understanding of this process stemmed solely from the crystal structure of VSV N bound to a N-terminal portion of P (Leyrat et al. 2011). The structure explains how P inhibits RNA-binding by inserting into the RNA-binding groove of N (**Fig. 1.16**). Puzzlingly, this VSV N^0-P complex crystallised as a decameric ring, in a conformational and oligomeric state

identical to that of N-RNA (Leyrat et al. 2011). In contrast, the solution state of N^0-P was shown to be that of a heterodimer. Furthermore, P sequences are not conserved between VSV and paramyxoviruses, leaving many open questions on how N-chaperoning is achieved in viral families beyond *Rhabdoviridae*.

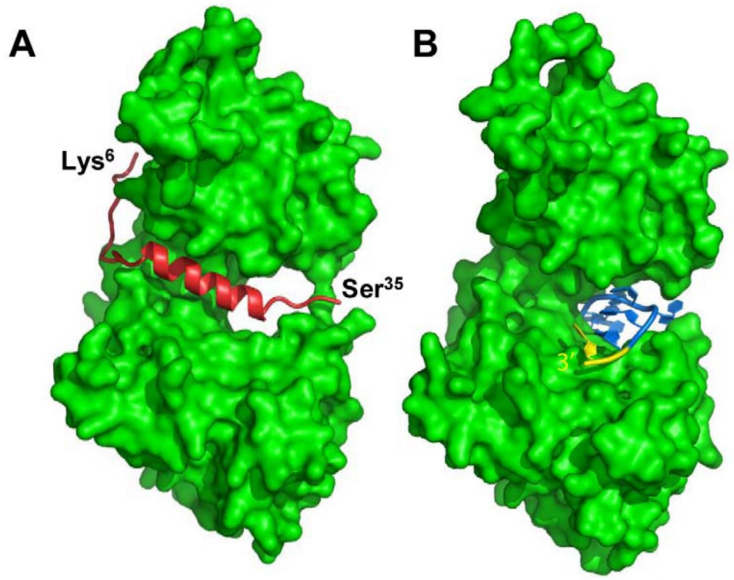


Figure 1.16: VSV N bound to P and RNA.
(A) Surface representation of VSV N (green) bound to a peptide of the N-terminal region of P (red). (B) VSV N bound to RNA (shown in stick representation). Figure taken from Leyrat et al. 2011.

1.3.3 The L Polymerase

The L proteins are the largest single-chain products encoded in the genomes of *Mononegavirales*. These highly flexible and multifunctional proteins often exceed lengths of 2000 residues and their recombinant expression and functional characterization constitutes a considerable challenge. As described above, L utilises the nucleocapsid as its template for transcription (generation of viral mRNAs) and replication (synthesis of antigenomes/progeny genomes). It is assumed that the N

proteins of the nucleocapsid transiently relinquish regions of the bound genomic RNA so that L can utilise these as template (Reguera et al. 2016). In addition to the RdRP function, L proteins add a cap structure to the produced mRNAs, subsequently methylate it, and also possess polyadenylation activity.

In the *Rhabdoviridae* (and, likely also in *Paramyxoviridae* and *Filoviridae*), capping of the 5'-end of the nascent product is initiated by transferring the mRNA to a His residue of the L protein (H1227 in VSV), resulting in the formation of a phosphoamide bond between L and 5'-monophosphorylated mRNA (Li et al. 2008b; Ogino and Banerjee 2007). This reaction is catalysed by a poly-ribonucleotidyltransferase (PRNTase) enzyme, a part of the capping domain of L (see below). Then, the mRNA is coupled to a GDP, followed by methylation of the first ribose at its 2'O via a methyltransferase domain of L. Finally, the cap structure is methylated at N7 (Rahmeh et al. 2009). This is in contrast to capping in eukaryotic cells, where a guanylyltransferase (GTase) transfers a monophosphorylated guanosine onto the nascent RNA. To clarify, in eukaryotic cells a monophosphorylated G is transferred onto a diphosphorylated RNA by a GTase, while in VSV a diphosphorylated G is coupled to a monophosphorylated RNA by the PRNTase (via an intermediate, covalently bound to L). Finally, polyadenylation of the transcript prior to termination of L is achieved by stuttering at a poly-U tract at the gene junctions (Barr and Wertz 2001; Barr et al. 1997).

Owing to the immense hurdles of sample production and high conformational heterogeneity, no structural data of full-length L were available until recently. Finally, one study reported the 3.8 Å resolution cryo-EM single particle reconstruction and chain trace of nearly all residues of full-length VSV L (Liang et al. 2015). This was

possible due to the many advancements in cryo-EM software and hardware, but also because the authors succeeded in rigidifying L by binding it to the VSV $P_{(35-106)}$ fragment (although this was found to be disordered in the resulting structure). In the structure all domains of the 2092 residue protein are assigned, which includes the RdRP, the capping domain, a connector domain, the methyltransferase and a C-terminal domain (**Fig. 1.17**).

The RdRP assumes a canonical right-handed configuration with palm, finger and thumb domains which together form a cavity. The catalytic site is located in the palm domain. The capping domain (which possesses the PRNTase activity) lies on top of the RdRP and includes a priming loop which is important for *de novo* initiation of RNA-synthesis (Reguera et al. 2016). No homologues of known structure could be identified for the capping domain. Next, a connector domain links to the methyltransferase. Unmodelled, low-resolution next to the connector domain was tentatively assigned to linkers 1 and 2 (see **Fig. 1.17A)** and also to the fragment of P (Liang et al. 2015). The domain may thus interact with P and help properly position the other domains of L during their catalytic cycles. The methyltransferase, which transfers methyl groups from the donor S-adenosyl methionine to the nascent mRNA, sits adjacent to the capping and connector domains. Finally, L concludes with a C-terminal domain which was suggested to mainly have an organizational role. Notably, the current structure does not possess clear exit channels and elongation will require considerable rearrangement of the domains to make room for the product (Liang et al. 2015). The VSV-L structure was thus interpreted as an early initiation state conformation. Further structural studies are necessary to understand the rearrangements L domains undergo during its catalytic cycle.

In addition to the VSV L protein, structures of segmented negative strand virus polymerases have also been determined for influenza A, B, and C viruses, and La Crosse orthobunyavirus (LACV) (Pflug et al. 2014; Reich et al. 2014; Hengrung et al. 2015; Gerlach et al. 2015). In the case of influenza, the polymerase complex consists of a heterotrimer, while the LACV polymerase is contained in a single protein chain. The available polymerase structures are compared in detail in Reguera et al. 2016. The most important differences shall be briefly stated here. While the VSV L structure did not contain nucleic acid, the LACV and influenza structures showed how the 5'- and 3'-extremities of the viral template RNA are bound in a sequence-specific manner on the surface of the polymerase (Reguera et al. 2016). It should be noted, that in segmented negative-strand viruses the genome segments form pseudo-circular ribonucleoprotein particles, spatially bringing together the 5'- and 3'-ends (Acheson 2006). Furthermore, the influenza polymerase features cap-binding and cap-snatching endonuclease domains, which are absent in VSV (as VSV replicates in the cytoplasm and produces its own cap). In turn, capping and methyltransferase domains are absent in influenza virus. To summarize, while the core architecture of the RdRPs are similar in all viruses, additional domains related to 5'-cap acquisition (nuclear replicating viruses, e.g. influenza) or synthesis (cytoplasmic replicating viruses, e.g. VSV) represent the major distinguishing features.

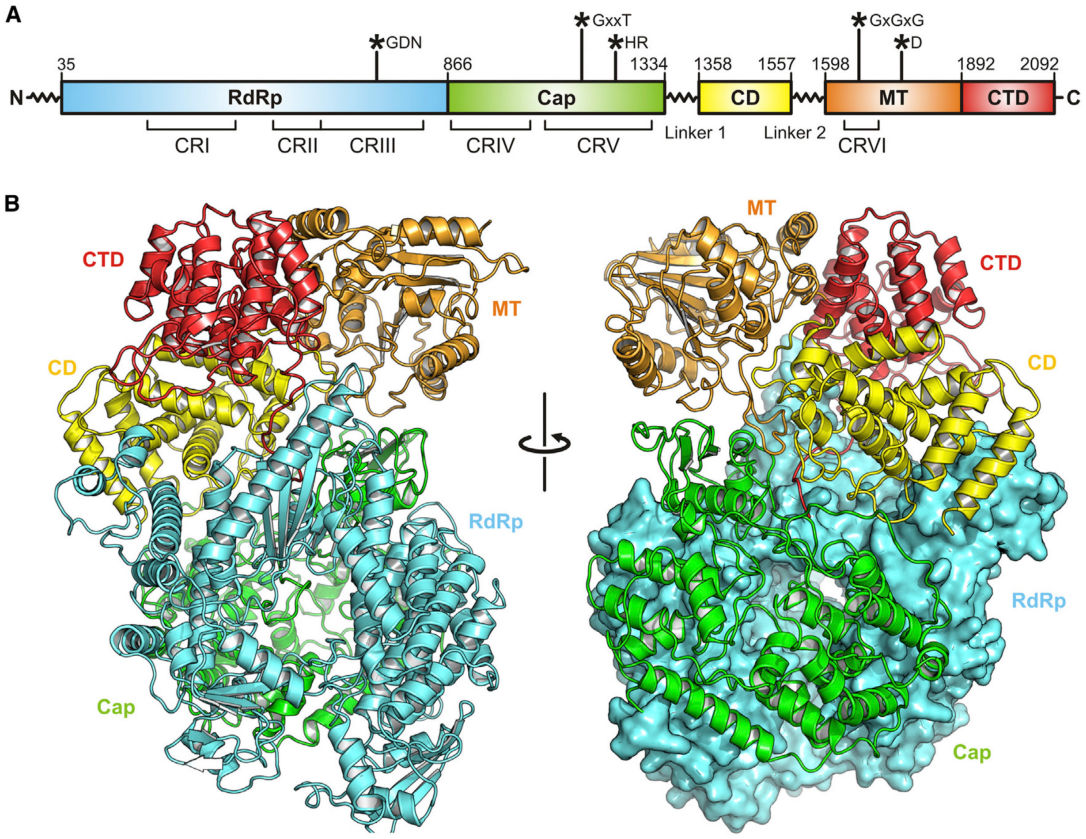


Figure 1.17: Structure of VSV-L by Cryo-EM Single Particle Reconstruction.
(A) Schematic of VSV-L domains. *RdRp*, RNA-dependent RNA polymerase, *Cap*, capping domain, *CD*, connector domain, *MT*, methyltransferase, *CTD*, C-terminal domain. Functional residues are indicated with stars. See original publication for more detailed information (Liang et al. 2015). (B) Overview of the full-length VSV-L structure at 3.8Å as determined by single particle cryo-EM. The functional domains are labelled and coloured as in (A). Figure taken from Liang et al. 2015.

A recent crystallographic and functional study from our own group has elucidated the structure of the methyltransferase domain of HMPV L (Paesen et al. 2015). The domain was shown to have 2'O- and N7-methylation activity *in vitro*. Importantly, the domain did not possess the expected cap binding site and, in addition, showed GTPase activity. This led the authors to suggest that the nascent, L-bound mRNA (covalently coupled to a His of the capping domain) is presented to the methyltransferase through domain rearrangements within L. This is consistent with

the flexible nature of L demonstrated via cryo-EM studies (Rahmeh et al. 2010, 2012). The methyltransferase would then supply the GDP cap through its GTPase activity and perform the methylation reactions.

1.3.4 The M2-1 Transcription Antiterminator

In all members of the *Pneumovirinae* efficient transcription of full-length mRNA requires the small basic protein M2-1. The M2-1 protein is a processivity cofactor of the transcription machinery which enhances synthesis of full-length mRNAs and read-through transcription at gene-junctions (Céspedes et al. 2016). For example, in RSV it has been shown that the absence of M2-1 leads to premature abortion of transcription (Collins et al. 1996). The presence of M2-1 also facilitates the synthesis of polycistronic read-through mRNAs by L, in which transcription does not terminate at gene-end (GE) sequences at the gene junctions (Fearns and Collins 1999; Hardy and Wertz 1998; Sutherland et al. 2001). The antitermination activity of M2-1 has been shown to have varying efficiency depending on the specific gene-junction (Fearns and Collins 1999; Hardy et al. 1999). Although the functional meaning of these polycistronic mRNAs has not been ultimately determined, they have been suggested to facilitate efficient access to genes further towards the 5'-end of the genome, thereby benefiting the virus by enabling fine-tuning of gene product levels (Cowton et al. 2006).

The RSV M2-1 protein has been shown to be a RNA-binding protein and it systematically prefers binding to A-rich, positive-sense viral RNA transcripts (Blondot et al. 2012; Cuesta et al. 2000). It has been suggested that M2-1 binds to nascent mRNAs and stabilises these, perhaps by inhibiting the formation of RNA

secondary structure (Blondot et al. 2012). M2-1 also interacts with P and is thus thought to be recruited to sites of viral transcription (Derdowski et al. 2008; Khattar et al. 2001; Tran et al. 2009). Interestingly, binding of P and RNA are competitive and the binding regions have been reported to partially overlap (Blondot et al. 2012; Tran et al. 2009).

M2-1 is a tetrameric protein that contains a CCCH-type zinc-finger domain, an oligomerisation domain, and a globular core domain (Blondot et al. 2012). Complex formation with P proceeds at a stoichiometry of one P tetramer to one M2-1 tetramer and results in an extended, non-globular dimer of tetramers (Esperante et al. 2012). Both P and RNA have been reported to interact with the globular core domain of M2-1 (Blondot et al. 2012). Interestingly, the core domain of M2-1 shows significant levels of homology to VP30 of *Filoviridae*, a protein which also binds RNA, contains a CCCH zinc-finger, and is important for transcription (John et al. 2007; Modrof et al. 2003; Mühlberger et al. 1999). This underlines the relatedness between filoviruses and the *Pneumovirinae* branch of paramyxoviruses as is often observed in phylogenetic trees constructed using sequences of N or L (see also **Fig. 1.1**). At the time my doctoral project was initiated the crystallographic structure of full-length HMPV M2-1 had just been solved in our laboratory by Dr Cedric Leyrat. Unexpectedly, the structure showed an asymmetric tetramer in which the core domain of one protomer is partially dissociated (**Fig. 1.18**). This conformational flexibility and the paucity of structural information regarding RNA-binding prompted an analysis of nucleotide recognition by M2-1 and its solution behaviour.

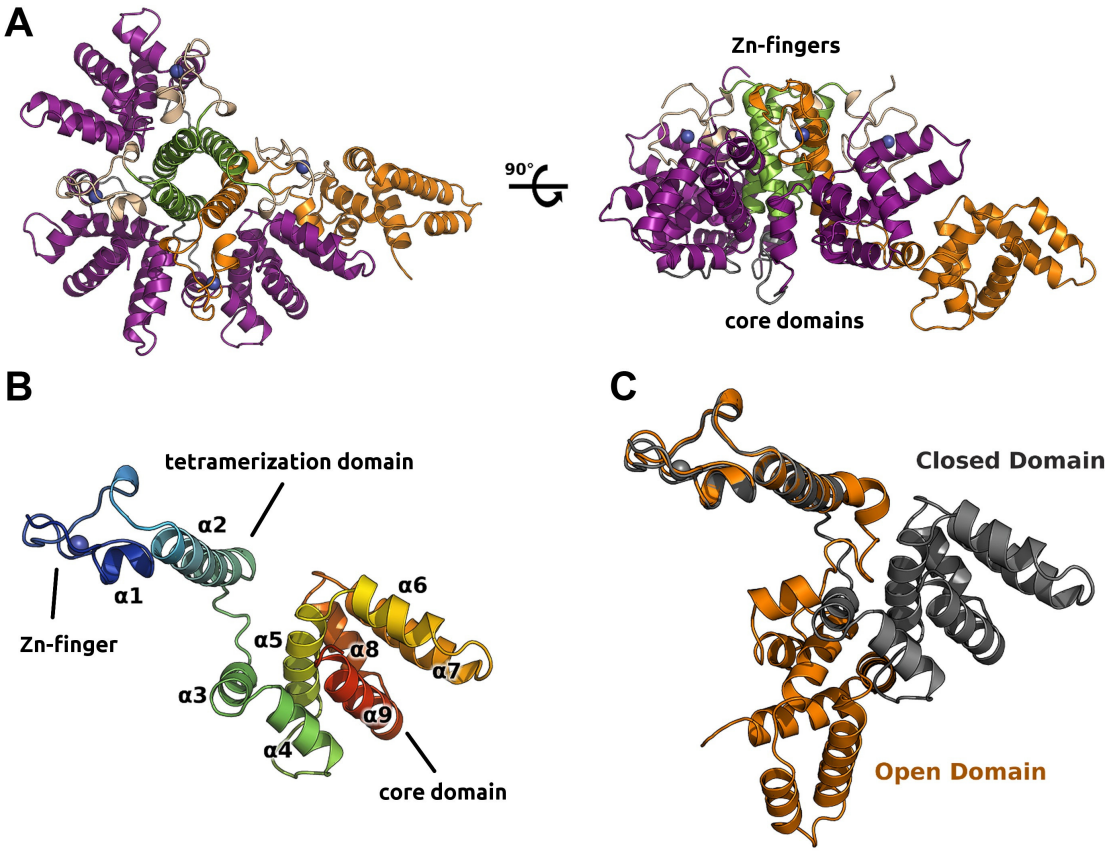


Figure 1.18: The Antitermination Factor M2-1 from HMPV.
(A) Overview of the crystallographic structure of the M2-1 antiterminator from HMPV. The structure is coloured according to its domains with the zinc finger in wheat, the oligomerisation helix in green, and the globular core domain in purple. Zinc atoms are coloured in blue. In the structure the core domain of one protomer is dissociated from the tetramer (indicated in orange), suggesting considerable inherent flexibility. (B) One monomer of M2-1 is depicted coloured from N-terminus to C-terminus (blue to red, respectively). The secondary structure elements are indicated. (C) Alignment of the disassociated domain (open, orange) with a non-disassociated domain (closed, grey). The comparison demonstrates how the core domain can swivel about the loop connecting α2 and α3 (see panel (B)). The figure is adapted from a published manuscript containing work carried out as part of this thesis (Leyrat et al. 2014b).

1.4 Project Aim

The background information provided in the previous sections should illustrate the complex interplay of the multiple components of the transcription and and replication machinery in HMPV and other single-stranded negative-sense viruses. Conformational flexibility, folding-upon-binding, and domain rearrangement are reoccurring themes among the players of these processes, making their study challenging. HMPV is a highly contagious human pathogen, which has been discovered only recently and has hardly been characterised structurally. For these reasons and, additionally, because of its relatedness to RSV, HMPV is an interesting target for further study.

At the time this project was initiated there was a great paucity of high-resolution structural information on pneumoviral transcription/replication components. Of special interest to me was the P protein, a key central organiser of these components and a challenging topic of study due to its high flexibility. No crystallographic structures were available for any region of P from *Pneumovirinae* and neither was there a structural basis for how P might chaperone the essential nucleocapsid building-block N in any *Paramyxoviridae*. A key question was how RNA-bound N (N-RNA) and P-bound N (N^0-P) might differ conformationally, and how N^0-P is transferred onto nascent nucleocapsids. For the reasons stated above the main objective in this work was the structural, comparative study of N-RNA and N^0-P (Chapter 2). In addition to this, I also worked on two side-projects which were concerned with the study of the solution behaviour and RNA-binding capabilities of M2-1 (Chapter 3), and also the structural characterization of the P oligomerisation domain and the M2-1-interacting domain (Chapter 4).

2 HMPV N

As discussed in the previous chapter, the nucleoproteins N of *Mononegavirales* are the principal components of the nucleocapsids and protect the viral genetic material from nuclease digestion. During replication, fresh N^0 is delivered to the polymerase by the P protein in form of a RNA-free N^0-P complex. The following chapter is concerned with the structural analysis of RNA-bound assembled N and RNA-free monomeric N^0-P from HMPV (strain NL1-00, A1).

2.1 RNA-Free N^0-P Complex

2.1.1 Construct Design Rationale

To facilitate the structural analysis of HMPV N^0-P, the development of a expression/purification strategy which would supply sufficient amounts of homogeneous material was required. The reconstitution of a monomeric N^0-P complex has proven to be challenging in the past, as recombinant expression of N proteins typically leads to irreversible oligomerisation and RNA-uptake (e.g. Albertini et al. 2006; Green et al. 2006; Tawar et al. 2009). A technique which has seen success previously with VSV is expressing N in which the NTD-arm (which facilitates N oligomerisation) has been truncated (Leyrat et al. 2011) and a similar strategy was attempted early in this project. Unfortunately, it was not possible to express HMPV N in which similar truncations had been introduced (these constructs did not show measurable expression, perhaps due to misfolding and immediate degradation in the expression host). Furthermore, co-expression of P fragments with

N did not lead to the formation of a N^0-P complex either. Instead, a strategy was followed in which naturally occurring binding is enhanced by the covalent fusion of the interacting partners in the form of a chimeric protein. Such approaches have seen great success in various systems (reviewed in Reddy Chichili et al. 2013), especially in the study of MHC-receptor complexes. The specific rationale in my case was, that by using a chimeric construct, the N-chaperoning region of P would be in direct vicinity of N at the point of protein translation. The hypothesis was that N would thus form a complex with P even before the irreversible oligomerisation and uptake of RNA could occur. Perhaps such an immediate, co-translational N chaperoning by P might be similar to what occurs in a physiological context.

Mononegavirales P proteins have been suggested to share a similar architecture (see **Fig. 1.12**) with a central oligomerisation domain, a C-terminal nucleocapsid-interacting domain, and a transiently folded N^0-chaperoning domain at the N-terminus. Although the N-terminal regions are little conserved, careful sequence analyses performed previously have revealed the presence of short semi-conserved sequence motifs (Karlin and Belshaw 2012). To design a chimeric N^0-P construct, disorder and secondary structure predictions of HMPV N and P were performed. Because different prediction servers use a variety of premises and algorithms to detect disordered regions, results are often inconsistent, complicating construct design for the molecular biologist. A powerful approach to mitigate these problems, is the use of prediction metaservers which call on a variety of available servers to calculate a consensus disorder profile, improving reliability and accuracy (Peng and Kurgan 2012). The consensus disorder predictor DisMeta (Huang et al. 2014) was utilised to generate profiles for HMPV P and N (**Fig. 2.1A** and **B**). Among

other predictors, DisMeta makes use of RONN, which calculates likelihoods of disorder based on alignments of sequences with known foldedness (Yang et al. 2005).

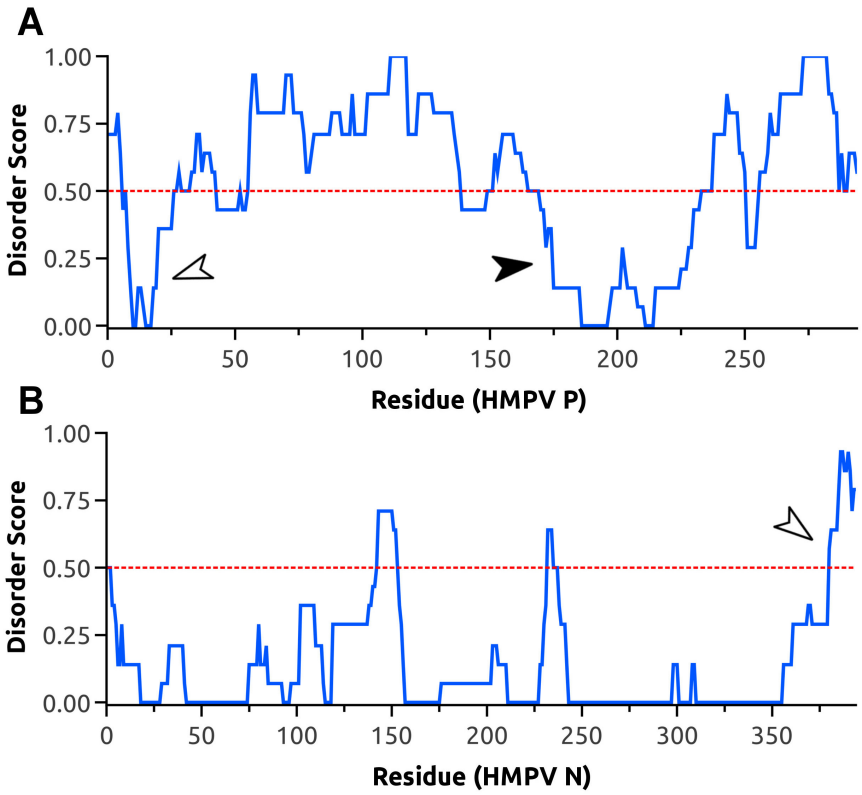


Figure 2.1: Disorder and Domain Analysis of HMPV N and P.
(A) and (B), consensus disorder profiles of HMPV P and N, respectively. The disorder values were calculated with the DisMeta server (Huang et al. 2014), with high values corresponding to a high probability of the given residue to be intrinsically disordered. The arrows in the panels are further discussed in the text.

The consensus profile of HMPV P (**Fig. 2.1A**) predicted the mostly disordered nature of the protein, consistent with the literature. The largest ordered region detectable in the profile is centred around residue 200 and most likely constitutes the oligomerisation domain of P (indicated with black arrow, **Fig. 2.1A**). A small ordered region of 25-30 residues was also detectable at the N-terminus of the protein and this was tentatively assigned as the transiently folded N^0-binding domain (indicated with

white arrow, **Fig. 2.1A**). The boundaries were chosen liberally, utilising the first 40 residues of P for the chimeric construct to make sure the hypothesised functional domain was not truncated. Furthermore, no additional linker residues were introduced to connect this region to N, but instead it was simply fused to the nucleoprotein C-terminus. The C-terminal region of N terminates with a predicted stretch of flexible and disordered residues (indicated with white arrow, **Fig. 2.1B**) which could be co-opted as a "natural" linker. Generation of the final construct thus was a simple matter of fusing the first 40 residues of P to the C-terminus of N, which was carried out via overlapping-primer PCR, followed by cloning into a pOPINE vector. The resulting $N_{1\text{-}394}$-$P_{1\text{-}40}$ hybrid (termed HMPV N^0-P for simplicity from here on) possessed a molecular weight of ~48 kDa.

2.1.2 Expression and Purification of the N^0-P Chimera

In the course of this project I was involved in the supervision of Master's student Mattia Bertinelli, who provided valuable assistance with the expression, purification and crystallographic screening of HMPV N^0-P. N^0-P was expressed in *E. coli* and purified via standard methods as described in Materials and Methods. Briefly, the chimeric protein (see **Fig. 2.2A** for construct details) was affinity-purified via a C-terminal $(His)_6$-tag, then an additional purification step over a Heparin column was carried out, followed by a final polishing step via size-exclusion chromatography (SEC). The purification was carried out in high-salt conditions (1 M NaCl) to ensure that there is no residual binding to RNA and, in addition, benzonase was added to the lysis buffer. The NaCl concentration was reduced to 100 mM in the final stages of the purification. The chromatogram of the final purification step

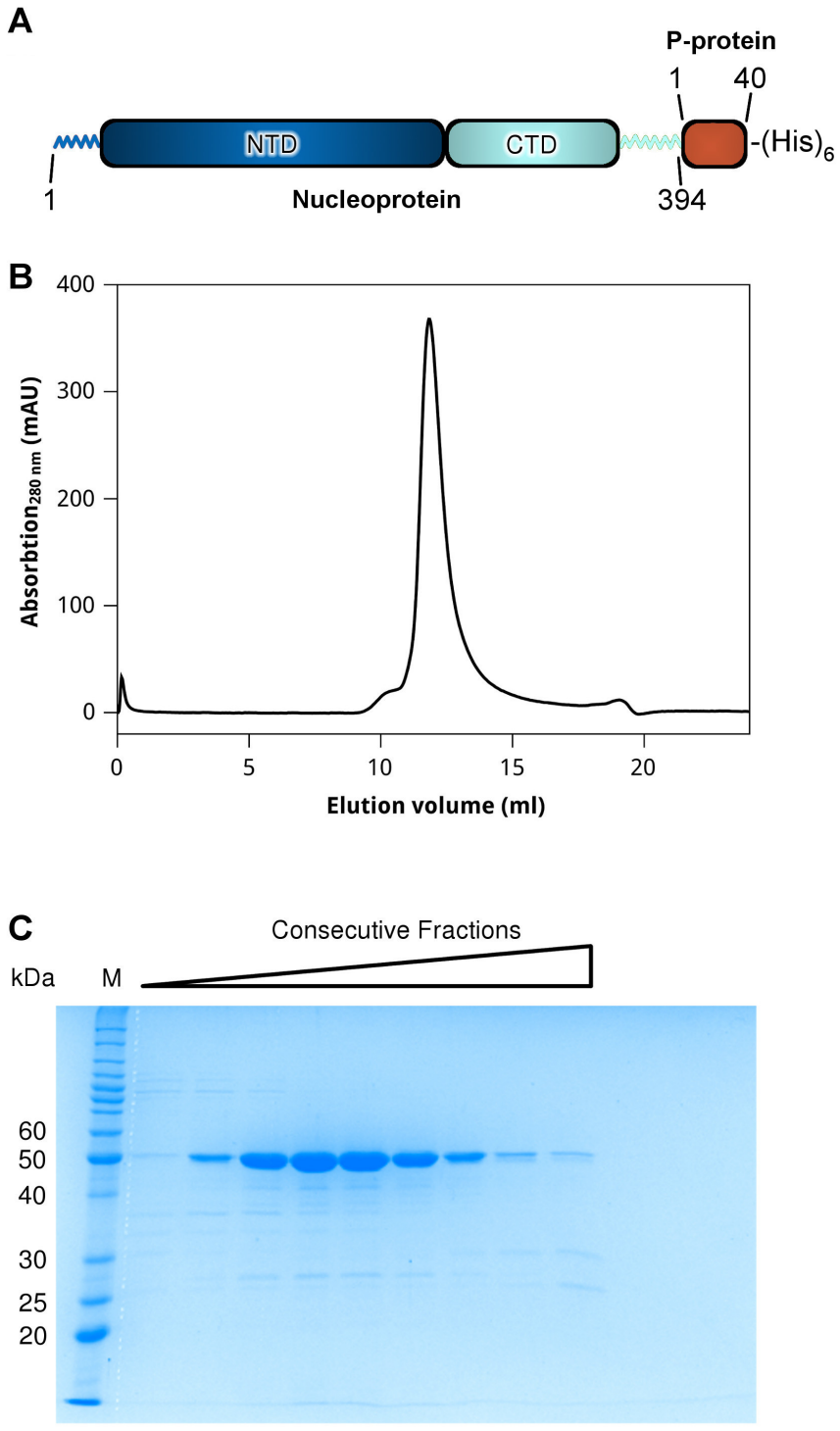


Figure 2.2: Purification of HMPV N^0-P.
(A) N^0-P chimeric construct designed for this work. The NTD, CTD, and corresponding exchange subdomains are indicated. The first 40 residues of P (orange) are fused to the C-terminus of N (blue). (B) Size-exclusion chromatography (SEC) profile of N^0-P (Superdex 75). (C) SDS-PAGE analysis of the N^0-P fractions obtained from SEC. Figure modified from Renner et al. 2016.

indicated a high level of monodispersity (**Fig. 2.2B**) and the peak was centred around an elution volume expected for a N^0-P monomer of the corresponding molecular weight.

2.1.3 Crystallisation, Structure Solution and Refinement

The purified HMPV N^0-P was concentrated to ~7 mg/ml using a centrifugal concentration device. The concentrated sample was subjected to crystallisation trials as described in Materials and Methods. Initially, no crystals were obtained. The construct of N^0-P included the NTD- and CTD-arms of N, which in absence of N-oligomerisation are expected to be disordered. Given that these flexible regions sample much conformational space it is likely that they impinge on crystallisation attempts. However, truncation of the construct was not an option as this would inhibit recombinant expression. Instead, limited proteolytic digestion was carried out to prune flexible regions in the hope of promoting crystal growth. Specifically, the method of *in situ* proteolysis was chosen, in which minute traces of protease are added just before crystallisation trials are set-up (Dong et al. 2007). In two systematic studies, *in situ* proteolysis led to the successful structure determination of 14.5% and 12.6% of targets, respectively, that had previously proven recalcitrant to crystallisation attempts (Dong et al. 2007; Wernimont and Edwards 2009). This method has proven to be somewhat mild and can result in protein that is nearly intact, save for cleaved loops. Indeed, a previous study has utilised a very similar approach, in which a Chikungunya virus chimeric p62–E1 fusion protein was expressed and subsequently cleaved with trypsin (Voss et al. 2010).

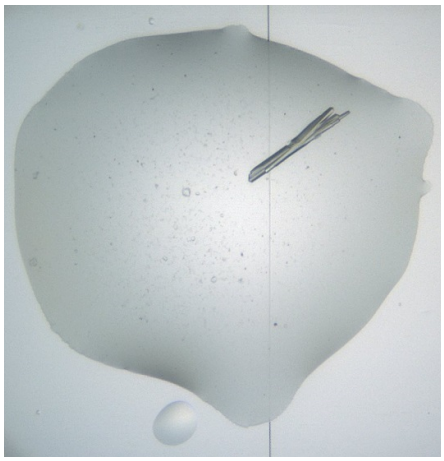


Figure 2.3: Crystal of HMPV N^0-P.
Crystal of HMPV N^0-P obtained after in situ proteolysis and additive screening (see accompanying text).

To facilitate *in situ* digestion of HMPV N^0-P, the sample was supplemented with 1 µg trypsin per 1000 µg of protein prior to crystallisation screening. Initial crystals with a thin, rod-shaped morphology were obtained overnight at 20°C with the reservoir condition 100 mM PCB System, pH 7, 25% polyethylene glycol (PEG) 1500. Using this condition as a basis additive screening was carried out with the Hampton Silver Bullets Screen (McPherson and Cudney 2006). Larger crystals (see **Fig. 2.3**, long rods, ~400 µm in length and ~40 µm thick) grew with the additive E11 of this screen (9 mM 1,2-diaminocyclohexane sulphate, 6 mM diloxanide furoate, 17 mM fumaric acid, 10 mM spermine, 9 mM sulphaguanidine and 20 mM HEPES, pH 6.8). For cryo-protection the crystals were immersed in reservoir solution supplemented with 25% glycerol. Subsequently, crystals were cooled to liquid nitrogen temperature. Diffraction data were collected at 100 K at Diamond Light Source, Didcot, UK, beamline I04. Diffraction data up to 1.9 Å were auto-processed with XIA2 (Winter et al. 2013). The structure was solved by molecular replacement (MR) using PHASER (McCoy et al. 2007) using a monomer of RSV N-RNA (pdbID: 2wj8) after deletion of the bound RNA. The structure was refined via iterative rounds of manual building in COOT (Emsley and Cowtan 2004) and atomic position, B-factor and TLS refinement with PHENIX (Adams et al. 2010). Two-fold torsion-angle non-crystallographic symmetry (NCS) restraints (there are

two copies of N^0-P in the asymmetric unit) as implemented in PHENIX were used during refinement (Adams et al. 2010). The geometry of the final model was validated with MOLPROBITY (Chen et al. 2010). Data collection and refinement statistics are given in Table 1.

Data Collection	
Wavelength (Å)	0.979
Number of images	3600
Ω Oscillation per image (°)	0.05
Exposure time per image (s)	0.05
Beam size (μm)	54x40
Transmission (%)	100
Space group	P 1
Cell dimensions	
a, b, c (Å)	40.9, 62.8, 86.7
α, β, γ (°)	91.0, 96.4, 109.0
Resolution (Å)	28.42-1.86 (1.91-1.86)
CC (1/2)	1.00 (0.47)
R_{merge}	0.055 (0.590)
I / σI	9.2 (1.1)
Completeness (%)	94.8 (75.0)
Redundancy	1.7 (1.6)
Refinement	
Resolution (Å)	28.42-1.86
No. reflections	64451 (3743)
R_{work} / R_{free}	17.1/20.5
No. atoms	
Protein	5707
Non-protein	588
B-factors	
Protein	34.5
Non-protein	42.6
RMSD	
Bond lengths (Å)	0.007
Bond angles (°)	1.000
Ramachandran plot quality	
Favoured (%)	99.72
Allowed (%)	0.28
Outliers (%)	0.00

Table 1: Data Collection and Refinement Statistics for HMPV N^0-P.
Data were collected at Diamond Light Source, Didcot, UK, beamline I04, using a Pilatus 6M-F detector.
Values in parentheses correspond to the highest resolution shell
RMSD, root mean square deviation from ideal geometry, CC, cross-correlation
R_{free} was calculated for a 5% subset of reflections

2.1.4 Structure of the HMPV N^0-P Complex

As described above the construct used to obtain the crystal constituted residues N_{1-394}-P_{1-40}. The asymmetric unit of the crystal contained two copies of the HMPV N^0-P complex. No RNA is found in the structure and the nucleoprotein is in an unassembled, non-oligomerised state (biological assembly was determined with PISA, Krissinel 2010). Fortuitously, the structure of N^0-P is very complete and showed only minimal amounts of digestion via the supplemented trypsin (**Fig. 2.5A**). Residues that were pruned are mostly located at the N-terminal and C-terminal ends. In the structure nucleoprotein residues 30-383 (of 394) are resolved, with an internal loop from residue 101 to 111 being disordered. In addition, it should be noted that the residues linking up the N protein with the P protein have also been cleaved, leaving the P protein bound to N exclusively through non-covalent interactions. This is supported by SDS-PAGE analysis of the trypsin-digested N^0-P chimera (**Fig. 2.4**). The molecular weight of the digested protein band (black arrow in **Fig. 2.4**) is only compatible with a cleaved linker between N and P (which would leave N at ~37.5 kDa vs. ~42 kDa for an uncleaved but disordered linker). For the P protein residues 1 to 28 (of originally 40) are visible in the crystal and bound to N.

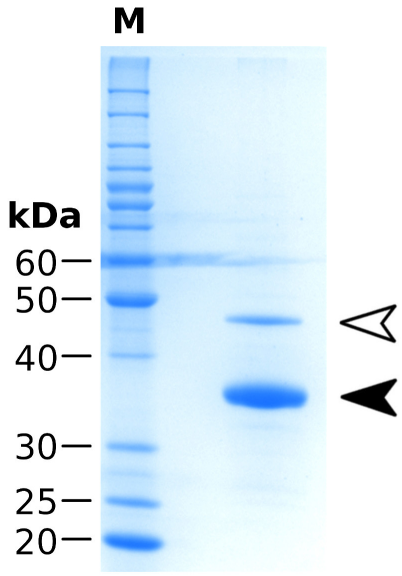


Figure 2.4: SDS-PAGE Analysis of Trypsin-Digested N^0-P.
The HMPV N^0-P chimera was subjected to limited proteolysis with trypsin and subsequently analysed by SDS-PAGE. The white arrow indicates residual uncleaved N^0-P. The black arrow indicates the degradation product at roughly 37 kDa.

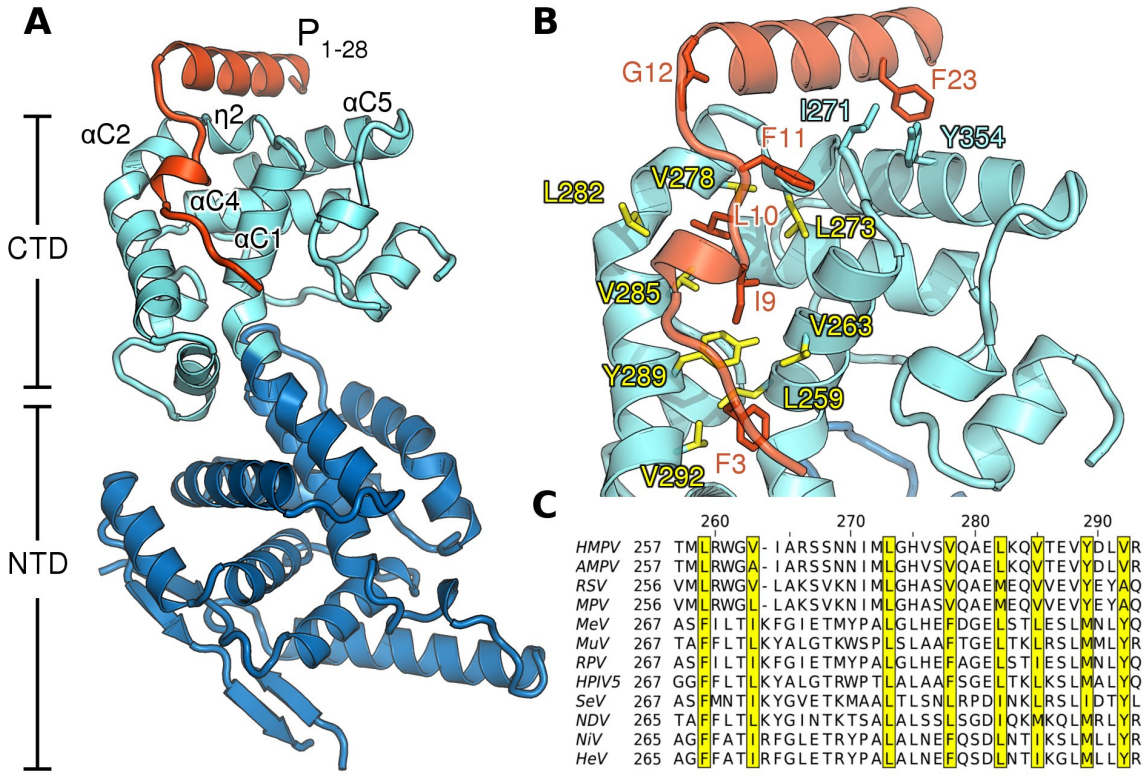


Figure 2.5: The RNA-free N^0-P complex from HMPV.
(A) Overview of the N^0-P structure in cartoon representation. The CTD (light blue) and NTD (dark blue) are indicated. The P peptide (orange) sits on top of the CTD. Secondary structure elements of N which form the interacting surface are indicated. (B) Close-up view of the P binding region. Key interacting residues of the P-peptide and the N binding surface are shown as sticks. Residues of the conserved hydrophobic P-binding surface of N are coloured in yellow. (C) The multiple sequence alignment (MSA) of *Paramyxoviridae* members showing residues belonging to the hydrophobic P-binding site. Conserved hydrophobics are shown in yellow. Figure adapted from Renner et al. 2016.

Like other nucleoproteins from mononegaviruses HMPV N consists of two globular α-helical bundle domains, the NTD and CTD, which together form a molecule with an overall bean-like shape (**Fig. 2.5A**). The structure reveals how the N-terminal region of P tightly attaches to and wraps around the CTD of the nucleoprotein. P interacts with N predominantly via hydrophobic residues, with the interface area between the two chains being roughly ~1030 $Å^2$. Helices αC1 and αC2 of N form a extensive hydrophobic surface, into which P residues Ile9, Ile10 and

Phe11 burrow, pinning down the P peptide (**Fig. 2.5B**). The P peptide then continues with a helical stretch starting from Gly12. This helix is fastened to the top of the CTD, again mainly via hydrophobic interactions. A noteworthy side-to-face interaction between Phe23 of P and Tyr354 of N plays a key role in pinning down the helix (**Fig. 2.5B**). The Phe residue of P is conserved in this position for all *Pneumovirinae* and is part of a sequence motif previously denoted as the *mir* motif (see Karlin and Belshaw 2012). In addition, a further study utilising a minireplicon system has demonstrated the importance of this Phe by substituting it with an alanine, causing a drop of polymerase activity by 75% (Galloux et al. 2015).

A multiple sequence alignment (MSA) of paramyxovirus N residues was performed using the sequences of αC1 and αC2 which form the bulk of the P-binding surface (**Fig. 2.5C**). The alignment suggests that the hydrophobic character of this surface is conserved throughout the viral family, pointing to a common P-recognition site for these viruses. Since initiating the work on the HMPV N^0-P complex, a flurry of new nucleoprotein structures from other mononegaviruses have been reported. This includes structures of RNA-free nucleoproteins bound to polymerase co-factors for Nipah virus (*Paramyxovirinae*, Yabukarski et al. 2014) and Ebola virus (*Filoviridae*, Leung et al. 2015).

In the case of the Ebola virus structure reported by Leung et al., expression screening of Ebola nucleoprotein (NP) truncation constructs identified one construct ($NP_{25\text{-}457}$ of 739 total residues) which showed a monodisperse peak in gelfiltration in addition to a peak in the void volume of the column (Leung et al. 2015). This partially non-oligomeric species was likely obtained by inadvertently truncating the N-terminal arm of NP, which is involved in oligomerisation. Using this monodisperse

sub-population, Leung et al. could reconstitute, *in vitro*, a complex of Ebola NP with the polymerase cofactor (VP35 in Ebola) peptide $VP35_{20-48}$ by mixing these two reagents in a 1:1 ratio. The reconstituted complex could then be crystallised. This strategy is in effect very similar to the one used by Leyrat et al. for the determination of the VSV N^0-P complex crystal structure. Here, the N-terminal arm (N residues 1-21) of VSV nucleoprotein was also truncated in the expression construct and the N^0-P complex reconstituted afterwards using a short peptide of the polymerase cofactor encompassing residues P_{1-60} (Leyrat et al. 2011). It should also be noted, that for both the Ebola virus and VSV work nucleoprotein truncation constructs were expressed as maltose binding protein (MBP) conjugated fusion proteins (Leyrat et al. 2011; Leung et al. 2015). In the case of the structure determination of Nipah virus N^0-P, both the N-terminal and C-terminal arms of N were truncated in the construct (Yabukarski et al. 2014). This core of Nipah N encompassed residues N_{32-383}. Again, the N^0-P complex was then reconstituted *in vitro* by mixing the truncated nucleoprotein with a P peptide (P_{1-50}), followed by crystallisation trials.

Comparison of the N^0-P forms shows that in all known complexes of this type the polymerase cofactor is bound to the CTD of the nucleoprotein (**Fig. 2.6**). Indeed, even when the directionality of P-binding varies (e.g. when comparing VSV with Nipah, indicated by arrows in **Fig. 2.6**), the region of this binding surface described above seems to be universal for all *Mononegavirales* (region is indicated with dotted circle in **Fig. 2.6**).

Not unexpectedly, a similarity search of HMPV N^0-P with the DALI server (Holm and Rosenström 2010) identified the RNA-bound form of N from RSV as closest homologue of known structure (RMSD 2.0 Å, sequence identities of

pneumoviral N proteins are shown in Table 2). The next closest N proteins were identified to be human parainfluenza virus 5 (RMSD 3.5 Å) and measles virus (RMSD 3.8 Å), both also RNA-bound forms. I will present a direct comparison of the RNA-bound and P-bound states of HMPV N later in this chapter.

HMPV	100%				
AMPV	88.32%	100%			
MPV	45.03%	43%	100%		
BRSV	42.45%	40.66%	60.1%	100%	
HRSV	42.45%	40.4%	60.61%	93.35%	100%
	HMPV	**AMPV**	**MPV**	**BRSV**	**HRSV**

Table 2: Sequence Identities of *Pneumovirinae* N Proteins.
Matrix showing sequence identities of N proteins from *Pneumovirinae*. Nucleoproteins for which structures have been determined are HMPV (this work, P-bound and RNA-bound forms) and HRSV (RNA-bound form only). Virus abbreviations: *HMPV*, human metapneumovirus, *AMPV*, avian metapneumovirus, *MPV*, murine pneumonia virus, *BRSV*, bovine respiratory syncytial virus, *HRSV*, human respiratory syncytial virus.

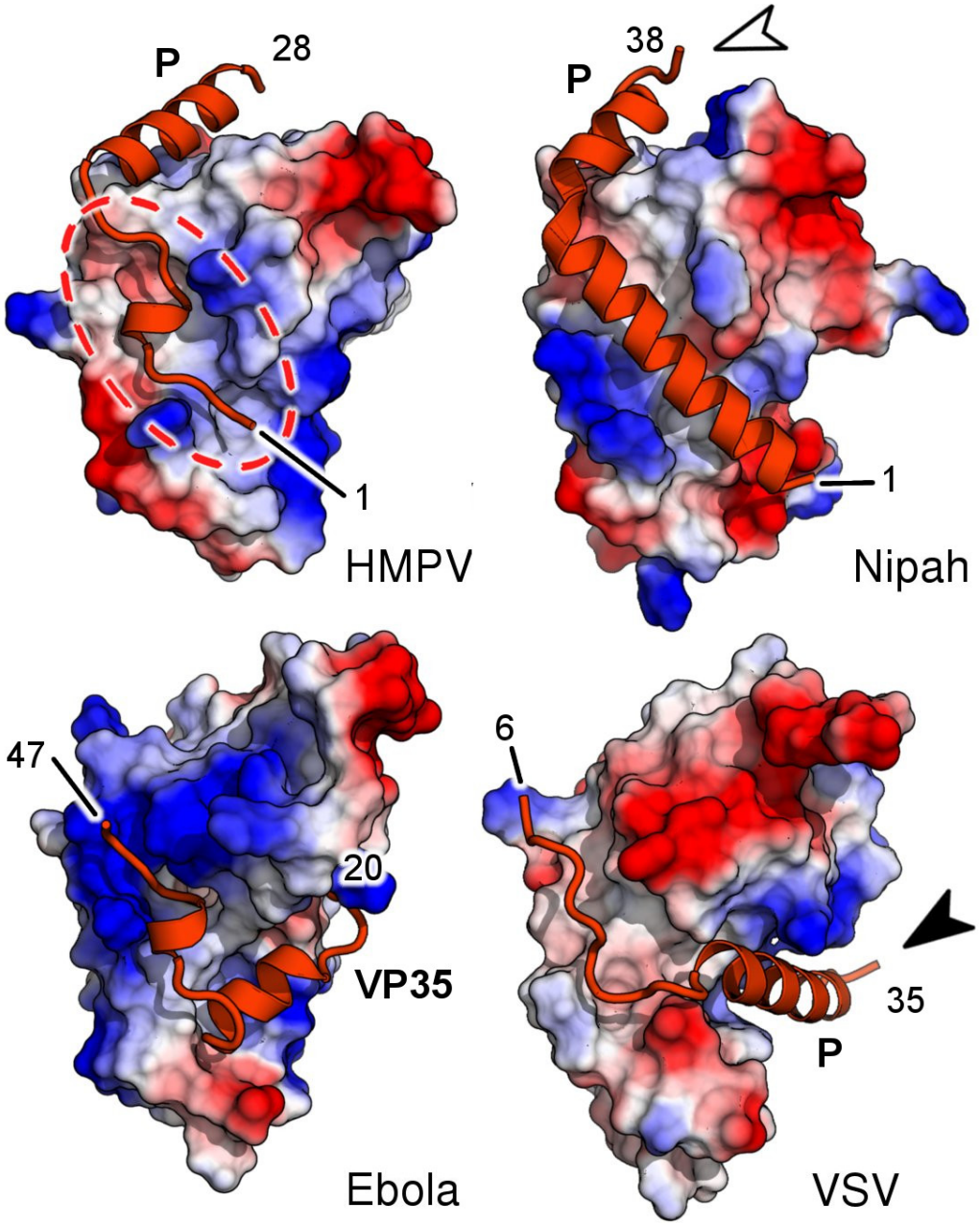


Figure 2.6: Mononegavirus Nucleoprotein-Phosphoprotein Complexes.
The CTDs of nucleoproteins bound to polymerase cofactors are shown in the same orientations. Surface representations of CTDs are coloured according to electrostatics from blue (positive) to red (negative). Polymerase cofactors are shown in cartoon representation and coloured in orange. The cofactors seem to wrap around the CTDs in two different modes, either sitting on top of the CTD (e.g. in Nipah virus, white arrow) or under it (e.g. VSV, black arrow). Notably, the directionality of P is reversed in the case of VSV. A binding surface on the lateral side of the CTD seems to be shared in all structures, however (indicated with dotted circle for HMPV). The pdb codes of the structures are: HMPV, 5fvd, Nipah, 4co6, Ebola, 4ypi, VSV, 3pmk. Figure adapted from Renner et al. 2016.

2.2 Assembled N-RNA

In parallel to the work on the N^0-P complex of HMPV, the determination of the structure of an assembled, RNA-bound form of N was also attempted. The rationale was that a direct residue-by-residue comparison of these two states for the same virus (without resorting to homologues of related viruses) would provide detailed information about conformational switching on a molecular level. To this end a structural characterization of HMPV N-RNA was carried out, which is described in the following section.

2.2.1 Initial Characterization of N-RNA from Different Expression Hosts

Previous studies have reported that recombinant expression of nucleoproteins in insect cells leads to the formation of nucleocapsid-like helical filaments of varying flexibility (Bhella et al. 2002; Ruigrok et al. 2011a). Recombinant nucleocapsids of *Paramyxovirinae* have been shown to rigidify after formaldehyde-fixing in buffer conditions containing high concentrations of salt and truncations of the N-tail region considerably stabilise a straight, rod-like conformation (Desfosses et al. 2011; Heggeness et al. 1980; Jensen et al. 2011; Heggeness et al. 1981). Indeed, this effect has facilitated the near-atomic resolution helical reconstruction from cryo-EM of measles virus nucleocapsids, which has been reported recently (Gutsche et al. 2015).

In the case of *Pneumovirinae,* N proteins do not possess comparable N-tail regions and recombinant nucleocapsids from RSV have proven to be highly flexible, rendering high-resolution reconstruction challenging (Bakker et al. 2013; Tawar et al. 2009). To assess the morphology of HMPV recombinant nucleocapsids, N-RNA was

purified from insect cells and the sample was imaged via negative-stain EM. Similar to what has been reported for RSV and not unexpectedly the helices proved to be conformationally heterogeneous (**Fig. 2.7**). This was also true if the sample was fixed (performed as described in Heggeness et al. 1980) in the presence of high salt concentrations (1 M NaCl) (**Fig. 2.8**). Although no detailed analysis was carried out (which is likely to be error-prone considering the sub-optimal quality of staining in the case of the data collected in low-salt), visual comparison of filaments in low-salt buffer with those fixed in high-salt buffer did not indicate a significant straightening or compaction of the helices.

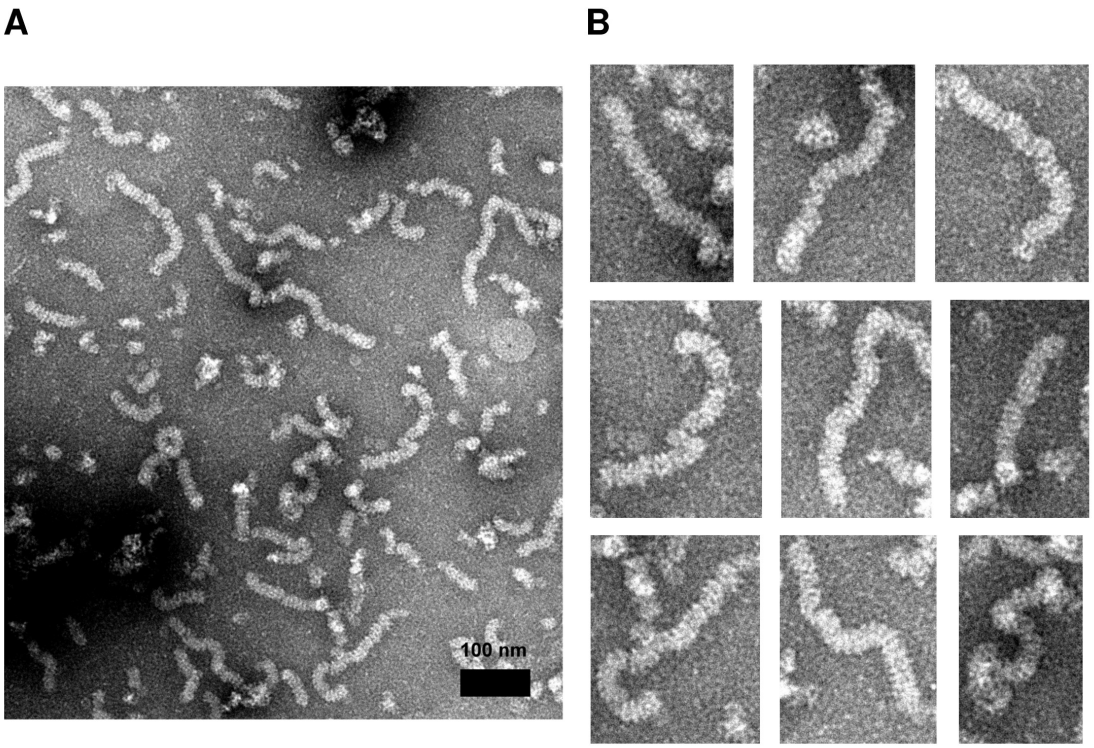


Figure 2.7: Recombinant Nucleocapsid-Like Filaments in Low-Salt Buffer.
(A) Negative-stain transmission EM image of recombinant HMPV N purified from insect cells in buffer containing 0.15 M NaCl. (B) Panel of close-up views of the recombinant filaments, emphasizing their inherent flexibility.

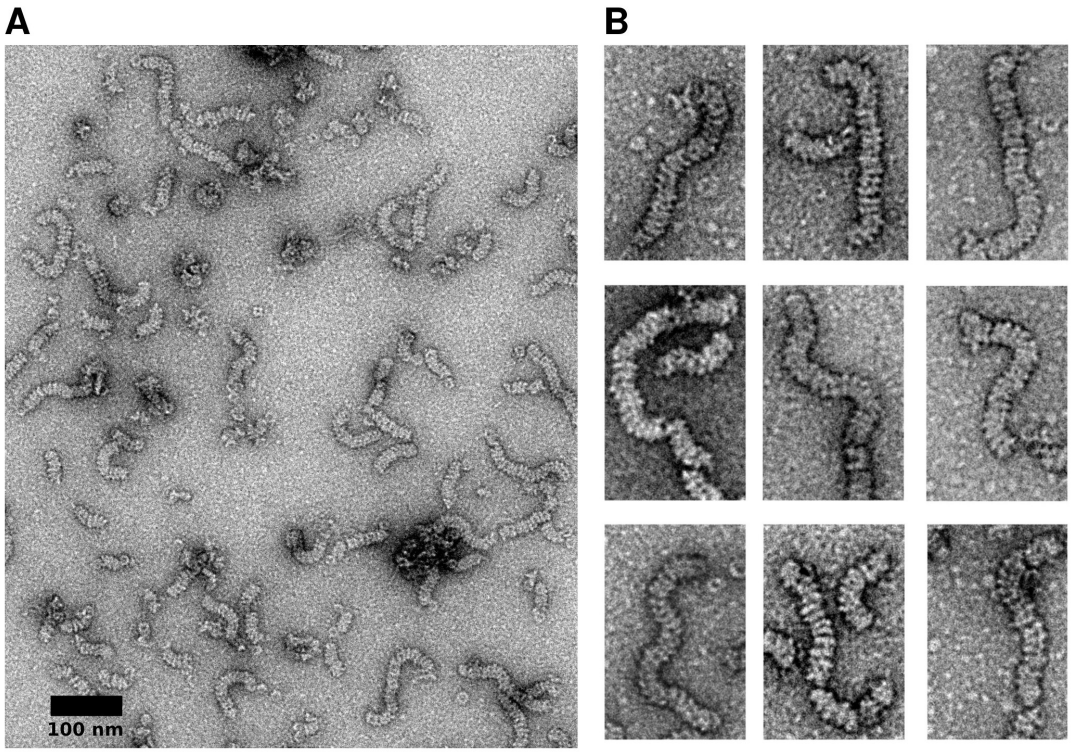


Figure 2.8: Recombinant Nucleocapsid-Like Filaments in High-Salt Buffer.
(A) Negative-stain transmission EM image of recombinant HMPV N purified from insect cells and formaldehyde-fixed in buffer containing 1 M NaCl. (B) Panel of close-up views of the recombinant filaments, emphasizing their inherent flexibility.

As the nucleocapsid-filaments proved to be too flexible for high-resolution structure determination to be feasible, I decided to go down a crystallographic route. Previous studies on other mononegavirus nucleoproteins showed that recombinant expression of N in bacteria leads to the formation of crystallisable, RNA-bound oligomeric rings (Alayyoubi et al. 2015; Albertini et al. 2006; Green et al. 2006; Tawar et al. 2009). HMPV N was expressed in *E. coli* and purified via standard methods, utilising a C-terminal $(His)_6$-tag (**Fig. 2.9**). The sample formed large assemblies and was bound to nucleic acids as judged by the elution volume and 260/280 nm absorption ratio in the gelfiltration chromatogram (**Fig. 2.9A**). Cryo-EM micrographs showed heterogeneous mixtures of ring-shaped assemblies (**Fig. 2.9C**).

2D-class averages revealed populations of rings with varying numbers of subunits, based on the number of radial projections. Specifically, nonameric, decameric and undecameric rings could be observed, with the decamers being in the majority (~70%) (**Fig. 2.9D**). This is in line with a previous study with RSV N in which decameric and undecameric rings could be detected in a ratio of 2/3 to 1/3 (Tran et al. 2007). No nonameric population was reported, however. Due to the lack of side views in the cryo-EM projection images, single particle reconstruction was not attempted. Such a preference for certain orientations is not an uncommon issue in single-particle cryo-EM (especially with ring-shaped objects) and is presumably caused by interactions with the air/water interface (Cheng et al. 2015). Screening of detergent additives, glow-discharging parameters and grid types is often carried out to overcome problems associated with preferred orientations of particles.

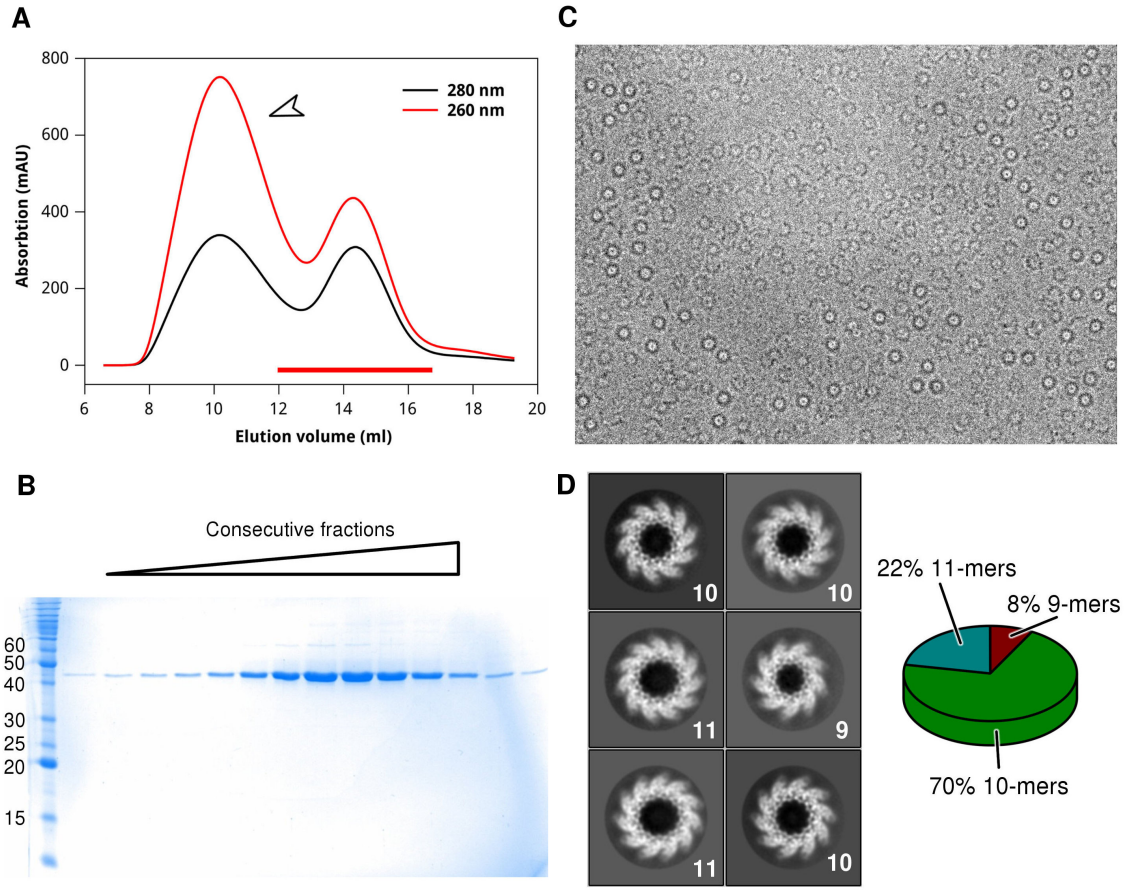


Figure 2.9: HMPV N Recombinantly Expressed in Bacteria.
(A) HMPV N-RNA was expressed in *E. coli* and subsequently purified via standard methods. A polishing step via size exclusion chromatography using a Superose 6 column is shown. Fractions containing the recombinant protein are marked with a red bar. The large peak eluting at around 10 ml (white arrow) does not contain protein and most likely is caused by bacterial nucleic acids which have been co-purified from the expression host (evidenced by the 260/280 ratio). (B) Marked fractions from (A) were separated via SDS-PAGE (C) Cryo-EM image of N-RNA purified from *E. coli*. The micrograph shows the formation of oligomeric rings. (D) 2D-classification of N-RNA rings demonstrated different populations of oligomeric states: rings containing either 9, 10, or 11 subunits (see white labels). The distribution of these populations is indicated in the pie chart. Figure adapted from Renner et al. 2016.

2.2.2 Crystallisation, Structure Solution and Refinement

To determine the structure of HMPV N-RNA crystallographic screening of the assembled form of N expressed in bacteria was carried out. Solubility of the N rings proved to be a major hurdle, with precipitation occurring at concentrations above 1 mg/ml. The solubility of the complex could be substantially improved by screening various buffer additives and testing if these improved the behaviour of the sample when concentrating. Specifically, inclusion of either the solubilising agents NDSB201 (at 500 mM) or NDSB256 (at 100 mM) and arginine (at 50-100 mM) allowed concentrations of ~4 mg/ml to be reached, which was sufficient to initiate crystallisation screening. These additives were not visible in the crystal structure, however.

Growing diffraction-quality crystals proved challenging with this sample and extensive screening needed to be carried out. Crystal hits of diamond-like morphology were obtained in a variety of conditions containing PEGs of varying molecular weights. Crystallogenesis was unreliable, however, and seemed to be dependent on the batch of protein. The MORPHEUS (Gorrec 2009) crystallisation screen produced the largest crystals and numerous screens of this type were set up using different batches of protein in different buffer conditions.

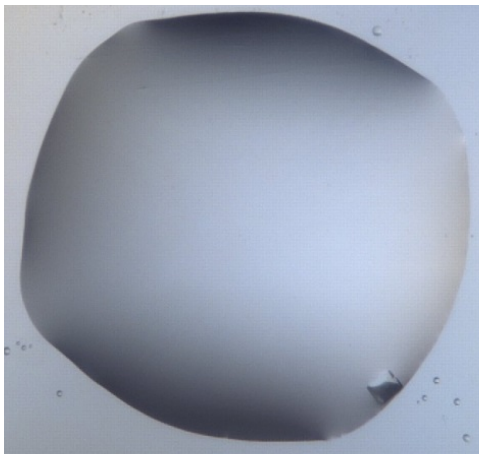


Figure 2.10: Crystal of HMPV N-RNA.
Crystallisation drop containing a single N-RNA crystal which grew after 132 days. The corresponding reservoir condition was C12 of the MORPHEUS screen (see Gorrec 2009).

The diffraction quality of the crystals varied greatly with no obvious discernible pattern with respect to batch, buffer conditions, crystal size or reservoir condition. Unfortunately, the majority of crystals showed only low-resolution diffraction (~20 Å) and very few gave rise to spots up to 6-8 Å. Crystallographic additive screens, crystal dehydration, and *in situ* proteolysis attempts were made but did not systematically improve the quality of the crystals. After screening roughly 300 crystals, a single well-diffracting specimen of small size (**Fig. 2.10**, approximate crystal size ~60 μm) was identified which had appeared after 132 days in mother liquor containing 100 mM Tris/Bicine, pH 8.5, 90 mM NPS ($NaN0_3$, Na_2HPO_4, $(NH_{4)2}SO_4$), 37.5% methyl-2 4-pentanediol (MPD), polyethylene glycol 1000 and polyethylene glycol 3350 (MORPHEUS crystal screen).

As the reservoir condition contained MPD, the crystal was cryo-cooled without addition of extra cryo-protectant. Diffraction data were collected at 100 K at Diamond Light Source, Didcot, UK, beamline I04-1. Diffraction data up to 4.17 Å were auto-processed with XIA2 (Winter et al. 2013, 2). The structure was solved by molecular replacement (MR) with PHASER (McCoy et al. 2007) using a decameric model of the previously determined HMPV N structure. This model was generated using the RSV decamer as a guide (Tawar et al. 2009). The structure was initially refined via iterative rounds of manual building in COOT (Emsley and Cowtan 2004) and atomic position and B-factor refinement with PHENIX (Adams et al. 2010, 201). Refinement was carried out using 10-fold NCS constraints (the 10 copies of N in the ASU were treated to be identical) as implemented in PHENIX to lower the parameter to observations ratio. To facilitate accurate placement of side-chains at this resolution, interactive B-factor sharpening in COOT (using a negative B-factor of around -150

Å^2) (Emsley and Cowtan 2004) and 10-fold NCS-averaged maps generated with PHENIX RESOLVE (Adams et al. 2010) and Parrot of the CCP4 suite (Winn et al. 2011) were utilised.

To further improve the Ramachandran quality of the model, the final rounds of refinement were carried out with BUSTER (Smart et al. 2012). These made use of TLS parameters, the high-resolution N^{0}-P structure (to generate reference model restraints), and 10-fold Local Structure Similarity Restraints (LSSR) to facilitate the use of NCS information as implemented in BUSTER (Smart et al. 2012). The geometry of the final model was validated with MOLPROBITY (Chen et al. 2010). Data collection and refinement statistics are given in Table 3. A sample of electron density is shown in **Fig. 2.11**.

Data Collection	
Wavelength (Å)	0.917
Number of images	3600
Ω Oscillation per image (°)	0.1
Exposure time per image (s)	0.1
Beam size (µm)	30x30
Transmission (%)	100
Space group	C 2 2 2_1
Cell dimensions	
a, b, c (Å)	202.0, 233.2, 203.6
α, β, γ (°)	90, 90, 90
Resolution (Å)	101.19-4.17 (4.28-4.17)
CC (1/2)	1.00 (0.38)
R_{merge}	0.220 (2.924)
I / σI	9.2 (1.0)
Completeness (%)	99.9 (100)
Redundancy	13.5 (13.8)
Refinement	
Resolution (Å)	101.19-4.17
No. reflections	36125 (2617)
R_{work} / R_{free}	19.1/23.0
No. atoms	
Protein	27957
Non-protein	1400
B-factors	
Protein	216.1
Non-protein	215.1
RMSD	
Bond lengths (Å)	0.010
Bond angles (°)	1.120
Ramachandran plot quality	
Favoured (%)	95.01
Allowed (%)	4.96
Outliers (%)	0.03

Table 3: Data Collection and Refinement Statistics for HMPV N-RNA.
Data were collected at Diamond Light Source, Didcot, UK, beamline I04-1, using a Pilatus 2M detector.
For use of NCS information during refinement see accompanying text.
Values in parentheses correspond to the highest resolution shell
RMSD, root mean square deviation from ideal geometry, CC, cross-correlation
R_{free} was calculated for a 5% subset of reflections

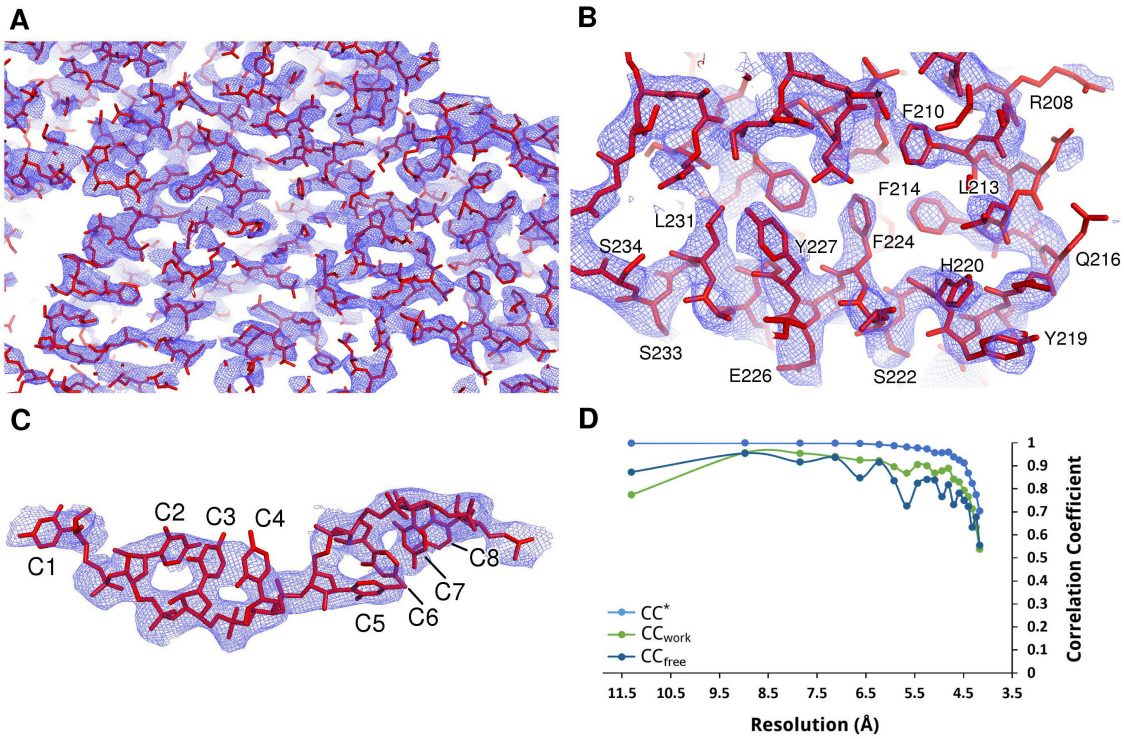


Figure 2.11: HMPV N-RNA Electron Density Maps and Model Validation.
(A-C) Maps after density modification with Parrot and B-factor sharpening shown at a contour level of 1.0 σ. Panel (A) shows an overview of the map, panel (B) a close-up view highlighting the side-chain fitting, and panel (C) the density for the RNA strand. (D) Plot of the correlation of the true signal CC*, CC_{free} and CC_{work}. CC* not dropping below CC_{work} or CC_{free} for any resolution demonstrates that the model is not overfitting the data (Karplus and Diederichs 2012). Figure adapted from Renner et al. 2016.

2.2.3 Structure of HMPV N-RNA

The structure of the RNA-bound and assembled nucleoprotein ring from HMPV is consistent with other structures of this type (Alayyoubi et al. 2015; Albertini et al. 2006; Green et al. 2006; Tawar et al. 2009). Residues 2 to 383 of HMPV N could be modelled into the density, with the exception of a disordered loop consisting of residues 99 to 112 (this region was also disordered in the N^0-P structure). In the N-RNA decamer the RNA wraps around the outer surface of the ring in a belt-like fashion (**Fig. 2.12A**). Oligomerisation of neighbouring nucleoproteins is

facilitated via the CTD- and NTD-arms. The whole complex has a diameter of ~160 Å, a height of 70 Å and a mass of roughly ~0.5 MDa. Like in the related RSV (Tawar et al. 2009), the RNA bases are sequestered in a positively charged cleft (**Fig. 2.12B**), with each N protein subunit contacting seven nucleotides (**Fig. 2.12C**). The nucleotides follow a "3-bases-in, 4-bases-out" topology as in RSV (Tawar et al. 2009). The "3-bases-in" correspond to nucleotides 2-3-4 while the outward-facing bases correspond to 5-6-7-1 (with base 1 being the first nucleotide of the neighbouring N subunit) (see **Fig. 2.12C**). In contrast, members of the *Paramyxovirinae* subfamily follow the so-called rule of six, requiring their genome lengths to be exact multiples of this number (Calain and Roux 1993). This is because the groove harbouring the "4-bases-out" in pneumoviruses is shorter in the case of *Paramyxovirinae* and can thus only accommodate 3 bases, leading to a "3-bases-in, 3-bases-out" configuration (Gutsche et al. 2015).

It is interesting to note that the region of the RNA-bound HMPV N between residues 99 to 114 appears disordered, as was the case with the P-bound form, whilst in RSV it forms the tip of a well-folded β-hairpin which protrudes from the surface of N (Tawar et al. 2009). These residues are also significantly less conserved between RSV and HMPV than the rest of the protein (< 20% identity for residues 90-150 vs. 42% for the whole protein). It is tempting to speculate that the β-hairpin in RSV serves an as of yet undetermined function which is not shared in HMPV. In line with this observation, amino acid substitutions within the β-hairpin of RSV have been identified as escape mutations for the N-inhibitor RSV604 (Chapman et al. 2007). The same inhibitor was shown to not be efficacious against HMPV (Chapman et al. 2007).

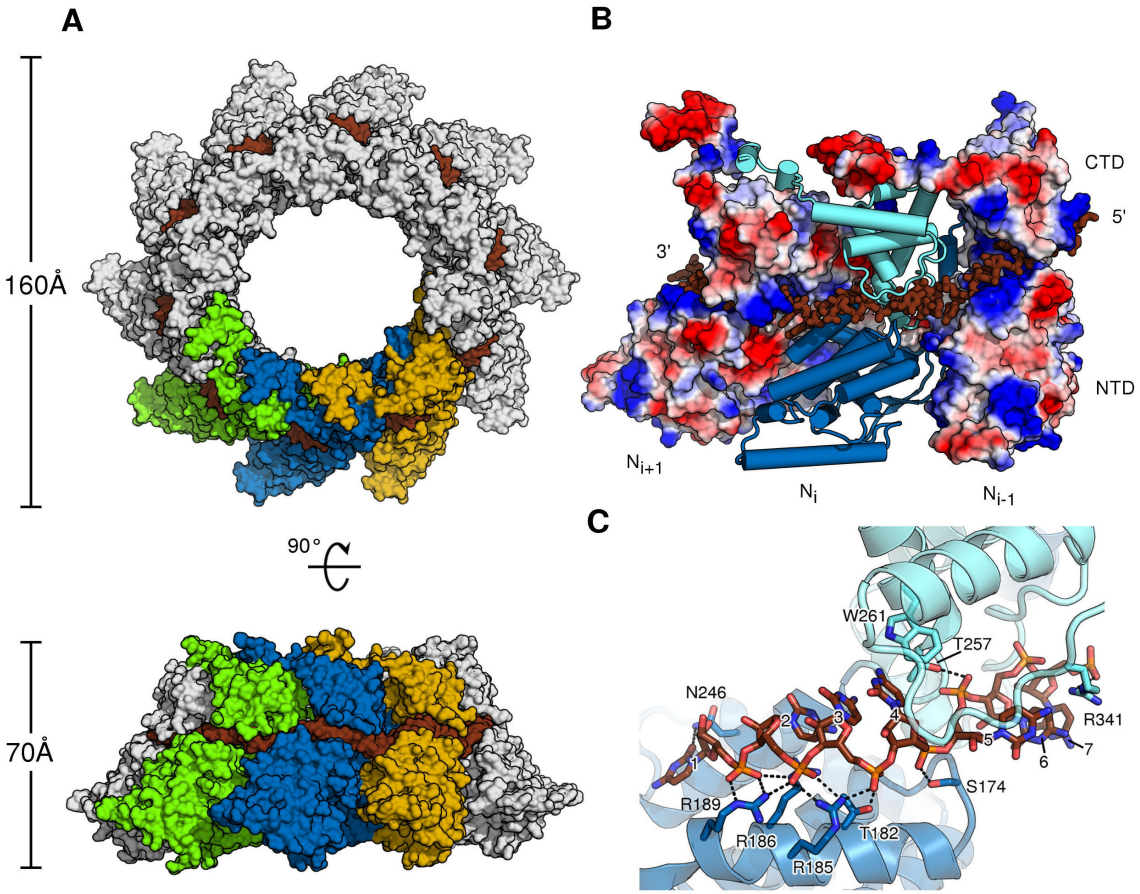


Figure 2.12: HMPV N-RNA.
(A) HMPV RNA-bound N rings as shown from the top and the side in surface representation. Three consecutive N subunits are coloured individually to highlight their boundaries within the ring. The RNA-belt is coloured in brown. (B) Tilted side view of three consecutive N subunits with inserted RNA. The two outer subunits are coloured according to their electrostatic surface charge to highlight the positively charged RNA-binding cavity. In the middle copy of N the CTD is coloured in light blue, while the NTD is shown in dark blue (indicated). RNA is shown as brown sticks. (C) RNA inserted into the binding-groove of HMPV N. The seven nucleotides are numbered as in comparable structures (Gutsche et al. 2015; Tawar et al. 2009). RNA and key interacting residues are shown as sticks. Figure adapted from Renner et al. 2016.

2.3 Comparison of N^0-P and N-RNA Crystal Structures

As is generally true for mononegavirus N proteins (Alayyoubi et al. 2015; Albertini et al. 2006; Green et al. 2006; Tawar et al. 2009), the NTD- and CTD-arms attach to the adjacent N subunits, thus facilitating lateral N-oligomerisation (**Fig. 2.13A**). The NTD-arm inserts into a groove of the previous N subunit (**Fig. 2.13A**, NTD-arm of N_{i+1} attaches to N_i). On the opposite side, the CTD-arm fastens onto the top of the CTD of the consecutive N (**Fig. 2.13A**, CTD-arm of N_{i-1} attaches to CTD of N_i). Importantly, in the structure of the P-bound N^0, the location of the P peptide coincides with the NTD- and CTD-arm attachment sites (**Fig. 2.13B**), rendering binding of P and N-oligomerisation mutually exclusive. The structures thus provide a direct and unambiguous picture of how P hampers subdomain exchange between neighbouring protomers in *Pneumovirinae*. In the course of this work this mechanism has also been suggested for other mononegaviruses (Alayyoubi et al. 2015; Guryanov et al. 2016; Yabukarski et al. 2014) and there is increasing support for the notion that it is shared within the entire viral order. Furthermore, one of these studies has suggested that N alternates between a rigidified open conformation which is bound to P and a closed conformation which is bound to RNA (Yabukarski et al. 2014).

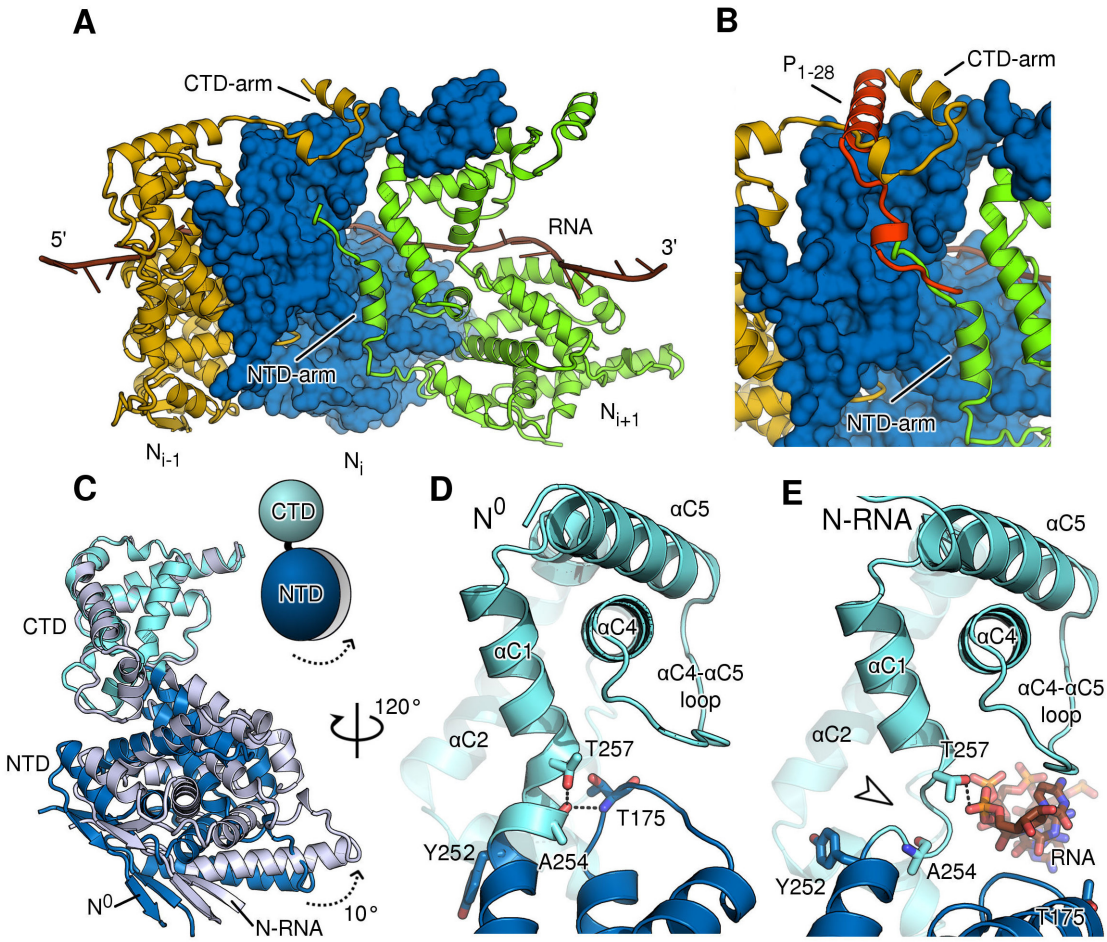


Figure 2.13: Comparison of HMPV N-RNA and N^0-P.
(A) Three assembled subunits of N-RNA are shown. The middle copy is depicted as surface and coloured in blue, while the outer protomers are shown in cartoon representation. The NTD- and CTD-arms facilitating the oligomerisation are labelled. The bound RNA is shown in brown. (B) Close-up of the middle protomer from (A) with the binding site of P overlaid. The attachment sites of the NTD- and CTD-arms overlap with the P-peptide (coloured in orange), causing major clashes. (C) Alignment of the cores of RNA-bound N (coloured in silver) and RNA-free N (coloured in blue) showing the hinge-like motion between NTD and CTD. (D) and (E) show close-up views of the central helix αC1 which connects the NTD and CTD. The helix partially unfolds upon RNA-binding (white arrow in E), thus enabling the hinge-motion. RNA and key residues are shown as sticks. Figure adapted from Renner et al. 2016.

Comparison of the global conformations of the structures of these two forms of N (N^0-P and N-RNA) for HMPV suggested a rigid body motion of the NTD and CTD relative to each other (**Fig. 2.13C**). This motion pivots the NTD towards the CTD by 10°, which is a less dramatic conformational change than has been proposed

for measles and Nipah virus (Guryanov et al. 2016; Yabukarski et al. 2014). The helix αC1 bridges the NTD and CTD of the nucleoprotein and may act as a hinge (**Fig. 2.13D**). A closer examination of helix αC1 in the RNA-free form reveals that residues Thr257 and Ala254 stabilise and maintain an α-helical conformation and interact with Thr175 in a close-by loop (**Fig. 2.13D**). In contrast, when RNA inserts into the binding-groove, the flap containing Thr175 moves downwards to accommodate the strand of nucleic acid and αC1 partially unravels, now acting as a flexible hinge (white arrow, **Fig. 2.13D**). In addition, Thr257 now contacts the RNA backbone instead of stabilizing αC1 as a helix. In this way hinge-like domain motions may be coupled to the binding of RNA by N.

It is also interesting to note that a highly conserved Tyr (Tyr252 in HMPV) is located just before this hinge (**Fig. 2.13D and E**). This Tyr tightly packs against helix αC3 in HMPV and an aromatic can be found in the same position in most other mononegavirus nucleoproteins (highlighted in red in **Fig. 2.14A-G**). The transition between RNA-free and RNA-bound states of HMPV N is accompanied by a dramatic shift of αC3 (located in the CTD) which pushes Tyr252 upwards (**Fig. 2.14A**). I propose that the binding or RNA induces a small-scale motion in αC3 of the CTD that is transmitted to the NTD via this Tyr (which is located just before the hinge, but on the NTD-side). Intriguingly, in N structures of HPIV5 and measles virus the aromatic side chain is flipped to the other side (**Fig. 2.14H and I**) and directly binds RNA (Gutsche et al. 2015). which may point towards a similar coupling of domain pivoting and RNA-uptake.

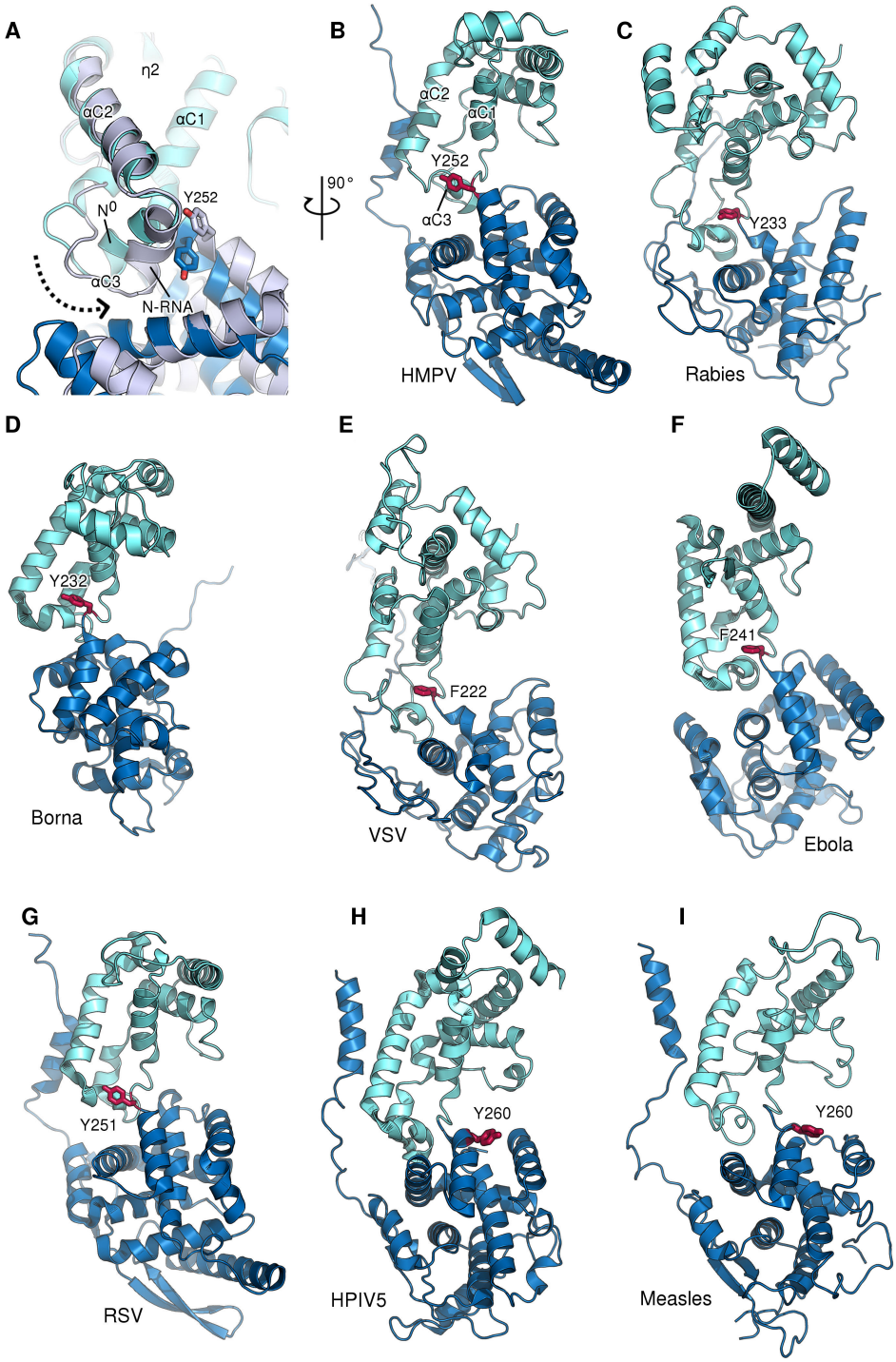


Figure 2.14: A Conserved Aromatic Residue Within the Hinge.
(A) Superposition of RNA-free N (blue) and RNA-bound N (silver) highlighting the motion (indicated by dotted arrow) of αC3 and the coupled Tyr252. (B-H) Nucleoproteins of mononegaviruses. The aromatic residues at the hinge are shown in stick representation and are coloured in red. The NTDs are coloured in dark blue and the CTDs in light blue. Figure taken from Renner et al. 2016.

Furthermore, a marked difference between the N-RNA and N^0-P states could be observed in the CTD-arm. The CTD-arm in HMPV roughly extends from residues 360-394, with good electron density being visible up to residue 383 in the N^0-P structure. In the assembled, N-RNA form the CTD-arm is in an "upwards" position (**Fig. 2.15A**), where it latches on to the adjacent protomer of N, facilitating assembly. In contrast, in the monomeric state and in absence of RNA, the CTD-arm folds downwards, against the core of N and reaches into the RNA-binding cleft (compare **Figs. 2.15B and 2.15C**). Structural alignments reveal that the downwards-positioned CTD-arm would hinder the path of the RNA strand along the binding cleft (**Fig. 2.15A**). In addition, the CTD-arm somewhat resembles a wedge, with the tip of the wedge exposing a cluster of negative charges (Asp373/Glu374/Glu375) into the RNA-binding groove (**Fig. 2.15A**), potentially repulsing RNA. The negatively charged residues are highly conserved in pneumoviruses and can also be found in several members of *Paramyxovirinae* (**Fig. 2.15D**). These structural data implicate a role of the CTD-arm in preventing premature RNA-uptake by N. This markedly contrasts with *Rhabdoviridae*, in which the P-protein itself is deeply inserted into the RNA binding groove, thus blocking binding of nucleic acids (Leyrat et al. 2011).

The interacting surface area between the CTD-arm and the core is 575 $Å^2$ which is a value typical for peptide-protein complexes (London et al. 2010). The interface includes many hydrophobic interactions (theoretical solvation free energy gain on formation of the interface Δ^iG = -4.9 kcal/mol, P = 0.349, Krissinel 2010) with Ile364, Phe366, Leu369, and Leu379 of the CTD-arm tightly packing against the CTD-core. Furthermore, prominent salt bridge pairs (Arg260/E375 and Arg186/E374) secure the CTD-arm in the downward position (**Fig. 2.15B**). One salt bridge between

E374 and R186 links the RNA-binding site with the CTD-arm. It is interesting to note that the residues Trp261, R186 and R260 are involved in RNA-binding and oligomerisation in the N-RNA state, whilst in N^0-P they contact the CTD-arm instead (compare **Figs. 2.15B and 2.15C**). This suggests that the transition into the upwards position of the CTD-arm (and thereby attachment to another N protomer and oligomerisation) are directly coupled to RNA uptake and that these processes may occur in a concerted fashion.

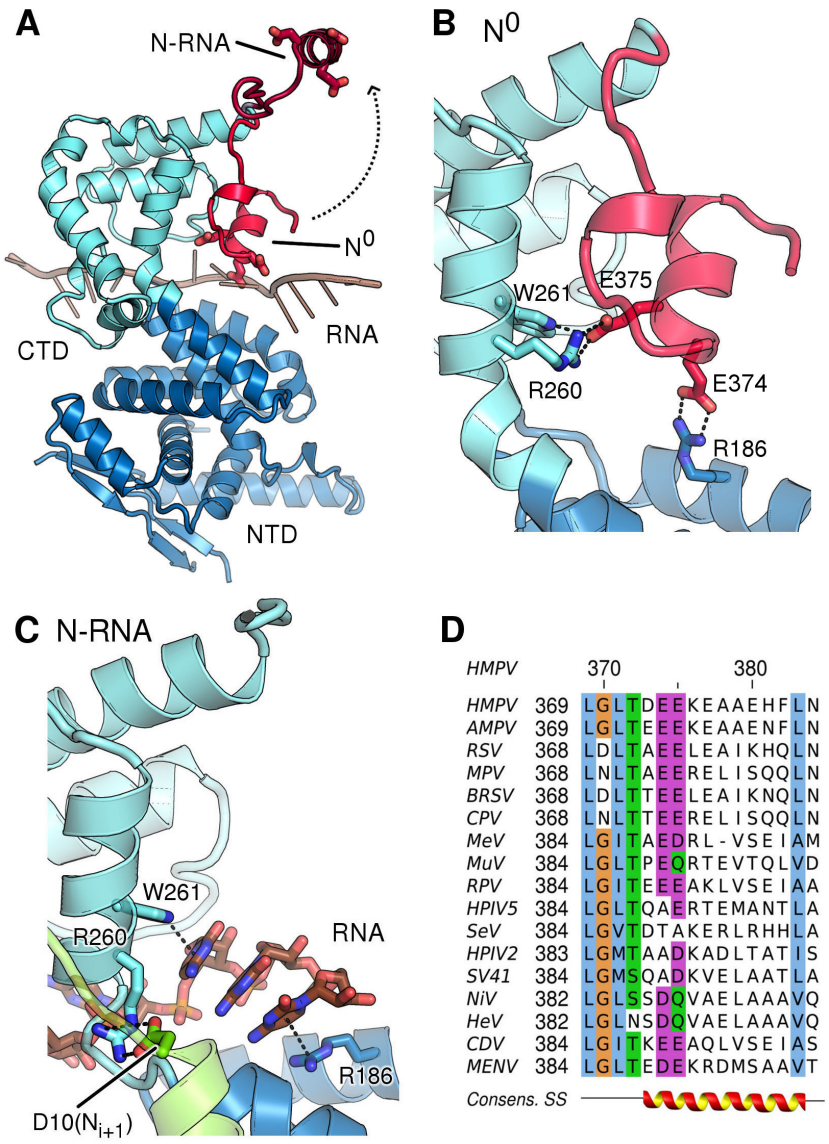


Figure 2.15: Conformational Changes of the CTD-Arm.
(A) Switching of the CTD-arm. The CTD-arm (red) assumes an upwards conformation in the N-RNA state and a downward conformation in the N^0 state (indicated). RNA (coloured in brown) is depicted to highlight the potential steric clash. For clarity, the P protein is not shown. (B) Key salt bridges stabilizing the downward conformation of the CTD-arm (red). Relevant side chains are indicated and shown as sticks. (C) The RNA-bound state of N is shown in the same orientation as in (B) for comparison. (D) Sequence alignment of paramyxovirus N proteins in the region of the CTD-arm. The alignment reveals a semi-conserved LGLT-motif which is immediately followed by a stretch of residues with helical propensity. Residues are coloured according to the ClustalX scheme. Virus abbreviations: *HMPV*, human metapneumovirus, *AMPV*, avian metapneumovirus, *RSV*, respiratory syncytial virus, *MPV*, murine pneumonia virus, *BRSV*, bovine respiratory syncytial virus, *CPV*, canine pneumonia virus, *MeV*, measles virus, *MuV*, mumps virus, *RPV*, rinderpest virus, *HPIV5*, human parainfluenza virus 5, *SeV*, Sendai virus, *HPIV2*, human parainfluenza virus 2, *SV41*, simian virus 41, *NiV*, Nipah virus, *HeV*, Hendra virus, *CDV*, canine distemper virus, *MENV*, menangle virus. Figure adapted from Renner et al. 2016.

2.4 MD Simulations of N^0-P

To test the stability of the N^0-P crystal structure in solution, I decided to carry out classical, explicit-solvent molecular dynamics simulations (MD). Specifically, the aim was to computationally validate both the P-binding site and the downward position of the CTD-arm. If the interacting surfaces within the crystal structure are sufficiently strong and physiologically relevant then it is to be expected that they would be stable within the course of the simulation (Fisette et al. 2016). However, if unbinding were to occur during the simulation (in the case of long simulations in the µs range) or the atomic coordinates around the interaction surface show strong fluctuations in shorter simulations, this may be an indicator for low affinities or unphysiological interfaces.

The structure of HMPV N^0-P determined in this work was missing some stretches of residues due to conformational flexibility and proteolysis. This encompasses the NTD-arm (residues 1-29), an internal loop (residues 102-110), and a flexible region at the end of the CTD-arm (residues 384-394). To generate a complete starting geometry for the simulation these missing regions were modelled based on the RNA-bound HMPV N structure and the corresponding structure from RSV. As the CTD-arm residues 384-394 are not accounted for in either of those structures and are predicted to be intrinsically disordered these were modelled in an unstructured, outstretched conformation. The final starting model consisted of the complete HMPV N chain, non-covalently bound to P_{1-30}.

Following preparation of the system, including energy minimisation and initial equilibration (see Materials and Methods for details), duplicate simulations were performed for an aggregated simulation time of ~1.2 µs using the GROMACS 4

MD package (Pronk et al. 2013). Per-residue root mean square fluctuations (RMSFs) for both trajectories are plotted independently for N and P (**Fig. 2.16A** and **B**, respectively). RMSF values for N core domains were generally low with spikes coinciding with positions of internal loops and the protein termini, indicating that these are flexible in solution, as expected (**Fig. 2.16A**). The high RMSFs around residue 105 are in line with the fact that no electron density was visible for this region in the N-RNA or N^0-P crystal structures. Importantly, the P-protein remained stably bound to the CTD of N for the entire course of both simulations. RMSFs of P residues were slightly higher than those within the core domains of N, not unexpectedly for a short peptide (**Fig. 2.16A**).

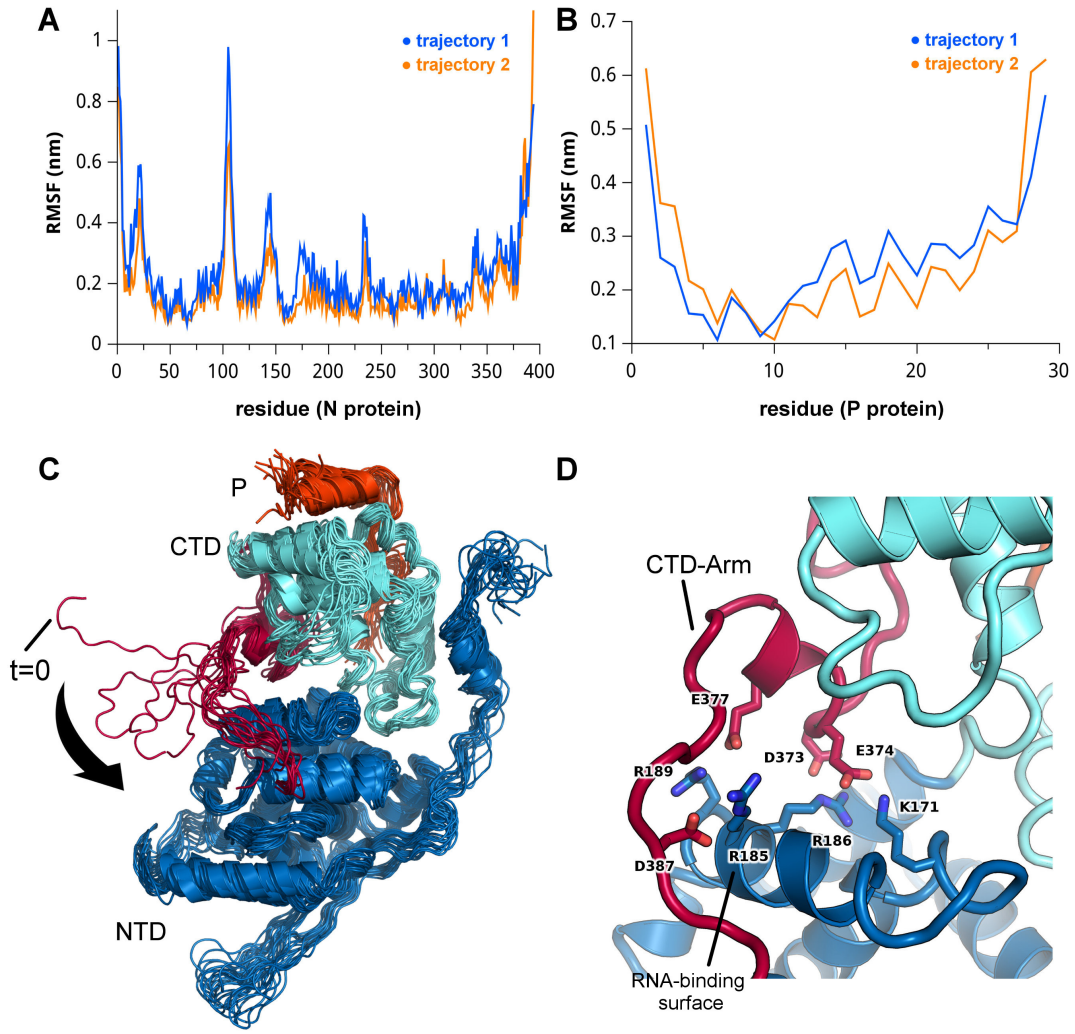


Figure 2.16: Molecular Dynamics Simulation of N^0-P.
Duplicates of classical, explicit-solvent MD simulations of HMPV N^0-P were carried out. Time-averaged per-residue RMSF values for the nucleoprotein (A) and the P protein (B) are shown for both trajectories. (C) Structural ensemble showing how the initial steps of one simulation evolved. Twenty sets of protein coordinates are aligned, representing the first 200 ns of the simulation. The earliest structure is indicated (t=0). The NTD of N is shown in dark blue, the CTD in light blue, the P protein in orange, and the CTD-arm in red. (D) Close-up view of a representative simulation snapshot (simulation time 700 ns) showing interactions between the CTD-arm and the RNA-binding region of N.

The downward conformation and the overall geometry of CTD-arm residues 360-380 was also found to be stable during the simulation. Notably, the salt bridge between Arg260 of the core domain and Glu375 of the CTD-arm (see **Fig. 2.15B**) remained in-tact in all time steps, fixing its position. The disordered tip of the CTD-

arm (residues 384-394), which was modelled in an outstretched conformation, rapidly folded back onto the core of N within the first 70 ns of the trajectory (indicated with black arrow in **Fig. 2.15C**) and remained fluctuating around this position. Indeed these residues wrap around the principal RNA binding surface of N, assisting in the formation of an extensive network of salt bridges which effectively neutralises the positively-charged RNA-binding cleft (**Fig. 2.15D)**. Taken together, these simulation data confirm the stability of the P-N interaction and lend additional support to the notion that the negatively charged CTD-arm shields N from premature RNA uptake in solution.

2.5 Solution Characterization by SAXS

To further probe the solution structure of the N^0-P chimera and RNA-bound N, a characterization by small angle x-ray scattering (SAXS) was carried out. The SAXS data were collected at Diamond Light Source, Didcot, UK, beamline B21. For solution scattering measurements a construct was utilised which was truncated by 8 residues at the N-terminus ($\Delta 8N^0$-P) as this improved sample monodispersity and led to more well-behaved SAXS curves (**Fig. 2.17A**). The radius of gyration (R_g) was determined via linear-fitting in the Guinier region (**Fig. 2.17B**) as implemented in the Software package ScÅtter (www.bioisis.net). The molecular mass calculated from our SAXS data (Rambo and Tainer 2013) closely represented the actual mass determined from the sequence (44 kDa vs. 47 kDa, respectively), demonstrating that the particle is monomeric in solution. Consistently, a single copy of the crystal structure fits

snugly into an ab-initio bead model (**Fig. 2.17C**) calculated from the SAXS data using the ATSAS package (Petoukhov et al. 2012).

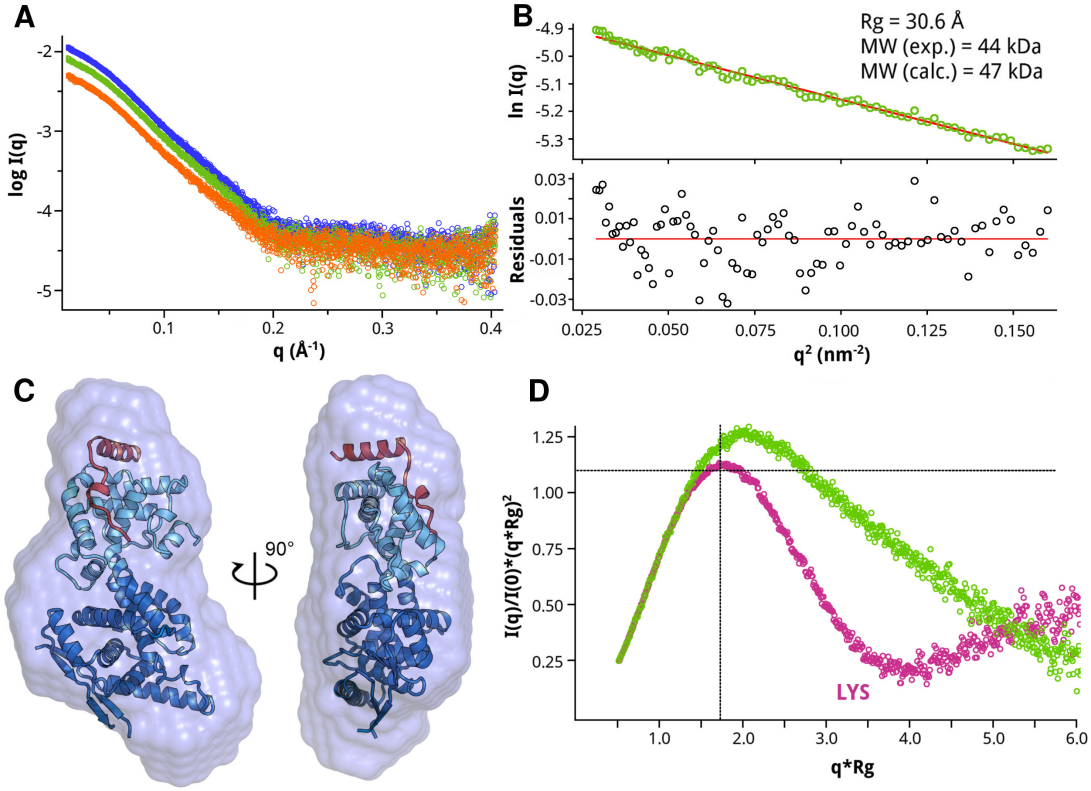


Figure 2.17: Characterization of the N^0-P Chimera in Solution.
(A) SAXS data of $\Delta 8N^0$-P were collected for protein concentrations of 1.2 mg/ml, 1.9 mg/ml and 2.7 mg/ml shown in orange, green and blue, respectively. (B) Representative Guinier fitting and corresponding residuals. The radius of gyration (Rg) determined from the fit is indicated, as well as the molecular weight (MW) calculated from the SAXS data and the mass determined from the protein sequence. (C) Crystallographic structure of N^0-P fitted into an ab-inito bead model calculated using DAMMIF and DAMAVER. (D) Flexibility analysis via normalised Kratky plots of $\Delta 8N^0$-P and a lysozyme standard (BID: LYSOZP, http://www.bioisis.net) in green and magenta, respectively. The Guinier-Kratky point ($\sqrt{3}$, 1.104) at the intersection of the dotted lines indicates the peak position of an ideal globular and rigid protein. The $\Delta 8N^0$-P is shifted upwards and to higher q*Rg values demonstrating deviation from ideal behaviour.

Particle rigidity and globularity can be semi-quantitatively assessed using a dimensionless Kratky plot (Rambo and Tainer 2011). The particular strength of dimensionless Kratky plots is that they allow for a direct comparison of different samples independent of size and aggregation state by normalization with R_g and the forward scattering intensity I_0. Kratky analysis of $\Delta 8N^0$-P revealed that the sample possessed considerable flexibility in solution (**Fig. 2.17D**). This is somewhat in contrast to the previous suggestion that the P peptide rigidifies the nucleoprotein C-terminal and N-terminal domains in an open conformation (Yabukarski et al. 2014). However, it is difficult to assess if the flexibility observed in the Kratky plot is due to large-scale hinge-motions between CTD and NTD or is caused by conformational heterogeneity of the protein termini, which would be in line with the MD simulations discussed above. In addition, the fact that in the chimeric construct the P protein and the N protein are covalently linked constitutes a limitation in the interpretation of the Kratky analysis. To mitigate this a construct was designed in which the P peptide could be specifically cleaved off from N via a 3C-protease site ($N\text{-}3C_{site}\text{-}P_{1\text{-}40}$). Unfortunately, this construct failed to recombinantly express. Similarly, purification of N^0-P, which was subjected to limited proteolysis, unfortunately yielded too degraded product to be used for SAXS. It seems likely that during *in situ* proteolysis, the formation of crystals protected the sample from too excessive proteolytic digestion.

Next, SAXS data of the RNA-bound subnucleocapsid rings of HMPV N were collected (**Fig. 2.18A**). As observed previously by cryo-EM, the sample is a heterogeneous mixture of rings with varying numbers of N subunits and the SAXS data should therefore be treated carefully. Because of this I refrained from attempting

to calculate any solution envelopes for N-RNA rings. In principle it should be possible to fit a theoretical SAXS curve calculated from a structural ensemble of N-RNA rings with 9, 10, and 11 subunits to the experimental scattering curve. This was not attempted here, however, as the rough composition of sub-populations was already estimated using cryo-EM (see above). Determination of R_g by Guinier analysis was complicated by small amounts of aggregation detectable in the low-q regime and by the large size of the scatterer (Rambo and Tainer 2011). To generate a dimensionless Kratky plot I instead determined the R_g value of the N-RNA (R_g = ~62 Å) from the P(r) function as implemented in the ScÅtter package. The Kratky analysis revealed an additional peak at a q*Rg value of ~4.8, consistent with the ring-shape of the molecule (**Fig. 2.18B**). A comparison of N-RNA with the N^0-P chimera indicated a significantly higher level of "foldedness" in the assembled and RNA-bound N. This is consistent with the C-terminal and N-terminal extensions of N tightly latching on to the adjacent subunits in the ring and the core domains clamping down on the RNA-strand, causing global rigidification. In conclusion, the SAXS analysis of N^0-P and N-RNA, although complicated by differences in constructs, points towards a greater flexibility of the unassembled and RNA-free state vs. the N-RNA state.

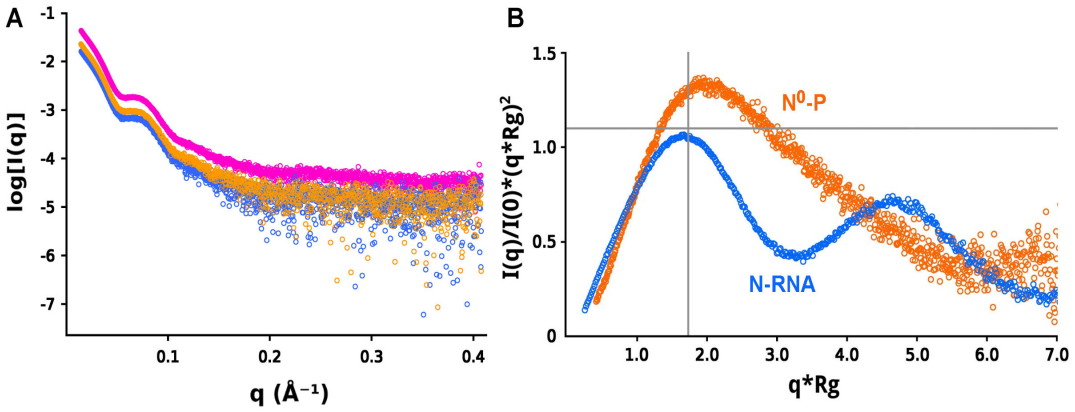


Figure 2.18: Solution Comparison of N-RNA and N^0-P via SAXS.
(A) SAXS data of RNA-bound N rings were collected for protein concentrations of 1.0 mg/ml, 1.5 mg/ml and 2.0 mg/ml shown in blue, orange and cyan, respectively. (D) Flexibility analysis via normalised Kratky plots of Δ8N^0-P (orange) and N-RNA (blue). The Guinier-Kratky point (√3, 1.104) at the intersection of the dotted lines indicates the peak position of an ideal globular and rigid protein.

2.6 Discussion

The RNA genome of members of *Mononegavirales* is never found naked in the cytoplasm of the host, but is instead encapsidated in a protective sheath of nucleoproteins N (Ruigrok et al. 2011a). The resulting ribonucleoprotein complex, termed nucleocapsid, is resistant to enzymatic digestion by host nucleases. It serves not only in storing the viral genetic material but also as the platform for the assembly of the replication and transcription machinery. During replication the polymerase cofactor P chaperones N in a RNA-free, monomeric state (N^0-P), feeding it to nascent nucleocapsids for co-transcriptional packaging and nucleocapsid assembly. This is essential for the virus as, otherwise, N would irreversibly take up unspecific host RNA, leading to a dead-end state.

To further the understanding of nucleocapsid assembly in *Pneumovirinae* and related viruses, a crystallographic study of assembled HMPV N bound to RNA (N-

RNA) and N^0-P was performed. The high-resolution structure of HMPV N^0-P clearly defines the interacting surfaces between the two proteins on the amino-acid level which can now serve as a basis in establishing structure-function relationships for HMPV and related viruses. Furthermore, the same interface is used not only by P but also by the CTD- and NTD-arms of N during assembly. Because of this, inhibitors targeting the P-binding site of the structural protein N may be less prone to the development of escape mutations. Such molecules may potentially compete with viral factors and thus inhibit the essential step of nucleocapsid assembly. The feasibility of such therapeutics has been demonstrated by the efficacy of peptides mimicking the P-protein. Indeed, such peptides have been demonstrated to have protective and antiviral activity in tissue culture (Castel et al. 2009; Yabukarski et al. 2014).

Structural studies of N^0-P complexes are complicated by the propensity of N to irreversibly oligomerise immediately upon recombinant expression. In this work a strategy to generate stable and soluble N^0-P complexes was utilised, in which a portion of the P protein was fused to the C-terminus of N. Previous studies have suggested that the regions of P responsible for chaperoning RNA-free N are located at the N-terminus of the polymerase cofactor for most, if not all mononegaviruses (Karlin and Belshaw 2012; Karlin et al. 2003). The N-binding portion of HMPV P was identified using disorder metaprediction and secondary structure predictions. I propose that this method may generally be applicable in the study of N^0-P complexes of mononegaviruses. A comparison of disorder profiles of P proteins from a selection of different viruses demonstrates that putative N-binding regions are straightforward to identify (**Fig. 2.19A to D**). The production of other viral N^0-P complexes could be achieved by simply attaching these regions to the flexible termini of the

corresponding N-proteins via molecular biology methods and subsequently expressing the chimeras. The techniques utilised here to stabilise the HMPV N^0-P complex and subsequently solve the structure may be effective in the structural elucidation of N-chaperoning by P in a host of related viruses. Indeed, initial work on N^0-P complexes from a diverse set of mononegaviruses has already been taken up by myself and other authors have reported studies on N proteins which utilised similar strategies (Guryanov et al. 2016).

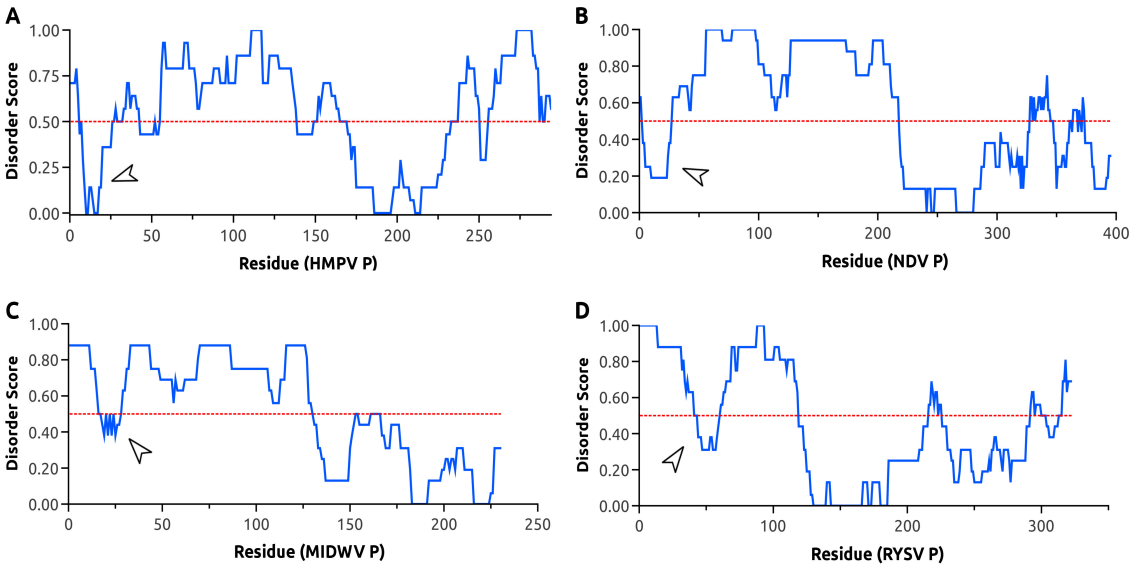


Figure 2.19: Examples for Disorder Metapredictions of P Proteins.
Consensus disorder profiles for the P proteins of human metapneumovirus (A), Newcastle disease virus (B), Midway nyavirus (C), and rice yellow stunt virus (D) calculated with the DisMeta server (Huang et al. 2014). The arrows indicate putative N-binding regions. These regions are often predicted to be partially ordered and possess helical secondary structure propensity.

With the availability of structures of both the assembled N-RNA and monomeric N^0-P states from HMPV it was possible to carry out a direct comparison. Similar to what has been recently suggested for other viruses (Guryanov et al. 2016;

Yabukarski et al. 2014), the structures revealed how in *Pneumovirinae* the P-peptide competes for the same binding site as the C-terminal and N-terminal extensions of N (CTD-arm and NTD-arm), thus inhibiting N-oligomerisation. In addition, a notable conformational difference in the CTD-arm between the two states could be observed. In the RNA-free N^0-P structure, CTD-arm residues 360-384 were in a downward conformation, reaching into the RNA binding cleft. The downward conformation is fixed in position through a considerable interacting surface with the N core domain and by salt-bridges formed through two negatively charged residues which were found to be strictly conserved in all pneumoviruses. MD simulations of a total aggregated length of 1.2 µs demonstrated that the CTD-arm positioning in the RNA-groove is stable throughout the trajectories. Furthermore, an interesting charge complementarity between the CTD-arm and the rest of N could be noted. In general, as RNA-binding proteins, nucleoproteins possess high isoelectric points (pI). In all pneumoviruses, however, the CTD-arms are markedly negatively charged in comparison (Table 4). These data led to the hypothesis that, prior to incorporation into a nascent nucleocapsid, the negatively charged CTD-arm functions in shielding the positively charged RNA-binding groove of N, thus preventing premature uptake of unspecific nucleic acids. In accordance with the simulation data, I propose that in solution, CTD-arm residues 360-384 of HMPV N^0-P remain tucked into the RNA-binding groove, while the remaining residues 384-394 sample conformational space around this site, thereby further contributing to blocking it.

	pI (CTD-arm)	pI (rest of protein)
AMPV	3.9	9.0
HMPV	4.0	9.0
HRSV	4.2	8.7
BRSV	4.0	8.7
MPV	3.5	8.7
CPV	3.5	8.6

Table 4: Isoelectric points of N protein CTD-arms in *Pneumovirinae*.
Virus abbreviations: *AMPV*, avian metapneumovirus, *HMPV*, human metapneumovirus, *HRSV*, human respiratory syncytial virus, *BRSV*, bovine respiratory syncytial virus, *MPV*, murine pneumonia virus, *CPV*, canine pneumonia virus

A lingering question in the nucleoprotein field is how the P-bound N is released and subsequently transferred onto new viral genomes during replication (Cox and Plemper 2015). In the following section a model is proposed which adds molecular detail to the growth mechanism of nascent nucleocapsids (**Fig. 2.20**). In this model, when naive N^0 is delivered to the replication site by P (green N^0 in **Fig. 2.20**) the CTD-arm of N_i (dark blue in **Fig. 2.20**) must flip upwards, attaching to the incoming N and therefore supplanting P. I suggest that the upwards flip of the CTD-arm and the insertion of nascent RNA into the binding cleft of N_i occur in concert with attachment of the CTD-arm of N_i to the next incoming N^0. The continuous synthesis of viral RNA through the RdRP and RNA-insertion into N would thereby drive the CTD-arm flipping upwards. The CTD-arm is then free to form new contacts to incoming N^0, thus liberating it from P. Because P-proteins attach to the nucleocapsid (P binds N^0 with its N-terminus and the nucleocapsid with its C-terminus), replication sites are characterised by high localised concentrations of nucleoproteins, P-proteins and viral RNA. The concerted displacement of the CTD-

arm from the RNA-site, displacement of P from N^0-P, and RNA-insertion may require these high local concentrations to exceed a certain threshold. This may explain why displacement of P from N^0 preferentially occurs at the replication site (high local concentrations of viral factors).

It is worth noting that in (mammalian) host cells infected with HMPV the N protein is expected to be kept RNA-free by P, whilst in bacteria recombinant co-expression of N and P is not sufficient to generate RNA-free N^0. Although this may be a consequence of the differences between the *E. coli* expression system and the genuine HMPV host cells it it tempting to speculate that additional mechanisms exist. For instance, it is feasible that the virus might subvert the host machinery in a way that spatially brings together P proteins with ribosomes in the process of translating N. This way P could grasp nascent N before it is able to contact unspecific RNA-molecules. However, further investigation is required to understand in all details how P inhibits the acquisition of unspecific RNA by N and delivers N^0 to the replication site in host cells.

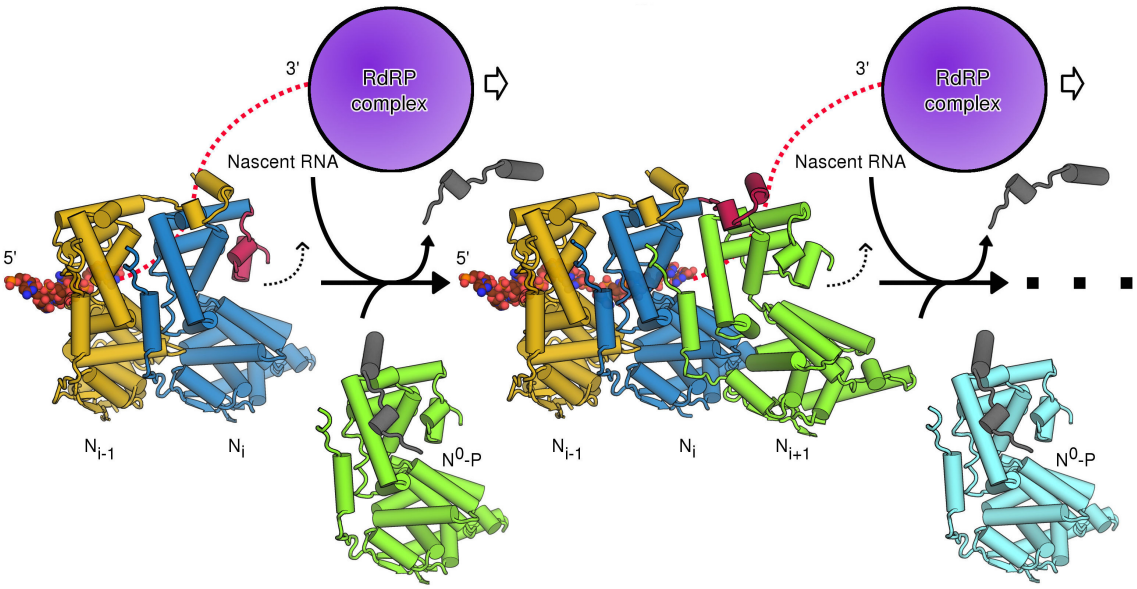


Figure 2.20: Hypothetical Model of Nucleocapsid-Filament Growth.
At the viral replication site nascent genomes are co-transcriptionally packaged into N proteins. Concerted insertion of nascent RNA (indicated) into N primes the CTD-arm (coloured in red) to flip upwards (dotted arrow) and attach to incoming fresh N^0-P, thereby displacing P (shown in grey). N^0-P complexes are indicated. The RdRP complex is shown schematically as purple sphere. For additional discussion see accompanying text. Figure adapted from Renner et al. 2016.

3 HMPV M2-1

3.1 Background

Apart from the main project of studying the N protein from human metapneumovirus I was also involved in the characterization of the antiterminator M2-1. The M2-1 protein is a RNA-binding transcription antiterminator which is present only within the *Pneumovirinae* subfamily. Although many questions remain, it has been suggested that M2-1 enables pneumoviruses to fine-tune the gradient of viral mRNAs (see Introduction) in a M2-1 concentration-dependent (and thus life-cycle stage-dependent) manner (Fearns and Collins 1999). Why it is unique to pneumoviruses (and debatably filoviruses in the form of VP30, see Concluding Remarks) and the evolutionary origin of M2-1 are unclear. M2-1 is delivered to transcription sites by the P protein and prevents the premature abortion of mRNA synthesis within genes and at gene junctions (intragenic and intergenic antitermination, respectively) (Collins et al. 1996; Hardy and Wertz 1998; Fearns and Collins 1999). M2-1 has a modular architecture consisting of a CCCH-type Zn-finger at the N-terminus, followed by a helical oligomerisation domain, and finally a globular core domain (Blondot et al. 2012).

At the time I joined the project the crystal structure of HMPV M2-1 had just been determined by Cedric Leyrat in our group. As discussed in the Introduction HMPV M2-1 crystallised as an asymmetric tetramer (see **Fig. 1.18**) in which one of the protomers displays an open conformation and three are closed. The overall organization of M2-1 places the four Zn-fingers on one side of the tetramer and the

four core domains on the opposite side, separated by a four helix bundle (please revisit **Fig. 1.18** for an overview of the M2-1 architecture). While our group was working on HMPV M2-1, the structure of the RSV homologue (38% sequence identity) was reported (Tanner et al. 2014). Contrasting with HMPV, RSV M2-1 possessed a symmetrical arrangement of protomers (**Fig. 3.1**), all of which were in the closed conformation. The conformational differences between these structures was intriguing and opened up the possibility that M2-1 can sample open and closed conformations in solution and that this may be functionally important.

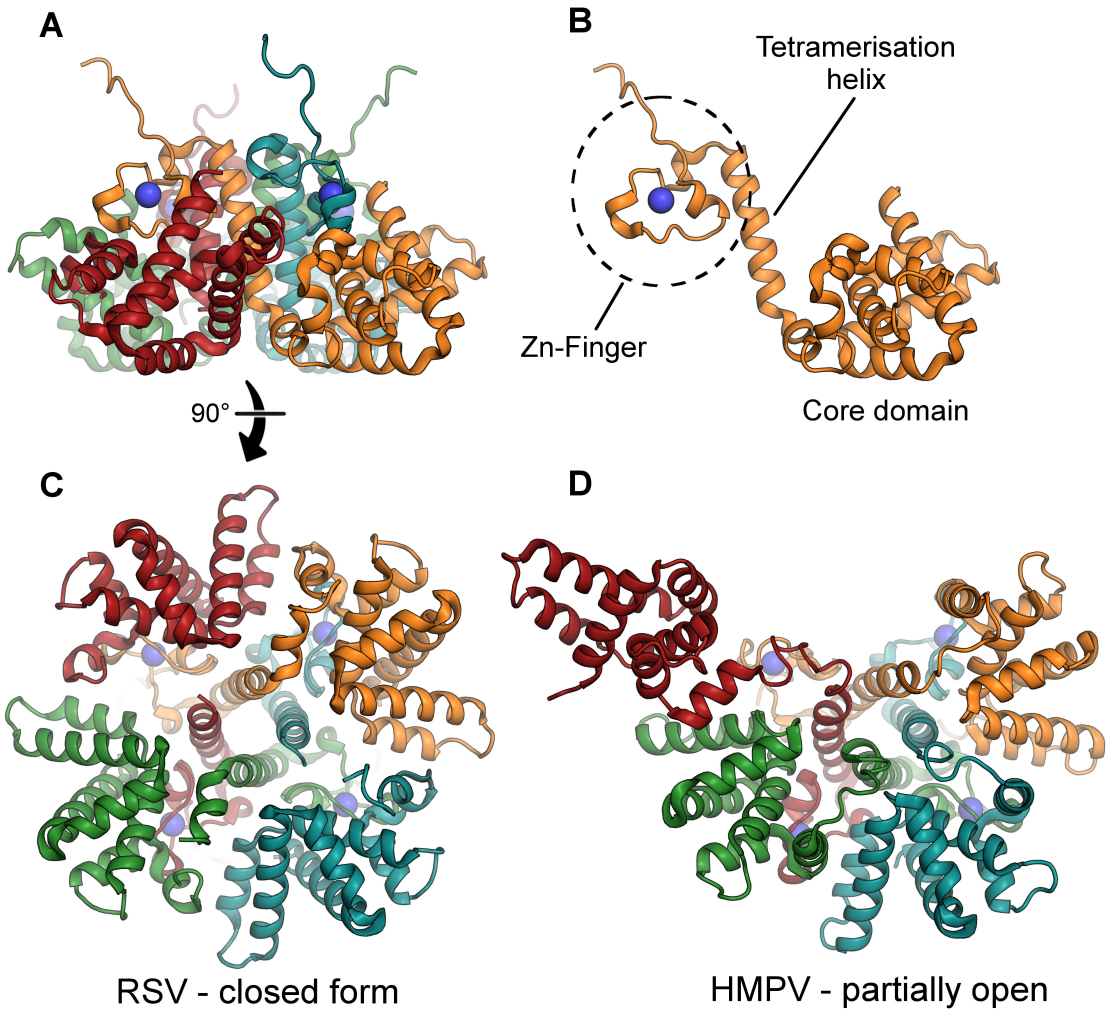


Figure 3.1: Comparison of RSV and HMPV M2-1 Proteins.
(A) Side-view cartoon representation of the RSV M2-1 protein (pdb ID: 4c3b). Each subunit of the tetrameric complex is coloured individually. (B) One monomer of RSV M2-1 is shown individually in the same orientation and colouring as before. Domains are indicated. (C) Top-view of RSV M2-1, highlighting the symmetric nature of the tetramer. In contrast to RSV, one core domain of the HMPV M2-1 tetramer is dissociated, leading to a partially open, asymmetric tetramer (D).

Besides the conformational dynamics of M2-1, a further underexplored aspect of M2-1 function was the structural basis of RNA recognition. A previous study has utilised NMR to assess the RNA binding properties of the recombinantly expressed core domain of RSV M2-1 (Blondot et al. 2012). Using chemical shift perturbation

analysis the authors could identify a cluster of positively charged residues exposed on the core domain which is involved in RNA binding and preferentially binds A-rich sequences. The functional importance of these residues was additionally validated by mutagenesis studies (Blondot et al. 2012). RNA-binding studies which used full-length M2-1 rather than only the core domain measured significantly higher binding affinities (Tanner et al. 2014). This may either indicate collaboration of the protomers of the tetramer and/or involvement of the Zn-finger domain in RNA binding. Indeed, later mutagenesis experiments showed that substitution of positively charged residues just before the Zn-finger negatively effected RNA-binding (Tanner et al. 2014). In general, the highest affinities for M2-1 could be observed with long oligonucleotides (around 13nt) with a high A-content, which is a feature of positive-sense mRNA transcripts (Blondot et al. 2012; Tanner et al. 2014). Furthermore, oligonucleotides of positive-sense gene-end sequences, which regulate termination and polyadenylation between consecutive viral genes (Hardy et al. 1999; Harmon et al. 2001; Sutherland et al. 2001), were found to be potent binders (nM range). Although these studies provide valuable information on RNA-binding preferences and affinities of M2-1, there was still much paucity in the structural basis of this recognition. Of special interest to our group was the question of if and how the conformational changes of the M2-1 core domain (open vs. closed) relate to the binding of ligand.

3.2 Purification of Recombinant Sample

HMPV M2-1 (molecular weight of the tetrameric complex: 92 kDa) was recombinantly expressed in bacteria and purified via Ni-IMAC utilising a N-terminal $(His)_6$-tag. High salt concentrations (1M) were used throughout the lysis and affinity-purification to disassociate the protein from *E.coli* nucleic acids. Subsequent reduction of the salt concentration in the buffer solution was found to lead to precipitation of the protein when at high concentrations. To mitigate this, the solubilising agent NDSB201 was added to the gelfiltration buffer at concentrations of 0.5 – 1.0 M. The final gelfiltration step yielded monodisperse sample (**Fig. 3.2**). Sample intended for crystallisation was additionally subjected to 3C-protease digestion to remove the affinity-tag and, finally, a further round of size-exclusion chromatography was performed.

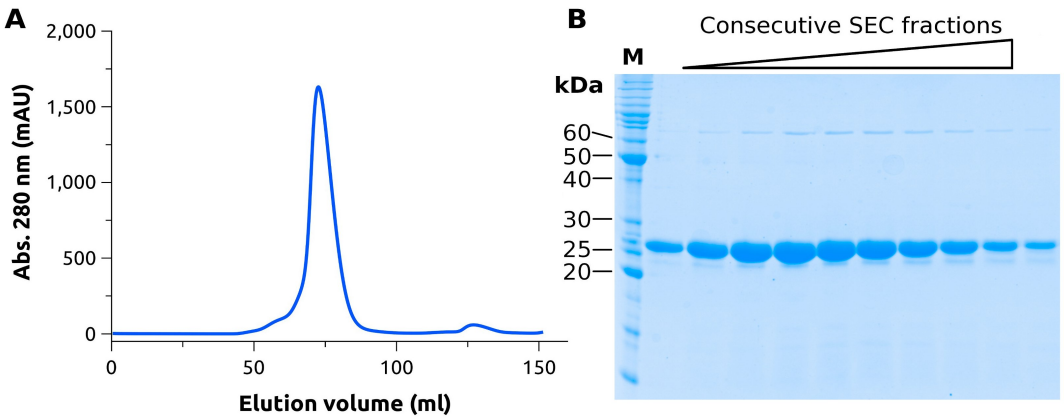


Figure 3.2: Purification of HMPV M2-1.
(A) Representative size-exclusion chromatogram (Superdex 200) of HMPV M2-1. (B) SDS-PAGE analysis of size-exclusion fractions.

3.3 M2-1 in Complex with Nucleotides

3.3.1 Crystal Production

Examination of the crystal packing of HMPV M2-1 (space group $P2_1$, one asymmetric tetramer in the ASU, ~50% solvent content) revealed that sufficiently large solvent channels are present within the crystals to potentially allow the diffusion of ligands (**Fig. 3.3**). A promising strategy to obtain structural data on nucleic acid binding by M2-1 would thus be the soaking (Hassell et al. 2007) of short oligonucleotides into M2-1 crystals. To this end large quantities of crystals for subsequent soaking experiments were produced.

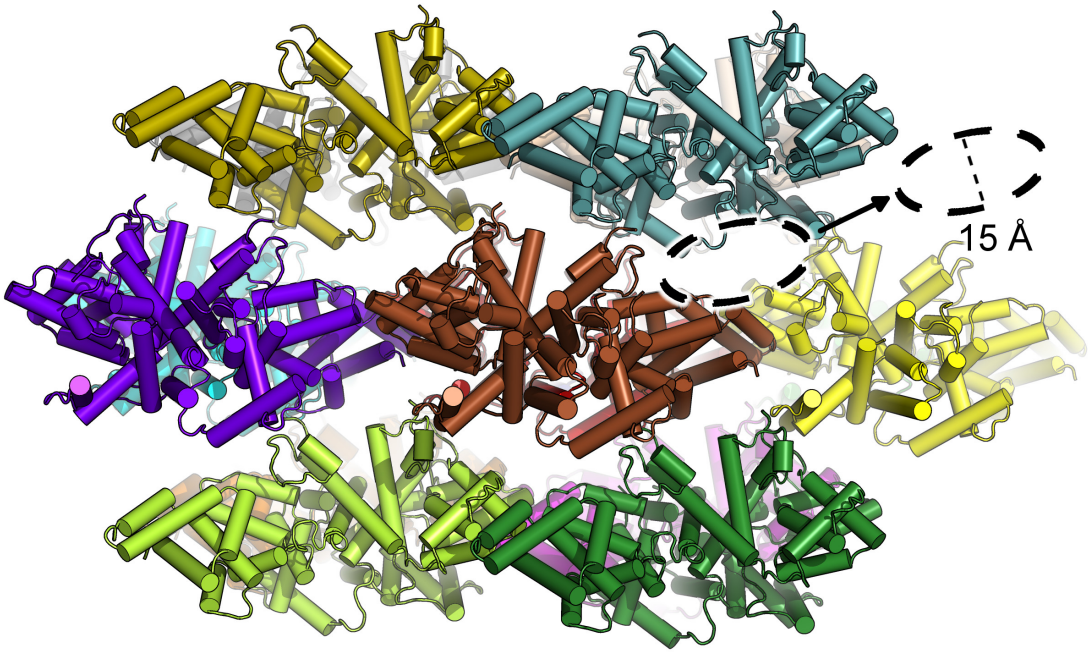


Figure 3.3: Section of the Crystallographic Packing of HMPV M2-1.
Each M2-1 tetramer is coloured individually and shown in cartoon representation. The crystallographic packing leads to solvent channels (indicated by dashed lines) between the tetramers, potentially allowing diffusion of ligands.

The initial apo-crystals obtained by Cedric Leyrat, which were used to solve the structure, were grown with high concentrations of NDSB201 and NaCl in the protein buffer (500 mM NDSB201 and 300 mM NaCl). These high concentrations of solutes in the sample vs. the reservoir solution led to a significant increase in the volume of the sitting drop within the course of the crystallisation experiment, limiting crystal size and reproducibility. I was able to alleviate this effect and reproducibly grow large single crystals (rod-shaped, typically ~250 μm long and ~50 μm thick) of M2-1 by supplementing the original reservoir condition (28% (wt/vol) PEG-MME 2000, 0.100 M bis-Tris pH 6.5) with 500 mM NDSB201. Large quantities of crystals for nucleotide-soaking could then be produced using optimization/dilution screens based on this reservoir formulation (**Fig 3.4**).

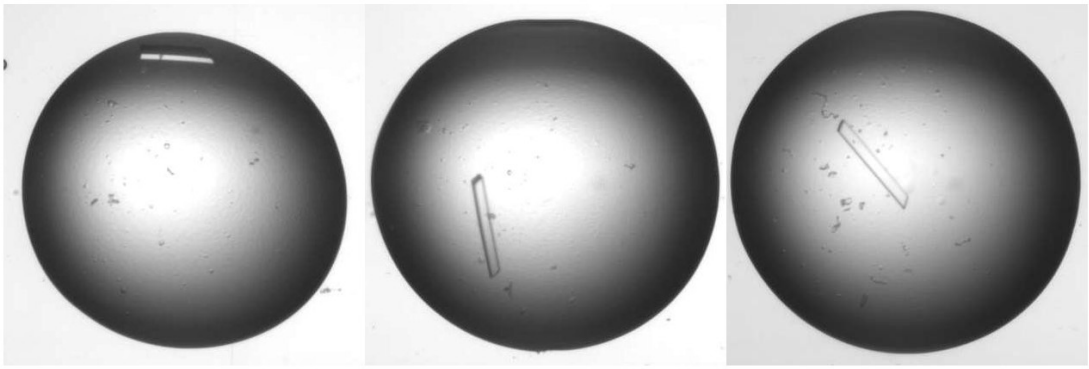


Figure 3.4: Representative Examples of M2-1 Crystals.
Crystals were obtained by optimization/dilution screening based on the reservoir condition which generated the initial crystals (28% (wt/vol) PEG-MME 2000, 0.100 M bis-Tris pH 6.5), supplemented with 500 mM NDSB201.

3.3.2 Nucleotide Soaking Experiments and Structure Refinement

M2-1 crystals were soaked by adding 1 µl of reservoir solution supplemented with the appropriate nucleotide and 25% glycerol on top of the sitting drops containing the crystals. The samples were then observed and incubated for various lengths of time and finally cryo-cooled in liquid nitrogen. Crystals were soaked with mononucleotides (AMP, UMP, CMP and GMP at 50 mM each) or longer DNA/RNA fragments ranging from di-/trinucleotides to whole GE sequences (positive-sense) of HMPV genes. In general, soaks with mononucleotides behaved well with no visible damage to the crystals even when incubated for extended time periods (multiple days). Soaks with oligonucleotides, however, showed the appearance of cracks in the crystals when incubated for more than a few hours. This was accompanied by loss of diffraction.

The soaked crystals, together with some unsoaked control crystals, were screened at Diamond Light Source, Didcot, UK, beamlines I03, I04, and I24. All collected data were autoprocessed with XIA2 (Winter et al. 2013). Structures were phased by molecular replacement with PHASER (McCoy et al. 2007) using the original apo-structure. This structure contains one tetramer of M2-1 per asymmetric unit, with 3 protomers in the closed conformation and one in the open conformation. Difference maps were assessed after molecular replacement and a single round of refinement.

Using this batch of crystals I was able to improve upon the resolution of the initial unbound M2-1 structure which was previously determined (2.1 Å vs 2.5 Å). In the case of the soaked samples no extra density was observed with crystals treated with UMP, CMP, GMP, dinucleotides or trinucleotides. Crystals soaked in AMP

revealed multiple regions near the Zn-fingers with additional density in F_O-F_C difference maps (**Fig. 3.5A**). Similarly, extra density for three or more contiguous nucleotides was observed in crystals subjected to short soaking with the AGTTA oligonucleotide (**Fig. 3.5B**), whereas other oligonucleotides of similar length did not yield additional density. The AGTTA DNA represents the DNA equivalent of the 5'-end of the conserved (positive-sense) GE sequence found at gene-junctions in HMPV (5'-AGUUAnnnAAAAA-3'). GE sequences have been previously shown to bind to RSV M2-1 with nanomolar affinity (Tanner et al. 2014).

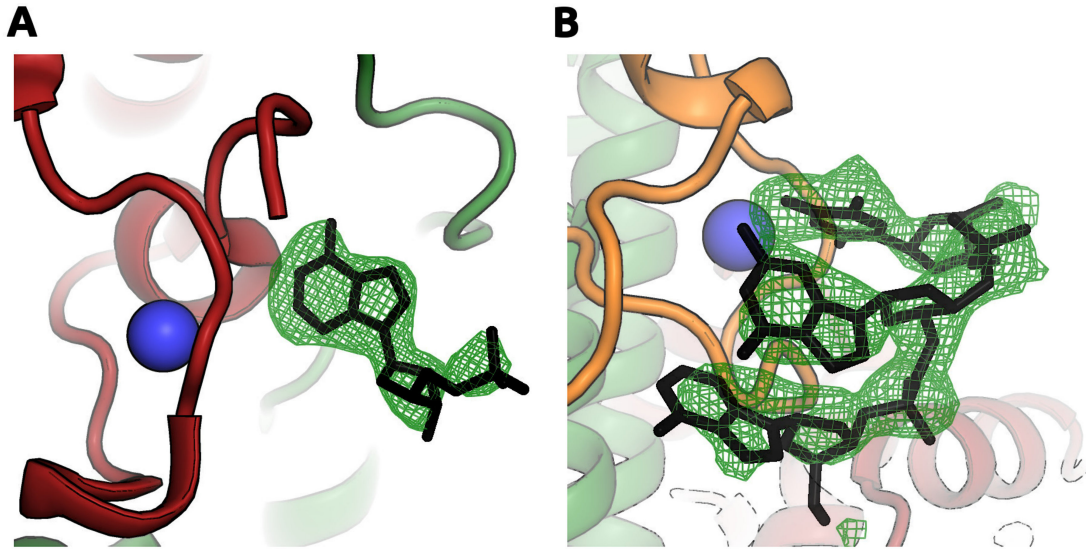


Figure 3.5: Sample Difference Map Density for Soaking Experiments.
Representative F_O-F_C difference maps for M2-1 crystals soaked with AMP (A) and AGTTA (B) with fitted nucleotides. The maps were obtained directly after molecular replacement and a single round of refinement in PHENIX without adding waters or ligands to the model. The density maps are contoured at 2.5σ. The extra density is in the vicinity of the Zn-fingers (Zn-atoms are shown as blue spheres).

The structures of the new apo M2-1 protein and those derived from the AMP/AGTTA soaked crystals were refined iteratively using BUSTER (making use of TLS parameters and 3-fold NCS via LSSR (Local Structure Similarity Restraints) as implemented in BUSTER) and COOT (Smart et al. 2012; Emsley and Cowtan 2004). In the final refined AMP-bound structure, for three of the four M2-1 protomers, a molecule of AMP was modelled into clear density adjacent to the Zn-finger (see **Fig. 3.6** for overview of the AMP locations and **Fig. 3.7A** for detailed view). Furthermore, at the Zn-finger where the open core domain is located, strong density for the purine ring of an additional AMP was observable, whilst the ribose and phosphate groups were mostly disordered (AMP2 in **Fig. 3.6** and **Fig. 3.7B**). In the case of the AGTTA-soaked crystal the four contiguous nucleotides 5'-AGTT-3' were modelled in the final structure. The terminal 'A' nucleotide did not show density and the penultimate 'T' was less well-defined than the others (see **Fig. 3.7C** and also **Fig. 3.8** for an overview highlighting the DNA oligonucleotide location) but could still be modelled at a lower density threshold. The structures are further discussed in the next section. Data collection and refinement statistics are shown below (Table 5).

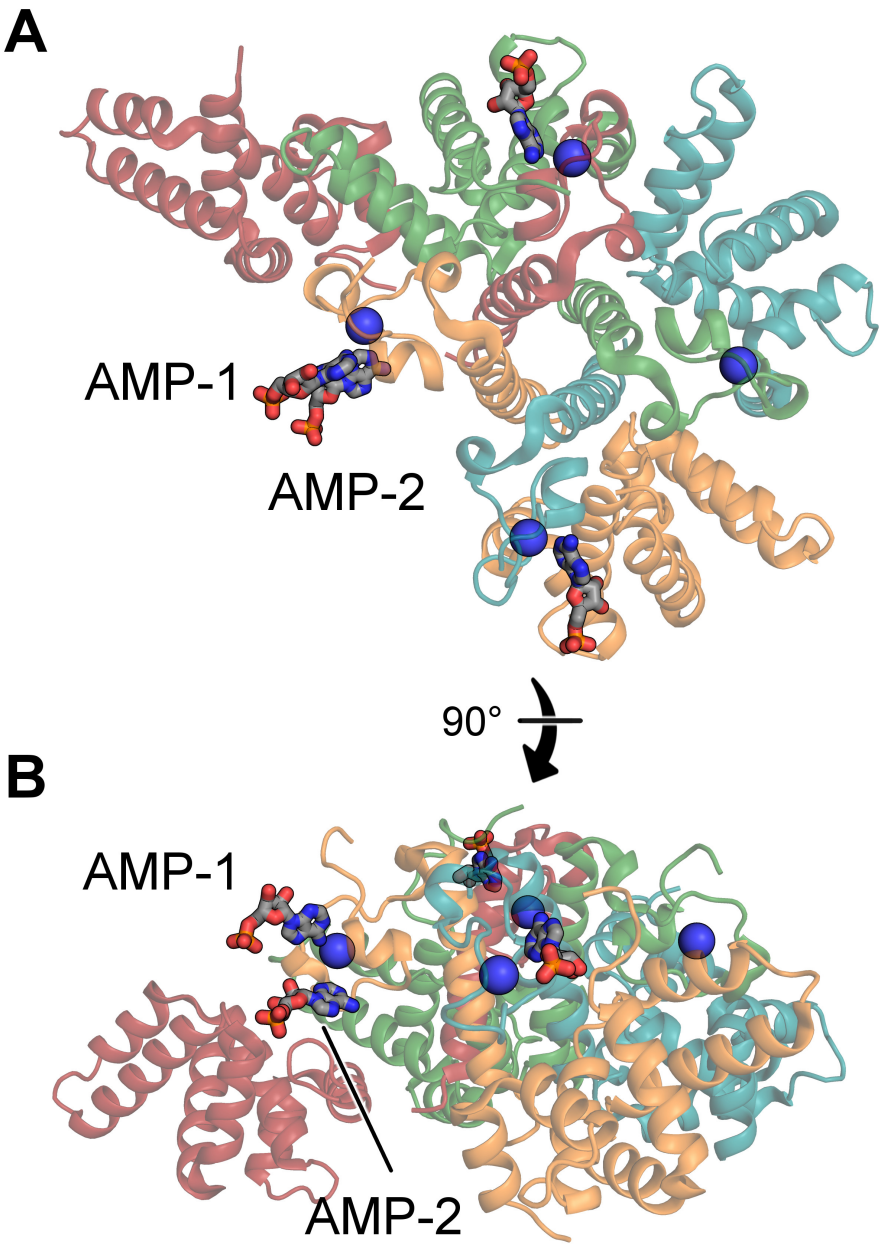


Figure 3.6: Overview of AMP-Soaked M2-1.
Overview of the structure of asymmetric HMPV M2-1 derived from the AMP-soaking experiments. (A) shows a top-view of M2-1 and (B) shows a side-view. M2-1 is shown in cartoon representation with each monomer coloured individually. Zinc atoms are represented by blue spheres. AMP molecules are drawn as sticks and coloured by element. Presumably due to solvent accessibility, AMPs are found at three of the four Zn-fingers. One of the Zn-fingers harbours two AMP molecules (labelled AMP-1 and AMP-2).

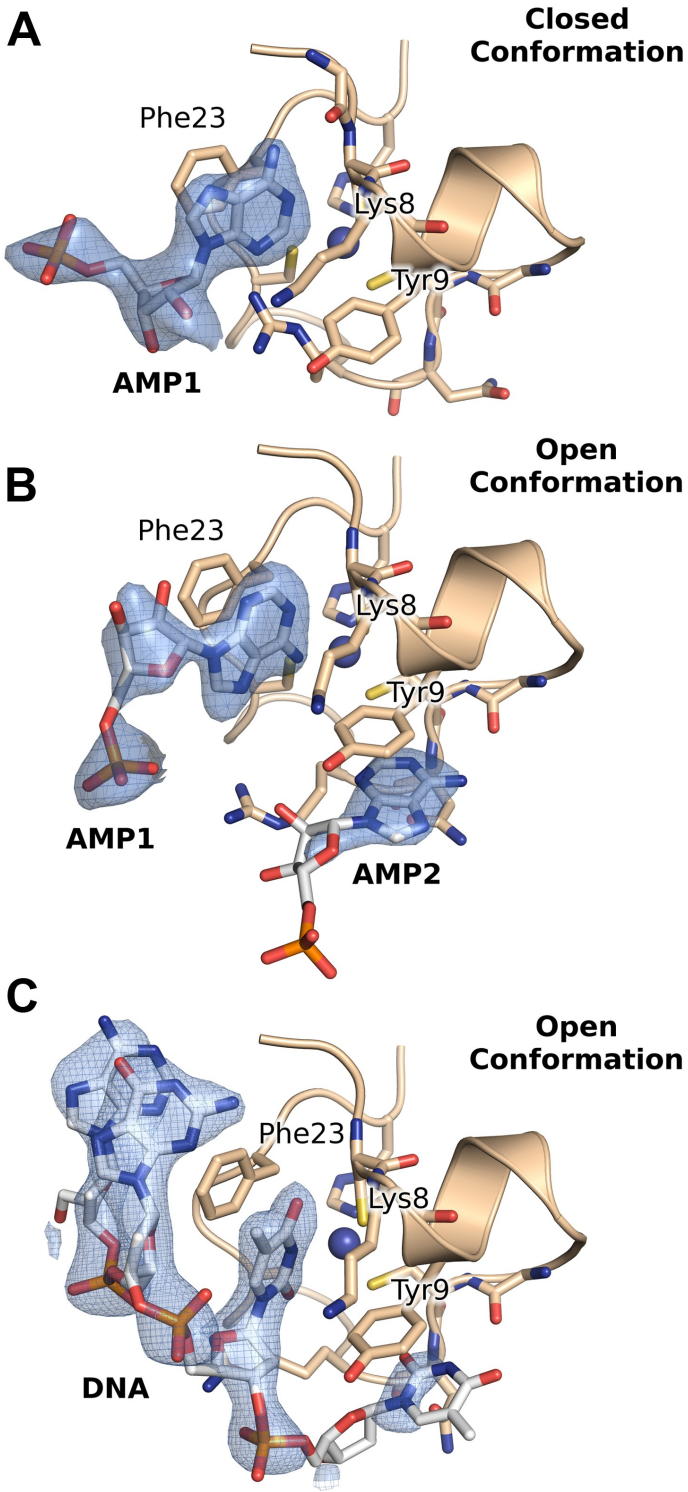


Figure 3.7: Electron Density of the Ligands in the Refined Structures.
Electron density $2F_O$-F_C maps contoured at 1σ. Close-ups of the Zn-finger regions are shown for the AMP-bound structure (A and B) and AGTT-bound structure of M2-1 (C). Figure taken from Leyrat et al. 2014b.

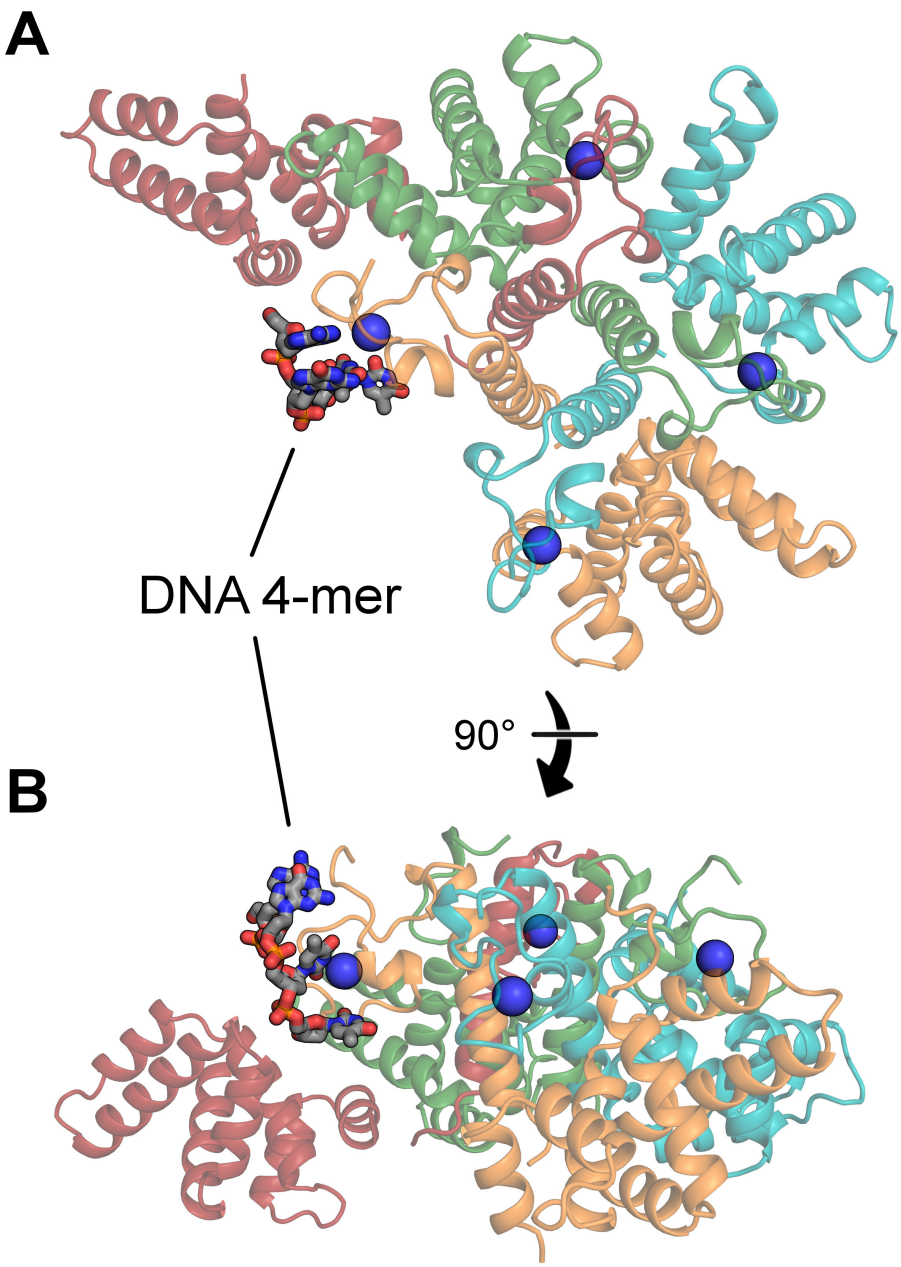


Figure 3.8: Overview of AGTTA DNA-Soaked M2-1.
Overview of the structure of asymmetric HMPV M2-1 derived from the DNA oligonucleotide soaking experiments. (A) shows a top-view of M2-1 and (B) shows a side-view. M2-1 is shown in cartoon representation with each monomer coloured individually. Zinc atoms are represented by blue spheres. The modelled AGTT 4-mer is drawn as sticks and coloured by element. Presumably due to solvent accessibility, the DNA oligonucleotide bound at the open subunit.

Data Collection	**Apo (new)**	**AMP-soak**	**DNA-soak**
Wavelength (Å)	0.97	0.97	0.98
Number of images	1800	1800	2400
Ω Oscillation per image (°)	0.2	0.2	0.15
Exposure time per image (s)	0.1	0.1	0.1
Beam size (µm)	20x20	20x20	50x45
Transmission (%)	100	100	100
DLS beam-line	I24	I24	I04
Detector	Pilatus 6M	Pilatus 6M	Pilatus 6M-F
Resolution (Å)	44.81–2.10	49.80–2.01	63.04–2.28
Space group	P 2_1	P 2_1	P 2_1
Cell dimensions			
a, b, c (Å)	50.0, 93.4, 85.2	50.1, 93.6, 85.6	50.1, 93.9, 85.5
α, β, γ (°)	90.0, 95.4, 90.0	90.0, 95.8, 90.0	90.0, 95.8, 90.0
CC (1/2)	1.00 (0.73)	1.00 (0.70)	1.00 (0.70)
R_{merge}	0.07 (0.94)	0.07 (0.84)	0.05 (0.84)
I / σI	23.1 (2.0)	22.5 (1.8)	19.6 (1.8)
Completeness (%)	99.8 (99.2)	98.8 (85.1)	99.7 (98.8)
Redundancy	11.8 (5.8)	12.8 (5.1)	6.5 (5.0)
Measured Reflections	538, 740	660, 719	235, 027
Unique Reflections	45, 494	51, 659	35, 938
Refinement			
R_{work} / R_{free}	19.2/22.2	18.6/20.8	19.4/22.2
RMSD			
Bond lengths (Å)	0.009	0.010	0.009
Bond angles (°)	1.00	0.96	1.03
Ramachandran plot quality			
Favoured (%)	97.8	97.3	97.3
Allowed (%)	2.2	2.4	2.4
Outliers (%)	0.0	0.3	0.3

Table 5: Data Collection and Refinement Statistics for HMPV M2-1.
Data were collected at Diamond Light Source (DLS), Didcot, UK.
Values in parentheses correspond to the highest resolution shell, *RMSD*, root mean square deviation from ideal geometry, R_{free} was calculated for a 5% subset of reflections.

3.3.3 Comparison of the Apo-Forms

Superposition of the two apo-structures of M2-1 (RMSD = 1.2 Å) reveals a significant shift of the open core domain (**Fig. 3.9A**). This underlines how the cores

and the tetramerisation helices are flexibly-linked and that the core-domain can adopt a range of degrees of "openness". A comparison of the protomers within one tetramer (RMSDs = 0.9 - 1.1 Å) shows that the positions of the closed core domains relative to the tetramerisation domains also vary (**Fig. 3.9B**). This indicates that even in the closed conformation, the interactions holding the domains in place relative to the tetramerisation helix are fuzzy and allow for relative movements (i.e. flexing). This is consistent with the observations made by Esperante et al. 2011 that M2-1 in solution is non-globular and extended.

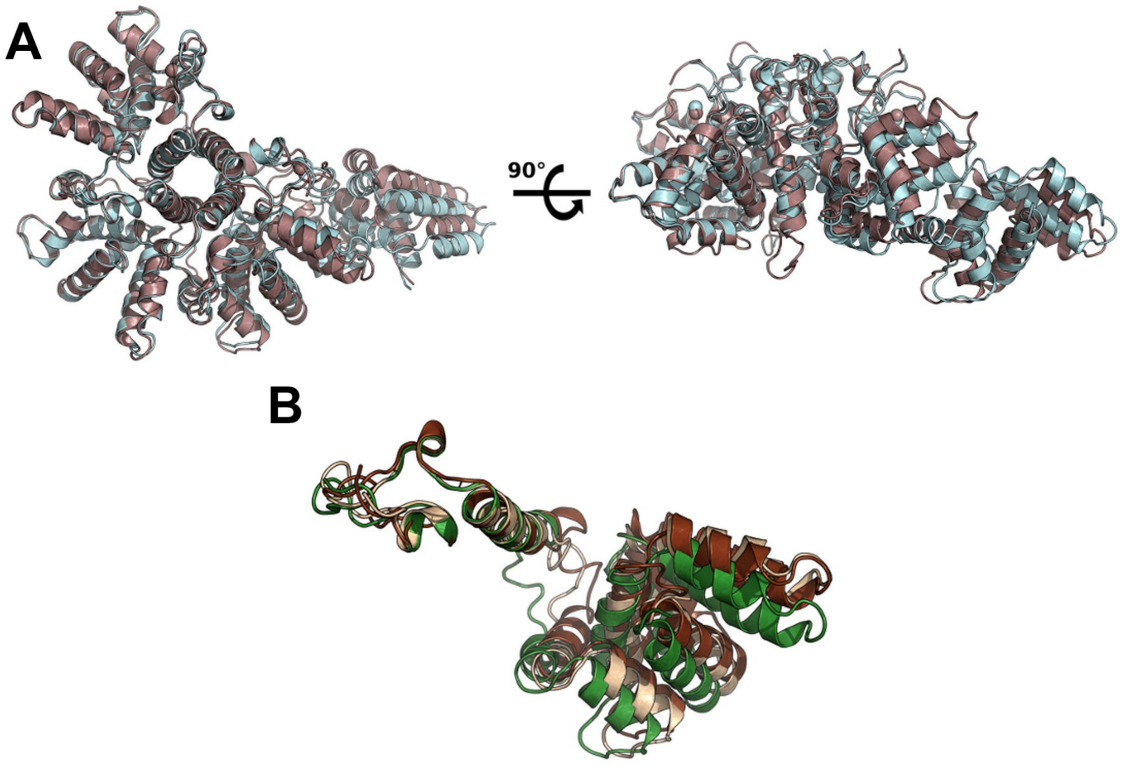


Figure 3.9: Interdomain motions within M2-1 Crystals.
(A) Structural alignment of the two apo-forms of M2-1 (light blue and light brown) showing movement of the open core domain. (B) Structural alignment of the four protomers of M2-1 from one crystal. The protomers are aligned onto the tetramerization domain. The superposition highlights the possible range of positions of the core domains. See accompanying text for RMSD values.

3.3.4 Nucleotide-Bound Structures

Analysis of the M2-1 complexes obtained from the soaking experiments reveals the structural basis of nucleotide-binding at the Zn-finger (**Fig. 3.10**). It is worth noting that these interactions are likely to be conserved from HMPV to RSV, taking into account the structural homology between the viruses. Specifically, both Zn-fingers possess aromatic residues and a conserved Lys protruding from the surface (**Fig. 3.10A**). The purine ring of the AMP inserts into the Zn-finger of HMPV via aromatic stacking with Phe23 (**Fig. 3.10B**). Simultaneously, the hydrophobic carbon-tail of Lys8 stacks on top of the purine ring, clamping the base in between these two residues. In addition, the C-terminal tail of a neighbouring (closed-state) protomer contributes in the stabilization of a ribose oxygen via Asp172. Furthermore, the heteroatoms of the AMP base form polar contacts with the backbone of multiple Zn-finger residues. Interestingly, at the Zn-finger associated with the open-conformation core, the same AMP binds in an inverted orientation (**Fig. 3.10C**). This ambiguity of orientation likely results from binding to a single nucleotide, rather than a longer stretch of consecutive nucleotides. The second AMP, which is only observed at one of the four protomers, inserts between Tyr9 and Asn16 (**Fig. 3.10C**). Again, the residues stack against the purine from both sides. Considering the ambiguity of orientation, it is not absolutely clear from the structure why AMP (vs. other mononucleotides) was preferentially bound in the mononucleotide soaking experiments. It may be that the preference is due to the larger purine ring of adenosine (compared to a six-membered pyrimidine ring) inserting more optimally (in terms of size) in the hydrophobic “slot” formed by Lys8 and Phe23. Furthermore, the -NH2 moiety attached at C6 of the adenosine purine ring (standard notation), which is not present in the guanine purine

ring, can additionally interact with backbone carbonyls. However, it should be noted that the previously reported (Tanner et al. 2014) binding preference for long A-rich oligonucleotides may also be determined by the binding site on the core domain. This is further explored later in this chapter.

The structure of AGTT-bound M2-1 reveals the 5'-3' directionality of a stretch of nucleic acid (**Fig. 3.10D**). In this case pyrimidine bases are sandwiched between Phe28/Lys8 and Tyr9/Asn16 instead of the purines in the AMP soak. The positively charged residues Arg17 and Lys22 contribute to the stabilization of the phosphate backbone of the oligonucleotide. The first adenosine forms polar contacts with the backbone of Ala5. Although not directly in contact, it should be noted that this brings the ligand into close proximity of Arg3 and Lys4, residues which have been determined to be important for RNA-binding by mutagenesis in RSV (Tanner et al. 2014). Modelling of larger nucleotide stretches will be explored later in this chapter. A structural comparison with the eukaryotic nucleic acid binding Zn-finger from Nab2 (Brockmann et al. 2012) shows a notable topological conservation of HMPV residues Phe23/Lys8/Tyr9 (**Fig. 3.11**), suggesting a shared binding mode between these proteins. Consistently, perturbation of the Nab2 Zn-finger by mutagenesis has been shown to decrease binding to oligonucleotides (Kelly et al. 2010).

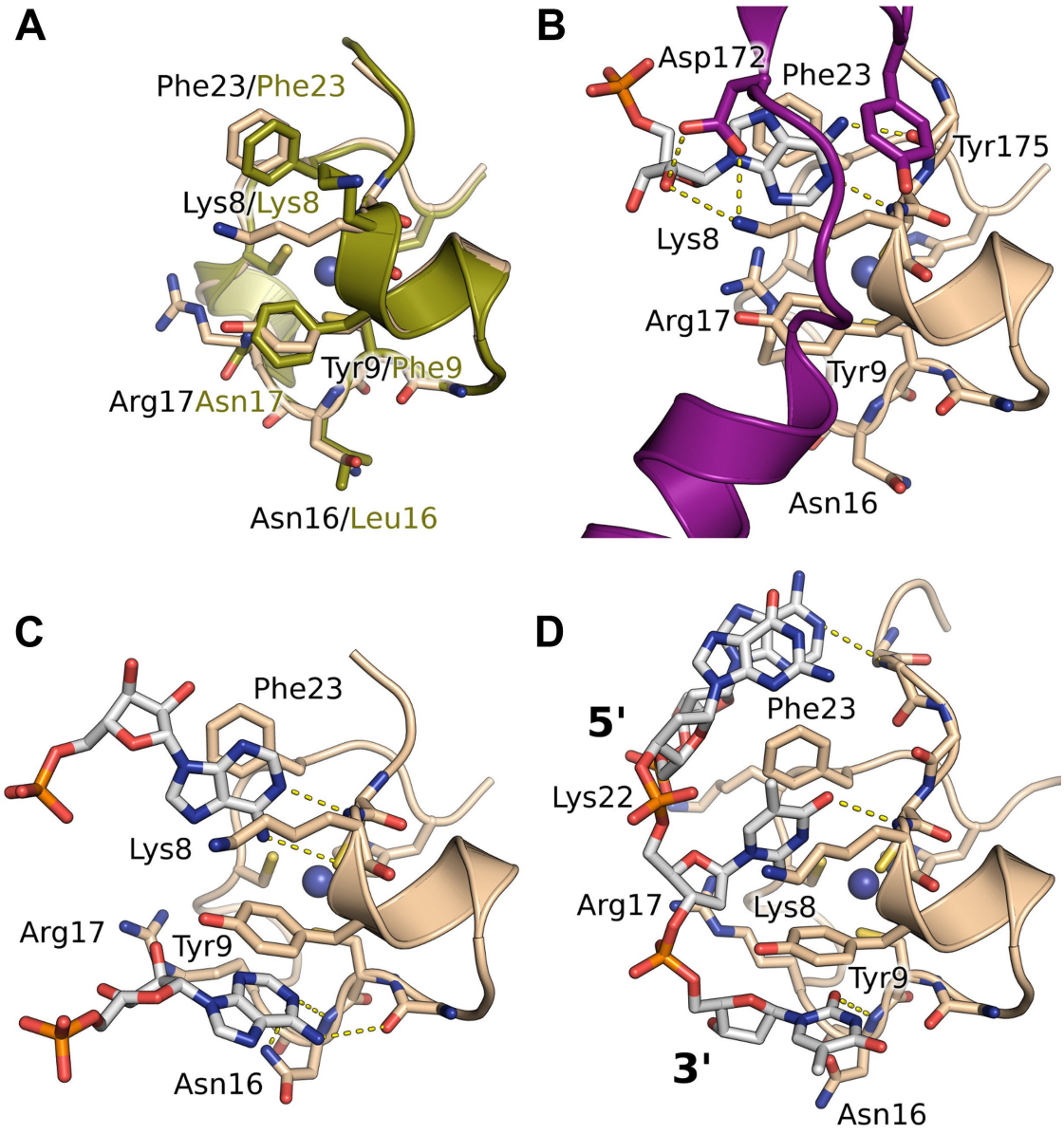


Figure 3.10: Structures of M2-1 bound to Nucleotides at the Zn-finger.
(A) Alignment of the M2-1 Zn-fingers of HMPV (coloured in wheat) and RSV (coloured in green, pdb ID: 4c3b. Key residues surrounding the Zn-finger are shown as sticks. (B) View of a AMP molecule bound to HMPV M2-1. The AMP and key residues are shown as sticks. A C-terminal loop (coloured in purple) of the adjacent protomer packs onto the nucleotide. (C) The HMPV M2-1 Zn-finger at the open core domain can accommodate two AMP molecules (shown as sticks). (D) Structure of the Zn-finger from HMPV M2-1 bound to four contiguous DNA nucleotides (AGTT). In all panels the Zn-atom is displayed as a blue sphere. Figure adapted from Leyrat et al. 2014b.

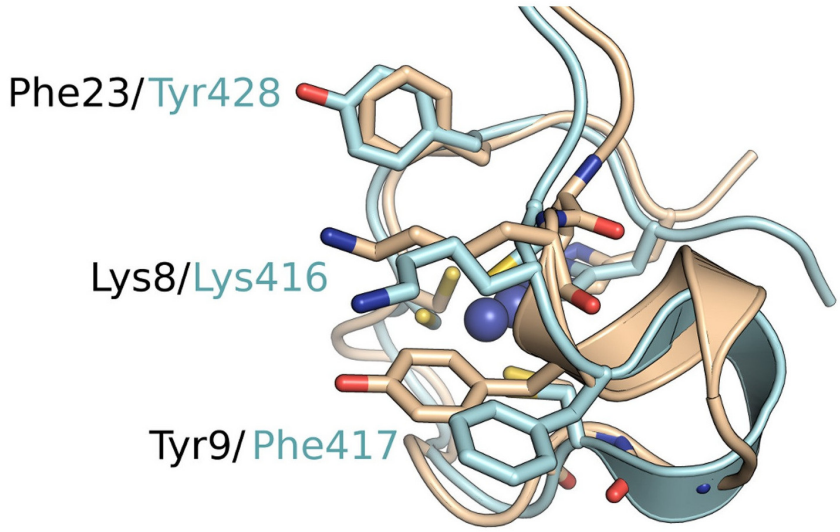


Figure 3.11: Comparison of HMPV M2-1 and Nab2 Zn-fingers.
Alignment of the Zn-finger of HMPV M2-1 (coloured in wheat) with that of the eukaryotic nucleic acid binding protein Nab2 (light blue). The Zn-atoms are displayed as blue spheres. Figure adapted from Leyrat et al. 2014b.

The RNA-binding study performed by Blondot et al. 2012 used chemical shift perturbations to map the RNA-interaction to a patch of positively charged residues protruding from the core domain. Blondot et al. showed that the core domain preferentially binds to adenine rich sequences. A comparison of the core domains of HMPV and RSV demonstrates that the positive patch is highly conserved with many residues being at topologically identical positions (**Fig. 3.12A**). Examination of the crystal packing in the structures of HMPV M2-1-nucleotide soaks reveals contacts of the nucleotides with crystallographically-related core domains. Specifically, while the AMP bases are inserted into the Zn-fingers as described above, the phosphate groups are buttressed against the core domains of M2-1 protomers related through crystallographic symmetry. In the AMP-bound structures the phosphates are stabilised by the Arg152 backbone and Arg149/Lys91 side-chains (**Fig. 3.12C**, **D**), residues

which have also been identified to interact with RNA in RSV by NMR (Blondot et al. 2012). Similarly, in the DNA-bound structure Arg149 of the core domain contacts the purine base of the AGTT oligonucleotide (**Fig. 3.12B**).

This observed simultaneous binding of multiple, different (crystallographically related) M2-1 tetramers to the same (oligo)-nucleotide very likely only occurs within the context of crystal packing and not in a cellular context. However, it is still worth noting the conserved positively charged surface of the core domain (see **Fig. 3.12A**) and its general ability to interact with phosphate groups (even though in a crystallographic context). A role of the core domain as a secondary nucleic acid binding site in addition to the Zn-finger will be further discussed later in this chapter and a binding model will be provided.

The information gained from the soaking experiments of HMPV M2-1 constitutes the only available high-resolution structural data of M2-1 ligand binding. The structures indicate how nucleotide bases insert next to aromatic residues exposed at the CCCH Zn-finger in conjunction with positively charged residues that are in close proximity. AGTT-bound M2-1 suggests the 5'-3' directionality that longer RNAs could adopt when attaching to the protein. Nucleic-acid binding at the Zn-finger is in line with numerous studies demonstrating the functional relevance of the M2-1 CCCH Zn-finger (Hardy and Wertz 2000; Tang et al. 2001; Zhou et al. 2003). In addition, the striking homology with the eukaryotic Zn-finger protein Nab2 further confirms its role in RNA recognition. Finally, crystal packing considerations are consistent with RNA binding to the core domain as shown previously (Blondot et al. 2012; Tanner et al. 2014). A picture is emerging in which M2-1 may use both the Zn-finger and the core domain to bind long RNA molecules. This theme will be further

explored in the following section using small angle x-ray scattering (SAXS) techniques. A model of this simultaneous binding will be presented later in this chapter.

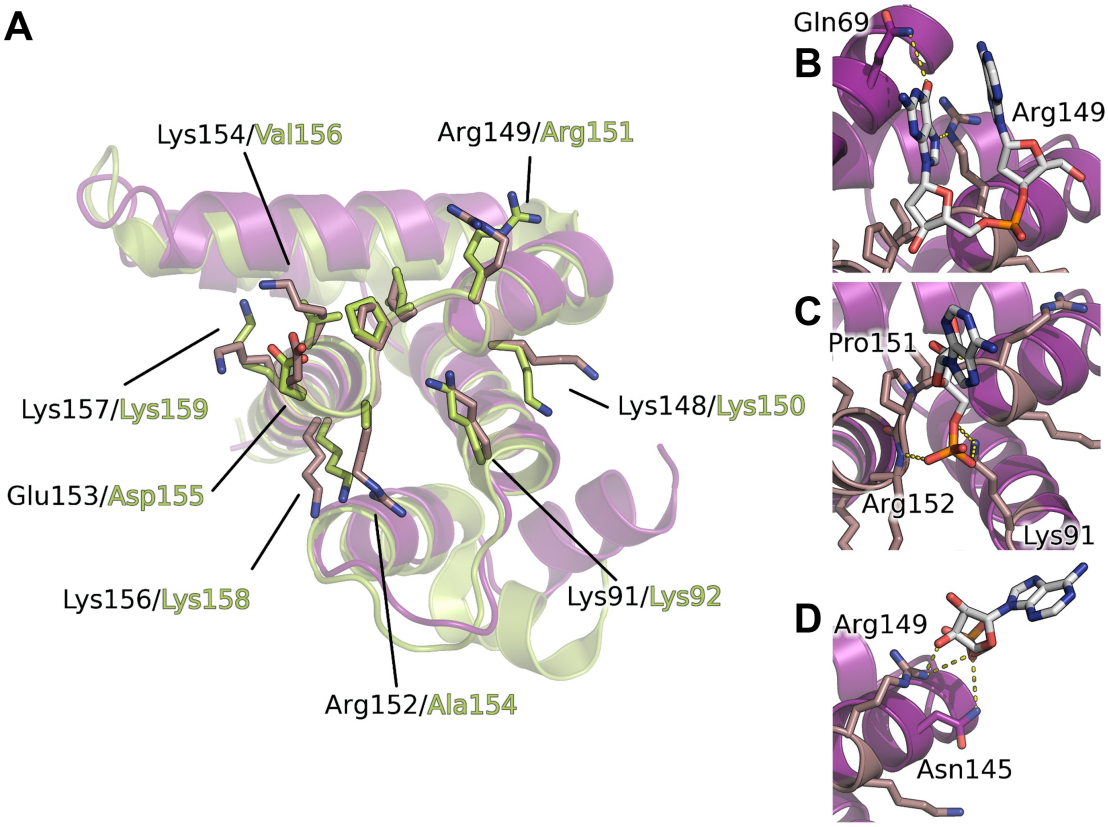


Figure 3.12: Core-domain Interactions with Crystallographically Related M2-1 Copies. (A) Alignment of the M2-1 core domains of HMPV (purple) and RSV (pdb ID: 4c3b, green). RSV residues which have been shown to interact with RNA and the homologous HMPV residues are shown in stick representation. (B-D) Highlight the interactions of the DNA (B) or AMP (C and D) molecules, which were soaked into the crystals, with crystallographically related core domains of M2-1. Figure taken from Leyrat et al. 2014b.

3.4 Analysis of RNA-binding in Solution

To better understand the structural consequences of RNA-binding to HMPV M2-1 in the solution state, RNA-titrations with positive-sense leader RNA (5'-ACGCGAAAAAAU-3') and F gene-end RNA (5'-AGUUAAUUAAAAA-3') oligonucleotides were carried out and measured by SAXS (**Fig. 3.13A**) at beamline BM29 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Adding increasing concentrations of RNA into solutions of M2-1 (1 mg/ml M2-1 with 150 mM NaCl, and 500 mM NDSB201) led to the rapid and unexpected formation of soluble aggregates (aggregation without visible precipitation) of the sample, rendering accurate R_g determination impossible due to the non-linearity of the Guinier region. The observed aggregation is likely to be unphysiological and due to unspecific electrostatic attraction between RNA-bound M2-1. It should be emphasised at this point, that in a physiological context M2-1 is found in complex with the tetrameric P protein of the transcriptase machinery. P in all probability inhibits the self-association of M2-1 observed *in vitro*.

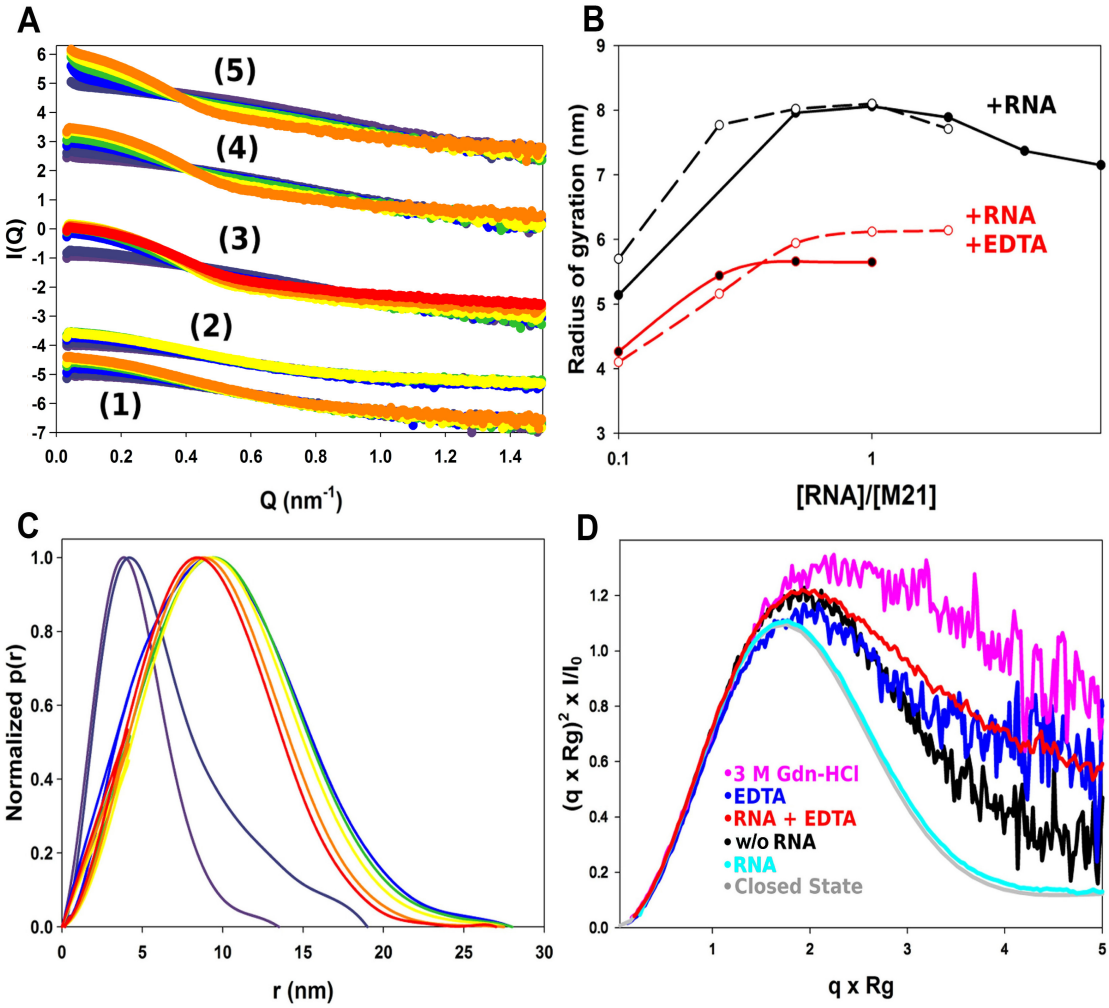


Figure 3.13: Solution Characterization of RNA-binding by SAXS.
(A) SAXS profiles of M2-1 samples with RNA. Either gene-end RNA of the F-gene was titrated at 0, 1, 5, 10, 20, 40, and 80 µM or leader RNA was titrated at 0, 1, 5, 10, and 20 µM, coloured from dark blue to red with increasing RNA concentrations. The five sets of curves are as follows: (1) M2-1 (0.2 mg/ml) + 5 mM EDTA + gene-end RNA, (2) M2-1 (0.2 mg/ml) + 5 mM EDTA + leader RNA, (3) M2-1 (0.2 mg/ml) + gene-end RNA, (4) M2-1 (0.2 mg/ml) + leader RNA, (5) M2-1 (1.0 mg/ml) + leader RNA. (B) Plots of radii of gyration calculated from the RNA titrations versus the concentration ratio of RNA to M2-1. The red plots were calculated using sample which was pre-incubated with 5 mM EDTA. Uninterrupted lines correspond to leader RNA titrations and dashed lines to F gene-end RNA titrations. (C) Pair distribution functions calculated from M2-1 samples with increasing amounts of gene-end RNA (see (A)). (D) Dimensionless Kratky-plots of M2-1 without additional treatment (black curve), with 20 µM RNA (cyan curve), with 20 µM RNA but pre-incubated with 5 mM EDTA (red curve), with 3 M Gdn-HCl (magenta curve), pre-incubated with 5 mM EDTA (blue curve). The grey curve is theoretically calculated for the fully-closed M2-1. Figure adapted from Leyrat et al. 2014b.

RNA-titrations into M2-1 solutions were repeated at lower protein concentrations (0.2 mg/ml M2-1) resulting in more well-behaved data and allowing reliable determination of Rg values. Nonetheless, a rise of Rg (up to ~8 nm) and size-increase could be measured with increasing RNA concentrations, due to the self-association of RNA-bound M2-1 (**Fig. 3.13B**, black curves and **Fig. 3.13C**). As the Rg of the RNA oligonucleotide alone was measured to be less than 2 nm, the observed increase in radius of gyration was not due to the scattering of the nucleotides. Dimensionless Kratky plots can be used to semiquantitatively assess the foldedness of a sample irrespective of aggregation state, with a sharp peak in the curve corresponding to high degrees of compactness whilst an upwards shift indicates dynamic behaviour (Doniach 2001; Bernadó 2010; Receveur-Brechot and Durand 2012). A set of dimensionless Kratky plots of M2-1 with different additives is shown in **Fig. 3.13D**. Partial unfolding of M2-1 by addition of 3 M Gdn-HCl exemplifies the changes in the curves associated with an increase in flexibility (compare black curve and magenta curve in **Fig. 3.13D**). The Kratky plot of the M2-1/RNA mixture indicated a stark globularisation (cyan curve in **Fig. 3.13D**). It is worth noting that at this RNA:M2-1 molar ratio (1:1) the curve is expected to be dominated by the scattering of the protein component due to M2-1 being a somewhat larger scatterer than the oligonucleotide (MW of M2-1 monomer: ~21.2 kDa; MW of oligonucleotide: ~3.6 kDa). Interestingly, a theoretical fully closed-state M2-1 curve (grey curve in **Fig. 3.13D**) closely represented the Kratky curve in presence of RNA, suggesting that RNA may induce closure of the core domain.

Samples which were stripped of zinc by pre-incubation with EDTA and then supplemented with RNA, showed a less pronounced R_g-increase (**Fig. 3.13B**, red

curves). This further indicates an involvement of the Zn-finger with RNA-recognition. Dimensionless Kratky analysis of M2-1/RNA mixtures showed that EDTA pre-treatment inhibited the compaction observed before (**Fig. 3.13D**, red curve), suggesting that simultaneous binding at the Zn-finger and the core-domain are necessary for the transition to the closed state. Indeed, the EDTA-treated curve indicated an increase of flexibility compared to apo M2-1 (compare black curve and red curve in **Fig. 3.13D**). A possible explanation for these phenomena is that, with the Zn-finger perturbed by EDTA-treatment, the RNA-molecules attach to the core domains only with one end, while the other end remains flexible in solution. This would be in contrast to the RNA being clamped down by the Zn-finger at one end and the core-domain at the other.

To summarise, addition of RNA to M2-1 led to self-association of copies of M2-1 tetramers, complicating further analysis. Nonetheless, using an approach based on Kratky analysis, a tentative interpretation of the SAXS data obtained at lower protein concentrations can still be made. The SAXS data indicate that RNA-binding induces a globularisation of M2-1. Kratky curves of experimental RNA-bound M2-1 and calculated closed-state M2-1 are basically superimposable, suggesting that RNA may shift the otherwise flexible M2-1 into the closed state. Perturbing the Zn-finger by stripping the ions with EDTA inhibited the globularisation upon RNA-binding. This strengthens the notion that both the Zn-finger and the core domain collaborate in binding.

3.5 Discussion

The RNA-binding antitermination protein M2-1 prevents the premature termination of transcription of viral mRNAs (intragenic antitermination) (Fearns and Collins 1999; Hardy et al. 1999; Hardy and Wertz 1998). M2-1 is thus important for the synthesis of full-length mRNAs and can be described as a processivity factor of the viral L protein. In addition, M2-1 also facilitates read-through transcription of consecutive genes (intergenic antitermination) leading to the synthesis of polycistronic mRNAs which consist of multiple concatenated gene-transcripts. This phenomenon may be important to fine-tune viral gene expression and facilitate access to genes which are at the end of the genome. *In vitro* binding studies have shown that M2-1 possesses much higher affinity (nM range) for poly-adenosine RNAs than for the poly-C/G/U counterparts (higher than μM) (Tanner et al. 2014). The P protein recruits M2-1 to the transcription site and it has been suggested that, there, M2-1 binds to the nascent mRNA transcripts (Tanner et al. 2014; Blondot et al. 2012). Importantly, the A-content of HMPV mRNA transcripts are appreciably high (Table 6). Finally, M2-1 also shows high affinity (nM range) to positive-sense gene-end sequences (consensus, sequence 5'-AGUUAnnnAAAAA-3') which are located directly after the stop codons of the viral mRNAs and direct transcription termination and polyadenylation via reiterative transcription (Tanner et al. 2014; Jacques and Kolakofsky 1991).

Gene	N	P	M	F	M2	SH	G	L
A-content of mRNA	36%	43%	36%	33%	37%	43%	40%	38%

Table 6: A-Content of HMPV mRNAs (Strain NL1-00, A1).

The data and literature presented in this chapter show how RNA can bind to the core-domain as well as to the Zn-finger of M2-1. The soaking experiments define the directionality of the strand and the Kratky analysis suggests that RNA-binding is accompanied with a conformational switch to the closed state. RNA-titration SAXS data obtained with M2-1 samples that were pre-incubated with EDTA demonstrate that this switch only occurs when the Zn-finger is also involved. The RNA-interacting amino acids which were identified in the course of this work and those that were identified in previous studies (Tanner et al. 2014; Blondot et al. 2012) form a near-contiguous surface on M2-1. Modelling of a conserved gene-end RNA (5'-AGUUAnnnAAAAA-3') onto these residues of the closed-state M2-1 molecule (using the AGTT-bound HMPV M2-1 structure as a starting point) shows that this length of oligonucleotide is highly compatible with the length needed to bridge a surface stretching from the Zn-finger to the core domain (**Fig. 3.14A**, surfaces shown to interact with RNA in this study and the studies cited above are coloured). In addition, a schematic model is shown for clarity (**Fig. 3.14B**). In this model, the length of the gene-end RNA would directly necessitate that the appropriate core domain of M2-1 transitions into the closed state to allow simultaneous binding, in line with the Kratky analysis. The model is also consistent with the observation that a 13-mer possessed nanomolar affinity while an 8-mer showed dramatically reduced affinity (μM range) as it would be too short to reach both binding sites simultaneously (Tanner et al. 2014). The less conserved nucleotides in the middle of the gene-end consensus sequence would be located between the Zn-finger and core-domain binding sites in the model (**Fig. 3.14**), providing a rational for their variability.

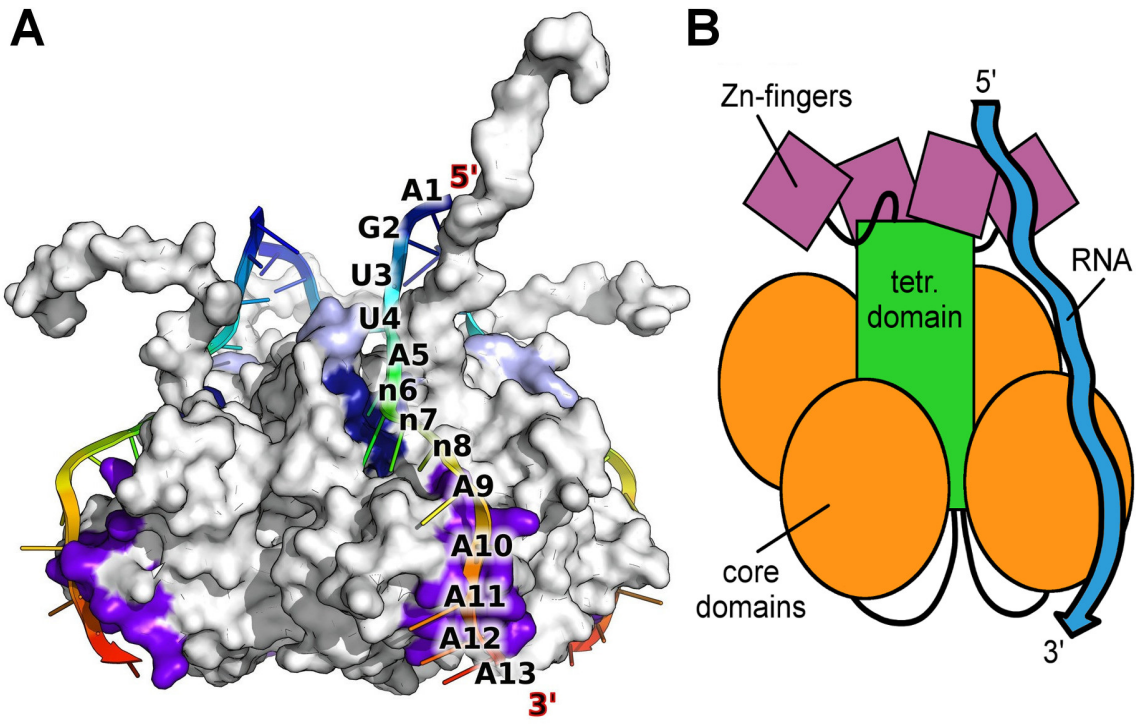


Figure 3.14: Model of Gene-End Recognition by M2-1.
(A) Hypothetical model of gene-end RNA (nucleotides labelled) bound to M2-1. The 5'-end of the gene-end sequence binds to the Zn-finger, whilst the poly-A stretch at the 3'-end binds to the core domain. This arrangement is sterically possible only in the closed state. Surfaces corresponding to amino acids located on the core domain (purple), the Zn-finger (dark blue), and those neighbouring the Zn-finger (light blue) which have been identified to interact with RNA are coloured. (B) Schematic representation of gene-end RNA bound to closed-state M2-1, highlighting the overall domain organisation. (A) taken from Leyrat et al. 2014b.

The open conformation of the core domain and the dynamic behaviour of unbound M2-1 observed in the Kratky analysis may be important features of the protein's function. The proposed model suggests how M2-1 would bind to two spatially separated RNA signals via two binding sites. The flexibility which the opening and closing confers on the distance between the Zn-fingers and core domains may allow for rapid scanning and binding rates of target sequences via a mechanism termed "fly-casting" (Shoemaker et al. 2000). Basically, this would allow the protein to sample a greater conformational space and volume prior to binding. An "ideal" sequence may then lock the core domain into the closed state, binding down on the

recognized motifs, which may in turn act as a signal to prevent the transcriptase from terminating and releasing the nascent transcript. Binding of M2-1 to mRNA emerging from the transcriptase may increase the affinity of the transcriptase (which is associated with M2-1) to the nascent RNA, favouring further attachment.

It is important to emphasise that M2-1 binds the polymerase cofactor P and RNA in a competitive manner with the putative binding sites overlapping (Tran et al. 2009; Blondot et al. 2012). As both P and M2-1 form tetramers, however, it is entirely conceivable that some protomers of M2-1 scan the nascent mRNA, while others are kept in proximity of the transcriptase via attachment to P. The close association of P, N-RNA, M2-1 and L within the transcriptase suggests a complex interplay between these viral factors during mRNA-synthesis. A schematic visualising the machinery at a site of RNA synthesis and illustrating the proposed dynamics of M2-1 is shown in **Fig. 3.15**, highlighting the coordinated multiple interactions and functions of the involved components. The schematic shows the P-mediated attachment of the transcriptase to the nucleocapsid template (blue) via the tetrameric P-protein (green). In the illustration P associates with three open-state M2-1 protomers (dark red). One M2-1 protomer (light red) is shown as scanning the nascent mRNA and switched to the closed state.

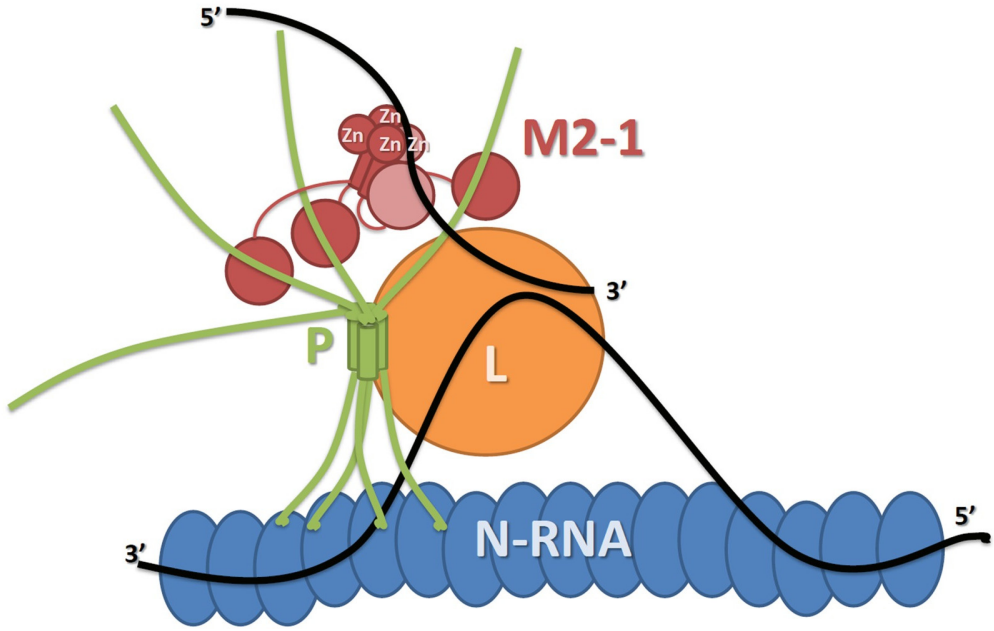


Figure 3.15: Schematic Model of the Pneumoviral Transcriptase.
The nucleocapsid template is shown in blue and is associated with the viral RNA-genome. The genome is proposed to loop out from the nucleocapsid to facilitate access for the L protein (orange). The polymerase co-factor P (green) is attached to the nucleocapsid and tightly associated with L. At the same time P recruits M2-1 (red) using another binding site and thus brings it to the transcription site. M2-1 flexibly scans the nascent RNA emerging from the L protein. Figure taken from (Leyrat et al. 2014b).

4 HMPV P

4.1 Background

As discussed in the Introduction, the polymerase cofactor P (VP35 in *Filoviridae*) is a critical component of the viral replication machinery in all *Mononegavirales*. In paramyxoviruses and rhabdoviruses the P protein is responsible for tethering the viral polymerase L to the nucleocapsid template by attaching to both the nucleoproteins and the L protein (Green and Luo 2009; Ouizougun-Oubari et al. 2015; Tran et al. 2007; Sourimant et al. 2015). For VSV it has been shown that the association with P increases the processivity of L (Morin et al. 2012). P is essential in facilitating access to the template for L and it has been suggested that P assists the L protein in translocating along the nucleocapsid (Cox and Plemper 2015). P also acts as a chaperone for RNA-free N^0 and delivers it to sites of viral replication, as discussed in a previous chapter. In addition, in the case of pneumoviruses like RSV and HMPV, P is also responsible for the recruitment of the aniterminator M2-1 to the transcription complex (Khattar et al. 2001; Derdowski et al. 2008).

4.2 Revisiting HMPV P Domain Organization

As noted earlier, P proteins across different viral families share similarities in their overall architecture, even though sequence identities may be negligible. In general, P proteins are highly disordered, but possess ordered functional domains (Karlin et al. 2003; Leung et al. 2010; Gerard et al. 2009). Additionally, a reoccurring

feature of P proteins are molecular recognition elements (MoREs) which fold-upon-binding of a specific partner. An example of such a MoRE is the N^0-chaperoning region found in P (Renner et al. 2016; Leyrat et al. 2011; Yabukarski et al. 2014). The domain organization of paramyxoviral P proteins has been suggested as early as 2003 by Karlin, et al., and has been described in Chapter 2. For the purposes of discussing construct design I will briefly revisit P architecture.

The N-terminal half of these P proteins is typically intrinsically disordered, with a short stretch of amino acids at the very N-terminus featuring residual alpha-helical propensity (**Fig. 4.1A**) (Karlin and Belshaw 2012). This portion constitutes the N^0-binding region (Galloux et al. 2015; Yabukarski et al. 2014). The C-terminal region of Paramyxoviridae P is associated with the attachment to the nucleocapsid template (**Fig. 4.1A**) (Johansson et al. 2003; Communie et al. 2013b; Kingston et al. 2004, 2008). All studied paramyxovirus P proteins form tetramers via a long twisted coiled-coil domain which is located at the centre of P. Structures of such domains have been solved for measles virus (Communie et al. 2013a), mumps virus (Cox et al. 2013), Nipah virus (Bruhn et al. 2014), and Sendai virus (Tarbouriech et al. 2000).

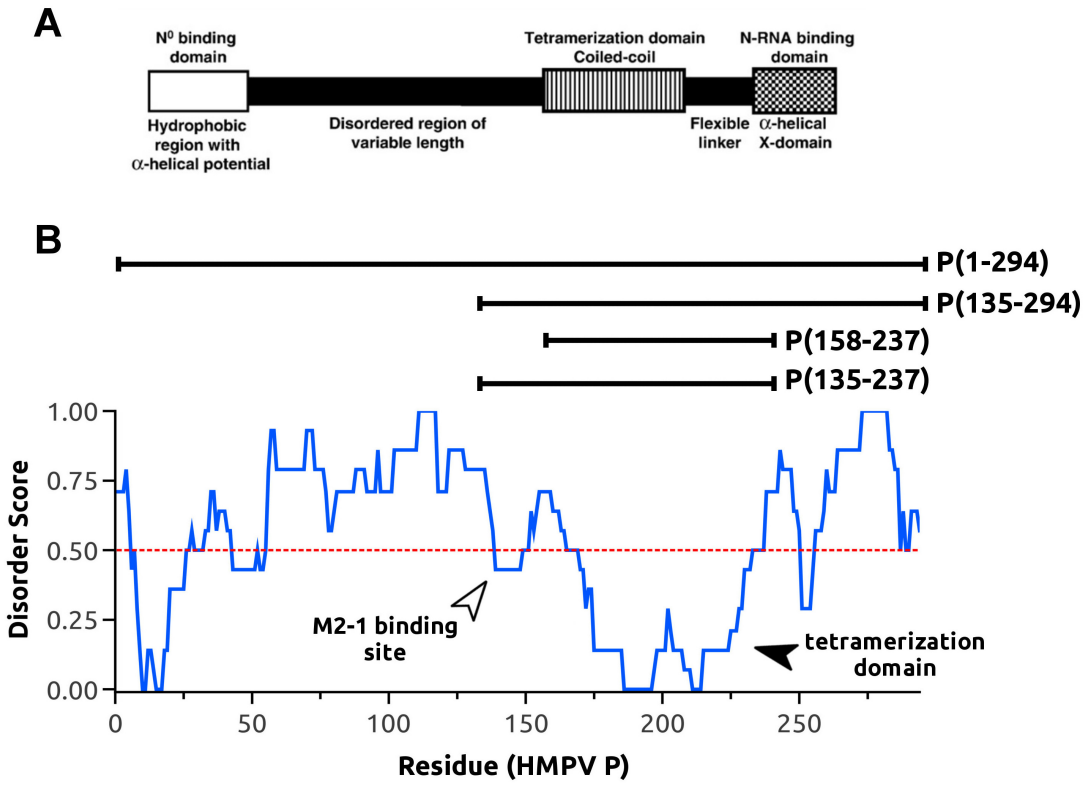


Figure 4.1: Construct-Design to Study P-Oligomerisation and M2-1 Recruitment.
(A) Schematic of the organization of P proteins in Paramyxovirinae (Karlin et al. 2003). (B) Consensus disorder profile of HMPV P generated with the DisMeta server (Huang et al. 2014). The locations of the putative M2-1 binding site and tetramerization domain are indicated. Key constructs used in this work are illustrated by black bars and the boundaries are indicated.

Examination of the disorder profile of HMPV P reveals a predicted ordered region from residues 158-237 which is likely to constitute the tetramerisation domain (**Fig. 4.1B**). Studies with RSV P have identified a proteolysis-resistant fragment ranging from residues 119-160 (roughly corresponding to residues 160-200 in HMPV P), which is consistent with a stable coiled-coil in this region (Llorente et al. 2006, 2008). Furthermore, just before this putative oligomerisation domain there is a small dip in the disorder profile (white arrow, **Fig. 4.1B**). The corresponding region in RSV (residues 100-120) has been associated with M2-1 binding via deletion mutant studies

(Mason et al. 2003). In addition, phosphorylation of Thr108 of RSV P has been shown to abrogate binding to M2-1, further demonstrating that this region is important for M2-1 recruitment (Asenjo et al. 2006).

Taking the above considerations into account, a range of truncated constructs were designed to study P-oligomerisation and interaction with M2-1 (key constructs are indicated as black bars in **Fig. 4.1B**). Targets for recombinant expression were the full-length P protein and the P fragments $P_{(135-294)}$, $P_{(135-237)}$, and $P_{(158-237)}$. The constructs $P_{(135-294)}$ and $P_{(135-237)}$ include the putative M2-1 binding region and were intended for the study the complex between M2-1 and P. The tetramerisation region was retained in both constructs as its truncation has resulted in loss of M2-1 binding in RSV (Mason et al. 2003). The construct $P_{(158-237)}$ includes only the putative tetramerisation domain.

4.3 The M2-1 · P Complex

4.3.1 Reconstitution of the Complex

To form a complex with M2-1, P constructs containing the putative M2-1-interacting region were expressed in bacteria and subsequently purified via Ni-IMAC and SEC filtration using standard methods. Full-length P preparations were problematic as substantial proteolytic degradation was occurring in the expression host, which could not be mitigated by the use of protease inhibitors or by performing all purification steps at 4°C. This is consistent with the disordered/flexible nature and thus high accessibility of P to proteases which may render the protein prone to digestion. However, the construct $P_{(135-294)}$, in which the predicted highly disordered N-terminal half of P is deleted, could be readily expressed and purified.

Complex formation between M2-1 and $P_{(135-294)}$ was complicated by the fact that HMPV M2-1 is prone to aggregation in buffers which do not contain high concentrations of NaCl/NDSB201 as mentioned in the previous chapter. However, at the same time, these additives seemed to inhibit formation of the complex through their solubilising capacity.

To get around this issue the two proteins were initially mixed in a ratio of 1:1 (assuming one tetramer of M2-1 binds one tetramer of P) in a buffer containing high amounts of NaCl to prevent precipitation of M2-1. Then the mixture was carefully and slowly diluted to a NaCl concentration of 150 mM to allow for the controlled formation of the complex. Using this procedure, no visible precipitation occurred. Finally the diluted sample was reconcentrated and subjected to SEC filtration (**Fig. 4.2**) to remove excess binding partners. The resulting chromatogram and

accompanying SDS-PAGE analysis indicated that a monodisperse M2-1·$P_{(135-294)}$ complex had formed. The complex eluted as a single peak centred around an elution volume of 11 ml on a Superose 6 column. Experimental molecular weight (MW) determination of the complex was carried out via SAXS (see below) as MW estimation from gelfiltration elution volumes is very inaccurate for non-globular samples such as this. The shoulder observed in **Fig. 4.2** was due to excess unbound M2-1 as evidenced in the accompanying gel.

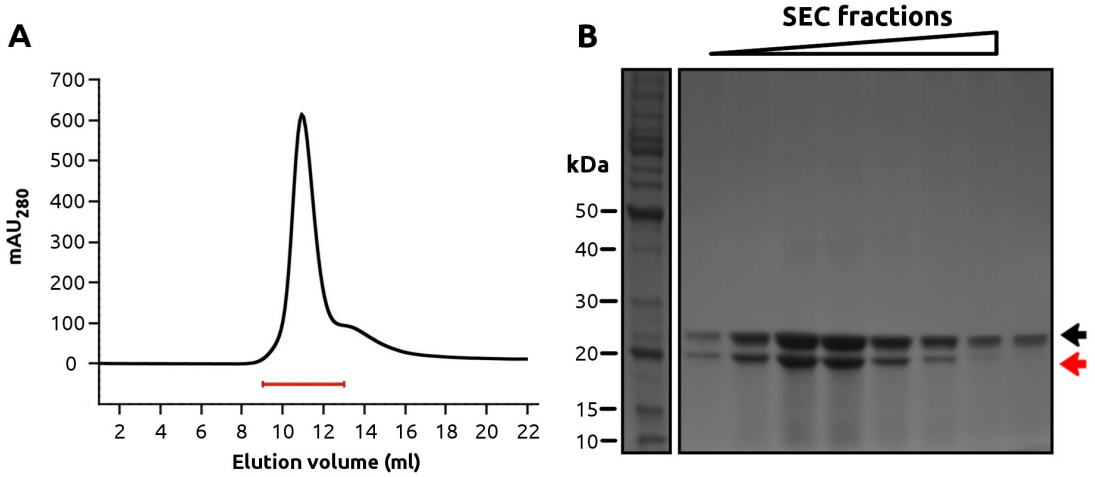


Figure 4.2: Purification of M2-1·$P_{(135-294)}$.
(A) Size-exclusion chromatogram of a mixture of M2-1 and P(135-294) run in buffer with low concentrations of NaCl (150 mM). The elution peak of the complex (red bar) was at ~11 ml on a Superose 6 column. The shoulder after the main peak was assigned to excess M2-1. (B) SDS-PAGE analysis of peak fractions accompanying the chromatogram from (A). The black arrow indicates protein bands corresponding to M2-1, while the red arrow marks $P_{(135-294)}$.

4.3.2 SAXS Characterization of M2-1 · $P_{135-294}$

Solution scattering data of the M2-1·$P_{(135-294)}$ complex were collected at beamline BM29 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France (**Fig. 4.3A**). Analysis of the Guinier fingerprint region (**Fig. 4.3B**) showed that the sample was well-behaved and free of aggregation (Putnam et al. 2007). The

molecular weight of the complex was calculated from the SAXS data via a concentration-independent method described previously (Rambo and Tainer 2013). The experimentally determined molecular weight (MW: 142 kDa) was close to the weight expected from the amino-acid sequences of a complex between a M2-1 tetramer and a $P_{(135-294)}$ tetramer (total MW: 156 kDa), taking into account the expected accuracy of the method (Rambo and Tainer 2013).

Next, a comparative Kratky analysis of the complex and its constituents was performed (**Fig. 4.3C**). The dimensionless Kratky plot of $P_{(135-294)}$ (grey curve, **Fig. 4.3C**) does not display the characteristic bell-shape of folded proteins, but instead possesses a plateau at high q*Rg values, typical for extended intrinsically disordered peptides (Kikhney and Svergun 2015) and consistent with other P proteins. Although M2-1 appears significantly more folded than the P construct, the peak position of its Kratky plot (pink curve, **Fig. 4.3C**) is still far removed from the ideal peak position of a globular scatterer (Guinier-Kratky point, marked by cross-hair) indicating flexible behaviour in solution (see also Chapter 3). The complex of M2-1 and $P_{(135-294)}$ is somewhat less disordered than the P construct on its own (see green curve, **Fig. 4.3C**). This is likely due to the contributions of the folded domains of M2-1. The distance distribution functions P(r) (**Fig. 4.3D**) of both the P fragment and the complex display peaks with extended tails, consistent with the notion that these species sample outstretched conformations (Kikhney and Svergun 2015). Comparison of P(r) of $P_{(135-294)}$ and the complex shows that formation of the complex is accompanied with a shift to a larger species, in terms of sampled volume (**Fig. 4.3D**). Taken together these data show that M2-1 and $P_{(135-294)}$ form an extended and flexible complex with a MW consistent with of a dimer of tetramers (see also **Fig. 4.3E**).

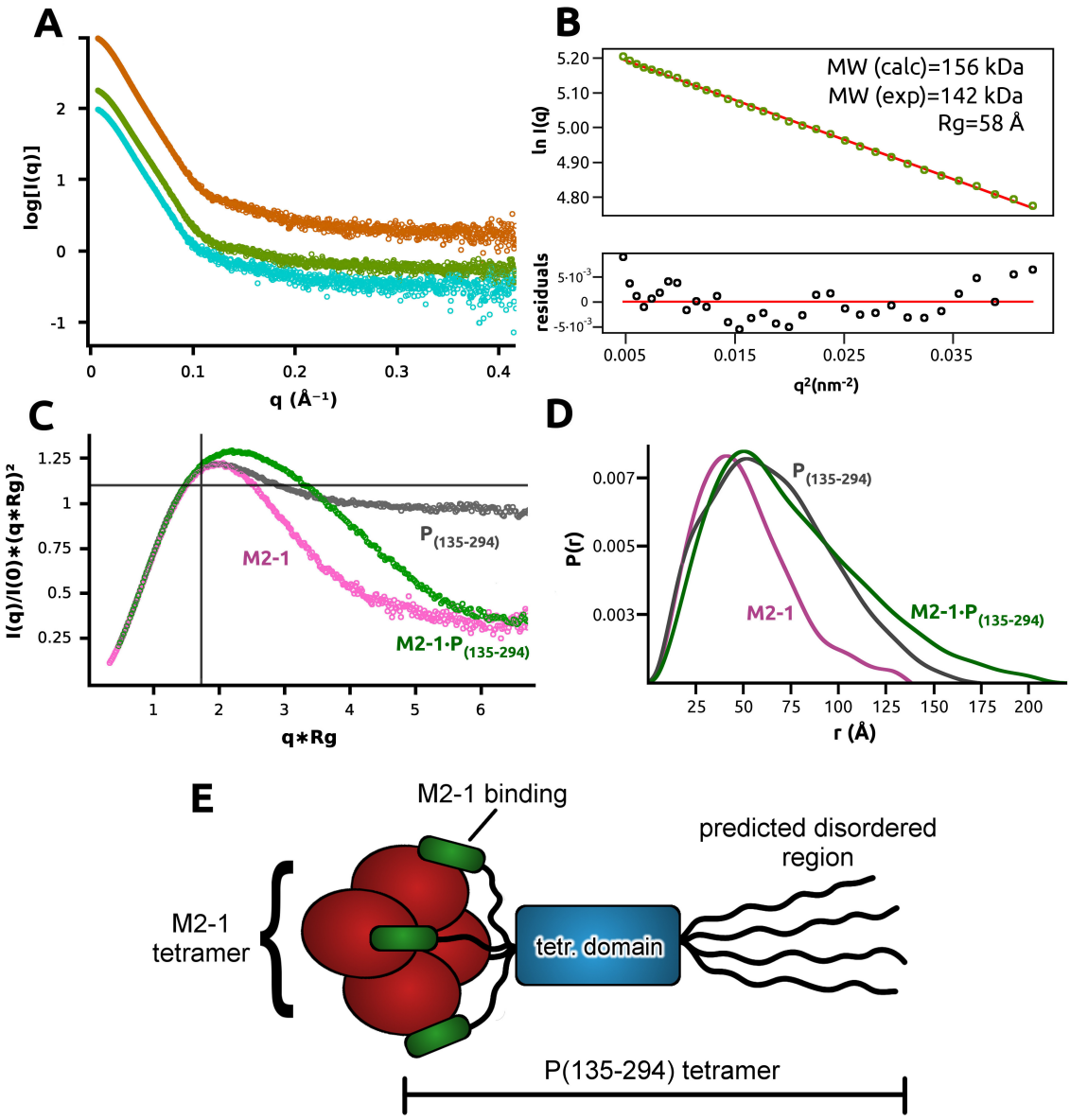


Figure 4.3: SAXS Analysis of the M2-1·P$_{(135-294)}$ Complex.
(A) SAXS scattering profiles of the complex at three different concentrations. (B) Guinier analysis demonstrating linearity in the low-q regime. The molecular weights experimentally determined from the SAXS data (Rambo and Tainer 2013) and calculated from the sequence (assuming four copies of P$_{(135-294)}$ and M2-1, each) are indicated, as is the Rg value. The residuals of the Guinier fitting are displayed at the bottom of the panel. (C) Dimensionless Kratky plots of M2-1, P(135-294) and the corresponding complex as indicated. The cross-hair denotes the Guinier-Kratky point. (D) Pair-wise distance distribution functions P(r) of M2-1, P(135-294) and the corresponding complex as indicated. The P(r) function resembles a histogram of interatomic distances and is relatied to the Patterson function in crystallography. (E) Suggested schematic domain arrangement of the M2-1/P$_{135-294}$ complex. The M2-1 tetramer is shown as four red ellipses. Domains of the P construct are indicated. Please note that the C-terminal predicted disordered region in actuality harbours further interaction sites (not shown for clarity), which are discussed in the next chapter.

4.3.3 Crystallisation Trials

Sitting drop vapour-diffusion crystallisation trials were set up with the M2-1·$P_{(135-294)}$ complex. In addition, due to the high flexibility of the sample observed in SAXS, the corresponding complex was also purified using the shorter P fragment $P_{(135-237)}$, although no high quality SAXS data was recorded with this shorter complex. This construct retains the M2-1 binding site and the putative tetramerisation domain, however, the disordered region at the C-terminus of P is deleted to increase chances of crystal formation (see **Fig. 4.1B**). Unfortunately, no crystals could be obtained for either the M2-1·$P_{(135-294)}$ or M2-1·$P_{(135-237)}$ complexes despite extensive screening. It is likely that the conformational heterogeneity of both the M2-1 protein and the P protein exacerbate crystal formation. It may be necessary to utilise a different strategy in this case, for instance directly linking the M2-1 binding peptide to M2-1 itself as was done for the N^0-P chimera described in Chapter 2. It should be noted, however, that one crystal grew during these trials after an extended duration, which contained a fragment of the P protein only. I will revisit this crystal later in the chapter.

4.4 P Tetramerisation Domain

4.4.1 Purification of $P_{(158-237)}$

In addition to the work carried out on P constructs which interact with M2-1, crystallisation was also attempted for the putative tetramerisation domain only. For this the P fragment $P_{(158-237)}$, that encompasses the predicted ordered region in the centre of P (see **Fig. 4.1B**), was recombinantly produced in bacteria. Because the protein was found to be expressed in the insoluble fraction of the bacterial

preparations a rapid on-column refolding technique was devised and successfully employed during the purifications. Briefly, the bacterial pellet was resuspended in high concentrations (8 M) of the denaturing agent Urea, sonicated, cleared by centrifugation, and the lysate was then loaded onto a Ni-IMAC column. After extensive washes the Urea concentration was rapidly reduced by dilution to facilitate refolding of the protein while bound to the Ni-beads. The sample was then eluted as usual (see **Fig. 4.4**) and subjected to a final round of gelfiltration. A rough estimate indicated a yield of about 10% of recovered, refolded protein.

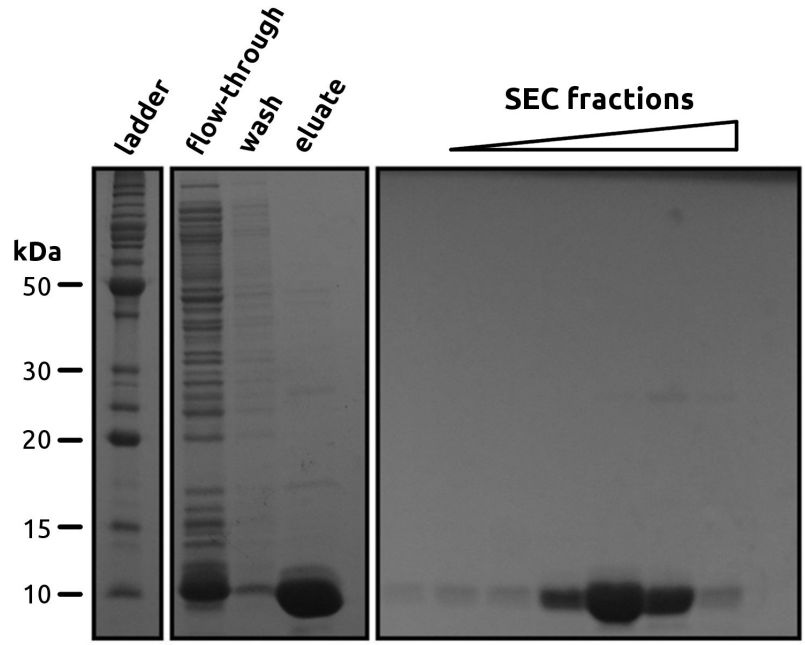


Figure 4.4: Purification of $P_{(158-237)}$.
SDS-PAGE documenting the purification of the central ordered fragment of P. The flow-through, wash, and elution fractions of the Ni-IMAC are shown, as well as consecutive fractions of a subsequent gelfiltration run.

4.4.2 SAXS Characterization of $P_{(158-237)}$

Small angle x-ray scattering data of the refolded central domain of P were collected at beamline BM29 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France (**Fig. 4.5A**). The sample was-well behaved and did not show signs of aggregation even at high concentrations. The radii of gyration of the samples were determined to be around 3.3 nm and the molecular weight was determined via the volume of correlation as implemented in ScÅtter (Rambo and Tainer 2013) to be ~33 kDa. This is in good agreement with the theoretical mass expected for a tetramer (4*8.8 kDa = 35.2 kDa) and confirms that the construct contains the tetramerisation domain. In contrast to the disorder prediction, a Kratky analysis indicated that $P_{(158-237)}$ remained highly flexible (blue Kratky plot, **Fig. 4.5B**) despite deletion of the disordered C-terminal region of P in this construct. Comparison with the longer P construct ranging from residues 135-294 did, however, point to a slightly more folded behaviour of the central domain (which constitutes this shorter construct) (compare blue and beige Kratky plots, **Fig. 4.5B**).

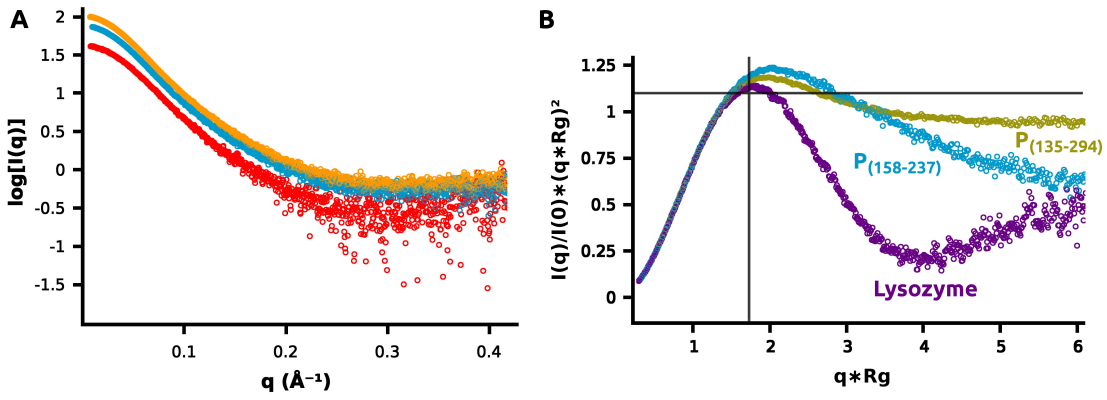


Figure 4.5: SAXS Characterization of the Central Domain of HMPV P.
(A) SAXS profiles of HMPV $P_{(158-237)}$ at three different concentrations. (B) Dimensionless Kratky plots of $P_{(158-294)}$ and the longer construct $P_{(135-294)}$ showing the deviation from ideal behaviour of a folded scatterer. The Kratky curve of a lysozyme standard is shown in comparison to highlight the difference between the disordered P constructs and the bell-shaped curve of a globular particle. The cross-hair denoted the Guinier-Kratky point.

4.4.3 Crystallisation, Structure Solution, and Refinement

Preparations of $P_{(158-237)}$ were concentrated to around 10 mg/ml and crystallisation trials were carried out. Crystals formed in one condition (25 %w/v polyethylene glycol 3350, 0.100 M HEPES, pH 7.5, 20°C) after 88-142 days (**Fig. 4.6**, the biggest crystal was approximately 60 μm large) indicating that degradation was necessary prior to crystallogenesis. Subsequently, crystals were cryo-protected with 25% glycerol and cooled to liquid nitrogen temperature. Diffraction data were collected at 100 K at Diamond Light Source, Didcot, UK, beamline I04-1. Data were autoprocessed up to 3.1Å with XIA2 (Winter et al. 2013).



Figure 4.6: Crystal of P_{core}-1. Crystal obtained with the construct $P_{(158-237)}$ after 88-142 days.

Solving the phase problem proved problematic in this case as crystal growth required extensive amounts of time and was not reproducible, precluding the production of further crystals for experimental phasing. In addition, the structure was likely to be composed of a helical coiled-coil arrangement, which is, paradoxically, known to be challenging to solve with molecular replacement in absence of a optimal search model (reviewed in Thomas et al. 2015). One of the reasons for this is that elongated coiled-coils often show tight lateral packing which makes it difficult to distinguish self and cross Patterson vectors. For a detailed discussion of the problem see Thomas et al. 2015. Specific approaches have been developed to phase coiled-coils using thousands of structures generated ab-initio as search models (Bibby et al. 2012; Thomas et al. 2015). The phasing of this crystal was performed in a similar manner, but guided by experimental SAXS data, by Dr Cedric Leyrat and is described

in detail in Leyrat et al. 2013. Briefly, Dr Leyrat utilised Rosetta fold-and-dock (Das et al. 2009) to generate tens of thousands of tetrameric models of the central region of HMPV P. Then theoretical SAXS profiles of these models were calculated and compared with the experimental SAXS data, yielding a goodness of fit value χ_{exp} (Svergun et al. 1995). Using one of the ten best-fitting ab-initio models, a high translation function Z score MR solution could be identified for the crystallographic dataset with PHASER (McCoy et al. 2007). Additionally, to eliminate steric clashes that violated the packing criteria, the search model needed to be truncated somewhat. Finally, the structure was refined with COOT (Emsley and Cowtan 2004) and BUSTER (Smart et al. 2012), making use of the eight-fold NCS (LSSR as implemented in BUSTER, the ASU contained two tetramers, i.e. eight monomers) and TLS parameters. The resulting structure (denoted as P_{core}-1) contained a much shorter-than-expected tetramerisation domain of HMPV P and will be discussed further below.

As noted earlier in this chapter, the crystallisation trials carried out with the $P_{(135-237)}$·M2-1 complex also yielded a crystal (**Fig. 4.7**, roughly 80 µm large) which grew only after a long period of time (between 135 and 182 days, reservoir condition: 25 %w/v polyethylene glycol 3350, 0.2 M NaCl, 0.1 M HEPES, pH 7.5). Diffraction data for this crystal were collected at 100 K at Diamond Light Source, Didcot, UK, beamline I03 and could be autoprocessed up to 1.9 Å with XIA2 (Winter

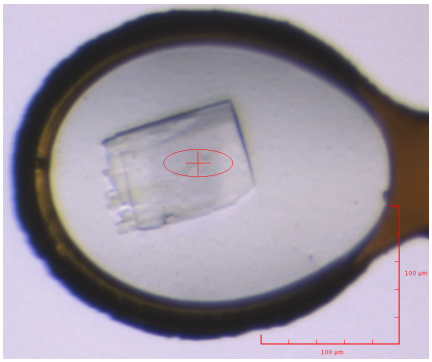


Figure 4.7: Crystal of P_{core}-2. Crystal obtained in trials with the complex $P_{(135-237)}$·M2-1. The crystal only contained a short fragment of HMPV P.

et al. 2013). The crystal could be phased by MR using the refined structure of P_{core}-1, and refinement was then performed with COOT (Emsley and Cowtan 2004) and PHENIX (Adams et al. 2010). making use of TLS parameters and eight-fold torsion-angle NCS restraints as implemented in PHENIX. Although the space group was different, this structure (denoted as P_{core}-2) was highly similar to the tetramerisation domain seen in P_{core}-1 (discussed further below). The data collection and refinement statistics are given in Table 7.

Data Collection	**P_{core}-1**	**P_{core}-2**
Wavelength (Å)	0.92	0.98
Number of images	1800	3600
Ω Oscillation per image (°)	0.2	0.1
Exposure time per image (s)	0.2	0.02
Beam size (μm)	30x30	50x20
Transmission (%)	100	50
DLS beam-line	I04-1	I03
Detector	Pilatus 2M	Pilatus 6M
Resolution (Å)	66.9-3.1	18.86-1.9
Space group	$P\ 2_12_12_1$	$P\ 2_1$
Cell dimensions		
a, b, c (Å)	66.9, 48.5, 64.45	28.9, 110.1, 38.4
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 95.6, 90.0
R_{merge}	0.35 (1.35)	0.10 (1.44)
I / σI	6.4 (1.4)	9.9 (1.2)
CC (1/2)	0.99 (0.32)	1.00 (0.62)
Completeness (%)	0.90 (0.57)	0.99 (1.00)
Redundancy	10.5 (6.1)	6.5 (6.1)
Measured Reflections	37, 981	121, 986
Unique Reflections	3, 602	18, 805
Refinement		
R_{work} / R_{free}	23.5/25.2	18.5/23.1
RMSD		
Bond lengths (Å)	0.01	0.02
Bond angles (°)	1.29	1.56
Ramachandran plot quality		
Favoured (%)	96.7	100
Allowed (%)	1.1	0.0
Outliers (%)	2.2	0.0

Table 7: Data Collection and Refinement Statistics for HMPV P_{core}.
Data were collected at Diamond Light Source (DLS), Didcot, UK.
Values in parentheses correspond to the highest resolution shell
RMSD, root mean square deviation from ideal geometry

4.4.4 Structure of the HMPV P Tetramerisation Domain

The oligomerisation domain observed in the P_{core}-1 structure consists of an unexpectedly short tetrameric coiled-coil (**Fig. 4.8A** and **B**). The asymmetric unit of the crystal contained two copies of the tetrameric coiled-coil oligomerisation domain, orthogonally packing against each other (i.e. there were 8 monomers in the ASU). From the P residues 158-237 which formed the construct used in the crystal trials, residues 158-167 and 195-237 were not visible. This can be attributed to degradation, as there would be no room for these regions in the crystal packing. This is also consistent with the prolonged time required for crystallogenesis as mentioned above. Furthermore, this also confirms that the majority of $P_{(158-237)}$ is disordered and flexible in solution, as observed in the Kratky analysis (**Fig. 4.5B**). Consequently, extensive pruning of the flexible residues, down to the stable coiled-coil core, needed to take place in the sitting drop before crystal contacts could form.

While it is somewhat unfortunate that the P_{core}-2 crystal did not contain a complex with M2-1, it nonetheless provides a large improvement in resolution as compared to the P_{core}-1 crystal (from 3.1 Å to 1.9 Å). In addition it was obtained without following a refolding protocol during the purification and thus validates that the renatured P structure obtained before is correct. The fact that both P fragments subjected to crystallisation trials ($P_{(135-237)}$ and $P_{(158-237)}$) degraded down to the same short region of residues prior to crystal growth further solidifies the notion that this core represents the sole folded domain in the centre of HMPV P.

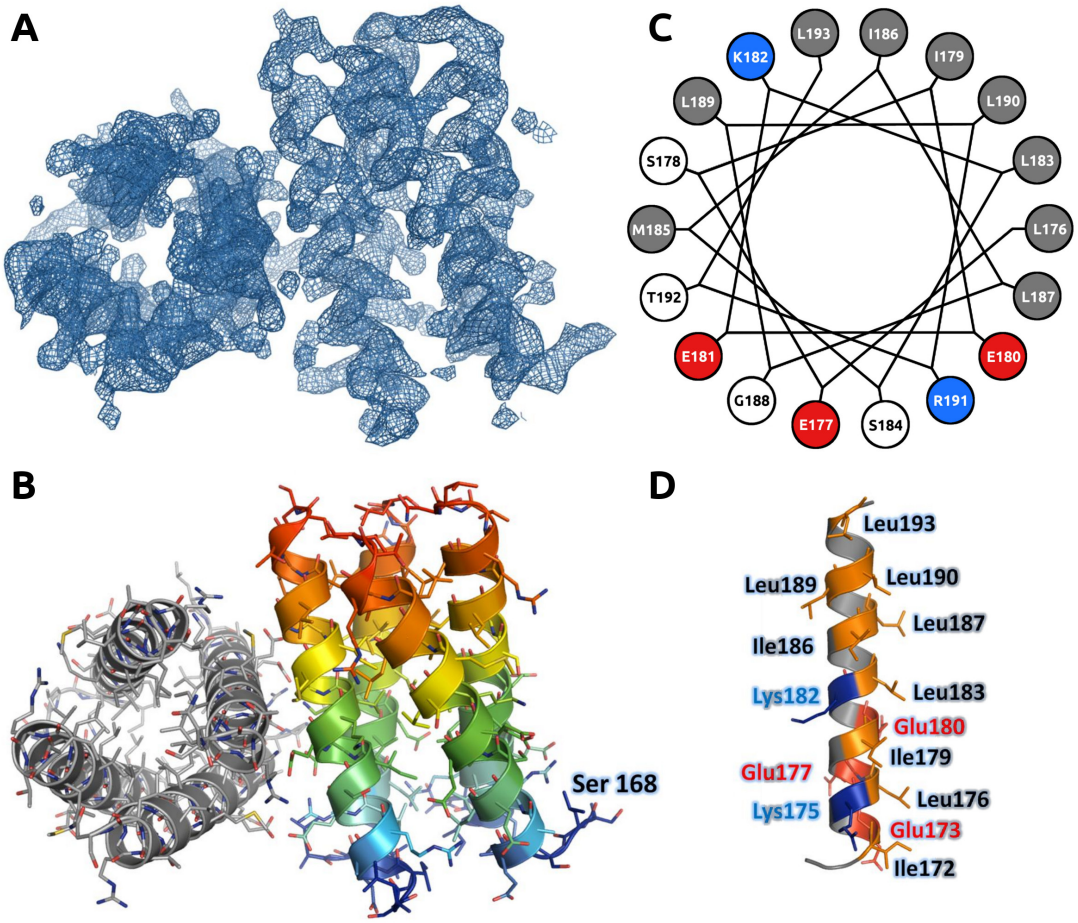


Figure 4.8: Structure of P_{core}-1.
(A) $2F_O - F_c$ electron density map of P_{core}-1 at 3.1 Å resolution, contoured at 1.4σ. (B) Cartoon representation of the two tetramers in the asymmetric unit, with side chains shown as sticks. One of the tetramers is coloured from N-terminus to C-terminus (blue to red, respectively), while the second is shown in grey. (C) Helical wheel representation of P residues 176 to 193, highlighting the clustering of hydrophobic residues (grey circles) at the oligomerisation interface. Negatively charged and positively charged residues are indicated as red or blue circles, receptively. (D) Cartoon representation of a single helix of P_{core}-1. Hydrophobic residues are shown as orange sticks, while negatively and positively charged residues are coloured in red or blue, respectively. Figure adapted from Leyrat et al. 2013.

Like P_{core}-1, the P_{core}-2 crystal contains two tetramers in the asymmetric unit (**Fig. 4.9**), although these pack against each other differently. The RMSDs between Pcore-1 and Pcore-2 are between 0.44 and 0.75 Å. Like other coiled-coils (Ramos and Lazaridis 2011), the tetramers of P_{core} are stabilised by a network of interdigitating hydrophobic residues at the interface, consisting mainly of Leu and Ile (**Fig. 4.8C** and

D). The interior of the tetramer is lined by the residues Leu176, Leu183, Leu187, Leu189, Leu190, Leu193, Ile172, Ile179, and Ile186. In addition, salt-bridges are formed between neighbouring monomers, involving solvent-exposed residues on the exterior of the assembly, specifically Glu173, Glu177, Arg175, Glu180, and Lys182. The shared total buried surface area of the tetramer is 4800 Å, with a predicted dissociation energy of ~21 kcal/mol (Krissinel 2010). This interaction surface is considerably less extensive than what is observed in the longer *Paramyxovirinae* P oligomerisation domains (Bruhn et al. 2014; Communie et al. 2013a; Tarbouriech et al. 2000). For instance, the value for the Nipah virus tetramerisation domain is between 15, 000 to 20, 000 Å with dissociation energies of 140 to 200 kcal/mol (Bruhn et al. 2014).

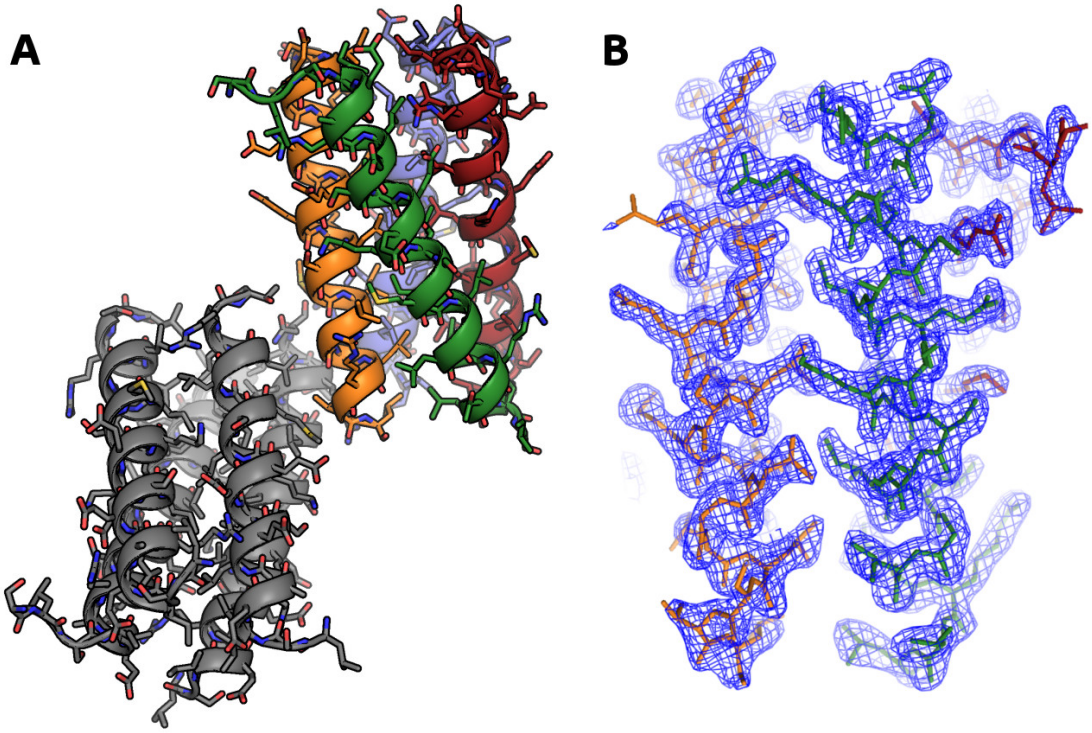


Figure 4.9: Structure of P_{core}-2.
(A) The two tetramers of P_{core}-2 contained in the asymmetric unit are displayed in cartoon representation. Side-chains are shown as sticks. One tetramer is coloured in grey while, in the second, each protomer is coloured individually. (B) Sample of the $2F_O$-F_c electron density map of P_{core}-2, displayed at a contour level of 1.5σ. The structure modelled into the density is shown in stick representation.

4.5 Discussion

The polymerase cofactor P is a multifunctional and essential component of the HMPV transcription machinery. By bringing together the RdRP-containing L protein, the nucleocapsid-template, and the M2-1 protein, P serves as an assembly-hub during viral RNA synthesis (Cox and Plemper 2015). P proteins possess a modular organization in which functional domains are separated by intrinsically disordered regions which do not conform to a single rigid fold but, instead, fluctuate spatially

and temporally, sampling conformational space (Dyson and Wright 2005; Dunker et al. 2008; Tompa 2011).

In the results presented here I have described the *in vitro* formation of a complex between HMPV M2-1 and fragments of P. The SAXS characterization of $P_{(135-294)}$ bound to M2-1 revealed a highly flexible assembly with a MW consistent with a dimer of tetramers. Analysis of the distance distribution function P(r) and corroboration with the Kratky curve indicated that the complex samples a large volume of space. Not unexpectedly, crystallisation attempts of M2-1 complexes with P fragments did not yield a structure of the constituents bound to each other. It is likely that the structural elucidation of the binding mode of HMPV P to M2-1 will require fine-mapping of a P-peptide with the highest affinity to M2-1. Binding studies may be complicated, however, by the low solubility of M2-1 in the absence of additives (NDSBs, high salt concentrations), which may in turn hamper binding. As described in this chapter, the M2-1·P complex does only appear to form at salt concentrations below 150 mM. To determine a structure of M2-1 bound to P I instead propose a strategy in which different parts of the M2-1 binding region of P are attached to M2-1 via a poly(GlyGlySer) linker to increase the naturally-occurring binding. The linker length can by chosen using the previously determined P-binding region on the M2-1 core domain as a guide (Blondot et al. 2012). These attempts can be made with full-length tetrameric M2-1 or, instead, with only the core-domain linked to a peptide of P. Work on M2-1-P chimeras is currently ongoing.

The Kratky analysis of the predicted central ordered region (residues 158 to 237) of HMPV P revealed an unexpectedly low level of foldedness. Crystallisation trials required extended durations and the degradation of the construct down to a short

folded core. This was consistent with crystallisation attempts carried out with the complex of M2-1 and a longer construct of P, further confirming that P_{core} (residues 168 to 195) is the only folded portion in the centre of P. In contrast, disorder prediction indicates that P residues beyond P_{core} (from 200 onwards, see **Fig 4.1B**) are also folded, whilst they were degraded in the presented crystal structures. Interestingly, the same residues in BRSV have been shown to be responsible for binding to N-RNA (Khattar et al. 2001). It is tempting to speculate that this extension of P_{core} constitutes a MoRE which folds-upon-binding to N-RNA, similar to the N^0-binding MoRE at the N-terminus of P.

The tetramerisation regions of the paramyxoviral P proteins from Sendai, Nipah, measles, and mumps virus show long helical coiled coils (**Fig. 4.10**). While most of these are arranged in a parallel orientation, with the C- and N-termini of the protomers on the same side, the mumps virus tetramerisation domain features a antiparallel orientation (Cox et al. 2013). The HMPV P oligomerisation domain described here is a parallel coiled-coil which is significantly shorter than those of *Paramyxoviridae*. Indeed, in terms of extent, it is comparable to the dimerisation domain of rabies virus P (Ivanov et al. 2010) which consists of a four-helix bundle formed by two helical hairpins (**Fig. 4.10**). A structure-based phylogenetic tree constructed with SHP (Abrescia et al. 2012), using the known multimerisation regions of P proteins clearly separates the the *Paramyxoviridae* and *Rhabdoviridae*, placing them at opposite ends (**Fig. 4.10**). Interestingly, the *Pneumovirinae* HMPV P protein clusters somewhat apart from the *Paramyxoviridae*, similar to what is found in traditional phylogenetic trees (see also **Fig. 1.1**).

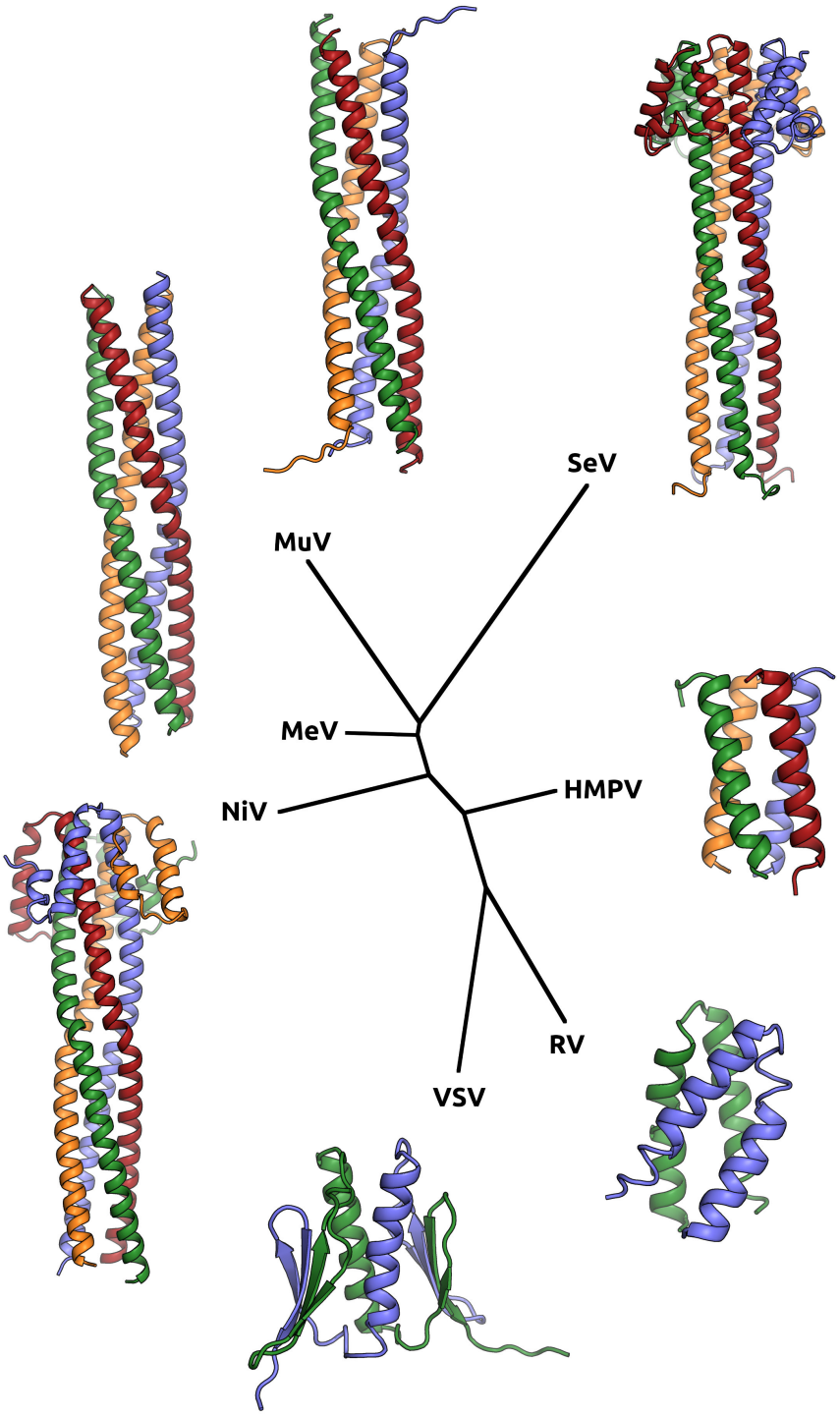


Figure 4.10: Phylogenetic Tree of P Multimerisation Domains.
The tree was calculated using SHP (Abrescia et al. 2012). The structures of the multimerisation domains are shown in cartoon representation and coloured by chain. The virus abbreviations are: *HMPV*, human metapneumovirus (pdb ID:4bxt), *MeV*, measles virus (pdb ID: 3zdo), *MuV*, mumps virus (pdb ID: 4eij), *NiV*, Nipah virus (pdb ID: 4n5b), *RV*, rabies virus (pdb ID:3l32), *SeV*, Sendai virus (pdb ID:1ezj), *VSV*, vesicular stomatitis virus (pdb ID:2fqm).

5 CONCLUDING REMARKS AND FUTURE WORK

5.1 The Numerous Interactions of Pneumovirus P

The core pneumoviral transcriptase/replicase possesses two multifunctional protein components, the L protein, which harbours the RdRP, and the polymerase cofactor P, which allows the complex to access the viral template (Fearns and Deval 2016; Morin et al. 2013). The RNA template is never naked in the host cytoplasm, but is instead sheathed in a polymer of nucleoproteins (N-RNA), providing protection from host-nucleases and anti-viral signalling (for instance via the RNA helicase retinoic acid-inducible gene I, RIGI) (Bakker et al. 2013; Tawar et al. 2009; Ruigrok et al. 2011b). By attaching to both the N-RNA template and the L protein, P serves as a hub, bridging components of the viral RNA-synthesis machine (Khattar et al. 2001; García-Barreno et al. 1996; Slack and Easton 1998; Tran et al. 2007). In addition, to allow for non-abortive synthesis of viral mRNAs during transcription mode (**Fig. 5.1A**), P also needs to recruit the antiterminator protein M2-1 (Collins et al. 1995; Fearns and Collins 1999; Hardy and Wertz 1998). In replication mode (**Fig. 5.1B**), however, P supplies the growing progeny (anti)nucleocapsid with fresh N^0, which is used for the immediate encapsidation of nascent (anti)genome. The oligomeric nature of P proteins is essential to coordinate these tasks, allowing it to simultaneously contact its many interaction partners (Schneider et al. 2004; Chen et al. 2006; Kolakofsky et al. 2004). By possessing extended intrinsically disordered segments separating the functional domains (Karlin et al. 2003), P can sample large volumes of space and thus can rapidly bring together all viral factors required RNA-synthesis.

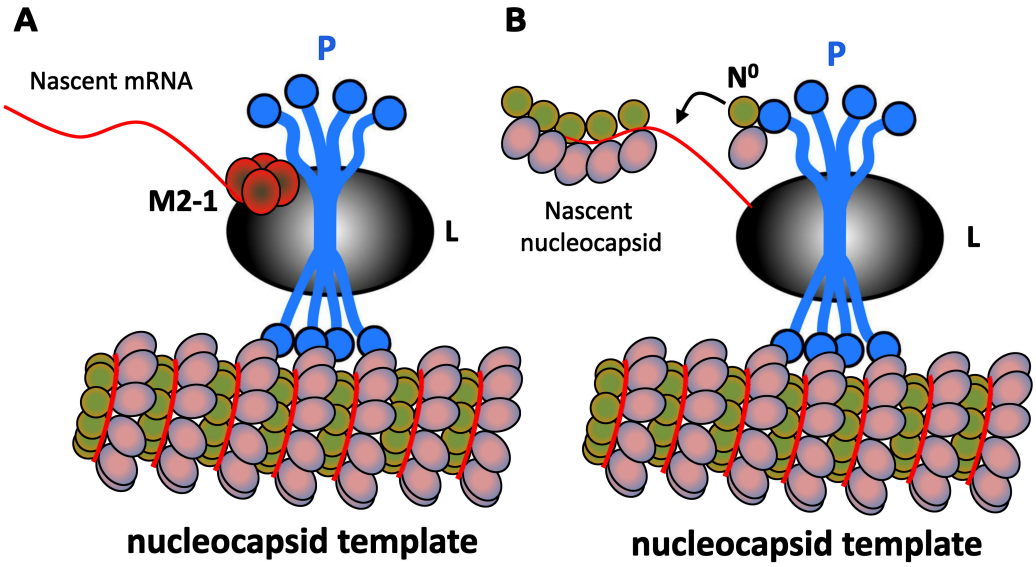


Figure 5.1: Schematic models of Pneumoviral RNA-Synthesis Machinery.
Schematic representations of the assembled transcriptase (A) and replicase (B) components. Note that in (A) the antiterminator M2-1 (coloured in red) is recruited to the polymerase in transcription mode and scans nascent viral mRNA. In replication mode (B) nascent genomic or antigenomic RNA is co-transcriptionally packaged in nucleoproteins.

In the results presented herein I have described the structures of two HMPV P domains. The first is the N-terminal N^0-chaperoning region bound to the CTD of the nucleoprotein (Chapter 2). The second constitutes the P tetramerisation domain (Chapter 4). Both structures have not been determined for any member of *Pneumovirinae* previously and are, due to the close relatedness, also relevant for the major human pathogen RSV. In addition, a solution characterization of an *in vitro* reconstituted complex of M2-1 and P has been carried out. In light of the general conservation of P protein architecture, the structural evidence from this work, and functional data published earlier, it is possible to produce a functionally annotated multiple sequence alignment for pneumoviral P proteins (**Fig. 5.2**). To directly accompany the MSA I have mapped the functional P domains onto a disorder prediction, schematically showing the putative domain architecture for pneumovirus P proteins, especially HMPV (**Fig. 5.3**).

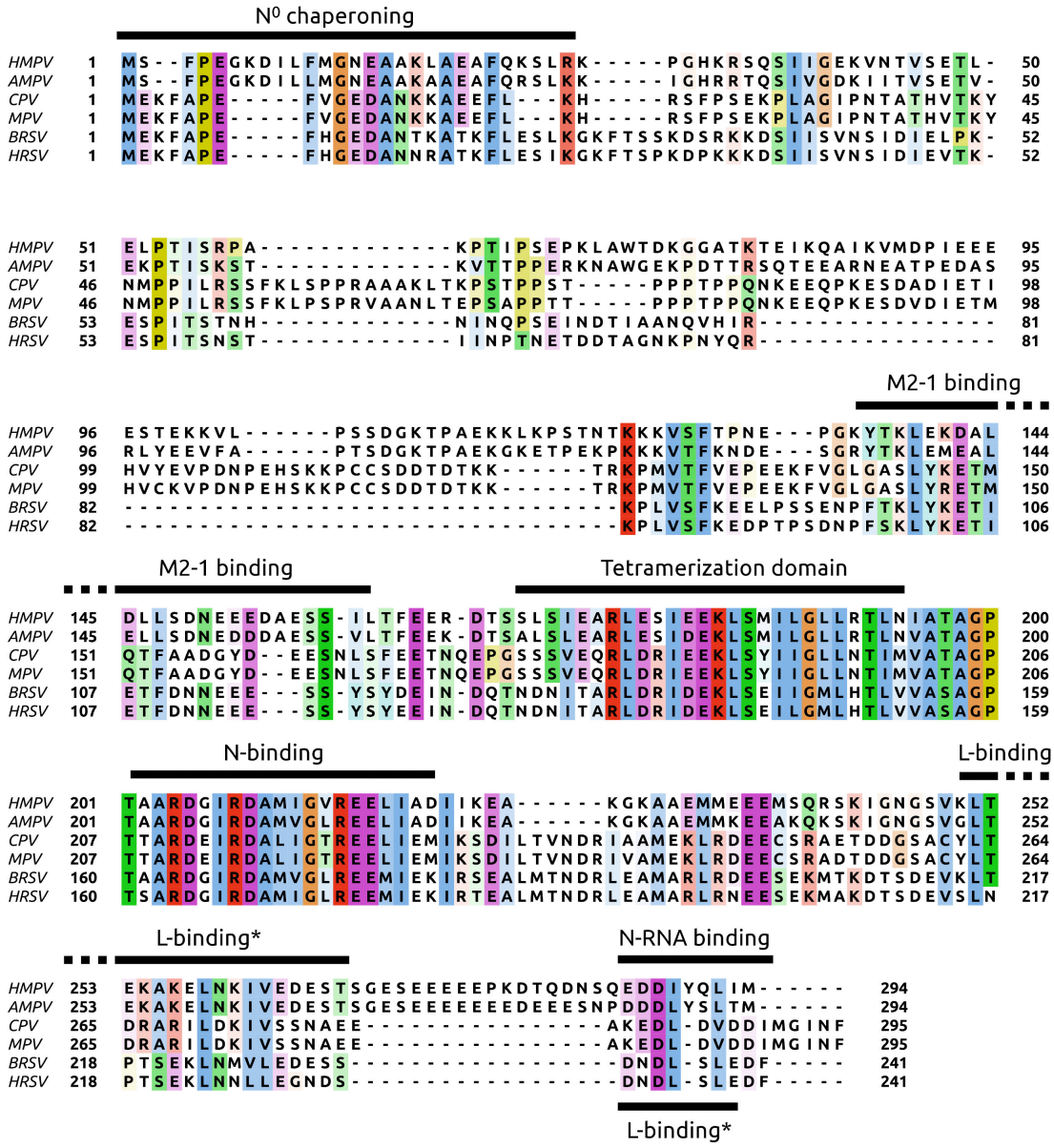


Figure 5.2: MSA of P Proteins from *Pneumovirinae* Family Members.
Annotated sequence alignment of P proteins from the pneumovirus subfamily members human metapneumovirus (HMPV), avian metapneumovirus (AMPV), canine pneumonia virus (CPV), murine pneumonia virus (MPV), bovine respiratory syncytial virus (BRSV), and human respiratory syncytial virus (HRSV). The sequences were aligned using PROMALS3D (Pei et al. 2008). The L-binding site (marked by asterisk) partially overlaps with the main N-RNA binding site. The annotations are further explained in the accompanying text. The functional domains (e.g. tetramerisation domain, N^0-chaperoning domain) are considerably more conserved than the disordered regions which act as linkers (e.g. between N^0-chaperoning domain and M2-1 binding domain). The variability of the disordered domains reveals that these are under less sequence constraints.

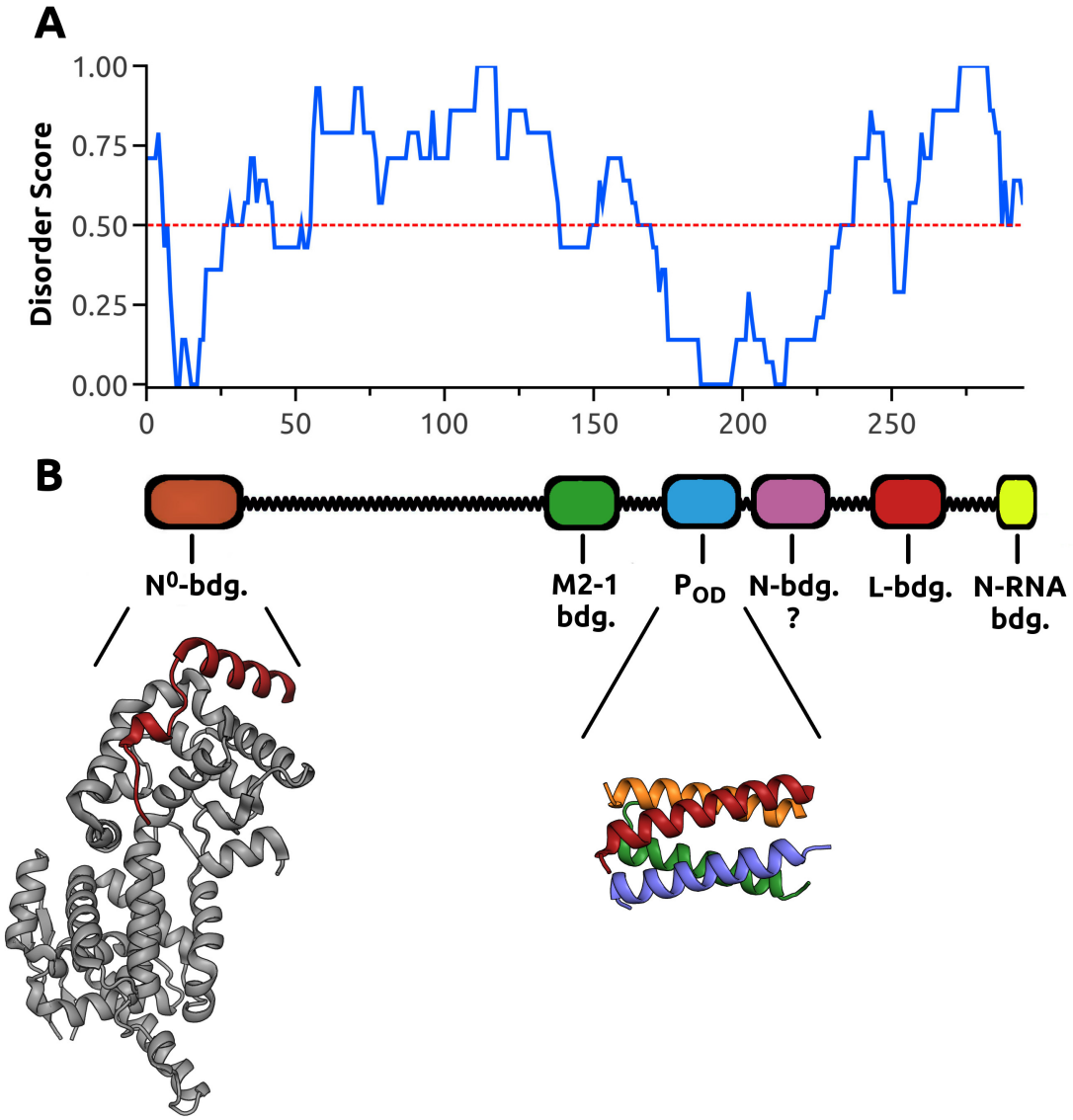


Figure 5.3: Schematic of HMPV P Domain Organization.
(A) Disorder metaprediction of HMPV P residues with low scores corresponding to ordered regions (Huang et al. 2014). (B) Schematic of HMPV P domains on approximately the same scale as the disorder prediction shown above. Structures that have been solved for pneumoviruses are shown in cartoon representation (both from this work). The P-protein portion of the N^0-P complex is coloured in red, while the N is shown in grey. Although the N-RNA binding portion of RSV P (yellow box in B) has been structurally characterised, it is omitted in the figure as only two residues were visible in the density (Ouizougun-Oubari et al. 2015). The domain annotation is discussed in detail in the accompanying text.

A conserved N^0-chaperoning region is found at the very N-terminus of P-proteins as determined in the N^0-P structure from Chapter 3 (**Figs. 5.2** and **5.3**). This is in also in line with previous bioinformatics studies and functional studies (Karlin and Belshaw 2012; Galloux et al. 2015). The region is then followed by an extended non-conserved area (HMPV residues ~30-135) which is likely to serve as an unstructured flexible linker, consistent with disorder metapredictions (**Fig. 5.3**). The residues responsible for recruiting the M2-1 protein (~135-158, green box in **Fig 5.3**, Mason et al. 2003; Asenjo et al. 2006; Blondot et al. 2012) are yet to be structurally characterised. Owing to the flexibility inherent in the M2-1 protein and the disorderedness of P, this is likely to be a highly challenging task. In addition, a tetrameric P protein seems to be essential for efficient binding to M2-1 (Mason et al. 2003). As discussed earlier I propose that more drastic construct engineering may be required for the crystallographic analysis of the M2-1 and P interaction.

The coiled-coil tetramerisation domain is located just after the putative M2-1 binding site (blue box in **Fig. 5.3**) and is highly conserved within all *Pneumovirinae* (see MSA in **Fig. 5.2**), indicating the more general relevance of the structure presented herein. Interestingly, an ordered region (HMPV residues 195-237) of high sequence identity between all aligned viruses is found just downstream of the tetramerisation domain (purple box in **Fig. 5.3**). We initially assumed that this area was part of a much longer coiled-coil, but later found it to be degraded in the presented crystal structures. Although the main N-RNA binding site is at the C-terminus of P (Tran et al. 2007), the residues 195-237 might also be involved in contacting the nucleoprotein, as has been shown for BRSV (corresponding BRSV residues 160-180, Mason et al. 2003). It is tempting to speculate that this constitutes a

MoRE which folds upon binding to N in collaboration with the classical N-RNA binding site composed of the nine C-terminal residues (yellow box in **Fig. 5.3**). In the future it would be interesting to test if a P construct encompassing this putative site is able to bind to N-RNA rings, potentially precipitating further structural investigations.

The interaction between the P protein and the large L polymerase is fundamental for replication and transcription, however, remains little characterised. Indeed, co-expression of the "right" fragment of P together with L may constitute a key factor in the stabilization and thus structural elucidation of L (Liang et al. 2015; Chattopadhyay and Banerjee 2009; Sourimant et al. 2015). This is because P has been shown to induce conformational changes in L and the addition of a fragment of P was essential in the structural elucidation of VSV L (Liang et al. 2015). The right fragment of P may be able to arrest L in a particular conformation, reducing the heterogeneity of the sample. A recent study for RSV which utilised serial deletions, point mutations and polymerase activity assays provided extensive evidence for the location of the L-binding region of P (Sourimant et al. 2015). The authors concluded that the site is located from RSV residues 216-239 (partially overlapping with the N-RNA site, see **Fig. 5.2**, also putatively assigned for HMPV, red box in **Fig. 5.3**) and constitutes a MoRE with some alpha-helical propensity. Importantly, mutagenesis of the RSV residues L216, L223, and L227 was found to reduce the polymerase activity by up to 90%. The multiple sequence alignment shown above demonstrates that the same residues are strictly conserved within *Pneumovirinae*. The determination of the structure of pneumoviral L would constitute an enormous gain in knowledge and P is likely to be essential in achieving this. These considerations highlight that there is still much to be learned about the structural biology of the multifunctional P protein.

5.2 Similarities of Filo- and Pneumoviruses

Several similarities exist between the *Filoviridae* and the *Pneumovirinae* subfamily of paramyxoviruses. For instance, there are numerous structural and functional similarities between the filovirus-specific VP30 and the pneumovirus-specific M2-1 proteins (see **Fig. 5.4**) (Blondot et al. 2012; Tanner et al. 2014; Leyrat et al. 2014b). As with M2-1, efficient transcription of filoviral mRNAs requires the presence of the RNA-binding VP30 (Mühlberger et al. 1999). VP30 also possesses a CCCH-type Zn-finger which is thought to be involved in the interaction with RNA (Modrof et al. 2003; John et al. 2007). Finally, VP30 multimerises via an oligomerisation helix downstream of the Zn-finger and features a core-domain which can be confidently aligned to that of M2-1 (Z-score of 5.7 using the DALI server, Blondot et al. 2012). No further homologues of M2-1/VP30 can be identified within *Mononegavirales*.

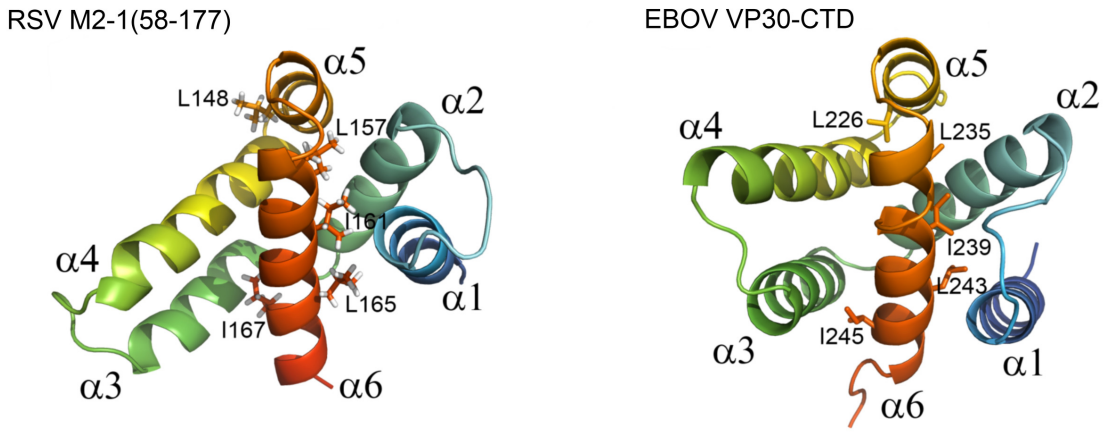


Figure 5.4: Similarity Between VP30 and M2-1.
Annotated structures of the RSV M2-1 and Ebola virus core domains shown in cartoon representation. The structures have been aligned using the DALI server (Holm and Rosenström 2010). Conserved hydrophobic residues are shown as sticks. The figure is taken from Blondot et al. 2012.

Curiously, although *Pneumovirinae* particles are usually classified as being pleomorphic, there is evidence that the most infectious virions are in fact filamentous, as are those of *Filoviridae (Liljeroos et al. 2013)*. Other members of the *Paramyxoviridae* are usually considered to form spherical particles (Loney et al. 2009), although filamentous virions have also been reported for parainfluenzaviruses (Yao and Compans 2000). Pneumovirus particles with filamentous morphologies have been previously observed and analysed by cryo-ET (Liljeroos et al. 2013; Kiss et al. 2014). The studies concluded that the presence and assembly of M proteins underneath the membrane layer correlated with the cylindrical virus morphology. Indeed, it has been suggested that the transformation of RSV virions from a native filamentous form to a more spherical form is actually caused by purification-dependent perturbation of the matrix protein layer (Liljeroos et al. 2013; Kiss et al. 2014). Furthermore, Marburg viruses imaged at late time-points postinfection possessed increasing numbers of spherical rather than filamentous particles, which also showed lower infectivity (Welsch et al. 2010). Consistent with these observations, previous studies have noted similarities between the matrix proteins of these viruses (Money et al. 2009; Leyrat et al. 2014c).

In line with the above observations, phylogenetic trees and sequence analyses reported in the literature (e.g. Feldmann et al. 1993) often place the *Pneumovirinae* close to the *Filoviridae* and on a separate branch apart from the *Paramyxovirinae* (see **Fig. 1.1**). Such mononegavirus phylogenies are often constructed using the sequences of the most highly conserved proteins, namely the L and N proteins. However, N- and L-based trees are often inconsistent, with marked differences in the clustering depending on which of the two genes is used. An example for this can be found in

Mihindukulasuriya et al. 2009. Here, the *Pneumovirinae* branch is located close to the *Paramyxovirinae* when the authors carried out the alignments using L sequences. However, using the amino acid sequences of N resulted in a very different tree, in which the *Pneumovirinae* are instead placed closer to the *Rhabdoviridae* than to their paramyxovirus relatives (Mihindukulasuriya et al. 2009).

I hypothesised that these inconsistencies reflect inaccuracies in the alignment rather than true phylogenetic differences. With the abundance of new N structures having been reported recently (including that of HMPV from this work), it is now possible to construct a fairly comprehensive structure-guided tree. The PROMALS3D server allows input of both sequences and structures and improves global alignment accuracy by utilising restraints based on structural alignments (Pei et al. 2008). A nucleoprotein-based phylogenetic tree including all families of *Mononegavirales* was calculated with alignments from PROMALS3D using available N-sequences and structures (**Fig. 5.5**). This N-based tree is now highly consistent with phylogenies obtained using sequences of the L-protein (e.g. **Fig 1.1** or Mihindukulasuriya et al. 2009) and places the *Pneumovirinae* close to the *Paramyxovirinae*. In addition, it further emphasises the relatedness between *Pneumovirinae* and *Filoviridae* and demonstrates how supplementing structural information can improve upon the accuracy of alignment results (Balaji and Srinivasan 2007).

The observations discussed above are fully consistent with changes in the classification of the *Pneumovirinae* subfamily presented by the International Committee on the Taxonomy of Viruses (ICTV) during the late stages of the preparation of this thesis (Afonso et al. 2016). This taxonomic update revised the status of pneumoviruses to that of an independent viral family, the *Pneumoviridae*.

The elevation to a *Mononegavirales* family was justified by the conclusion that pneumoviruses are as closely related to *Filoviridae* as they are to paramyxoviruses (Afonso et al. 2016). The structural clues presented herein, which point to the distinctness of pneumoviruses in respect to the paramyxoviruses, are thus in-line with the recent reclassification.

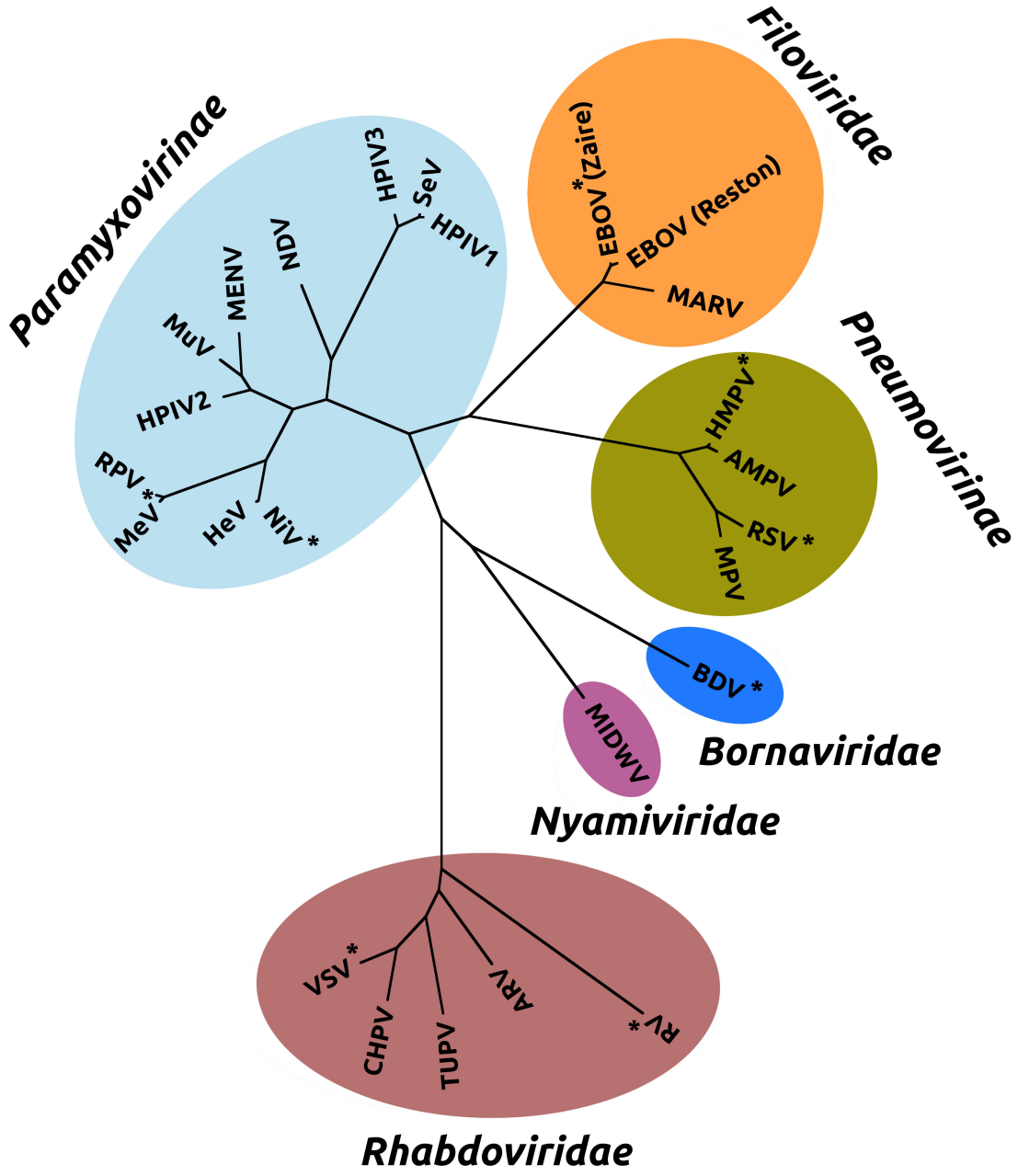


Figure 5.5: Nucleoprotein-Based Phylogenetic Tree of Mononegavirales.
Nucleoprotein sequences and structures were aligned using PROMALS3D (Pei et al. 2008). A maximum-likelihood phylogenetic tree was then calculated from the alignment using MEGA6 (Tamura et al. 2013). Asterisks mark viruses for which structural data was used to restrain the alignment. The virus abbreviations are: *AMPV*, avian metapneumovirus, *ARV*, Adelaide River virus, *BDV*, Borna disease virus, *CHPV*, Chandipura virus, *EBOV*, Ebola virus, *HeV*, Hendra virus, *HMPV*, human metapneumovirus, *HPIV*, human parainfluenza virus, *MARV*, Marburg virus, *MENV*, menangle virus, *MeV*, Measles virus, *MIDWV*, Midway virus, *MuV*, mumps virus, *MPV*, murine pneumonia virus, *NDV*, Newcastle disease virus, *NiV*, Nipah virus, *RPV*, Rinderpest virus, *RSV*, respiratory syncytial virus, *RV*, rabies virus, *SeV*, Sendai virus, *TUPV*, Tupaia rhabdovirus, *VSV*, vesicular stomatitis virus.

5.3 Future Directions in Nucleocapsid Research

High-resolution structural work on isolated components of the replication/transcription machinery has significantly enhanced our understanding of the mononegavirus life-cycle. At the same time cryo-electron tomographic studies of intact virus particles have shown the organization of their constituents (Loney et al. 2009; Liljeroos et al. 2013; Kiss et al. 2014). Fluorescent microscopy has demonstrated how the components of pneumoviral replication machine organise into cytoplasmic inclusion within infected cells (Li et al. 2008a; Ghildyal et al. 2002). However, to unravel the interplay between these components, large higher-order assemblies need to be further examined. Recent high-resolution studies on helical filaments of measles virus nucleoprotein have demonstrated the unprecedented power of modern cryo-EM detectors and image processing software in tackling such projects (Gutsche et al. 2015). In the study by Gutsche et al., modern single particle-based helical reconstruction procedures were used to obtain a 4.3 Å structure of rigidified, measles nucleocapsids. Similarly, another group has utilised cryo-EM to examine how the mumps virus P protein is able to bind to authentic nucleocapsids and facilitate their uncoiling (Cox et al. 2014). Cox et al. showed how addition of a N-terminal region of mumps virus P induced conformational changes in authentic nucleocapsids. Furthermore, binding of a C-terminal domain of mumps P yielded additional density in helical reconstructions, providing clues on how P associates with the nucleocapid.

Sample preparation of mononegavirus L protein still remains a major hurdle for structural studies. Production of L often is complicated by very low expression levels and frequent aggregation. Fortuitously, some recent reports describe successful

preparations of L (Liang et al. 2015; Sourimant et al. 2015). Specifically, RSV L could only be successfully expressed if P was co-expressed with it (Noton et al. 2012). It is a truly exciting prospect that the *in vitro* reconstitution of L-P complexes onto nucleocapsids for structural investigations might be possible in the near future. Combined with novel localised reconstruction techniques (Ilca et al. 2015) and classification procedures (Bai et al. 2015) this may finally yield a complete picture of the holo-replicase enzyme. Continuing on from there even higher order structures could be examined. For instance, in *Pneumovirinae*, the M2-1 protein has been reported to also serve structural purposes, coating the nucleocapsids and mediating between the nucleocapsids and the matrix layer (Liljeroos and Butcher 2013; Li et al. 2008a; Kiss et al. 2014). Gradually the field is inching forward to a holistic understanding of the structural virology of pleomorphic viruses.

6 MATERIALS AND METHODS

6.1 Molecular Biology and Microbiology Methods

6.1.1 Preparation of Solutions and Media

Buffers and solutions were sterilised either by autoclaving at 121°C or by filtering through a 0.22 µm filter (Merck-Millipore). Bacterial growth media, glass- and plasticware were autoclaved at 121°C prior to use. Luria-Broth (LB) and Terrific Broth (TB) media were prepared from ready-to-use powder as per the manufacturer's instructions (Thermo-Fisher) (Kram and Finkel 2015; Bertani 1951).

6.1.2 Cultivation and Storage of *E. coli*

Overnight cultures of *E. coli* were grown in LB medium supplemented with the appropriate antibiotics (see Table 8) while shaking (160-180 rpm) at 37°C. Sterilised antibiotics stocks were prepared at 1000x of the working concentrations and were stored at -20°C. For long-term storage, *E. coli* strains were kept as glycerol stocks. To this end, 30% (v/v) glycerol were added to overnight cultures and the bacteria were then flash-frozen and kept at -80°C.

Antibiotic	Concentration (µg/ml)
Ampicillin	100
Kanamycin	50
Chloramphenicol	34

Table 8: Antibiotics used in this work.

6.1.3 Polymerase Chain Reaction

Amplification of target genes and corresponding truncated fragments was performed by PCR (Mullis and Faloona 1987; Saiki et al. 1985). Template DNA for HMPV genes (strain NL1-00, A1) was kindly provided by Prof Ron Fouchier and Dr Bernadette van den Hoogen. Primer oligonucleotides were ordered from Eurofins MWG Operon (Ebersberg, Germany). When performing In-Fusion molecular cloning with the pOPIN family of vectors, primers were designed with the 5'- and 3'-extensions appropriate for the specific plasmid in question (see Bird et al. 2014). The Phusion® High-Fidelity DNA polymerase formulated as 5x PCR master mix (New England Biolabs) was used throughout this work. Typically, 50 µl PCR reactions were prepared. A representative thermal cycling protocol is shown below.

Step	Temperature	Duration
1. Initialization	95°C	1:30 min
2. Denaturation	95°C	30 s
3. Annealing	T_A°C*	30 s
4. Elongation	72°C	30 s per kb
5. Final Elongation	72°C	5 min
6. Final Hold	10°C	∞

Table 9: Representative PCR Protocol.
Steps 2 to 4 are repeated 32 to 36 times.
*Optimal annealing temperature for primer pair

6.1.4 Agarose Gel Electrophoresis

Agarose gel electrophoresis was utilised to monitor the success of various steps during the molecular cloning of HMPV constructs. Gels were prepared by adding 1% (w/v) of agarose powder to TAE-buffer (Tris-acetate-EDTA) and heating

the mixture until dissolved. SYBR® Safe DNA gel stain (Thermo-Fisher) was added before the gel had cooled and set. Analyte DNA was loaded onto the gel and separated by electrophoresis for ~30 min at 100 V, followed by visualisation of the bands on a UV transilluminator at a wavelength of 302nm.

6.1.5 In-Fusion Ligation Independent Cloning

Molecular cloning was carried out using the proprietary, ligation independent In-Fusion® system (Clontech), as per instructions of the manufacturer, and with expression vectors of the pOPIN family, developed by the Oxford Protein Production facility (Bird et al. 2014; Berrow et al. 2007). Briefly, target DNA was amplified by PCR using primers with appropriate extensions (Bird et al. 2014). After quality control by agarose gel electrophoresis and subsequent band visualisation, amplified DNA was purified using a PCR Purification Kit (QIAGEN). In addition, pOPIN vectors were linearised with the appropriate restriction enzymes (see Bird et al. 2014). Restriction digests were performed as per the instructions on the manufacturer's website (New England Biolabs, https://www.neb.com/). Linearised plasmids were purified via PCR Purification Kit (QIAGEN), as before.

Lyophilised enzyme from the In-Fusion® Dry-Down PCR Cloning Kit (Clontech) was dissolved in a mixture of purified insert and vector (2:1 molar ratio, respectively, in 10 µl of deionised water). The reactions were incubated in a thermal cycler for 15 min at 37°C and then for 15 min at 50°C. Subsequently, the mixture was diluted with 40 µl of TE (Tris-EDTA) buffer and 2.5 µl of this dilution were used for the transformation of chemically-competent *E. coli*.

6.1.6 Transformation of Chemically-Competent *E. coli*

Pre-aliquoted chemically-competent *E. coli* were stored at -80°C. Competent cells were gently thawed on ice and pipetted into sterile 10 ml culture tubes prior to supplementation with plasmid DNA. A minimum amount of 50 ng of plasmid DNA was added to an aliquot of 25 µl of cells and these were then incubated on ice for 5 min. Subsequently, the cells were subjected to a heat shock for 45 s at 42°C using a water bath and the tubes were quickly placed back into ice. After further incubation on ice for 5 min, 200 µl of SOC-medium (Super Optimal broth with catabolite repression) were added to the cells. To establish resistances, the *E. coli* cells were incubated at 37°C while shaking (180 rpm) for 1 hour and, finally, plated onto LB-agar plates with the appropriate antibiotics.

6.1.7 Preparation of Plasmid-DNA from *E. coli*

Plasmid DNA was prepared from overnight cultures of freshly-transformed *E. coli*. Typically, a single colony of a transformation LB-agar plate was used to inoculate 5 ml of LB medium with the appropriate antibiotics. The culture was then grown overnight at 37°C while shaking (180 rpm). On the subsequent day, plasmid DNA was extracted via the QIAprep® Miniprep Kit (QIAGEN). The integrity of all inserts was confirmed by Sanger-sequencing (Source Bioscience).

6.2 Protein Production

6.2.1 Recombinant Expression in *E. coli*

For recombinant expression of HMPV constructs, the corresponding pOPIN vectors containing the sequence-verified inserts were transformed into Rosetta2™(DE3)pLysS *E. coli* cells as described above. The cells were spread onto LB agar plates supplemented with appropriate antibiotics for selection and grown over night at 37°C. Colonies were subsequently used directly to inoculate precultures in LB medium or to create glycerol stocks for later use. Expression cultures (in TB medium with appropriate antibiotics and 1% glucose) were started from overnight precultures at an OD_{600} of roughly 0.1. The expression cultures were grown at 37°C while shaking (160 rpm) to an OD_{600} of 0.8 to 1.0 and then they were cooled to 18°C. Protein synthesis was induced by addition of 1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) to the cultures. Following induction the cultures were left shaking at 18°C for 18 to 24 hours and cell pellets were then harvested by centrifugation (18°C, 20 min, 4000x g). The pellets were either processed immediately or flash-frozen and stored at -20°C for later use.

6.2.2 Purification of Recombinant Proteins from *E. coli*

<u>N^0-P Hybrid:</u>

The chimeric HMPV N^0-P (N_{1-394} - P_{1-40}) construct was expressed as a C-terminally $(His)_6$-tagged fusion protein from a pOPINE vector (Bird et al. 2014) as described above. Expression culture pellets were resuspended in 20 mM Tris, pH 7,

1M NaCl. The cells were lysed by sonication and subsequently the lysate was cleared by centrifugation (4°C, 45 min, 50000 x g). The supernatant was loaded onto a column containing pre-equilibrated Ni^{2+}-nitrilotriacetic acid (NTA) agarose (QIAGEN). After extensive washes the recombinant protein was eluted with 20 mM Tris, pH 7, 1M NaCl, 300 mM imidazole. The N^0-P hybrid was then buffer exchanged by careful dilution into 20 mM Tris, pH 7, 100 mM NaCl to lower the salt concentration. The sample was then loaded onto a HiTrap Heparin HP column (GE Healthcare) and subsequently eluted via a NaCl gradient (from 150 mM NaCl to 1 M NaCl). Finally, a gel-filtration step was carried out using a Superdex 75 column (GE Healthcare) and 20 mM Tris, pH 7, 100 mM NaCl as running buffer.

N-RNA:

Full-length HMPV N was expressed as a C-terminally $(His)_6$-tagged fusion protein from a pOPINE vector (Bird et al. 2014) as described above. Pellets of expression cultures were resuspended in 40 ml of 25 mM Tris, pH 8, 1 M NaCl per l of culture. The resuspended cells were then lysed by sonication. The lysate was cleared by centrifugation (4°C, 45 min, 50000 x g) and loaded onto pre-equilibrated Ni^{2+}-NTA agarose (QIAGEN). After extensive washes, the protein was eluted in 25 mM Tris, pH 8, 1 M NaCl, 400 mM imidazole. The sample was further polished by gel-filtration using a Superose6 10/300 column (GE Healthcare) equilibrated in 25 mM Tris, pH 8, 1 M NaCl. N-RNA preparations were then buffer exchanged into 25 mM Tris, pH 8, 150 mM NaCl, 500 mM NDSB201, 50 mM Arginine via a PD10 column (GE Healthcare).

M2-1:

Full-length HMPV M2-1 was expressed as a N-terminally $(His)_6$-tagged fusion protein from a pOPINF vector (Bird et al. 2014) as described above. Pellets of the expression cultures were resuspended in 20 mM Tris, pH 7.5, 1 M NaCl and subsequently lysed by sonication. To clear the lysate the sample was centrifuged for 45 min at 4°C and 50000×*g*. The cleared lysate was loaded onto pre-equilibrated Ni^{2+}-NTA Agarose (QIAGEN). Contaminants were removed by extensive washing of the beads. Finally, the protein was eluted with 20 mM Tris, pH 7.5, 1 M NaCl, 400 mM imidazole. This was followed by a polishing step via SEC, using a S200 column equilibrated in either 20 mM Tris, pH 7.5, 1 M NaCl, or in 20 mM Tris, pH 7.5, 300 mM NaCl, 1 M NDSB-201. For crystallisation, the $(His)_6$-tag was cleaved by proteolysis with 3C-protease. To further improve the purity, the 3-C digested sample was subjected to reverse Ni^{2+}-IMAC, removing the His-tagged 3C protease. This was followed by an additional gel-filtration step as above. The protein could be concentrated to up to 10 mg/ml using a Millipore spin-filter when in presence of 500 mM or more of NDSB-201. Otherwise, precipitation could be observed at concentrations beyond ~1mg/ml.

$P_{(135-294)}$/$P_{(135-237)}$/$P_{(158-237)}$:

P constructs were expressed with an N-terminal $(His)_6$-tag from a pOPINF vector (Bird et al. 2014) as described above. The resulting pellets were gently resuspended in 20 mM Tris, pH 7.5, with 500 mM NaCl and lysed by sonication. In the case of $P_{(158-237)}$ the lysis buffer was supplemented with 8 M Urea. The lysate was subsequently centrifuged for 45 min at 4°C and 50000×g. The cleared lysate was then

filtered using a pore size of 0.22 µm (Millipore) and loaded onto pre-equilibrated Ni^{2+}-NTA Agarose (QIAGEN) and several washes were carried out. For $P_{(158-237)}$ the bound recombinant protein was refolded on the beads by rapid dilution of the Urea. The recombinant proteins were eluted in Tris, pH 7.5, 150 mM NaCl using a step gradient of up to 400 mM imidazole. The proteins were finally subjected to size exclusion chromatography on a S200 column equilibrated in 20 mM Tris, pH 7.5, 150 mM NaCl.

6.2.3 Production and Purification of N-RNA Filaments in Insect Cells

Sf21 cells were transferred from a suspension culture to a 6-well plate ($1x10^6$ cells/well) and allowed to adhere for 20-30 minutes. The suspension-culture medium (S-CM; Sf900II, supplemented with 3.5% FCS and 100 U of a Penicillin/Streptomycin antibiotics mix) was replaced with plating medium (SF900II medium supplemented with 1.5% FCS; 2.5 ml/well). 0.5 µg of the pOPINF vector containing the HMPV N gene and 0.15 µg of the linearised bacmid (donated by Prof Ian Jones, Reading University, Zhao et al. 2003) were combined in 100 µl Sf900II, to which the transfection reagent solution was added (100 µl of Sf900II containing 8 µl Cellfectin II, Life Technologies). Following gentle mixing, the DNA/Cellfectin II solution was kept for 20-30 minutes at room temperature, before adding it drop-wise to a well containing Sf21 cells. The 6-well plate was incubated at 28°C for 5 hours, the medium was then replaced with 2 ml of S-CM, and the plate put back at 28°C for 72 hours. 1 ml of the medium (now containing the recombinant baculovirus) was used to inoculate a 2L-flask containing 40 ml a Sf21 suspension culture (1.2 x10 6 cells/ml), and the virus was allowed to amplify for 76 h (28 °C, 135 rpm). The

amplified virus was clarified (1000x g, 5 min, 20 °C), and 5 ml of it was added to 300 ml Sf21 suspension cultures (1.2 x10^6 cells/ml) for expression for 72 hours. The cultures were then centrifuged for 5 minutes at 1000xg and the resulting pellets were either processed immediately or frozen for later use.

Pellets were resuspended in 25 mM Tris, pH 8.0, 50 mM NaCl, 1 % Triton and lysed by vortexing rigorously. The lysate was cleared by centrifugation for 45 minutes at 5000xg. and the supernatant was loaded onto a HiTrap Heparin HP column (GE Healthcare). Recombinant N protein was eluted in 25 mM Tris, pH 8.0 using a step gradient of up to 2 M NaCl. The protein was further purified by size exclusion chromatography using a Superose 6 column (GE Healthcare) equilibrated in 25 mM Tris, pH 8.0, 250 mM NaCl.

6.2.4 SDS-PAGE

The purity of protein preparations was monitored via denaturing sodium dodecyl sulphate polyacrylamide gel-electrophoresis (SDS-PAGE). To denature the proteins these were incubated with Laemmli sample buffer for 10 min at 95°C (Laemmli 1970). The denatured samples were loaded onto 12% or 15% polyacrylamide gels and separated by electrophoresis using tris-glycine buffer in a BioRad Mini-PROTEAN apparatus until the dye running front reached the end of the gel. The BenchMark™ Protein Ladder (Thermo-Fisher) was run alongside the analytes to estimate molecular weights. Protein bands were visualised by staining with SimplyBlue™ SafeStain (Thermo-Fisher).

6.3 Structural Biology Methods

6.3.1 Crystallisation

Crystallisation trials were carried out using a 96-well, vapour-diffusion, sitting drop format. The full range of crystallisation screens available at The Division of Structural Biology, Oxford, UK was used (obtained from Hampton Research, Emerald BioSystems and Molecular Dimensions Limited). Typically, screening was performed by mixing 100 nl of reservoir solution with 100 nl of protein solution in the central shelves of CrystalQuick 96-well Greiner plates using a Cartesian Technologies robotics system (Genomic Solutions) (Walter et al. 2005). Reservoirs contained 95 µl of mother liquor.

In general, initial crystallisation conditions were improved by carrying out optimization-dilution screens (the mother liquor concentrations were screened in diluted form in 3% steps from 67% to 100%). In addition, mother liquor to sample ratios were varied. This work also made use of the Silver Bullets and Silver Bullets Bio additive screens from Hampton Research (McPherson and Cudney 2006). More detailed information on crystallisation and sample conditions can be found in the relevant Chapters.

6.3.2 Cryo-Protection

To cryo-protect crystals (if necessary), solutions of mother liquor, supplemented with 25% glycerol (v/v) were carefully pipetted on top of the crystallisation drop. The crystal was then mounted and rapidly flash-cooled by plunging into liquid nitrogen.

6.3.3 Crystal Ligand Soaking

To soak nucleotides and oligonucleotides into crystals, solutions of mother liquor were supplemented with 50 mM of mononucleotides or 10 mM of oligonucleotides and 25% glycerol. These solutions were then carefully pipetted onto the crystallisation drops. The crystals were then periodically monitored to assess if any morphological changes occurred. Crystals were harvested after different incubation times and rapidly flash-cooled by plunging into liquid nitrogen.

6.3.4 Data Collection and Processing

All crystal diffraction data were collected at Diamond Light Source, Didcot, UK. Diffraction data were auto-processed with XIA2 (Winter et al. 2013). In general, programs of the PHENIX (Adams et al. 2010) and CCP4 (Winn et al. 2011) suites were used for structure solution, refinement and analysis. In addition, extensive use was also made of AUTOBUSTER (Smart et al. 2012) for structure refinement. Manual model-building was performed with COOT (Emsley and Cowtan 2004). More detailed information, including used beamlines and software, and crystallographic tables, can be found in the relevant sections in Chapters 2 to 4.

6.3.5 Small-Angle X-ray Scattering (SAXS)

SAXS solution scattering data were collected on beamline BM29 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France and on beamline B21, Diamond Light Source, Didcot, UK. At BM29, ESRF, samples were kept at 20°C, and data were collected using a wavelength of 0.0995 nm and a sample-to-

detector distance of 1 m. Scattering curves were generated and buffer subtraction was carried out by the automated processing pipeline available at the beamline (Pernot et al. 2010).

At B21 measurements were carried out using a sample-to-detector distance of 4.09 m at a wavelength of 1 Å and samples were kept at 16°C. The in-house beamline processing pipeline (GDA) performed beam-current corrections and normalizations of the recorded images. Prior to data merging and buffer subtraction, individual images were checked for radiation damage and, if necessary, discarded using the program ScÅtter (http://www.bioisis.net).

Subsequent analyses of the SAXS data was carried out with programs of the ATSAS suite (Petoukhov et al. 2012) or with ScÅtter. For M2-1 samples, the pairwise distance distributions P(r) were obtained with the program GNOM (Svergun 1992). Otherwise, P(r) curves were calculated with ScÅtter. Molecular weights were obtained in a concentration independent manner described previously (Rambo and Tainer 2013). A set of 23 independent *ab initio* bead models were calculated from P(r) via DAMMIF (Franke and Svergun 2009) and subsequently averaged using DAMAVER (Volkov and Svergun 2003). Fitting of crystal structures into bead model solution envelopes was performed with SUPCOMB (Kozin and Svergun 2001).

6.3.6 Electron Microscopy

Recombinant HMPV N-RNA rings were imaged using electron cryo-microscopy (cryo-EM). Preparations of bacterially-produced N-RNA were pipetted (3 µl aliquots) onto glow-discharged Cflat holey carbon grids (Protochips, Raleigh, NC). Subsequently, excess sample was blotted away with Whatman filter papers for 3 s. A

CP3 plunging device (Gatan) was utilised for plunge-freezing of grids into an ethane-propane mixture cooled to liquid nitrogen temperature. Cryo-EM images were recorded on a 300-kV Polara TEM (FEI) equipped with a K2 Summit direct electron detector (Gatan). Data were recorded at defocus values ranging from -2.0 to -6.0 µm at a calibrated magnification of 37,000x, corresponding to a pixel size of 1.35 Å. The contrast transfer function (CTF) parameters were obtained with CTFFIND3 (Mindell and Grigorieff 2003) and 2D-classification was performed with RELION (Scheres 2012).

HMPV N-RNA filaments expressed in insect cells were stained using 2% uranyl acetate and imaged using a transmission electron microscope (T12, FEI) operated at 80 kV and equipped with a CCD camera (Eagle, FEI). Microscopy grids of the M2-1 and RNA mixture (ratio 1:2) were stained with 2% uranyl acetate.

6.3.7 Molecular Dynamics Simulations

Molecular dynamics simulations of the N-P complex were carried out with the GROMACS 4 software package (Hess et al. 2008), utilising the AMBER99SB-ILDN* force field (Lindorff-Larsen et al. 2010; Best and Hummer 2009). Prior to simulation, the system was solvated using a box of SPC/E water. Ions were added with genion to obtain an electrically neutral system with 150 mM NaCl. Long range electrostatics were modelled with the particle-mesh Ewald (PME) summation (Essmann et al. 1995). Covalent bond lengths were constrained using P-LINCS (Hess 2008). The simulations used a time step of 5 fs. The v-rescale thermostat was used to maintain a temperature of 300 K (Bussi et al. 2007) and the Parrinello–Rahman barostat was utilised to control the pressure at 1 atm (Parrinello and Rahman 1981).

Initially, systems were energy minimised via 1000 steps of steepest descent and subsequently equilibrated for 200 ps. Independent replicated of simulations were calculated by varying initial velocities. RMSF calculations were performed using the g_rmsf program within GROMACS (Hess et al. 2008).

6.4 Amino Acid Sequences of Studied Proteins

HMPV N (Strain NL1-00, A1, GenBank: AAK62966.1):

```
        10          20          30          40
MSLQGIHLSD  LSYKHAILKE  SQYTIKRDVG  TTTAVTPSSL
        50          60          70          80
QQEITLLCGE  ILYAKHADYK  YAAEIGIQYI  STALGSERVQ
        90         100         110         120
QILRNSGSEV  QVVLTRTYSL  GKIKNNKGED  LQMLDIHGVE
       130         140         150         160
KSWVEEIDKE  ARKTMATLLK  ESSGNIPQNQ  RPSAPDTPII
       170         180         190         200
LLCVGALIFT  KLASTIEVGL  ETTVRRANRV  LSDALKRYPR
       210         220         230         240
MDIPKIARSF  YDLFEQKVYH  RSLFIEYGKA  LGSSSTGSKA
       250         260         270         280
ESLFVNIFMQ  AYGAGQTMLR  WGVIARSSNN  IMLGHVSVQA
       290         300         310         320
ELKQVTEVYD  LVREMGPESG  LLHLRQSPKA  GLLSLANCPN
       330         340         350         360
FASVVLGNAS  GLGIIGMYRG  RVPNTELFSA  AESYAKSLKE
       370         380         390
SNKINFSSLG  LTDEEKEAAE  HFLNVSDDSQ  NDYE
```

HMPV N was either expressed with a 3C-cleavable N-terminal tag (MAHHHHHHSSGLEVLFQGP) from the pOPINF vector, or with a C-terminal tag (KHHHHHH) using pOPINE (Bird et al. 2014).

HMPV $N_{(1-394)} - P_{(1-40)}$ chimera (Strain NL1-00, A1):

```
        10          20          30          40
MSLQGIHLSD  LSYKHAILKE  SQYTIKRDVG  TTTAVTPSSL
        50          60          70          80
QQEITLLCGE  ILYAKHADYK  YAAEIGIQYI  STALGSERVQ
        90         100         110         120
QILRNSGSEV  QVVLTRTYSL  GKIKNNKGED  LQMLDIHGVE
       130         140         150         160
KSWVEEIDKE  ARKTMATLLK  ESSGNIPQNQ  RPSAPDTPII
       170         180         190         200
LLCVGALIFT  KLASTIEVGL  ETTVRRANRV  LSDALKRYPR
       210         220         230         240
MDIPKIARSF  YDLFEQKVYH  RSLFIEYGKA  LGSSSTGSKA
       250         260         270         280
ESLFVNIFMQ  AYGAGQTMLR  WGVIARSSNN  IMLGHVSVQA
       290         300         310         320
ELKQVTEVYD  LVREMGPESG  LLHLRQSPKA  GLLSLANCPN
       330         340         350         360
FASVVLGNAS  GLGIIGMYRG  RVPNTELFSA  AESYAKSLKE
       370         380         390         400
SNKINFSSLG  LTDEEKEAAE  HFLNVSDDSQ  NDYEMSFPEG
       410         420         430
KDILFMGNEA  AKLAEAFQKS  LRKPGHKRSQ  SIIG
```

The N-P chimera was expressed with a C-terminal tag (KHHHHHH) from the pOPINE vector (Bird et al. 2014). The underlined sequences denote residues stemming from P.

HMPV P (Strain NL1-00, A1, GenBank: AAK62966.1):

```
        10          20          30          40
MSFPEGKDIL  FMGNEAAKLA  EAFQKSLRKP  GHKRSQSIIG
        50          60          70          80
EKVNTVSETL  ELPTISRPAK  PTIPSEPKLA  WTDKGGATKT
        90         100         110         120
EIKQAIKVMD  PIEEEESTEK  KVLPSSDGKT  PAEKKLKPST
       130         140         150         160
NTKKKVSFTP  NEPGKYTKLE  KDALDLLSDN  EEEDAESSIL
       170         180         190         200
TFEERDTSSL  SIEARLESIE  EKLSMILGLL  RTLNIATAGP
       210         220         230         240
TAARDGIRDA  MIGVREELIA  DIIKEAKGKA  AEMMEEEMSQ
       250         260         270         280
RSKIGNGSVK  LTEKAKELNK  IVEDESTSGE  SEEEEEPKDT
       290
QDNSQEDDIY  QLIM
```

Constructs of HMPV P used in this study were: $P_{(1-294)}$, $P_{(135-294)}$, $P_{(135-237)}$, and $P_{(158-237)}$. All P constructs were expressed with a 3C-cleavable N-terminal tag (MAHHHHHHSSGLEVLFQGP) using a pOPINF vector (Bird et al. 2014).

HMPV M2-1 (Strain NL1-00, A1, GenBank: AAM12941.1):

```
        10          20          30          40
MSRKAPCKYE  VRGKCNRGSE  CKFNHNYWSW  PDRYLLIRSN
        50          60          70          80
YLLNQLLRNT  DRADGLSIIS  GAGREDRTQD  FVLGSTNVVQ
        90         100         110         120
GYIDDNQSIT  KAAACYSLHN  IIKQLQEVEV  RQARDNKLSD
       130         140         150         160
SKHVALHNLV  LSYMEMSKTP  ASLINNLKRL  PREKLKKLAK
       170         180
LIIDLSAGAE  NDSSYALQDS  ESTNQVQ
```

M2-1 was expressed with a 3C-cleavable N-terminal tag (MAHHHHHHSSGLEVLFQGP) from the pOPINF vector (Bird et al. 2014).

7 REFERENCES

Abrescia NGA, Bamford DH, Grimes JM, Stuart DI. 2012. Structure unifies the viral universe. *Annu Rev Biochem* **81**: 795–822.

Acheson NH. 2006. *Fundamentals of Molecular Virology*. Wiley.

Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**: 213–221.

Afonso CL, Amarasinghe GK, Bányai K, Bào Y, Basler CF, Bavari S, Bejerman N, Blasdell KR, Briand F-X, Briese T, et al. 2016. Taxonomy of the order Mononegavirales: update 2016. *Arch Virol* **161**: 2351–2360.

Alayyoubi M, Leser GP, Kors CA, Lamb RA. 2015. Structure of the paramyxovirus parainfluenza virus 5 nucleoprotein-RNA complex. *Proc Natl Acad Sci U S A* **112**: E1792-1799.

Albertini AAV, Wernimont AK, Muziol T, Ravelli RBG, Clapier CR, Schoehn G, Weissenhorn W, Ruigrok RWH. 2006. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313**: 360–363.

Amarasinghe GK, Dutch RE. 2014. A calcium fortified viral matrix protein. *Struct Lond Engl 1993* **22**: 5–7.

Asenjo A, Calvo E, Villanueva N. 2006. Phosphorylation of human respiratory syncytial virus P protein at threonine 108 controls its interaction with the M2-1 protein in the viral RNA polymerase complex. *J Gen Virol* **87**: 3637–3642.

Assenberg R, Delmas O, Ren J, Vidalain P-O, Verma A, Larrous F, Graham SC, Tangy F, Grimes JM, Bourhy H. 2010. Structure of the Nucleoprotein Binding Domain of Mokola Virus Phosphoprotein. *J Virol* **84**: 1089–1096.

Bai X, McMullan G, Scheres SHW. 2015. How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci* **40**: 49–57.

Bakker SE, Duquerroy S, Galloux M, Loney C, Conner E, Eléouët J-F, Rey FA, Bhella D. 2013. The respiratory syncytial virus nucleoprotein-RNA complex forms a left-handed helical nucleocapsid. *J Gen Virol* **94**: 1734–1738.

Balaji S, Srinivasan N. 2007. Comparison of sequence-based and structure-based phylogenetic trees of homologous proteins: Inferences on protein evolution. *J Biosci* **32**: 83–96.

Barr JN, Wertz GW. 2001. Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3'-AUAC-5' tetranucleotide: implications for the mechanism of transcription termination. *J Virol* **75**: 6901–6913.

Barr JN, Whelan SP, Wertz GW. 1997. cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation. *J Virol* **71**: 8718–8725.

Batonick M, Wertz GW. 2011. Requirements for Human Respiratory Syncytial Virus Glycoproteins in Assembly and Egress from Infected Cells. *Adv Virol* **2011**.

Battisti AJ, Meng G, Winkler DC, McGinnes LW, Plevka P, Steven AC, Morrison TG, Rossmann MG. 2012. Structure and assembly of a paramyxovirus matrix protein. *Proc Natl Acad Sci U S A* **109**: 13996–14000.

Belyi VA, Levine AJ, Skalka AM. 2010. Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. *PLoS Pathog* **6**: e1001030.

Bernadó P. 2010. Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur Biophys J EBJ* **39**: 769–780.

Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DI, Owens RJ. 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res* **35**: e45.

Bertani G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J Bacteriol* **62**: 293–300.

Best RB, Hummer G. 2009. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B* **113**: 9004–9015.

Bhella D, Ralph A, Murphy LB, Yeo RP. 2002. Significant differences in nucleocapsid morphology within the Paramyxoviridae. *J Gen Virol* **83**: 1831–1839.

Biacchesi S, Pham QN, Skiadopoulos MH, Murphy BR, Collins PL, Buchholz UJ. 2005. Infection of Nonhuman Primates with Recombinant Human Metapneumovirus Lacking the SH, G, or M2-2 Protein Categorizes Each as a Nonessential Accessory Protein and Identifies Vaccine Candidates. *J Virol* **79**: 12608–12613.

Biacchesi S, Skiadopoulos MH, Boivin G, Hanson CT, Murphy BR, Collins PL, Buchholz UJ. 2003. Genetic diversity between human metapneumovirus subgroups. *Virology* **315**: 1–9.

Biacchesi S, Skiadopoulos MH, Yang L, Lamirande EW, Tran KC, Murphy BR, Collins PL, Buchholz UJ. 2004. Recombinant Human Metapneumovirus Lacking the Small Hydrophobic SH and/or Attachment G Glycoprotein: Deletion of G Yields a Promising Vaccine Candidate. *J Virol* **78**: 12877–12887.

Bibby J, Keegan RM, Mayans O, Winn MD, Rigden DJ. 2012. AMPLE: a cluster-and-truncate approach to solve the crystal structures of small proteins using rapidly computed ab initio models. *Acta Crystallogr D Biol Crystallogr* **68**: 1622–1631.

Bird LE, Rada H, Flanagan J, Diprose JM, Gilbert RJC, Owens RJ. 2014. Application of In-Fusion™ cloning for the parallel construction of E. coli expression vectors. *Methods Mol Biol Clifton NJ* **1116**: 209–234.

Black RE, Cousens S, Johnson HL, Lawn JE, Rudan I, Bassani DG, Jha P, Campbell H, Walker CF, Cibulskis R, et al. 2010. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet Lond Engl* **375**: 1969–1987.

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

Boeckh M, Erard V, Zerr D, Englund J. 2005. Emerging viral infections after hematopoietic cell transplantation. *Pediatr Transplant* **9 Suppl 7**: 48–54.

Boivin G, Abed Y, Pelletier G, Ruel L, Moisan D, Côté S, Peret TCT, Erdman DD, Anderson LJ. 2002. Virological features and clinical manifestations associated with human metapneumovirus: a new paramyxovirus responsible for acute respiratory-tract infections in all age groups. *J Infect Dis* **186**: 1330–1334.

Boivin G, De Serres G, Côté S, Gilca R, Abed Y, Rochette L, Bergeron MG, Déry P. 2003. Human metapneumovirus infections in hospitalized children. *Emerg Infect Dis* **9**: 634–640.

Briese T, de la Torre JC, Lewis A, Ludwig H, Lipkin WI. 1992. Borna disease virus, a negative-strand RNA virus, transcribes in the nucleus of infected cells. *Proc Natl Acad Sci U S A* **89**: 11486–11489.

Brockmann C, Soucek S, Kuhlmann SI, Mills-Lujan K, Kelly SM, Yang J-C, Iglesias N, Stutz F, Corbett AH, Neuhaus D, et al. 2012. Structural basis for polyadenosine-RNA binding by Nab2 Zn fingers and its function in mRNA nuclear export. *Struct Lond Engl 1993* **20**: 1007–1018.

Bruhn JF, Barnett KC, Bibby J, Thomas JMH, Keegan RM, Rigden DJ, Bornholdt ZA, Saphire EO. 2014. Crystal structure of the nipah virus phosphoprotein tetramerization domain. *J Virol* **88**: 758–762.

Buijs PRA, Verhagen JHE, van Eijck CHJ, van den Hoogen BG. 2015. Oncolytic viruses: From bench to bedside with a focus on safety. *Hum Vaccines Immunother* **11**: 1573–1584.

Bussi G, Donadio D, Parrinello M. 2007. Canonical sampling through velocity rescaling. *J Chem Phys* **126**: 014101.

Calain P, Roux L. 1993. The rule of six, a basic feature for efficient replication of Sendai virus defective interfering RNA. *J Virol* **67**: 4822–4830.

Cane PA, van den Hoogen BG, Chakrabarti S, Fegan CD, Osterhaus ADME. 2003. Human metapneumovirus in a haematopoietic stem cell transplant recipient with fatal lower respiratory tract disease. *Bone Marrow Transplant* **31**: 309–310.

Castel G, Chtéoui M, Caignard G, Préhaud C, Méhouas S, Réal E, Jallet C, Jacob Y, Ruigrok RWH, Tordo N. 2009. Peptides that mimic the amino-terminal end of the rabies virus phosphoprotein have antiviral activity. *J Virol* **83**: 10808–10820.

Céspedes PF, Palavecino CE, Kalergis AM, Bueno SM. 2016. Modulation of Host Immunity by the Human Metapneumovirus. *Clin Microbiol Rev* **29**: 795–818.

Challa S, Scott AD, Yuzhakov O, Zhou Y, Tiong-Yip CL, Gao N, Thresher J, Yu Q. 2015. Mechanism of action for respiratory syncytial virus inhibitor RSV604. *Antimicrob Agents Chemother* **59**: 1080–1087.

Chang A, Masante C, Buchholz UJ, Dutch RE. 2012. Human metapneumovirus (HMPV) binding and infection are mediated by interactions between the HMPV fusion protein and heparan sulfate. *J Virol* **86**: 3230–3243.

Chapman J, Abbott E, Alber DG, Baxter RC, Bithell SK, Henderson EA, Carter MC, Chambers P, Chubb A, Cockerill GS, et al. 2007. RSV604, a novel inhibitor of respiratory syncytial virus replication. *Antimicrob Agents Chemother* **51**: 3346–3353.

Chattopadhyay S, Banerjee AK. 2009. Phosphoprotein, P of human parainfluenza virus type 3 prevents self-association of RNA-dependent RNA polymerase, L. *Virology* **383**: 226–236.

Chen M, Ogino T, Banerjee AK. 2007. Interaction of vesicular stomatitis virus P and N proteins: identification of two overlapping domains at the N terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA. *J Virol* **81**: 13478–13485.

Chen M, Ogino T, Banerjee AK. 2006. Mapping and functional role of the self-association domain of vesicular stomatitis virus phosphoprotein. *J Virol* **80**: 9511–9518.

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**: 12–21.

Cheng Y, Grigorieff N, Penczek PA, Walz T. 2015. A primer to single-particle cryo-electron microscopy. *Cell* **161**: 438–449.

Chiu CY, Alizadeh AA, Rouskin S, Merker JD, Yeh E, Yagi S, Schnurr D, Patterson BK, Ganem D, DeRisi JL. 2007. Diagnosis of a critical respiratory illness caused by human metapneumovirus by use of a pan-virus microarray. *J Clin Microbiol* **45**: 2340–2343.

Collins PL, Hill MG, Camargo E, Grosfeld H, Chanock RM, Murphy BR. 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 development. *Proc Natl Acad Sci U S A* **92**: 11563–11567.

Collins PL, Hill MG, Cristina J, Grosfeld H. 1996. Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus. *Proc Natl Acad Sci U S A* **93**: 81–85.

Communie G, Crépin T, Maurin D, Jensen MR, Blackledge M, Ruigrok RWH. 2013a. Structure of the tetramerization domain of measles virus phosphoprotein. *J Virol* **87**: 7166–7169.

Communie G, Habchi J, Yabukarski F, Blocquel D, Schneider R, Tarbouriech N, Papageorgiou N, Ruigrok RWH, Jamin M, Jensen MR, et al. 2013b. Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus. *PLoS Pathog* **9**: e1003631.

Coronel EC, Takimoto T, Murti KG, Varich N, Portner A. 2001. Nucleocapsid incorporation into parainfluenza virus is regulated by specific interaction with matrix protein. *J Virol* **75**: 1117–1123.

Cowton VM, McGivern DR, Fearns R. 2006. Unravelling the complexities of respiratory syncytial virus RNA synthesis. *J Gen Virol* **87**: 1805–1821.

Cox R, Green TJ, Purushotham S, Deivanayagam C, Bedwell GJ, Prevelige PE, Luo M. 2013. Structural and functional characterization of the mumps virus phosphoprotein. *J Virol* **87**: 7558–7568.

Cox R, Pickar A, Qiu S, Tsao J, Rodenburg C, Dokland T, Elson A, He B, Luo M. 2014. Structural studies on the authentic mumps virus nucleocapsid showing uncoiling by the phosphoprotein. *Proc Natl Acad Sci U S A* **111**: 15208–15213.

Cox R, Plemper RK. 2015. The paramyxovirus polymerase complex as a target for next-generation anti-paramyxovirus therapeutics. *Front Microbiol* **6**: 459.

Cox RG, Livesay SB, Johnson M, Ohi MD, Williams JV. 2012. The Human Metapneumovirus Fusion Protein Mediates Entry via an Interaction with RGD-Binding Integrins. *J Virol* **86**: 12148–12160.

Cox RG, Mainou BA, Johnson M, Hastings AK, Schuster JE, Dermody TS, Williams JV. 2015. Human Metapneumovirus Is Capable of Entering Cells by Fusion with Endosomal Membranes. *PLoS Pathog* **11**: e1005303.

Cox RG, Williams JV. 2013. Breaking in: human metapneumovirus fusion and entry. *Viruses* **5**: 192–210.

Cseke G, Maginnis MS, Cox RG, Tollefson SJ, Podsiad AB, Wright DW, Dermody TS, Williams JV. 2009. Integrin alphavbeta1 promotes infection by human metapneumovirus. *Proc Natl Acad Sci U S A* **106**: 1566–1571.

Cuesta I, Geng X, Asenjo A, Villanueva N. 2000. Structural Phosphoprotein M2-1 of the Human Respiratory Syncytial Virus Is an RNA Binding Protein. *J Virol* **74**: 9858–9867.

Cureton DK, Massol RH, Saffarian S, Kirchhausen TL, Whelan SPJ. 2009. Vesicular stomatitis virus enters cells through vesicles incompletely coated with clathrin that depend upon actin for internalization. *PLoS Pathog* **5**: e1000394.

Curran J, Marq JB, Kolakofsky D. 1995. An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. *J Virol* **69**: 849–855.

Das R, André I, Shen Y, Wu Y, Lemak A, Bansal S, Arrowsmith CH, Szyperski T, Baker D. 2009. Simultaneous prediction of protein folding and docking at high resolution. *Proc Natl Acad Sci U S A* **106**: 18978–18983.

de Graaf M, Herfst S, Schrauwen EJA, Choi Y, van den Hoogen BG, Osterhaus ADME, Fouchier RAM. 2008a. Specificity and functional interaction of the polymerase complex proteins of human and avian metapneumoviruses. *J Gen Virol* **89**: 975–983.

de Graaf M, Osterhaus ADME, Fouchier RAM, Holmes EC. 2008b. Evolutionary dynamics of human and avian metapneumoviruses. *J Gen Virol* **89**: 2933–2942.

Derdowski A, Peters TR, Glover N, Qian R, Utley TJ, Burnett A, Williams JV, Spearman P, Crowe JE. 2008. Human metapneumovirus nucleoprotein and phosphoprotein interact and provide the minimal requirements for inclusion body formation. *J Gen Virol* **89**: 2698–2708.

Desfosses A, Goret G, Farias Estrozi L, Ruigrok RWH, Gutsche I. 2011. Nucleoprotein-RNA orientation in the measles virus nucleocapsid by three-dimensional electron microscopy. *J Virol* **85**: 1391–1395.

Ding H, Green TJ, Lu S, Luo M. 2006. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J Virol* **80**: 2808–2814.

Dong A, Xu X, Edwards AM, Midwest Center for Structural Genomics, Structural Genomics Consortium, Chang C, Chruszcz M, Cuff M, Cymborowski M, Di Leo R, et al. 2007. In situ proteolysis for protein crystallization and structure determination. *Nat Methods* **4**: 1019–1021.

Doniach S. 2001. Changes in biomolecular conformation seen by small angle X-ray scattering. *Chem Rev* **101**: 1763–1778.

Dunker AK, Silman I, Uversky VN, Sussman JL. 2008. Function and structure of inherently disordered proteins. *Curr Opin Struct Biol* **18**: 756–764.

Dürrwald R, Ludwig H. 1997. Borna disease virus (BDV), a (zoonotic?) worldwide pathogen. A review of the history of the disease and the virus infection with comprehensive bibliography. *Zentralblatt Für Veterinärmedizin Reihe B J Vet Med Ser B* **44**: 147–184.

Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**: 197–208.

El Najjar F, Schmitt AP, Dutch RE. 2014. Paramyxovirus glycoprotein incorporation, assembly and budding: a three way dance for infectious particle production. *Viruses* **6**: 3019–3054.

Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126–2132.

Erales J, Blocquel D, Habchi J, Beltrandi M, Gruet A, Dosnon M, Bignon C, Longhi S. 2015. Order and Disorder in the Replicative Complex of Paramyxoviruses. *Adv Exp Med Biol* **870**: 351–381.

Esperante SA, Chemes LB, Sánchez IE, de Prat-Gay G. 2011. The respiratory syncytial virus transcription antiterminator M(2-1) is a highly stable, zinc binding tetramer with strong pH-dependent dissociation and a monomeric unfolding intermediate. *Biochemistry (Mosc)* **50**: 8529–8539.

Esperante SA, Paris G, de Prat-Gay G. 2012. Modular unfolding and dissociation of the human respiratory syncytial virus phosphoprotein p and its interaction with the m(2-1) antiterminator: a singular tetramer-tetramer interface arrangement. *Biochemistry (Mosc)* **51**: 8100–8110.

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995. A smooth particle mesh Ewald method. *J Chem Phys* **103**: 8577–8593.

Falsey AR, McElhaney JE, Beran J, Essen GA van, Duval X, Esen M, Galtier F, Gervais P, Hwang S-J, Kremsner P, et al. 2014. Respiratory Syncytial Virus and Other Respiratory Viral Infections in Older Adults With Moderate to Severe Influenza-like Illness. *J Infect Dis* jit839.

Fearns R, Collins PL. 1999. Role of the M2-1 transcription antitermination protein of respiratory syncytial virus in sequential transcription. *J Virol* **73**: 5852–5864.

Fearns R, Deval J. 2016. New antiviral approaches for respiratory syncytial virus and other mononegaviruses: Inhibiting the RNA polymerase. *Antiviral Res* **134**: 63–76.

Feldmann H, Klenk HD, Sanchez A. 1993. Molecular biology and evolution of filoviruses. *Arch Virol Suppl* **7**: 81–100.

Feschotte C. 2010. Virology: Bornavirus enters the genome. *Nature* **463**: 39–40.

Fisette O, Wingbermühle S, Tampé R, Schäfer LV. 2016. Molecular mechanism of peptide editing in the tapasin-MHC I complex. *Sci Rep* **6**: 19085.

Förster A, Maertens GN, Farrell PJ, Bajorek M. 2015. Dimerization of Matrix Protein Is Required for Budding of Respiratory Syncytial Virus. *J Virol* **89**: 4624–4635.

Franke D, Svergun DI. 2009. DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J Appl Crystallogr* **42**: 342–346.

Galloux M, Gabiane G, Sourimant J, Richard C-A, England P, Moudjou M, Aumont-Nicaise M, Fix J, Rameix-Welti M-A, Eléouët J-F. 2015. Identification and characterization of the binding site of the respiratory syncytial virus phosphoprotein to RNA-free nucleoprotein. *J Virol* **89**: 3484–3496.

Gan SW, Ng L, Lin X, Gong X, Torres J. 2008. Structure and ion channel activity of the human respiratory syncytial virus (hRSV) small hydrophobic protein transmembrane domain. *Protein Sci Publ Protein Soc* **17**: 813–820.

García-Barreno B, Delgado T, Melero JA. 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. *J Virol* **70**: 801–808.

Ge P, Tsao J, Schein S, Green TJ, Luo M, Zhou ZH. 2010. Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science* **327**: 689–693.

Geisbert TW, Feldmann H. 2011. Recombinant vesicular stomatitis virus-based vaccines against Ebola and Marburg virus infections. *J Infect Dis* **204 Suppl 3**: S1075-1081.

Gerard FCA, Ribeiro Jr E de A, Leyrat C, Ivanov I, Blondel D, Longhi S, Ruigrok RWH, Jamin M. 2009. Modular Organization of Rabies Virus Phosphoprotein. *J Mol Biol* **388**: 978–996.

Gerlach P, Malet H, Cusack S, Reguera J. 2015. Structural Insights into Bunyavirus Replication and Its Regulation by the vRNA Promoter. *Cell* **161**: 1267–1279.

Ghildyal R, Li D, Peroulis I, Shields B, Bardin PG, Teng MN, Collins PL, Meanger J, Mills J. 2005. Interaction between the respiratory syncytial virus G glycoprotein cytoplasmic domain and the matrix protein. *J Gen Virol* **86**: 1879–1884.

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

Gorrec F. 2009. The MORPHEUS protein crystallization screen. *J Appl Crystallogr* **42**: 1035–1042.

Govindarajan D, Buchholz UJ, Samal SK. 2006. Recovery of avian metapneumovirus subgroup C from cDNA: cross-recognition of avian and human metapneumovirus support proteins. *J Virol* **80**: 5790–5797.

Green TJ, Luo M. 2009. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. *Proc Natl Acad Sci U S A* **106**: 11713–11718.

Green TJ, Zhang X, Wertz GW, Luo M. 2006. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* **313**: 357–360.

Guryanov SG, Liljeroos L, Kasaragod P, Kajander T, Butcher SJ. 2016. Crystal Structure of the Measles Virus Nucleoprotein Core in Complex with an N-Terminal Region of Phosphoprotein. *J Virol* **90**: 2849–2857.

Gutsche I, Desfosses A, Effantin G, Ling WL, Haupt M, Ruigrok RWH, Sachse C, Schoehn G. 2015. Structural virology. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science* **348**: 704–707.

Haas LEM, de Rijk NX, Thijsen SFT. 2012. Human metapneumovirus infections on the ICU: a report of three cases. *Ann Intensive Care* **2**: 30.

Haas LEM, Thijsen SFT, van Elden L, Heemstra KA. 2013. Human Metapneumovirus in Adults. *Viruses* **5**: 87–110.

Habchi J, Longhi S. 2012. Structural disorder within paramyxovirus nucleoproteins and phosphoproteins. *Mol Biosyst* **8**: 69–81.

Hardy RW, Harmon SB, Wertz GW. 1999. Diverse gene junctions of respiratory syncytial virus modulate the efficiency of transcription termination and respond differently to M2-mediated antitermination. *J Virol* **73**: 170–176.

Hardy RW, Wertz GW. 2000. The Cys(3)-His(1) motif of the respiratory syncytial virus M2-1 protein is essential for protein function. *J Virol* **74**: 5880–5885.

Hardy RW, Wertz GW. 1998. The product of the respiratory syncytial virus M2 gene ORF1 enhances readthrough of intergenic junctions during viral transcription. *J Virol* **72**: 520–526.

Harmon SB, Megaw AG, Wertz GW. 2001. RNA Sequences Involved in Transcriptional Termination of Respiratory Syncytial Virus. *J Virol* **75**: 36–44.

Hassell AM, An G, Bledsoe RK, Bynum JM, Carter HL, Deng S-JJ, Gampe RT, Grisard TE, Madauss KP, Nolte RT, et al. 2007. Crystallization of protein–ligand complexes. *Acta Crystallogr D Biol Crystallogr* **63**: 72–79.

Hastie E, Grdzelishvili VZ. 2012. Vesicular stomatitis virus as a flexible platform for oncolytic virotherapy against cancer. *J Gen Virol* **93**: 2529–2545.

Hasvold J, Sjoding M, Pohl K, Cooke C, Hyzy RC. 2016. The role of human metapneumovirus in the critically ill adult patient. *J Crit Care* **31**: 233–237.

Heggeness MH, Scheid A, Choppin PW. 1980. Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proc Natl Acad Sci U S A* **77**: 2631–2635.

Heggeness MH, Scheid A, Choppin PW. 1981. the relationship of conformational changes in the Sendai virus nucleocapsid to proteolytic cleavage of the NP polypeptide. *Virology* **114**: 555–562.

Henao-Restrepo AM, Preziosi M-P, Wood D, Moorthy V, Kieny MP, WHO Ebola Research, Development Team. 2016. On a path to accelerate access to Ebola vaccines: The WHO's research and development efforts during the 2014-2016 Ebola epidemic in West Africa. *Curr Opin Virol* **17**: 138–144.

Hengrung N, El Omari K, Serna Martin I, Vreede FT, Cusack S, Rambo RP, Vonrhein C, Bricogne G, Stuart DI, Grimes JM, et al. 2015. Crystal structure of the RNA-dependent RNA polymerase from influenza C virus. *Nature* **527**: 114–117.

Herfst S, de Graaf M, Schickli JH, Tang RS, Kaur J, Yang C-F, Spaete RR, Haller AA, van den Hoogen BG, Osterhaus ADME, et al. 2004. Recovery of human metapneumovirus genetic lineages a and B from cloned cDNA. *J Virol* **78**: 8264–8270.

Herfst S, Mas V, Ver LS, Wierda RJ, Osterhaus ADME, Fouchier RAM, Melero JA. 2008. Low-pH-induced membrane fusion mediated by human metapneumovirus F protein is a rare, strain-dependent phenomenon. *J Virol* **82**: 8891–8895.

Herrel M, Hoefs N, Staeheli P, Schneider U. 2012. Tick-borne Nyamanini virus replicates in the nucleus and exhibits unusual genome and matrix protein properties. *J Virol* **86**: 10739–10747.

Hess B. 2008. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput* **4**: 116–122.

Hess B, Kutzner C, van der Spoel D, Lindahl E. 2008. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **4**: 435–447.

Holm L, Rosenström P. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**: W545-549.

Holzmann H, Hengel H, Tenbusch M, Doerr HW. 2016. Eradication of measles: remaining challenges. *Med Microbiol Immunol (Berl)* **205**: 201–208.

Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, Ikuta K, Jern P, Gojobori T, Coffin JM, et al. 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* **463**: 84–87.

Huang YJ, Acton TB, Montelione GT. 2014. DisMeta: a meta server for construct design and optimization. *Methods Mol Biol Clifton NJ* **1091**: 3–16.

Ilca SL, Kotecha A, Sun X, Poranen MM, Stuart DI, Huiskonen JT. 2015. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat Commun* **6**: 8843.

Ivanov I, Crépin T, Jamin M, Ruigrok RWH. 2010. Structure of the dimerization domain of the rabies virus phosphoprotein. *J Virol* **84**: 3707–3710.

Iwasaki M, Takeda M, Shirogane Y, Nakatsu Y, Nakamura T, Yanagi Y. 2009. The matrix protein of measles virus regulates viral RNA synthesis and assembly by interacting with the nucleocapsid protein. *J Virol* **83**: 10374–10383.

Jacques JP, Kolakofsky D. 1991. Pseudo-templated transcription in prokaryotic and eukaryotic organisms. *Genes Dev* **5**: 707–713.

Jensen MR, Communie G, Ribeiro EA, Martinez N, Desfosses A, Salmon L, Mollica L, Gabel F, Jamin M, Longhi S, et al. 2011. Intrinsic disorder in measles virus nucleocapsids. *Proc Natl Acad Sci U S A* **108**: 9839–9844.

Johansson K, Bourhis J-M, Campanacci V, Cambillau C, Canard B, Longhi S. 2003. Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J Biol Chem* **278**: 44567–44573.

John SP, Wang T, Steffen S, Longhi S, Schmaljohn CS, Jonsson CB. 2007. Ebola virus VP30 is an RNA binding protein. *J Virol* **81**: 8967–8976.

Karlin D, Belshaw R. 2012. Detecting remote sequence homology in disordered proteins: discovery of conserved motifs in the N-termini of Mononegavirales phosphoproteins. *PloS One* **7**: e31719.

Karlin D, Ferron F, Canard B, Longhi S. 2003. Structural disorder and modular organization in Paramyxovirinae N and P. *J Gen Virol* **84**: 3239–3252.

Karplus PA, Diederichs K. 2012. Linking crystallographic model and data quality. *Science* **336**: 1030–1033.

Kelly SM, Leung SW, Apponi LH, Bramley AM, Tran EJ, Chekanova JA, Wente SR, Corbett AH. 2010. Recognition of Polyadenosine RNA by the Zinc Finger Domain of Nuclear Poly(A) RNA-binding Protein 2 (Nab2) Is Required for Correct mRNA 3′-End Formation*. *J Biol Chem* **285**: 26022–26032.

Khattar SK, Yunus AS, Samal SK. 2001. Mapping the domains on the phosphoprotein of bovine respiratory syncytial virus required for N-P and P-L interactions using a minigenome system. *J Gen Virol* **82**: 775–779.

Kikhney AG, Svergun DI. 2015. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett* **589**: 2570–2577.

Kingston RL, Gay LS, Baase WS, Matthews BW. 2008. Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization. *J Mol Biol* **379**: 719–731.

Kingston RL, Hamel DJ, Gay LS, Dahlquist FW, Matthews BW. 2004. Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc Natl Acad Sci U S A* **101**: 8301–8306.

Kinnunen PM, Inkeroinen H, Ilander M, Kallio ER, Heikkilä HP, Koskela E, Mappes T, Palva A, Vaheri A, Kipar A, et al. 2011. Intracerebral Borna disease virus infection of bank voles leading to peripheral spread and reverse transcription of viral RNA. *PloS One* **6**: e23622.

Kinnunen PM, Palva A, Vaheri A, Vapalahti O. 2013. Epidemiology and host spectrum of Borna disease virus infections. *J Gen Virol* **94**: 247–262.

Kiss G, Holl JM, Williams GM, Alonas E, Vanover D, Lifland AW, Gudheti M, Guerrero-Ferreira RC, Nair V, Yi H, et al. 2014. Structural analysis of respiratory syncytial virus reveals the position of M2-1 between the matrix protein and the ribonucleoprotein complex. *J Virol* **88**: 7602–7617.

Knipe DM, Howley PM. 2007. *Fields' Virology*. Lippincott Williams & Wilkins.

Kolakofsky D, Le Mercier P, Iseni F, Garcin D. 2004. Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology* **318**: 463–473.

Kozin MB, Svergun DI. 2001. Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* **34**: 33–41.

Kram KE, Finkel SE. 2015. Rich Medium Composition Affects Escherichia coli Survival, Glycation, and Mutation Frequency during Long-Term Batch Culture. *Appl Environ Microbiol* **81**: 4442–4450.

Krissinel E. 2010. Crystal contacts as nature's docking solutions. *J Comput Chem* **31**: 133–143.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.

Leung DW, Borek D, Luthra P, Binning JM, Anantpadma M, Liu G, Harvey IB, Su Z, Endlich-Frazier A, Pan J, et al. 2015. An Intrinsically Disordered Peptide from Ebola Virus VP35 Controls Viral RNA Synthesis by Modulating Nucleoprotein-RNA Interactions. *Cell Rep* **11**: 376–389.

Leung DW, Prins KC, Basler CF, Amarasinghe GK. 2010. Ebolavirus VP35 is a multifunctional virulence factor. *Virulence* **1**: 526–531.

Leyrat C, Paesen GC, Charleston J, Renner M, Grimes JM. 2014a. Structural Insights into the Human Metapneumovirus Glycoprotein Ectodomain. *J Virol* **88**: 11611–11616.

Leyrat C, Renner M, Harlos K, Grimes JM. 2013. Solution and Crystallographic Structures of the Central Region of the Phosphoprotein from Human Metapneumovirus. *PLOS ONE* **8**: e80371.

Leyrat C, Renner M, Harlos K, Huiskonen JT, Grimes JM. 2014b. Drastic changes in conformational dynamics of the antiterminator M2-1 regulate transcription efficiency in Pneumovirinae. *eLife* **3**: e02674.

Leyrat C, Renner M, Harlos K, Huiskonen JT, Grimes JM. 2014c. Structure and self-assembly of the calcium binding matrix protein of human metapneumovirus. *Struct Lond Engl 1993* **22**: 136–148.

Leyrat C, Yabukarski F, Tarbouriech N, Ribeiro EA, Jensen MR, Blackledge M, Ruigrok RWH, Jamin M. 2011. Structure of the vesicular stomatitis virus N^0-P complex. *PLoS Pathog* **7**: e1002248.

Li D, Jans DA, Bardin PG, Meanger J, Mills J, Ghildyal R. 2008a. Association of respiratory syncytial virus M protein with viral nucleocapsids is mediated by the M2-1 protein. *J Virol* **82**: 8863–8870.

Li J, Rahmeh A, Morelli M, Whelan SPJ. 2008b. A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping. *J Virol* **82**: 775–784.

Liang B, Li Z, Jenni S, Rahmeh AA, Morin BM, Grant T, Grigorieff N, Harrison SC, Whelan SPJ. 2015. Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. *Cell* **162**: 314–327.

Liao RS, Appelgate DM, Pelz RK. 2012. An outbreak of severe respiratory tract infection due to human metapneumovirus in a long-term care facility for the elderly in Oregon. *J Clin Virol Off Publ Pan Am Soc Clin Virol* **53**: 171–173.

Liljeroos L, Butcher SJ. 2013. Matrix proteins as centralized organizers of negative-sense RNA virions. *Front Biosci Landmark Ed* **18**: 696–715.

Liljeroos L, Krzyzaniak MA, Helenius A, Butcher SJ. 2013. Architecture of respiratory syncytial virus revealed by electron cryotomography. *Proc Natl Acad Sci U S A* **110**: 11133–11138.

Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**: 1950–1958.

Ling R, Sinkovic S, Toquin D, Guionie O, Eterradossi N, Easton AJ. 2008. Deletion of the SH gene from avian metapneumovirus has a greater impact on virus production and immunogenicity in turkeys than deletion of the G gene or M2-2 open reading frame. *J Gen Virol* **89**: 525–533.

Llorente MT, García-Barreno B, Calero M, Camafeita E, López JA, Longhi S, Ferrón F, Varela PF, Melero JA. 2006. Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization. *J Gen Virol* **87**: 159–169.

Llorente MT, Taylor IA, López-Viñas E, Gomez-Puertas P, Calder LJ, García-Barreno B, Melero JA. 2008. Structural properties of the human respiratory syncytial virus P protein: evidence for an elongated homotetrameric molecule that is the smallest orthologue within the family of paramyxovirus polymerase cofactors. *Proteins* **72**: 946–958.

Lo MS, Brazas RM, Holtzman MJ. 2005. Respiratory syncytial virus nonstructural proteins NS1 and NS2 mediate inhibition of Stat2 expression and alpha/beta interferon responsiveness. *J Virol* **79**: 9315–9319.

London N, Movshovitz-Attias D, Schueler-Furman O. 2010. The Structural Basis of Peptide-Protein Binding Strategies. *Structure* **18**: 188–199.

Loney C, Mottet-Osman G, Roux L, Bhella D. 2009. Paramyxovirus ultrastructure and genome packaging: cryo-electron tomography of sendai virus. *J Virol* **83**: 8191–8197.

Loo LH, Jumat MR, Fu Y, Ayi TC, Wong PS, Tee NW, Tan BH, Sugrue RJ. 2013. Evidence for the interaction of the human metapneumovirus G and F proteins during virus-like particle formation. *Virol J* **10**: 294.

Ludwig K, Schade B, Böttcher C, Korte T, Ohlwein N, Baljinnyam B, Veit M, Herrmann A. 2008. Electron cryomicroscopy reveals different F1+F2 protein States in intact parainfluenza virions. *J Virol* **82**: 3775–3781.

Mann KS, Dietzgen RG. 2014. Plant rhabdoviruses: new insights and research needs in the interplay of negative-strand RNA viruses with plant and insect hosts. *Arch Virol* **159**: 1889–1900.

Martinez N, Ribeiro EA, Leyrat C, Tarbouriech N, Ruigrok RWH, Jamin M. 2013. Structure of the C-Terminal Domain of Lettuce Necrotic Yellows Virus Phosphoprotein. *J Virol* **87**: 9569–9578.

Martins KA, Jahrling PB, Bavari S, Kuhn JH. 2016. Ebola virus disease candidate vaccines under evaluation in clinical trials. *Expert Rev Vaccines* **15**: 1101–1112.

Mas V, Herfst S, Osterhaus ADME, Fouchier RAM, Melero JA. 2011. Residues of the human metapneumovirus fusion (F) protein critical for its strain-related fusion phenotype: implications for the virus replication cycle. *J Virol* **85**: 12650–12661.

Masante C, El Najjar F, Chang A, Jones A, Moncman CL, Dutch RE. 2014. The human metapneumovirus small hydrophobic protein has properties consistent with those of a viroporin and can modulate viral fusogenic activity. *J Virol* **88**: 6423–6433.

Mason SW, Aberg E, Lawetz C, DeLong R, Whitehead P, 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. *J Virol* **77**: 10670–10676.

Masters PS, Banerjee AK. 1988. Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA. *J Virol* **62**: 2658–2664.

Mavrakis M, McCarthy AA, Roche S, Blondel D, Ruigrok RWH. 2004. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J Mol Biol* **343**: 819–831.

Mavrakis M, Méhouas S, Réal E, Iseni F, Blondel D, Tordo N, Ruigrok RWH. 2006. Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA. *Virology* **349**: 422–429.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr* **40**: 658–674.

McLellan JS, Chen M, Leung S, Graepel KW, Du X, Yang Y, Zhou T, Baxa U, Yasuda E, Beaumont T, et al. 2013. Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340**: 1113–1117.

McLellan JS, Yang Y, Graham BS, Kwong PD. 2011. Structure of Respiratory Syncytial Virus Fusion Glycoprotein in the Postfusion Conformation Reveals Preservation of Neutralizing Epitopes▽. *J Virol* **85**: 7788–7796.

McPhee HK, Carlisle JL, Beeby A, Money VA, Watson SMD, Yeo RP, Sanderson JM. 2011. Influence of lipids on the interfacial disposition of respiratory syncytical virus matrix protein. *Langmuir ACS J Surf Colloids* **27**: 304–311.

McPherson A, Cudney B. 2006. Searching for silver bullets: an alternative strategy for crystallizing macromolecules. *J Struct Biol* **156**: 387–406.

Mihindukulasuriya KA, Nguyen NL, Wu G, Huang HV, da Rosa APAT, Popov VL, Tesh RB, Wang D. 2009. Nyamanini and midway viruses define a novel taxon of RNA viruses in the order Mononegavirales. *J Virol* **83**: 5109–5116.

Mindell JA, Grigorieff N. 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **142**: 334–347.

Modrof J, Becker S, Mühlberger E. 2003. Ebola virus transcription activator VP30 is a zinc-binding protein. *J Virol* **77**: 3334–3338.

Modrow S, Falke D, Truyen U, Schätzl H. 2010. *Molekulare Virologie*. Spektrum Akademischer Verlag.

Moll R, Reece S, Cosford P, Kessel A. 2016. The Ebola epidemic and public health response. *Br Med Bull* **117**: 15–23.

Money VA, McPhee HK, Mosely JA, Sanderson JM, Yeo RP. 2009. Surface features of a Mononegavirales matrix protein indicate sites of membrane interaction. *Proc Natl Acad Sci U S A* **106**: 4441–4446.

Morin B, Kranzusch PJ, Rahmeh AA, Whelan SPJ. 2013. The polymerase of negative-stranded RNA viruses. *Curr Opin Virol* **3**: 103–110.

Morin B, Rahmeh AA, Whelan SPJ. 2012. Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase. *EMBO J* **31**: 1320–1329.

Mühlberger E, Weik M, Volchkov VE, Klenk HD, Becker S. 1999. Comparison of the transcription and replication strategies of marburg virus and Ebola virus by using artificial replication systems. *J Virol* **73**: 2333–2342.

Mullis KB, Faloona FA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* **155**: 335–350.

Neubauer J, Ogino M, Green TJ, Ogino T. 2016. Signature motifs of GDP polyribonucleotidyltransferase, a non-segmented negative strand RNA viral mRNA capping enzyme, domain in the L protein are required for covalent enzyme-pRNA intermediate formation. *Nucleic Acids Res* **44**: 330–341.

Neumann P, Lieber D, Meyer S, Dautel P, Kerth A, Kraus I, Garten W, Stubbs MT. 2009. Crystal structure of the Borna disease virus matrix protein (BDV-M) reveals ssRNA binding properties. *Proc Natl Acad Sci U S A* **106**: 3710–3715.

Noton SL, Deflubé LR, Tremaglio CZ, Fearns R. 2012. The respiratory syncytial virus polymerase has multiple RNA synthesis activities at the promoter. *PLoS Pathog* **8**: e1002980.

Noton SL, Fearns R. 2015. Initiation and regulation of paramyxovirus transcription and replication. *Virology* **479–480**: 545–554.

Noton SL, Fearns R. 2011. The first two nucleotides of the respiratory syncytial virus antigenome RNA replication product can be selected independently of the promoter terminus. *RNA N Y N* **17**: 1895–1906.

Ogino T, Banerjee AK. 2007. Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol Cell* **25**: 85–97.

Ortín J, Martín-Benito J. 2015. The RNA synthesis machinery of negative-stranded RNA viruses. *Virology* **479–480**: 532–544.

Ouizougun-Oubari M, Pereira N, Tarus B, Galloux M, Lassoued S, Fix J, Tortorici MA, Hoos S, Baron B, England P, et al. 2015. A Druggable Pocket at the Nucleocapsid/Phosphoprotein Interaction Site of Human Respiratory Syncytial Virus. *J Virol* **89**: 11129–11143.

Paesen GC, Collet A, Sallamand C, Debart F, Vasseur J-J, Canard B, Decroly E, Grimes JM. 2015. X-ray structure and activities of an essential Mononegavirales L-protein domain. *Nat Commun* **6**: 8749.

Panda S, Mohakud NK, Pena L, Kumar S. 2014. Human metapneumovirus: review of an important respiratory pathogen. *Int J Infect Dis IJID Off Publ Int Soc Infect Dis* **25**: 45–52.

Parrinello M, Rahman A. 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **52**: 7182–7190.

Pei J, Kim B-H, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* **36**: 2295–2300.

Peng Z, Kurgan L. 2012. On the complementarity of the consensus-based disorder prediction. *Pac Symp Biocomput Pac Symp Biocomput* 176–187.

Perez M, García-Barreno B, Melero JA, Carrasco L, Guinea R. 1997. Membrane permeability changes induced in Escherichia coli by the SH protein of human respiratory syncytial virus. *Virology* **235**: 342–351.

Pernot P, Theveneau P, Giraud T, Fernandes RN, Nurizzo D, Spruce D, Surr J, McSweeney S, Round A, F Felisaz, et al. 2010. New beamline dedicated to solution scattering from biological macromolecules at the ESRF. *J Phys Conf Ser* **247**: 012009.

Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HDT, Konarev PV, Svergun DI. 2012. New developments in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* **45**: 342–350.

Pfaller CK, Cattaneo R, Schnell MJ. 2015. Reverse genetics of Mononegavirales: How they work, new vaccines, and new cancer therapeutics. *Virology* **479–480**: 331–344.

Pflug A, Guilligay D, Reich S, Cusack S. 2014. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516**: 355–360.

Pham QN, Biacchesi S, Skiadopoulos MH, Murphy BR, Collins PL, Buchholz UJ. 2005. Chimeric recombinant human metapneumoviruses with the nucleoprotein or phosphoprotein open reading frame replaced by that of avian metapneumovirus exhibit improved growth in vitro and attenuation in vivo. *J Virol* **79**: 15114–15122.

Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, et al. 2013. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinforma Oxf Engl* **29**: 845–854.

Putnam CD, Hammel M, Hura GL, Tainer JA. 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **40**: 191–285.

Rahmeh AA, Li J, Kranzusch PJ, Whelan SPJ. 2009. Ribose 2'-O methylation of the vesicular stomatitis virus mRNA cap precedes and facilitates subsequent guanine-N-7 methylation by the large polymerase protein. *J Virol* **83**: 11043–11050.

Rahmeh AA, Morin B, Schenk AD, Liang B, Heinrich BS, Brusic V, Walz T, Whelan SPJ. 2012. Critical phosphoprotein elements that regulate polymerase architecture and function in vesicular stomatitis virus. *Proc Natl Acad Sci U S A* **109**: 14628–14633.

Rahmeh AA, Schenk AD, Danek EI, Kranzusch PJ, Liang B, Walz T, Whelan SPJ. 2010. Molecular architecture of the vesicular stomatitis virus RNA polymerase. *Proc Natl Acad Sci U S A* **107**: 20075–20080.

Rambo RP, Tainer JA. 2013. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**: 477–481.

Rambo RP, Tainer JA. 2011. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**: 559–571.

Ramos J, Lazaridis T. 2011. Computational analysis of residue contributions to coiled-coil topology. *Protein Sci Publ Protein Soc* **20**: 1845–1855.

Rawlinson WD, Waliuzzaman Z, Carter IW, Belessis YC, Gilbert KM, Morton JR. 2003. Asthma exacerbations in children associated with rhinovirus but not human metapneumovirus infection. *J Infect Dis* **187**: 1314–1318.

Receveur-Brechot V, Durand D. 2012. How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr Protein Pept Sci* **13**: 55–75.

Reddy Chichili VP, Kumar V, Sivaraman J. 2013. Linkers in the structural biology of protein–protein interactions. *Protein Sci Publ Protein Soc* **22**: 153–167.

Reguera J, Gerlach P, Cusack S. 2016. Towards a structural understanding of RNA synthesis by negative strand RNA viral polymerases. *Curr Opin Struct Biol* **36**: 75–84.

Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crépin T, Hart D, Lunardi T, Nanao M, Ruigrok RWH, et al. 2014. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* **516**: 361–366.

Renner M, Bertinelli M, Leyrat C, Paesen GC, Saraiva de Oliveira LF, Huiskonen JT, Grimes JM. 2016. Nucleocapsid assembly in pneumoviruses is regulated by conformational switching of the N protein. *eLife* **5**: e12627.

Ruigrok RWH, Crépin T, Kolakofsky D. 2011a. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol* **14**: 504–510.

Ruigrok RWH, Crépin T, Kolakofsky D. 2011b. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol* **14**: 504–510.

Russell CJ, Luque LE. 2006. The structural basis of paramyxovirus invasion. *Trends Microbiol* **14**: 243–246.

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**: 1350–1354.

Saso A, Kampmann B. 2016. Vaccination against respiratory syncytial virus in pregnancy: a suitable tool to combat global infant morbidity and mortality? *Lancet Infect Dis*.

Scheres SHW. 2012. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**: 519–530.

Schildgen V, van den Hoogen B, Fouchier R, Tripp RA, Alvarez R, Manoha C, Williams J, Schildgen O. 2011. Human Metapneumovirus: lessons learned over the first decade. *Clin Microbiol Rev* **24**: 734–754.

Schmitt AP, Lamb RA. 2004. Escaping from the cell: assembly and budding of negative-strand RNA viruses. *Curr Top Microbiol Immunol* **283**: 145–196.

Schmitt PT, Ray G, Schmitt AP. 2010. The C-terminal end of parainfluenza virus 5 NP protein is important for virus-like particle production and M-NP protein interaction. *J Virol* **84**: 12810–12823.

Schneider U, Blechschmidt K, Schwemmle M, Staeheli P. 2004. Overlap of interaction domains indicates a central role of the P protein in assembly and regulation of the Borna disease virus polymerase complex. *J Biol Chem* **279**: 55290–55296.

Schneider U, von Messling V, Devaux P, Cattaneo R. 2002. Efficiency of measles virus entry and dissemination through different receptors. *J Virol* **76**: 7460–7467.

Schowalter RM, Chang A, Robach JG, Buchholz UJ, Dutch RE. 2009. Low-pH triggering of human metapneumovirus fusion: essential residues and importance in entry. *J Virol* **83**: 1511–1522.

Schowalter RM, Smith SE, Dutch RE. 2006. Characterization of human metapneumovirus F protein-promoted membrane fusion: critical roles for proteolytic processing and low pH. *J Virol* **80**: 10931–10941.

Shaikh FY, Cox RG, Lifland AW, Hotard AL, Williams JV, Moore ML, Santangelo PJ, Crowe JE. 2012. A critical phenylalanine residue in the respiratory syncytial virus fusion protein cytoplasmic tail mediates assembly of internal viral proteins into viral filaments and particles. *mBio* **3**.

Shoemaker BA, Portman JJ, Wolynes PG. 2000. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. *Proc Natl Acad Sci* **97**: 8868–8873.

Slack MS, Easton AJ. 1998. Characterization of the interaction of the human respiratory syncytial virus phosphoprotein and nucleocapsid protein using the two-hybrid system. *Virus Res* **55**: 167–176.

Smart OS, Womack TO, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, Bricogne G. 2012. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr D Biol Crystallogr* **68**: 368–380.

Sourimant J, Rameix-Welti M-A, Gaillard A-L, Chevret D, Galloux M, Gault E, Eléouët J-F. 2015. Fine mapping and characterization of the L-polymerase-binding domain of the respiratory syncytial virus phosphoprotein. *J Virol* **89**: 4421–4433.

Sutherland KA, Collins PL, Peeples ME. 2001. Synergistic effects of gene-end signal mutations and the M2-1 protein on transcription termination by respiratory syncytial virus. *Virology* **288**: 295–307.

Svergun D, Barberato C, Koch MHJ. 1995. CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J Appl Crystallogr* **28**: 768–773.

Svergun DI. 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* **25**: 495–503.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725–2729.

Tang RS, Nguyen N, Cheng X, Jin H. 2001. Requirement of Cysteines and Length of the Human Respiratory Syncytial Virus M2-1 Protein for Protein Function and Virus Viability. *J Virol* **75**: 11328–11335.

Tanner SJ, Ariza A, Richard C-A, Kyle HF, Dods RL, Blondot M-L, Wu W, Trincão J, Trinh CH, Hiscox JA, et al. 2014. Crystal structure of the essential transcription antiterminator M2-1 protein of human respiratory syncytial virus and implications of its phosphorylation. *Proc Natl Acad Sci U S A* **111**: 1580–1585.

Tarbouriech N, Curran J, Ruigrok RW, Burmeister WP. 2000. Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nat Struct Biol* **7**: 777–781.

Tawar RG, Duquerroy S, Vonrhein C, Varela PF, Damier-Piolle L, Castagné N, MacLellan K, Bedouelle H, Bricogne G, Bhella D, et al. 2009. Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science* **326**: 1279–1283.

Te Wierik MJ, Nguyen DT, Beersma MF, Thijsen SF, Heemstra KA. 2012. An outbreak of severe respiratory tract infection caused by human metapneumovirus in a residential care facility for elderly in Utrecht, the Netherlands, January to March 2010. *Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull* **17**.

Thomas JMH, Keegan RM, Bibby J, Winn MD, Mayans O, Rigden DJ. 2015. Routine phasing of coiled-coil protein crystal structures with AMPLE. *IUCrJ* **2**: 198–206.

Tompa P. 2011. Unstructural biology coming of age. *Curr Opin Struct Biol* **21**: 419–425.

Tran T-L, Castagné N, Bhella D, Varela PF, Bernard J, Chilmonczyk S, Berkenkamp S, Benhamo V, Grznarova K, Grosclaude J, et al. 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 protomer. *J Gen Virol* **88**: 196–206.

Tran T-L, Castagné N, Dubosclard V, Noinville S, Koch E, Moudjou M, Henry C, Bernard J, Yeo RP, Eléouët J-F. 2009. The respiratory syncytial virus M2-1 protein forms tetramers and interacts with RNA and P in a competitive manner. *J Virol* **83**: 6363–6374.

Tremaglio CZ, Noton SL, Deflubé LR, Fearns R. 2013. Respiratory syncytial virus polymerase can initiate transcription from position 3 of the leader promoter. *J Virol* **87**: 3196–3207.

van den Hoogen BG, Bestebroer TM, Osterhaus ADME, Fouchier RAM. 2002. Analysis of the genomic sequence of a human metapneumovirus. *Virology* **295**: 119–132.

van den Hoogen BG, de Jong JC, Groen J, Kuiken T, de Groot R, Fouchier RA, Osterhaus AD. 2001. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat Med* **7**: 719–724.

van den Hoogen BG, Herfst S, Sprong L, Cane PA, Forleo-Neto E, de Swart RL, Osterhaus ADME, Fouchier RAM. 2004. Antigenic and genetic variability of human metapneumoviruses. *Emerg Infect Dis* **10**: 658–666.

van den Hoogen BG, van Doornum GJJ, Fockens JC, Cornelissen JJ, Beyer WEP, de Groot R, Osterhaus ADME, Fouchier RAM. 2003. Prevalence and clinical symptoms of human metapneumovirus infection in hospitalized patients. *J Infect Dis* **188**: 1571–1577.

Volkov VV, Svergun DI. 2003. Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* **36**: 860–864.

Voss JE, Vaney M-C, Duquerroy S, Vonrhein C, Girard-Blanc C, Crublet E, Thompson A, Bricogne G, Rey FA. 2010. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* **468**: 709–712.

Walsh EE, Peterson DR, Falsey AR. 2008. Human metapneumovirus infections in adults: another piece of the puzzle. *Arch Intern Med* **168**: 2489–2496.

Walter TS, Diprose JM, Mayo CJ, Siebold C, Pickford MG, Carter L, Sutton GC, Berrow NS, Brown J, Berry IM, et al. 2005. A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr D Biol Crystallogr* **61**: 651–657.

Wei Y, Zhang Y, Cai H, Mirza AM, Iorio RM, Peeples ME, Niewiesk S, Li J. 2014. Roles of the putative integrin-binding motif of the human metapneumovirus fusion (f) protein in cell-cell fusion, viral infectivity, and pathogenesis. *J Virol* **88**: 4338–4352.

Welsch S, Kolesnikova L, Krähling V, Riches JD, Becker S, Briggs JAG. 2010. Electron Tomography Reveals the Steps in Filovirus Budding. *PLoS Pathog* **6**. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2861712/ (Accessed September 28, 2016).

Wen X, Krause JC, Leser GP, Cox RG, Lamb RA, Williams J, Crowe JE, Jardetzky TS. 2012. Structure of the Human Metapneumovirus Fusion Protein with Neutralizing Antibody Identifies a Pneumovirus Antigenic Site. *Nat Struct Mol Biol* **19**: 461–463.

Wernimont A, Edwards A. 2009. In situ proteolysis to generate crystals for structure determination: an update. *PloS One* **4**: e5094.

Whelan SPJ, Barr JN, Wertz GW. 2004. Transcription and replication of nonsegmented negative-strand RNA viruses. *Curr Top Microbiol Immunol* **283**: 61–119.

White JM, Whittaker GR. 2016. Fusion of Enveloped Viruses in Endosomes. *Traffic Cph Den* **17**: 593–614.

Williams JV, Crowe JE, Enriquez R, Minton P, Peebles RS, Hamilton RG, Higgins S, Griffin M, Hartert TV. 2005a. Human Metapneumovirus Infection Plays an Etiologic Role in Acute Asthma Exacerbations Requiring Hospitalization in Adults. *J Infect Dis* **192**: 1149–1153.

Williams JV, Edwards KM, Weinberg GA, Griffin MR, Hall CB, Zhu Y, Szilagyi PG, Wang CK, Yang C-F, Silva D, et al. 2010. Population-based incidence of human metapneumovirus infection among hospitalized children. *J Infect Dis* **201**: 1890–1898.

Williams JV, Harris PA, Tollefson SJ, Halburnt-Rush LL, Pingsterhaus JM, Edwards KM, Wright PF, Crowe JE. 2004. Human metapneumovirus and lower respiratory tract disease in otherwise healthy infants and children. *N Engl J Med* **350**: 443–450.

Williams JV, Martino R, Rabella N, Otegui M, Parody R, Heck JM, Crowe JE. 2005b. A Prospective Study Comparing Human Metapneumovirus with Other Respiratory Viruses in Adults with Hematologic Malignancies and Respiratory Tract Infections. *J Infect Dis* **192**: 1061–1065.

Williams JV, Wang CK, Yang C-F, Tollefson SJ, House FS, Heck JM, Chu M, Brown JB, Lintao LD, Quinto JD, et al. 2006. The Role of Human Metapneumovirus in Upper Respiratory Tract Infections in Children: A 20-Year Experience. *J Infect Dis* **193**: 387–395.

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, et al. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**: 235–242.

Winter G, Lobley CMC, Prince SM. 2013. Decision making in xia2. *Acta Crystallogr D Biol Crystallogr* **69**: 1260–1273.

Yabukarski F, Lawrence P, Tarbouriech N, Bourhis J-M, Delaforge E, Jensen MR, Ruigrok RWH, Blackledge M, Volchkov V, Jamin M. 2014. Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. *Nat Struct Mol Biol* **21**: 754–759.

Yang ZR, Thomson R, McNeil P, Esnouf RM. 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinforma Oxf Engl* **21**: 3369–3376.

Yao Q, Compans RW. 2000. Filamentous particle formation by human parainfluenza virus type 2. *J Gen Virol* **81**: 1305–1312.

Zhao Y, Chapman DAG, Jones IM. 2003. Improving baculovirus recombination. *Nucleic Acids Res* **31**: E6-6.

Zhou H, Cheng X, Jin H. 2003. Identification of Amino Acids That Are Critical to the Processivity Function of Respiratory Syncytial Virus M2-1 Protein. *J Virol* **77**: 5046–5053.

8 APPENDIX

8.1 List of Publications

Wittmann S, **Renner M**, Watts B, Adams O, Huseyin M, Baejen C, El Omari K, Kilchert C, Heo DH, Kecman T, Cramer P, Grimes JM, Vasiljeva L. 2016. The conserved termination factor Seb1 bridges RNA polymerase II and nascent RNA. *Nat Commun* (accepted).

Renner M, Bertinelli M, Leyrat C, Paesen GC, Saraiva de Oliveira LF, Huiskonen JT, Grimes JM. 2016. Nucleocapsid assembly in pneumoviruses is regulated by conformational switching of the N protein. *eLife* **5**: e12627.

Leyrat C, Paesen GC, Charleston J, **Renner M**, Grimes JM. 2014. Structural Insights into the Human Metapneumovirus Glycoprotein Ectodomain. *J Virol* **88**: 11611–11616.

Liebherr RB, **Renner M**, Gorris HH. 2014. A Single Molecule Perspective on the Functional Diversity of in Vitro Evolved β-Glucuronidase. *J Am Chem Soc* **136**: 5949–5955.

Leyrat C, **Renner M**, Harlos K, Huiskonen JT, Grimes JM. 2014. Drastic changes in conformational dynamics of the antiterminator M2-1 regulate transcription efficiency in Pneumovirinae. *eLife* **3**: e02674.

Huiskonen JT, Parsy M-L, Li S, Bitto D, **Renner M**, Bowden TA. 2014. Averaging of viral envelope glycoprotein spikes from electron cryotomography reconstructions using Jsubtomo. *J Vis Exp JoVE* e51714.

Leyrat C, **Renner M**, Harlos K, Huiskonen JT, Grimes JM. 2014. Structure and self-assembly of the calcium binding matrix protein of human metapneumovirus. *Struct Lond Engl* **22**: 136–148.

Leyrat C, **Renner M**, Harlos K, Grimes JM. 2013. Solution and Crystallographic Structures of the Central Region of the Phosphoprotein from Human Metapneumovirus. *PLOS ONE* **8**: e80371.