Structural analysis and biochemical characterization of human metapneumovirus L protein

Yeo, Tiong Han

2018

Yeo, T. H. (2018). Structural analysis and biochemical characterization of human metapneumovirus L protein. Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/89014>

<https://doi.org/10.32657/10220/46071>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**STRUCTURAL ANALYSIS AND BIOCHEMICAL
CHARACTERIZATION OF HUMAN METAPNEUMOVIRUS
L PROTEIN**

YEO TIONG HAN

SCHOOL OF BIOLOGICAL SCIENCES

2018

**STRUCTURAL ANALYSIS AND BIOCHEMICAL
CHARACTERIZATION OF HUMAN METAPNEUMOVIRUS
L PROTEIN**

YEO TIONG HAN

School of Biological Sciences

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of

Masters of Science

2018

Acknowledgement

I would express the deepest appreciation to my supervisor Assoc Prof Julien Lescar for the opportunity to take on this project and for his valuable advice that was given throughout the course of this project. Without his support, this dissertation would not have been possible. I would also like to express my gratitude to Dr Qian Xinlei for her mentoring throughout this project and imparting skills and techniques.

I would like to thank members of the JL team- Dr Abbas El Sahili, Dr Goh Boon Chong, Dr Akshita Kumar Dhawan, Wong Yee Hwa, Lin Jianqing, Tan Siok Thing, Swathi Pasunooti, Farhana Abu Bakar, Jia Huan, Liu Chong and Grace Chan for their support as we worked together in a team, fighting alongside in this long and dark battle of research.

I would like to thank the Nanyang Institute of Structural Biology (NISB) for the support and facilities that were made available to me for my project.

I would like to thank Dr Muhammad Fursham Bin Hamid for guidance on radiolabelled work and generous use of radioactive facilities.

I would like to extend my gratitude to Dr Shashi Bushan, for his expertise in cryo electron microscopy and single particle analysis.

I would like to thank my family for the continuous support and love that was given unconditionally during this period of uncertainty. This is a long and difficult journey about self-discovery, as I seek to understand more about myself.

To friends and others whom I might have missed out, thank you for your unwavering support.

“Science is the observation of things possible, whether present or past”

Leonardo da Vinci

Table of Contents

TABLE OF CONTENTS	I
ABBREVIATIONS	III
LIST OF FIGURES	V
LIST OF TABLES	VI
ABSTRACT	VII
CHAPTER 1 INTRODUCTION	1
1.1 HUMAN METAPNEUMOVIRUS	1
1.2 THE INFECTION CYCLE	5
1.3 L PROTEIN	7
1.4 P PROTEIN	12
1.5 REPLICATION AND TRANSCRIPTION	13
1.6 TREATMENT OF HMPV INFECTION	15
1.7 AIMS AND OBJECTIVES	16
CHAPTER 2 MATERIALS AND METHODS	18
2.1 MATERIALS	18
2.2 METHODS	23
2.2.1 PROTEIN EXPRESSION 2.2.1.1 HMPV P <i>E. COLI</i> EXPRESSION	23
2.2.1.2 <i>E. COLI</i> CELL HARVEST	24
2.2.1.3 HMPV P PURIFICATION	24
2.2.1.4 BACMID PREPARATION AND ISOLATION	25
2.2.1.5 INSECT CELL CULTURE	26
2.2.1.6 TRANSFECTION OF INSECT CELLS	26
2.2.1.7 AMPLIFICATION OF BACULOVIRUS	27
2.2.1.8 TITRATION OF EMBAC ^Y	28
2.2.1.9 INSECT CELL PROTEIN EXPRESSION	28
2.2.1.10 PURIFICATION OF HMPV L AND P PROTEIN COMPLEX	28
2.2.1.11 WESTERN BLOTTING	29
2.2.2 IN VITRO RNA SYNTHESIS	30
2.2.3 STRUCTURAL ANALYSIS	31
2.2.3.1 NEGATIVE STAINING AND IMAGE ANALYSIS OF ELECTRON MICROSCOPY ANALYSIS	31
2.2.3.2 DATA PROCESSING	31
CHAPTER 3 RESULTS	32
3.1 PURIFICATION OF HMPV L AND P PROTEINS	32
3.1.1 PURIFICATION OF HMPV P	32
3.1.2 PURIFICATION OF HMPV L USING HMPV P PROTEIN EXPRESSED BY <i>E. COLI</i>	36

3.1.3 COEXPRESSION AND PURIFICATION OF HMPV L AND P PROTEIN FROM INSECT CELLS	39
3.2 BIOCHEMICAL CHARACTERIZATION OF HMPV L TRANSCRIPTION	46
3.2.1 ISOLATED RECOMBINANT HMPV L IS CAPABLE OF <i>DE NOVO</i> RNA SYNTHESIS	46
3.2.2 EFFECT OF NTP CONCENTRATION ON RNA SYNTHESIS	49
3.2.3 EFFECTS OF DIVALENT CATIONS ON RNA TRANSCRIPTION OF HMPV L	52
3.3 STRUCTURAL STUDY OF HMPV L	56
3.3.1 PREDICTIVE MODELING OF HMPV L PROTEIN	56
3.3.2 MOLECULAR ARCHITECTURE OF THE HMPV L AND P PROTEIN COMPLEX	60
CHAPTER 4 DISCUSSION	63
4.1 ISOLATION OF HMPV L	63
4.2 BIOPHYSICAL CHARACTERISATION OF HMPV POLYMERASE COMPLEX	65
4.3 CHARACTERIZATION OF HMPV L CATALYTIC RdRp SITE AND FACTORS AFFECTING RNA REPLICATION	66
4.4 STRUCTURAL ANALYSIS	68
CHAPTER 5 FUTURE WORK AND CONCLUSION	70
5.1. FUTURE WORK	70
5.2 CONCLUSION	72
CHAPTER 6 REFERENCES	73
CHAPTER 7 APPENDIX	83

Abbreviations

<u>Abbreviation</u>	<u>Full name</u>
AA	Amino acid
ATI	Acute respiratory tract infection
BSA	Bovine serum albumin
CryoEM	Cryo-electron microscopy
CTD	C-terminal domain
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
GTP	Guanosine Triphosphate
HMPV	Human Metapneumovirus
HRP	Horseradish peroxidase
IEX	Ion exchange
IPTG	Isopropyl β -D-1-Thiogalactopyranoside
kDa	kilodalton
L protein	Large Protein
LC/MS	Liquid chromatography mass spectrometry
Le	Leader
LIC	Ligation independent cloning
min	Minutes
MOI	Multiplicity of infection
mRNA	Messenger Ribonucleic Acid
NNS	Non-segmented negative sensed
nt	Nucleotide
NTD	N-terminal domain

NTP	Nucleoside Triphosphate
P protein	Phospho-protein
P ₂	Virus Passage 2
P ₃	Virus Passage 3
PAGE	Polyacrylamide Gel Electrophoresis
PC	Positive control
PCR	Polymerase Chain Reaction
PDB	Protein Database
RdRp	RNA dependent RNA polymerase
RNA	Ribonucleic Acid
RNP	Ribonucleoprotein
rpm	Rotations per minute
RSV	Respiratory Syncytial Virus
SDS	Sodium Dodecyl Sulfate
SEC	Size exclusion chromatography
TBS	Tris buffered saline
TEV	Tobacco Etch Virus
UTP	Uridine Triphosphate
VSV	Vesicular Stomatitis Virus
X-gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
YFP	Yellow fluorescence protein

List of figures

FIGURE 1 HUMAN METAPNEUMOVIRUS	3
FIGURE 2 THE LIFE CYCLE OF HMPV	5
FIGURE 3 VSV L PROTEIN	9
FIGURE 4 PURIFICATION OF HMPV P PROTEIN EXPRESSED IN <i>E. COLI</i>	34
FIGURE 5 COPURIFICATION OF HMPV L USING HMPV P PURIFIED FROM <i>E. COLI</i>	37
FIGURE 6 CO-EXPRESSION AND PURIFICATION OF HMPV L AND P FROM INSECT CELLS	41
FIGURE 7 PURIFICATION OF HMPV L AND P AND CONFIRMATION OF THEIR IDENTITY.....	44
FIGURE 8 <i>DE NOVO</i> RNA SYNTHESIS BY ISOLATED HMPV POLYMERASE	47
FIGURE 9 EFFECTS OF NTP CONCENTRATION OF HMPV POLYMERASE <i>DE NOVO</i> RNA SYNTHESIS.	50
FIGURE 10 EFFECTS OF CATIONS ON HMPV POLYMERASE ACTIVITY	54
FIGURE 11 PREDICTED STRUCTURE OF THE HMPV L PROTEIN	58
FIGURE 12 SINGLE PARTICLE ANALYSIS OF HMPV L	61

List of tables

TABLE 1 COMPOSITION OF BACTERIA GROWTH MEDIA AND SELECTION PLATES	18
TABLE 2 BUFFER AND THEIR APPLICATION AND COMPOSITION	19
TABLE 3 LIST OF PRIMERS.....	22

Abstract

Human metapneumovirus (HMPV) is a nonsegmented negative sense RNA virus (NNS), first described in 2001, that causes acute respiratory tract infection in children and elderly. As HMPV is relatively new, most of our current understanding on this virus has been inferred from homologous viruses of the NNS family. The large (L) protein is a multi-domain protein involved in replication and transcription of viral RNA. The replication and transcription process is mediated by the viral (P) phosphoprotein. The L protein is expressed in low abundance in the infected cells due to the position of the cognate gene on the viral genome and little is known about the L protein. Our study shows that HMPV L protein interacts with P protein to form a properly folded and active HMPV polymerase. Using recombinant protein expression techniques, we describe the expression and purification of functional HMPV polymerase complex. The purified polymerase complex was found to be functional in *de novo* RNA synthesis and to recognize HMPV leader (le) sequence for replication. Structural analysis of HMPV polymerase complex using electron microscopy reveals ring like structures presumably representing the RdRp and capping domains. Due to the recent advances in cryo-electron microscopy techniques, the recombinant HMPV polymerase complex as obtained here can be used for high-resolution cryo-electron microscopy analysis to determine the structure at atomic resolution. Knowledge gained will assist the development of anti-HMPV L therapeutics targeting the replication and transcription of the virus.

Chapter 1 Introduction

1.1 Human Metapneumovirus

Acute respiratory tract infection (RTI) is an important cause of disease and death globally. It accounts for about 20% of mortality and is the second leading cause of death in children aged under 5 (B. G. Williams, Gouws, Boschi-Pinto, Bryce, & Dye, 2002). Rhinovirus, Influenza virus, human parainfluenza virus (PIV) and respiratory syncytial virus (RSV) have been well established as causing respiratory tract infections and bronchiolitis. Human Metapneumovirus (HMPV), a previously unknown virus, was isolated in 2001 from the respiratory tract of a pediatric patient in the Netherlands (van den Hoogen et al., 2001). Unlike many respiratory tract-infecting viruses, HMPV infections usually occur from late winter to early spring (van den Hoogen et al., 2003). HMPV infections have been reported in both developed (Mackay et al., 2006) and developing countries (Al-Sonboli et al., 2006). HMPV infects all age groups. However, children under 5 (Al-Ayed, Asaad, Qureshi, & Ameen, 2014; Kusel et al., 2007) and elderly people are more susceptible to infections and developing of more severe form respiratory diseases (Falsey, Erdman, Anderson, & Walsh, 2003). HMPV is transmitted via infectious airborne droplets and aerosol. Patients infected with HMPV exhibit clinical manifestations similar to RSV infected patients (Kahn, 2006). HMPV causes acute respiratory infection in the upper respiratory tract. However, manifestation of severe form of disease such as lower respiratory tract infection (Caracciolo et al., 2008), pneumonia (Johnstone, Majumdar, Fox, & Marrie, 2008) and chronic lung disease (J. V. Williams, Tollefson, et al., 2005) have also been reported. In children, HMPV infection has been shown to induce

wheezing, leading to the development of asthma (Chung et al., 2006; Inoue & Shimojo, 2013). Although there are limited studies conducted on adult population, HMPV has been detected in 6.9% of adults with more severe or with less-well controlled asthma (J. V. Williams, Crowe, et al., 2005). Currently, there are no FDA approved drugs for treating HMPV infection and patients rely on supportive care and oxygen therapy. Although treating HMPV infection using Ribavirin has been reported (Kitanovski et al., 2013), administration of Ribavirin is a challenge for children as dosing is based on the technique of inhalation.

Figure 1 Human Metapneumovirus

Figure 1(Top): Electron micrograph of HMPV particle (Adapted from: Verena Schildgen et al. Clin. Microbiol. Rev. 2011;24:734-754)

Figure 1 (Middle): Taxonomy tree of NNS viruses.

Figure 1 (Bottom): Schematic of HMPV genome (Adapted from Cheemarla et al., Pathogens. 2015 Sep 18;4(3):682-96. doi: 10.3390/pathogens4030682.)

The size of the HMPV virion varies in size between 150 nm to 600 nm and is pleomorphic in nature (van den Hoogen et al., 2001). Little is known about HMPV, and our current understanding of HMPV is inferred from studies of other non-segmented negative-strand (NNS) viruses, mainly RSV and Vesicular Stomatitis Virus (VSV). HMPV is an enveloped, non-segmented, negative sensed RNA virus of *Pneumoviridae* family, within the order of *Mononegavirales* (Afonso et al., 2016). HMPV has a 13 kb genome that codes for 8 virus-specific genes and 9 viral proteins (van den Hoogen, Bestebroer, Osterhaus, & Fouchier, 2002). HMPV genome is translated in the following order: the nucleoprotein (N protein), the phosphoprotein (P protein), the matrix protein (M protein), the fusion glycoprotein protein (F protein), the putative transcription factor (M2-1, M2-2 protein), the small hydrophobic glycoprotein (SH protein), the attachment glycoprotein (G protein) and the large polymerase protein (L protein) (Biacchesi, Murphy, Collins, & Buchholz, 2007).

1.2 The infection cycle

Figure 2 The life cycle of HMPV

Figure 2 Schematic representation of the HMPV life cycle (Adapted from Verena Schildgen et al. Clin. Microbiol. Rev. 2011;24:734-754)

To facilitate virus entry, HMPV G protein first attaches to host cell receptor. This interaction will trigger the F protein (presented as F₀ precursor protein) to be cleaved by proteases to yield F₁ and F₂ subunits (Lamb & Jardetzky, 2007), which results in the fusion of the viral envelope with the host cell membrane (Chang, Masante, Buchholz, & Dutch, 2012). Following membrane fusion, the virus ribonucleoprotein (RNP) complex is released into the host cell cytoplasm (Collins, Fearn, & Graham, 2013). The RNP complex is important both for virus mRNA transcription and genome replication. It is comprised of the RNA wrapped by the N protein, P protein, M2 protein and also the L protein (Garcia, Garcia-Barreno, Vivo, & Melero, 1993). Transcription and replication start when the RNA wrapped-N protein is presented to the L protein by the P protein (Murphy et al., 2003; Murray, Loney, Murphy, Graham, & Yeo, 2001; Tran et al., 2007). The transcription complex is also further stabilized by the presence of the M2-1 protein (Collins, Hill, Cristina, & Grosfeld, 1996). The transcribed mRNAs are released from the RNP complex and subsequently translated to viral proteins via the host translation machinery. Replication of the viral genome is initiated in presence of the M2-2 protein and high amount of unbound N protein (N₀) (Bermingham & Collins, 1999). The newly synthesized genomic RNA, which is not subject to any modification, becomes encapsidated by the N₀ protein to form the N₁ protein (Horikami, Curran, Kolakofsky, & Moyer, 1992). Newly formed N₁, P and L protein associate to form the RNP complex. The newly formed RNP complex is trafficked to the membrane, where other newly synthesized virus structural proteins are located. RNP and other structural proteins are then packaged into a virus particle and progeny virus will eventually bud off from the cell to infect surrounding cells.

1.3 L Protein

The L protein is the largest protein encoded by the mononegavirales. It is responsible for virus genome replication and mRNA elongation, capping and methylation, with the help of accessory proteins, P and N proteins (Horikami et al., 1992; Morin, Rahmeh, & Whelan, 2012). The L protein of the NNS virus family has been categorized into 4 major domains and 6 conserved regions (Harcourt et al., 2001). Amino acid sequence alignments showed conserved residues amongst the NNS L protein, indicating that the L protein function might be conserved across viruses from various families (Govindarajan & Samal, 2004; Poch, Blumberg, Bougueleret, & Tordo, 1990). Biochemical studies have shown that the L protein of NNS virus contains enzymatic functions, such as RNA polymerization, polyribonucleotidyltransferase (PRNT) and methyltransferase (MTase). Currently, most studies of L protein have been conducted using purified L protein from VSV, Chandipura Virus, Sendai Virus and RSV (Morin et al., 2012; Noton, Deflube, Tremaglio, & Fearn, 2012; Ogino & Banerjee, 2010; Ogino, Kobayashi, Iwama, & Mizumoto, 2005). Due to the difficulties in expression and purification of the L protein, there is little structural and biochemical information available about HMPV L protein. Studies of the L protein is challenging due to the large protein size which precludes its expression in a bacterial protein expression system. However, thanks to the development of Baculovirus Expression Vector Systems (BEVS) over the past two decades, production of large proteins has been made possible (Fitzgerald et al., 2006). Full-length recombinant L protein can now be expressed by using BEVS (Noton et al., 2012; Rahmeh et al., 2010; Song et al., 2015). Although BEVS have made expression of the L protein possible, the reported expression levels of the L

protein remain low. Codon-optimization of the L synthetic gene improves the expression level of the L protein (Noton et al., 2012; Sourimant et al., 2015).

Figure 3 VSV L protein

Figure 3 (Top): Schematic organization of L protein domains

Figure 3 (Bottom): Ribbon diagram of VSV-L domains: polymerase domain (RdRp) is in cyan; capping domain (Cap), green; connector domain (CD), yellow; methyltransferase (MT), orange; C-terminal domain (CTD), red (Adapted from Liang et al., *Cell*, Volume 162, Issue 2, Pages 314-327 (July 2015) DOI: 10.1016/j.cell.2015.06.018)

Determining the atomic structure of large proteins has been hindered due to the limitation in technology in the past. However, development of new technologies has made it possible to solve the structure of large molecules at near atomic resolution. Although there are no crystal structures of NNS L protein, a recent study published in 2015 reported the structure of the VSV L protein at a resolution of 3.8 Å using cryo-electron microscopy (cryo-EM) (Song et al., 2015). VSV is a NNS virus belonging to the *Rhabdoviridae* family and was the first published structural report of a full-length L protein of any NNS viruses. VSV L is composed of 5 domains, the RNA-dependent RNA polymerase domain (RdRp domain); the capping domain (Cap domain); the connector domain (CD domain); the methyltransferase domain (MT domain) and the C-terminal domain (CTD domain). Mutational studies indicated that the N-terminal region of the L protein interacts with the P protein (Cevik, Holmes, et al., 2004; Chandrika, Horikami, Smallwood, & Moyer, 1995; Parks, 1994). However, there are variations of the P binding domain of L region between the different viruses.

Similar to other RNA polymerases, the VSV L protein resembles a right-hand with fingers, palm and thumb subdomains (Song et al., 2015). The RdRp domain resembles a doughnut-ring under negative staining conditions (Rahmeh et al., 2010). There are 3 conserved regions within the RdRp domain of the L protein. Although little information is known about conserved region I and II, some clues have been inferred from homologous proteins. Conserved region I is involved in Mg^{2+} coordination while conserved region II is involved in the selection of ribose from deoxyribose (Hansen, Long, & Schultz, 1997; Jacobo-Molina et al., 1993). GDNQ is an important conserved motif that lies within conserved region III of the L protein. The GDNQ motif has been reported to be essential for phosphodiester

bond formation, while mutation of the conserved motif has been shown to abolish RNA polymerization (Chattopadhyay, Raha, & Shaila, 2004; Noton et al., 2012; Sleat & Banerjee, 1993).

A priming loop from the capping domain protrudes into the RdRp domain suggesting that RNA synthesis and capping are coordinated in NNS viruses. In NNS virus, the viral mRNA is capped by a polyribonucleotidyltransferase (PRNTase) rather than a guanylyltransferase (Ogino & Banerjee, 2007). Newly synthesized mRNA is capped by the capping domain located in conserved region V of the L protein. During capping, a histidine residue in the conserved HR motif of the L protein forms a covalent linkage to the 5' end of the RNA (Ogino & Banerjee, 2010). The binding of RNA and histidine will result in the removal of γ and β phosphate from the 5' end of RNA, forming a 5'-monophosphate. A GTPase activity of the L protein removes a phosphate from a GTP forming an acceptor GDP (Ogino & Banerjee, 2007). The GDP acceptor binds to PRNTase via a GxxT motif located about 70 residues downstream of the HR motif. The new GDP motif is subsequently transferred onto the 5'-monophosphate via nucleophilic attack of the 5' α -phosphate of RNA and β -phosphate of the GDP, resulting in formation and release of the capped RNA (Barik, 1993; Ogino & Banerjee, 2007; Ogino et al., 2005).

Using S-adenosylmethionine as substrate, the methyl group is transferred onto the cap structure of the ribose-2'-O position of the first nucleotide. Another methyl group is subsequently transferred onto the N7 position of the guanosine cap. The methylation is done by the methyltransferase located in conserved region VI (Li, Wang, & Whelan, 2006). Conserved residues, GxGxG and K-D-K-E, have been identified from the VSV L protein as vital for the methylation of capped RNA (Li,

Fontaine-Rodriguez, & Whelan, 2005; Li et al., 2006). A structure of truncated HMPV L protein consisting of 406 amino acids residues at the C-terminal domain was solved at 2.2 Å using X-ray crystallography (Paesen et al., 2015). The truncation consists of methyltransferase domain of HMPV L. Biochemical assay confirmed the function through methylation of capped and uncapped RNA by the MTase activity of the truncated protein. A K-K-G motif was identified as a vital motif for methylation rather than the K-D-K-E motif in VSV. Unlike the VSV L, a novel pocket that can accommodate nucleoside was observed in the HMPV L MTase domain. A novel discovery of NTPase activity was discovered in the truncated protein, indicating possible role in capping of RNA.

1.4 P Protein

The P protein is a phosphoprotein of the mononegavirales virus that contains multiple serine and threonine residues. It is an essential co-factor for viral RNA synthesis. The P protein comprises 3 major regions, an intrinsically disordered N terminal domain (NTD); an oligomerisation domain and an intrinsically disordered C terminal domain (CTD). The NTD and CTD of P protein are highly variable amongst the various NNS viruses. The oligomerisation domain contains conserved sequences that have been shown to form an alpha helix. Four alpha-helices of the P protein associate to form a homo-tetrameric assembly (Chen, Ogino, & Banerjee, 2007; Leyrat, Renner, Harlos, & Grimes, 2013; Rahaman, Srinivasan, Shamala, & Shaila, 2004; Tarbouriech, Curran, Ruigrok, & Burmeister, 2000). No known catalytic functions have been reported for the P protein which has been shown to interact with the N protein (N₀, N₁) (Garcia et al., 1993; Renner et al., 2016), the M2 protein (Mason et al., 2003) and the L protein (Gao & Lenard, 1995; Noton et al., 2012; Sourimant et al., 2015) to aid in

transcriptional activities. Mapping studies revealed that the P protein of different viruses binds to their respective L protein via different domains (Chenik, Schnell, Conzelmann, & Blondel, 1998; Khattar, Yunus, & Samal, 2001; Sourimant et al., 2015).

1.5 Replication and Transcription

To initiate mRNA transcription from genomic RNA, the L/P complex first recognizes and binds the leader sequence on the 3' of the viral genome (Tremaglio, Noton, Deflube, & Fearn, 2013). The synthesis of RNA is controlled by specific promoter sequences within the le sequence (Cowton & Fearn, 2005). The L/P complex moves along the 3' of viral genomic RNA in search of specific signals till it reaches the 5' of the genome (Gattone et al., 2004). These specific signals are known as *cis*-acting gene start (GS) and gene end (GE) sequences, that controls the transcription activity of the polymerase (Kuo, Grosfeld, Cristina, Hill, & Collins, 1996). The GS sequence initiates mRNA transcription, while the GE sequence halts transcription, as it primes the newly synthesized RNA for polyadenylation and subsequently release of mRNA (Liuzzi et al., 2005). The polymerase then continues to scan for the next GS sequence for synthesis of the next gene mRNA (Fearn & Collins, 1999). During this process, the L/P transcribing complex might fall off from the RNA template in the intergenic sequences, resulting in a gradual decrease of mRNA expression for genes located nearer to the 5' end of the virus genome. The transcribed mRNA becomes capped and methylated by the corresponding subunits within the L protein and eventually released into the cytoplasm to be translated by host cell ribosomes.

To initiate replication, the L/P complex binds to the leader sequence and

synthesize nascent RNA from the 3' viral genomic RNA producing the antigenomic RNA. The polymerase does not respond to the cis-acting GE and GS signals during RNA synthesis to produce a full-length positive-sense RNA complement of the genome, the antigenome (Cowton, McGivern, & Fearn, 2006). In presence of antigenomic RNA, L/P complex binds to the promoter sequence on the trailer complementary (TrC) to initiate synthesis the genomic RNA. During replication, more genomic RNA is synthesized than antigenomic RNA (Hanley et al., 2010). Higher levels of genomic RNA is produced as the genomic promoter is more efficient than antigenomic promoter as the genome RNA has to support both RNA replication and transcription (Fearn, Collins, & Peeples, 2000). Unlike mRNA, both genome and antigenome RNA are not modified by the polymerase but are encapsidated with the virus nucleoprotein (Cowton et al., 2006).

Recent studies have shown that the recombinant L/P complex is able to synthesize RNA from the leader (Tremaglio et al., 2013) and the trailer complementary sequence (TrC) (Noton et al., 2012). RNA synthesis was detected at positions +1 and +3 of the RNA initiation site of the leader sequence (Tremaglio et al., 2013). The detection of RNA synthesis at +1 site indicates that the L protein recognizes specific sequences to initiate antigenomic RNA synthesis. However, a larger amount of transcribed RNA initiated at the +3 leader sequences was observed compared to +1 products. This suggests that the L polymerase recognizes specific sequences that are similar to GS sequence (Kuo, Fearn, & Collins, 1997). Likewise, the purified L/P complex was able to synthesize new leader RNA, using the trailer complementary sequence (Noton et al., 2012). However, modification of RNA sequences at the 3' end was observed.

This suggests that the L protein is able to modify the sequence that causes the RNA to form secondary structures, hence preventing RNA replication in the absence of the N protein.

A clear understanding of initiation of replication and transcription of the NNS virus is still lacking. However, models of NNS RNA virus transcription and replication initiation have been proposed (Noton & Fearn, 2015). None of the published works involved investigating the mechanism of replication and transcription by the HMPV L protein, but rather by the VSV and RSV L proteins. Mutation studies of other NNS viruses revealed differences in the amino acid residues on both the L and P protein at the putative interaction sites of the L protein (Cevik, Holmes, et al., 2004; Cevik, Kaesberg, Smallwood, Feller, & Moyer, 2004; Chandrika et al., 1995; Khattar et al., 2001; Parks, 1994; Sourimant et al., 2015) (**Supp Fig 3 and 4**). Hence, more information is needed to understand the mechanism that controls the L protein for replication and transcription and the interactions between HMPV L and HMPV P protein. Although one structure of VSV L protein was solved, there is still relatively little structural information of full-length NNS L proteins available. Currently, no full-length high-resolution structure of NNS L protein from a human pathogen is known (VSV is not a human pathogen).

1.6 Treatment of HMPV infection

Similar to RSV, the current treatment of HMPV infection is mainly supportive. Palivizumab (Synagis®), a humanized monoclonal antibody, that was developed against RSV has shown to bind to conserved neutralising epitope on HMPV (Ulbrandt et al., 2008). Another human monoclonal antibody (MPE8) have shown to neutralise both RSV and HMPV as it binds to conserved regions on the F protein (Corti et al., 2013). Although both monoclonal antibodies provide

protection against HMPV infection through preventing fusion of virus with host cell, the administration of antibody is invasive as Palivizumab is administered via intramuscular injection.

Ribavirin, a nucleoside analog, is effective in inhibiting transcription in a variety of RNA and DNA viruses, including HMPV (Haas, Thijsen, van Elden, & Heemstra, 2013). Ribavirin has shown to increase interleukin-2 production but down regulates T-cell cytokine production *in vitro* (Sookoian, Castano, Flichman, & Cello, 2004). Although Ribavirin can inhibit viral transcription, it has also shown to modulate T-cell response (Graci & Cameron, 2006).

Although intravenous immune globulin (IVIG) and oral or aerosolized ribavirin combination treatment is effective in severe HMPV infection treatment (Shahda, Carlos, Kiel, Khan, & Hage, 2011), the combination therapy is expensive and have many disadvantages. Ribavirin is administered via nebulisation by a small particle aerosol generator. HMPV primarily infects the infants and elderly, hence, making administration of Ribavirin difficult. IVIG has also shown to be associated with side effects in children with congenital heart disease (Wyde, Chetty, Jewell, Boivin, & Piedra, 2003). Therefore, new antiviral against HMPV is needed to bridge the shortcoming of current therapeutics.

1.7 Aims and objectives

There is currently a lack of understanding of HMPV L function and structure. It will be interesting to determine the 3D structure of HMPV L to better understand HMPV RNA replication. Therefore, this study aims at:

1. to produce HMPV L in a functional state and in a quantity compatible with structure determination

2. to understand factors influencing the transcription and replication profile of HMPV L polymerase especially the interaction between L and P.

Recent research has shown that recombinant functional L protein recognizes viral specific sequences and is able to synthesize RNA. Therefore, in our study, we hypothesized that the recombinant full-length HMPV L/P complex is able to recognize HMPV specific sequences. We also hypothesized that the 3D structure of HMPV L will be similar to VSV-L as both L protein shares 22% similarities. To our knowledge, there is no published method for purification of the full length HMPV L protein. Overexpression of HMPV L will give structural insights as well as afford biochemical studies of HMPV polymerase. Knowledge obtained from our study may be used for screening potential therapeutic compounds against NNS virus replication and transcription process. Development of antiviral therapeutics will aid in decreasing the incidences of HMPV infection on human health. Information from this study can be used for development of antiviral therapeutics that are specific towards targetting HMPV transcription and replication process, thus reducing the economic cost associated to the treatment of HMPV infection.

Chapter 2 Materials and methods

2.1 Materials

Table 1 below summarizes the recipe of growth media and selection plate for bacterial plasmid and protein expression.

Table 1 Composition of bacteria growth media and selection plates

Growth media/agar	Composition (per litre (L))
<i>Luria-Bertani (LB) broth</i>	10 g Tryptone 5 g Yeast Extract 10 g Sodium Chloride
2x YT broth	20 g Tryptone 10 g Yeast Extract 10 g Sodium Chloride
pNIC28-Bsa4 selection plate	LB agar supplemented with: 25 µg/mL Chloramphenicol 50 µg/mL Kanamycin
Bacmid selection plate	LB agar supplemented with: 50 µg/mL Kanamycin 7 µg/mL Gentamycin 10 µg/mL Tetracycline 40 µg/mL IPTG 45 µg/mL Xgal

Table 2 below summarizes the application and composition of buffers.

Table 2 Buffer and their application and composition

Buffer/Solutions	Application	Composition (per liter (L))
HMPV P lysis buffer	Protein purification	20 mM Na HEPES pH 7.5 500 mM NaCl 1 mM DTT 10 % (v/v) glycerol 10 mM imidazole Protease Inhibitor (1: 1000) (v/v)
HMPV P buffer A	Protein Purification	20 mM Na HEPES pH 7.5 500 mM NaCl 1 mM DTT 10 % (v/v) glycerol
HMPV P elution buffer	Protein purification	20 mM Na HEPES pH 7.5 500 mM NaCl 1 mM DTT 10 % (v/v) glycerol 500 mM imidazole
HMPV P IEX buffer A	Protein purification	20 mM Na HEPES pH 7.5 500 mM NaCl 1 mM DTT 10 % (v/v) glycerol
HMPV P IEX buffer B	Protein Purification	20 mM Na HEPES pH 7.5

		1 M NaCl 1 mM DTT 10 % (v/v) glycerol
HMPV P SEC buffer	Protein Purification	20 mM Na HEPES pH 7.5 300 mM NaCl 1 mM DTT 10 % (v/v) glycerol
HMPV L lysis buffer HMPV L/P lysis buffer	Protein purification	20 mM Na HEPES pH 7.4 300 mM NaCl 1 mM DTT 5 % (v/v) glycerol 10 mM imidazole Protease Inhibitor (1: 1000) (v/v)
HMPV L buffer A HMPV L/P buffer A	Protein purification	20 mM Na HEPES pH 7.4 300 mM NaCl 1 mM DTT 5 % (v/v) glycerol
HMPV L elution buffer HMPV L/P elution buffer	Protein purification	20 mM Na HEPES pH 7.4 300 mM NaCl 1 mM DTT 5 % (v/v) glycerol 500 mM imidazole
10x gel running buffer	SDS-Page	30.2 g Tris base 250 mM) 144 glycine (1.92 M) 10 g SDS (1%)

10x transfer buffer	Western blot	30.2 g Tris (250 mM) 144 g glycine (1.92M)
10x Tris buffer saline (TBS)	Western blot	60.6 Tris base (0.5 M) 8.7.8 g NaCl (1.5M) pH 7.4

Table 3 below summarizes the list of primers used and the application of the primers.

Table 3 List of primers

Name	Sequence (5'-3')	Application
pFBD_HMPV_L _{D745A} _For	GATGGACTGGTTCGCGCCGTTGA GC	Mutagenesis
pFBD_HMPV_L _{D745A} _Rev	GACGTGTCCAAGCCTGTGAAGCT GTC	
pFBD_HMPV_L_10533- 10544_for	GTATACGGACCTTTAATTCAAC	Sequencing
pFBD_HMPV_L_9668- 9682_for	GGGTTTCCAACAGCC	Sequencing
pFBD_HMPV_L_8804- 8818_for	TCAAGACCCGTCGTG	Sequencing
pFBD_HMPV_L_7950- 7964_for	TCCCGTGACCTGCAG	Sequencing
pFBD_HMPV_L_7093- 7112_for	GCTATTAACGGCGAAGATAT	Sequencing
pFBD_HMPV_L_6243- 6257_for	AGCGACCAACACATC	Sequencing
pFBD_HMPV_L_5387- 5407_for	GTAAC TACATGATCTTGCTGC	Sequencing
pFBD_HMPV_L_4299- 4284_rev	GCTATGCATCAGCTGC	Sequencing
pFBD_HMPV_P_10654- 10676_for	CCTATAAATATTCCGGATTATTC	Sequencing
pFBD_HMPV_P_11693- 11678_rev	GGCCTTTGAATTCCGC	Sequencing

2.2 Methods

2.2.1 Protein Expression

2.2.1.1 HMPV P *E. coli* expression

The HMPV P gene (Genescript Biotech) was chemically synthesized and cloned into pUC57. HMPV P sequence was subcloned downstream of TEV cleavage sequence on pNIC28-Bsa4 from pUC57_HMPV_P using ligation independent cloning by protein production platform. Modified plasmid was sequenced and subsequently transformed into BL21 (DE3) Rosetta T1R via heat shock. 400 µL of LB was added to heat shocked cells and allowed to recover at 37 °C with agitation at 225 rpm for 1 hour. The recovered cells were plated onto LB plate supplemented with kanamycin and chloramphenicol and incubated in 37 °C overnight. A single colony of the transformed cells were grown in a LB starter culture supplemented with kanamycin and chloramphenicol at 37 °C overnight, with agitation at 225 rpm. Cells were inoculated into larger culture flask containing 2x YT supplemented with kanamycin and chloramphenicol, at a dilution of 1:100 and cultured at 37 °C with agitation at 225 rpm. Optical density at 600 nm of the cells was checked every 2 hours or till cells reach a OD of 0.8. The culture was chilled in the cold room at 4 °C for 1 hour. The chilled cells were induced with 0.5 mM of IPTG at 16 °C for 16 hours.

2.2.1.2 *E. coli* cell harvest

Cells were harvested via centrifugation at 1200g at 4 °C for 15 min. Cell pellet was collected to store at -20 °C or resuspended in cold HMPV lysis buffer for protein purification.

2.2.1.3 HMPV P purification

HMPV P cell pellets were resuspended in cold HMPV lysis buffer at a ratio of 1:10 (w:v). Cells were lysed via sonication at 1 min, 5 sec pulse (on and off) and 45 amp for every 10 mL of resuspension mixture. Lysed cell mixture was centrifuged at 22,000 g at 4°C for 30 min to pellet down insoluble cell extract. Supernatant was incubated with nickel-NTA beads washed and pre-equilibrated with 2 x BV of HMPV P lysis buffer for 1 hour. Beads were packed in gravity flow column and subjected to imidazole washes with HMPV P buffer A at 10 mM and 25 mM imidazole (30x and 20x BV respectively). HMPV P was eluted in HMPV elution buffer and protein was diluted to 50 mM NaCl with HMPV IEX buffer A. Diluted HMPV P was loaded onto HiTrap Q HP anion exchange column (GE Healthcare) at 1 mL/min. Column was washed with 5 CV at 50 mM NaCl to elute unbounded protein. A gradient elution was performed and all fractions were collected. Collected fractions were analyzed on SDS PAGE and stained with coomassie blue stain for visualisation of proteins. Fractions containing HMPV P were pooled and concentrated using Vivaspın15R 100 kDa cutoff concentrator to a volume of 5 mL (Sartorius). The concentrated samples were loaded onto a HMPV P buffer A pre-equilibrated HiLoad 16/600 Superdex 200 pg (GE Healthcare) for size exclusion chromatography. Protein was eluted into HMPV buffer A. Eluted HMPV P was aliquoted into 20 µL and snapped frozen in liquid nitrogen. Snapped freeze HMPV P was stored at -80 °C.

2.2.1.4 Bacmid preparation and isolation

HMPV L gene (Genescript Biotech) was chemically synthesized and cloned into pFASTBAC-LIC-BseR1, downstream of 6x His-tag TEV sequence by Genescript Biotech. 1 ng of pFASTBAC-HMPV_L was used to transform MAX Efficiency™ DH10Bac™ competent cells (ThermoFisher Scientific) via heat shock. 1 mL of SOC (ThermoFisher Scientific) was added to the heat shocked competent cells and allowed to recover for 6 hours at 37 °C with agitation at 225 rpm. The cells were pelleted by centrifugation at 1200 g for 5 min. Supernatant was discarded and cells were resuspended in 100 µL of LB. Resuspended cells were plated onto bacmid selection plates. The colonies were left to grow in a 37 °C incubator for 48 hours. White colonies were streaked onto bacmid selection plates and left to grow in 37 °C incubator for 48 hours. 5 white colonies are selected for screening of HMPV L gene transposition by colony PCR. Colonies with correct transposition of genes was grown in 5 mL of LB at 37 °C for 16 hours with agitation at 225 rpm. Cells were pelleted by centrifugation at 1200 g for 15 min. Pellet was resuspended, lysed and neutralised using P1, P2 and P3 buffer from QIAprep Spin Miniprep Kit (Qiagen). Neutralised cell lysate was pelleted via centrifugation at 12,000 g for 15 min at room temperature. Supernatant was transferred into a clean tube and bacmid was precipitated through isopropanol precipitation. Precipitated bacmid DNA was washed with 70% (v/v) ethanol and eluted in 50 µL of PCR grade H₂O, pre-warmed at 60 °C. Concentration of purified bacmid was measured using a nanodrop (ThermoFisher Scientific).

New HMPV L and P gene (named HMPV_L_GA and HMPV_P_GA) were codon optimized by (Geneart) and chemically synthesized and cloned into modified pFastBacDual (Biobasic). HMPV L_GA was inserted behind a Strep-tag-II

followed by a TEV cleavage sequence. Strep-TEV-HMPV L_GA was inserted downstream of polyhedrin promoter on pFastBacDual. HMPV P_GA was inserted behind a 8x His-tag sequence followed by a TEV cleavage sequence. 8xHis-TEV-HMPV P_GA was inserted downstream of p10 promoter sequence on pFastBacDual. pFastBacDual_HMPV_L/P_GA was mutated by plasmid PCR to generate pFastBacDual_HMPV_L_{D745A}/P_GA, using primers listed in **Table 3**. KOD Xtreme™ Hot Start DNA Polymerase was used for plasmid PCR of pFastBacDual_HMPV_L_{D745A}/P_GA. Mutant plasmid was confirmed by sequencing. 1 ng of plasmid was used to transform EMDH10BacY via heat shock. Generation and isolation of bacmid of HMPV_L/P_GA and HMPV_L_{D745A}/P_GA followed the same protocol as for HMPV_L bacmid DNA.

2.2.1.5 Insect cell culture

Spodoptera frugiperda 9 cells (Novagen), Sf-9, was resuscitated and cultured at 27 °C in T175 flask in SFM growth media (1% (v/v) fetal bovine serum supplemented Sf 900 III SFM media) (ThermoFisher Scientific). At confluency, cells were scrapped and 2×10^7 cells were transferred into a 125 mL flask with 20 mL of pre-warmed SFM growth media for adaption to suspension growth. The cells were cultured in a suspension at 27 °C with shaking at 150 rpm. Cells were maintained at 0.4×10^6 cells/mL and passaged at 6×10^6 cells/mL. Cells were discarded at passage 50 and new cells resuscitated for protein expression.

2.2.1.6 Transfection of insect cells

1×10^6 cells was seeded into a 6 well plate at 2 mL per well and incubated at 27 °C for 1 hour to allow attachment of cells to the bottom of the well. Cells were washed with Grace's Insect Medium (ThermoFisher Scientific) twice to remove unattached cells. 800 µL of Grace's Insect Medium was overlaid onto cell for

transfection. 8 μL of Cellfectin II reagent (ThermoFisher Scientific) was added 100 μL of Grace's Insect Medium and incubated for 10 min at room temperature. Concurrently, 6 μg of isolated bacmid DNA was added to 100 μL of Grace's Insect Medium was added to and incubated for 10 min at room temperature. 108 μL of Cellfectin mixture was added to bacmid mixture and allowed to incubate for another 15 min at room temperature. The transfection mix was added drop by drop onto washed *Sf*-9 cells in 800 μL overlay. The transfected cells were incubated with transfection mix for 8 hours at 27 °C to allow transfection of bacmid into cells. Cells were washed twice with SFM growth media and overlaid with 3 mL of SFM growth media. Transfected cells were incubated for 5 days to allow packaging of baculovirus, P_1 .

2.2.1.7 Amplification of Baculovirus

To generate more P_2 and P_3 baculovirus, cells were infected at 2.5×10^6 cells/mL at a MOI of 0.1. Estimated viral titre at 5×10^6 and 5×10^7 pfu/mL were used for generation of P_2 and P_3 baculovirus respectively. Virus was harvested in supernatant day 3 post infection by centrifugation at 1000 g for 10 min at 4°C.

2.2.1.8 Titration of EMBacY

0.4×10^5 cells/well were seeded into 96 wells plate in 200 μ L of SFM growth media. EmBacY backboneed baculovirus was titrated by infecting cells with a serial dilution of P₃ baculovirus, from 10^{-2} to 10^{-9} . Infected cells were incubated at 27 °C. Diluted wells were scored for presence of YFP day 3 post infection. The titre of baculovirus was quantified based on TCID₅₀ calculation from scoring of wells.

2.2.1.9 Insect cell protein expression

Sf-9 cells were infected at a MOI of 1 using pfu/ml for EMBacY or with an estimated titre of 5×10^7 pfu/mL for HMPV L. The infected cells were incubated at 27 °C and harvested 72 hours post infection. Infected insect cells were harvested by centrifugation at 1000 g, for 15 min at 4°C.

2.2.1.10 Purification of HMPV L and P protein complex

Sf-9 cells pellet was resuspended in 1/10 of cultured volume in cold HMPV L/P lysis buffer. The resuspended cells were lysed by sonication with the same parameters as described for *E. coli* HMPV P cell lysis. Cell debris were pelleted by centrifugation at 20,000 g, for 30 min at 4 °C. Supernatant was incubated with washed/pre-equilibrated 3 mL of nickel-NTA beads per litre of cell pellet for 2 hours at 4 °C. The beads were collected by centrifugation at 1000 g for 10 min at 4 °C. The pelleted beads were subjected to washes at 10 mM, 25 mM and 50 mM imidazole with 40 BV, 30 BV and 20 BV respectively. Washed beads were transferred into gravity flow column for elution. HMPV L and P protein complex is eluted in HMPV L/P buffer B through gravity flow at 4 °C. Eluted fractions were analyzed using SDS PAGE stained with InstantBlue (Expedeon) to visualise presence of HMPV L and P protein. Eluted fraction containing HMPV L and P

protein complex were pooled and subjected to proteolytic cleavage by TEV protease. Pooled proteins were dialyzed against HMPV L/P buffer A to 15 mM of imidazole at 4 °C, overnight along with the cleavage of protein. Cleaved HMPV L and P was incubated with 1 mL nickel-NTA beads and collected as reverse flow-through. Collected sample was concentrated using Vivaspin15R 100 kDa cutoff concentrator (Sartorius) to 100 µL. Concentrated sample was loaded onto HMPV L/P buffer A pre-equilibrated Superose 6 5/150 GL for SEC. Fractions corresponding to the peak observed in chromatogram are separated on SDS PAGE and visualised by instant blue staining. HMPV L concentration was estimated against bovine serum albumin as reference.

2.2.1.11 Western blotting

Protein separated by electrophoresis was transferred onto PVDF membrane (Bio-Rad) at 30 V, at 4 °C and for 16 hours. Membrane was blocked in 5% BSA in 1x TBST (TBS supplemented with 0.1% Tween20) at room temperature for 1 hour, Blocked membrane was incubated with anti-His antibody conjugated to HRP (Sigma) or with rabbit anti-strep-tag II antibody (Abcam) for 1 hour at room temperature. Membrane was washed thrice with 1x TBST at 10 min for each wash. Rabbit anti-strep-tag II antibody was probed with anti-rabbit goat antibody conjugated with HRP (Santa Cruz) and incubated at room temperature for 1 hour. Membrane was washed with TBST thrice at 10 min for each wash. Clarity Western™ ECL Blotting reagent A and B were mixed in a 1: 1 ratio and sprayed onto the probed membrane. Chemiluminescence detection of band was performed on ChemiDoc (Bio-Rad).

2.2.2 In vitro RNA synthesis

RNA oligonucleotides representing nts 1-25 of the Le sequence were purchased (Sigma). The RNA oligonucleotides were mixed with HMPV L/P (containing 200 ng of L protein) in standard transcription buffer containing 20 mM Na HEPES, 50 mM NaCl, 5 mM MgCl₂, 1 mM DTT, 10 U Rnasin (Promega), NTPs, 0.4 μM of Le RNA template and 1 μL of [α-³²P] UTP or GTP (as indicated in figure legend), in a volume of 50 μL. Different experiments were carried out with varying concentrations of NTP, MgCl₂ and substitution of MnCl₂ as indicated in the figure legend presented in the result section. Reactions were incubated at 30 °C for 3 hours and heated to 90 °C for 3 minutes to inactivate HMPV L/P and cooled on ice for 5 min. RNA was isolated by phenol chloroform extraction and followed by ethanol precipitation. Precipitated RNA was resuspended in 20 μL of 2x BBXC loading dye and heated at 90 °C for 5 min. RNA was separated by electrophoresis on a 20% polyacrylamide gel containing 7 M urea in tris-borate-EDTA buffer and analysed by phosphorimaging (Typhoon Trio, GE Healthcare). Radiolabeled RNA decade marker (Ambion) was generated by incubating [Y-³²P] CTP, 3000 Ci/mmol at 10 mCi/mL, according to manufacturer's protocol. 0.2 mM of Le RNA template was labeled using reagent from RNA decade marker (Ambion), according to manufacturer's protocol. Labeled Le RNA template was then serially diluted to 10⁻³. 4 μL of RNA and labeled markers were loaded into each well for electrophoresis for each RNA transcription assay. The bottom of the gel was cropped to remove any nonspecific signal from unincorporated radiolabelled NTP that was not effectively removed during purification or electrophoresis.

2.2.3 Structural analysis

2.2.3.1 Negative staining and image analysis of electron microscopy analysis

Protein from a single fraction of SEC was obtained and diluted to 1 µg/ml. Copper coated carbon EM grids (200 mesh) were first glow-discharged for 60 sec (PDC-32G, HARRICK). 4 µL of diluted protein sample was applied onto the glow-discharged carbon surface for 1 min at room temperature and dry blotted with whatman paper. Following blotting, 4 µL of freshly prepared 2% (wt/v) uranyl acetate solution was applied onto the grid for 1 min. Uranyl acetate was blotted off the grid and the grid allowed to air dry for 10 min before data collection or for grid storage. Electron microscopy was performed using a T12 cryo-electron microscope (FEI) operating at 120 keV equipped with FEI Eagle 4k x 4K CCD camera. Image were collected at a defocus of -1.2 µm to -1.6 µm and at a dose of 25.6 e⁻/Å², at magnification of 49 000 (2.11 Å/pixel).

2.2.3.2 Data processing

1691 particles were manually picked 30 electron micrographs using Eman 2.0 Boxer tool, box size 120x120, on Eman 2.0 (Ludtke, Baldwin, & Chiu, 1999) at 2x binning. Picked particles were exported to Relion 2.0 for 2D classification and averaging (Scheres, 2012). Exported particles were sorted into 20 classes at 4.22 Å/pixel, with parameter of 150 Å for particle diameter.

Chapter 3 Results

3.1 Purification of HMPV L and P proteins

3.1.1 Purification of HMPV P

Baculovirus-infected insect cells expressed HMPV L protein, however, purification of HMPV L protein by nickel-NTA affinity chromatography was unsuccessful. NNS P proteins mediate transcription and bind to NNS L protein during replication and transcription. Therefore a strategy of purifying HMPV P protein expressed in *E. coli* was developed to purify the HMPV L protein.

NTD-histidine tagged HMPV P was expressed in *E. coli* cells and purified successfully by nickel-NTA affinity purification (Fig 4A). Due to the weak binding affinity to the nickel-NTA beads, beads were washed with 10 and 25 mM imidazole before eluting protein in HMPV elution buffer. Eluted HMPV P protein from Nickel-NTA purification was diluted in 20 mM Na HEPES (pH 7.5), 50 mM NaCl, 1 mM DTT and 5% glycerol and injected onto a Q column for ion exchange chromatography (Fig 4B and 4C). A gradient of increasing NaCl concentration was performed to remove impurities from the injected sample. A single symmetrical peak was detected at 260 mM NaCl concentration (Fig 4B).

Fractions corresponding to the single peak on IEX chromatogram were resolved on a SDS PAGE. SDS PAGE was stained with coomassie blue. The stained gel revealed bands migrating slightly below 48 kDa in denaturing conditions. Fraction from the IEX chromatography showed an improved protein purity when compared to injected sample for IEX. HMPV P protein has been reported to form tetramers in solution. Size exclusion chromatography was performed to further purify HMPV P protein and as a buffer exchange step for HMPV P protein for HMPV L buffer

A. A single symmetrical peak (Fig 4D) was observed in SEC, indicating that the separated protein is a homogenous population. Fractions corresponding to the peak were subjected to SDS PAGE. Coomassie blue stained gel showed bands of molecular weight smaller than HMPV P protein under denaturing conditions. All bands were excised and subjected to LC/MS analysis. LC/MS analysis confirmed all bands to belong to HMPV P (Fig 4E).

Figure 4 Purification of HMPV P protein expressed in *E. coli*

Figure 4A: Coomassie blue stained SDS PAGE gel of elutions from Nickel-NTA affinity chromatography purification of *E. coli* expressed HMPV P protein. Protein was eluted in HMPV P elution buffer at 1x bead volume per elution. 4 elutions at 1x BV/elution was carried out to ensure complete all HMPV P were eluted from nickel-NTA beads.

Figure 4B: Ion Exchange chromatogram of Nickel-NTA purified HMPV P. UV260 detected a peak at 260 mM NaCl outlined by red box. Eluted fractions corresponds to purified HMPV P as seen in Fig 1C.

Figure 4C: Coomassie blue stained SDS PAGE gels of fraction corresponding to peak from ion-exchange chromatography. Lane labeled with Inj represents injected sample for IEX-chromatography. Eluted fractions from peak in Fig 4B is marked by red box.

Figure 4D: Size exclusion chromatogram of HMPV P. HMPV P was injected onto HiLoad 16/60 Superdex S200. A single peak symmetrical peak was detected via absorbance at UV260. Peak indicates purified HMPV P from *E.coli* is a tetramer.

Figure 4E: Coomassie blue stained SDS PAGE gel of eluted fractions from single peak in SEC. Inj denotes injected samples for SEC. HMPV P is indicated by the black arrow. Fine protein bands outlined by red box indicates HMPV P is highly unstable and prone to degradation.

3.1.2 Purification of HMPV L using HMPV P protein expressed by *E. coli*

Purified HMPV P protein expressed in *E. coli* was added to clarified lysate from baculovirus infected Sf-9 cells expressing HMPV L protein. The clarified lysate and HMPV P protein were incubated with nickel-NTA beads in an effort to co-purify HMPV L protein through its interaction with the HMPV P protein. HMPV P protein bound to the nickel beads was successfully eluted using the HMPV L elution buffer (Fig 5Bi, lane 7 and 8). Bands with molecular weight of HMPV L protein were detected in the soluble fraction and flow through upon 5 min of chemiluminescence exposure (Fig 5Bii, lane 2 and 4). This result indicates that HMPV L protein is indeed expressed in insect cells albeit in low amount. However, no HMPV L band was detected in the eluted fractions in western blot either during both short and long chemiluminescence exposure, indicating that HMPV L protein was not copurified in presence of *E. coli* expressed HMPV P protein. Therefore, a new strategy was needed to purify HMPV L.

A

		Soluble fraction		FT			Elution (concentrated)		
	(Control) Purified HMPV P					Protein Standards			Control (High MW)
Infected		+	-	+	-		+	-	
Uninfected		-	+	-	+		-	+	
Ln	1	2	3	4	5	6	7	8	9

Figure 5 Copurification of HMPV L using HMPV P purified from *E. coli*

Figure 5A: Table of contents describing samples in respective lanes in Figure 5B and 5C.

Figure 5B: Western blot detection of HMPV L and HMPV P protein using HRP conjugated anti histidine antibody after 1 minute exposure (Fig 5Bi) and after 5 minutes exposure (Fig 5Bii). HMPV L is expressed in low amount by baculovirus infected insect cells as outlined in the red box. Purified HMPV P protein from *E. coli* is unable to pull-down HMPV L expressed by insect cells.

3.1.3 Coexpression and purification of HMPV L and P protein from insect cells

A new codon-optimised gene was synthesized and cloned into pFastBacDual. pFastBacDual_HMPVL/P_GA was transformed into EMDh10BacY competent cells and bacmid was purified for transfection of insect cells. EMDh10BacY is genetically modified to encompass a YFP in the viral genome that will produce YFP following baculovirus infection. The use of EmDh10BacY aided in monitoring viral infection in insect cells. Under 488nm (Excitation) and 509 (Emission), YFP is visualised as GFP. YFP signal was observed on day 3-post transfection, indicating successful transfection. YFP signal was able to spread to adjacent cells from day 3 to day 5-post transfection indicating successful baculovirus packaging and infection of adjacent cells (Fig 6A). Supernatant was harvested and virus was amplified until passage 3 before infection for protein expression. Infected insect cells were harvested 72 hours post infection and lysed. Despite being tagged with a N terminal Strep-tag-II, purification of HMPV L with streptavidin beads was unsuccessful; confirming that the N-terminal tags on HMPV L protein is not exposed to the external environment. Nickel-NTA affinity chromatography was used to purify HMPV P protein. Elution of nickel-NTA affinity chromatography showed bands migrating near 37 kDa and 245 kDa markers (Fig 6B) on a 4-20% PAGE. Western blot analysis showed successful expression and purification of HMPV P when probed with an anti-histidine antibody (Fig 6Cii). More importantly, a band at 245 kDa was observed when probed with an anti-streptavidin antibody, indicating successful co-purification of HMPV L protein (Fig 6Ci). Although some HMPV L and P proteins were detected in the insoluble fraction, much of the remaining HMPV L and P proteins could be purified as a

complex in a soluble form. Decreased HMVP L signal was observed on western blot when eluted protein was heated. This indicates that HMPV L is unstable at high temperature even when boiled in SDS loading buffer as seen in elution lanes figure 6Ci.

Figure 6 Co-expression and purification of HMPV L and P from insect cells

Figure 6A: Photo of transfected insect cells with purified bacmid from EMDH10BacY day 5 post infection. Presence of YFP indicated successful transfection and production of baculovirus encoding expression vector of HMPV L and HMPV P genes.

Figure 6B: Coomassie blue stained SDS PAGE for purification of HMPV L and P protein from infected cell pellets. Eluted fractions showed presence of two bands near the 245 kDa and 37 kDa markers suggesting successful purification of HMPV L and P protein complex. (P- Pellet, T- Total cell lysate, S- Soluble supernatant, W- Imidazole wash, E- Elution)

Figure 6C: Western blot detection with Anti Strep-tag-II (Fig 6i) and Anti Histidine (Fig 6ii). Samples were separated via electrophoresis on a 4 – 20 % gradient gel.

Fig 6i: Western blot detection of Strep-tag-II with rabbit anti Strep-tag-II antibody and goat anti-rabbit HRP antibody on HMPV L purified from baculovirus infected insect cells. Detection of Strep-tag-II protein at 245 kDa suggests recombinant HMPV L is expressed and can be purified by co-purification of HMPV P protein. Eluted HMPV L is thermolabile as observed by the decrease in band intensity after boiling in 5x SDS buffer.

Fig 6Cii: Western blot detection of 6x his tag with anti-histidine antibody conjugated with HRP. Detection of bands at 37 kDa with HRP conjugated anti-histidine antibody suggests HMPV P can be expressed by insect cells expression system and purified by nickel-NTA.

Eluted fractions from nickel-NTA affinity chromatography was pooled and cleaved by TEV protease. A reverse his-trap was performed to remove contaminants and the flow-through was collected, concentrated and injected onto a SEC column. Size exclusion chromatography showed 3 peaks eluting respectively at 1.06 mL, 1.67 mL and 2.09 mL on a 3 mL column (Fig 7A). Proteins from eluted fractions were separated by SDS PAGE and visualized by InstantBlue staining. Protein bands were observed at 245 kDa and 37 kDa in elution at 1.06 mL along with contaminants as indicated by an arrow on Figure 4A and 4B. This indicated that HMPV L and P protein might form soluble aggregate with other soluble cellular contaminants as seen in the early elution volume. A symmetrical peak was observed at 1.67 mL. Stained gel showed bands migrating at 245 kDa and 37 kDa that correspond to the molecular weight of HMPV L and P protein respectively under denaturing conditions along with an additional band migrating at 75 kDa. A calibration curve was used to estimate the molecular mass of the protein complex eluting at 1.67 returned a value of 644 kDa. Bands at 250 kDa, 75 kDa and 37 kDa (highlighted by red circle) in Figure 7B was excised and subjected to LC/MS. LC/MS analysis confirmed protein identity of the 250 kDa band as HMPV L protein and the band at 37 kDa as the HMPV P protein. Protein band at 75 kDa was identified as Hsp70 protein (Fig 7C) with confirmation by LC/MS analysis (Supp data 1). The estimated size from SEC indicate that purified protein may be a complex of dimeric L protein along with tetrameric P protein and a single Hsp70 protein.

Figure 7 Purification of HMPV L and P and confirmation of their identity.

Figure 7A: Analytical SEC of purified HMPV L and P complex. Purified protein was injected onto a Superose 6 5/150 Increase column and eluted as a single

symmetrical peak outlined by a red box. Red arrow indicates the presence of a larger protein complex. Right inset: Calibration of the analytical column.

Ribonuclease A (13.7 kDa), aldolase (158 kDa), ferritin (440 kDa), thyroglobulin (669 kDa) were used as reference proteins. The protein is estimated to elute at 644 kDa.

Figure 7B: Visualization of Instant Blue stained SDS PAGE of eluted fractions from SEC. Lane marked by a red arrow corresponds to fraction eluted at V_0 of Superose 6 5/150 Increase column (indicated by red arrow in Fig 7A). Some HMPV L and P complex elutes with other impurity as soluble aggregates at V_0 . Fractions corresponding to eluted fractions outlined by red box in Fig 7A was visualized on SDS PAGE stained by Instant Blue staining. Protein band near 250 kDa and 37 kDa suggest formation of protein complex that cannot be separated by size exclusion chromatography. Protein outlined by red circle was subjected to LC/MS analysis. Inj lane represents sample that was loaded onto the Superose 6 5/150 Increase column.

Figure 7C: Confirmatory results from LC/MS. Report from LC/MS analysis indicated purified samples as HMPV L and HMPV P protein.

3.2 Biochemical characterization of HMPV L transcription

3.2.1 Isolated recombinant HMPV L is capable of *de novo* RNA synthesis

HMPV RNA replication occurs when the L protein recognises the 3' le sequence specific to HMPV. The first 25 nt of HMPV genomic 3' RNA was chemically synthesised for the RdRp transcription assay. A functional HMPV L polymerase will replicate and yield a RNA product of 25 nt in length (Fig 8A). GDNQ motif has been reported to be vital for RNA transcription in the NNS RdRp domain (Fig 3). Therefore, a substitution in HMPV L targeting the GDNQ motif, named L_{D745A}, was introduced and purified (Fig 8B). Isolated polymerase was incubated with 0.25 mM of NTP, $\alpha^{32}\text{P}$ -GTP or $\alpha^{32}\text{P}$ -UTP, Le RNA template and Mg^{2+} at 30 °C for 3 hours. Radiolabeled RNA transcripts were separated on a 7 M Urea 20% PAGE and RNA was detected via autoradiography (Fig 8C). No RNA transcripts were detected in reaction without the polymerase indicating that the L and P protein complex forms the basic active unit of HMPV polymerase. RNA transcripts were detected in reactions incubated with NTP, polymerase and Mg^{2+} . No RNA product was detected in reactions with L_{D745A} indicating that substitution of aspartate residue 745 to alanine abolishes of HMPV polymerase activity. No RNA products were observed in reactions lacking Mg^{2+} , suggesting that the Mg^{2+} divalent ion is essential for RNA replication. A comparison of RNA products between $[\alpha^{32}\text{P}]$ -GTP and $[\alpha^{32}\text{P}]$ -UTP labeling showed more RNA was synthesized in presence of $[\alpha^{32}\text{P}]$ -UTP instead of $[\alpha^{32}\text{P}]$ -GTP. Replication of RNA by HMPV polymerase also yields products that were larger than 25 nt, indicating HMPV polymerase possess additional processivity activity that may add nucleotides to the RNA products through back priming.

A

Template 3' UGCGCUUUUU UUGCGCAUUA UAAAU 5'

Product 5' ACGCGAAAAA AACGCGUAUA AAUUA 3'

B

C

Figure 8 *De novo* RNA synthesis by isolated HMPV polymerase

Figure 8A: Sequence of the 25 nt Le RNA used in the *in vitro* assay and the expected complementary antigenomic product.

Figure 8B: InstantBlue stained gel showing isolated wt HMPV L and P complex and mutant L_{D745A} and P complex. Band migrating near 245 kDa corresponds to HMPV L and band migrating slightly above 37 kDa corresponds to HMPV P protein.

Figure 8C: RNA product synthesized by purified HMPV L RdRp and comparison between $\alpha^{32}\text{P}$ -GTP and $\alpha^{32}\text{P}$ -UTP. *De novo* RNA transcription was carried out in a standard reaction of 20 mM Na HEPES pH 7.4, 60 mM NaCl, 1 mM DTT, 5 mM MgCl, 10 U RNAase inhibitor, 200 ng of L protein, 0.4 μg of template RNA, 0.2 μM of NTP and 0.025 μM of radiolabelled NTP (10 mCi/ml). Lane 1 shows the gamma labelled 25 nt template RNA used in the RNA synthesis assay and lane 2 shows a molecular weight ladder. In lane 4, HMPV L with D745A substitution was used instead of wt HMPV L. Mg^{2+} (Lane 6) and HMPV L (Lane 7) was omitted from the reaction as negative controls for the transcription reaction. The labeled products were separated by electrophoresis on 20% polyacrylamide 7 M Urea PAGE and visualized by autoradiography. Lane 1 shows a 25 nt positive control and lane 2 shows the molecular weight ladder.

3.2.2 Effect of NTP concentration on RNA synthesis

NTP is the essential building block needed for RNA transcription. To understand the replication profile of HMPV polymerase, *in vitro* RNA synthesis was performed at various NTP concentrations. Concentrations of NTP used in this experiment ranged from 0.25 mM to 2 mM. No RNA transcription was detected in a reaction mixture incubated with HMPV L_{D745A}. Similarly, no RNA transcription was detected in a reaction mixture with 2 mM of NTP. Increase level of RNA products were observed with increased concentrations of NTP, up to 1 mM (Fig 9A). Data normalised to reaction with 0.25 mM NTP concentration showed 2 fold increase in RNA synthesis in reaction with 1 mM NTP concentration, while reaction with 0.5 mM NTP concentration showed a 0.5 fold increase (Fig 9B). More products were generated with 1 mM NTP concentration as compared to products generated using 0.25 mM and 0.5 mM NTP concentration.

Figure 9 Effects of NTP concentration of HMPV polymerase *de novo* RNA synthesis.

Figure 9A: RNA product synthesized by purified HMPV L RdRp. RNA products were synthesized in 20 mM Na HEPES pH 7.4, 60 mM NaCl, 1 mM DTT, 5 mM MgCl, 10 U RNasin, 0.4 μ g of template RNA, 0.025 μ M of [α^{32} P]-UTP (10 mCi/ml) and different concentration of non-radiolabeled NTP, ranging from 0.25 mM to 2 mM. The labeled products were separated by electrophoresis on 20% polyacrylamide 7 M Urea PAGE and visualized by autoradiography. Lane 1 shows a gamma-labeled 25 nt template RNA and lane 2 shows the molecular weight ladder.

Figure 9B: Bar chart showing synthesis of RNA product at varying concentrations of NTP used. The data show the mean and range of 2 different sets of

independent experiments. The data from each experiment were normalized to RNA synthesis at 1 mM NTP concentration, which was set arbitrarily to a value of 1. Bar represents mean and error bars represent range from 2 independent sets of experiments. ANNOVA Statistical analysis was conducted on PRISM Graphpad software, (** $P \leq 0.01$, * $P \leq 0.05$)

3.2.3 Effects of divalent cations on RNA transcription of HMPV L

Divalent metal cations have been reported to be essential for polymerization activity as it coordinates incoming NTPs and allow phosphoryl transfer through chelation with Asp residues in the RdRp active site (Yang, Lee, & Nowotny, 2006). RNA replication was performed in conditions lacking cations and increasing concentration of cations, from a range of 2.5 mM to 15 mM. No RNA product was observed in reaction incubated with L_{D745A} or in reaction performed in absence of Mg²⁺ (Fig 10A). This indicated that Mg²⁺ is crucial for *de novo* RNA synthesis. Reaction with 5 mM Mg²⁺ has the highest yield in terms of RNA products, compared to reactions performed with various concentration of Mg²⁺. Lowest RNA product was observed from reaction performed at 2.5 mM Mg²⁺. Reactions performed with Mg²⁺ at concentration of 2.5 mM, 7.5 mM, 10 mM and 15 mM showed reduced transcription when compared to reaction performed in 5 mM Mg²⁺ (Fig 10C). This indicated that optimal RNA replication occurs in presence of 5 mM Mg²⁺.

Similarly, no RNA product was observed in reaction with L_{D745A} in presence of Mn²⁺. No RNA product was detected in reaction performed in absence of Mn²⁺. This indicates that cation is essential for HMPV polymerase activity. Faint RNA products were observed in experiments performed in 7.5 mM, 10 mM and 15 mM of Mn²⁺. However, densitometry analysis was unable to detect the bands due to the low contrast between the signal and background. Contrast of Mn²⁺ radiography was adjusted, as the signal from RNA product at 2.5 mM Mn²⁺ was too intense and masks the signal from the markers and positive control. Labeled RNA transcripts were detected in reaction performed in 2.5 mM Mn²⁺ (Fig 10B), suggesting manganese can be used as an alternative divalent cation during RNA

transcription in HMPV L, as anticipated. Same amount of RNA markers and PC was used for experiment with Mg^{2+} and Mn^{2+} . The result showed that reaction with 2.5 mM Mn^{2+} was able to generate more RNA product than reaction with 5 mM Mg^{2+} , suggesting a more active polymerase in the presence of Mn^{2+} .

Figure 10 Effects of cations on HMPV polymerase activity

Figure 10A: Effects of varying concentration of Mg²⁺ on HMPV L *de novo* RNA synthesis.

Figure 10B: Effects of varying concentration of Mn²⁺ on HMPV L *de novo* RNA synthesis. RNA products were synthesized in standard reaction condition with varying concentration of Mg²⁺ or a substitution of Mn²⁺ at concentrations, ranging from 0 mM to 15 mM. mt HMPV L was carried out in presence of 5 mM of Mg²⁺ or

Mn²⁺. The labeled products were separated by electrophoresis on 20% polyacrylamide 7 M Urea PAGE and visualized by autoradiography. Lane 1 shows a gamma labeled 25nt template RNA and lane 2 shows the molecular weight ladder. Lane 1 shows the molecular weight ladder and lane 9 shows gamma-labeled 25 nt template RNA as a positive control. Lane 8 is a negative control in which substrates were incubated with L protein with D745A substitution at the catalytic site of the RdRp domain of the L protein.

Figure 10C: Bar chart showing RNA product synthesised at varying concentrations of Mg²⁺. The data show the mean and range of 2 different repeats of independent experiments. The data from each experiment were normalized to RNA synthesis at 5 mM Mg²⁺ concentration, which was set arbitrarily to a value of 1. Bar represents mean and error bar represents range from 2 independent repeats of experiments. ANNOVA Statistical analysis was conducted on PRISM Graphpad software.

3.3 Structural study of HMPV L

3.3.1 Predictive modeling of HMPV L protein

HMPV L shares similar function to VSV L and the sequence identity is 22%. A sequence alignment between the L protein from HMPV and VSV is shown in Supp data 4. Hence a three-dimensional model of the HMPV L was generated using RaptorX (Kallberg et al., 2012). The predicted HMPV L model (Fig 11A) showed a less compact protein configuration when compared to VSV L (Fig 11B), due to the presence of a long and flexible linker that joins the connector domain to the methyltransferase domain. Similar to VSV L, the RdRp domain is predicted to adopt a classical right-handed configuration. The predicted model showed a “doughnut ring” like structure, made up of the RdRp and capping domains. The predicted HMPV L shows the connector domain is positioned at the bottom right of the molecule near the RdRp, causing the shape of the doughnut ring to be oval instead of circular (Fig 11A). The methyltransferase and CTD domain of HMPV L is connected to the oval-shaped “doughnut ring” structure by a flexible linker. Structural regions of predicted methyltransferase and CTD domain model were superimposed onto the published X-ray crystallographic structure (Paesen et al., 2015), showing a perfect fit (Supp data 2). However, the flexible linker connecting connector domain to methyltransferase domain in the published structure are of different length as compared to the predicted. It is unclear if the different length of the loop may affect the methyltransferase activity of HMPV L.

Conserved residues (GDNQ) involved in RdRp enzymatic activity and capping of newly synthesized RNA (HR) observed at the core of the doughnut-like structure, are represented by sphere in magenta and blue respectively. This suggests that catalytic activity occurs at the core of the doughnut and RNA is threaded,

elongated and capped within the core of the doughnut-ring. Although the GxxT and HR motif are 71 amino acids apart, the predicted model showed both motif separated by a 13.4 Å distance (Fig 11C). The presence of poorly ordered loops and the close proximity of these two motifs (HR/GxxT) suggests a conformational change in HMPV capping domain as newly synthesized RNA is capped via a PRNTase.

Figure 11 Predicted structure of the HMPV L protein

Figure 11A: A homology-based three dimensional structure of the HMPV L protein highlighting the RdRp domain, capping domain, connector domain, methyltransferase domain and C-terminal domain. Orientation: front (8Ai), back (8Aii), top (8Aiii) and bottom (8Aiv). The RdRp domain is in cyan; capping domain is in green; connector domain; yellow; methyltransferase domain is in orange and C-terminal domain is in red. Predicted model shows HMPV L connector domain to be linked to methyltransferase via a flexible loop. Conserved motif for RdRp (GDNQ) is shown in magenta sphere and conserved motif of capping domain (GxxT and HR) is shown by blue sphere

Figure 11B: Published VSV L protein structure (Liang et al., 2015) PDB code: 5A22

Figure 11C: Magnified view of the conserved motif of the capping domain. Important residues of the capping domain are outlined in green.

3.3.2 Molecular architecture of the HMPV L and P protein complex

Negatively stained HMPV L and P protein complex revealed mono-dispersed particles when viewed under the electron microscope (Fig 12A). The images revealed small, 10-20 nm globular particles and larger proteins particles (indicated by white arrows), likely to be soluble aggregates. Some particles (circled in white) have an appearance of a “doughnut-ring” with an appendage that resembles the predicted HMPV L model.

Smaller particles were selected for single particle analysis. 2D class averaging showed a prominent 10-11 nm oval ring and particles with appendage attached to the oval ring to be at 15 nm (Fig 12B). Data observed from the 2D classification matched the shape of the predicted HMPV L. Unlike VSV L, only 1 appendage attached to the isolated HMPV L was observed. This indicates that there are structural differences between HMPV L and VSV L.

A

B

Figure 12 Single particle analysis of HMPV L

Figure 12A: Representative negative stained image of HMPV L and P protein complex. Scale bar at the bottom:100 nm.

Figure 12B: Class averages of HMPV L and P complex obtained by classification of 1691 particles. The side length of the individual panels is 55 nm.

Chapter 4 Discussion

4.1 Isolation of HMPV L

Thus far, the only recombinant full-length L protein of a NNS virus that have been successfully purified are those of VSV, Chandipura, Sendai virus and RSV (Noton, Aljabr, Hiscox, Matthews, & Fearn, 2014; Noton et al., 2012; Ogino & Banerjee, 2010; Ogino et al., 2005; Song et al., 2015; Tremaglio et al., 2013). In this study, we report the first successful expression, purification and biochemical characterization of a full-length active recombinant HMPV RNA polymerase. HMPV L and P protein were coexpressed in insect cells and purified through nickel affinity chromatography by binding to 8 x histidine tag at the N-terminal end of HMPV P protein.

Although N-terminal his-tagged VSV L protein can be purified by nickel affinity chromatography (Song et al., 2015), purification of HMPV L protein alone via nickel affinity chromatography was not successful despite confirmation of its presence in the milieu by western blot. Similar to our RSV L protein, N-terminal tagged HMPV L protein can only be purified as a complex with P protein through affinity (Noton et al., 2012). Detection of HMPV L when probed with anti-histidine antibody rules out the possibility of N-terminal degradation that may result in failure of purification. Failed nickel purification of HMPV L protein and positive detection of HMPV L on western blot suggests that affinity tag at the N-terminal region of HMPV L protein is not exposed in solution. More recently, an insertion of affinity tag between amino acid residue 1738 and 1739 in RSV L showed that it is possible to purify RSV L protein without the presence of RSV P protein (Sourimant et al., 2015) but the insertion of the tag at this position affected the expression of L protein. Based on sequence alignment, the position for tag

insertion places the insertion site on the flexible linker that connects the connector domain to the methyltransferase domain (Fig 11). Although RSV and HMPV L protein share high sequence homology of 62%, it is unclear if insertion of affinity tag at predicted linker region would affect the function or expression of HMPV L protein.

When expressed alone, HMPV L protein was only detected after 5 minutes of chemiluminescence exposure on western blot. This indicates that there is only pico-mole quantity of HMPV L protein, when expressed alone. Only 36 µg of HMPV L protein is obtained at the end of purification from 1 L of infected insect cell culture. This suggests that HMPV L is expressed at low concentration in insect cells and the expression is dramatically improved when coexpressed with the P protein.

HMPV P binds to HMPV N protein (Renner et al., 2016) to form inclusion bodies in the host cell cytoplasm. Although both HMPV N and P proteins do not contain enzymatic activity essential for RNA transcription, newly synthesized viral RNA was also detected as inclusion bodies at the early stage of infection (Cifuentes-Munoz, Branttie, Slaughter, & Dutch, 2017). This suggests that HMPV L protein is present as inclusion bodies for transcription of RNA. *Ex vivo* experiments have shown RSV as L and P protein of RSV colocalising in the cytoplasm (Sourimant et al., 2015). L proteins of the NNS viruses have been shown to interact with P protein (Ogino & Banerjee, 2010; Song et al., 2015; Sourimant et al., 2015; Tremaglio et al., 2013). Therefore the results presented prove that HMPV L and P protein interacts with each other. Hence, a coexpression and copurification strategy was developed in an effort to purify HMPV L protein. HMPV L protein

was copurified in the purification of HMPV P protein from baculovirus infected cells.

HMPV P protein forms a tetramer by binding other P proteins through its alpha-helical core domain. The core domain is flanked by intrinsically disordered regions at the N-terminal and C-terminal ends (Leyrat et al., 2013). When expressed alone, the less ordered N-terminal and C-terminal ends are prone to degradation. However, structural studies showed N-terminal of HMPV P protein forming into alpha-helical loop when interacting with HMPV N protein (Renner et al., 2016). Degradation of protein is often due to presence of less structured loop. The result suggests that intrinsically disordered region of HMPV P protein may be stabilized during HMPV L protein interaction, thus decreasing degradation of HMPV P protein.

Hsp70 is a cellular chaperone protein that is involved in protein folding processes (Mayer & Bukau, 2005). It was previously reported that RSV polymerase complex interacts with Hsp70 (Brown et al., 2005) and Hsp70 was co-purified in the purification of RSV L and P protein complex (Noton et al., 2012). The characterization of Hsp70 and its association with HMPV L and P protein complex remains unclear. It is tempting to speculate that Hsp70 is needed as a chaperone for folding of the L protein.

4.2 Biophysical characterisation of HMPV polymerase complex

Analysis of purified HMPV complex indicates a ratio of 4 P proteins per L protein. Despite having a calculated molecular weight 440 kDa, with the addition of Hsp70, size exclusion chromatography estimated the purified protein complex to be 644 kDa. Although L protein has been reported to interact together as dimers

(Cevik, Holmes, et al., 2004; Rahmeh et al., 2010), our data from electron microscopy showed isolated and individual HMPV L protein. Size exclusion chromatography is effective in determining the molecular Stokes radius of globular protein. The early elution at 1.67 mL on SEC suggests HMPV L and P protein complex might not be globular. This is supported by the predicted model of HMPV L in Fig 11A and the data from the 2D classification of single particle analysis. HMPV L and P protein complex is elongated in length as seen by the attachment on the top of oval shaped doughnut ring.

4.3 Characterization of HMPV L catalytic RdRp site and factors affecting RNA replication

Transcription of NNS virus requires the presence of N protein along with an anti-termination factor, M2-1 (Cai et al., 2016). The result shows that HMPV L and P form the basic unit of the polymerase complex needed for *de novo* RNA synthesis.

GDNQ is a conserved motif of the RdRp domain (Poch, Sauvaget, Delarue, & Tordo, 1989) involved in the elongation of RNA NNS virus (Sleat & Banerjee, 1993). Alanine substitution at D745 abolished the RdRp activity of HMPV L protein presumably by altering the metal binding site essential for RNA elongation (Hansen et al., 1997).

Results showed larger molecular weight of *de novo* RNA synthesis products. Viral polymerase is able to cap newly synthesized viral mRNA (Ogino & Banerjee, 2007) but does not cap antigenomic viral RNA. The *de novo* RNA synthesis utilizes 25 nt of the Le sequence, mimicking antigenomic RNA replication. Larger

products suggest that isolated HMPV polymerase complex modifies the RNA through adding nucleotides. No initiation of transcription at +3 site may suggest the lack of secondary start site on HMPV Le sequence. The result presented is different from purified RSV L polymerase enzymatic activity (Tremaglio et al., 2013). Reaction with isolated RSV L and P protein complex and nt 1-14 of the Le sequence showed RNA products ranging from 7 to 12 nt in length and no products bigger than 14 nt. RSV polymerase also showed initiation of transcription at +3 of Le sequence. The RNA products of HMPV polymerase may form hairpin like secondary structure which was suggested by Norton et.al when a larger RNA products was observed due to the back priming effect of RSV L polymerase on the TrC template (Norton et al., 2012).

More RNA products labeled with [$\alpha^{32}\text{P}$]-UTP were detected when compared to products labeled with [$\alpha^{32}\text{P}$]-GTP. There is an equal number of UTP and GTP in nt 1-25 of the Le sequence. Therefore it is unclear if HMPV polymerase has a preference of radiolabelled NTP. However, the lower yield of [$\alpha^{32}\text{P}$]-GTP labeled RNA product may be due to the difference in half-life of radiolabelled NTP as the different radiolabelled NTP were synthesised on different dates. Reaction with [$\alpha^{32}\text{P}$]-UTP and [$\alpha^{32}\text{P}$]-GTP yielded the same results but [$\alpha^{32}\text{P}$]-UTP was cheaper. Hence, the subsequent radiolabeled experiments were carried out using [$\alpha^{32}\text{P}$]-UTP.

NTP is present in the cells at picomolar concentration (Huang, Zhang, & Chen, 2003). Isolated HMPV polymerase is capable of *de novo* RNA synthesis. The amount of RNA products is dependent on the concentration of NTP present, up to 1 mM. However, no RNA products were detected in reactions using 2 mM NTP.

This may be due to low concentration of [$\alpha^{32}\text{P}$]-UTP in presence of high amount of cold NTP that resulted in limited the formation of RNA products.

Mg^{2+} is a cofactor for RNA synthesis in RdRp as it participates in the coordination of nucleic acid substrate during elongation (Yang et al., 2006).

While most RNA synthesis requires Mg^{2+} , there is an increase in reports of Mn^{2+} binding site in RdRp (Butcher, Grimes, Makeyev, Bamford, & Stuart, 2001; Poranen et al., 2008; Wright, Poranen, Bamford, Stuart, & Grimes, 2012). Mg^{2+} and Mn^{2+} have different chemical properties; hence it is not surprising that substitution to Mn^{2+} yielded a different RNA replication profile when compared to data using Mg^{2+} as a cofactor. Despite Mg^{2+} being in abundance intracellularly, binding Mn^{2+} to catalytic site may represent a rare phenomenon. More RNA product is yielded in presence of low concentration Mn^{2+} as compared to Mg^{2+} . Mn^{2+} has been reported to lower the K_m for initiation nucleotide in flavivirus RdRp, resulting in a less stringent but more products being synthesised (Ranjith-Kumar et al., 2002; Selisko et al., 2012). It is tempting to speculate Mn^{2+} is involved in antigenomic replication and initiation of RNA binding due to the lack of discrimination between sequences for transcription. However, this suggests Mn^{2+} needs to be replaced by Mg^{2+} for the transcription of viral mRNA. No experiment was carried in presence of both Mg^{2+} and Mn^{2+} . Therefore, further testing to understand to elucidate the mechanism of divalent ion selection has to be done in order to confirm this hypothesis.

4.4 Structural analysis

The connector domain of the predicted structure HMPV L protein is at a different position as VSV L (Fig 11). This difference maybe due to the folding of protein

and will not affect the activity of HMPV L as there is no known catalytic site in the connector domain. Connector domain has been speculated to be a structural scaffold to support the L protein (Song et al., 2015). GxxT and HR motifs are conserved residue of the PRNTase in NNS virus (Ogino & Banerjee, 2010). The close proximity of both motifs may suggest HMPV L protein caps newly synthesized RNA via PRNTase activity.

2D classification from single particle analysis showed views of HMPV L protein similar to the predicted models and topology that resembles the VSV L protein. A higher resolution data will have to be collected from cryo-electron microscopy to full understand the domain organisation and the differences between HMPV L protein and VSV L protein.

Chapter 5 Future work and conclusion

5.1. Future work

This study shows the isolation of a functional HMPV polymerase complex. While L binding site of P protein of other viruses such as BRSV, RSV and VSV has been identified (Khattar et al., 2001; Song et al., 2015; Sourimant et al., 2015), exact residues involved in HMPV L binding remains unknown (Renner et al., 2017). Currently, there are no available reported assays to test for polymerase inhibition for HMPV. The isolated HMPV polymerase is able to synthesize RNA through *in vitro de novo* RNA synthesis. This is a valuable system that can be modified for screening polymerase inhibitors. Specific nucleotides within the RSV le promoter sequences have shown to be important to the polymerase activity of RSV polymerase (Cowton & Fearn, 2005). These specific sequences control the polymerase processivity. In this study, purified HMPV L/P complex has shown to be able to process RNA through a 25 nt sequence from the le sequence, it may be used as a model to determine specific promoter sequences involved in HMPV transcription and replication, so as to characterize template sequences involved in polymerase function. This will allow a deeper understanding on how the polymerase is able to form switch from transcription to replication as well as to determine the important factors needed in each step. In the above experiments, a product larger than 25 nt is generated by the polymerase. Nucleotidyltransferase activities have been reported in purified RSV polymerase when using short template to initiate transcription (Noton et al., 2012), however, nucleotidyltransferase activity was not reported during *in vivo* transcription (Noton & Fearn, 2011). Further studies can be conducted to determine HMPV polymerase's RNA processivity to determine the differences between HMPV and

RSV polymerase. A minigenome system may provide comparison for the function of recombinant expressed HMPV polymerase *in vitro* and *ex vitro*. This will provide validation support for polymerase activities *in vivo*, as cellular factors can be associated with the polymerase.

A cryo-EM, atomic resolution of HMPV polymerase complex should be obtainable from data collected with a higher-powered cryoEM machine. Although the L protein of NNS shares conserved residues and motifs, it will be interesting to determine the differences between various polymerase complexes and for development of anti-polymerase drug compounds. With the advancement of technology in cryo-EM, understanding of the binding between HMPV L and P protein may be achieved. Hence, providing better structural understanding and relationship between both proteins.

5.2 Conclusion

HMPV is a relatively new virus of NNS family and much of the information known for HMPV has been inferred from other NNS viruses, especially RSV. This report has hinted that there are some differences in RNA synthesis between HMPV and RSV polymerase. More information about HMPV is needed in order to develop strategies to block HMPV replication and transcription. This study demonstrated the isolation of functional full length HMPV polymerase, in a complex of HMPV L and P protein. Although HMPV L protein is present at low concentration, it can be purified through coexpression and copurification along with HMPV P protein in baculovirus infected insect cells. Isolated HMPV L and P protein complex is functional as it is able to synthesize RNA using nt 1-25 from the Le sequence. The RNA synthesis of HMPV polymerase is NTP concentration. As anticipated, the divalent cation and concentration of ion available do affect the amount of RNA synthesized. Negatively stained HMPV L and P protein can be visualised using cryo-electron microscopy. 2D classification and averaging reveals HMPV L and P protein to share similar “doughnut ring” liked structure reported in VSV L and P protein. Both viruses belong to different family, Rhabdovirus for VSV and Metapneumovirus for HMPV, there is difference in the domain organisation of the L protein as seen in the 2D classes.

Chapter 6 References

- Afonso, C. L., Amarasinghe, G. K., Banyai, K., Bao, Y., Basler, C. F., Bavari, S., . . . Kuhn, J. H. (2016). Taxonomy of the order Mononegavirales: update 2016. *Arch Virol*, 161(8), 2351-2360. doi:10.1007/s00705-016-2880-1
- Al-Ayed, M. S., Asaad, A. M., Qureshi, M. A., & Ameen, M. S. (2014). Viral etiology of respiratory infections in children in southwestern Saudi Arabia using multiplex reverse-transcriptase polymerase chain reaction. *Saudi Medical Journal*, 35(11), 1348-1353.
- Al-Sonboli, N., Hart, C. A., Al-Aghbari, N., Al-Ansi, A., Ashoor, O., & Cuevas, L. E. (2006). Human metapneumovirus and respiratory syncytial virus disease in children, Yemen. *Emerg Infect Dis*, 12(9), 1437-1439. doi:10.3201/eid1209.060207
- Barik, S. (1993). The structure of the 5' terminal cap of the respiratory syncytial virus mRNA. *J Gen Virol*, 74 (Pt 3), 485-490. doi:10.1099/0022-1317-74-3-485
- Bermingham, A., & Collins, P. L. (1999). The M2-2 protein of human respiratory syncytial virus is a regulatory factor involved in the balance between RNA replication and transcription. *Proc Natl Acad Sci U S A*, 96(20), 11259-11264.
- Biacchesi, S., Murphy, B. R., Collins, P. L., & Buchholz, U. J. (2007). Frequent frameshift and point mutations in the SH gene of human metapneumovirus passaged in vitro. *J Virol*, 81(11), 6057-6067. doi:10.1128/JVI.00128-07
- Brown, G., Rixon, H. W., Steel, J., McDonald, T. P., Pitt, A. R., Graham, S., & Sugrue, R. J. (2005). Evidence for an association between heat shock protein 70 and the respiratory syncytial virus polymerase complex within lipid-raft membranes during virus infection. *Virology*, 338(1), 69-80. doi:10.1016/j.virol.2005.05.004
- Butcher, S. J., Grimes, J. M., Makeyev, E. V., Bamford, D. H., & Stuart, D. I. (2001). A mechanism for initiating RNA-dependent RNA polymerization. *Nature*, 410(6825), 235-240. doi:10.1038/35065653
- Cai, H., Zhang, Y., Lu, M., Liang, X., Jennings, R., Niewiesk, S., & Li, J. (2016). Phosphorylation of Human Metapneumovirus M2-1 Protein Upregulates Viral Replication and Pathogenesis. *J Virol*, 90(16), 7323-7338. doi:10.1128/JVI.00755-16
- Caracciolo, S., Minini, C., Colombrita, D., Rossi, D., Miglietti, N., Vettore, E., . . . Fiorentini, S. (2008). Human metapneumovirus infection in young children hospitalized with acute respiratory tract disease: virologic and clinical features. *Pediatr Infect Dis J*, 27(5), 406-412. doi:10.1097/INF.0b013e318162a164

- Cevik, B., Holmes, D. E., Vrotsos, E., Feller, J. A., Smallwood, S., & Moyer, S. A. (2004). The phosphoprotein (P) and L binding sites reside in the N-terminus of the L subunit of the measles virus RNA polymerase. *Virology*, 327(2), 297-306. doi:10.1016/j.virol.2004.07.002
- Cevik, B., Kaesberg, J., Smallwood, S., Feller, J. A., & Moyer, S. A. (2004). Mapping the phosphoprotein binding site on Sendai virus NP protein assembled into nucleocapsids. *Virology*, 325(2), 216-224. doi:10.1016/j.virol.2004.05.012
- Chandrika, R., Horikami, S. M., Smallwood, S., & Moyer, S. A. (1995). Mutations in conserved domain I of the Sendai virus L polymerase protein uncouple transcription and replication. *Virology*, 213(2), 352-363. doi:10.1006/viro.1995.0008
- Chang, A., Masante, C., Buchholz, U. J., & Dutch, R. E. (2012). Human metapneumovirus (HMPV) binding and infection are mediated by interactions between the HMPV fusion protein and heparan sulfate. *J Virol*, 86(6), 3230-3243. doi:10.1128/JVI.06706-11
- Chattopadhyay, A., Raha, T., & Shaila, M. S. (2004). Effect of single amino acid mutations in the conserved GDNQ motif of L protein of Rinderpest virus on RNA synthesis in vitro and in vivo. *Virus Res*, 99(2), 139-145.
- Chen, M., Ogino, T., & Banerjee, A. K. (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(24), 13478-13485. doi:10.1128/JVI.01244-07
- Chenik, M., Schnell, M., Conzelmann, K. K., & Blondel, D. (1998). Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. *J Virol*, 72(3), 1925-1930.
- Chung, J. Y., Han, T. H., Kim, B. E., Kim, C. K., Kim, S. W., & Hwang, E. S. (2006). Human metapneumovirus infection in hospitalized children with acute respiratory disease in Korea. *J Korean Med Sci*, 21(5), 838-842. doi:10.3346/jkms.2006.21.5.838
- Cifuentes-Munoz, N., Branttie, J., Slaughter, K. B., & Dutch, R. E. (2017). Human Metapneumovirus Induces Formation of Inclusion Bodies for Efficient Genome Replication and Transcription. *J Virol*, 91(24). doi:10.1128/JVI.01282-17
- Collins, P. L., Fearn, R., & Graham, B. S. (2013). Respiratory syncytial virus: virology, reverse genetics, and pathogenesis of disease. *Curr Top Microbiol Immunol*, 372, 3-38. doi:10.1007/978-3-642-38919-1_1
- Collins, P. L., Hill, M. G., 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(1), 81-85.

- Corti, D., Bianchi, S., Vanzetta, F., Minola, A., Perez, L., Agatic, G., . . . Lanzavecchia, A. (2013). Cross-neutralization of four paramyxoviruses by a human monoclonal antibody. *Nature*, 501(7467), 439-443. doi:10.1038/nature12442
- Cowton, V. M., & Fearn, R. (2005). Evidence that the respiratory syncytial virus polymerase is recruited to nucleotides 1 to 11 at the 3' end of the nucleocapsid and can scan to access internal signals. *J Virol*, 79(17), 11311-11322. doi:10.1128/JVI.79.17.11311-11322.2005
- Cowton, V. M., McGivern, D. R., & Fearn, R. (2006). Unravelling the complexities of respiratory syncytial virus RNA synthesis. *J Gen Virol*, 87(Pt 7), 1805-1821. doi:10.1099/vir.0.81786-0
- Falsey, A. R., Erdman, D., Anderson, L. J., & Walsh, E. E. (2003). Human metapneumovirus infections in young and elderly adults. *J Infect Dis*, 187(5), 785-790. doi:10.1086/367901
- Fearn, R., & Collins, P. L. (1999). Model for polymerase access to the overlapped L gene of respiratory syncytial virus. *J Virol*, 73(1), 388-397.
- Fearn, R., Collins, P. L., & Peebles, M. E. (2000). Functional analysis of the genomic and antigenomic promoters of human respiratory syncytial virus. *J Virol*, 74(13), 6006-6014.
- Fitzgerald, D. J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T. J., & Berger, I. (2006). Protein complex expression by using multigene baculoviral vectors. *Nat Methods*, 3(12), 1021-1032. doi:10.1038/nmeth983
- Gao, Y., & Lenard, J. (1995). Cooperative binding of multimeric phosphoprotein (P) of vesicular stomatitis virus to polymerase (L) and template: pathways of assembly. *J Virol*, 69(12), 7718-7723.
- Garcia, J., Garcia-Barreno, B., Vivo, A., & Melero, J. A. (1993). Cytoplasmic inclusions of respiratory syncytial virus-infected cells: formation of inclusion bodies in transfected cells that coexpress the nucleoprotein, the phosphoprotein, and the 22K protein. *Virology*, 195(1), 243-247. doi:10.1006/viro.1993.1366
- Gattone, V. H., 2nd, Tourkow, B. A., Trambough, C. M., Yu, A. C., Whelan, S., Phillips, C. L., . . . Peterson, R. G. (2004). Development of multiorgan pathology in the wpk rat model of polycystic kidney disease. *Anat Rec A Discov Mol Cell Evol Biol*, 277(2), 384-395. doi:10.1002/ar.a.20022
- Govindarajan, D., & Samal, S. K. (2004). Sequence analysis of the large polymerase (L) protein of the US strain of avian metapneumovirus indicates a close resemblance to that of the human metapneumovirus. *Virus Res*, 105(1), 59-66. doi:10.1016/j.virusres.2004.04.014
- Graci, J. D., & Cameron, C. E. (2006). Mechanisms of action of ribavirin against distinct viruses. *Rev Med Virol*, 16(1), 37-48. doi:10.1002/rmv.483

- Haas, L. E., Thijsen, S. F., van Elden, L., & Heemstra, K. A. (2013). Human metapneumovirus in adults. *Viruses*, 5(1), 87-110. doi:10.3390/v5010087
- Hanley, L. L., McGivern, D. R., Teng, M. N., Djang, R., Collins, P. L., & Fearn, R. (2010). Roles of the respiratory syncytial virus trailer region: effects of mutations on genome production and stress granule formation. *Virology*, 406(2), 241-252. doi:10.1016/j.virol.2010.07.006
- Hansen, J. L., Long, A. M., & Schultz, S. C. (1997). Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure*, 5(8), 1109-1122.
- Harcourt, B. H., Tamin, A., Halpin, K., Ksiazek, T. G., Rollin, P. E., Bellini, W. J., & Rota, P. A. (2001). Molecular characterization of the polymerase gene and genomic termini of Nipah virus. *Virology*, 287(1), 192-201. doi:10.1006/viro.2001.1026
- Horikami, S. M., Curran, J., Kolakofsky, D., & Moyer, S. A. (1992). Complexes of Sendai virus NP-P and P-L proteins are required for defective interfering particle genome replication in vitro. *J Virol*, 66(8), 4901-4908.
- Huang, D., Zhang, Y., & Chen, X. (2003). Analysis of intracellular nucleoside triphosphate levels in normal and tumor cell lines by high-performance liquid chromatography. *J Chromatogr B Analyt Technol Biomed Life Sci*, 784(1), 101-109.
- Inoue, Y., & Shimojo, N. (2013). Epidemiology of virus-induced wheezing/asthma in children. *Front Microbiol*, 4, 391. doi:10.3389/fmicb.2013.00391
- Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., . . . et al. (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc Natl Acad Sci U S A*, 90(13), 6320-6324.
- Johnstone, J., Majumdar, S. R., Fox, J. D., & Marrie, T. J. (2008). Human metapneumovirus pneumonia in adults: results of a prospective study. *Clin Infect Dis*, 46(4), 571-574. doi:10.1086/526776
- Kahn, J. S. (2006). Epidemiology of human metapneumovirus. *Clin Microbiol Rev*, 19(3), 546-557. doi:10.1128/CMR.00014-06
- Kallberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nat Protoc*, 7(8), 1511-1522. doi:10.1038/nprot.2012.085
- Khattar, S. K., Yunus, A. S., & Samal, S. K. (2001). Mapping the domains on the phosphoprotein of bovine respiratory syncytial virus required for N-P and P-L interactions using a minigenome system. *J Gen Virol*, 82(Pt 4), 775-779. doi:10.1099/0022-1317-82-4-775
- Kitanovski, L., Kopriva, S., Pokorn, M., Dolnicar, M. B., Rajic, V., Stefanovic, M., & Jazbec, J. (2013). Treatment of severe human metapneumovirus (hMPV) pneumonia in an immunocompromised child with oral ribavirin and

IVIG. *J Pediatr Hematol Oncol*, 35(7), e311-313.
doi:10.1097/MPH.0b013e3182915d2d

- Kuo, L., Fearn, R., & Collins, P. L. (1997). Analysis of the gene start and gene end signals of human respiratory syncytial virus: quasi-templated initiation at position 1 of the encoded mRNA. *J Virol*, 71(7), 4944-4953.
- Kuo, L., Grosfeld, H., Cristina, J., Hill, M. G., & Collins, P. L. (1996). Effects of mutations in the gene-start and gene-end sequence motifs on transcription of monocistronic and dicistronic minigenomes of respiratory syncytial virus. *J Virol*, 70(10), 6892-6901.
- Kusel, M. M., de Klerk, N. H., Keadze, T., Vohma, V., Holt, P. G., Johnston, S. L., & Sly, P. D. (2007). Early-life respiratory viral infections, atopic sensitization, and risk of subsequent development of persistent asthma. *J Allergy Clin Immunol*, 119(5), 1105-1110. doi:10.1016/j.jaci.2006.12.669
- Lamb, R. A., & Jardetzky, T. S. (2007). Structural basis of viral invasion: lessons from paramyxovirus F. *Curr Opin Struct Biol*, 17(4), 427-436. doi:10.1016/j.sbi.2007.08.016
- Leyrat, C., Renner, M., Harlos, K., & Grimes, J. M. (2013). Solution and crystallographic structures of the central region of the phosphoprotein from human metapneumovirus. *PLoS One*, 8(11), e80371. doi:10.1371/journal.pone.0080371
- Li, J., Fontaine-Rodriguez, E. C., & Whelan, S. P. (2005). Amino acid residues within conserved domain VI of the vesicular stomatitis virus large polymerase protein essential for mRNA cap methyltransferase activity. *J Virol*, 79(21), 13373-13384. doi:10.1128/JVI.79.21.13373-13384.2005
- Li, J., Wang, J. T., & Whelan, S. P. (2006). A unique strategy for mRNA cap methylation used by vesicular stomatitis virus. *Proc Natl Acad Sci U S A*, 103(22), 8493-8498. doi:10.1073/pnas.0509821103
- Liang, B., Li, Z., Jenni, S., Rahmeh, A. A., Morin, B. M., Grant, T., . . . Whelan, S. P. J. (2015). Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. *Cell*, 162(2), 314-327. doi:10.1016/j.cell.2015.06.018
- Liuzzi, M., Mason, S. W., Cartier, M., Lawetz, C., McCollum, R. S., Dansereau, N., . . . Simoneau, B. (2005). Inhibitors of respiratory syncytial virus replication target cotranscriptional mRNA guanylation by viral RNA-dependent RNA polymerase. *J Virol*, 79(20), 13105-13115. doi:10.1128/JVI.79.20.13105-13115.2005
- Ludtke, S. J., Baldwin, P. R., & Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol*, 128(1), 82-97. doi:10.1006/jsbi.1999.4174
- Mackay, I. M., Bialasiewicz, S., Jacob, K. C., McQueen, E., Arden, K. E., Nissen, M. D., & Sloots, T. P. (2006). Genetic diversity of human

- metapneumovirus over 4 consecutive years in Australia. *J Infect Dis*, 193(12), 1630-1633. doi:10.1086/504260
- Mason, S. W., 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(19), 10670-10676.
- Mayer, M. P., & Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci*, 62(6), 670-684. doi:10.1007/s00018-004-4464-6
- Morin, B., Rahmeh, A. A., & Whelan, S. P. (2012). Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase. *EMBO J*, 31(5), 1320-1329. doi:10.1038/emboj.2011.483
- Murphy, L. B., Loney, C., Murray, J., Bhella, D., Ashton, P., & Yeo, R. P. (2003). Investigations into the amino-terminal domain of the respiratory syncytial virus nucleocapsid protein reveal elements important for nucleocapsid formation and interaction with the phosphoprotein. *Virology*, 307(1), 143-153.
- Murray, J., Loney, C., Murphy, L. B., Graham, S., & Yeo, R. P. (2001). Characterization of monoclonal antibodies raised against recombinant respiratory syncytial virus nucleocapsid (N) protein: identification of a region in the carboxy terminus of N involved in the interaction with P protein. *Virology*, 289(2), 252-261. doi:10.1006/viro.2001.1150
- Noton, S. L., Aljabr, W., Hiscox, J. A., Matthews, D. A., & Fearn, R. (2014). Factors affecting de novo RNA synthesis and back-priming by the respiratory syncytial virus polymerase. *Virology*, 462-463, 318-327. doi:10.1016/j.virol.2014.05.032
- Noton, S. L., Deflube, L. R., Tremaglio, C. Z., & Fearn, R. (2012). The respiratory syncytial virus polymerase has multiple RNA synthesis activities at the promoter. *PLoS Pathog*, 8(10), e1002980. doi:10.1371/journal.ppat.1002980
- Noton, S. L., & Fearn, R. (2011). The first two nucleotides of the respiratory syncytial virus antigenome RNA replication product can be selected independently of the promoter terminus. *RNA*, 17(10), 1895-1906. doi:10.1261/rna.2813411
- Noton, S. L., & Fearn, R. (2015). Initiation and regulation of paramyxovirus transcription and replication. *Virology*, 479-480, 545-554. doi:10.1016/j.virol.2015.01.014
- Ogino, T., & Banerjee, A. K. (2007). Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus. *Mol Cell*, 25(1), 85-97. doi:10.1016/j.molcel.2006.11.013

- Ogino, T., & Banerjee, A. K. (2010). The HR motif in the RNA-dependent RNA polymerase L protein of Chandipura virus is required for unconventional mRNA-capping activity. *J Gen Virol*, 91(Pt 5), 1311-1314. doi:10.1099/vir.0.019307-0
- Ogino, T., Kobayashi, M., Iwama, M., & Mizumoto, K. (2005). Sendai virus RNA-dependent RNA polymerase L protein catalyzes cap methylation of virus-specific mRNA. *J Biol Chem*, 280(6), 4429-4435. doi:10.1074/jbc.M411167200
- Paesen, G. C., Collet, A., Sallamand, C., Debart, F., Vasseur, J. J., Canard, B., . . . Grimes, J. M. (2015). X-ray structure and activities of an essential Mononegavirales L-protein domain. *Nat Commun*, 6, 8749. doi:10.1038/ncomms9749
- Parks, G. D. (1994). Mapping of a region of the paramyxovirus L protein required for the formation of a stable complex with the viral phosphoprotein P. *J Virol*, 68(8), 4862-4872.
- Poch, O., Blumberg, B. M., Bougueleret, L., & Tordo, N. (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J Gen Virol*, 71 (Pt 5), 1153-1162. doi:10.1099/0022-1317-71-5-1153
- Poch, O., Sauvaget, I., Delarue, M., & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J*, 8(12), 3867-3874.
- Poranen, M. M., Salgado, P. S., Koivunen, M. R., Wright, S., Bamford, D. H., Stuart, D. I., & Grimes, J. M. (2008). Structural explanation for the role of Mn²⁺ in the activity of phi6 RNA-dependent RNA polymerase. *Nucleic Acids Res*, 36(20), 6633-6644. doi:10.1093/nar/gkn632
- Rahaman, A., Srinivasan, N., Shamala, N., & Shaila, M. S. (2004). Phosphoprotein of the rinderpest virus forms a tetramer through a coiled coil region important for biological function. A structural insight. *J Biol Chem*, 279(22), 23606-23614. doi:10.1074/jbc.M400673200
- Rahmeh, A. A., Schenk, A. D., Danek, E. I., Kranzusch, P. J., Liang, B., Walz, T., & Whelan, S. P. (2010). Molecular architecture of the vesicular stomatitis virus RNA polymerase. *Proc Natl Acad Sci U S A*, 107(46), 20075-20080. doi:10.1073/pnas.1013559107
- Ranjith-Kumar, C. T., Kim, Y. C., Gutshall, L., Silverman, C., Khandekar, S., Sarisky, R. T., & Kao, C. C. (2002). Mechanism of de novo initiation by the hepatitis C virus RNA-dependent RNA polymerase: role of divalent metals. *J Virol*, 76(24), 12513-12525.
- Renner, M., Bertinelli, M., Leyrat, C., Paesen, G. C., Saraiva de Oliveira, L. F., Huiskonen, J. T., & Grimes, J. M. (2016). Nucleocapsid assembly in

- pneumoviruses is regulated by conformational switching of the N protein. *Elife*, 5, e12627. doi:10.7554/eLife.12627
- Renner, M., Paesen, G. C., Grison, C. M., Granier, S., Grimes, J. M., & Leyrat, C. (2017). Structural dissection of human metapneumovirus phosphoprotein using small angle x-ray scattering. *Sci Rep*, 7(1), 14865. doi:10.1038/s41598-017-14448-z
- Scheres, S. H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol*, 180(3), 519-530. doi:10.1016/j.jsb.2012.09.006
- Selisko, B., Potisopon, S., Agred, R., Priet, S., Varlet, I., Thillier, Y., . . . Canard, B. (2012). Molecular basis for nucleotide conservation at the ends of the dengue virus genome. *PLoS Pathog*, 8(9), e1002912. doi:10.1371/journal.ppat.1002912
- Shahda, S., Carlos, W. G., Kiel, P. J., Khan, B. A., & Hage, C. A. (2011). The human metapneumovirus: a case series and review of the literature. *Transpl Infect Dis*, 13(3), 324-328. doi:10.1111/j.1399-3062.2010.00575.x
- Sleat, D. E., & Banerjee, A. K. (1993). Transcriptional activity and mutational analysis of recombinant vesicular stomatitis virus RNA polymerase. *J Virol*, 67(3), 1334-1339.
- Song, J., Liang, B., Han, D., Tang, X., Lang, Q., Feng, R., . . . Liu, A. (2015). Bacterial cell-surface displaying of thermo-tolerant glutamate dehydrogenase and its application in L-glutamate assay. *Enzyme Microb Technol*, 70, 72-78. doi:10.1016/j.enzmictec.2014.12.002
- Sookoian, S., Castano, G., Flichman, D., & Cello, J. (2004). Effects of ribavirin on cytokine production of recall antigens and phytohemagglutinin-stimulated peripheral blood mononuclear cells. (Inhibitory effects of ribavirin on cytokine production). *Ann Hepatol*, 3(3), 104-107.
- Sourimant, J., Rameix-Welti, M. A., Gaillard, A. L., Chevret, D., Galloux, M., Gault, E., & Eleouet, J. F. (2015). Fine mapping and characterization of the L-polymerase-binding domain of the respiratory syncytial virus phosphoprotein. *J Virol*, 89(8), 4421-4433. doi:10.1128/JVI.03619-14
- Tarbouriech, N., Curran, J., Ruigrok, R. W., & Burmeister, W. P. (2000). Tetrameric coiled coil domain of Sendai virus phosphoprotein. *Nat Struct Biol*, 7(9), 777-781. doi:10.1038/79013
- Tran, T. L., Castagne, N., Bhella, D., Varela, P. F., Bernard, J., Chilmonczyk, S., . . . Eleouet, J. F. (2007). The nine C-terminal amino acids of the respiratory syncytial virus protein P are necessary and sufficient for binding to ribonucleoprotein complexes in which six ribonucleotides are contacted per N protein protomer. *J Gen Virol*, 88(Pt 1), 196-206. doi:10.1099/vir.0.82282-0

- Tremaglio, C. Z., Noton, S. L., Deflube, L. R., & Fearn, R. (2013). Respiratory syncytial virus polymerase can initiate transcription from position 3 of the leader promoter. *J Virol*, 87(6), 3196-3207. doi:10.1128/JVI.02862-12
- Ulbrandt, N. D., Ji, H., Patel, N. K., Barnes, A. S., Wilson, S., Kiener, P. A., . . . McCarthy, M. P. (2008). Identification of antibody neutralization epitopes on the fusion protein of human metapneumovirus. *J Gen Virol*, 89(Pt 12), 3113-3118. doi:10.1099/vir.0.2008/005199-0
- van den Hoogen, B. G., Bestebroer, T. M., Osterhaus, A. D., & Fouchier, R. A. (2002). Analysis of the genomic sequence of a human metapneumovirus. *Virology*, 295(1), 119-132. doi:10.1006/viro.2001.1355
- van den Hoogen, B. G., de Jong, J. C., Groen, J., Kuiken, T., de Groot, R., Fouchier, R. A., & Osterhaus, A. D. (2001). A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat Med*, 7(6), 719-724. doi:10.1038/89098
- van den Hoogen, B. G., van Doornum, G. J., Fockens, J. C., Cornelissen, J. J., Beyer, W. E., de Groot, R., . . . Fouchier, R. A. (2003). Prevalence and clinical symptoms of human metapneumovirus infection in hospitalized patients. *J Infect Dis*, 188(10), 1571-1577. doi:10.1086/379200
- Williams, B. G., Gouws, E., Boschi-Pinto, C., Bryce, J., & Dye, C. (2002). Estimates of world-wide distribution of child deaths from acute respiratory infections. *Lancet Infect Dis*, 2(1), 25-32.
- Williams, J. V., Crowe, J. E., Jr., Enriquez, R., Minton, P., Peebles, R. S., Jr., Hamilton, R. G., . . . Hartert, T. V. (2005). Human metapneumovirus infection plays an etiologic role in acute asthma exacerbations requiring hospitalization in adults. *J Infect Dis*, 192(7), 1149-1153. doi:10.1086/444392
- Williams, J. V., Tollefson, S. J., Heymann, P. W., Carper, H. T., Patrie, J., & Crowe, J. E. (2005). Human metapneumovirus infection in children hospitalized for wheezing. *J Allergy Clin Immunol*, 115(6), 1311-1312. doi:10.1016/j.jaci.2005.02.001
- Wright, S., Poranen, M. M., Bamford, D. H., Stuart, D. I., & Grimes, J. M. (2012). Noncatalytic ions direct the RNA-dependent RNA polymerase of bacterial double-stranded RNA virus varphi6 from de novo initiation to elongation. *J Virol*, 86(5), 2837-2849. doi:10.1128/JVI.05168-11
- Wyde, P. R., Chetty, S. N., Jewell, A. M., Boivin, G., & Piedra, P. A. (2003). Comparison of the inhibition of human metapneumovirus and respiratory syncytial virus by ribavirin and immune serum globulin in vitro. *Antiviral Res*, 60(1), 51-59.
- Yang, W., Lee, J. Y., & Nowotny, M. (2006). Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol Cell*, 22(1), 5-13. doi:10.1016/j.molcel.2006.03.013

Chapter 7 Appendix

Rank	Protein Name	Species	Accession No.	Protein MW	Protein Score	Protein Score C. I. %	Total Ion Score	Total Ion C. I. %	Confirmed
1	Heat shock 70 kDa protein cognate 4 OS=Manduca sexta PE=2 SV=1		sp Q9U639 HSP70_MANSE	71672.4	82	99.676	59	98.807	.T.

Supp Fig 1 Report from LC/MS analysis

LC/MS analysis confirms band at 75 kDa as heat shock 70 protein.

Supp Fig 2 Superimposition of methyltransferase and CTD domain from predicted model and VSV L

Superimposition of models showed alignment of secondary structure of predicted model (red) and VSV (cyan).

Supp Fig 3 Phosphoprotein alignment of Pneumovirus

Supp Fig 3: Amino acid alignment of HMPV, RSV and VSV (Uniprot accession number Q8B9Q8, P03421 and P33454). Conserved sequence is highlighted in red. HMPV oligomerisation domain is indicated by the alpha helix representation ($\alpha 1$). Sequence alignment of P protein was made by Uniprot and prepared with ESPript 3.

	VSV		α18	α19	η4	β6	β7
		490	500	510	520	530	540
1	VSV	NPT	S	L	D	T	K
2	HMPV	NAS	S	L	D	T	K
3	HRSVA	KFS	S	L	D	T	K
	VSV		α20	η5	α21		
		550	560	570	580	590	
1	VSV	L	M	S	K	F	P
2	HMPV	M	O	P	G	K	Q
3	HRSVA	M	O	P	G	M	F
	VSV		η6	β8	α22	α23	η7
		600	610	620	630	640	650
1	VSV	K	S	Y	E	A	I
2	HMPV	N	N	Y	I	A	R
3	HRSVA	N	N	Y	I	A	R
	VSV		η8	β10	β11	β12	α25
		660	670	680	690	700	710
1	VSV	G	R	P	D	L	M
2	HMPV	H	A	P	E	T	K
3	HRSVA	H	A	P	E	T	K
	VSV		β14	α26	β15		
		720	730	740	750	760	770
1	VSV	A	Q	G	D	N	Q
2	HMPV	L	N	G	D	N	Q
3	HRSVA	L	N	G	D	N	Q
	VSV		β16	β17	β18	α27	α28
		780	790	800	810	820	830
1	VSV	A	D	Y	L	N	Y
2	HMPV	R	D	L	Q	F	I
3	HRSVA	R	D	L	Q	F	I
	VSV		α29	α30			
		830	840	850	860	870	880
1	VSV	N	P	I	N	A	M
2	HMPV	S	I	I	V	S	L
3	HRSVA	S	I	I	V	S	L
	VSV		α31	α32			
		880	890	900	910	920	930
1	VSV	Y	I	L	D	P	S
2	HMPV	N	I	P	M	Q	F
3	HRSVA	N	I	P	M	Q	F
	VSV		α33	α34			
		940	950	960	970	980	990
1	VSV	R	I	.	.	.	.
2	HMPV	F	K	A	L	L	S
3	HRSVA	L	T	C	I	I	T
	VSV		α35	α36	α37	α38	η9
		990	1000	1010	1020	1030	1040
1	VSV	A	T	I	Y	L	Y
2	HMPV	S	A	I	H	Y	S
3	HRSVA	S	A	I	H	Y	S
	VSV		α39	α40			
		1050	1060	1070	1080	1090	1100
1	VSV	H	R	E	L	D	I
2	HMPV	A	I	N	G	E	D
3	HRSVA	A	I	N	G	E	D

Supp Fig 4 Amino acid alignment of the L protein of the NNS virus

Supp Fig 4: L protein alignment of the VSV, HMPV and RSV (Uniprot accession number P03523, Q6WB93 and P28887). The RdRp domain is in cyan: capping domain is in green; connector domain; yellow; methyltransferase domain is in orange and C-terminal domain is in red. Sequence alignment of P protein was made by Uniprot and prepared with ESPrpt 3.

Rank	Protein Name	Species	Accession No.	Protein MW	Protein Score	Protein Score C. I. %	Total Ion Score	Total Ion C. I. %	Confirmed
1	Phosphoprotein OS=Human metapneumovirus (strain CAN97-83) GN=P PE=3 SV=1		sp Q8B9Q8 PHOS_P_HMPVC	32727.8	919	100	715	100	.T.
Peptide Information									
Calc. Mass	Obsrv. Mass	± da	± ppm	Start Seq.	End Sequence Seq.	Ion Score	C. I. %	Modification	Rank Result Type
806.4406	806.4017	-0.0389	-48	19	25 LAEAFQK				Mascot
861.4676	861.4296	-0.038	-44	35	42 SQSIIGEK				Mascot
1015.6332	1015.5822	-0.051	-50	183	191 LSMILGLLR	44	87.655		Mascot
1015.6332	1015.5822	-0.051	-50	183	191 LSMILGLLR				Mascot
1031.6282	1031.5748	-0.0534	-52	183	191 LSMILGLLR			Oxidation (M)[3]	Mascot
1078.5375	1078.4852	-0.0523	-48	166	175 DTSSLSIEAR				Mascot
1108.6698	1108.6151	-0.0547	-49	26	34 SLRKPSHKR				Mascot
1159.647	1159.5891	-0.0579	-50	69	79 LAWTDKGGAIK				Mascot
1208.598	1208.5388	-0.0592	-49	8	18 DILFMGNEAAK				Mascot
1216.6466	1216.5884	-0.0582	-48	205	215 DGIIRDAMIGIR				Mascot
1256.6958	1256.6354	-0.0604	-48	192	204 TLNIATAGPTAAR	96	100		Mascot
1256.6958	1256.6354	-0.0604	-48	192	204 TLNIATAGPTAAR				Mascot
1328.7056	1328.6351	-0.0705	-53	102	114 VLPSSDGKTPAEK				Mascot
1467.7478	1467.677	-0.0708	-48	126	138 VSFTPNPEPGKYTK				Mascot
1468.5865	1468.5162	-0.0703	-48	230	241 AAEMMEEEEMNQR	56	99.225		Mascot
1468.5865	1468.5162	-0.0703	-48	230	241 AAEMMEEEEMNQR				Mascot
1484.5814	1484.5035	-0.0779	-52	230	241 AAEMMEEEEMNQR			Oxidation (M)[4]	Mascot
1500.5763	1500.4949	-0.0814	-54	230	241 AAEMMEEEEMNQR			Oxidation (M)[4,5]	Mascot
1516.5713	1516.4862	-0.0851	-56	230	241 AAEMMEEEEMNQR			Oxidation (M)[4,5,9]	Mascot
1588.8694	1588.7924	-0.077	-48	69	83 LAWTDKGGAIKTEAK	103	100		Mascot
1588.8694	1588.7924	-0.077	-48	69	83 LAWTDKGGAIKTEAK				Mascot
1595.7145	1595.7545	0.04	25	88	100 VMDPIEEEFTEK				Mascot
1653.7029	1653.6223	-0.0806	-49	228	241 GKAAEMMEEEEMNQR	89	100		Mascot
1653.7029	1653.6223	-0.0806	-49	228	241 GKAAEMMEEEEMNQR				Mascot
1669.6979	1669.6162	-0.0817	-49	228	241 GKAAEMMEEEEMNQR			Oxidation (M)[6]	Mascot
1685.6927	1685.6118	-0.0809	-48	228	241 GKAAEMMEEEEMNQR			Oxidation (M)[6,7]	Mascot
1697.7291	1697.6532	-0.0759	-45	230	243 AAEMMEEEEMNQRTK				Mascot
1697.9293	1697.847	-0.0823	-48	192	208 TLNIATAGPTAARDGIR				Mascot
1751.8157	1751.7399	-0.0758	-43	88	101 VMDPIEEEFTEKR	112	100		Mascot
1751.8157	1751.7399	-0.0758	-43	88	101 VMDPIEEEFTEKR				Mascot
1767.8105	1767.7222	-0.0883	-50	88	101 VMDPIEEEFTEKR	99	100	Oxidation (M)[2]	Mascot
1767.8105	1767.7222	-0.0883	-50	88	101 VMDPIEEEFTEKR			Oxidation (M)[2]	Mascot
1799.9935	1799.8951	-0.0984	-55	209	224 DAMIGIRELIADIIK				Mascot
1837.9694	1837.8792	-0.0902	-49	126	141 VSFTPNPEPGKYTKLEK				Mascot
1844.0562	1843.9648	-0.0914	-50	176	191 LESIEEKLISMILGLLR				Mascot
1860.051	1859.953	-0.098	-53	176	191 LESIEEKLISMILGLLR			Oxidation (M)[10]	Mascot
2128.1682	2128.0693	-0.0989	-46	209	227 DAMIGIRELIADIIKEAK				Mascot
2222.1008	2222.0139	-0.0869	-39	84	101 QTIKVMPIEEEFTEKR				Mascot
2238.0959	2238.0066	-0.0893	-40	84	101 QTIKVMPIEEEFTEKR	65	99.915	Oxidation (M)[6]	Mascot
2238.0959	2238.0066	-0.0893	-40	84	101 QTIKVMPIEEEFTEKR			Oxidation (M)[6]	Mascot
2740.2319	2740.1216	-0.1103	-40	142	165 DALDLLSDNEEEDAESSI LTFEER				Mascot
2850.5823	2850.5034	-0.0789	-28	43	68 VNTVSETLELPTISRPTKP TILSEPK	107	100		Mascot
2850.5823	2850.5034	-0.0789	-28	43	68 VNTVSETLELPTISRPTKP TILSEPK				Mascot
3110.4536	3110.3469	-0.1067	-34	139	165 LEKDALLSDNEEEDAE SSILTFEER				Mascot
3693.032	3693.0273	-0.0047	-1	35	68 SQSIIGKVNTVSETLELP TISRPTKPTILSEPK	44	89.124		Mascot
3693.032	3693.0273	-0.0047	-1	35	68 SQSIIGKVNTVSETLELP TISRPTKPTILSEPK				Mascot
3988.7136	3988.7087	-0.0049	-1	261	294 IVEDESTSGESEEELK DTQENNQEDDIYQLIM				Mascot

Supp Fig 5 Results from mass spectrometry analysis confirming target protein as HMPV Phosphoprotein

Rank	Protein Name	Species	Accession No.	Protein MW	Protein Score	Protein Score C. I. %	Total Ion Score	Total Ion C. I. %	Confirmed
1	RNA-directed RNA polymerase L OS=Human metapneumovirus (strain CAN97-83) GN=L PE=3 SV=1		sp Q6WB93 L_HM PVC	232778.6	334	100	89	100	.T.
Peptide Information									
Calc. Mass	Obsrv. Mass	± da	± ppm	Start Seq.	End Seq.	Sequence	Ion Score	C. I. %	Modification
817.4752	817.5665	0.0913	112	1791	1796	MVILWR			
839.4774	839.576	0.0986	117	899	904	FFEIKR			
860.4988	860.602	0.1032	120	1197	1203	LVPVYNR			
875.412	875.5164	0.1044	119	428	433	WTMYFK			
944.5523	944.6616	0.1093	116	1071	1078	SITNLLQR			
951.4683	951.5773	0.109	115	1491	1497	IDNTFWR	9	0	
951.4683	951.5773	0.109	115	1491	1497	IDNTFWR			
979.4843	979.6026	0.1183	121	597	604	YGDLQLR			
982.5866	982.7015	0.1149	117	1248	1255	CVKPLPR			Carbamidomethyl (C)[1]
988.5938	988.7113	0.1175	119	1196	1203	KLVVYVNR			
1005.5767	1005.6906	0.1139	113	483	490	LIWSVYPK			
1031.5132	1031.6361	0.1229	119	427	433	RWTMYFK			
1049.5231	1049.6456	0.1225	117	1122	1129	LICQISR			Carbamidomethyl (C)[3,4]
1077.5588	1077.6843	0.1255	116	1760	1767	THWDLIHR			
1105.6438	1105.7062	0.0624	56	1502	1510	VMFEPKVKK			
1150.5826	1150.7173	0.1347	117	1256	1264	FMSVNFLLHR			
1150.5826	1150.7173	0.1347	117	1256	1264	FMSVNFLLHR			
1161.6779	1161.8124	0.1345	116	482	490	RLIWSVYPK			
1172.6633	1172.7949	0.1316	112	967	976	ALLSIEKNER			
1189.5961	1189.7351	0.139	117	283	292	ITEHAQFSTR			
1195.6317	1195.7673	0.1356	113	789	798	LKEGETYISR			
1200.7496	1200.725	-0.0246	-20	770	779	LAVKMLKEIR			
1232.6674	1232.8115	0.1441	117	1049	1058	VLYESLPFHK			
1271.6089	1271.7598	0.1509	119	1874	1884	IAVCNDFYAAK			Carbamidomethyl (C)[4]
1375.7692	1375.9287	0.1595	116	48	59	VAIENPVIEHVR	13	0	
1375.7692	1375.9287	0.1595	116	48	59	VAIENPVIEHVR			
1385.6556	1385.8243	0.1687	122	619	629	RNDSYNNYIAR			
1394.7249	1394.8876	0.1627	117	1822	1832	DCNVKLPPFVR			Carbamidomethyl (C)[2]
1394.7249	1394.8876	0.1627	117	1822	1832	DCNVKLPPFVR			Carbamidomethyl (C)[2]
1433.728	1433.8965	0.1685	118	112	122	LNMICDWLQK			Carbamidomethyl (C)[5]
1465.6351	1465.8049	0.1698	116	706	717	YHMGIGEGWCQK			Carbamidomethyl (C)[10]
1516.7166	1516.8929	0.1763	116	501	513	YLEETFNASDSLK			
1533.8127	1533.9919	0.1792	117	1678	1690	TCIGKLMKDLNPK			Carbamidomethyl (C)[2], Oxidation (M)[7]
1553.7815	1553.9637	0.1822	117	75	87	VVEPVNMQHEIMK			
1560.842	1561.0223	0.1803	116	1049	1061	VLYESLPFHKAKE			
1615.8114	1615.9976	0.1862	115	350	362	NLENAEYIYFR			
1615.8439	1615.9976	0.1537	95	1182	1196	GPKSPWGSSTQEKK			
1661.7992	1661.996	0.1968	118	1706	1718	TACEYPDIFVYR			Carbamidomethyl (C)[3]
1663.92	1664.0903	0.1703	102	806	820	VIQSEGMHPTPIKK			
1683.8312	1684.051	0.2198	131	1691	1705	VLYFIGEGAGNWMAR	7	0	
1683.8788	1684.051	0.1722	102	1650	1662	NYMILLPWQHVNR			
1699.8737	1700.048	0.1743	103	1650	1662	NYMILLPWQHVNR			Oxidation (M)[3]
1718.8894	1719.0885	0.1991	116	1575	1589	ITVLNYTDMHALTR			
1733.9545	1734.1476	0.1931	111	293	307	FRNTLLNGLDQLTK			
1734.8844	1735.1437	0.2593	149	1575	1589	ITVLNYTDMHALTR			Oxidation (M)[9]
1744.9341	1745.0281	0.094	54	784	798	NIGHKLKEGETYISR			
1802.9589	1803.1454	0.1865	103	197	210	RVSFFTYNQLLTKW			
1818.0048	1818.208	0.2032	112	581	596	LLADNIVPPFETLTK			
1845.9487	1846.1638	0.2151	117	977	993	ATLTLMRDPAVGSER			
1914.908	1915.1365	0.2285	119	690	705	GEYDIDKIEQSGLYR			

1930.9149	1931.1512	0.2363	122	363	379	IFGHPMVDERDAMDAVK						Mascot
1959.0481	1959.2614	0.2133	109	88	103	NVHSCELTLKQFLTR				Carbamidomethyl (C)[5]		Mascot
1994.0164	1994.2498	0.2334	117	1265	1282	LSVSSRPMEFPASVPAYR						Mascot
2010.0114	2010.2477	0.2363	118	1265	1282	LSVSSRPMEFPASVPAYR				Oxidation (M)[8]		Mascot
2015.0656	2015.2959	0.2303	114	1474	1490	NIKDEDIVDESIDKLLR						Mascot
2028.9774	2029.2131	0.2357	116	1719	1734	SLKDLDLHHYPLEYQR	22	0				Mascot
2029.115	2029.2131	0.0981	48	179	196	LVFIVSSYGCVKSNKSK				Carbamidomethyl (C)[10]		Mascot
2093.9888	2094.2349	0.2461	118	1283	1300	TTNYHFDTSPIQALSER						Mascot
2094.0901	2094.2349	0.1448	69	1768	1785	VSKDALLITLCDAEFKDR				Carbamidomethyl (C)[11]		Mascot
2159.1091	2159.3579	0.2488	115	539	556	QEYLNDKEHIVSLTGKER						Mascot
2160.0828	2160.3469	0.2641	122	70	87	ISDYKVVEPVNMQHEIMK						Mascot
2173.1394	2173.4045	0.2651	122	977	996	ATLTLMRDPQAVGSERQAK						Mascot
2250.1111	2250.374	0.2629	117	316	334	VHSTVLENNDYPMYEVV LK						Mascot
2350.2036	2350.4824	0.2788	119	1028	1048	NEEEVGIIAENITPVYPHG LR						Mascot
2478.2986	2478.5928	0.2942	119	534	554	SYVVRQEYLNDKEHIVSL TGK						Mascot
2675.2947	2675.6218	0.3271	122	683	705	HAPPETKGEYDIDKIEEQ SGLYR	40	71.832				Mascot
2675.2947	2675.6218	0.3271	122	683	705	HAPPETKGEYDIDKIEEQ SGLYR						Mascot
2714.4434	2714.7668	0.3234	119	1360	1383	LVDKITSQHFSPDKIDM LTLGK						Mascot

Supp Fig 6 Results from mass spectrometry analysis confirming target protein as HMPV L protein