THE UNIVERSITY OF READING

Coronavirus proteins and their directed evolution to inhibit virus replication

A thesis submitted for the degree of Doctor of Philosophy in Microbiology/ Virology

By

BAN ODAY ABDULSATTAR

Supervisors:

Dr. Benjamin W. Neuman

Dr. Geraldine Mulley

School of Biological Sciences

The University of Reading, UK

July 2017

Declaration

I confirm that this is my own work and that it has not been previously submitted for a degree at any university or institution. The use of all materials from other sources has been properly and fully acknowledged. All drawings not cited are original artwork by the author.

Ban Oday Abdulsattar

Abstract

Coronaviruses are enveloped, positive sense, RNA viruses that infect many species of animals, including humans. Of the six coronaviruses that can infect humans, SARS-CoV and MERS-CoV are the etiological agents of most concern currently. Coronaviruses possess the most complex and largest RNA genomes among all RNA viruses. The genome contains up to 15 genes with multiple open reading frames (ORFs) encoding both structural and non-structural proteins. Coronaviruses encode about 30 proteins that play specific, and often essential, roles in viral replication and assembly.

This thesis presents work done to express Murine hepatitis virus strain A59 (MHV-A59) proteins such as Nucleoprotein and membrane genes, non-structural proteins (5,6,7,8,9,10,16) from gene1, part of non-structural proteins (PI^{pro}, Y-domain from nsp3 and N-terminal from nsp12) and RdRp from C-terminal part of nsp12 in *E. coli* BL21 cells and mammalian 17clone-1 cells, the latter of which are permissive for MHV-A59. The efficiency of transfection and expression of the proteins in mammalian cells was evaluated. SUMOStar (small ubiquitin-like modifier) fusion technology was used to enhance protein expression in the eukaryotic system. Expressed proteins were detected by Western blot with an anti-His tag antibody. The ability of virus-expressed proteins to interfere with virus infection was tested and an inhibitory effect was detected by plaque assay.

The coronavirus nucleoprotein (N) is an important component for both viral replication and transcription. Error-prone PCR (ep-PCR) was used with the N protein as template to introduce random error and the number of mutations

introduced was calculated after 100 random colonies were sequenced to validate the mutagenesis. Transient expression of N protein was shown to increase the efficiency of infection and virus yield. The function of N was investigated by screening for dominant-negative N mutants, using a library of N variants constructed using ep-PCR. The cytotoxicity of N variants was tested by MTT assay. Expressed N variants showed a range of effects ranging from a 10-fold increase in virus yield associated with the wild type N to 10-fold inhibition of virus growth. One particular N variant, mut38, was non-toxic, but reproducibly inhibited virus growth. The potential to screen for dominant-negative N variants using cell survival was also assessed using different N libraries. The thesis also investigated different strategies aimed at purification of non-structural protein 16 (nsp16). The overall findings suggest an ability of virus-expressed proteins in eukaryotic cells to interfere with virus infection and demonstrate that such antiviral activity can be generated by mutating an important viral protein.

Dedication

This thesis is dedicated to my father and mother, for their endless love, support and encouragement. I am truly thankful for having you in my life. This work is also dedicated to my brother and sisters, Mohammed, Hanan and Jwan, who stand by me when things look bleak and to all the people in my life who touch my heart.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Ben Neuman for his constant support, advice and useful scientific discussions. His guidance helped me in all the time of research. I could not have imagined having a better advisor and mentor for my Ph.D study. I specially thank Dr. Mulley for the assistance she provided at all levels of the research project and during writing of this thesis.

I would like to express my special appreciation and thanks to Professor Ian Jones who gave access to his laboratory and for brilliant suggestions and also for his guidance and help during writing of this thesis.

A very special thanks goes out to Dr. Hawaa Almulla who always kindly advised and helped me. A would also like to thank Saad Mutlk for his help and support.

I would also like to thank all the staff in knight building for their kindness and help, with special thanks to Mrs Jessica del Rio.

My sincere gratitude goes to my sponsor Ministry of Higher Education and Scientific Research in Iraq and Al-Mustansiriyah University; I recognize that this research would not have been possible without their financial support.

Finally, I would like to thank my parents, my brother and sisters. They were always supporting me and encouraging me with their best wishes. Your prayer for me was what sustained me thus far.

And…. To all who I have mentioned and everyone that I forgot, I thank you all once again.

Presentations

1- A novel approach in gene therapy by using directed evolution to generate antiviral drug. Society of General Microbiology annual conference 2015. The international convention centre Birmingham. Presented poster.

2- Suppression of MHV infectivity by virus proteins and dominant negative mutants. Society of General Microbiology annual conference 2016. ACC Liverpool. Presented poster.

3- A novel approach in gene therapy by using directed evolution to generate antiviral drug. Society of General Microbiology annual conference 2017. The Edinburgh International Conference Centre (EICC), Edinburgh. Presented poster.

Table of contents

Declaration ... **1**

Abstract ... **2**

Dedication ... **4**

Acknowledgments ... **5**

Presentations ... **6**

Table of contents ... **7**

List of Figures ... **11**

List of Tables ... **15**

Abbreviations ... **16**

1 Introduction ... **20**

1.1 Coronaviruses ... **20**

1.2 Taxonomy ... **21**

1.3 Morphology ... **25**

1.4 Coronavirus life cycle ... **26**

1.5 Coronavirus genome ... **30**

1.5.1 Non structural proteins ... 31

1.5.2 Structural proteins ... 37

1.5.2.1 Spike protein (S) ... 37

1.5.2.2 Membrane protein (M) ... 40

1.5.2.3 N protein (N) ... 42

1.5.2.4 Envelope protein (E) ... 44

1.6 Antiviral drugs ... **44**

1.7 Therapies against coronavirus infection ... **48**

1.8 Directed evolution ... **52**

1.9 Aims of this research ... **55**

2 Materials and methods ... **56**

2.1 Plasmid construction and cloning of desired DNA fragment ... **56**

2.1.1 Primers ... 56

2.1.2 pTriEx1.1 vector map ... 59

2.1.3 pTriEx1.1-recombinant map 61
2.1.4 SUMO-pTriEx1.1 61
2.1.5 Amplification of DNA fragments 62
2.1.6 Double digest and gel extraction of pTriEx1.1 vector 63
2.1.7 In-Fusion cloning of desired proteins to the pTriEx1.1 vector 64
2.2 DNA agarose electrophoresis 64
2.3 Transformation of *E. coli* competent cells 64
2.4 Colony PCR screening of transformant bacteria 65
2.4.1 Colony PCR analysis 65
2.4.2 Plasmid DNA purification and sequencing 65
2.5 BL21 (DE3)-pLysS transformation 66
2.6 Cloning with SUMOStar 67
2.6.1 Amplification of SUMOStar and desired protein fragments 67
2.6.2 Restriction digest of pTriEx1.1 plasmid and DNA fragments 67
2.6.3 Ligation of PCR fragments in pTriEx1.1 68
2.7 Transfection of mammalian cells 68
2.8 Treatment with proteasome inhibitor 69
2.9 SDS Polyacrylamide gel electrophoresis (PAGE) 69
2.10 Protein staining with coomassie blue 70
2.11 Western blotting analysis 70
2.12 Immunoprecipitation 71
2.13 Error-prone PCR 72
2.14 Directed evolution 73
2.15 Cell culture 74
2.16 Determination of viral titre by plaque assay 74
2.17 Evaluation of cytotoxicity using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide assay (MTT) 75
2.18 Nsp16 protein purification 76
2.18.1 Preparation of LOBSTR *E. coli* competent cells 76
2.18.2 Heat shock transformation of competent LOBSTR *E. coli* 77
2.18.3 Small scale protein expression 77
2.18.4 Large scale protein expression 78
2.18.5 Cell lysis 78
2.18.6 Purification of protein 79

2.18.7 Determination of protein concentration by Bradford Assay 79
2.18.8 Statistical analysis 80

3 Cloning and expression of MHV genes and proteins in *E. coli* BL21(DE3)-pLysS competent cells and mammalian cells. 81

3.1 Introduction 81

3.2 Results 86

3.2.1 PCR amplification of DNA fragments from cDNA of MHV-A59 ... 88

3.2.2 Transformation of In-Fusion products into competent *E.coli* cells90

3.2.3 Correction of random mutation in nsp12_N-terminal 92

3.2.4 Protein expression in *E. coli* BL21(DE3)-pLysS competent cells by IPTG induction 95

3.2.5 Expression and optimization of transfection in mammalian cells……..101

3.2.5.1 Optimization of transfection reagents 102

3.2.5.2 Optimization of cell density 103

3.2.5.3 Optimization of incubation time..... 103

3.3 Discussion 109

4 Inhibition of MHV-A59 infectivity by virus proteins *in vitro*..... 114

4.1 Introduction 114

4.2 Results 117

4.2.1 Cloning strategy for SUMOStar-MHV fragments in pTriEx1.1 vector 117

4.2.2 Amplification of MHV-A59 and SUMOStar fragments 119

4.2.3 Generation of SUMOStar-MHV fragments by overlap extension PCR…….. 120

4.2.4 Ligation of SUMOStar-MHV fragments in pTriEx1.1 vector 121

4.2.5 Transformation in Stellar^{TM} *E. coli* competent cells and restriction digest with *Sac*I and *Xho*I restriction enzymes..... 121

4.2.6 Transfection and expression of SUMOStar-MHV fusion proteins in mammalian cells 123

4.2.7 Suppression of MHV infectivity by virus proteins 124

4.3 Discussion 128

5 Fitness landscape of N protein variants..... 132

5.1 Introduction 132
5.2 Results 138
5.2.1 Plaque assay of virus yield in the presence of N protein expressed *in trans* 138
5.2.2 Error prone PCR 140
5.2.3 Construction of library variant 145
5.2.4 Phylogenetic tree analysis 150
5.2.5 Effect of mutated N proteins on virus replication 152
5.2.6 MTT assay 158
5.3 Discussion 160
6 Directed evolution 165
6.1 Introduction 165
6.2 Results 168
6.2.1 Directed evolution for N protein and mut38 168
6.2.1.1 Round 1: wild type and mut38 168
6.2.1.2 Round 2: wild type and mut38 172
6.2.1.3 Round 3: wild type and mut38 173
6.2.1.4 Round 4: wild type and mut38 173
6.2.2 Directed evolution for 10 variants 174
6.2.3 Directed evolution of a complex N protein library 177
6.3 Discussion 180
7 Expression and purification of MHV-A59 Mtase protein 183
7.1 Introduction 183
7.2 Results 187
7.2.1 Optimizing conditions of nsp16 expression and purification 187
7.3 Discussion 191
8 General discussion 193
9 Appendix 198
References 203

List of Figures

Figure 1.1. Phylogeny of nidoviruses in comparison to the Tree of life (ToL). 22
Figure 1.2. Schematic representation of the taxonomy of *coronaviridae* according to the International Committee on Taxonomy of Viruses.....24
Figure 1.3. (A) Electron microscopy image of coronavirus. (B) Schematic diagram of a coronavirus particle 26
Figure 1.4. A schematic representation of MHV-A59 genome organization and expression 27
Figure 1.5. Life cycle of SARS-CoV 29
Figure 1.6. Coronavirus genome organization 31
Figure 1.7. Schematic of the S protein domains 38
Figure 1.8. Coronavirus M protein structure 40
Figure 1.9. Domain organization of coronavirus nucleocapsid protein.....43
Figure 2.10. Map of the pTriEx1.1 vector showing the cloning sites. 60
Figure 2.11. Vector map of the pTriEx1.1 vector showing the site of insertion61
Figure 2.12. Vector map of the pTriEx1.1 vector showing the site of protein and SUMO tag insertion. 62
Figure 3.13. The In-Fusion system.....84
Figure 3.14. Cloning design using In-Fusion protocol 87
Figure 3.15. Amplification of membrane gene, nsp3_Y-domain, nsp8 and nsp9 from cDNA of MHV-A59 88
Figure 3.16. Amplification of nsp12, nsp12_C-terminal, N protein, nsp5, nsp6 and nsp3_Pl^{Pro} from cDNA of MHV-A59 89
Figure 3.17. Agarose gel electrophoresis of nsp10, nsp12_N-terminal and nsp16 amplified by PCR from cDNA of MHV-A59 89
Figure 3.18. Double digest of pTriEx1.1 90
Figure 3.19. N protein screened by PCR 91
Figure 3.20. Gel electrophoresis of double digested pTriEx1.1 containing nsp16 and N protein. 92
Figure 3.21. Gel electrophoresis of PCR products following amplification of nsp12_N-terminal for one transformant using protein-specific primers....93
Figure 3.22. Agarose gel electrophoresis of two joined parts of nsp12_N-terminal amplified by PCR.....94

Figure 3.23. Gel electrophoresis of colony PCR of Nsp12N in pTriEx1.1 plasmid 95
Figure 3.24. Western blot analysis of recombinant expression of nsp16 in *E. coli* BL21(DE3)-pLysS competent cells 96
Figure 3.25. Western blot analysis of recombinant expression of N protein in *E. coli* BL21(DE3)-pLysS competent cells 97
Figure 3.26. Western blot analysis of recombinant expression of nsp8 in *E. coli* BL21(DE3)-pLysS competent cells 97
Figure 3.27. Western blot analysis of recombinant expression of nsp3_Y-domain in *E. coli* BL21(DE3)-pLysS competent cells 98
Figure 3.28. Western blot analysis of recombinant expression of nsp12_C-terminal in *E. coli* BL21(DE3)-pLysS competent cells 98
Figure 3.29. Western blot analysis of recombinant expression of nsp7 in *E. coli* BL21(DE3)-pLysS competent cells 99
Figure 3.30. Western blot analysis of recombinant expression of membrane protein (M) in *E. coli* BL21(DE3)-pLysS competent cells 100
Figure 3.31. Western blot analysis of recombinant expression of nsp3_Pl^{Pro} in *E. coli* BL21(DE3)-pLysS competent cells 100
Figure 3.32. Western blot analysis of recombinant expression of nsp12_N-terminal in *E. coli* BL21(DE3)-pLysS competent cells 101
Figure 3.33. Western blot analysis of recombinant expression of nsp9 in *E. coli* BL21(DE3)-pLysS competent cells 101
Figure 3.34. *In vitro* transfection of 17clone-1 cells with pTriEx1.1-GFP using different transfection reagents 102
Figure 3.35. *In vitro* transfection of 17clone-1 cells with pTriEx1.1-GFP using different cells density 103
Figure 3.36. N protein expressed in mammalian cells 105
Figure 3.37. Western blot analysis of MHV-A59 proteins using conjugated and unconjugated antibody 106
Figure 3.38. Protein inhibitor does not stimulate protein expression 107
Figure 3.39. Nsp16 protein is not detected by immunoprecipitation 108
Figure 4.40. Cloning design 118
Figure 4.41. Gel electrophoresis of nsp7, nsp8, nsp9, nsp10, nsp16, nsp12_C-terminal, nsp3_Y-domain amplified by PCR from cDNA of MHV-A59.. . 119

Figure 4.42. Gel electrophoresis of SUMOStar PCR product following amplification of SUMOStar fragment 120
Figure 4.43. Nsp7 screened by PCR 122
Figure 4.44. Gel electrophoresis of double digest pTriEx1.1 contains SUMOStar-nsp7 fragment 123
Figure 4.45. Protein expression in mammalian cells using SUMOStar fusion tag 124
Figure 4.46. *In vitro* effect of transient expression of MHV-SUMO fusion proteins on MHV virus replication 126
Figure 4.47. Effect of the MHV-A59 proteins on the virus replication *in vitro*. 127
Figure 5.48. N protein effect on MHV-A59 virus replication 139
Figure 5.49. Affect of N protein expression on MHV replication in 17clone-1 mammalian cells 140
Figure 5.50. Gel electrophoresis of N protein amplified by first error-prone PCR protocol from cDNA of MHV-A59 141
Figure 5.51. N protein screened by PCR 141
Figure 5.52. Agarose gel electrophoresis of N protein amplified by second PCR protocol from cDNA of MHV-A59 142
Figure 5.53. Gel electrophoresis of error-prone libraries double digest containing N protein 143
Figure 5.54. Error-prone libraries screened by colony PCR 144
Figure 5.55. Distribution of amino acid mutations in 2 variants from 50 μ M $MnCl_2$ and 100 μ M $MnCl_2$ libraries generated with error-prone PCR. 146
Figure 5.56. Distribution of amino acid mutations in 10 variants generated with error-prone PCR 149
Figure 5.57. Phylogenetic analysis between non-mutated and mutated N protein 151
Figure 5.58. Effect of mutated N protein on MHV-A59 virus replication. 153
Figure 5.59. Inhibitory activity of different mutated N protein on virus replication tested by plaque assay 154
Figure 5.60. Inhibition of MHV-A59 infectivity by mut38 155
Figure 5.61. Comparison of wild type and mut38 nucleotides sequences 156

Figure 5.62. Comparison of wild type and mut38 amino acid sequences. 157
Figure 5.63. Cytotoxicity of wild type and mut38 159
Figure 6.64. Work flow for directed evolution of MHV N protein in 17clone-1 cells 166
Figure 6.65. Gel electrophoresis of N protein extracted from survived cell. ..169
Figure 6.66. Gel electrophoresis of error-prone library double digest containing variants (wild type and mut38) of N protein 170
Figure 6.67. N protein (wild type and mut38) screened by PCR 171
Figure 6.68. Sequence alignment of 10 isolates with wild type. 171
Figure 6.69. Gel electrophoresis of N protein extracted from survived cell (2nd round of selection) 172
Figure 6.70. Gel electrophoresis of N protein extracted from survived cell. ..175
Figure 6.71. Gel electrophoresis of error-prone library (10 variants) double digest 175
Figure 6.72. N protein (10 variants library) screened by colony PCR 176
Figure 6.73. Gel electrophoresis of N protein generated by error prone PCR (50 μ M $MnCl_2$) 178
Figure 6.74. Inhibition of virus replication tested by plaque assay. 179
Figure 7.75. Expression and purification of nsp16 protein from BL21 (DE3)-pLysS competent cells 188
Figure 7.76. Expression and purification of nsp16 protein from LOBSTR competent cells 189
Figure 7.77. Expression and purification of nsp16 protein from LOBSTR competent cells in AIM media 190
Figure 9.78. The sequence of SUMOStar tag protein 199
Figure 9.79. Distribution of nucleotide mutations in 10 variants generated by ep-PCR 200
Figure 9.80. Distribution of nucleotide mutations in 10 variants generated by ep-PCR 201
Figure 9.81. Distribution of amino acid mutations in 37 variants generated with ep-PCR and subjected to plaque assay 202

List of Tables

Table 2.1. Oligonucleotides used for cloning in pTriEx1.1 plasmid. 57
Table 2.2. Sequencing primers 57
Table 2.3. Primers used for correction of random mutation in nsp12_N-terminal. 58
Table 2.4. Primers used for SUMO cloning in pTriEx1.1 plasmid. 58
Table 2.5. Primers for SUMOStar cloning 59
Table 5.6. Sequence analysis and mutation rate of 10 mutated colonies generated by error-prone PCR. 147
Table 5.7. Point mutation sequence analysis and mutation rates of variant library generated by ep-PCR. 148
Table 5.8. Point mutation sequence analysis and mutation rates of 100 variants library generated by error-prone PCR. 150
Table 6.9. Screening of N protein variants after 3 rounds of directed evolution. 177
Table 9.10. Coronavirus MHV-A59 protein sizes 198

Abbreviations

Amino acid

A Alanine	G Glycine	M Methionine	S Serine
C Cysteine	H Histidine	N Asparagine	T Threonine
D Aspartic acid	I Isoleucine	P Proline	V Valine
E Glutamic acid	K Lysine	Q Glutamine	W Tryptophan
F Phenylalanine	L Leucine	R Arginine	Y Tyrosine

A	Adenine
Ab	Antibody
Ag	Antigen
AIM	Auto induction media
Amp	Ampicillin
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
°C	Degree Celsius
cAMP	Cyclic adenosine monophosphate
$CaCl_2$	Calcium chloride
CEACAM1	Carcinoembryonic antigen-related cell adhesion molecule1
CMV	Cytomegalovirus
CoV	Coronavirus
Cryo-EM	Electron cryomicroscopy
CTD	C-terminal dimerization domain
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dH_2O	Distilled water
DMEM	Dulbecco's modified Eagle medium

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotidet riphosphat
dTTP	Deoxythymidine triphosphate
ECL	Enhanced chemiluminescence
E. coli	*Escherichia coli*
EDTA	Ethylene Diamine Tetra Acetic acid
EM	Electron microscopy
ep-PCR	Error-prone polymerase chain reaction
ERGIC	Endoplasmic Reticulum-Golgi Intermediate Compartment
FBS	Fetal bovine serum
G	Guanine
GFP	Green Florescent Protein
GST	Glutathione-S -transferase
HCoV-HKU1	Human coronavirus- Hong Kong University strain1
HEK293 cells	Human embryonic kidney cells 293
HRP	Horse reddish peroxidase
IBV	Infectious Bronchitis Virus
IDT	Integrated DNA Technology
IFN	Interferon
IMAC	Immobilized metal ion chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilo base pairs
kDa	Kilo Daltons
LB	Luria- Bertani
LOBSTR	Low background strain competent cells
M	Molar
MBP	Maltose binding protein
MERS	Middle Eastern Respiratory Syndrome
mg	Milligram(s)
µg	Micro gram
$MgCl_2$	Magnesium chloride
MHV	Mouse Hepatitis Virus

ml	Milliliter(s)
µl	Micro liter(s)
mM	Millimolar
µM	Micro molar
$MnCl_2$	Manganese (II) chloride
MOI	Multiplicity of infection
M protein	Membrane protein
mRNA	Messenger RNA
MTases	Methyltransferases
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5- diphenyltetrazolium bromide
N	N protein protein
NEAA	Non-essential amino acids
ng	Nanogram(s)
nm	Nanometer(s)
NMR	Nuclear magnetic resonance
nsp	Non structural protein
NTD	N-terminal RNA-binding domain
OD	Optical density
oligo	Oligonucleotides
ORF	Open Reading Frame
ori	Origin of replication
PBS	Phosphate Buffer Saline
PCR	Polymerase chain reaction
PEDV	Porcine epidemic diarrhea virus
PFU	Plaque forming unit
Pfu-Pol	*Pyrococcus furious* polymerase
pH	Hydrogen potential
pi	Post infection
PI^{pro}	Papain-like proteinases
PVDF	Polyvinylidene fluoride
RdRp	RNA dependent RNA polymerase
RNA	Ribonucleic acid
RNP	Ribonucleoprotein complex

rpm	Revolutions per minute
RTC	Replication transcription complex
SARS	Sever Acute Respiratory Syndrome
SAXS	Small-angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SiRNA	Small interfering RNA
SOC	Super Optimal Broth
SR	Serine/Arginine
ssDNA	Single stranded deoxyribonucleic acid
ssRNA	Single stranded ribonucleic acid
SUMOStar	Small ubiquitin-related modifier
T	Thiamine
Taq-Pol	*Thermus aquaticus* polymerase
TBST	Tris-buffered saline Tween20
TGEV	Transmissible gastroenteritis virus
TPase	RNA Triphosphatase
Trx	Thioredoxin
Ub	Ubiquitin
v/v	Concentration, volume to volume
VLPs	Virus-like particles
WT	Wild type
w/v	Concentration, weight to volume

1 Introduction

1.1 Coronaviruses

Coronaviruses are a group of important pathogens that infect both humans and a variety of domesticated animals. Infections caused by this group have resulted in significant economic losses for the domestic animal industry worldwide and has recently threatened human health with multiple outbreaks, becoming potentially pandemic. Coronaviruses have been studied since the 1930's; the first described virus was called avian infectious bronchitis virus (IBV) followed by murine hepatitis virus (MHV) and transmissible gastroenteritis virus (TGEV) (Beaudette and Hudson, 1937; Doyle and Hutchings, 1946; JB, 1949; Schalk and Hawn, 1931). However, the relationships between these viruses was not clearly realized until the 1960's when the human coronaviruses HCoV-229E and HCoV-OC43 were described (Hamre and Procknow, 1966). All of these viruses possessed a distinct morphology of a crown-like appearance and so were called *coronaviruses* (Tyrell *et al*., 1968). Coronaviruses (CoVs) may infect and cause diseases in a wide variety of animals, including bats, birds, cats, dogs, pigs, mice, horses and whales in addition to humans (De Groot,RJ *et al*., 2012). They can cause respiratory, enteric, hepatic, and neurological diseases with highly variable severity and cause acute or persistent infections.

The first human coronavirus epidemic occurred in 2003 with the outbreak of severe acute respiratory syndrome (SARS) caused by SARS-CoV (Marra *et al*., 2003; Rota *et al*., 2003). Then, in 2012 a novel coronavirus emerged in the Middle East (Zaki *et al*., 2012). The novel Middle East Respiratory Syndrome coronavirus

(MERS-CoV) causes severe pneumonia as well as renal failure, with a high fatality rate.

Over the past ten years significant effort has resulted in the discovery of new additional human coronaviruses such as HCoV-NL63 and HCoV-HKU1 (van der Hoek *et al*., 2004; Woo *et al*., 2005b) and in the development of effective therapies by both the academic sector and pharmaceutical industries. Despite this and although more than 10 years have passed since the SARS-CoV outbreak, there are presently still no available approved vaccines or antiviral drugs for human coronavirus infection. Therefore, revealing any undiscovered aspects of the coronavirus genome and it's replication might help improve our understanding of the role of each gene and suggest useful antiviral drug targets.

1.2 Taxonomy

Coronaviruses are a group of enveloped complex ssRNA viruses that belong to the order *Nidovirales*. The order *Nidovirales* is a large group of RNA viruses that posses the largest genomes known to date. The phylogenetic analysis of nidovirus members shows a great genetic distance, which is almost equivalent to that of the archaea, eubacteria and eukaryote combined **Figure 1.1,** and due to the extremely high mutation rates and lack of fossil records this group remains difficult to study.

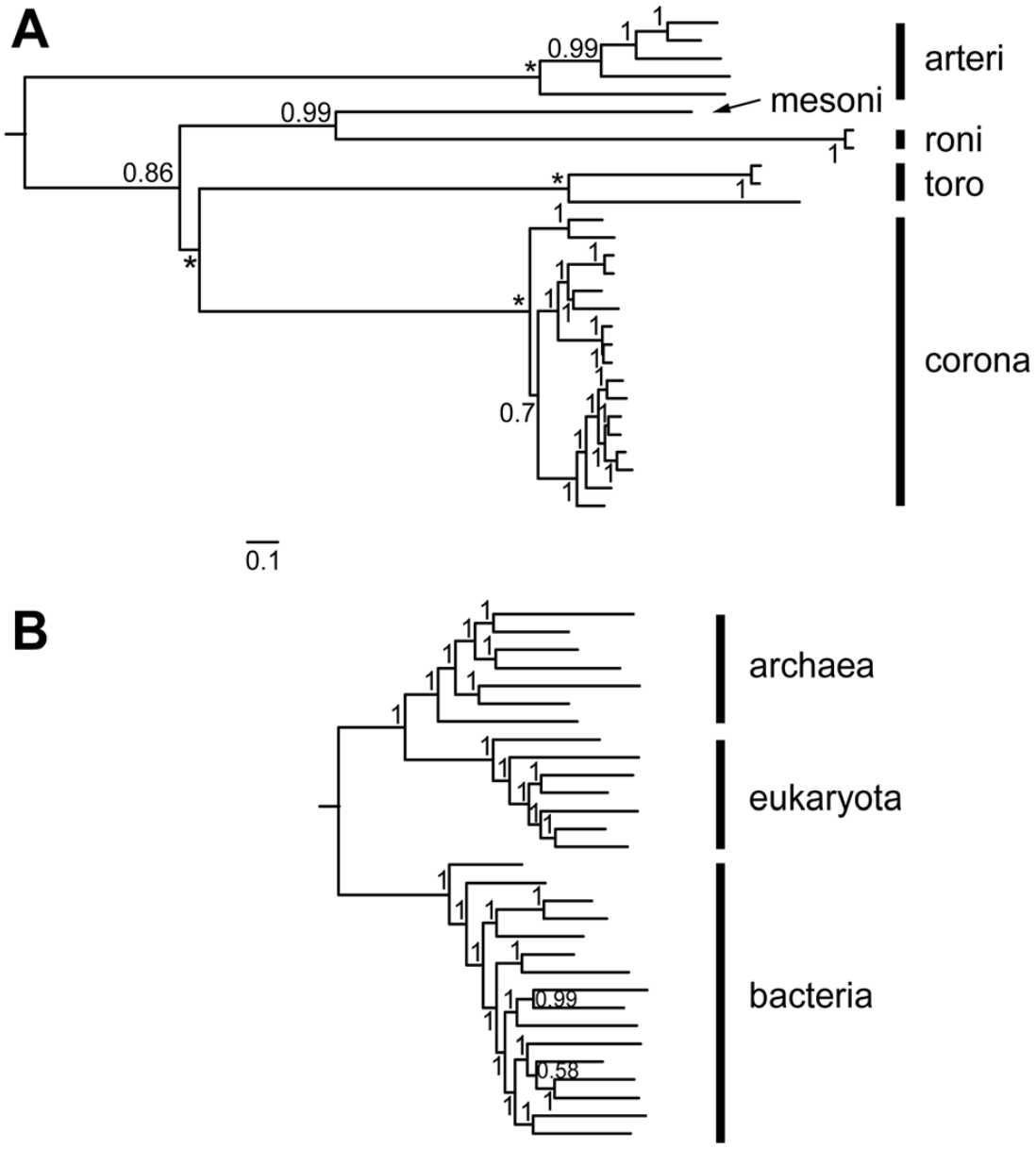

Figure 1.1. Phylogeny of nidoviruses in comparison to the Tree of life (ToL). Bayesian phylogenies of nidoviruses (A) and ToL (B) are drawn to a common scale of 0.1 amino acid substitutions per position. Major lineages are indicated by vertical bars and names; arteri: *Arteriviridae*, mesoni: *Mesoniviridae*, roni: *Roniviridae*, toro: *Torovirinae*, corona: *Coronavirinae*. Rooting was according to either (A) domain- specific outgroups (Nga *et al*., 2011) or (B) as described (Boussau *et al*., 2008). A common normalized scale of (0,1) is used. Posterior probability support values and fixed basal branch points (*) are indicated. The nidovirus and ToL alignments include, respectively, three enzymes and 56 single-gene protein families, 604 and 3336 columns, 2.95% and 2.8% gaps. Taken from (Lauber *et al*., 2013).

Nidoviruses are named for their eminent feature of a set of multiple 3'-nested subgenomic RNAs (nidus = nest). Nidoviruses possess a linear 5'-capped positive sense single stranded, non-segmented RNA with two large open reading frames (ORFs) 1a and 1b located at the 5'-end of the genome. The order historically included three families, which are *coronaviridae*, *roniviridae* and *arteriviridae*, the first two families with large genome size (26.3-31.3 kb) and the third one with smaller genome sizes (12.7-15.7 kb) (De Groot,RJ *et al*., 2012). Recently a fourth family *Mesonviridae* joined the previous three families with an intermediate genome size between those of the *coronaviridae* and *arteriviridae* (Gorbalenya *et al*., 2006; Lauber *et al*., 2012; Nauwynck *et al*., 2012). *Coronavirinae* and *torovirinae* are two subfamilies of *coronaviridae*. Until a few years ago the *coronavirinae* family was divided into three groups (I, II and III) based on genotypic and serological differences. However these groups have since been elevated to genus level and named alphacoronavirus, betacoronavirus and gammacoronavirus, which respectively represent groups I, II and III (Gonzalez *et al*., 2003) and a new genus, deltacoronavirus, was added recently (King *et al*., 2011). The genus alphacoronavirus is further separated into a and b clades while betacoronavirus are subdivided into clades a to d. The alphacoronavirus genus is exclusively found in mammalian hosts, and includes the human pathogens HCoV-229E and HCoV-NL63, in addition to porcine Epidemic Diarrhea Virus (PEDV) and Transmissible Gastroenteritis Virus (TGEV) in pigs **Figure 1.2**.

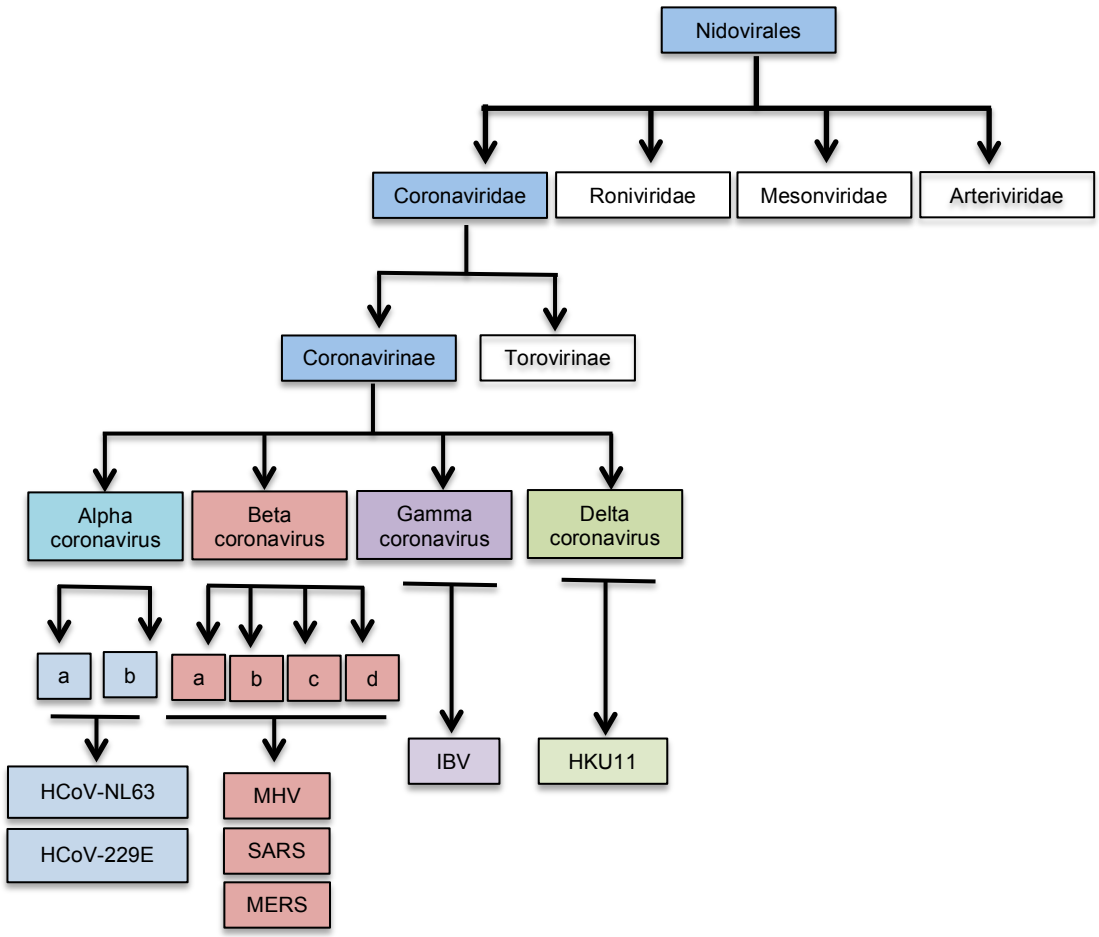

Figure 1.2. Schematic representation of the taxonomy of *coronaviridae* according to the International Committee on Taxonomy of Viruses. Human coronavirus-NL63 (HCoV-NL63), Human coronavirus-229E (HCoV-229E), Murine hepatitis virus (MHV), Severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), Avian infectious bronchitis virus (IBV), Bulbul coronavirus HKU11 (HKU11).

Betacoronaviruses are also associated exclusively with mammalian hosts. Examples of pathogenic betacoronaviruses include the human pathogen HCoV-OC43, murine hepatitis virus (MHV) and SARS-CoV, which causes severe acute respiratory syndrome. Intensive studies led to the finding of several more novel coronavirus in human and animals after the 2003 SARS-CoV epidemic outbreak (Poon *et al*., 2005). New studies indicate that bats are the natural reservoirs of these two genera and suggest that bats are likely to play an important role in the introduction of coronaviruses to other species as well as the evolution and

dissemination of coronaviruses (Tong *et al.*, 2009). Conversely, the majority of gammacoronavirus are isolated from birds such as Infectious Bronchitis Virus (IBV) in chickens and Grey Goose coronavirus (GCoV) in geese. A new genus, deltacoronavirus, was detected in pigs and birds, and includes Bulbul coronavirus HKU11, Thrush coronavirus HKU12 and Munia coronavirus HKU13 (Chu *et al*., 2011). Both gammacoronavirus and deltacoronavirus reservoirs are birds (Woo *et al*., 2012).

1.3 Morphology

The virion of coronavirus is spherical or pleomorphic, approximately 50-150nm in diameter (Neuman, 2008). The virion membrane as shown in **Figure 1.3** contains three viral proteins, the spike protein (S) which gives the virus it's crown shape under the electron microscope and plays a role in viral attachment to the cell and the subsequent fusion process, the membrane protein (M) with multi spanning membrane domains, large carboxy terminus and small amino terminus, and the envelope protein (E) with highly hydrophobic amino and carboxy-terminal regions (Bond *et al*., 1979). Both E protein and M proteins are involved in the virus assembly process (Hsieh *et al*., 2005). Some coronaviruses from the betacoronavirus group also contain a hemagglutinin esterase (HE) protein on the virion surface that may serve as another protein for binding or release from the host cell. Moreover, SARS-CoV contains accessory proteins such as 3a, 6 and 7a, that are involved in cellular processes or modulating virus-host interactions (McBride and Fielding, 2012). Inside the virion there is a helical nucleocapsid that contains the largest known viral RNA genome associated with the nucleoprotein (N) (Lai and Anderson, 2007).

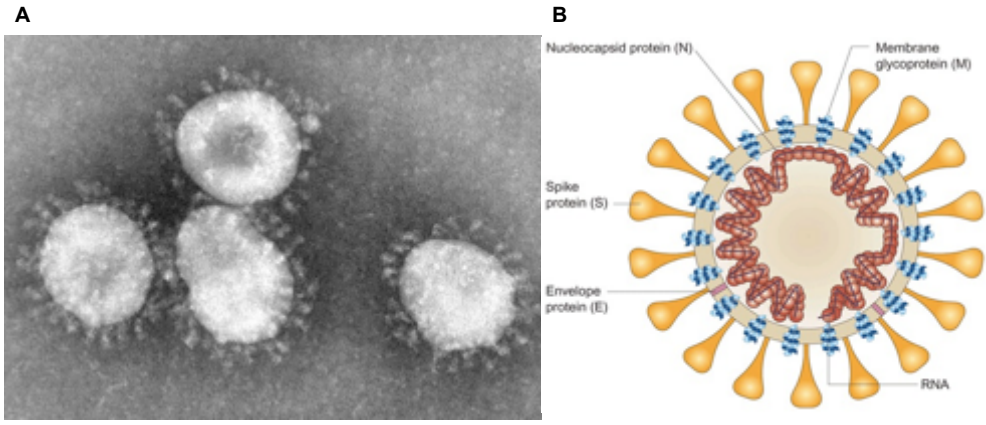

Figure 1.3. (A) Electron microscopy image of coronavirus. Adapted from (Davies and Macnaughton, 1979). **(B) Schematic diagram of a coronavirus particle.** The viral RNA is associated with the nucleocapsid protein (N). The lipid bilayer includes the spike protein (S), the membrane protein (M) and the envelope protein (E) (Figure is adapted from (Stadler *et al*., 2003)).

1.4 Coronavirus life cycle

The first step of the coronavirus life cycle starts by interaction of the spike protein with a specific receptor on the host cell surface. There are different receptors across the coronavirus family: carcinoembryonic antigen 1 (CEACAM1) for MHV (Tan *et al*., 2002), human angiotensin converting enzyme 2 (hACE2) for SARS-CoV (Li, 2013) and NL63 (Glowacka *et al*., 2010; Wu *et al*., 2009), dipeptidyl peptidase 4 (DPP4) for MERS-CoV (Wang *et al*., 2013) and aminopeptidase N (APN) for TGEV (Shahwan *et al*., 2013). After attachment and uptake into a vesicle, a major conformational change occurs in the spike protein (S) (Gallagher and Buchmeier, 2001). The spike glycoprotein typically can be divided into the S1 domain (the amino-proximal half), which contains the receptor-binding domain and the S2 domain (the carboxyl- proximal half), which contains elements involved in membrane fusion (Heald-Sargent & Gallagher, 2012). The two domains are cleaved from each other by the activity of a cellular furin-like enzyme (de Haan *et*

al., 2004). Like most RNA viruses, coronaviruses replicate in the cytoplasm of an infected cell. Once the genome has entered the cytoplasm it serves as an mRNA since it has a 5'-methyl cap structure and a polyadenylated tail, which mimics mRNAs of eukaryotes. Two thirds of the genome at the 5' end is occupied by the two large open reading frames (ORFs 1a and 1b) which together encode the replicase gene while the rest of the genome encodes structural and accessory genes in the 3' one-third. A frame shift region in the replicase gene connects ORF1a and ORF1b and will direct the expression of ORF1b to facilitate the formation of an ORF1ab polyprotein (pp1ab) (Bredenbeek *et al*., 1990; Brian and Baric, 2005; Brierley and Dos Ramos, 2006) **Figure 1.4**.

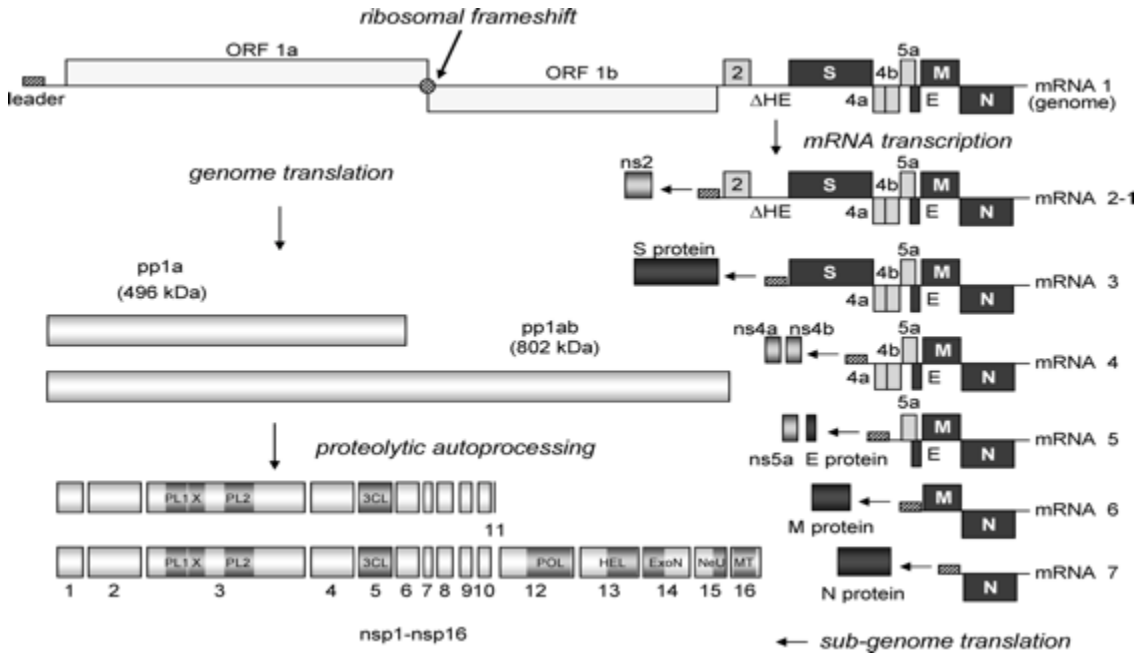

Figure 1.4. A schematic representation of MHV-A59 genome organization and expression. Replicase components are processed from two forms of polyprotein translated from the genomic RNA into non-structural proteins (nsp1- nsp16) by autoproteolytic processing. Nested sets of positive sense sub-genomic sized mRNAs are translated by host ribosomes into viral structural (S, E, M, N) and accessory proteins. Figure adapted from (Sawicki *et al*., 2005).

After translation, the papain-like protease(s) (PL^{Pro}) and main protease (M^{Pro}) proteins, which are encoded in ORF1ab, cleave the replicase polyprotein into 15 to 16 non-structural proteins (nsps) (Woo *et al*., 2007, 2005a). The non-structural proteins are anchored onto double membrane vesicles (DMVs) to make up the bulk of the replication/transcription complex (RTC) and this is the site where all events for virus transcription/replication take place (Brockway *et al*., 2003; Gosert *et al*., 2002; Snijder *et al*., 2006). A full-length negative-stranded RNA intermediate serves as template for the synthesis of more full-length positive sense RNA (Sawicki *et al*., 2007). While not completely understood, subgenomic mRNA synthesis in the *nidovirales* involves a unique discontinuous transcription mechanism which produces a 3' co-terminal nested set of mRNAs (La Monica *et al*., 1992). The mRNA synthesis is regulated by transcription-regulating sequences (TRSs) present in the genomic RNA upstream of most open reading frames (La Monica *et al*., 1992). Both new copies of the complete genome and sub-genomic mRNA species, which are synthesized from the negative strand RNA intermediate later, are translated into structural and accessory proteins. After translation of the structural proteins, the N protein wraps the genomic RNA to form the nucleocapsid (a helical structure). All structural proteins, and the HE protein in some species, are located in the endoplasmic reticulum (ER) and aggregate with the nucleocapsid in the endoplasmic reticulum Golgi intermediate compartment (ERGIC) for virion assembly, which is driven by M and E proteins (Hsieh *et al*., 2005; Neuman *et al*., 2011; Tseng *et al*., 2005). Later, virions are released extracellularly either from the basolateral and apical surfaces for MHV (Rossen *et al*., 1996) or from the apical surface of the host cell for TGEV and SARS-CoV (Rossen *et al*., 1994; Tseng *et al*., 2005) **Figure 1.5**.

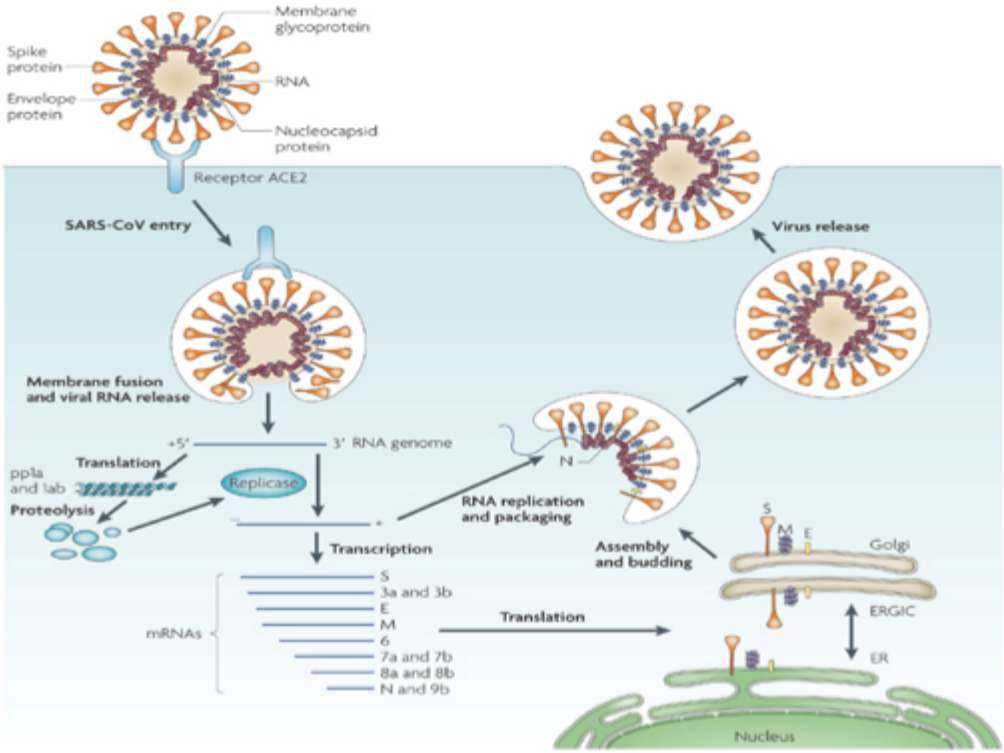

Figure 1.5. Life cycle of SARS-CoV. First spike protein binds to the specific receptor (ACE2) to form ACE2-virus complex. Then the complex is translocated to endosomes in which S protein is cleaved by the endosomal acid protease (cathepsin L) for activation of it's fusion activity. The viral genome is released inside the cell, and then replicase gene is translated to polyproteins pp 1a and 1ab, which is cleaved into small products by viral proteinases. Subgenomic negative-strand templates are synthesized from discontinuous transcription on the plus-strand genome and act as templates for mRNA synthesis to produce negative-sens RNA and subgenomic RNAs. Nucleocapsid is assembled from genomic RNA and N protein in the cytoplasm. The structural proteins S, E, M are translated and inserted in the endoplasmic reticulum (ER) to meet nuclecapsid in the endoplasmic reticulum (ER)-Golgi intermediate compartment ERGIC. Virions are released from the cell by exocytosis. Figure adapted from (Du *et al*., 2009).

1.5 Coronavirus genome

Coronaviruses possess the largest and most complex RNA genome among all RNA viruses ranging from 26.4 kb - 31.7 kb (Woo *et al*., 2009). The genome contains up to 15 genes with multiple open reading frames (ORFs) encoding both structural and non-structural proteins (Sawicki *et al.*, 2005; Ziebuhr, 2005). The RNA genome comprises a 5' leader sequence (60-80 nucleotides), followed by an un-translated region (UTR) (200-600 nucleotides). The first gene (replicase) comprises two-thirds of the genome with the two large ORFs (ORF 1a and ORF 1b) translated to give polyprotein 1a and polyprotein 1ab, the latter via a frameshift mechanism. Despite it's positive strand nature the replicase is the only translated product derived from the genome. The last third of the genome is occupied by genes for structural proteins with the order S-E-M-N, all of which are expressed following the production of several subgenomic RNAs. Between these genes there are a variable numbers of ORFs encoding accessory proteins (Sawicki *et al*., 2007). The transcriptional regulatory sequences (TRSs) that are present at the 3' end of each gene represent a signal for subgenomic RNA transcription of the following gene. At the 3' end of the genome, there are 270-500 nucleotides of UTR followed by the poly A tail **Figure 1.6**.

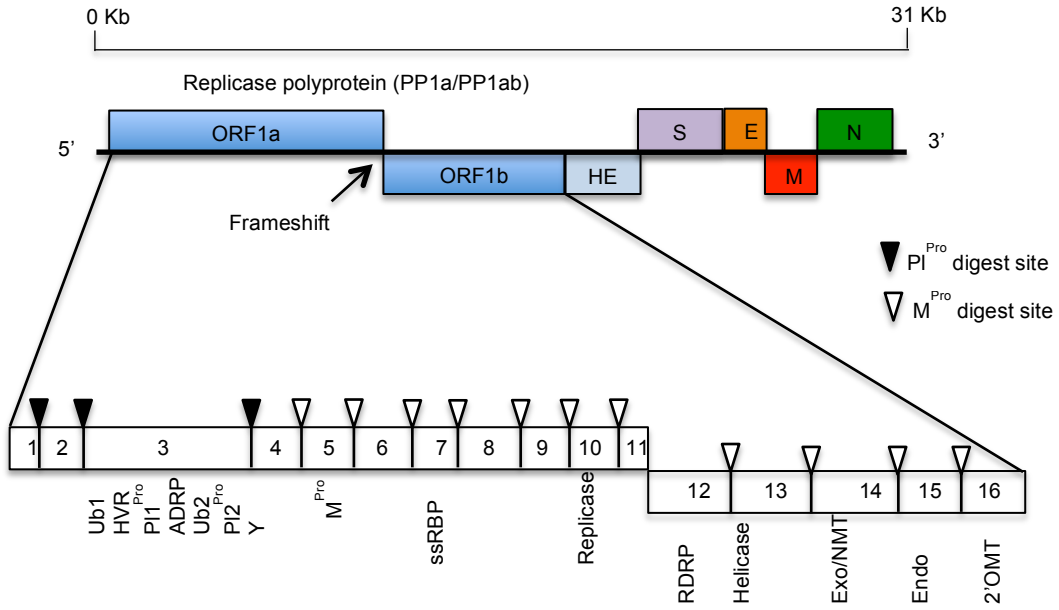

Figure 1.6. Coronavirus genome organization. ORF1a and ORF1b are located at the 5' two thirds and encode two polyproteins, namely pp1a and pp1ab which include 16 nonstructural proteins (nsp1 to nsp16) contain ubiquitin-like (Ub1), hypervariable region (HVR), papain-like protease ($PI1^{pro}$), ADP-ribose 1″ phosphatase (ADRP), papain-like protease (PI^{pro2}) Y-domain, 3C-like main proteinase (M^{Pro}), single-strand RNA binding protein (ssRDP), RNA-dependent RNA polymerase (RDRP), replicase, exonuclease (ExoN), N7 methyltransferase (NMT), endoribonuclease (Endo), 2′-*O*-methyltransferase (2′-*O*-MTase). The first 3 non-structural proteins are cleaved by viral papin-like proteinase(s) (PL^{pro}), the other 13 non-structural proteins are cleaved by the 3C-like main proteinase (M^{Pro}). The four structural proteins are coded by the remaining one third of the genome in coronaviruses and include the spike (S) protein, envelope (E) protein, membrane (M) protein, nuclepcapsid (N) protein and HE (hemagglutinin) is found on some betacoronaviruses.

1.5.1 Non structural proteins

The nonstructural proteins required for the replication/transcription complex (RTC) are encoded by the replicase gene, which is expressed from two overlapping ORFs. The first three non structural proteins in ORF1a are released by cleavage by a viral papin-like proteinase(s) (PI^{pro}) while the 3C-like main proteinase (M^{pro}) cleaves 11 sites in polyprotein 1ab to release 13 non-structural proteins (Denison *et al*., 1992; Namy *et al*., 2006; Prentice *et al*., 2004; Snijder *et al*., 2003; Ziebuhr *et al*., 2000). The second ORF, 1b, is translated via a ribosomal frameshift signal

which is stimulated by a slippery sequence (UUUAAAC) and a pseudoknot structure, resulting in higher levels of the truncated ORF1a product relative to the larger protein 1ab which contains 5 additional non-structural proteins (nsp12-nsp16) (Subissi *et al*., 2014). ORF1a includes the viral proteinase, anti-host activities, membrane-anchoring domains, RNA-binding and RNA-modifying activities while ORF1b encodes enzymes required in RNA replication, transcription, proofreading and capping (Decroly *et al*., 2008).

The first region (ORF1a) of the replicase gene encodes non-structural proteins (nsp) 1-11. The first mature protein released is nsp1. This protein is ~110 residues in betacoronaviruses and can be used as a group specific marker due to it's high sequence variability and the absence of nsp1 in both gammacoronaviruses and deltacoronaviruses. Nsp1 has been reported to inhibit host gene expression most likely by promoting host mRNA degradation as described for SARS-CoV (Kamitani *et al*., 2006; Narayanan *et al*., 2008; Züst *et al*., 2007), it also prevents type I interferon production in infected cells (Wathelet *et al*., 2007). The nsp2 possesses a highly divergent sequence across the coronaviruses and it's function remains unknown. Nsp3 is the largest coronavirus-encoded protein. It is a glycosylated, transmembrane, multidomain protein that interacts with several proteins involved in the replication and transcription processes (Barretto *et al*., 2005; Imbert *et al*., 2008; Kanjanahaluethai *et al*., 2007; Neuman *et al*., 2008; von Brunn *et al.*, 2007). Nsp3 proteins in coronavirus contain 10-16 domains (Neuman, 2016). Some of the nsp3 domains are duplicated and are conserved in all coronaviruses. The N-terminal region of nsp3 contains ubiquitin (Ub1), a hypervariable region (HVR) and a papain-like protease (PLP) domain $PL1^{Pro}$ (Neuman, 2016). In MHV, Ub1 initiates viral RNA synthesis by interacting with the N protein (Hurst *et al*., 2013; Hurst-Hess

et al., 2015) while $PL1^{Pro}$ is responsible for the 3 cleavage sites at the N-terminal sites in the replicase to release nsp1, nsp2 and nsp3 respectively. However, there are different mechanisms of processing in different coronaviruses (Gadlage and Denison, 2010; Ziebuhr *et al*., 2007). In some cases one PLP is sufficient while in others two active PLP may have specialized or overlapping functions (Baker *et al*., 1993; Bonilla *et al*., 1997; Graham and Denison, 2006). The ADRP like domain is a conserved domain in coronavirus and is associated with proteins involved in ADP-ribosylation or poly (ADP-ribose) polymerization and ATP-dependent chromatin re-modeling (Egloff *et al*., 2004; Saikatendu *et al*., 2005). Many coronaviruses encode two papain-like proteases, however SARS-CoV has only a single copy of a papain-like cysteine protease ($PL2^{pro}$), which cleaves polyprotein 1a at three sites to release the three non-structural proteins (Harcourt *et al*., 2004; Thiel *et al*., 2003). The Y domain at the C-terminus of nsp3 is highly conserved in all coronaviruses, but the function of this domain is unclear (Neuman *et al*., 2008). Nsp4 is about 500 amino acids in length, contains four transmembrane helices and a carboxy-terminal domain (Oostra *et al*., 2008). This protein is fundamental for cytoplasmic membrane modification with the assistance of nsp3 and nsp6 (Angelini *et al.*, 2013). Nsp5 is the 3C-like main protease (M^{pro}) which has a three-domain structure that mediates maturation cleavages of nsp4 to nsp16 (Perlman and Netland, 2009; Ziebuhr *et al*., 2000). Nsp6 is involved in the activation of autophagy which induces vesicles that contain Atg5 and LC3-II (Cottam *et al*., 2011). LC3 is present in the cell in a cytoplasmic form (LC3-I) that is converted into an active lipidated form (LC3-II) by specific covalent linkage upon autophagy induction (Mizushima *et al*., 2004). The conversion of LC3-I into LC3-II requires several proteins including Atg5 (Mizushima *et al*., 2001; Yoshimori and Noda, 2008) and Atg7 (Komatsu *et al*.,

2005). Furthermore, nsp6 interacts with other proteins such as nsp2, nsp3, nsp4, nsp8 and nsp9 (Krogh *et al*., 2001; von Brunn *et al*., 2007). In SARS-CoV, the crystal structure of the nsp7 (12kDa) and nsp8 (22-kDa) complex revealed a hollow, cylindrical hexadecamer composed of eight copies of nsp7 and eight copies of nsp8. This complex might provide a platform which improves the processivity of RNA synthesis by nsp12 and the increased binding of nsp12 to RNA (Zhai *et al.*, 2005). However the role of nsp8 is not yet clear despite some studies indicating that nsp8 may act as an RNA primase for nsp12 in SARS-CoV, since it polymerizes small oligomers via it's C-terminal domain, similar to those that can bind the palm subdomain of RNA dependent RNA polymerase (RdRp) (Subissi *et al*., 2014). Nsp9 is possibly involved in binding single stranded RNA as part of the viral replication complex as well as other proteins required for this complex (Sutton *et al*., 2004).

Nsp10 represents a novel fold that consists of a pair of antiparallel N-terminal helices, ß-strand, a loop at C-terminus, and two zinc fingers (Krishna *et al*., 2003). Studies proposed that nsp10 might play important roles in the synthesis of the viral RNA and in polyprotein processing through interaction with the nsp5 (Donaldson *et al*., 2007). Recently, studies proposed that it may act as a co-factor for nsp16 2`-*O*-Meythyltransferase activity for the regulation of viral RNA capping (Bouvet *et al*., 2010) and/or enhancing ExoN activity by interaction with nsp14 (Bouvet *et al*., 2012; Chen *et al*., 2011; Decroly *et al*., 2011). Nsp10 is followed by nsp11, a short peptide of highly variable sequence that is not conserved in different coronavirus genomes (Neuman *et al*., 2014). Nsp11 of SARS-CoV is attached to nsp10 but appears not to cause a significant difference in the core nsp10 structure (Su *et al*., 2006) and there is no indication of nsp11 interaction with nsp10 or any other

protein.

The second region of gene1 (ORF1b) encodes non-structural proteins 12-16. Nsp12 is about 102 kDa (932 amino acid residues), and is the most conserved protein in coronavirus, produced by M^{pro} (nsp5) action to result in pp1ab cleavage. The C-terminal domain of nsp12 contains the canonical RNA-dependent RNA polymerase (RdRp) motifs. RdRp is the core catalytic subunit that synthesizes negative strand RNA, new genome molecules and subgenomic messenger RNAs (mRNA) in many groups of RNA viruses (Ahlquist *et al*., 2003; Miller and Koev, 2000). Nsp12 therefore plays a key role in the viral replication/transcription complex. This enzyme carries a conserved active site (Ser-Asp-Asp - motif C) that is conserved in all *nidovirales*. Another motif, G, has a SXGXP conserved sequence and is consistently followed by a conserved basic residue (Gorbalenya *et al*., 2002). Motif G in SARS-CoV RdRp has also been shown to initiate RNA synthesis in a primer-dependent manner (te Velthuis *et al*., 2010). Several studies have shown a validated direct interaction between nsp8 and nsp12 (Imbert *et al.*, 2006; von Brunn *et al.*, 2007), while new studies indicate that nsp7/nsp8 complexes increase binding of nsp12 to RNA as a result of a major increase in the number of nucleotides synthesized per binding event. Nsp8 carries a second, non-canonical RdRp activity (Imbert *et al*., 2006; te Velthuis *et al*., 2012; Xiao *et al*., 2012). The ability of nsp8 to polymerize small oligomers in a sequence-specific manner suggests it acts as an RNA primase for nsp12. The interaction of RdRp with other viral proteins could be either directly or indirectly. Taken together, these data suggest nsp12 as a good candidate for the development of an antiviral drug. Nsp13 (66.5kDa) is a multi-functional protein that contains a zinc-binding domain at the amino-terminus and a helicase at the carboxy-terminus (Gorbalenya *et al*.,

1989) and unwinds both dsDNA and dsRNA in a 5' to 3' direction (Ivanov and Ziebuhr, 2004; Ivanov *et al*., 2004a; Lee *et al*., 2010; Tanner *et al*., 2003). This protein is essential for viral replication (Fang *et al*., 2007) and the presence of nsp12 stimulates helicase activity (2-fold) through the direct interaction of both proteins (Adedeji *et al*., 2012). Nsp13 has been also shown to exhibit RNA triphosphatase (TPase) activity *in vitro* and may be involved in the RNA capping reaction (Ivanov and Ziebuhr, 2004; Ivanov *et al*., 2004b), however, this proposed role requires further experimental evidence. Nsp14 is a N-terminal 3'-5'exonuclease domain (ExoN) which is unique to nidoviruses (Minskaia *et al*., 2006). This enzyme is capable of hydrolyzing single-stranded and double-stranded RNAs to final products of 8-12 and 5-7 nucleotides respectively and acts as a proofreading system that is involved in improving the fidelity of the large coronavirus genome during replication (Lauber *et al*., 2013). Furthermore, the presence of nsp10 was found to stimulate the activity of nsp14 *in vitro* by more than 35-fold (Bouvet *et al*., 2012) in a reaction which appears to be dsRNA-specific and is able to excise one 3´ mismatched nucleotide, mimicking a polymerase-mediated misincorporation product, which strongly indicates a role for the 5'-exonuclease activity in RNA synthesis proofreading. The nsp15 is a uridine-specific endoribonuclease (Snijder *et al*., 2003) and forms a hexamer (Guarino *et al*., 2005; Ricagno *et al*., 2006) with the active site at the C-terminus. Neither the exact function of nsp15 in viral replication nor the stimulating effect of Mn^{2+} on nsp15 activity is well understood. Finally, nsp16 is a 2´-O-methyltransferase (Chen *et al.*, 2011; Decroly *et al*., 2008) that plays a key role in the coronavirus life cycle by preventing virus detection by the cell innate sensing mechanisms. In MERS-CoV, 2′O-methyltransferase is an *S*-adenosyl-L-methionine (SAM)-dependent 2′-*O*-

methyltransferase (2′-O-MTase) which is stimulated by nsp10 which acts as an allosteric activator of the nsp16 2′-*O*-methyltransferase activity (Aouadi *et al*., 2017a). This protein is expected to be involved in the final step in cap synthesis by adding the final methyl group to complete the cap structure (Bouvet *et al*., 2010; Chen *et al*., 2011). The cap structure is a distinct feature of eukaryotic mRNAs, essential for it's translation and stability (Cougot *et al*., 2004; Furuichi and Shatkin, 2000; Lewis and Izaurflde, 1997; Schwer *et al*., 1998). Uncapped virus RNAs may be detected as 'non-self' by the host cell, triggering an antiviral innate immune response through the production of interferons (Züst *et al*., 2011). Therefore, many viruses that replicate in the cytoplasm of eukaryotes have evolved the means to mimic host mRNA by modifying their mRNA through capping. Capping involves the sequential activity of three enzymes. First, an RNA triphosphatase (TPase) will remove the γ- phosphate group from the 5'-triphosphate end (pppN) of the nascent mRNA chain to generate the diphosphate 5'-ppN. The second step, a RNA guanylyltransferase (GTase) transfers a GMP to the 5'-diphosphate end to yield the cap core structure (GpppN). The cap is then methylated at the N-7 position of it's guanine by a N7-methyltransferase (MTase) to produce a cap-0 structure (m7GpppN) (Furuichi and Shatkin, 2000).

1.5.2 Structural proteins

1.5.2.1 Spike protein (S)

The first step in viral infection of host cells is receptor recognition (Baranowski *et al*., 2001). Coronavirus entry into host cells is mediated by the envelope-anchored spike protein, first by binding to a specific receptor on the host cell surface and then by fusing viral and host membranes (Bosch *et al*., 2003; Spaan *et al*., 1988).

The spike protein is the largest of the coronavirus structural proteins (Dye and Siddell, 2005) and is a member of the class 1 fusion proteins (Eckert and Kim, 2001; Harrison, 2008; Skehel and Wiley, 2000; Wilson *et al*., 1981). The spike varies from 1,160 to 1,450 amino acids in length and is heavily glycosylated with 21 to 35 N-glycosylation sites (Belouzard *et al*., 2012). The spike protein is a trimer located at the surface of the virion and gives the distinctive corona shape recognized by EM (Xu *et al*., 2004), consists of three segments; a large ectodomain, a single-pass transmembrane anchor, and a short intracellular tail (Beniac *et al*., 2006; Li *et al*., 2006). The ectodomain of coronavirus S protein can be further divided into two domains, binding S1 subunit (variable domain) and a membrane-fusion S2 subunit (more conserved domain) **Figure 1.7**.

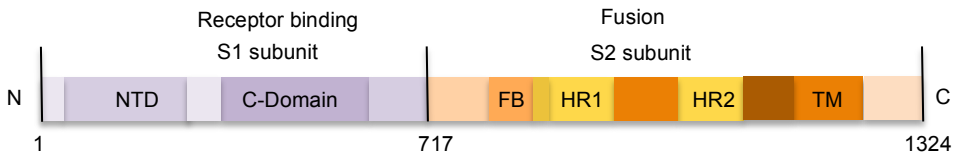

Figure 1.7. Schematic of the S protein domains. Domain structure of MHV spike protein. NTD: N-terminal domain; FB: fusion protein; HR: Heptad repeat; TM: transmembrane domain.

The globular S1 gives the virus it's crown-like appearance and is exposed on the outside the virus. It contains the receptor binding domain (RBD) within the first 330 amino acids at the amino terminus that binds to a variety of proteins and sugars, responsible for cellular attachment and therefore cell tropism. In coronaviruses, the interaction between the RBD and it's receptor is one of the most important determinants of host range and cross-species infection (Li, 2013; Li *et al*., 2006). Furthermore, RBDs contain major neutralization epitopes, which induce most of the host immune response and may serve as subunits for vaccine development

against coronavirus infections (Du *et al*., 2014, 2009; He *et al*., 2005; Jiang *et al.*, 2014; Sui *et al.*, 2004; Ying *et al*., 2014). The S2 subunit forms the stalk region of the spike protein that is anchored in the membrane and contains two heptad repeats HR1 and HR2, a typical feature of class I viral fusion proteins, a putative fusion peptide (FP) and one short transmembrane domain (Chambers *et al*., 1990). The S2 domain mediates fusion of the viral and host membranes. In addition, cell-cell fusion is activated by the expressed viral fusion proteins to form giant multinucleated cell named syncytia (Belouzard *et al*., 2012; Bosch *et al*., 2003; Luo and Weiss, 1998). Some betacoronavirus and gammacoronavirus groups also posses a small hemagglutinin-esterase (HE) protein, standing 5-7 nm on the envelope along with the S protein (Kienzle *et al*., 1990). The HE protein contains a disulfide link to form a homodimer with both hemagglutinating and esterase activity (Brian *et al*., 1995; Kienzle *et al*., 1990). The HE protein is believed to have been gained from Influenza C virus, where it is the sole glycoprotein, as a result of recombination with a coronavirus ancestor (Vlasak *et al*., 1988). This protein may play a role in the entry and/or release of the virus. However, it doesn't appear to be essential for viral replication or virulence (Kazi *et al*., 2005; Popova and Zhang, 2002).

Since the S protein plays an essential role in virus entry and determines tissue and cell tropism as well as host range (Lu *et al*., 2015), this is deemed an excellent target for the development of vaccines and antiviral drugs. The S protein proves an important target for T cell responses and epitopes located in it's N- terminal portion trigger the production of virus-neutralizing antibodies. The coronavirus S protein, when inoculated alone, can induce protective immunity for a number of viruses (Cavanagh, 2005). For example, studies on SARS-CoV revealed that vaccines

based on the S protein could induce antibodies to block virus binding and fusion to neutralize virus infection (Du *et al*., 2009). For therapeutic targeting, the S protein RBD or S2 regions so far appear to be the most investigated targets to identify various specific antivirals (Du *et al*., 2014; Gao *et al*., 2013; Jiang *et al*., 2014; Li *et al*., 2015; Lu *et al*., 2014; Pascal *et al*., 2015).

1.5.2.2 Membrane protein (M)

The membrane protein (M) is about 25-30 kDa in size (221-262 amino acids) and is the most abundant protein embedded within the coronavirus envelope. M contains a short amino terminus located outside the virion, followed by three transmembrane domains (Tm), and a large carboxy-terminal domain that is usually located inside the virion (Hogue and Machamer, 2008) **Figure 1.8**.

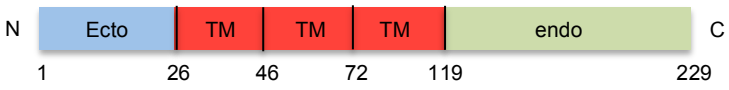

Figure 1.8. Coronavirus M protein structure. A schematic representation of MHV-A59 M protein domains. Ecto: ectodomain; TM: transmembrane domain; endo: endodomain.

As yet, studies have revealed limited structural information for the M protein, it's transmembrane nature making it difficult to express and purify. The protein is moderately well conserved within each group but completely divergent across the three groups (Den Boon *et al*., 1991). M proteins of coronavirus exclusively localize to the ER/Golgi area with the exception of SARS-CoV, TGEV and Feline Infectious Peritonitis virus, where it is capable of reaching the plasma membrane (Jacobse-Geels and Horzinek, 1983; Krijnse-Locker *et al*., 1994; Laviada *et al*., 1990; To *et*

al., 1991; Tooze and Tooze, 1985; Tooze *et al*., 1984; Voß *et al*., 2006). Alphacoronavirus and gammacoronavirus M proteins are believed to be O-glycosylated, while M is N-glycosylated in betacoronavirus (de Haan *et al*., 2003). Glycosylation may play a role in virus-host interaction but does not appear to be important for virus assembly or infectivity (de Haan *et al*., 2003). M protein is involved directly in viral assembly and budding in addition to it's functions in host interaction, along with E, S and N (de Haan *et al*., 1999; Haan *et al*., 1998; Neuman *et al*., 2011). Studies showed that the endodomain is the locus for M-N (Escors *et al*., 2001; Fang *et al*., 2005; Hurst *et al*., 2005; Kuo and Masters, 2002; Luo *et al*., 2006a; Verma *et al*., 2007, 2006) and S-M interaction (de Haan *et al*., 1999; McBride and Machamer, 2010). Cryo-electron micrographic (cryo-EM) and tomographic reconstructions (Neuman *et al*., 2011) and inferences drawn from a genetic study of evolved M mutants (Kuo and Masters, 2010) indicate that M-M monomer interactions occur via the transmembrane (Tm) domains, whilst higher-order oligomerization of M dimers is predominatly driven by the endodomain. In innate antiviral responses, type I interferons (IFNs) are the prime effector cytokines. IFN production is induced by pathogen-associated molecular patterns such as viral double-stranded RNA, which are sensed by host pattern recognition receptors. M protein is postulated to suppress type I IFN production potently by preventing the formation of a functional TRAF3–TANK–TBK1/IKKe complex which signals IFN gene expression downstream of interaction of virus RNA with innate sensors such as RIG-I or MDA-5 (Siu *et al*., 2009). A study suggested that interaction between the domain N3 of N protein and endodomain of M protein is involved in the gRNA packaging process (Kuo *et al*., 2016). For these reasons, the M protein could be an attractive target for development as an antiviral.

1.5.2.3 N protein (N)

N protein (N) protein is one of the most abundant coronavirus structural proteins, located inside the virus particle. The N protein's primary function is to enclose the viral RNA genome in a helically symmetric ribonucleoprotein (RNP) (Masters and Sturman, 1990; Masters *et al*., 1990). The RNP complex is important to maintain the RNA genome in an ordered conformation for replication and transcription. N protein is a helical, highly basic and phosphorylated protein, about 50-60 kDa in size. Based on MHV strain sequence comparisons, the N protein has three conserved domains with a high serine content (7-11%) (Tan *et al*., 2006). The N protein consists of two structural domains: N-terminal domain (NTD) and C-terminal domain (CTD) linked by a poorly structured linkage region (Linker) containing a serine/arginine- rich (S/R) domain (SRD) (Chang *et al*., 2006; Q. Huang *et al*., 2004; Luo *et al*., 2006a, 2005) **Figure 1.9**. The NTD domain of the N protein in coronaviruses is involved in RNA binding while the CTD domain is involved in RNA binding and self-association of the protein to form higher-order oligomers (Chang *et al*., 2014, 2013; Lo *et al*., 2013; Ma *et al*., 2010; Yu *et al*., 2006). A critical determinant for recognition of the genomic RNA packaging signal has been mapped to the NTD of the N protein in MHV (Kuo *et al*., 2014). A domain called N3 can be found at the carboxy terminus of the N protein and has been reported in many studies, but not all, to be a locus for N-M interaction (Fang *et al*., 2006; Hatakeyama *et al*., 2008; He *et al*., 2004; Hurst *et al*., 2005; Kuo and Masters, 2002; Luo *et al*., 2006b; Verma *et al*., 2007, 2006). In addition to it's primary function in packaging the RNA genome in a helical nucleocapsid structure during the encapsidation process, N protein plays a role in viral replication and discontinuous transcription (Baric *et al*., 1988; Compton *et al*., 1987). The N protein

of MHV and SARS-CoV posseses chaperone activity (Thiel *et al*., 2001; Zúñiga *et al*., 2007) that appears important during template switching events. Furthermore, the N protein plays a structural role in virus assembly (Hurst *et al*., 2005).

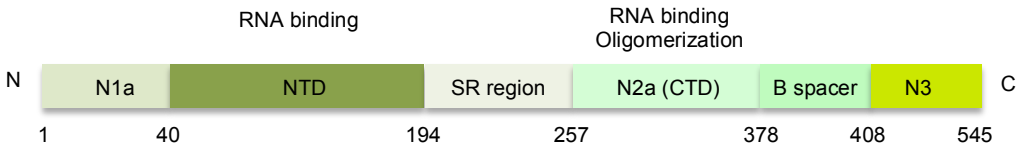

Figure 1.9. Domain organization of coronavirus nucleocapsid protein. Schematic of the coronavirus N protein divided into domains: N1a, NTD: N-terminal RNA-binding domain, SR: serine/ arginine region, CTD: C-terminal dimerization domain, N3: N3 domain.

Development of antiviral drugs traditionally targets enzymatic proteins such as proteases, polymerases and helicases. However, the multifunctional and crucial role of N protein during the viral life cycle could make it an attractive target for antiviral design. In addition, detailed high-resolution structural information about N proteins from various coronavirus species provides a good starting point for structure-based drug discovery, while most replicase protein structures are still unknown. For instance, mutations in the center of the NTD domain in HCoV-OC43 significantly decreased the RNA-binding affinity of the N protein and subsequently decreased viral replication which is consistent with it being considered a target for the development of RNA-binding inhibitors (Lin *et al.*, 2014). Removal of 40 amino acids from the C-terminal tail of the SARS-CoV N protein decreased protein oligomerization (Luo *et al*., 2006a). N protein has been also considered an eligible target for DNA or recombinant-protein-mediated vaccination. N protein does not elicit neutralizing antibodies because it resides inside the virus particle so the goal of N based vaccines is to induce the generation of cytotoxic T lymphocytes (CTLs) capable of destroying infected cells. Studies reported that SARS-CoV N protein

could induce high CTL activity when introduced into mice (Zhao *et al.*, 2005; Zhu *et al*., 2004).

1.5.2.4 Envelope protein (E)

The envelope protein (E) is the smallest structural protein. Envelope protein is a small monomeric, non-glycosylated protein of 76-109 amino acids with a single hydrophobic domain (HD). This protein is usually encoded either by the second or third ORF in a bi- or tri cistronic mRNA (Boursnell *et al*., 1985; Budzilowicz and Weiss, 1987) and plays an important role in viral assembly. However, recent studies indicate that E protein is also essential for efficient trafficking of the virions through the secretory pathway, which may be due to an ion channel activity (Nieto-Torres *et al*., 2011; Ruch and Machamer, 2012). The expression of E protein only or together with M protein can induce virus like particle synthesis (VLP) but some reports indicate that M and N (Y. Huang *et al*., 2004), M and E (Hsieh *et al*., 2005) or only M protein are able to drive the production of released vesicles (Tseng *et al*., 2010), all effectively enabling study of virus-like particle production in different cell types and expression systems. E protein is not the best target for development as an antiviral as it is the smallest and least abundant of the coronavirus structural proteins. It is also poorly conserved across the coronavirus genera (Masters, 2006) and lacks confirmed direct interactions with the other structural proteins.

1.6 Antiviral drugs

Since the SARS-CoV outbreak and it's worldwide spread, intensified research efforts targeting the coronavirus group have resulted in the discovery of two new human coronaviruses, and most recently the discovery of MERS-CoV. However, high throughput screening is still ongoing for the discovery of antivirals targeting

coronavirus proteins. In fact, most known inhibitors are focusing on the main antiviral drug targets and are designed on the basis of the structural knowledge of these targets. Based on their broad mechanism of action, antiviral therapeutics can be divided into three groups: virucides, biological response modifiers, or direct acting antivirals. The first group includes chemical or physical agents that are capable of physically inactivating a virus. However, owing to toxicity problems, agents in this group are used primarily as disinfectants for inanimate objects. The second group is a diverse group of agents with the common property of modulating the host immune response. They include substances produced naturally in the body, like cytokines and synthetically produced compounds (Ford, 1986), or non-endogenous biological components such as bacterial cell wall extracts. Drugs in this group can have an indirect antiviral effect by stimulating the host's innate or adaptive immune response (Bergman *et al*., 2011). The third group is the direct acting antivirals, which is the largest and most important group of antiviral therapeutics. Drugs in this group exert their antiviral effect by targeting essential viral or cellular factors involved in replication. As mentioned previously, coronaviruses have the largest RNA genome known and encode a large number of proteins that are involved in viral replication and assembly - about 30 different proteins for each virus. Each protein has a specific function or functions. Most of the viral proteins are associated with other proteins or with themselves to carry out their functions and the interaction between viral proteins plays a crucial role during the viral infection cycle. Due to the large genome and high diversity among coronaviruses, there are no effective structure based pan-coronavirus drug inhibitors. Even if there were, all designed drugs would need further optimization and validation before they could be approved (Barnard and Kumaki, 2011). For

example, despite the fact that SARS-CoV emerged over a decade ago and over 3500 publications have been published on it since 2002, the FDA has not yet approved an antiviral agent for the treatment of SARS-CoV infection. Understanding the etiology, pathology, and possible therapeutic targets of coronaviruses will improve the development of an effective antiviral, which may be important for future outbreak control.

There are many challenges in the development of anti-coronavirus therapeutics. One of the challenges is the lack of natural infection models that makes it hard to evaluate any proposed drug efficiency. Another challenge is to make the availability of any broad-spectrum anti-coronavirus drugs and/or vaccines beneficial if a new coronavirus with pandemic potential emerges at any time. This complexity makes the development of any antiviral a difficult process. On the other hand, many of the proteins involved directly in the virus life cycle have been shown to be valid targets for antivirals (Mielech *et al*., 2014).

There are many factors to bear in mind before considering any coronavirus protein as a potential target for rational drug design. Firstly, the specificity of the protein(s) (i.e. non-existence of a similar cellular target) is one of the most important requirements because a similar cellular target could be affected by any antiviral drug and cause serious side effects. Fortunately, most viral enzymes are unique in their folding, organization and mechanisms of action and that provides a large space for drug design and drug selectivity. Secondly, the potency of the expected outcome of viral target inhibition is another important parameter. Lastly, an amplifying mechanism means the number of events that the protein is involved in during the virus replication cycle. For example, RNA capping events can vary from a single capping event to many RNA capping events in *nidovirales*. In contrast,

RdRp enzymatic action incorporates several thousand nucleotides to produce a single RNA genome and inhibition of the viral RdRp at each nucleotide incorporation step should exhibit a powerful antiviral effect. As noted the large CoV genome contains two open reading frames, connected by a ribosomal frame shift that encodes two large overlapping replicase polyproteins from which the functional proteins are produced by extensive proteolytic processes (Dougherty and Semler, 1993; Pillaiyar *et al*., 2016; Ziebuhr *et al*., 1997). Coronaviruses utilize from one to three proteases for such proteolytic processing (Harcourt *et al*., 2004; Hilgenfeld, 2014). These enzymes are indispensable to the viral replication and infection process, making them an attractive target for the development of an antiviral drug. Targeting the 2′ O-MTase activity and corresponding immune responses has been suggested as an approach to ablate the capping process in a variety of coronavirus. Cap formation is an important post-transcriptional process in coronavirus RNA synthesis to ensure that the viral RNAs can be translated by host ribosomes as well as being indistinguishable from host mRNA (Menachery *et al*., 2014). The strategy described for development of coronavirus antivirals drug so far therefore has good starting points based on sound experimental data. However, to be prepared for future zoonotic transmissions of coronaviruses into the human population, or in the case of new adpated coronavirus outbreaks, more effort is required to explore new targets and to develop a novel strategy for drug design against all coronaviruses.

Mouse hepatitis virus (MHV) is commonly used as a model system to study the replication and pathogenesis of coronaviruses (Weiss and Leibowitz, 2011) for several reasons. MHV is easy to grow and maintain. MHV causes hepatic and central nervous system diseases of varying severity depending on the strain and

therefore is used as a model for hepatitis, viral encephalitis, and demyelination (Weiss and Navas-Martin, 2005). MHV is used as a model to study the interaction of coronaviruses with the alpha/beta interferon response since it induces delayed IFN response (Roth-Cross *et al*., 2007) which help in studying general immune response. Another reason is that N protein of MHV is interchangeable with N protein of counterpart from the closely related bovine coronavirus, which help to study infectivity of coronavirus genomic RNA (Hurst *et al*., 2010). Both MHV and SARS-CoV N proteins share the same IFN-β antagonizing mechanism by having the ability to disorganize the function of cellular protein activator of protein kinase R (PACT) to abolish the innate antiviral response. However, this strategy does not appear to be used by all coronavirus N proteins (Ding *et al*., 2017).

1.7 Therapies against coronavirus infection

Therapeutic options for coronavirus infections can be divided to vaccine and antiviral drug development. Several strategies have been developed to produce effective vaccines against SARS-CoV infection such as recombinant vectored SARS-CoV S protein, DNA vaccines, inactivated whole-virus vaccines and recombinant-protein vaccines (Gillim-Ross and Subbarao, 2006). However there is still no progress towards a credible SARS vaccine and further work is required for them to be evaluated for safety profile and treatment effects in patients. In the field of developing antiviral drugs, there are many difficulties in developing antiviral drugs against coronavirus infection. Three general approaches are used to discover potential anti-CoV treatment options for human-pathogenic CoVs. The first approach is to test existing broad-spectrum antiviral drugs, which have been used to treat infection of other viruses by using standard assays to test effects of these drugs on virus yield and plaque formation of live and/or pseudotyped coronaviruses

such as interferon alfa, interferon beta, interferon gamma, ribavirin and cyclophilin inhibitors (Chan *et al.*, 2013; Cheng *et al*., 2013, 2007; Cinatl *et al.*, 2003; de Wilde *et al*., 2013; Falzarano *et al*., 2013; Pfefferle *et al*., 2011; Tanaka *et al*., 2013). Although these drugs are available with known pharmacokinetic and pharmacodynamic properties, side effects and dosing regimens they do not have specific anti-CoV effects and may be associated with severe adverse effects. For example, the HIV-protease inhibitor lopinavir, often combined with ritonavir, appeared to show some benefit for SARS patients (Chu *et al*., 2004), and the effect of these compounds was also observed in cell culture. However the coronavirus genome does not code for an aspartic protease related to the HIV protease (Wu *et al*., 2004) but codes for a cysteine protease (3C-like protease). More information on the proposed binding mode of lopinavir to the SARS-CoV main protease and attempts to improve it's inhibitory potency might be a good starting point for anti-SARS drug design. The second approach involves the screening in chemical libraries that is compromised from large numbers of existing compounds or databases that contain information on transcriptional signatures in different cell lines (Chan *et al*., 2013; de Wilde *et al*., 2014; Dyall *et al*., 2014; Elshabrawy *et al*., 2014; Kindrachuk *et al*., 2015; Liu *et al*., 2015). This approach provides rapid, high-throughput screening of many readily available compounds that can be further evaluated by anti- viral assays. However, most of these drugs are not clinically useful because they are either associated with immunosuppressive effects or they have high anti-CoV half-maximal effective concentrations (EC_{50}) (J. F.-W. Chan *et al*., 2015; Chan *et al*., 2003; Chu *et al*., 2004). The third approach involves the de novo development of novel, specific agents based on the genomic and biophysical understanding of the individual coronaviruses such as small interfering RNA

(siRNA) molecules or inhibitors against specific viral enzymes involved in the viral replication cycle, monoclonal antibodies which target the host receptor, inhibitors of host cellular proteases, inhibitors of virus endocytosis by the host cell, human or humanized monoclonal antibodies that target the S1 subunit RBD and antiviral peptides that target the S2 subunit (Lu *et al*., 2014).

The main drug targets among the viral nonstructural proteins are the main protease, the RdRp (Nsp12) and the helicase (Nsp13). Many inhibitors against coronavirus proteins have been designed on the basis of crystal structures. There are many inhibitors that have been designed and synthesized targeting the coronavirus M^{pro}, but few of them have undergone systematic toxicity and other preclinical studies and are still yet not available for clinical trials (Xue *et al*., 2008). Numerous SARS-CoV PL^{Pro} inhibitors belonging to different classes have been identified such as small-molecule inhibitors, thiopurine compounds, natural products, zinc ion and zinc conjugate inhibitors and naphthalene inhibitors (Báez-Santos *et al*., 2015) by using high throughput screening and structure-based rational design (Hilgenfeld and Peiris, 2013; Kumar *et al*., 2013; Kuo and Liang, 2015; Pillaiyar *et al*., 2016; Tong, 2009; Zhao *et al*., 2013) but some of these drugs only inhibit the enzymatic activities of PL^{pro}, not viral replication, or vice versa (Báez-Santos *et al*., 2015; Báez-Santos *et al*., 2014; Ratia *et al*., 2008) and none have been validated in animal or human studies (Báez-Santos *et al.*, 2015; Báez-Santos *et al*., 2014). Further animal studies needs to be conducted for developing one of these potent inhibitors into an antiviral drug. Another target is RNA-dependent RNA polymerase (RdRp) and helicase, which are required for the transcription and replication of the virus (J. F. Chan *et al*., 2015; van Boheemen *et al*., 2012). However, obtaining an active form, of RdRp (Nsp12) and the helicase

(Nsp13) and many attempts to crystallize them has failed.

On the other hand, CoV structural proteins have also been the targets for antiviral development. The spike protein is of particular interest for antiviral development because the spike protein is a major immunogenic antigen and is essential for the interaction between the virus and the host cell receptor, which make it a perfect target for vaccine and antiviral development. Several monoclonal antibodies have been developed which target S1, S2 and the RBD. Most of these monoclonal antibodies target specific epitopes on the S1 subunit RBD to inhibit virus–cell receptor binding, whereas others bind to the S2 subunit to interrupt virus–cell fusion (Coughlin and Prabhakar, 2012). Another strategy is targeting different regions of S by antiviral peptides. For example, antiviral peptides analogous to the N terminus, pre-transmembrane domain or the loop region separating the HR1 and HR2 domains of SARS- CoV can inhibit virus plaque formation by 40–80% at micromolar concentrations (Sainz *et al*., 2006; Zheng *et al*., 2005).

The N protein of SARS-CoV has been widely used as a diagnostic target of SARS infection (Surjit and Lal, 2008). Viral N protein shows genetic stability with the least variation in the gene sequence, which is a primary requirement for an efficient drug target candidate. The N protein has also become a therapeutic target in antiviral therapy by disrupting RNP formation through inhibition of either protein oligomerization or nucleic acid binding activity in different viruses. For example, inhibition of influenza virus by targeting it's nucleocapsid protein through nucleozin and it's analogues (Hung *et al*., 2012; Kao *et al*., 2010), compounds targeting the interaction between N protein and nucleic acids have been developed against HIV-1 virus (Musah, 2004) and a peptide that interferes with the CTD oligomerization of

the HCoV-229E N protein and inhibits virus production have been also discovered (Lo *et al*., 2013).

E, M and some of the accessory proteins, which are essential for virion assembly, also have additional functions like suppressing the host immune response to facilitate viral replication, siRNAs were developed against E, M, ORF3a, ORF7a or ORF7b of SARS-CoV inhibited viral replication *in vitro* (AAkerström *et al*., 2007; He *et al*., 2009), none of these siRNAs is ready for human use until better delivery methods become available. Huge progress has been made in the elucidation of the functions and structures of coronavirus proteins especially SARS and MERS coronavirus and research on vaccine development has also progressed, with a number of strategies being developed and evaluated in experimental animal models but still more effort is needed to develop effective antivirals against current and re-emergence coronavirus infection.

1.8 Directed evolution

Protein engineering is one of the most popular methods used to improve the properties of enzymes or proteins by genetic changes. There is an intimate relationship between the amino acid sequence and the structure of a folded protein but the relationship between a protein structure and it's function is less well defined and understood. Therefore, manipulating or mutating the sequence of a protein can alter proteins function and properties, often in unforeseen ways. A successful and widely used example of protein design and engineering is the directed evolution method (Otten and Quax, 2005), which has become a powerful approach over the last two decades (Yuan *et al*., 2005). This method is inspired by the Darwinian concept of natural evolution by mimicking the process of natural selection on a

protein (Romero and Arnold, 2009; Skandalis *et al*., 1997). *In vitro* evolution is a laboratory method applied to evolve molecules with desired properties in a short time and has been used to optimize enzymes, improve drug resistance, and in the development of novel pharmaceuticals and vaccines (Arnold, 1998, 1996; Patten *et al*., 1997). Gene recombination and random mutagenesis are the two natural evolutionary processes, which have been adapted for *in vitro* evolution experiments. An *in vitro* evolution experiment firstly requires building a library of variants from the protein of interest followed by screening and/or selection of the protein products with desired properties. Libraries can be created *in vitro* either by site directed mutagenesis or by random mutagenesis; the latter of which is the most popular and widely used method. Random mutagenesis requires no structural or mechanistic information of the target and can uncover unexpected beneficial mutations. There are several methods to generate genetic diversity by random mutagenesis, by using chemical mutagens (Cox, 1976; Kadonaga and Knowles, 1985; Shortle and Nathans, 1978), passing cloned genes through a mutator strain (Cox, 1976; Greener *et al*., 1996) or by error-prone PCR mutagenesis (Cadwell and Joyce, 1992; Leung *et al*., 1989; Patrick and Firth, 2005; Vartanian *et al*., 1996). Error-prone PCR (ep-PCR) techniques are the most popular due to their low-cost and simplicity. Various methods are available to generate random mutations using ep-PCR such as using a low-fidelity DNA polymerase such as *Taq*, which lacks proofreading efficiency so that mismatched bases are not removed (Tindall and Kunkel, 1988). For higher rates of mutation, Mn^{2+} ions and/or changes in the Mg^{2+} ion concentration, along with unbalanced available nucleotides pools, can effectively increase the rate of mutation (Beckman *et al*., 1985; Cadwell and Joyce, 1992; El-Deiry *et al*., 1984). The degree of mutation can be directed by adjusting

the number of amplification cycles to allow easy manipulation of the final libaray complexity. In the next step, a selection process to be applied to a library of clones is required for any directed evolution experiment and again, several methods have been developed for screening and selection. For example, libraries of clones may be directly separated into individual cultures, expressed, and screened for the desired property or whole libraries may be plated and screen *en masse*. Overall a number of available tools required for gene variant generation by directed evolution should allow an experimenter to gain results of high impact with moderate effort. Directed evolution is commonly used in the field of protein engineering as an alternative to rationally designing modified proteins as well as studying fundamental evolutionary principles in a controlled, laboratory environment. For example, directed evolution-based strategies are most commonly used in protein engineering projects, to increase activity, improve stability or provide greater specificity to the desired protein, which is an alternative to rational design (Romero and Arnold, 2009). Computational design has been used in recent years as a promising tool for creating protein catalysts with tailored activities and specificities and by using directed evolution of a computationally designed enzyme, dramatic molecular changes can also drive the optimization of protein active sites with no need to understand the mechanism of the desired activity or how mutations would affect it (Giger *et al*., 2013). Another example is directed evolution may also be employed to 'tune' genetic circuits, or adjust regulatory elements within a stretch of DNA incorporating one or more genes (Cobb *et al*., 2013).

1.9 Aims of this research

The aim of this project is to develop small protein inhibitors that interfere with the coronavirus replication cycle as a novel mechanism to treat infections caused by CoVs. The first aim is to clone several proteins from MHV-A59 in vectors capable of protein expression in prokaryotic and eukaryotic cells. For that, protein constructs will be expressed in *E. coli* to confirm correct MHV protein production before transfection into a mammalian cell line (17clone-1) which is permissive for MHV. The second aim is to examine the effects of expressed non-mutated proteins on virus replication based on their ability to protect cells from virus-induced cytopathic effects as measured by plaque assay. The third aim is to create a library of variants using random mutagenesis via error prone PCR and to optimize the library to control the number of mutations generated and to assess the potential of the mutated MHV protein varaints as inhibitors of MHV replication. The fourth aim is to screen for dominant-negative N variants using cell survival following mammalian cell line (17clone-1) transfection with different libraries; variants from survivors will be extracted, re-cloned, and transfected again into 17clone-1 and this process will be repeated for several rounds. A last aim of this project is to optimize purification conditions for difficult to express coronavirus proteins such as nsp16 so that they too may be considered as targets. The overall goal of this is to contribute to the development a small protein inhibitors using, as exemplars, mutated coronavirus proteins that have been selected by directed *in vitro* evolution.

2 Materials and methods

2.1 Plasmid construction and cloning of desired DNA fragment

2.1.1 Primers

The oligonucleotides primers were designed using Gene runner and DNADynamo software using the sequence of Murine hepatitis virus strain A59 (GenBank accession number AY700211.1). Primers were purchased from Integrated DNA Technology (IDT). Forward and reverse primers were reconstituted as 50 µM solutions in nuclease-free water and stored at -20°C. All primers used in this study are listed below **in Tables 2.1– 2.5**. Red colour nucleotide in **Table 2.3** represents nucleotide change for mutation correction. In **Table 2.4**, red colour represents vector sequence.

Table 2.1. Oligonucleotides used for cloning in pTriEx1.1 plasmid. Highlight sequence represents start codon.

Primer	Sequence	bp
Fw_nsp3_PL^{Pro}	AGGAGATATACCATGGCCTCCGAGGCAGATCTTG	34
Rv_nsp3_ PL^{Pro}	GATGGTGGTGCTCGAGCGAAAAAGTTTGCTTTAAATTTT	39
Fw_nsp5	AGGAGATATACCATGGCCTCTGGTATAGTGAAGATG	36
Rv_nsp5	GATGGTGGTGCTCGAGCTGTAGCTTGACACCAGC	34
Fw_nsp6	AGGAGATATACCATGGCCTCAAAGCGCACAAGAG	34
Rv_nsp6	GATGGTGGTGCTCGAGTTGAATTTGAGATACTTCAATG	38
Fw_nsp7	AGGAGATATACCATGGCCTCAAGATTGACGGATGT	35
Rv_nsp7	GATGGTGGTGCTCGAGTTGCAAGACAGTATTGTCG	36
Fw_nsp8	AGGAGATATACCATGGCCGCCTTACAGAGTGAATT	35
Rv_nsp8	GATGGTGGTGCTCGAGCTGCAAAACAACAGTAGACA	36
Fw_nsp9	AGGAGATATACCATGGCCAACAATGAGTTGATGCC	35
Rv_nsp9	GATGGTGGTGCTCGAGCTGCAATCTCACTGTCGAG	35
Fw_nsp10	AGGAGATATACCATGGCCGCGGGTACGGCAACTG	34
Rv_nsp10	GATGGTGGTGCTCGAGGCCTACACAGGAACAGCT	34
Fw_nsp12C	AGGAGATATACCATGGCCAAGGACTTGCTTTTGTAT	36
Rv_nsp12C	GATGGTGGTGCTCGAGTTGCAGCACTGCACTTCTT	35
Fw_nsp12N	AGGAGATATACCATGGCCTCAAAAGACACGAACTTT	36
Rv_nsp12N	GATGGTGGTGCTCGAGAAGAGACAAGCGATAACGAT	36
Fw_nsp16	AGGAGATATACCATGGCCGCTGCTGCTGACTGGA	34
Rv_nsp16	GATGGTGGTGCTCGAGTTTGACATTTACTAGGCTATC	37
Fw_M	AGGAGATATACCATGGCCATGAGTAGTACTACTCAG	36
Rv_M	GATGGTGGTGCTCGAGGATTCTCAACAATGCGGTG	35
Fw_N	AGGAGATATACCATGGCCATGTCTTTTGTTCCTGG	35
Rv_N	GATGGTGGTGCTCGAGCACATTAGAGTCATCTTCTAA	37
Fw_Y	AGGAGATATACCATGGCCTGGAGTGCTCGTTTGT	34
Rv_Y	GATGGTGGTGCTCGAGGCCCCCTTTAAGAGAGAAC	35

Table 2.2. Sequencing primers

Primer	Sequence	bp
T7 Forward	TAATACGACTCACTATAGGG	20
TriExUP	GGTTATTGTGCTGTCTCATCA	21
TriExDOWN	TCGATCTCAGTGGTATTTGTG	21

Table 2.3. Primers used for correction of random mutation in nsp12_N-terminal. Red colour nucleotide represents nucleotide change for mutation correction.

Primer	Sequence	bp
Fw_nsp12N-new	CACGAACTTTTTAAACCGGATTCGGGGTACA	31
Rv_nsp12N-new	TGTACCCCGAATCCGGTTTAAAAAGTTCGTG	31

Table 2.4. Primers used for SUMO cloning in pTriEx1.1 plasmid. Red colour represents vector sequence.

Primer	Sequence	bp
Fw_nsp3_ PL^{Pro} _SUMO	CACAGAGAACAGATTGGTGGTTCCGAGGCAGATCTT	36
Rv_nsp3_ PL^{Pro} _SUMO	GATGGTGGTGCTCGAGCGAAAAAGTTTGCTTTAAATTTT	39
Fw_nsp5_SUMO	CACAGAGAACAGATTGGTGGTTCTGGTATAGTGAAG	36
Rv_nsp5_SUMO	GATGGTGGTGCTCGAGCTGTAGCTTGACACCAGC	34
Fw_nsp6_SUMO	CACAGAGAACAGATTGGTGGTTCAAAGCGCACAAGA	36
Rv_nsp6_SUMO	GATGGTGGTGCTCGAGTTGAATTTGAGATACTTCAATG	38
Fw_nsp7_SUMO	CACAGAGAACAGATTGGTGGTTCAAGATTGACGGAT	36
Rv_nsp7_SUMO	GATGGTGGTGCTCGAGTTGCAAGACAGTATTGTCG	36
Fw_nsp8_SUMO	CACAGAGAACAGATTGGTGGTGCCTTACAGAGTGAA	36
Rv_nsp8_SUMO	GATGGTGGTGCTCGAGCTGCAAAACAACAGTAGACA	36
Fw_nsp9_SUMO	CACAGAGAACAGATTGGTGGTAACAATGAGTTGATG	36
Rv_nsp9_SUMO	GATGGTGGTGCTCGAGCTGCAATCTCACTGTCGAG	35
Fw_nsp10_SUMO	CACAGAGAACAGATTGGTGGTGCGGGTACGGCAACT	36
Rv_nsp10_SUMO	GATGGTGGTGCTCGAGGCCTACACAGGAACAGCT	34
Fw_nsp12C_SUMO	CACAGAGAACAGATTGGTGGTAAGGACTTGCTTTTG	36
Rv_nsp12C_SUMO	GATGGTGGTGCTCGAGTTGCAGCACTGCACTTCTT	35
Fw_nsp12N_SUMO	CACAGAGAACAGATTGGTGGTTCAAAAGACACGACC	36
Rv_nsp12N_SUMO	GATGGTGGTGCTCGAGAAGAGACAAGCGATAACGAT	36
Fw_nsp16_SUMO	CACAGAGAACAGATTGGTGGTGCTGCTGCTGACTGG	36
Rv_nsp16_SUMO	GATGGTGGTGCTCGAGTTTGACATTTACTAGGCTATC	37
Fw_M_SUMO	CACAGAGAACAGATTGGTGGTAGTAGTACTACTCAG	36
Rv_M_SUMO	GATGGTGGTGCTCGAGGATTCTCAACAATGCGGTG	35
Fw_N_SUMO	CACAGAGAACAGATTGGTGGTATGTCTTTTGTTCCT	36
Rv_N_SUMO	GATGGTGGTGCTCGAGCACATTAGAGTCATCTTCTAA	37
Fw_Y_SUMO	CACAGAGAACAGATTGGTGGTTGGAGTGCTCGTTTG	36
Rv_Y_SUMO	GATGGTGGTGCTCGAGGCCCCCTTTAAGAGAGAAC	35

Table 2.5. Primers for SUMOStar cloning

Primer	Sequence	bp
Fw_SacII	TGGCTGCGTGAAAGCCTTG	19
Rv_SacII	ACCACCAATCTGTTCTCTGTG	21

2.1.2 pTriEx1.1 vector map

The pTriEx1.1 (Novagen) was kindly provided by Dr. Ian Jones. This vector contains mammalian, bacterial and insect promoters upstream of the cloning cassette. This vector contains an ampicillin resistance gene for positive colony selection, a HSV tag sequence upstream of the cloning site and a 8x His tag downstream of the cloning site to enable construction of N-terminal HSV-tagged and/or C-terminal His-tagged fusion proteins if desired **Figure 2.10**.

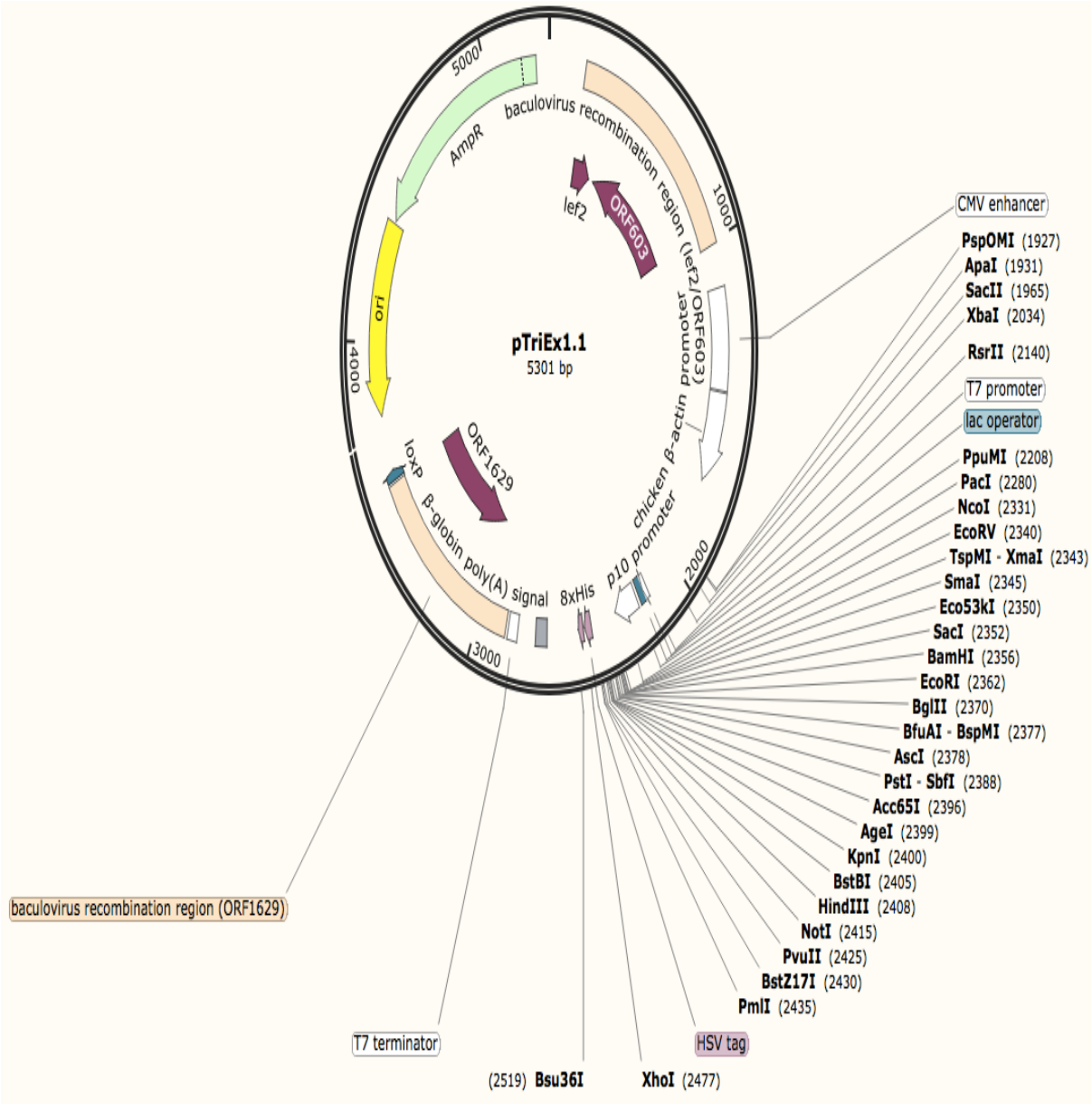

Figure 2.10. Map of the pTriEx1.1 vector showing the cloning sites.

2.1.3 pTriEx1.1-recombinant map

All MHV proteins in this study were cloned as C-terminal His-tagged fusion proteins by insertion between *Nco*I and *Xho*I restriction sites **Figure 2.11**.

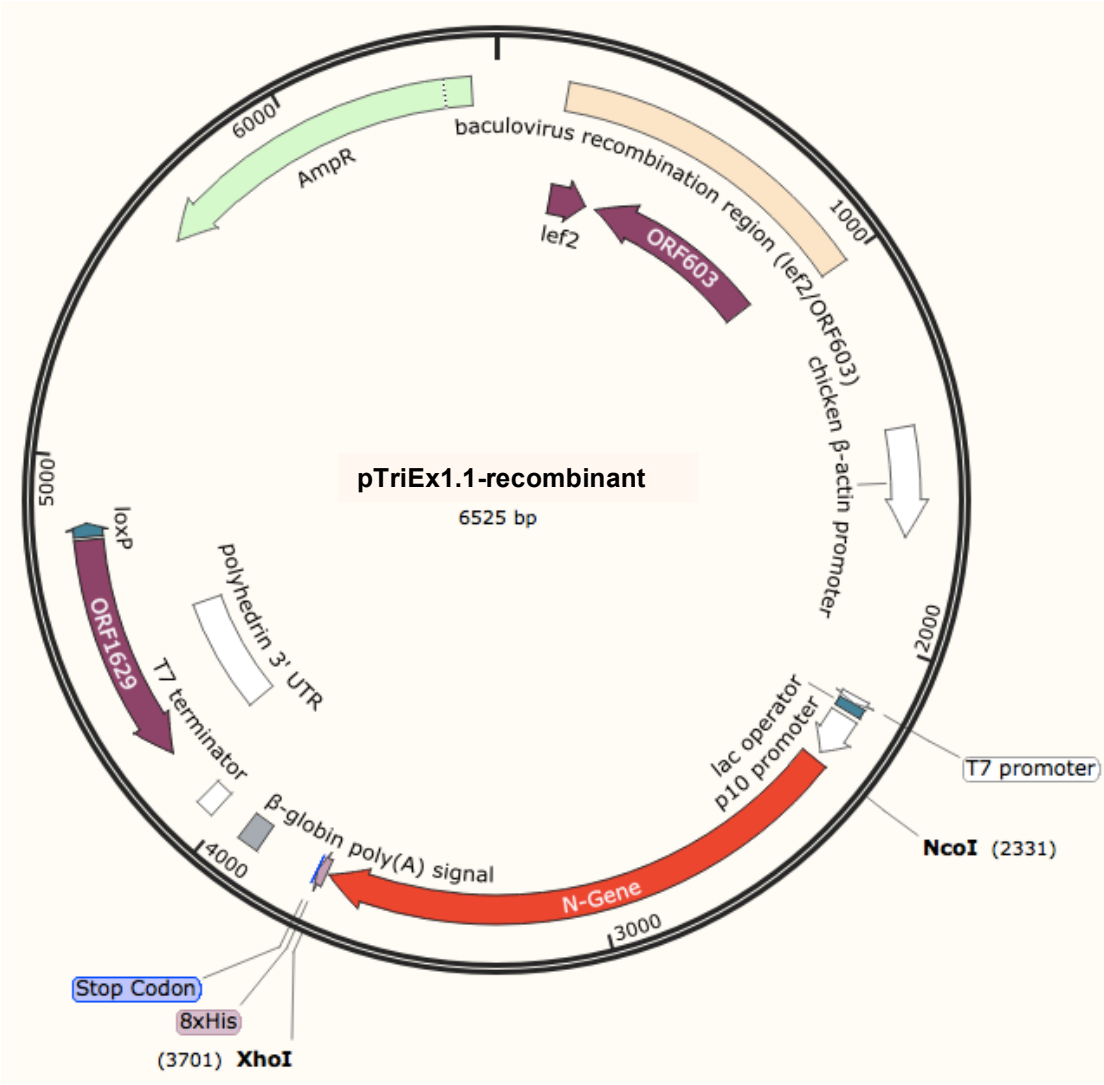

Figure 2.11. Vector map of the pTriEx1.1 vector showing the site of insertion

2.1.4 SUMO-pTriEx1.1

All MHV proteins in this study were expressed as N-terimal SUMOStar fusion proteins by insertion between *Nco*I and *Xho*I restriction sites in pTriEx1.1 vector **Figure 2.12**.

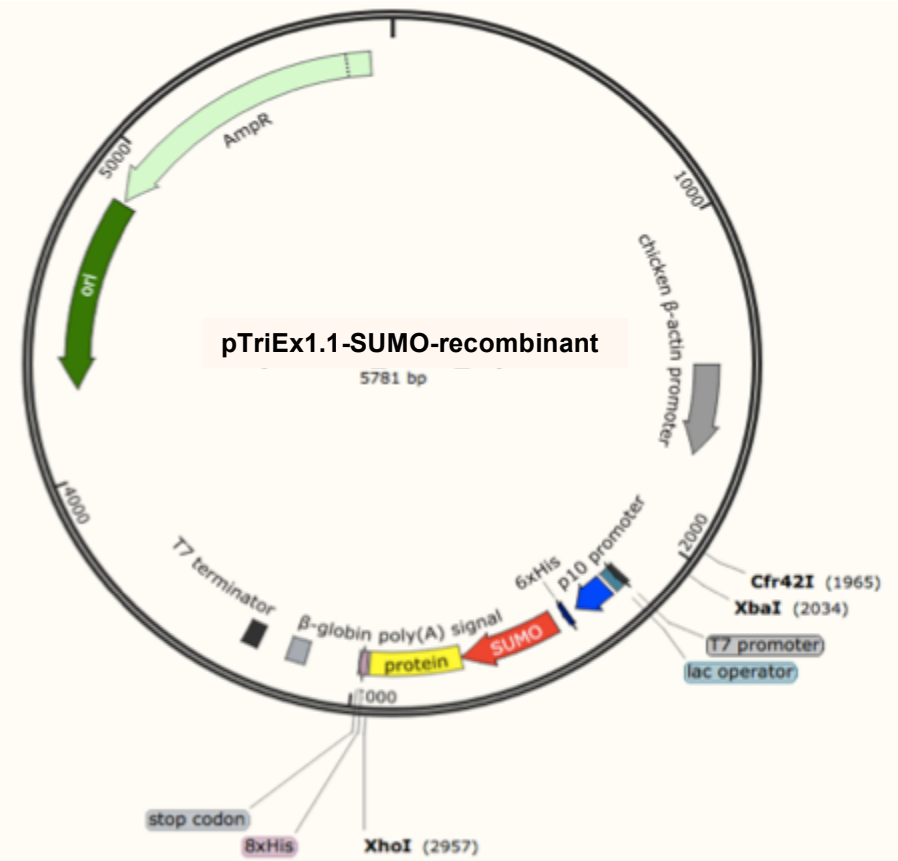

Figure 2.12. Vector map of the pTriEx1.1 vector showing the site of protein and SUMO tag insertion.

2.1.5 Amplification of DNA fragments

The coding sequences of structural proteins N protein and M protein, non structural proteins (nsp5, nsp6, nsp7, nsp8, nsp9, nsp10 and nsp16) from gene1, part of nsp (Y-domain, PI^{Pro} from nsp3 and N-terminal domain from RdRp and the C-terminal domain of nsp12 of MHV strain A59 (Accession No. AY700211.1) were amplified by polymerase chain reaction (PCR) from cDNA kindly provided by Dr. Volker Thiel. The amplification reaction contained the following components recommended by the manufacturer's protocol: CloneAmp HiFi PCR Premix, 100 µM forward and reverse primers, >100 ng DNA template and volume completed to 50 µl with dH_2O. Placed in PCR machine with the following thermal cycling

protocol:

Step	Temperature	Time	No. of cycles
Initial denaturation	95°C	3:00	1x
Denaturation	98°C	0:10	30x
Annealing	55°C	0:10	
Extension	72°C	0:30	
Final extension	72°C	5:00	1x

For optimal annealing temperature, gradient PCR was applied with the following protocol: 98°C for 30 sec, 95°C for 8 sec, 55°C - 65°C for 15 sec, 72°C for 20 sec (30x) and 72°C for 5 min. PCR products were purified using cleanup kit according to the manufacturer's protocol (MACHERER NAGEL), eluted in ultra-pure water. The concentrations were determined by ND-1000 Nanodrop spectrophotometer and stored at -20°C.

2.1.6 Double digest and gel extraction of pTriEx1.1 vector

The pTriEx1.1 vector was double digested with *Nco*I and *Xho*I (Thermo Fisher SCIENTIFIC) restriction enzymes according to the following protocol: 10x Green buffer, template (150> ng), (10 U/µl) *Nco*I, (10 U/µl) *Xho*I restriction enzymes and volume completed to 100 µl, incubated for 30 min at 37°C. After incubation, 100 µl was subjected to agarose gel electrophoresis, extracted and purified by cleanup kit (MACHERER NAGEL) following the manufacturer's instructions. The double digestion was used for extraction and confirmation of correct clones.

2.1.7 In-Fusion cloning of desired proteins to the pTriEx1.1 vector

Different genes were amplified using primers designed for In-Fusion cloning, as listed in **Table 1**. In a total volume of 10 µl, purified DNA (10-200 ng) was mixed with linearized vector (50-200 ng), 2 µl 5X In-Fusion HD Enzyme Premix (Clontech) and dH_2O following the manufacturer's protocol. Incubated for 15 min in 50°C.

2.2 DNA agarose electrophoresis

Gel electrophoresis was used to visualise DNA fragments and to purify by extracting the desired fragment. Agarose was dissolved in 1x TAE (Fisher Thermo SCIENTIFIC) buffer by heating in a microwave until boiling. After cooling Gel red (Cambridge Bioscience) was added to 1% (w/v). Electrophoresis was carried out at 120v for 45 min. Gels were imaged using G:BOX Chemi XL (Syngene). The approximate size of DNA bands in the agarose gel was determined using 1kb HyperLadder (Bioline). Images were photographed and printed.

2.3 Transformation of *E. coli* competent cells

For transformation, 50 µl of Stellar competent cells (Clontech) were thawed on ice and 5-50 ng of plasmid DNA was added and gently mixed. The DNA mixture was incubated for 30 min on ice. Next step was a heat shock at 42°C for 45 sec, which opens pores of the cell membrane to allow plasmid entry. The cells tube was placed back on ice for 1-2 min. Then, 450 µl of pre-warmed SOC (super optimal broth) recovery media (Clontech) was added to allow the cells to recover from heat shock and to express antibiotic resistance gene. After 1 hr incubation in the shaker (225 rpm) at 37°C, $1/5^{th}$-1/100 of each transformation reaction were separated and volume completed to 100 µl with SOC medium. Later, cells were plated onto Luria-Bertani (LB) agar supplemented with 100 µg/ml of ampicillin and incubated at 37°C

overnight. Approximately 30 colonies were selected randomly and transferred onto LB agar with ampicillin, left for ~5 hours at 37°C. Colonies were then analysed by colony PCR.

2.4 Colony PCR screening of transformant bacteria

2.4.1 Colony PCR analysis

Each single colony from LB agar transferred to 30 µl water incubated for 2 min at 100°C in PCR machine, centrifuged for 2 min and the supernatant (5 µl) used for PCR reaction with GoTaq Master mix (promega), (0.5 µM) T7 Forward or TriExUP and (0.5 µM) TriExDOWN primers **Table 2**. The PCR protocol was 1 cycle at 95°C for 2 min, 35 cycles with sequential incubation for 30 sec at 95°C, 30 sec at 52°C and 1 min at 72°C. Finally, an extension cycle for 10 min at 72°C was used to ensure complete extension of all PCR products.

2.4.2 Plasmid DNA purification and sequencing

Clones identified by colony PCR as containing the correct size insert were selected from LB agar plates and inoculated into LB broth containing 100 µg/ml ampicillin and incubated overnight in the shaker (225 rpm) at 37°C. Thermo Fisher SCIENTIFIC Miniprep kit was used for isolation of the plasmid DNA from the 10 ml cultures according to the manufacturer's guidelines for plasmid DNA purification. The samples concentrations were measured using a NanoDrop ND_1000 spectrophotometer. Later, pure plasmid DNA samples were sent to Source BioScience for sequencing and also double digested with restriction enzymes for confirmation. Cell suspension was used to create a stock for each clone in 50% glycerol and stored at -80°C.

2.5 BL21 (DE3)-pLysS transformation

Plasmids were transferred in *E. coli* BL21 (DE3)-pLysS (Invitrogen) competent cells. 0.1 µg plasmid DNA mixed gently with 50 µl competent cells, incubated on ice for 20-30 min, heat shocked for 45 sec at 42°C water bath, re-placed on ice for 1-2 min, 450 µl of pre-warmed SOC medium were added to plasmid-cells mixture and incubated in the shaker at 37°C for 40-45 min (225 rpm). After incubation, 400 µl and 100 µl from the total mixture were plated on LB plates containing 100 µg/ml ampicillin and incubated overnight at 37°C. The next day 2 transformants were grown under the selection of chloramphenicol (34 µg/ml) and 100 µg/ml ampicillin in 10 ml LB broth at 37°C in the shaker overnight. The overnight culture was inoculated (100 µl) into 10 ml fresh LB broth with ampicillin (100 µg/ml), incubated in the shaker at 37°C for 1-3 hr. The expression of MHV fusion proteins was induced when cultures had reached an OD_{600} of 0.3 by adding IPTG (isopropyl-β-D-thiogalactopyranoside) at a final concentration of 100 mM. For optimal expression, 1 ml was taken at time points of 1,2,3, and 4 hr, centrifuged and cells pellet were subjected for SDS-PAGE.

2.6 Cloning with SUMOStar

2.6.1 Amplification of SUMOStar and desired protein fragments

The gene encoding SUMOStar was amplified from the pTriExSUMO plasmid using primers listed in **Table 5**. The reaction contained the following components: 1 x phusion^{R} High-Fidelity PCR Master mix (BioLabs), (0.5 µM) forward and reverse primers, ~ 500 ng DNA template and dH_2O added to a final reaction volume of 50 µl. PCR products were amplified using method described in section 2.1.5. MHV coding sequences of interest were amplified using the same protocol with specific primers for each sequence **Table 4**. Gel electrophoresis was carried out to separate and visualise DNA fragments. The PCR products were purified using gel extraction kit (Thermo Fisher SCIENTIFIC). The PCR amplified SUMOStar fragment was mixed with the MHV PCR products using the following protocol: 1 x Phusion^{R} High-Fidelity PCR Master Mix, SUMOStar fragment (0.1 ng/bp) and MHV PCR product (0.1 ng/bp) and adjusted to a final volume of 45 µl with dH_2O. The reactions were placed in a PCR machine and 5 cyles of the program described in 2.1.5 was used before adding the forward primer Fw_SacII (0.5 µM) and reverse primer specific to the MHV coding sequence **Table 4** (2.5 µl). The PCR protocol in 2.1.5 was carried out for a further 30 cycles. The gel extraction kit from (Thermo Fisher SCIENTIFIC) was used according to the manufacturer's instructions to purify and extract DNA fragments after gel electrophoresis.

2.6.2 Restriction digest of pTriEx1.1 plasmid and DNA fragments

In separate tubes, the plasmid vector and the DNA fragments from 2.6.1 were digested first using the restriction endonuclease *Cfr*421 (*Sac*II) (Thermo Fisher SCIENTIFIC) restriction enzyme according to the following protocol: 1x buffer B

(Thermo Fisher SCIENTIFIC), 0.5-1 µg/ml DNA template, (10 U/µl) *Cfr*421 (*Sac*II) restriction enzymes and volume adjusted to 50 µl using dH_2O, and incubated for 60 min at 37°C. After incubation, products were purified using a PCR purification kit (Thermo Fisher SCIENTIFIC). A second restriction digest was carried out using fast digest *Xho*I (Thermo Fisher SCIENTIFIC) according to the following protocol: 1x Green buffer, 0.5-1 µg/ml DNA template from first digest, (10 U/µl) *Xho*I restriction enzymes in a final volume of 100 µl adjusted using dH_2O, and incubated for 30 min at 37°C. Following visualisation by agarose gel electrophoresis, the appropriate DNA size was purified using the gel extraction kit (Thermo Fisher SCIENTIFIC) according to the manufacturer's instructions.

2.6.3 Ligation of PCR fragments in pTriEx1.1

Ligation reactions were prepared using an approximate 3:1 molar ratio of the insert and vector together from 2.6.2 with 1x ligation buffer (Thermo Fisher SCIENTIFIC), and 1 unit T4 DNA ligase in a final volume of 10 µl. The reaction was incubated at room temperature for 1-3 hours. The ligation mix was then ready for transformation into Stellar competent cells (Clontech).

2.7 Transfection of mammalian cells

Transfection is a method used to introduce nucleic acids into cells. 17-clone1 were cultured in Dulbecco's modified Eagle medium (DMEM) (Sigma Aldrich) supplemented with 10% fetal bovine serum (FBS) (GE Healthcare), and antibiotics penicillin/streptomycin (penicillin 100 U/ml, streptomycin 0.1 mg/ml; Gibco/Invitrogen) in 6 well plate and were transfected with plasmid DNA for the expression of viral proteins by using different transfecting reagents to compare their efficiency Turbofect (Thermo Fisher SCIENTIFIC), FugeneHD (Promega),

Fugene6 (Roche), MirusTransIT-LT1 (Mirus), Lipofectamine2000 (Invitrogen), and Lipofectamine3000 (Invitrogen) following the manufacturer's instructions for each reagent. After incubation for 18, 24 and 48 hr, cell culture plates were placed on ice and by using a cold plastic cell scraper to scrape adherent cells of each well gently. The suspensions were transferred into pre-cooled tubes, centrifuged for 4 min (2000 rpm) at 4°C. Finally, the supernatants were aspirated and the pellets re-suspended with 50 µl ice-cold phosphate-buffered saline (PBS) and stored at -20°C.

2.8 Treatment with proteasome inhibitor

MG132 proteasome inhibitor treatment is performed to determine if coronavirus proteins in this study are degraded in mammalian cells. 17 clone-1 cells were seeded at $1.9x10^5$ in 24 well plates and the next day, cells were transiently transfected with plasmid DNA using different transfection reagents. Transfected mammalian cells were incubated for 4 hr, and then MG132 (Sigma-Aldrich) was added to each well at final concentration of 50 µM. Untreated cells were incubated for 24 and 48 hr, and then harvested as described in section 2.7. The MG132 proteasome inhibitor was added to transfected cells after 20 hr incubation and incubated for 4 hr before harvesting. Cell lysates were subjected to immunoblotting. Untreated cells were used as a control.

2.9 SDS Polyacrylamide gel electrophoresis (PAGE)

The expressed proteins were detected by Western blot. Transfected cells were aliquoted in 50 µl chilled PBS, 30 µl of samples were mixed with 10 µl of LDS sample buffer (980 µl 4x LDS sample buffer (Novex) and 20 µl β-mercaptoethanol), incubated at 100°C for 10 min before being loaded into 4-12% SDS-PAGE gel (Life

Technologies). Later, samples were centrifuged for 2 min and 10 µl loaded in mini gel tank (Life Technologies). Samples were separated at a constant voltage of 170v for 30 min in a 1x MES (50 mM MES, 50 mM Tris, 0.1% SDS, 1% EDTA, pH 7.25) running buffer (Life Technologies). A sample (10 µl) of sharp pre-stained protein standard or SeeBlue Plus2 Prestained Standard (Life Technologies) was used for molecular size estimation. Once electrophoresed, the gel will be transferred either onto PVDF membranes for Western blotting analysis or stained with coomassie blue.

2.10 Protein staining with coomassie blue

Coomassie blue dye was used to visualise proteins separated by SDS PAGE. After electrophoresis, the gel was incubated at room temperature on rocking platform (25 rpm) in a solution of coomassie brilliant blue stain (0.025% coomassie Brillant Blue R-250, 45% methanol and 10% glacial acetic acid) for 30 min. The gel was then washed with a destaining solution (10% methanol, 10% glacial acetic acid) at room temperature on the rocker for 10 minutes. The washing step was repeated three times. The gel kept overnight on rocking platform (25 rpm) to enable the visualisation of individual protein bands.

2.11 Western blotting analysis

Following electrophoresis, SDS-PAGE gels were incubated for 1 min in transfer buffer (400 ml 1.5x Tris/glycine (Tris base 25 mM, Glycine 190 mM pH 8.3) buffer, 20% methanol (Fisher)) for 3 min. The SDS-PAGE gels were transferred onto PVDF membranes (Merck) and electroblotted for 1 hour and 20 minutes/ at 150mA in transfer buffer using semi-dry Western blotting apparatus (ATTO). Following transfer, the membranes were incubated in blocking buffer (5% non-fat milk in

TBST buffer (500 ml 1xTBS (50 mM Tris-Cl, 150 mM NaCl, pH 7.6) and 2% Tween20 (Fisher BioReagents)), kept in the fridge at 4°C to reduce nonspecific binding for overnight. The next day, membranes were washed with TBST buffer 3 times for 5 min on a rocking platform (25-30 rpm) (Stuart). The primary 6x His tag antibody (Abcam) was diluted 1/10,000 in blocking buffer, which was added to the membrane in a plastic pouch. Membranes were incubated for 1 hr at room temperature on rocking platform (25-30 rpm). After incubation, membranes were washed with TBST buffer 3 times for 5 min. Membranes were then incubated for 1 hr at RT with the secondary antibody (polyclonal Goat Anti-Rabbit Immunoglobulins /HRP) dissolved in 1% TBST-milk on rocking platform (25-30 rpm), followed by 3 times washing steps in TBST as previously described and the protein bands were detected using equal volumes of ECL reagents A (0.25 ml) and B (0.25 ml) enhanced chemiluminescent substrate (GE healthcare). Luminescence from his-tag labeled proteins was visualized using a G: BOX Chemi XL(Syngene).

2.12 Immunoprecipitation

Cell extracts were prepared on ice by incubating the cell pellets in 500 µl of NP40 lysis buffer (150 mM sodium chloride, 1.0% NP-40, 50 mM Tris,pH 8.0), vortexed and incubated for 30 min on ice. Following centrifugation, clarified lysate were ready for immunoprcipitation. 50 µl of Dynabeads protein A (invitrogen) was aliquoted into Eppendorfs and placed in a magnetic rack for 1 min and the supernatant was removed. Tubes were removed from the rack and the Dynabeads re-suspended in 200 µl of PBS with 0.02% Tween20 containing Anti-His tag antibody (Ab) (1-10 µg) and incubated for 10 min with slow rotation at 4°C. The supernatant was removed after placing in the magnetic rack and Dynabeads-Ab copmlex re-suspended in 200 µl PBS with 0.02% Tween20. The supernatant was

removed again and 200 µl of the sample containing Ag was added, incubated on ice for 30 min to allow antigen (Ag) to bind to the beads-Ab complex. After removing supernatant, the beads were washed with wash buffer (PBS and 0.1% Tween-20). This wash cycle was repeated 3 times with 200 µl wash buffer and each wash included separation by magnet, supernatant removel and resuspention by gentle pipetting. Then, beads-Ab-Ag complex were re-suspended in 100 µl of washing buffer and placed on magnetic to remove supernatant. Then, 30 µl of elution buffer (LDS sample buffer (4X) (novex)) containing loading dye was added and incubated at 100°C for 10 min. The supernatant was taken to a new microcentrifuge tube after using the magnetic rack and subjected to SDS-PAGE. The gel was transferred to PVDF and immunoblotted as described in section 2.10.

2.13 Error-prone PCR

Error-prone PCR was performed using two protocols to create a library of variants. In the first protocol a 100 µl reaction solution containing 10 mM Tris.Cl (pH 8.3), 50 mM KCl, 7 mM $MgCl_2$, 1 mM dCTP, 1 mM dTTP, 0.2 mM dATP, 0.2 mM dGTP, 2 µM Forward and Reverse PCR primers, 20 pg/µl DNA template, 0.5 mM $MnCl_2$ and 0.05 U/µl *Taq* DNA polymerase (New England Biolabs). The reaction was placed in PCR machine with following thermal cycling conditions: 95°C for 10 min (initial denaturation), 94°C for 1 min, 60°C for 1 min, 72°C for 3 min (15 cycle), 72°C for 5 min (final extension), which will achieve a mutation rate of 0.66% per nucleotide position. In the second protocol, a set of 5 tubes prepared and each tube contain: Phusion^R High-Fidelity PCR (BioLabs), (0.5 µM) forward and reverse primers **Table 1**, DNA template and volume completed to 50 µl by dH_2O. Different concentrations of $MnCl_2$ (50 µl, 100 µl, 200 µl .300 µl, 400 µl) were added to each tube before amplification. PCR products were amplified using the protocol described in 2.1.5.

The PCR products were analyzed by agarose gel electrophoresis and the correct sizes of the PCR products were extracted using a gel extraction kit according to the kit manufacturer's instructions. The products were cloned into pTriEx1.1 vector, transformed in Stellar cells (section 2.3) and subsequently plated on solid LB medium with 100 µg/ml ampicillin and the mutation rate was confirmed by sequencing 10 randomly selected colonies.

2.14 Directed evolution

The 17 clone-1 mammalian cells were grown in a T25 tissue culture flask and when cells confluence reached ~70%, cells were transfected with a library of variants generated by ep-PCR second protocol (50 µl $MnCl_2$) (section 2.13) using Turbofect reagent according to the manufacturer's instructions. After 24 hr incubation, the adherent cells were washed with PBS twice and infected with MHV-A59 (MOI=3). The flask was incubated at 37°C on a low speed platform rotator (Grant-bio) for 45 min. After incubation, the inoculum was removed and adherent cells gently washed with 2 ml PBS. A fresh 5 ml DMEM media added to the flask and incubated at 37°C for 16 hr. The supernatant was collected and kept at -80°C. The flask was washed carefully with 2 ml PBS 3 times to eliminate all dead cells. Intact cells were recovered by scraping and low speed centrifugation at 4°C and any plasmid DNA present in the surviving cells was extracted using a plasmid miniprep kit (Thermo Fisher SCIENTIFIC), amplified with specific primers **Table 1** using PCR amplification protocol described in section 2.1.5. After amplification, gel electrophoresis was carried out to visualize DNA fragments. As described previously, the PCR reaction products were purified using PCR purification kit (Thermo Fisher SCIENTIFIC). The DNA fragments were then inserted in plasmid through In-Fusion protocol, transformed in Stellar competent cells and each 100 µl

from transformed mixture was placed in 10 ml LB broth with 100µg/ml ampicillin. After 16 hr incubation, Miniprep kit (Thermo Fisher SCIENTIFIC) was used for isolation of the plasmid DNA from the culture cells according to the manufacturer's instructions and concentration was measured. The library was then ready for second round of directed evolution. This process was repeated for several passages and the supernatants were collected from each passage and applied for plaque assay.

2.15 Cell culture

17clone-1 mouse fibroblast cells were grown at 37°C in Dulbecco's modified Eagle medium (DMEM; Gibco/Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum (FBS), penicillin/streptomycin (penicillin 100 U/ml, streptomycin 0.1 mg/ml; Gibco/Invitrogen), 1% L-Glutamine (Gibco/Invitrogen) and 1% non-essential amino acids 100× concentrate (NEAA; Gibco/Invitrogen). For the maintenance of cells, cells were split once they reached ~90% confluence. The media was removed and the cells were washed twice with sterile 1x PBS. The cells were detached from the plastic using 0.2% Trypsin-EDTA (PAA; Sigma). The cells were then suspended in fresh DMEM media to inactivate the trypsin. An aliquot of the cell suspension was added to a new flask containing fresh DMEM media and this sub-cultivation cycle were routinely repeated.

2.16 Determination of viral titre by plaque assay

17clone-1 cells were seeded into 24 well plate (Greiner Bio-one) at approximately 1.9×10^5 cells per well, and incubated at 37°C in 5% CO_2 for two days, at which a monolayer reached ~90-100% confluence. Serial 10-fold dilutions of the MHV-A59 were prepared in serum free medium (Sigma Aldrich). Following the aspiration of

culture media from the wells, 100 µl of each viral dilution was added per well in four wells. Plates were incubated at 37°C for 45 min on a low speed platform rotator to allow virus to infect the cells and avoid plates drying. 1.4 gm of agarose was dissolved in 100 ml distilled water and autoclaved to set agarose medium. Mixing 2x DMEM (composed of 13.38 gm Dulbecco's modified eagles medium powder (gibco), 3.7 gm $NaHCo_3$ (BDH AnalaR) in 500 ml nH_2O, Filtered in 0.22 µm filter (Sarstedt)) with 4% FBS, 2% penicillin/streptomycin (Gibco, Invitrogen), L-Glutamine (Invitrogen) and non essential amino acid (Sigma Aldrich) for preparing overlay. Equal volumes of agarose and overlay were mixed and kept at 42°C in water bath. After incubation, the inoculum was removed and the cells were washed with PBS. 1-1.5 ml of agarose overlay media was added in each well, allowed to solidified and incubated at 37°C in humidified CO_2 (5%) for three days. After 72 hr, cells were fixed by adding fixation solution (25 ml 37% formalin (Sigma-Aldrich) and 75 ml PBS) for one hour at room temperature. Then, agarose overlay were removed by flipping out and monolayers were stained with staining solution (35 ml 1% crystal violate (ACROS), 35 ml ethanol and volume completed to 100 ml with distilled water) for 10 min, washed, and stained plaques of 4 wells of each appropriate dilution were counted. PFU/ml were calculated to determine the viral infectivity using the following formula: average/dilution factor*volume of diluted virus added to the well.

2.17 Evaluation of cytotoxicity using 3-(4,5-dimethylthiazol-2-yl)-2,5- diphenyltetrazolium bromide assay (MTT)

The MTT assay method was used to assess the cytotoxicity of wild type and mutated proteins *in vitro*. Cells viability was measured by monitoring the conversion of MTT (Sigma) to formazan. 17clone-1 cells were seeded into 24 well plates

(Greiner Bio-one) (100 µl/ well) and allowed to adhere for 24 hr. The next day, each well was transfected with a plasmid containing wild type and mutated protein in triplicate. After incubation, 20 µl of filter sterilized MTT (5 mg/ml in PBS) was added to each well. Following 4 hr incubation period with MTT at 37°C, media was removed and 100 µl of sterile DMSO was added to dissolve blue formazan crystals trapped inside cells. The absorbance at 595 nm was measured with a plate reader (Tecan GENios Microplate Reader) after 1 hour and 24 hours incubation at 37°C.

2.18 Nsp16 protein purification

2.18.1 Preparation of LOBSTR *E. coli* competent cells

LOBSTR *E. coli* competent cells was made chemically competent using the following protocol; A glycerol stock of LOBSTR *E. coli* strain was kindly provided from Prof. Ian Jones and streaked in fresh LB agar plate to obtain single colonies. A single colony was then inoculated into 10 ml of LB medium and grown at 37°C in the shaker (225 rpm) overnight. The following day, 1 ml of the starter culture was diluted 1/50 to a final volume of 50 ml with fresh LB medium and incubated with shaking at 37°C until an OD_{600} of 0.3- 0.5 was reached. The cells were harvested by centrifugation at 4,500 rpm for 20 min at 4°C in falcon tube. The supernatant was discarded and the bacterial cell pellet was re-suspended in 5 ml ice-cold sterile 0.1M $CaCl_2$ and incubated on ice for 3 min. The cells were harvested as above and the pellet re-suspended in 5 ml 0.1M $CaCl_2$ containing 15% glycerol (v/v). The cell suspension was divided into 50 µl aliquots in sterile 1.5ml eppendorf tubes, snap frozen in liquid nitrogen and stored at -80°C.

2.18.2 Heat shock transformation of competent LOBSTR *E. coli*

Chemically competent LOBSTR cells (50 µl) were thawed on ice for 5-10 minutes. 50-100 ng Plasmid DNA was added and incubated for 25 min on ice. “Heat shock” was carried out at 42°C for 45 sec followed by immediate transfer to ice for 2 min. 450 µl of pre-warmed SOC media was added to allow the bacteria to recover from the heat shock. After one hour incubation at 37°C in the shaker (225 rpm), 100 µl and 400 µl were plated on LB agar plates containing ampicillin and incubated at 37°C overnight. After 18 hr incubation, single colonies were inoculated in 10 ml LB broth supplemented with 100 µg/ml ampicillin and grown overnight at 37°C in the shaker (225 rpm). Glycerol stocks (15% v/v) were prepared and stored at -80°C.

2.18.3 Small scale protein expression

Levels of expression and solubilites of nsp16 protein were assessed in different *E. coli* host strains (BL21 (DE3)-pLysS and LOBSTR *E. coli*) growing in either LB or Auto induction media (AIM), with different induction temperatures (16°C and 37°C) and different time for induction (3 hours and 16 hours). The expression plasmid was transformed into the cloning hosts; BL21 (DE3)-pLysS and LOBSTR *E. coli* as descried in sections 2.3 and 2.18.2, previously. The next day, single colonies were inoculated to 10 ml LB with ampicillin, incubated overnight at 37°C in the shaker (225 rpm). The following day, the starter culture was diluted 1/1,000 to a final volume of 50 ml with fresh LB and AIM media and incubated with shaking until an OD_{600} of 0.6 was reached. For cells in LB media, IPTG was added to a final concentration of 1 mM to induce expression. The culture was left to grow at 37°C for three hours or overnight at 16°C. Then, the cells were harvested by centrifugation (4500 rpm for 45 min at 4°C). The supernatant was discarded and

the cell pellets were then re-suspended in 1x binding buffer (50 mM Tris-HCL, 150 mM NaCl, 40 mM imidazole, pH 8.0) with 1x EDTA-free, cOmplete ULTRA protease inhibitor cocktail (Roche), stored at -20°C. For AIM media, the culture was left to grow overnight at 16°C with agitation. Then, the cells were harvested by centrifugation (4500 rpm for 45 min at 4°C), the supernatant was discarded and the cell pellets were then re-suspended in 1x binding buffer with 1x EDTA-free, cOmplete ULTRA protease inhibitor cocktail, stored at -20°C.

2.18.4 Large scale protein expression

For large-scale expression in LOBSTR *E. coli,* transformed cells were recovered from glycerol stock (section 2.18.2) by streaking a loopful onto LB agar containing ampicillin (100 µg/ml). A single colony was picked and used to inoculate a 10 ml overnight LB culture. The following day, the overnight culture was diluted 1/500 in fresh 500 ml of selective LB medium with ampicillin and incubated at 37°C in the shaker until an OD_{600} of 0.6 was reached. Adding IPTG induced protein expression. The culture was left to grow at 37°C for three hours. 1 ml was taken after each hour to optimize level of induction. Cell pellets were harvested and re-suspended in binding buffer containing proteinase inhibitor tablets and were ready for sonication, filtration, and purification by HisTrap™ HP column.

2.18.5 Cell lysis

Cell stocks were left to thaw on ice before proceeding with purification; a pinch of lysozyme (SIGMA) and 10 µl of benzonase (Expedeon) was added to concentration of 250 unit/µl. Then, sample was sonicated for 5 min on ice (30 sec on: 30 sec off) with 85% amplitude. 500 µl of Triton 100x (Fisher) was added and samples were incubated on ice for 30-60 minutes. The samples were centrifuged at

4,500 rpm for 20 min at 4°C in falcon tube. The soluble fraction was filtered through a 0.8 µm or 0.45 µm syringe filter unit (Millipore Merck). The supernatant was ready for purification.

2.18.6 Purification of protein

Proteins was purified with a Ni^{2+} affinity column (HisTrap™ HP column; GE Healthcare), according to the manufacturers' instructions', the 5 ml column was equilibrated with 5 column volumes (CV) of binding buffer containing low imidazole concentration. The protein sample was loaded onto the column using a syringe. Then, the column was washed with 5 CV of binding buffer to remove unbounded materials. The sample was eluted with elution buffer (50 mM Tris-HCL, 150 mM NaCl, 400 mM imidazole, pH 8.0). The 5 ml fractions were collected in separated falcon tubes and those containing the protein of interest were concentrated using the vivaspin 10kDa molecular weight (GE Healthcare). SDS-PAGE (section 2.9) and coomassie staining (section 2.10) were used to check purification of protein at different stages. The column was immediately re-equilibrated with 5-10 CV binding buffer.

2.18.7 Determination of protein concentration by Bradford Assay

Bradford assays were used to measure protein concentrations. Bovine serum albumin (BSA) (Sigma) was dissolved in PBS buffer to prepare standard curve. The stock solution was diluted with PBS to obtain serial of dilutions.10 µl of each standard dilution was added into a 96 well plate in duplicate and 200 µl of Bradford reagent (BioRad) was added. 10 µl of unknown protein solution was combined with 200 µl of Bradford reagent in duplicate. Two more different dilutions (1x, 5x) of unknown protein were added in duplicate and the absorbance of each sample was

measured at OD_{595}. The unknown protein concentration is deterimed by blotting a BSA protein standard curve against the absorbance 595 nm.

2.18.8 Statistical analysis

The error bars represent the standard deviation of the mean. Statistical significance was determined using the t-test and differences were considered significant if the *P*-value was less than or equal to 0.05.

3 Cloning and expression of MHV genes and proteins in *E. coli* BL21(DE3)-pLysS competent cells and mammalian cells.

3.1 Introduction

Murine Hepatitis Virus (MHV) is used as a typical model system for the study of coronavirus replication and transcription (Snijder *et al*., 2003) and is closely related to SARS-CoV. MHV has historically served as a model system for studying infections caused by CoVs since it has the undeniable advantage of easy growth in cell culture in comparison to other CoVs such as human coronaviruses. Several MHV-A59 proteins have been investigated in the past two decades to clarify protein functions and protein-protein interaction. It has been observed in a number of studies that the presence of some coronavirus proteins and genes may alter the infectivity of the virus (Cui *et al*., 2015), some proteins stimulate other proteins, which is a common mechanism for coronaviruses as they form many protein-protein complexes and targeting these proteins can suppress coronavirus replication (Wang *et al*., 2015). To investigate any potential inhibitory effect of different CoV proteins requires transient expression of these proteins in suitable mammalian cells.

The expression of stable protein in mammalian cell requires several factors to be considered. For instance, selection of a proper vector is an essential factor that needs to be considered during experiment design. The plasmid pTriEx1.1 is designed to allow rapid characterization of target proteins in multiple expression systems (*E. coli*, insect and vertebrate). Expression of the gene of interest in mammalian cells is mediated by a hybrid promoter composed of the

cytomegalovirus (CMV) immediate early enhancer fused to the chicken β-actin promoter, termed a CAG promoter. In addition, pTriEx1.1 recombinant plasmids can be transferred into certain *E. coli* strains engineered to express the T7 polymerase allowing IPTG induction of T7 promoter driven transcripts. Another critical factor is the selection of proper cloning method. Various techniques have been developed to introduce DNA into cultured prokaryotic and eukaryotic cells efficiently and selection of a suitable cloning system depends on several important numbers such as high fidelity, ease of use, less time consumption, low cost and high yield and solubility of any protein in a single batch (Marsischky and LaBaer, 2004). Traditional cloning involves any DNA constructs that are joined by a ligation enzyme at restriction enzyme sites. However, the available unique sites in the vector and gene limit construct options. Such traditional DNA constructs are also undesirable for fusion proteins because this method sometimes includes undesired amino acids encoded by the restriction sites engineered to provide a joining point which can lead to reduced protein expression. In contrast, an In-Fusion enzyme reaction has the ability to join any two pieces of DNA that have 15 bp of identity at their ends. In addition, four or more pieces of DNA can be joined by In-Fusion in a single reaction (Zhu *et al*., 2007).

The In-Fusion method is based on ligation independent cloning and depends on the unique properties of the 3′–5′ exonuclease activity of poxvirus DNA polymerase (Hamilton *et al*., 2007; Zhu *et al*., 2007). This system requires linear duplex DNAs with homologous ends that are incubated in the presence of Mg^{2+} and low concentrations of dNTP. The 3′–5′ proofreading activity of poxvirus DNA polymerase removes nucleotides from the 3′ end to produce single-stranded ends. Through base pairing, the complementary regions are annealed spontaneously to

generate a joined molecule that contains a hybrid region flanked by nicks, 1–5 nucleotide gaps, or short overhangs. The low affinity of poxvirus DNA polymerase toward nicked or gapped DNA ends than for duplex ends permits the formation of metastable annealed structures. Finally, after transformation *E. coli* will repair any single-stranded gaps when introduced and the result is equivalent to a recombination event at the ends of the DNAs **Figure 3.13**. The In-Fusion cloning method was therefore selected as the most appropriate method to clone selected coding regions from the MHV-A59 genome into the pTriEx1.1 vector.

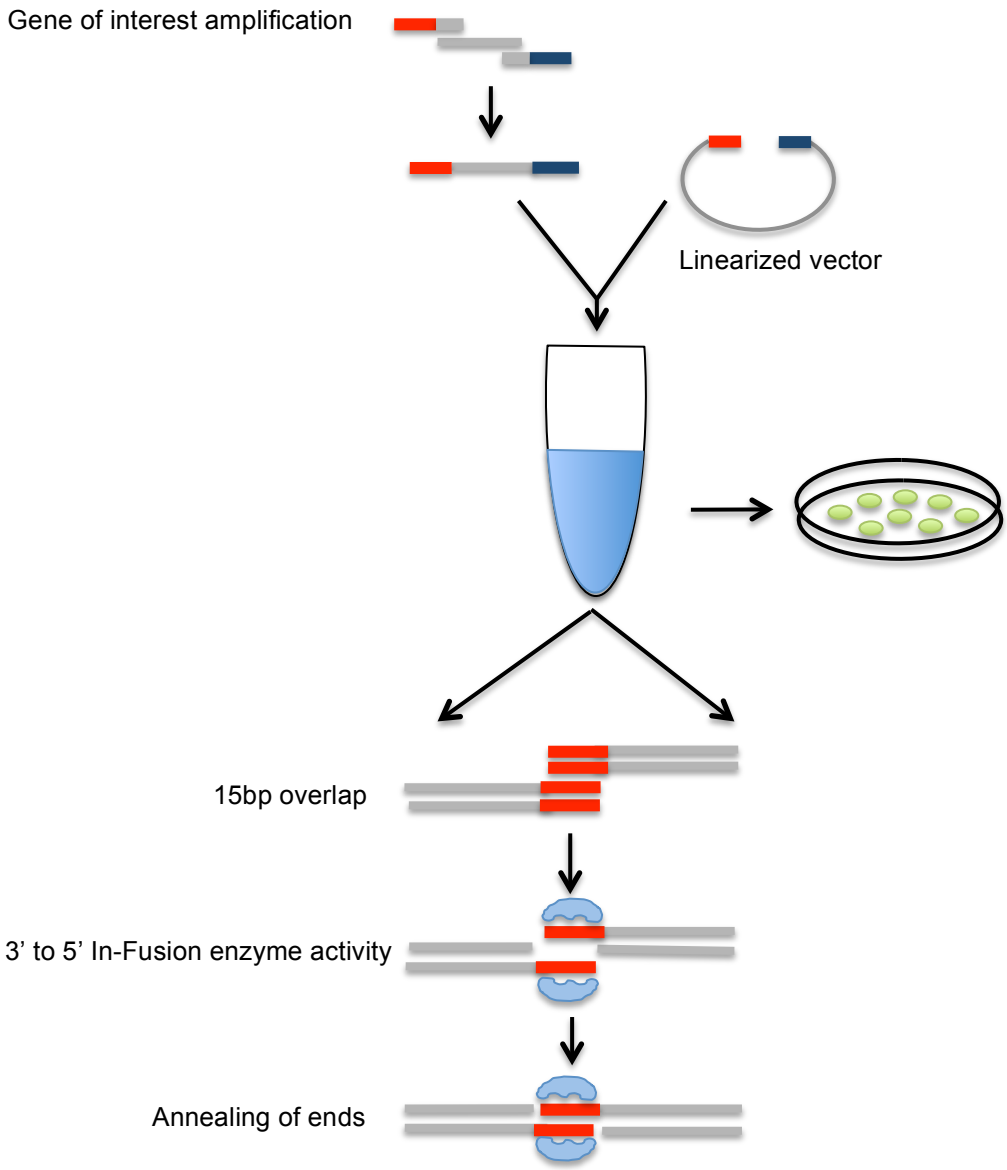

Figure 3.13. The In-Fusion system. The gene of interest is amplified with specific designed primers that contain 15bp extensions homologous to vector ends. The In-Fusion cloning reaction is set up for 15 minutes at 50°C. The Vaccinia virus DNA polymerase that has 3’ 5’ exonuclease activity will anneal the 15bp fragment of single stranded DNA. The In-Fusion mixture then, transformed into competent cell and transformed cells are grown on agar plate supplemented with antibiotic. Transformation of the construct into *E. coli* will repairs any gaps or nicks in the DNA sequence (Adapted from Zhu *et al*., 2007).

A gene delivery system is required to achieve cytoplasmic gene expression and the delivery of plasmid DNA by transfection reagents is one of the simplest and widely used methods. There are several protein transfection reagents that are commercially available however and a major key to success in protein delivery to cells is choosing the appropriate transfection reagent with high efficiency and low cytotoxicity. There are several factors that are important in selecting transfection reagent. Firstly, surface charge density, which is an important factor that can be related to the transfection efficiency. Several studies showed that the highest values of transfection occurred for the most positively charged complexes (Farrow *et al*., 2006), while other studies observed that the highest transfection efficiency was with negatively charged complexes. Secondly, some manufacturer's protocols recommend serum-free medium for best results (Oba and Tanaka, 2012) as size, chemical structure and mechanism for gene delivery are also factors to be considered when choosing protein transfection reagents for experiments.

This aim of this chapter is to generate constructs for several MHV-A59 proteins in the pTriEx1.1 vector, demonstrate expression in *E. coli* and develop tranfection of 17clone-1 mammalian cells, which are chosen as they are permissive for MHV-A59 infection and so, by definition, contain any necessary cellular factors required by the virus.

3.2 Results

In-Fusion cloning technology was used in this study. This technology ensures easy, directional cloning of DNA fragments into any vector with high cloning efficiency for any DNA fragment size. This strategy relies on recognition of 15 bp overlap sequences at the ends of each of the desired DNA fragments and linearized vector and then fuses these DNA fragments at these ends. These 15 bp overlap sequence can be simply generated by PCR amplification after addition them to the requisite primers for the genes concerned. The vector of choice, in this case pTriEx1.1, is linearized with *Nco*I and *Xho*I in such a way that the amplified ORF in the final construct will allow expression to include a C-terminal His tag already present in the pTriEx1.1 vector **Figure 3.14**. The His tag in the pTriEx1.1 vector enables detection of all expressed proteins using SDS-PAGE analysis and Western blot with a His tag specific antibody. Structural proteins such as N and M proteins, non structural proteins (5,6,7,8,9,10,16) from gene1, non-structural proteins Pl^{Pro}, Y-domain from nsp3 and the N-terminus and RdRp from the C-terminal region of nsp12 from MHV-A59 cDNA were PCR amplified with primers designed specifically for each coding region using the In-Fusion primer design tool **Table 2.1**. Each PCR product was mixed with linearized vector in the In-Fusion reaction mix. This cloning strategy should derive vectors that allow for expression of MHV-His tagged fusion proteins in both *E. coli* and 17clone-1 mammalian cells via the promoters present.

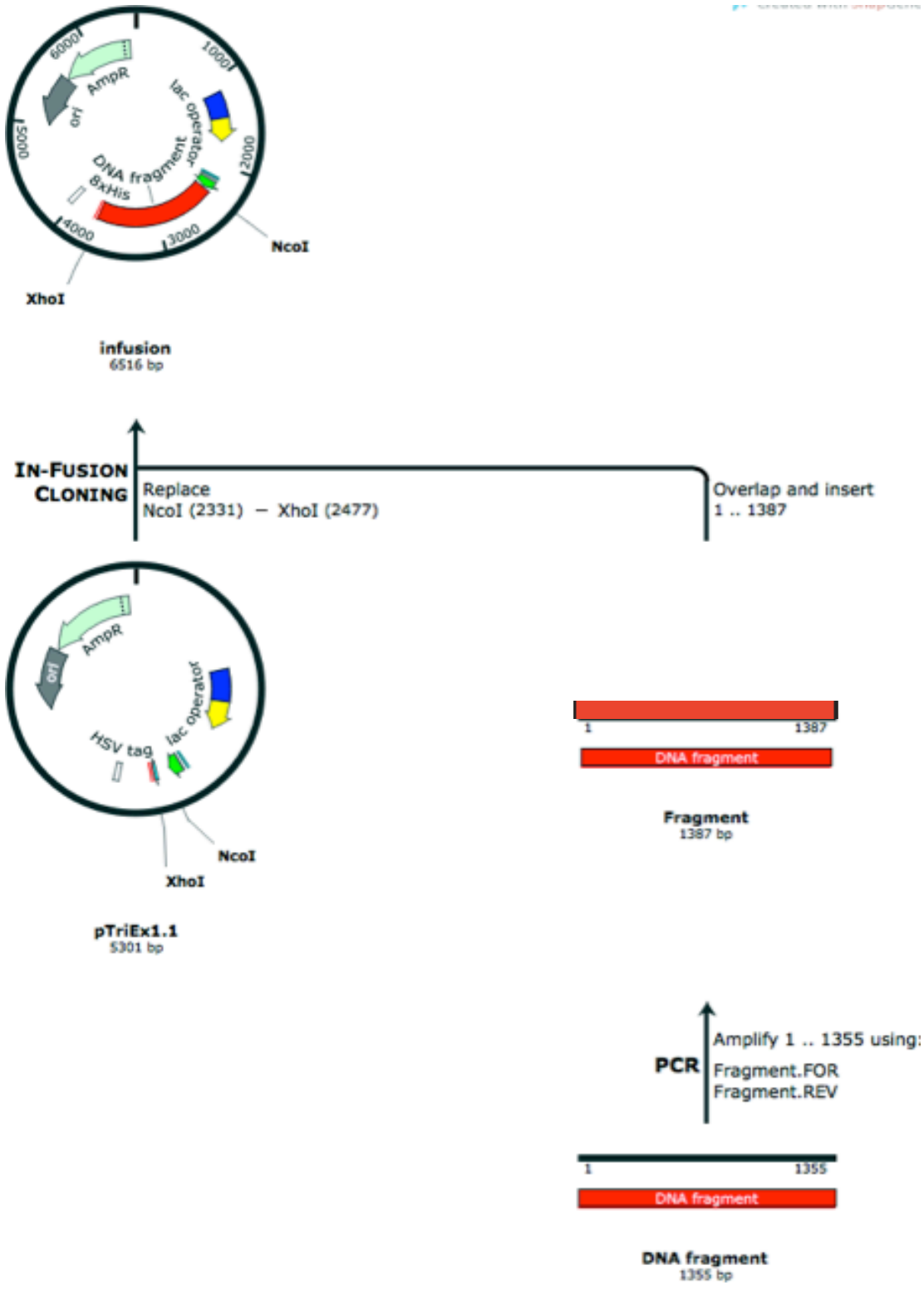

Figure 3.14. Cloning design using In-Fusion protocol. Cloning site located between *Nco*I and *Xho*I sites.

3.2.1 PCR amplification of DNA fragments from cDNA of MHV-A59

The cloning process was carried out by amplifying the gene of interest by PCR. After amplifying with specific forward and reverse primers for each coding region **Table 2.1**, the PCR products were analyzed by agarose gel electrophoresis and the correct sizes of the PCR products (see **Table 9.10**) were compared against a 1kbp DNA ladder, the results of which are shown in **Figure 3.15**, **Figure 3.16** and **Figure 3.17**. The obtained products were purified using gel extraction kit and quantified by a Nanodrop spectrophotometer.

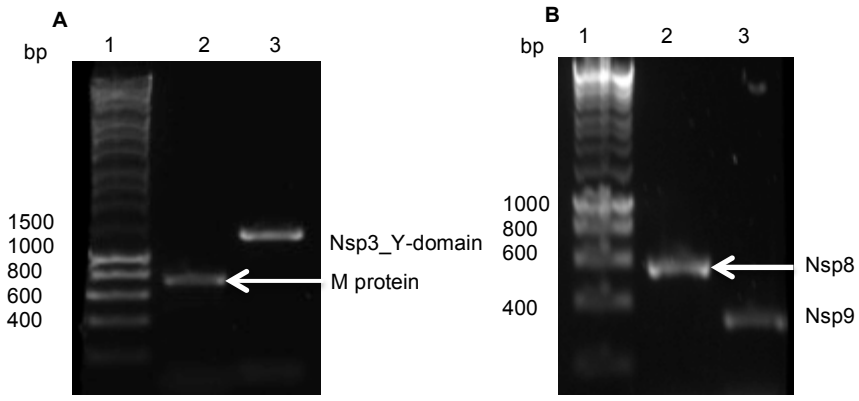

Figure 3.15. Amplification of membrane gene, nsp3_Y-domain, nsp8 and nsp9 from cDNA of MHV-A59. cDNA of MHV-A59 was used as a template to amplify the M, nsp3_Y-domain from nsp3, nsp8 and nsp9 coding regions. **(A)** Lane 1: Marker (1kb) DNA ladder, Lane 2: M protein (687bp), Lane 3: nsp3_Y-domain (1200bp). **(B)** Lane 1: Marker (1kb) DNA ladder, Lane 2: nsp8 (591bp), Lane 3: nsp9 (330bp).

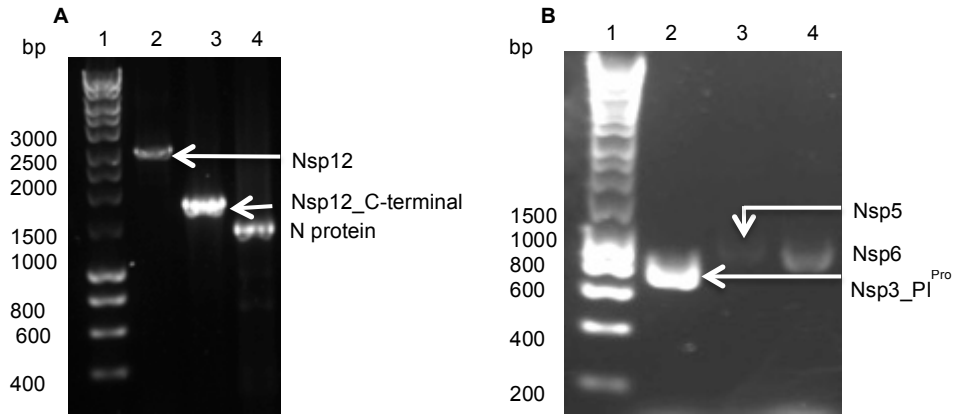

Figure 3.16. Amplification of nsp12, nsp12_C-terminal, N protein, nsp5, nsp6 and nsp3_Pl^{Pro} from cDNA of MHV-A59. cDNA of MHV-A59 was used as a template to amplify nsp12, nsp12_N-terminal, N protein, nsp5, nsp6 and nsp3_Pl^{Pro}. **(A)** Lane 1: Marker (1kb) DNA ladder, Lane 2: nsp12 (2784bp), Lane 3: nsp12_C-terminal (1692bp), Lane 4: N protein (1365bp). **(B)** Lane 1: Marker (1kb) DNA ladder, Lane 2: nsp3^{Pro} (744bp) Lane 3: nsp5 (909bp), Lane 4: nsp6 (861bp).

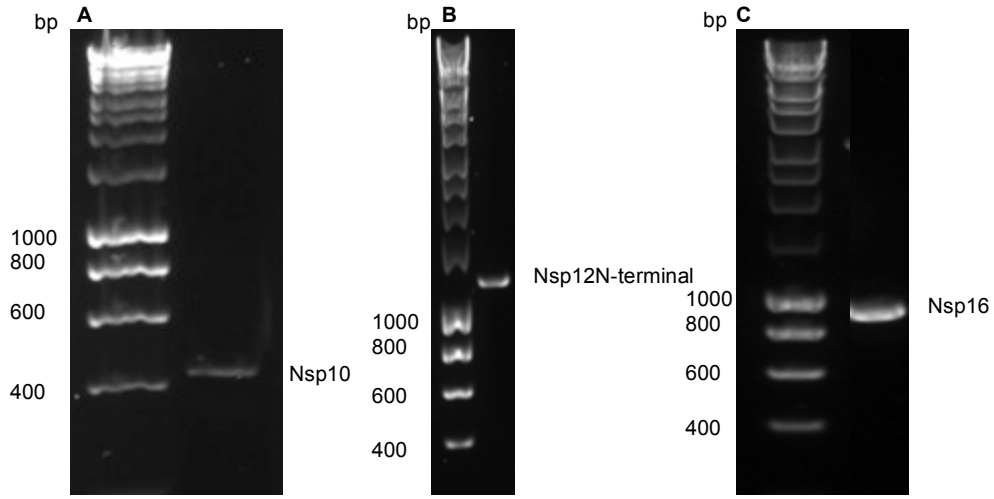

Figure 3.17. Agarose gel electrophoresis of nsp10, nsp12_N-terminal and nsp16 amplified by PCR from cDNA of MHV-A59. cDNA of MHV-A59 was used as a template to amplify nsp10, nsp12_N-terminal and nsp16 **(A)** Lane 1: Marker (1kb) DNA ladder, Lane 2: nsp10 (393bp), **(B)** Lane 1: Marker (1kb) DNA ladder, Lane 2: nsp12_N-terminal (1092bp), **(C)** Lane 1: Marker (1kb) DNA ladder, Lane 2: nsp16 (900bp).

The pTriEx1.1 plasmid was double digested with *Nco*I and *Xho*I restriction enzymes and visualized following gel electrophoresis as in **Figure 3.18**. Linearized vector was purified using a gel extraction kit according to the manufacturer's instructions. Then, each amplified MHV PCR product was combined with the

linearized pTriEx1.1 vector through an In-Fusion cloning reaction to allow efficient and precise recombination of the vector and the desired DNA fragment.

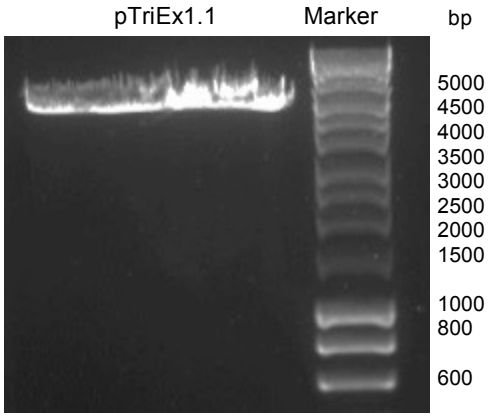

Figure 3.18. Double digest of pTriEx1.1. An agarose gel showing the digested pTriEx1.1. Marker (1kb) ladder, pTriEx1.1 plasmid digested with *Nco*I and *Xho*I restriction enzymes (5155bp).

3.2.2 Transformation of In-Fusion products into competent *E.coli* cells

The In-Fusion reaction mixes were transformed into chemically competent Stellar™ *E. coli* cells and plated on LB agar containing ampicillin to select for transformants. About 30 colonies were selected randomly and screened by colony PCR using either T7 forward or TriExUP as a forward primer and either TriExDOWN or fragment specific primers as the reverse primer. The amplification products were analyzed following agarose gel electrophoresis; **Figure 3.19** shows 13 colonies screened by colony PCR for the N protein gene using T7 forward and Rv_N primers (correct size 1546bp) where all but one of the screened colonies was positive for the correct sized insert.

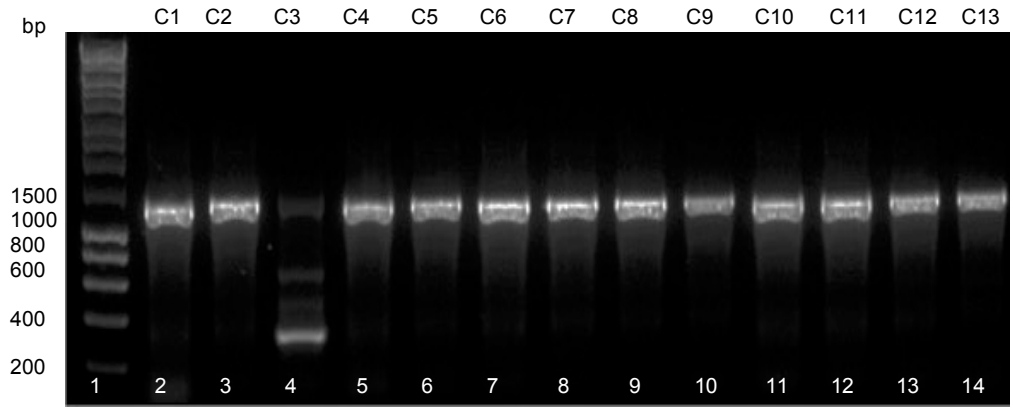

Figure 3.19. N protein screened by PCR. 13 colonies were screened for the presence of N protein insert. Lane 1: Marker (1kb) DNA ladder, Lanes 2-14: colonies numbered from 1-13.

Two positive colonies were transferred into 10 ml LB broth supplemented with ampicillin and grown overnight at 37°C with shaking. Plasmid DNA was purified and a restriction digest performed using *Nco*I and *Xho*I as shown in **Figure 3.20**. Plasmid DNA from the transformants that contained the expected size insert were sequenced via Sanger sequencing at Source BioScience using T7 forward and TriExDOWN primers **Table 2.2**. Alignments were performed in Snapgene software to confirm the expected sequence of each insert and to check that the inserts were in frame with respect to the ATG start site and the C-terminal His tag. A glycerol stock of each positive transformant was prepared and stored at -80°C.

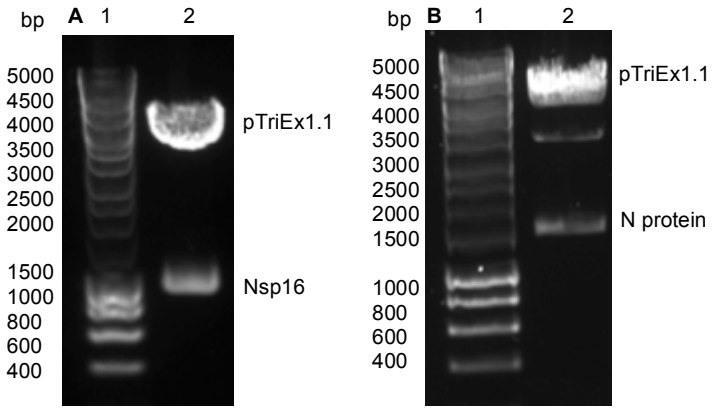

Figure 3.20. Gel electrophoresis of double digested pTriEx1.1 containing nsp16 and N protein. (A) Lane 1: Marker (1kb) DNA ladder, Lane 2: pTriEx1.1_nsp16 double digest with *Nco*l and *Xho*l (900bp), **(B)** Lane 1: Marker (1kb) DNA ladder, Lane 2: pTriEx1.1_ N protein double digest with *Nco*l and *Xho*l (1365bp).

3.2.3 Correction of random mutation in nsp12_N-terminal

All proteins sequences were identical to the sequences obtained from GeneBank (accession number AY700211.1) for MHV-A59 except for the nsp12_N-terminal coding region where the sequence analysis showed a C deletion at position 26. In order to correct this mutation in nsp12_N-terminal, which appeared to be a random error in the cDNA as it was present in all the transformants screened, additional forward (Fw_nsp12N-new) and reverse (Rv_nsp12N-new) primers **Table 2.3** were designed to correct the mutation by overlapping PCR. One original nsp12_N-terminal transformant was amplified by PCR with forward primer (Fw_SacII) and reverse primer (Rv_nsp12N-new) in reaction PCR1. Another PCR was carried out with forward (Fw_nsp12N-new) and reverse primer Rv_12N **Table 2.1,** reaction PCR2. The PCR products were analysed by agarose electrophoresis, and the correct sizes (396bp and 1070bp bp) were obtained for both PCR1 and PCR2 respectively **Figure 3.21**.

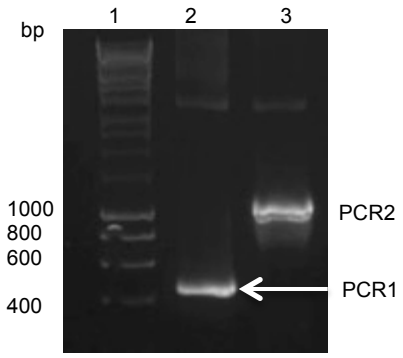

Figure 3.21. Gel electrophoresis of PCR products following amplification of nsp12_N-terminal for one transformant using protein-specific primers. Lane 1: Marker (1kb) DNA ladder, Lane 2: PCR1 (396bp), Lane 2: PCR2 (1070bp).

The DNA products were purified using a PCR purification kit and quantified by Nanodrop sectrophotometer, the concentrations were 195.6ng/µl for PCR1 and 397.7ng/µl for PCR2. Then, 0.1ng/bp PCR from each product were mixed and subjected to PCR in a PCR reaction for five cycles without primers. Then, forward primer (Fw_SacII) and reverse primer (Rv_12N) were added and another thirty cycles were completed. The correct size 1466bp of overlap PCR product was analysed by agarose gel electrophoresis **Figure 3.22**, and extracted by gel extraction kit as before.

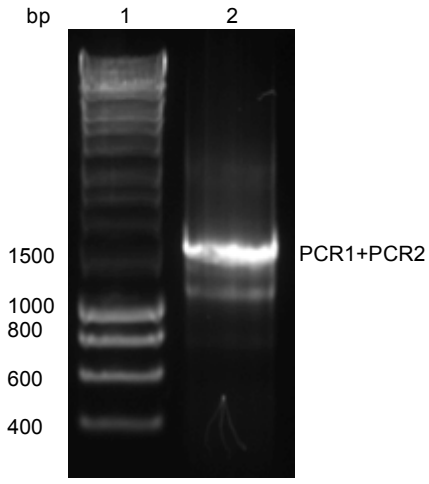

Figure 3.22. Agarose gel electrophoresis of two joined parts of nsp12_N-terminal amplified by PCR. Lane 1: Marker (1kb) DNA ladder, Lane 2: PCR1+PCR2 (1445bp).

The purified product was quantified by Nanodrop spectrophotometer and the concentration was 216.8 ng/µl. The PCR product was first digested with *Sac*II restriction enzyme for one hour at 37°C and then a clean up kit was used to purify the product. A second digest was applied to the PCR product with restriction enzyme *Xho*I for 45 minutes at 37°C. The new digested product was extracted by gel extraction kit after agarose gel electrophoresis and quantified by Nanodrop sectrophotometer, the concentrations was 35.2 ng/µl. Empty pTriEx1.1 plasmid was also digested with *Sac*II and *Xho*I restriction enzymes as described previously. A T4 ligation kit was used for ligation according to the manufacturer`s instruction. The ligation mix was transformed into Stellar^{TM} *E. coli* competent cells with empty plasmid transformed as a control to check the efficiency of transformation. Colony PCR was performed using forward primer and reverse primer as in **Table 2.1** to check for the presence of the nsp12N-terminal protein and three positive colonies with the most intense bands of the expected size were selected **Figure 3.23**.

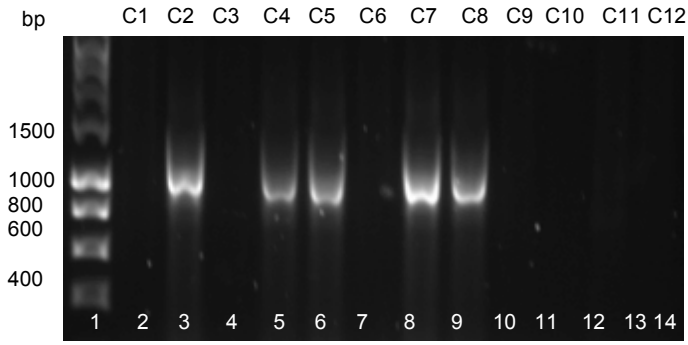

Figure 3.23. Gel electrophoresis of colony PCR of Nsp12N in pTriEx1.1 plasmid. Lane 1: Hyperladder, Lanes 2-13: The colony PCR product corresponding to 1-12 transformants, the correct size of Nsp12N is 1092bp.

Transformants were grown at 37°C overnight in 10 ml LB broth with ampicillin and plasmid DNA extracted as before. Finally the three isolates were sent for sequencing to confirm the sequence correction. A glycerol stock was prepared and samples stored at -80°C freezer.

3.2.4 Protein expression in *E. coli* BL21(DE3)-pLysS competent cells by IPTG induction

All positive plasmids used in this study were transformed into *E. coli* BL21 (DE3)-pLysS competent cells and plated on LB agar plates containing ampicillin. Two colonies selected randomly were grown in 10 ml LB broth with ampicillin (10 µg/ml) and chloramphenicol (34 µg/ml) at 37°C overnight in a shaking incubator. Of these cultures, 200 µl was transferred to a fresh 10 ml LB broth and grown to OD_{600} 0.4-0.6. Protein expression was induced by the addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) and allowed to continue for 4 hours for optimal expression. Both before and after IPTG induction, aliquots of the bacterial culture were harvested for SDS-PAGE analysis. Western blots were performed and bands detected using anti His tag antibody. Cells only and *E. coli* BL21 (DE3)-pLysS carrying the empty pTriEx1.1 vector were used as controls. The induction was

successful for all proteins and the predicted sizes of each expressed protein were detected. Induction of Nsp16 protein showed a clear band (34.3kDa) after one hour induction with IPTG which became less distinct with time **Figure 3.24**.

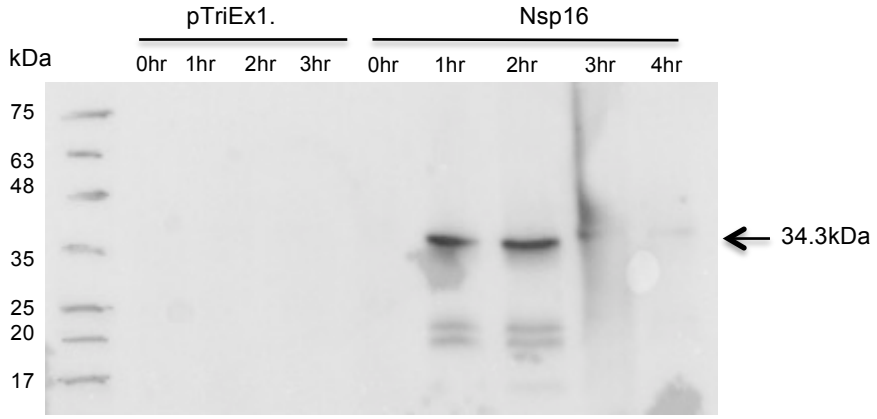

Figure 3.24. Western blot analysis of recombinant expression of nsp16 in *E. coli* BL21(DE3)-pLysS competent cells. Protein sample was extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after 1hr for 4 hours. pTriEx1.1 vector was used as a control.

Similarly, Western blot analysis showed a band of 50.8kDa for the N protein and nothing of the pTriEx1.1 vector used as a control **Figure 3.25**.

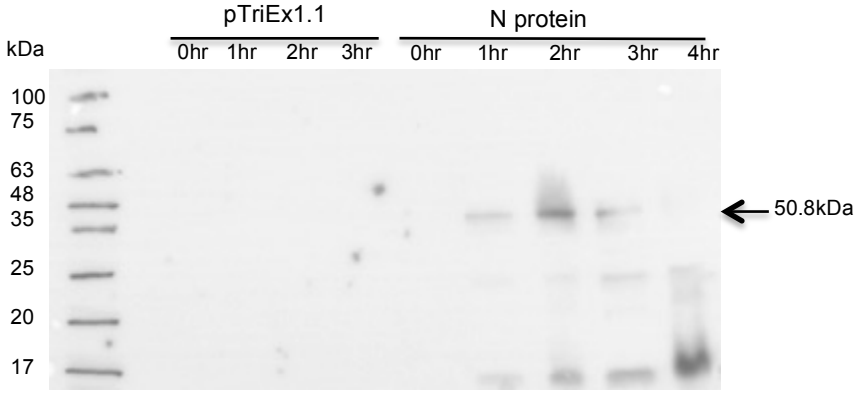

Figure 3.25. Western blot analysis of recombinant expression of N protein in *E. coli* BL21(DE3)-pLysS competent cells. Protein sample was extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after 1hr for 4 hours. pTriEx1.1 vector was used as a control.

Figure 3.26 and Figure 3.27 shows a clear band of 23kDa and 45.3kDa for nsp8 and the Y-domain of nsp3 respectively in Western blot analysis.

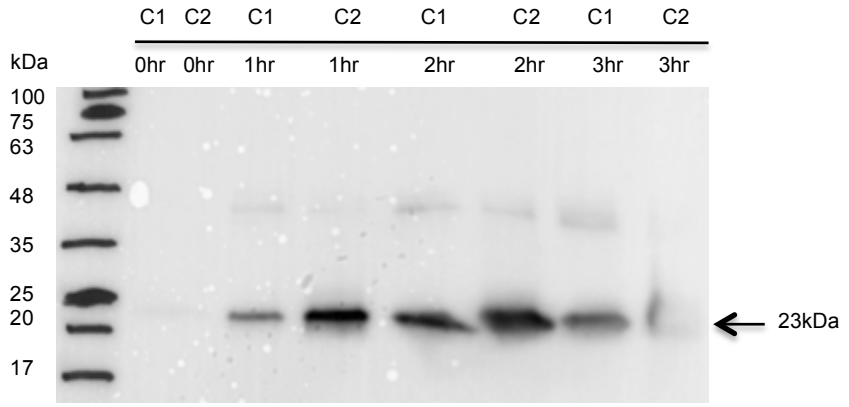

Figure 3.26. Western blot analysis of recombinant expression of nsp8 in *E. coli* BL21(DE3)-pLysS competent cells. Protein sample was extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after 1hr for 3 hours. Nsp8 protein size (indicated by an arrow) was detected after 1hr; C1 and C2 represent two colonies.

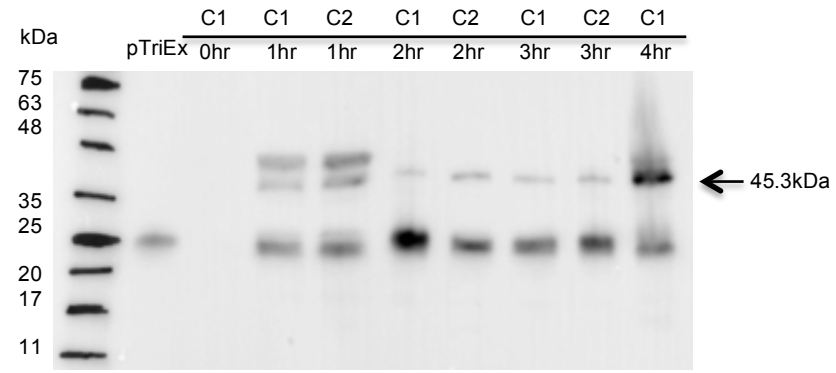

Figure 3.27. Western blot analysis of recombinant expression of nsp3_Y-domain in *E. coli* BL21(DE3)-pLysS competent cells. Protein sample was extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after 1hr for 4 hours. Nsp3_Y-domain protein size (indicated by an arrow) was detected after 1hr; C1 and C2 represent two colonies, pTriEx1.1 vector was used as a control.

The nsp12_C-terminal showed a band of about 68.2kDa in Western blot analysis after one hour induction. Plasmid vector served as a control **Figure 3.28**.

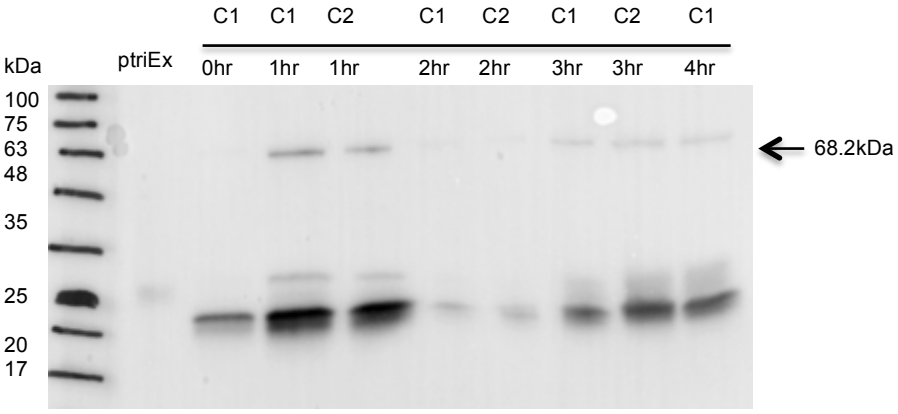

Figure 3.28. Western blot analysis of recombinant expression of nsp12_C-terminal in *E. coli* BL21(DE3)-pLysS competent cells. Protein sample was extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after 1hr for 4 hours. Nsp12_C-terminal protein size (indicated by an arrow) was detected after 1hr; C1 and C2 represent two colonies; pTriEx1.1 was used as a control.

Similarly, Western blotting was performed to detect expression of nsp7 and the result showed a clear band (10.9kDa) after one hour induction in both

transformants analysed **Figure 3.29**. A stronger band at around twice the molecular weight appears to be an *E. coli* protein cross reacting with the antibody as it was also present in some control tracks (e.g. **Figure 3.28**).

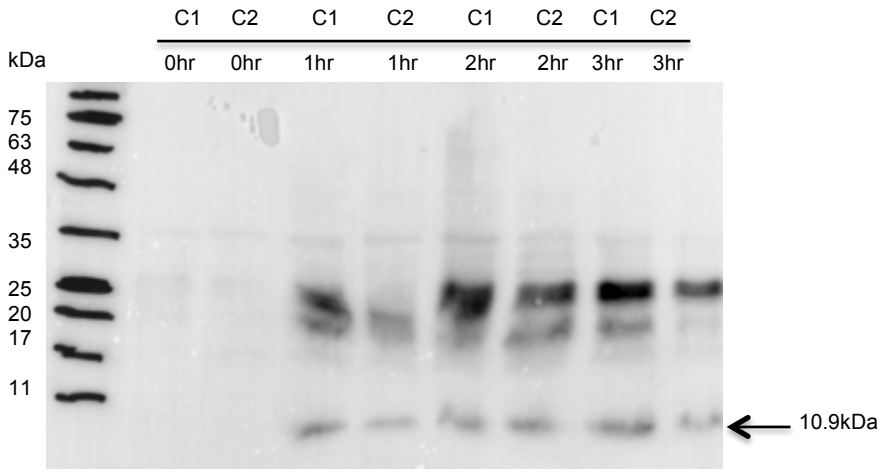

Figure 3.29. Western blot analysis of recombinant expression of nsp7 in *E. coli* BL21(DE3)-pLysS competent cells. Protein sample was extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after 1hr for 3 hours. Nsp7 protein size (indicated by an arrow) was detected after 1hr; C1 and C2 represent two colonies.

Another two MHV-A59 proteins, M protein and nsp3_PL^{Pro} were detected by Western blotting, showing the predicted sizes of 27.1kDa and 29kDa for M protein and nsp3_PL^{Pro} respectively **Figure 3.30 and Figure 3.31** although the M protein migrated close to the *E. coli* contaminant making its identity tentative at this stage.

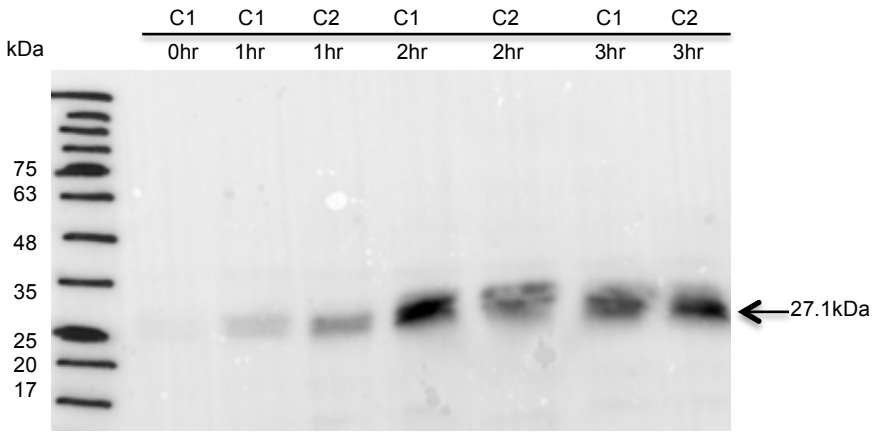

Figure 3.30. Western blot analysis of recombinant expression of membrane protein (M) in *E. coli* BL21(DE3)-pLysS competent cells. Protein sample was extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after 1hr for 3 hours. Membrane gene size (indicated by an arrow) was detected after 1hr; C1 and C2 represent two colonies.

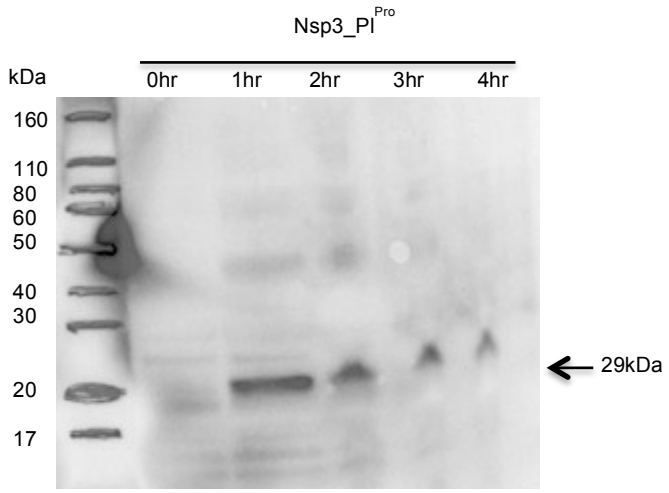

Figure 3.31. Western blot analysis of recombinant expression of nsp3_Pl^{Pro} in *E. coli* BL21(DE3)-pLysS competent cells. Protein sample was extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after 1hr for 4 hours. Nsp3^{Pro} protein size (indicated by an arrow) was detected after 1hr.

Finally, the expression of nsp12_N-terminal and nsp9 were also successful in *E. coli* BL21(DE3)-pLysS and Western blot analysis for both proteins showed clear bands of 43.3kDa for nsp12_N-terminal and 13.3kDa for nsp9 after one hour induction with IPTG **Figure 3.32 and Figure 3.33**.

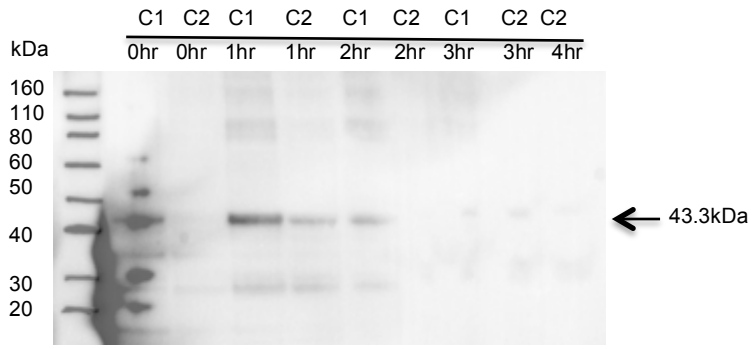

Figure 3.32. Western blot analysis of recombinant expression of nsp12_N-terminal in *E. coli* BL21(DE3)-pLysS competent cells. Protein samples were extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after 1hr for 4 hr. Nsp12_N-terminal protein size (indicated by an arrow) was detected after 1hr; C1 and C2 represent two colonies

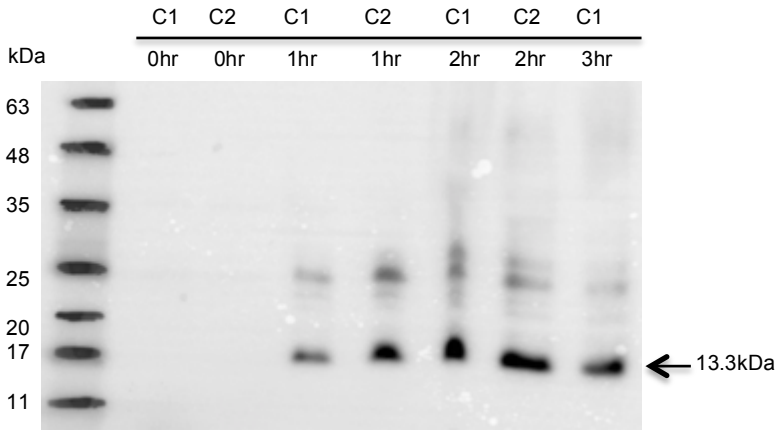

Figure 3.33. Western blot analysis of recombinant expression of nsp9 in *E. coli* BL21(DE3)-pLysS competent cells. Protein samples were extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after 1hr for 3 hours. Nsp9 protein size (indicated by an arrow) was detected after 1hr; C1 and C2 represent two colonies.

3.2.5 Expression and optimization of transfection in mammalian cells

17 clone-1 mammalian cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM). Transfections of the protein expressing constructs were carried out in 6 well plates using a range of different transfection reagents (FugeneHD, Fugene6, Mirus TransIT-LT1, Turbofect, Lipofectamine2000, Lipofectamine3000)

in accordance with the manufacturer's instructions to determine the optimal conditions for plasmid transfection in this cell line. A control vector pTriEx1.1-GFP, which carries the Green Florescent Protein (GFP) gene inserted between the *Nco*I and *Bsu*361 sites in pTriEx1.1 was used to visualize the efficiency of transfection (O'Flynn, 2011).

3.2.5.1 Optimization of transfection reagents

Transfections of 17clone-1 with the selection of transfection reagents showed the best levels of GFP expression from the control vector were obtained using Lipofectamine2000, Lipofectamine3000 and Turbofect **Figure 3.34**.

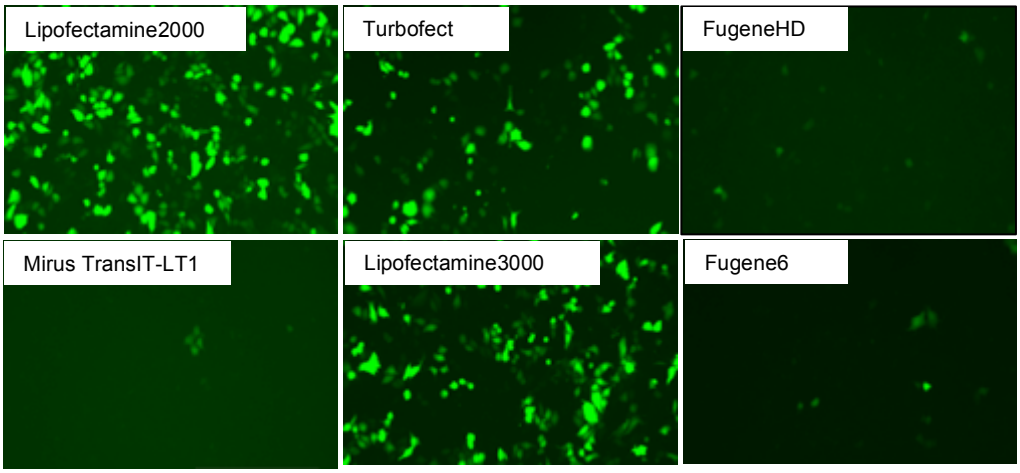

Figure 3.34. *In vitro* transfection of 17clone-1 cells with pTriEx1.1-GFP using different transfection reagents. Cells were grown in equal conditions and transfected with pTriEx1.1-GFP using different transfection reagents. The intensity of GFP was observed under fluorescence microscopy after 24 hours.

A next step of optimization is determining the optimal DNA concentration for each 3 transfection reagent that showed best level of GFP expression and results showed that Lipofectamine2000 and Lipofectamine3000 gave best result with low DNA concentration while Turbofect best result was with high DNA concentration.

3.2.5.2 Optimization of cell density

The initial cell density is an important variable to optimize transfection. Low cell density may result in poor transfection and a long time for expression, leading to a correspondingly low level of expression. Conversely, high cell density may lead to cell overgrowth, which may result in a plateauing of the signal that may obscure a true expression result. To determine the optimal cell density, 24 well plates were seeded with 17clone-1 mammalian cells so that the next day the confluence was 50%, 60%, 70%, 80% and 90% in each well. At these stages of confluency plates were transfected with pTriEx1.1-GFP using Turbofect, Lipofectamine2000 and Lipofectamine3000 transfection reagents as before. The optimal cell density was found to be 60%, while higher cell densities increased the numbers of dead cells, particularly with Lipofectamine2000 and Lipofectamine3000 **Figure 3.35**.

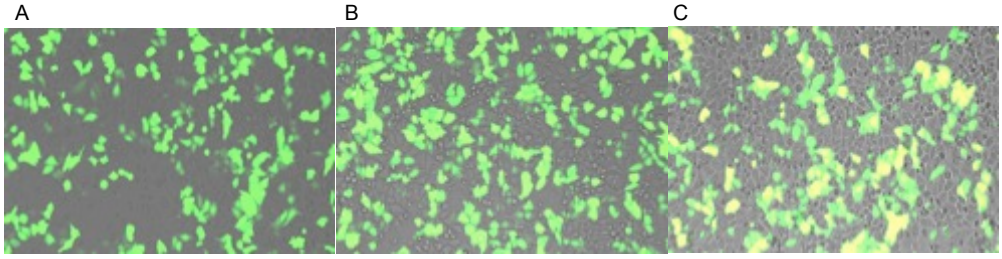

Figure 3.35. *In vitro* transfection of 17clone-1 cells with pTriEx1.1-GFP using different cells density. Cells were grown in different density and transfected with pTriEx1.1-GFP using Turbofect reagent. The intensity of GFP was observed under fluorescence microscopy after 24 hours. Figure (A) represent 50% cells density; figure (B) 60% cells density; and figure (C) 90% cells density. The intensity of GFP was observed under fluorescence microscopy.

3.2.5.3 Optimization of incubation time

It was also necessary to optimize the incubation time with the three most promising transfection reagents, Turbofect, Lipofectamine2000 and Lipofectamine3000. 17clone-1 mammalian cells were seeded in 24 well plates in DMEM containing

10% FBS for 24 hours before transfection. For Lipofectamine2000 and Lipofectamine3000 transfection reagents, the DMEM media was changed to serum free opti-MEM and the transfecting reagent containing DNA was added and incubated for 24 and 48 hours at 37°C. For Turbofect transfection reagent, DMEM media was replaced by serum free DMEM media, then Turbofect reagent containing DNA in serum free DMEM media was added and incubated for 24 and 48 hours at 37°C. The transfection efficiency was evaluated by fluorescence microscopy for GFP as before. pTriEx1.1-GFP showed a good expression signal with Turbofect, Lipofectamine2000 and Lipofectamine3000 after 24 hours incubation and these optimised conditions were then used for transfection of each of the expression positive clones obtained as above. However, disappointingly, when seeded 6 well plates at the optimum density were transfected under optimal consdtions as established using the GFP reporter plasmid, no MHV protein bands were detected for any of the clones tested except for the N protein following SDS-PAGE and Western blot. The N protein was the only protein that showed a clear band of 50.8kDa with the His tag antibody, in this case using Turbofect as the transfection reagent **Figure 3.36**.

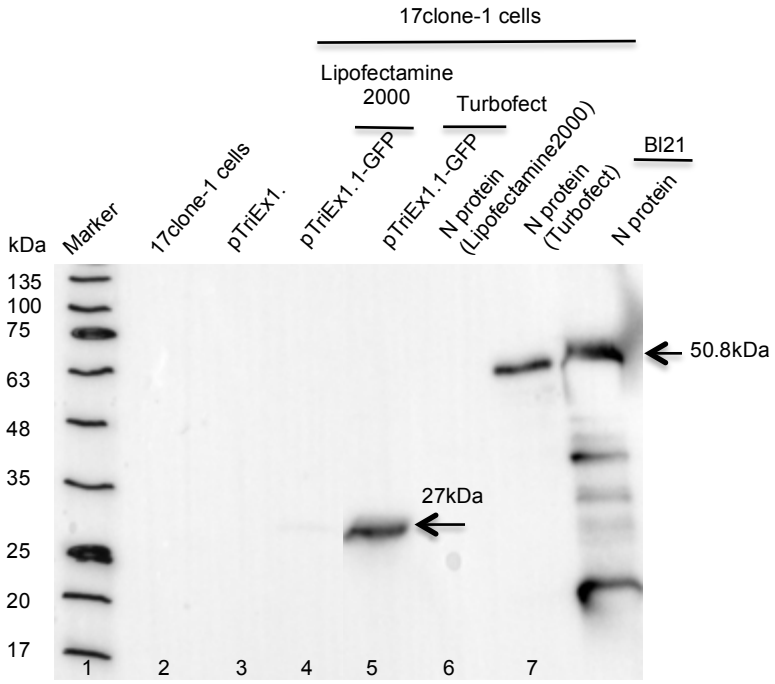

Figure 3.36. N protein expressed in mammalian cells. Cells were grown in equal conditions and transfected with the N protein construct using different transfection reagents, after 24 hours cells they were harvested and total cell extracts were prepared. Protein expression of total cell lysates was detected by Western blot using anti His tag-conjugated antibody. N protein was detected only with Turbofect. Molecular weight marker in kDa is indicated on the left, while N protein in *E. coli* BL21(DE3)-pLysS competent cells was used as a positive control.

The lack of product in Western blot may be due to poor expression or protein instability due to degradation by cellular proteases, either of which make it a challenge to detect the proteins after SDS-PAGE and Western blot. One option was to increase the sensivity of the Western blot using conjugated Anti-His versus unconjugated Anti-His antibody plus a secondary HRP labeled conjugate and to harvesting cell on ice to reduce protease enzyme activity prior to analysis. However, as before the results showed no bands after using conjugated and unconjugated antibody for all proteins except for N protein **Figure 3.37**.

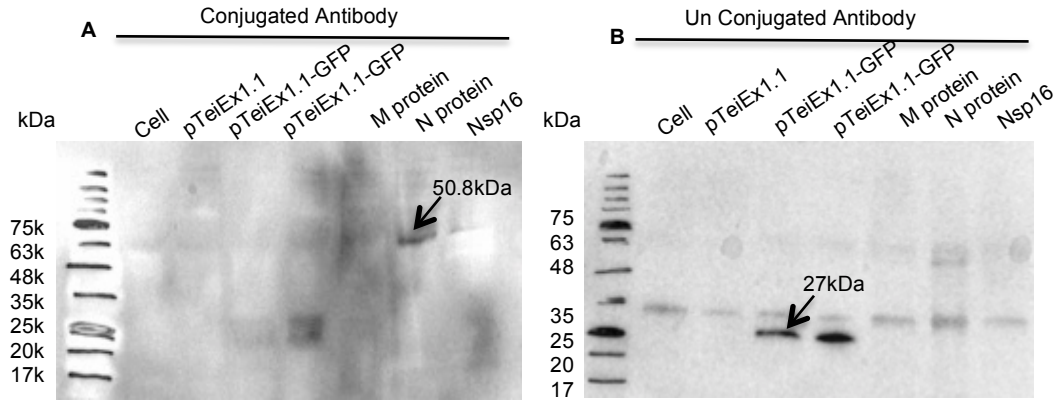

Figure 3.37. Western blot analysis of MHV-A59 proteins using conjugated and unconjugated antibody. Cells were grown in equal conditions and transfected with membrane gene, N protein and nsp16 protein, after 24 hours cells were harvested then total cell extracts were prepared. (A) Protein expression of total cell lysates was detected by Western blot using anti-His tag-conjugated antibody. (B) Protein expression of total cell lysates was detected by Western blot using anti-His tag-unconjugated antibody. Molecular weight marker in kDa is indicated on the left, while cells, pTriEx1.1, pTriEx1.1-GFP were used as a control.

It is possible that the presence of intracellular proteases could affect stability of MHV-A59 proteins in 17clone1 cell line and lead to their degradation following expression. A protease inhibitor, MG132, was used to try and improve the stability of expressed MHV proteins. MG132 protease inhibitor is a peptide aldehyde that has the ability to effectively block the proteolytic activity of the 26S proteasome complex. Mammalian 17clone-1 cells were treated with MG132 protease inhibitor (50 mM) four hours after transfection with pTriEx1.1-nsp3_Y-domain (as a test case) and incubated for 24 and 48 hours. Cells in another plate were treated with MG132 protease inhibitor four hours before transfection and cultured similarly. However, cells harvested from either experiment showed no improvement in the expression of nsp3_Y-domain at 24 or 48 hours post transfection **Figure 3.38**. The addition of MG132 was compatible with expression as a GFP control was positive throughout. A control *E. coli* extract showed a positive signal at the predicted molecular weight on this blot, indicating the nsp3_Y-domain could be detected by the His antibody used in this experiment.

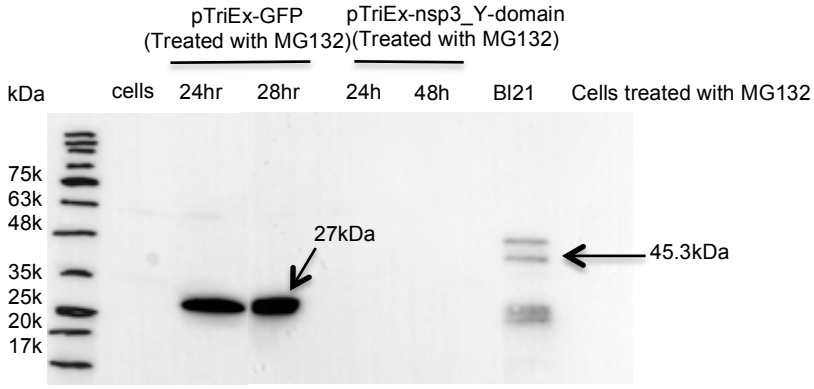

Figure 3.38. Protein inhibitor does not stimulate protein expression. 17 clone-1 cells were grown in a 6 well plate for 24 hours and transfected using Turbofect and pTriEx-nsp3_Y-domain or pTriEx_GFP as a positive control. After 24 and 48 hours cells were harvested then total cell extracts were prepared for analysis of protein expression via SDS-PAGE and Western blot. Molecular weight marker in kDa is indicated on the left. Nsp3_Y-domain in *E. coli* BL21(DE3)-pLysS competent cells was also used as a positive control.

Since there was no improvement in protein expression signal in two optimisation experiments, a further approach assessed was to try using immunoprecipitation to concentrate any low level of expressed proteins present in the transfected lysates in order to increase the expression signal on the blot. However, while this method did improve detection of nsp16 (as an example) expressed in *E. coli* BL21 (DE3)-pLysS it again failed to show any evidence of expression in 17clone-1 mammalian cells **Figure 3.39**.

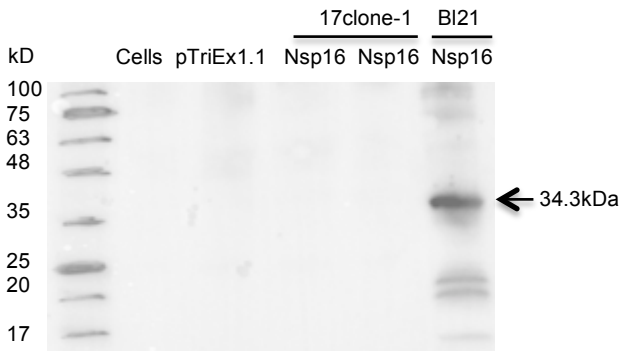

Figure 3.39. Nsp16 protein is not detected by immunoprecipitation. 17 clone-1 cells were grown in a 6 well plate for 24 hours, transfected with Turbofect containing nsp16, after 24 hours cells were harvested then total cell extracts were prepared. Protein expression of total cell lysates was detected by immunoprecipitation method. Nsp16 could not be detected. Molecular weight marker in kDa is indicated on the left, while pTriEx1.1-GFP and nsp16 in *E. coli* BL21 (DE3)-pLysS competent cells was used as a positive control.

3.3 Discussion

This chapter reports the successful cloning of N protein, M protein, non structural proteins (7,8,9,10,16) from gene1, Pl^{Pro}, Y-domain from nsp3, RdRp from C-terminal of nsp12 and N-terminal from nsp12 from cDNA of MHV-A59 virus in the pTriEx vector using the In-Fusion cloning technique to generate clones capable of expression of His tagged fusion proteins. Protein expression was induced in *E. coli* BL21(DE3)-pLysS competent cells using IPTG and the MHV-His-tagged proteins were detected at the expected molecular weights in Western blots for all cloned constructs after one hour induction. The levels thereafter were variable, perhaps indicating issues of protein stability. By contrast with *E. coli*, transfections of the mammalian cell line 17clone-1 showed only a clear band for the N protein.

As a first step to examining the effect of different proteins and genes on virus infectivity, this study set out to obtain expressed recombinant MHV proteins to determine whether the existing presence of these proteins in cells could affect virus infectivity or not. The In-Fusion technique was used to construct the necessary vectors and each amplified gene was cloned and inserted between the *Nco*I and *Xho*I restriction enzyme sites in the pTriEx1.1 vector. The In Fusion technique was very efficient with the majority of clones obtained showing positive inserts. The success of the cloning design was confirmed when all proteins were expressed in *E. coli* after one hour induction with IPTG. Clear bands were detected for most proteins and in some cases expression appeared stable even after 4 hours induction. In others however later expression was variable and in the cases of M protein the validation of expression was confounded by a host cross-reactive band at a similar molecular weight.

The preliminary sequence of the nsp12_N-terminal fragment showed a deletion of one nucleotide (C) at position 26. This mutation may be due to the framshifting events associated with this region, when the ribosome is directed to move into the −1, +1 or +2 reading frames from the reference “0” frame (Harger *et al*., 2002; Stahl *et al*., 2002) and N-terminal of nsp12 is the adjacent domain to the frameshift. An overlap PCR extension (Warrens *et al*., 1997) was utilized to correct this mutation to generate a *bona fide* sequence capable of expression of the nsp12_N-terminal protein. An overlap PCR technique was successfully employed to correct a mutation in a DNA fragment without the need to use a site directed mutagenesis kit. Although both ways provide correction for any mutation, including deletion, the overlap PCR provides a simple alternative giving the same result of site directed mutagenesis kits that are commercially available.

In order to test whether MHV-A59 proteins interfere with virus replication, the detection of transient expression in 17clone-1 mammalian cells is required as a first step. One of the most critical features of expressing proteins in mammalian cells was establishing the most suitable conditions and in this study various conditions were optimized to provide the highest levels of expressed proteins. The first stage of optimization involved determining the best transfection reagent for this cell line and several different transfection reagents were tested. The transfection efficiency was evaluated by GFP expression, which showed the highest fluorescence profiling with Turbofect, lipofectamine2000 and lipofectamine3000. The next stage of optimization involved determining the proper DNA concentration for efficient expression with Turbofect, lipofectamine2000 and lipofectamine3000 transfection reagents. Then, cell density, which is essential to achieve accurate results with different transfection reagents were also tested, as low or high cells density will

affect protein yield, resulting in an unexpressed or low protein yield in mammalian cells. Increasing the incubation time for expression up to 48 hours was also tested. Results indicated that Lipofectamine2000 and lipofectamine3000 showed the highest fluorescence profiling with low DNA concentration but they also showed the highest toxic effect on cells. In contrast, Turbofect, which is a cationic polymer, showed less intensity but was the only reagent to express N protein in 17clone-1 mammalian cells, a permissive cell line for MHV-A59 virus **Figure 3.36**.

Western blot analysis was performed to detect the C-terminal His tag routinely using a directly conjugated anti His tag antibody. Western blot analysis revealed that only N protein was detectable. The undetectable proteins may due to the fact that these reagents were toxic to the cells and cause changes such as cell shrinking, reduced number of mitoses and vacuolization of the cytoplasm (Salvati *et al*., 2006) although no obvious effects were observed following transfection by any construct. Another possible explanation is that the CAG promoter poorly expressed these proteins, although the GFP control was expressed well during transfection reagents optimisiation and used the same promoter. A furthter possibility is that these proteins are inherently unstable.

Conjugated versus unconjugated antibody use to detect any His tag proteins did not improve sensitivity to allow any additional bands to be visualised **Figure 3.37** leaving inherent instability the most likely explanation. In eukaryotic cells, proteins are degraded through the ubiquitin-proteasomal system, which is the major pathway of non-lysosomal proteolysis of intracellular proteins. It plays a critical role in cell cycling, division, differentiation and apoptosis (D. Voges *et al*., 1999; Orlowski, 1999). For better protein detection, tolerated protein inhibitors might have been an option in this study. MG132 was chosen to test whether the presence or

absence of this protein inhibitor would improve the protein expression of the reading frames cloned from MHV. This possibility was strengthened by the fact that proteasome inhibitor MG132 has been shown to block the proteolytic activity of proteasome complex in live cells (Lee and Goldberg, 1998). Transfected cells were treated with MG132 protease inhibitor both before and after 4 hour of transfection but the expressed proteins were again not detected by immunoblotting **Figure 3.38**. However, additional experiment could have been performed as a positive control to confirm that MG132 was functional.

This result suggests that the MHV proteins when expressed alone might be unstable. As noted in the introduction, many of CoV proteins form part of protein complexes in infected cells and may require another partner for stability. It is also notable that the one positive expression was of a structural protein (N protein). Proteins intended as non-structural may have intrinsically short half-lives as they are never required to accumulate to a high level. For example, SARS M protein is strongly detected until 1 h post synthesis but only weakly at 3 h, SARS S protein remains detectable at 12 h however, the E protein has a short half-life of 30 min when expressed in baby hamster kidney (BHK)-21 cell line (Nal *et al*., 2005). Nsp proteins have shorter half lives and some enzymatic activity of some nsp depends on other domains in gene 1. For example the activity of MHV-A59 PLP is affected by mutations to the adjacent ubiquitin-like domain (Ubl-2) and leading to the loss of enzymatic activity over time (half-life of ~30 min at 30°C) (Mielech *et al*., 2015).

In this study, the immunoprecipitation was also performed from whole cell extract to improve the sensitivity of MHV-His tagged protein detection. Trace amount of positive protein are “pulled down” from a cytoplasmic extract that represents many more cells than could be analysed by direct gel electrophoresis. However, this

treatment did not improve detection of any MHV protein **Figure 3.39** further suggesting that the expression was very low, consistent with degradation following translation or at certain steps in the protocol. Concluding from these experiments, no clear explanation was found for the undetected proteins other than intrinsic instability.

There are many challenges in delivering a foreign DNA to mammalian cells, both at the intracellular and extracellular level. In order for a gene or a protein of interest to show an effect on virus replication, they must be able to be in active form for the desired length of time, present in many copies inside the selected cell and not toxic on host cells. In order to overcome the problems in transfection/ infection several parameters of the transfection process were optimized however N protein was the only detected protein. Further experiments could be conducted to improve expression condition such as altering cell line or lysis method, use of immunofluorescence to look for protein expression and coomassie gel or PonceauS to check cellular protein loading and membrane transfer.

In summary, all the genes cloned are expressed in *E. coli*, which indicates the success of the cloning strategy and the functional ability of the clones. However, the efficiency of recombinant protein expression in 17-clone1 mammalian cells depends on different parameters and was most likely thwarted by protein stability. Initial attempts were made to optimize transfection of the mammalian cells by varying the parameters affecting protein expression (cell density/ reagents/ time). However, the optimized conditions failed to increase the detection of any MHV proteins except the N protein. It is clear therefore that an alternative technique is required to increase protein stability, which will be discussed in the next chapter.

4 Inhibition of MHV-A59 infectivity by virus proteins *in vitro*

4.1 Introduction

Although all the MHV proteins selected in this study were expressed in the whole lysates of induced bacterial cells, the levels of expression was poor or not detectable in mammalian cells, with only N protein showing any indication of soluble expression. For maximizing soluble and functional protein yields in mammalian cells, the SUMO-based protein expression system was then selected to try and provide an improvement in the solubility and stability of each recombinant protein. To overcome problems of expressing proteins of interest in different systems like bacteria, yeast, insect and mammalian cells, a number of fusion tag technologies have been developed over the years, such as NusA, maltose binding protein (MBP), glutathione-S -transferase (GST), ubiquitin (UB), thioredoxin (Trx) and Small ubiquitin-related modifier (SUMO) (Catanzariti *et al*., 2004; Malakhov *et al*., 2004; Nilsson *et al*., 1997). Fusion of target proteins to SUMO is used to enhance expression and protect against proteolytic degradation and so facilitate the purification of difficult to express proteins (Changsen *et al*., 1999; De Marco *et al*., 2004; Pryor and Leiting, 1997).

SUMO is a member of a group of proteins that are part of the ubiquitin-like protein family. SUMO and Ubiquitin are very similar in their three-dimensional conformation (Melchior, 2000) however, SUMO shares only ~18% similarity at the sequence level with Ubiquitin. SUMO is absent from prokaryotes (Hochstrasser, 2000; Jentsch and Pyrowolakis, 2000; Müller *et al*., 2001) but present in all eukaryotes and is highly conserved from yeast to humans (Jentsch and

Pyrowolakis, 2000; Melchior, 2000; Müller *et al*., 2001). The first SUMO protein (SMT3) was discovered in 1995 in *Saccharomyces cerevisiae,* which has only one SUMO gene (SMT3) (Meluh and Koshland, 1995). A human homologue was discovered the following year (Mannen *et al*., 1996) and the mammalian SUMO family now has four members (SUMO-1, SUMO-2, SUMO-3 and SUMO-4) (Bohren *et al*., 2004; Saitoh and Hinchey, 2000). The SUMO protein is a 101 amino acid polypeptide (~10-20kDa) in mammals, slightly bigger than ubiquitin (8.5kDa). Although SUMO and Ubiquitin share similar structures and global folding, SUMO differs by the presence of a flexible N-terminal extension (Bayer *et al.*, 1998; Skilton *et al*., 2009) and a different surface charge distribution (Gill, 2004).

In use as an expression tag, the SUMO ORF is fused to the N-terminus of the target ORF and serves to generally protect the overall protein by acting in a chaperone like manner (Malakhov *et al*., 2004). The SUMO fusion system has been described to enhance both expression level and the solubility of proteins that are difficult to express in addition to protecting the protein of interest from proteolytic degradation (Butt *et al*., 2005). In eukaryotic cells, SUMO can be naturally conjugated to hundreds of protein species (Flotho and Melchior, 2013; Wilkinson and Henley, 2010) and is believed to participate in numerous protein/protein and protein/DNA interactions. However, the precise regulatory mechanisms it engenders via conjugation remains unknown in most cases (Flotho and Melchior, 2013).

In eukaryotic cells, SUMO tags are recognized by endogenous SUMO proteases when expressed as fusions and the isopeptide bond formed between the carboxy-terminal glycine of SUMO and the protein is cleaved (Li and Hochstrasser, 2000,

1999). This means that SUMO fusion proteins expressed in mammalian cells quickly have the SUMO domain removed, defeating the objective of the fusion construct. This limitation is eliminated with the development of the universal “SUMOStar” tag, which is a SUMO based tag that is double mutated (R64T/R71E) such that it will not be recognized or cleaved by the activity of cellular SUMO proteases (Peroutka *et al*., 2008).

4.2 Results

4.2.1 Cloning strategy for SUMOStar-MHV fragments in pTriEx1.1 vector

This section describes the strategy applied to generate stable expressed proteins in the pTriEx1.1 vector. The MHV-A59 genes encoding the M protein, the non-structural proteins (5,6,7,8,9,10,16), nsp3_Y-domain and nsp12_C-terminal were inserted into pTriEx1.1 using *Sac*II and *Xho*I restriction sites as described. A SUMOStar encoding fragment (Appendix **Figure 9.78**) was incorporated at the 5' end of each MHV gene using overlap PCR extension. The amplified overlapped PCR fragments containing the SUMO and the MHV ORFs in-frame were digested using *Sac*II and *Xho*I, and cloned into the pTriEx1.1 vector using T4 DNA ligase **Figure 4.40**. The ligation reaction was transformed into chemically competent Stellar^{TM} *E. coli* cells and colonies were selected randomly for screening by colony PCR using TriExUP as the forward primer and TriExDOWN as the reverse primer. The construction of each SUMOStar-MHV fusion protein is described in detail in the next section.

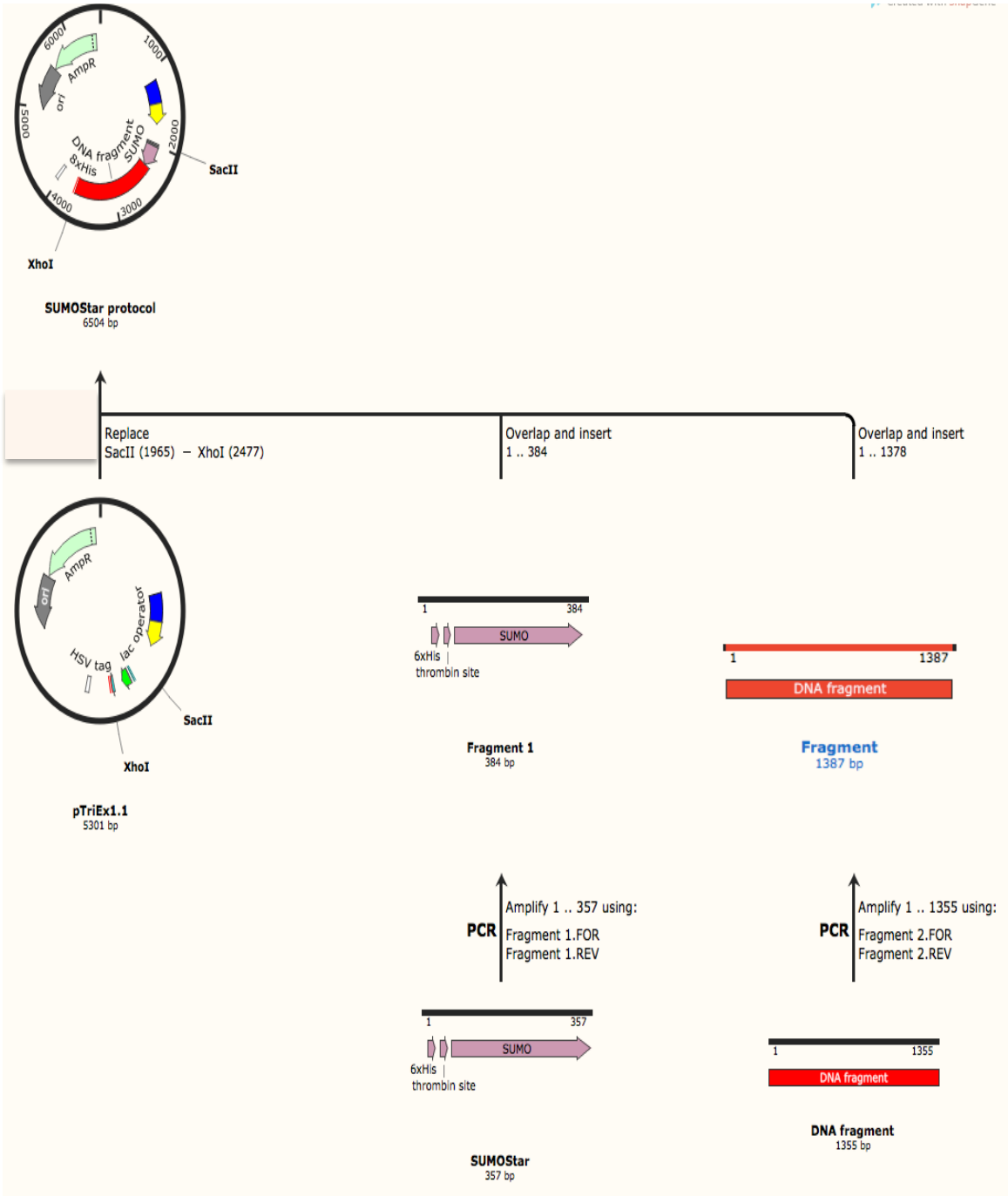

Figure 4.40. Cloning design. Overlap extension was used to incorporate a SUMOStar fragment into each MHV gene and the complete unit inserted between *Sac*II and *Xho*I restriction sites in the pTriEx1.1 vector.

4.2.2 Amplification of MHV-A59 and SUMOStar fragments

The first step was amplifying the protein coding regions of MHV-A59 virus by PCR using specific forward and reverse primers as described in **Table 2.4**. The results of these PCR reactions are shown in **Figure 4.41**. A SUMOStar fragment (795bp) was amplified from a donor plasmid pTriEx-SUMO kindly provided by Dr. Ian Jones with specific (0.5µM) Fw_SacII forward and Rv_SacII reverse primers **Table 2.5 Figure 4.42**. After visualization by agarose gel electrophoresis, PCR products were extracted from the gel.

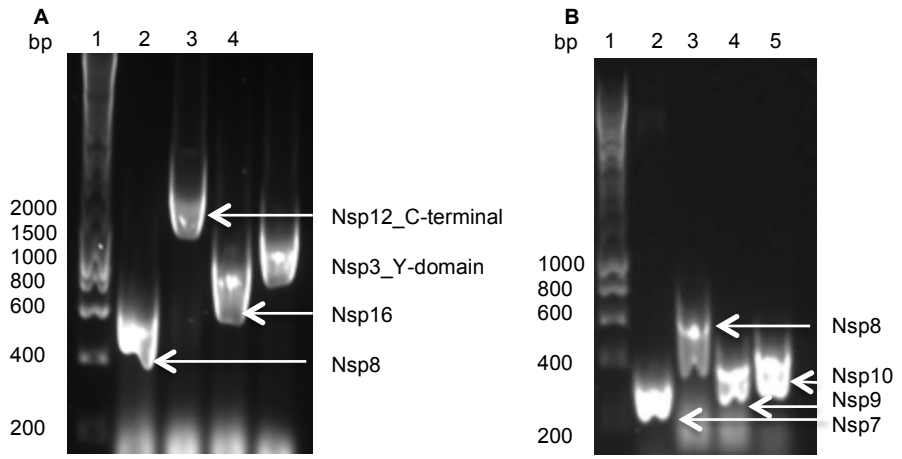

Figure 4.41. Gel electrophoresis of nsp7, nsp8, nsp9, nsp10, nsp16, nsp12_C-terminal, nsp3_Y-domain amplified by PCR from cDNA of MHV-A59. cDNA of MHV-A59 was used as a template to amplify nsp8, nsp16 , nsp3_Y-domain, nsp12_C-terminal, nsp7, nsp9 and nsp10. **(A)** Lane 1: Marker (1kb) DNA ladder, Lane 2: nsp8 (591bp), Lane 3: nsp12_C-terminal (1692bp), Lane 4: nsp16 (900bp), Lane 5: nsp3_Y-domain (1200bp). **(B)** Lane 1: Marker (1kb) DNA ladder, Lane 2: nsp7 (267bp), Lane 3: nsp8 (591bp), Lane 4: nsp9 (330bp), Lane 5: nsp10 (393bp).

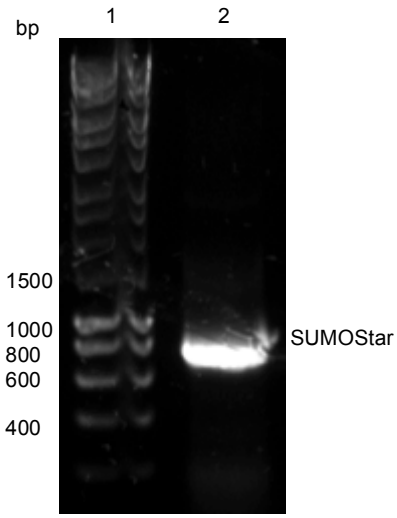

Figure 4.42. Gel electrophoresis of SUMOStar PCR product following amplification of SUMOStar fragment. The SUMOStar fragment was amplified from pTriEx-SUMO vector by PCR. A 1kb DNA ladder molecular marker was used to confirm the correct fragment size. Lane 1: Marker (1kb) DNA ladder, Lane 2: SUMOStar fragment (795bp).

4.2.3 Generation of SUMOStar-MHV fragments by overlap extension PCR.

The PCR reactions were set up in a PCR tube by mixing each PCR product (0.1ng/bp) from each amplified protein coding regions, the SUMOStar fragment (0.1ng/bp), 1x Phusion® High-Fidelity PCR polymerease master mix (BioLabs) with dH_2O to a final volume of 45µl. The mixture was placed in a thermal cycler with the PCR program consisting of five rounds without primers in order to combine the MHV encoding DNA fragment with the SUMOStar fragment. Then, (0.5 µM) Fw_SacII **Table 4.5** and reverse primers for each gene **Table 2.4** were added to the mixture in 2.5 µl volume for each primer. The program continued for 35 cycles and the PCR products were analysed on a 1% agarose gel and extracted using a gel extraction kit.

4.2.4 Ligation of SUMOStar-MHV fragments in pTriEx1.1 vector

For the cloning of the SUMOStar-MHV fragments into the pTriEx1.1 vector it was necessary to treat the cloned fragments and vector with restriction enzymes to produce compatible sticky ends. The SUMOStar-MHV fragments and pTriEx1.1 vector were first digested using *Sac*II (37°C for 60 min) and purified with a clean up kit following the manufacturer's protocol. They were then digested with *Xho*I (37°C for 30 min) and extracted from the agarose gel following gel electrophoresis. The SUMOStar-MHV fragments and linearized pTriEx1.1 vector were mixed at a molar ratio of 3:1 and ligated using T4 DNA ligase.

4.2.5 Transformation in Stellar™ *E. coli* competent cells and restriction digest with *Sac*II and *Xho*I restriction enzymes

The ligation product for each pTriEx1.1-SUMOStar-MHV construct was transformed into Stellar™ competent cells. Colony PCR was performed with the Fw_SacII forward primer and specific reverse primers for each MHV gene **Table 2.4** to check if the vector contained the overlapped PCR product. As an example **Figure 4.43** shows that the SUMOStar-nsp7 fusion ORF was present in 70% of the 20 randomly picked colonies.

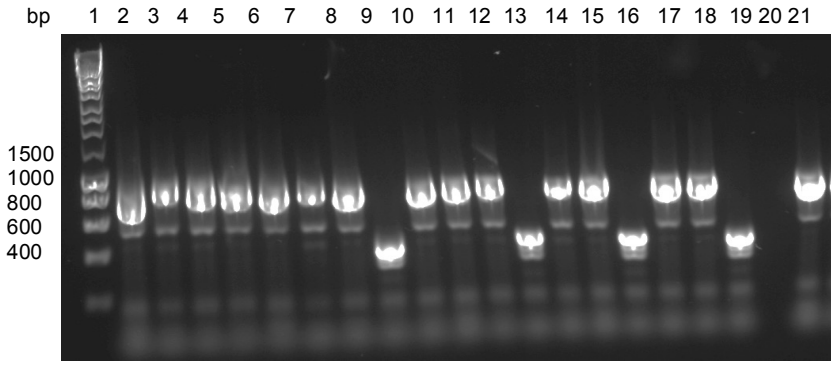

Figure 4.43. Nsp7 screened by PCR. 20 colonies were screened for the presence of SUMOStar-nsp7 insert (1062bp). Lane 1: Marker (1kb) DNA ladder, Lanes 2-21: colonies numbered from 1-20.

To confirm the insertion of the correct fragments, plasmid DNA was extracted from 4 colonies that were positive for the correct sized insert and digested with *Xbal* and *Xho*I restriction enzymes for 30 min and then analysed by gel electrophoresis using a 1% agarose gel at 120v to confirm the expected size fragments **Figure 4.44**. Two plasmids for each transformation were sequenced with TriExUP and TriExDOWN primers to confirm the SUMOStar fusion was in-frame in all cases. Glycerol stocks were prepared for long-term storage at -80°C. In this way SUMOStar fusion constructs were generated for all the MHV target ORFs previously cloned for non-fusion expression.

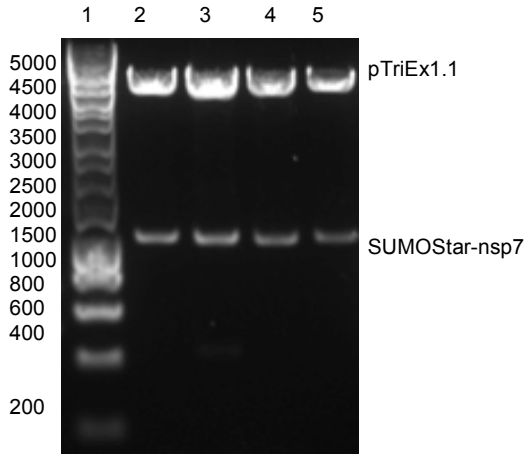

Figure 4.44. Gel electrophoresis of double digest pTriEx1.1 contains SUMOStar-nsp7 fragment. Lane 1: Marker (1kb) DNA ladder, Lanes 2-5: selected pTriEx1.1-SUMOStar-nsp7 transformants double digested with *Xba*I and *Xho*I, pTriEx1.1 (4789bp) and SUMO-nsp7 (1062bp)

4.2.6 Transfection and expression of SUMOStar-MHV fusion proteins in mammalian cells

17clone-1 cells were seeded in 6 well plates at 1.5 x 10^5 cells/well and incubated in 5% CO_2 at 37°C. The next day, cells were transfected with each pTriEx1.1-SUMOStar-MHV constructs as before. For DNA transfection, the optimal transfection reagents (see Chapter 3) Lipofectamine2000 or Turbofect were used according to the manufacture's instructions. After transfection, 17clone-1 cells were incubated for 24 hr, plates were then placed on ice and cells were harvested, transferred into 1.5 ml microcentrifuge tubes and then centrifuged for 2 min. The pellets were either stored at -20°C or used immediately to verify expression of the proteins by Western blotting.

Samples of the cell lysates transfected with Turbofect showed no bands for any of the SUMO tagged protein while cell lysate samples transfected with Lipofectamine2000 showed that fusion of the MHV ORFs to the SUMOStar tag enhanced gene expression in several cases. In particular, the result in **Figure 4.45**

shows that M protein, nsp9, nsp10 and nsp16 were detected for the first time while nsp7 and nsp8 were not detected. The plasmid pTriExSUMO was used as a control.

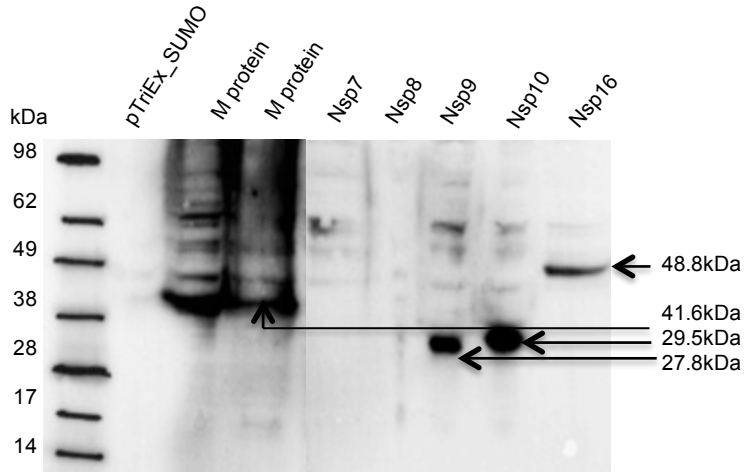

Figure 4.45. Protein expression in mammalian cells using SUMOStar fusion tag. Cells were grown in equal conditions and each well was transfected with nsp7, nsp8, nsp9, nsp10, nsp16 and M protein using lipofectamine2000 transfection reagent, after 24 hours cells were harvested then total cell extracts were prepared. Protein expression of total cell lysates was detected by Western blot using anti-His tag ab. Nsp9, nsp10, nsp16 and membrane were detected. Molecular weight marker in kDa is indicated on the left, while pTriEx-SUMO was used as a control.

4.2.7 Suppression of MHV infectivity by virus proteins

The primary reason to express MHV ORFs in mammalian cells was to establish an assay for the detection of MHV proteins with the potential to interfere with MHV replication when produced *in trans* in infected cells. The M, nsp9, nsp10 and nsp16 plasmid showed positive expression as a SUMO fusion was used to examine the effect on virus replication, based on its ability to protect cells from virus-induced cytopathic effects. The pTriEx1.1-SUMO vector, Lipofectamine2000 and virus only were used as acontrol. 17clone-1 cells were seeded in a 24 well plate and after incubation at 37°C in 5% CO_2, each well was transfected with one protein encoding

construct for 18 hours. The next day, media was removed from each well and washed twice with PBS and the monolayers were infected with MHV-A59 virus at a MOI of 3. The MOI reflects the ratio of the number of virus particles that will enter the target cell, which is a statical process, and according to Poisson distribution, when MOI increase, the percentage of cells infected with at least one virus will increase also. With MOI of 3, there are a percentage of 95% of target cells will be infected with one virus. After incubation at 37°C for 45 min, the inoculums were removed and the monolayers were washed twice with PBS. Then, the plate was incubated with fresh DMEM media for 16 hr at 37°C. The plasmid pTriEx1.1-SUMO was used as a control. As shown in **Figure 4.46**, there was some general interference in virus induced killing by the assay design as the cytopathic effect caused by MHV infection of 17clone-1 mammalian cells in non-transfected cells was different from the cytopatic effect of the virus in the presence of transfected pTriEx1.1-SUMO. The virus yield from all transfections was used as a quantitative measure of replication inhibition by using plaque assay.

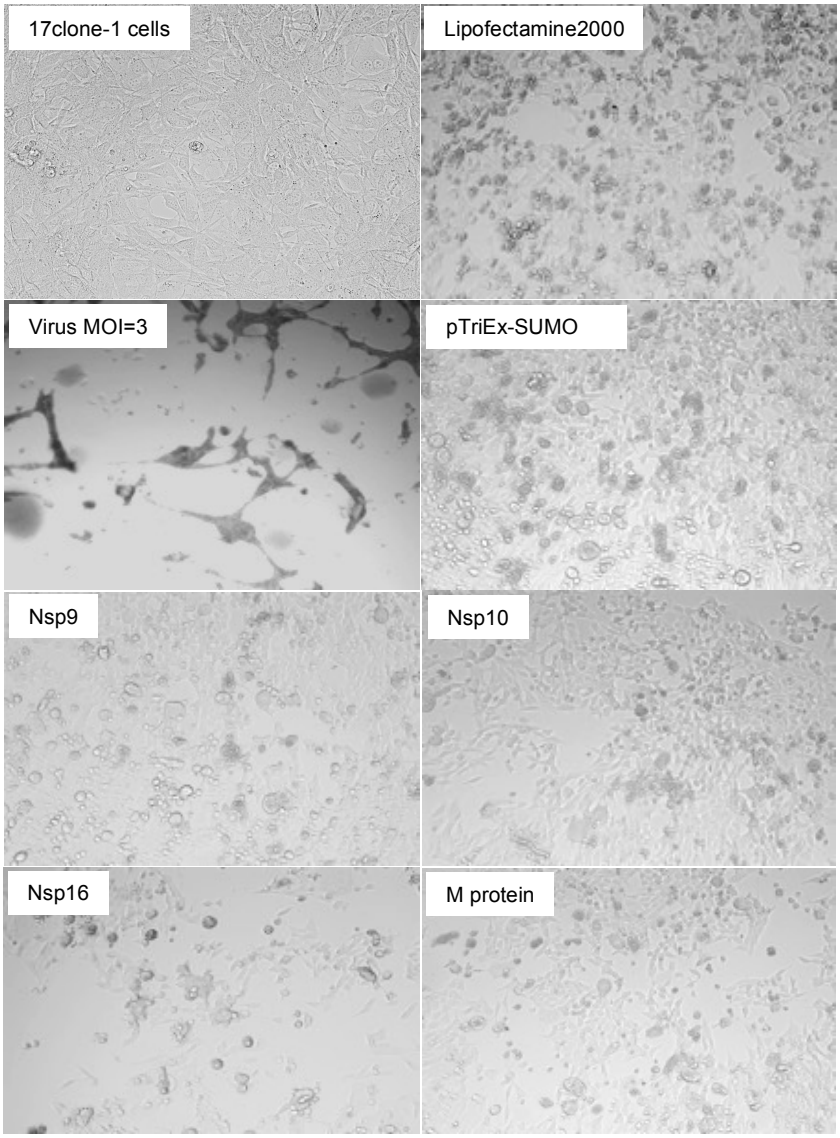

Figure 4.46*. In vitro* effect of transient expression of MHV-SUMO fusion proteins on MHV virus replication. Cells were grown in equal conditions and each well was transfected with nsp9, nsp10, nsp16 and membrane protein using lipofectamine2000 transfection reagent, after 24 hours cells in each well were infected with MHV-A59 at MOI=3 except control. After 16 hours incubation, supernatants were collected and virus titre was measured by plaque assay.

To do this, the supernatants were removed from each transfected/infected well and used for virus titre determination by plaque assay and this experiment was repeated in triplicate to ensure reproducability.

The result of the plaque assay of three indepanant experimants shows that the expression of M protein, nsp9, nsp10 and of nsp16, as SUMOStar fusion proteins

did not reduce virus titre significantly when the replicates were taken into account. The *p*-value from each sample compared to the control (pTriEx1.1-SUMO vector) were >0.05 **Figure 4.47**.

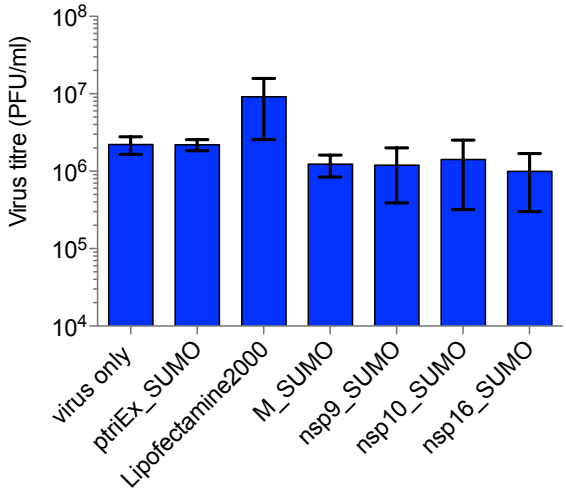

Figure 4.47. Effect of the MHV-A59 proteins on the virus replication *in vitro*. Plaque titrations were carried out on 17clone-1; monolayers were transfected with SUMOStar-MHV-A59 plasmids, 24 hours after transfection, monolayers were infected with MHV-A59 at MOI=3, 16 hours post infection, supernatants were collected and virus titers were measured with plaque assay. Lipofectamine2000, pTriEx-SUMO were used as a control. Error bar represent mean and SD from three independent experiments.

4.3 Discussion

Coronavirus proteins are multifunctional and play a crucial role at different stages during the viral replication cycle. The data here show that transient expression of M and nsp16 MHV proteins in 17clone-1 cells appears to reduce the cytopathic effect caused by MHV-A59 *in vitro* but that a reduction in virus titre was not significant.

As described in the previous chapter, the expression of functionally active forms of many MHV-A59 proteins in 17clone-1 mammalian cells was difficult to achieve at detectable levels, with the exception of the N protein. Yet to evaluate the potential inhibitory effect of MHV-A59 proteins expressed *in trans* in 17clone-1 mammalian cells requires the presence of the expressed proteins during a replication cycle. It was found that use of the SUMOStar tag was benefical to demonstrate expression of a number of MHV-A59 proteins.

Non-structural proteins nsp9, nsp10, nsp16 and M protein of MHV-A59 were expressed at detectable levels using the SUMOStar system in this project **Figure 4.45**. However, the SUMOStar system did not result in production of detectable levels of nsp3_PL^{Pro}, nsp5, nsp6, nsp7, nsp8, nsp12_C-terminal, nsp12_N-terminal, or nsp3_Y-domain in 17clone-1 mammalian cells. There is a possibility that the SUMOStar tag was not suitable for these proteins due to their poor solubility, slow folding or instability. As discussed, it may also be the case that these proteins are produced at low level in mammalian cells but are maintained below the detection threshold by proteolysis. It has been reported that SARS 3CL protease, N and S protein can be expressed with SUMO-fusion technology in *E. coli* strain BL21 (DE3) (Zuo *et al*., 2005). The results here indicate that some MHV-A59 proteins can also be expressed in mammalian cells using the SUMOStar tag. Rescuing the expression of MHV-A59 proteins in mammalian cells, as presented in this study, is

a first step towards studying the effect of these proteins during the virus replication cycle. Preliminarly data suggested that some of the virus proteins might interfere with cytopathic effect although no consistent reduction in infectivity of the virus was demonstrable.

The virus titre appeared to be unaffected by the presence of expressed MHV proteins, implying no effect on the viral replication cycle. A trend in all the replicated experiments was that transient expression of MHV M protein appeared to reduce virus induced cytopathic effect. Thus, it is possible that the constant presence of M protein may interfere with replication by interacting with other virus proteins in the early stages of the viral life cycle before virus encoded M is synthesised. Another possibility is that nsp16 and M protein may interact and bind with virus proteins that participate in virus formation. From the virus titre analysis **Figure 4.47**, the presence of M, nsp9, nsp10 and nsp16 protein had no signifinant inhibitory effect on MHV-A59 virus infectivity.

M protein is the most abundant constituent of the virus outside the virion and is responsible for virus shape (De Haan *et al*., 2000), in association with other virus components. The expression of M protein and E protein alone are sufficient for the formation of virus-like particles (VLPs) in coronaviruses (Bos *et al*., 1996; Corse and Machamer, 2000; Vennema *et al*., 1996). M protein also interacts with N protein through its endodomain (Escors *et al*., 2001; Fang *et al*., 2005; Hurst *et al*., 2005; Kuo and Masters, 2002; Luo *et al*., 2006b; Verma *et al*., 2007, 2006). Cryo-EM and tomographic reconstructions studies found that the M protein endodomain of MHV and SARS-CoV virions are connected to the nucleocapsid via a single thread-like connection (Bárcena *et al*., 2009; Neuman *et al*., 2006). It has

also been reported that the M protein is a multifaceted molecule that interacts with diverse proteins such as S protein (de Haan *et al*., 1999; McBride and Machamer, 2010), E protein (Corse and Machamer, 2003; Venkatagopalan *et al*., 2015) and also M-M interaction (Kuo and Masters, 2010). The interactions of M protein with other viral proteins and possibly cellular proteins may be consistent with the idea that an excess of this protein, especially at early times of the replication cycle, may bind with other proteins of the virus and slow virus growth or lead to the formation of defective viruses. It has been shown that SARS-CoV and MERS-CoV M protein could suppress type I IFN production in infected HEK293 cells (Lui *et al*., 2016; Siu *et al*., 2009), leading to evasion of the host innate antiviral response although this immunosuppressive activity was not observed for the M protein of HCoV-HKU1 (Siu *et al*., 2014), implying that this inhibitory activity is not found in all coronaviruses.

Another protein that is important in virus replication is the nsp16 protein. This protein is responsible for coronavirus capping (Snijder *et al*., 2003), interacts with nsp10 in SARS-CoV (Bouvet *et al*., 2010) and has other functions (reviewed in (Neuman *et al*., 2014)). It has been demonstrated that in some cases, the presence of exogenously expressed coronavirus proteins such as N protein can increase virus replication (Cui *et al*., 2015). However, evidence is lacking in regard to whether nsp16 can function as an inhibitory or stimulatory factor. No significant impact of nsp9 and nsp10 was observed on the virus titre. Although the data revealed no potential role for M, nsp9, nsp10 and nsp16 proteins as inhibitors, further experiments are required to confirm this and to clarify the mechanism involved. Using fluorescence and coomassie stain could be conducted to check whether these proteins are expressed or not and check also for half life of these

proteins.

In conclusion, a revised recombinant expression strategy employed in this work was designed to express MHV-A59 proteins as SUMOStar fusion proteins, in order to rescue detectable levels of expression and enable an investigation of the potential inhibitory effect of these proteins on virus replication. Successful expression of nsp9, nsp10, nsp16 and M proteins in mammalian cells was used as the basis for inhibitory effects by transfection of each protein coding plasmid with infection by the wild-type MHV virus in 17clone-1 cells. The results showed no significant effect of these proteins on virus replication.

In light of potential future studies, introducing reandom mutations by EP-PCR for M, nsp9, nsp10 and nsp16 proteins would be the next step for screening/selection for dominant negative variant(s). Another approach could be studying the effect of mutations introduced to nsp10 and nsp16 by using Gel-shift mobility assay since nsp10 and nsp16 interact with each other and this step is fundamental for the capping of the viral genome. The overarching aim of the work presented within this chapter was to investigate potential inhibitory effect MHV-A59 proteins on virus replication.

5 Fitness landscape of N protein variants

5.1 Introduction

Since N protein was the only protein well expressed in 17clone-1 mammalian cells (as described in chapter 3), it was the only real candidate to investigate the potentially inhibitory effects of non-mutated and mutated MHV protein on virus replication, based on their ability to protect cells from virus-induced cytopathic effects. The N protein is an important component for both replication and transcription by coronaviruses. It is a highly basic protein, encoded near the 3' end of the coronavirus genome, with a molecular weight ranging from 45 to 60kDa among the coronavirus groups and is synthesized in large amounts during infection (Lai and Cavanagh, 1997; Stohlman and Lai, 1979). The N protein binds with the viral RNA and forms the nucleocapsid to protect the viral genome from degradation and, following S mediated fusion, the nucleocapsid enters the host cell to facilitate replication (Almazán *et al*., 2004; Baric *et al*., 1988; Grossoehme *et al*., 2009; Tylor *et al*., 2009). However, the exact mechanisms and role of N in replication are poorly defined. The N protein is an important diagnostic marker for coronavirus disease and induces host immune responses (Leung *et al*., 2004; Xiong *et al*., 2004). Antibodies against N protein strongly inhibit RNA transcription *in vitro* (Compton *et al*., 1987), while coronavirus replication is strongly stimulated by inclusion of N protein at an early step of infection (Schelle *et al*., 2005). N protein was identified to be dynamically associated with the RTC (Verheije *et al*., 2010). This transient association may allow N protein molecules to change places with one another during transcription/replication and to carry out a variety of distinct functions. Reflecting its overall basic charge, N protein was also found to have nonspecific binding activity toward nucleic acids, including ssRNA, ssDNA, and

dsDNA (Takeda *et al*., 2008; Tang *et al*., 2005). At the level of amino acid sequences, coronavirus N proteins are quite diverse across the *nidovirales,* explaining their role as serospecific markers of infection. Coronavirus N proteins contain three domains: an N-terminal RNA-binding domain (NTD), a C-terminal dimerization domain (CTD), and a Ser/Arg (SR) rich linker (Chang *et al*., 2013; I.-J. Chen *et al*., 2013; Lo *et al*., 2013). The crystal structures have been solved for the N- terminal domain (NTD) of SARS (Saikatendu *et al*., 2007), infectious bronchitis virus (IBV) (Fan *et al*., 2005; Jayaram *et al*., 2006), HCoV-OC43 (I.-J. Chen *et al*., 2013), and mouse hepatitis virus (MHV) (Grossoehme et al., 2009). The N-terminal domain (NTD) of coronavirus N protein contains several critical residues for RNA binding and virus infectivity (Keane *et al*., 2012; Saikatendu *et al*., 2007; Tan *et al*., 2006) but the structural and mechanistic basis for RNA binding and ribonucleoprotein complex (RNP) formation remain unclear. The central region of the N (Hurst *et al*., 2013; Snijder *et al*., 2003) protein contains an RNA-binding region and the primary phosphorylation sites (Chang *et al*., 2006; Wootton *et al*., 2002). Studies have identified that this region in MHV binds with the ubiquitin-like domain of nonstructural protein 3 (nsp3), and that this is an important interaction in early coronavirus replication (Hurst *et al*., 2013; Snijder *et al*., 2003). The C-terminal domain (CTD) is known to be a critical determinant for recognition of the genomic RNA packaging signal in MHV N protein (Kuo *et al*., 2014), mediating N-N dimerization as well as longer-range interactions in the nucleocapsid (Chang *et al*., 2013). The spacer B is a linker that joins the CTD with the N3 domain, this domain is an acidic domain which binds to the endodomain of the M protein during virion assembly (Hurst *et al*., 2005; Kuo *et al*., 2016; Verma *et al*., 2007, 2006).

The N3 domain at the carboxy terminus is found to be a site for N-M interactions

suggested to be required to form virions (Hurst *et al*., 2005; Kuo and Masters, 2002; Luo *et al*., 2006a; Verma *et al*., 2007, 2006) but other studies have disagreed with this conclusion (Fang *et al*., 2006; Hatakeyama *et al*., 2008; He *et al*., 2004).

A library of variants is required to study any potentially inhibitory effect of mutated virus proteins and directed evolution can be used to generate and screen for such potential dominant negative variant(s). The term "directed evolution" or "*in vitro* evolution" refers to a general strategy that mimics natural evolution in the laboratory (Zhao *et al*., 2004). Selection of the most effective method to generate molecular diversity is an important step in an *in vitro* evolution experiment. Two evolutionary methods are used commonly for *in vitro* directed evolution experiments, gene recombination and random mutagenesis (Arnold, 2001; Bershtein and Tawfik, 2008; Chaput *et al.*, 2008; Deshler, 1992; Fox and Huisman, 2008; Hou, 2009; Jin *et al*., 2009; Johannes and Zhao, 2006; Kotzia and Labrou, 2009; Xu *et al*., 2008; Yu *et al*., 2008). The first method depends on the exchange of any blocks of sequences among two or more DNA strands. By contrast, random mutagenesis depends on introducing random point mutations into the whole gene. Error-prone PCR (ep-PCR) has been widely used to generate libraries with random mutations by using the low fidelity of DNA polymerase and changing conditions in standard PCR experiments so that they favor error during the polymerization reaction. In bacteria there are three main polymerase families, A, B and C. Most family A and B bacterial polymerases are composed of a single subunit and have three domains: the 3'-5' exonuclease (proofreading) domain, the 5'-3' exonuclease domain and the polymerase domain. The combination of domains will affect the proofreading ability of the polymerase. *Taq* polymerase from *Thermus aquaticus* (*Taq-Pol*) lacks 3'-5' exonuclease activity and exhibits a higher error rate than *Pfu*

polymerase from *Pyrococcus furious* (*Pfu-Pol*). As a result, the error rate of *Taq* DNA polymerase (measured at 2.0 x 10^{-5} per nucleotide per cycle) is usually tenfold higher than the error rate of *Pfu* DNA polymerase (1.6 x 10^{-6} per nucleotide per cycle). Because of its higher fidelity, *Pfu* DNA polymerase is often the preferred polymerase in PCRs used in molecular cloning experiments (Lundberg *et al*., 1991). Early error-prone PCR experiments were proposed in 1989 (Leung *et al*., 1989) and later modified by Cadwell and Joyce in 1992 (Cadwell and Joyce, 1992) who used *Taq* DNA polymerase to introduce random mutations at each PCR cycle. In standard PCR reactions, mutations occur at different places each time and the rate is usually reasonable but in error-prone PCR, the mutation rate becomes relatively higher and occurs earlier in the cycles.

There are several ways to modify the mutation rate over and above the endogenous rate of the polymerase. One of the easiest ways is to decrease the fidelity of the DNA polymerase furhter by including Mn^{2+} in the PCR reaction mixture, which reduces base pairing specificity (Beckman *et al.*, 1985; Lin-Goerke *et al.*, 1997). For example, adding $MnCl_2$ to 0.7 mM final concentration can increase the error rate up to 25-fold (Beckman *et al*., 1985). Another way to modify the mutation rate is to use different ratios of nucleotides in the reaction mixture leading the polymerase “hunting” for the correct nucleotide and so more likely to make a mis-incorporation (Fromant *et al*., 1995; Nishiya and Imanaka, 1994; Shafikhani *et al*., 1997). Increasing or decreasing the number of cycles can also generate a library of variants. Mutation frequencies from 0.11-2% are achieved by varying the nucleotide ratio and the amount of $MnCl_2$ (Cadwell and Joyce, 1992; Zhao *et al*., 1999). Accurate estimation and control of the mutation rate during

error-prone PCR has become more and more exact (Wang *et al*., 2000), but there are few comparative studies of different random mutagenesis strategies for their efficiency in producing mutations for directed evolution experiments.

Directed evolution also requires screening and selection methods for library analysis and these will depend on the new molecule(s)' properties of interest. Mutation rate, population size and selection strategy are all parameters that can potentially effect and alter the outcome of directed evolution experiments and an additional consideration when developing strategies for antiviral drugs based on mutated virus proteins is their potential toxicity on the physiology and viability of the transduced cells. Non-mutated or mutated proteins can affect cellular properties and cell growth in an unpredictable manner and even subtle differences in cell growth rates can result in a major impact on different biological assays during *in vitro* experiments. The 3-(4,5-dimethylthiazol-2-yl)-2,5- diphenyltetrazolium bromide (MTT) cell assay is a simple and convenient method to judge general toxicity. This assay is a colorimetric assay used to measure *in vitro* cytotoxicity and cell proliferation following exposure to test compounds. The MTT assay is used in mammalian cell studies to measure cell activation and viability (Gerlier and Thomasset, 1986; Price and McMillan, 1990), cell growth and survival (Ficken *et al*., 1991; Mosmann, 1983), the bactericidal activity of macrophages (Ferrari *et al*., 1990; Peck, 1985) and fungal viability (Levitz and Diamond, 1985). The MTT assay is based on the ability of the mitochondrial enzyme, succinate-dehydrogenase, of viable cells to reduce the MTT tetrazolium salt to purple formazan crystals when added to the cell culture. The level of MTT formazan produced is therefore correlated with the viability of cells (Denizot and Lang, 1986), i.e. when there are live cells in the culture, active reductase enzymes will convert MTT tetrazolium salt

to MTT formazan. The absorbance of the formazan product can be measured with a spectrophotometer, high absorbance values result from more live cells in a culture over time.

The aim of this chapter is to generate a library of variants using error-prone PCR, to test the effect of some variants on virus replication and to investigate mutated N protein toxicity on 17clone-1 cells by the MTT assay.

5.2 Results

5.2.1 Plaque assay of virus yield in the presence of N protein expressed *in trans*

As previously described in chapter 3, N protein was expressed following In-Fusion cloning of the N ORF into the pTriEx1.1 plasmid vector. Expression of N protein was detected in BL21 (DE3)-pLysS and in 17clone-1 cells by Western blotting **Figure 3.25 and Figure 3.36**. To examine the effect of N protein on virus replication, 17clone-1 cells were seeded in 24 well plates and transfected with the plasmid encoding N protein. The next day, monolayers were infected with MHV-A59 virus at an MOI of 3; incubated for 45 min, washed with PBS and then incubated with fresh DMEM media. Images were taken at 18, 24 and 48 hours post infection **Figure 5.48** before supernatants were removed to quantify viral yield by plaque assay. The empty plasmid was used as a control to check whether the presence of the plasmid affect virus replication irrespective of N expression. The data show that the observed cytopathic effect of the untransfected cells infected with MHV is different from the cytopathic effects caused by the virus in 17clone-1 cells expressing MHV N protein **Figure 5.48**.

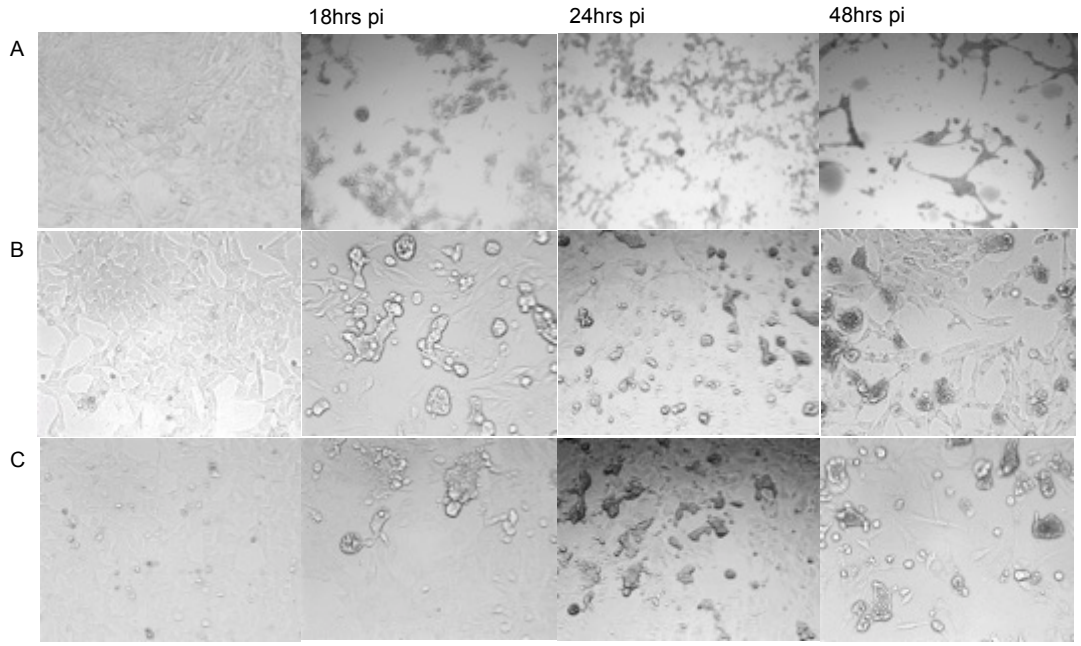

Figure 5.48. N protein effect on MHV-A59 virus replication. 17 clone-1 monolayers were transfected with N protein and empty vector. After 24 hours post transfection (0 hr), cells were infected with MHV-A59 virus MOI=3 and imaged at 18, 24 and 48 hours post infection. (A) Untransfected 17clone-1 cells, (B) 17clone-1 cells transfected with pTriEx1.1 vector as a control, and (C) 17clone-1 cells transfected with pTriEx1.1-N protein.

The result of the plaque assay showed that the presence of N protein increased virus infectivity when expressed whereas the plasmid vector only had no effect on virus replication **Figure 5.49**. The *p*-value was significant after 24 hrs post infection (0.00198936) confirming that N protein was expressed during that period and after 48 hrs, the *p*-value was (0.00296086). As N protein expression has an observable effect on MHV replication, it was taken as the basis for library generation with a view to the isolation of mutants with increased interference in MHV replication

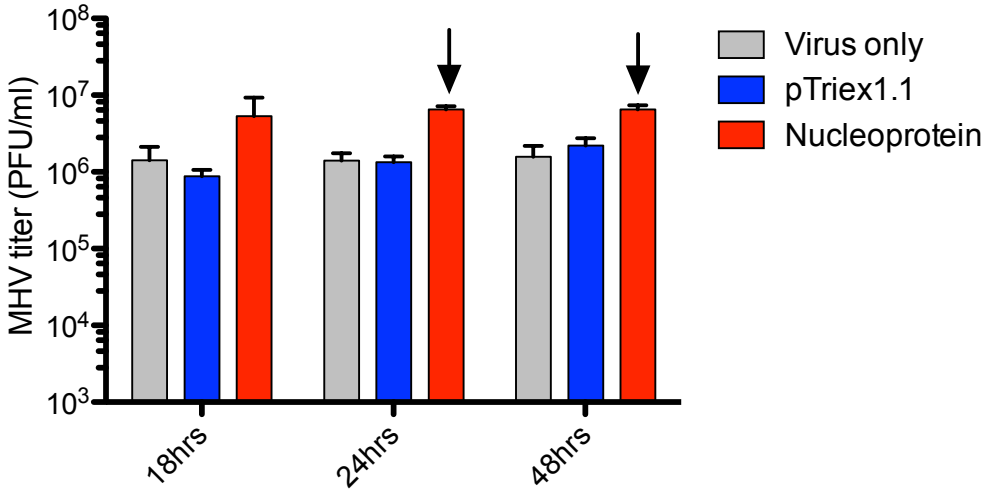

Figure 5.49. Affect of N protein expression on MHV replication in 17clone-1 mammalian cells. Titrations were carried out on 17clone-1 cells; monolayers were transfected with N protein in triplicate, 24 hours after transfection, media was removed, washed twice with PBS, monolayers were infected with MHV-A59 at MOI=3, inoculums removed after 45min and monolayer washed with PBS twice, fresh media added, supernatants were collected after 18, 24 and 48 hours post infection, and virus titer was measured with plaque assay. Empty vector and cells were used as a control. The arrows represent $p<0.05$.

5.2.2 Error prone PCR

The first step in the construction of a library of N protein variants was to determine which PCR amplification protocol to use to achieve a low and evenly distributed rate of mutations in the N protein coding region. To do this error-prone PCR reactions were carried out using two different protocols in the presence of $MnCl_2$. The first protocol was carried out with *Taq* polymerase; different concentration of nucleotides and $MnCl_2$ were added immediately before the thermal cycling was initiated. This protocol has been reported to result in a mutation rate of ~0.66% of the nucleotide positions in the DNA template (Cadwell and Joyce, 1992). PCR products were analyzed by agarose gel electrophoresis **Figure 5.50**.

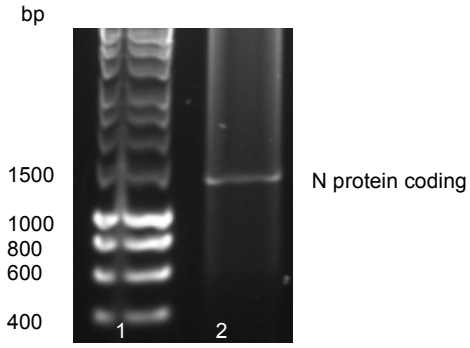

Figure 5.50. Gel electrophoresis of N protein amplified by first error-prone PCR protocol from cDNA of MHV-A59. cDNA of MHV-A59 was used as a template to amplify N protein using first error-prone protocol. Lane 1: Marker (1kb) DNA ladder, Lane 2: N protein (1365bp).

To characterise the PCR products for the level of mutation, the amplified band was extracted from the gel and cloned into *Nco*I-*Xho*I digested pTriEx1.1 vector and transformed into Stellar competent cells (Clontech) as described in section 2.3. Ten randomly selected colonies were screened by colony PCR for the presence of the N protein coding region using T7 forward primer and Rev_N primer **Table 2.1** the exact size predicted is 1546bp **Figure 5.51**.

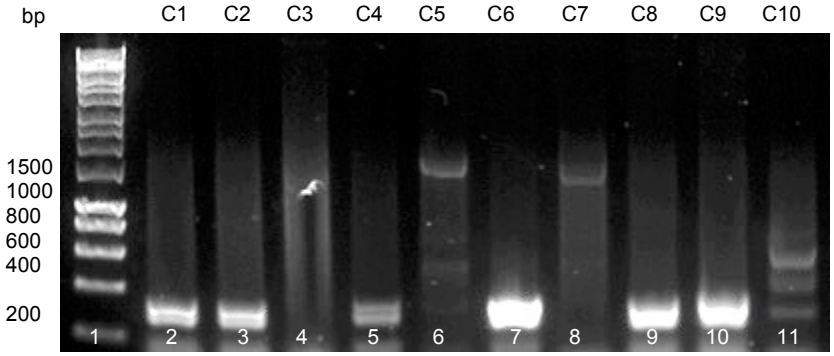

Figure 5.51. N protein screened by PCR. 10 colonies were screened for the presence of N protein insert. Lane 1: Marker (1kb) DNA ladder, Lanes 2-11: colonies numbered from 1-10.

The result showed that only two out of 10 colonies (colony 5 and colony 7) carried a band consistent with the expected size of the N protein ORF, the remainder were much smaller plasmids, consistent with deletions of the cloning vector. This result indicated that the efficiency of the ep-PCR reaction for single mutations within the N ORF was low and that many errors resulted in unclonable fragments or fragments that lost sequence once transformed into *E. coli*. However, a second error prone PCR protocol was carried out using reagents from the In-Fusion HD Cloning Kits (Clontech) mostly according to the manufacturer's instructions but with slight modifications. Before the thermal cycling was initiated $MnCl_2$ was added to the reaction in different concentrations (50 µM, 100 µM, 200 µM, 300 µM, and 400 µM $MnCl_2$). As previously described, section 2.1.5, the PCR protocol was completed in the normal way and the PCR products were analyzed by agarose electrophoresis as shown in **Figure 5.52**.

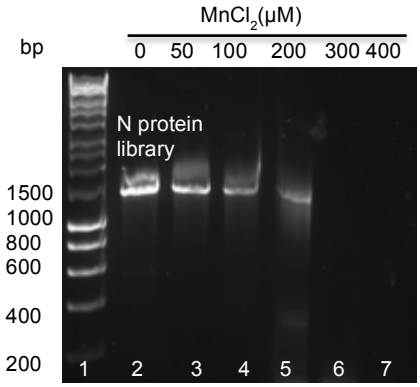

Figure 5.52. Agarose gel electrophoresis of N protein amplified by second PCR protocol from cDNA of MHV-A59. cDNA of MHV-A59 was used as a template to amplify N protein using different concentrations of $MnCl_2$. Lane 1: Marker (1kb) DNA ladder, Lane 2: N protein with 0 µM $MnCl_2$(control) ,Lane 3: N protein with 50 µM $MnCl_2$, Lane 4: N protein with 100 µM $MnCl_2$, Lane 5: N protein with 200 µM $MnCl_2$, Lane 6 : N protein with 300 µM $MnCl_2$, Lane 7 : N protein with 400 µM $MnCl_2$.

The PCR amplified band encoding N was clearer in these amplifications and so the bands amplified from error prone libraries of the 50 µM, 100 µM and 200 µM $MnCl_2$ conditions were all gel extracted, cloned by In-Fusion ligation-independent recombination, independently transformed into Stellar competent cells and cultured *en masse* in LB broth with ampicillin overnight. The next day, libraries were purified using a miniprep plasmid isolation kit. All libraries were double digested with *Nco*I and *Xho*I restriction enzymes as demonstrated in section 2.1.6. The results showed that as the $MnCl_2$ concentration increased the intensity of the band encoding N fell suggesting a lower rescue rate for replication competent plasmids with an increased error rate **Figure 5.53**, a result, in principle, similar to that obtained with the initial ep-PCR conditions.

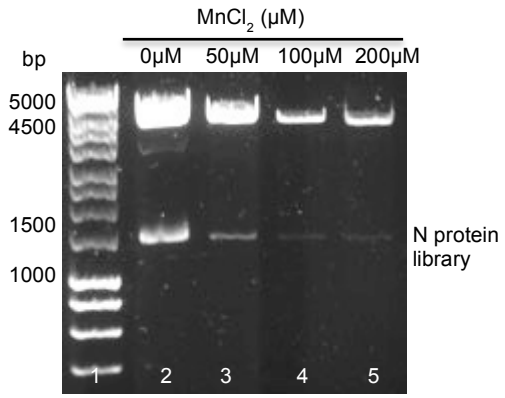

Figure 5.53. Gel electrophoresis of error-prone libraries double digest containing N protein. Error-prone libraries were double digested with *Nco*I and *Xho*I. Lane 1: Marker (1kb) DNA ladder, Lane 2: N protein with 0 µM $MnCl_2$ library (control), Lane 3: N protein with 50 µM $MnCl_2$ library, Lane 4: N protein with 100 µM $MnCl_2$ library, Lane 5: N protein with 200 µM $MnCl_2$ library.

Colonies were selected randomly from platings of the error prone libraries (50 µM, 100 µM and 200 µM $MnCl_2$) on LB agar plates containing ampicillin and screened by colony PCR for presence of the full length N protein ORF. The result showed that a band expected of the N protein ORF was present in 10 colonies when 50 µM

$MnCl_2$ was added to the PCR reaction and present in 9 colonies when 100 µM $MnCl_2$ was added to the PCR reaction. However, in the 200 µM $MnCl_2$ library a band consistent with the N protein sequence was absent in all 16 colonies. The In-Fusion cloning reaction described in chapter 3 was also used as a control with a unmutated amplified N sequence to compare the effiency of N protein region cloning without $MnCl_2$ and every colony screened had the expected band size **Figure 5.54**.

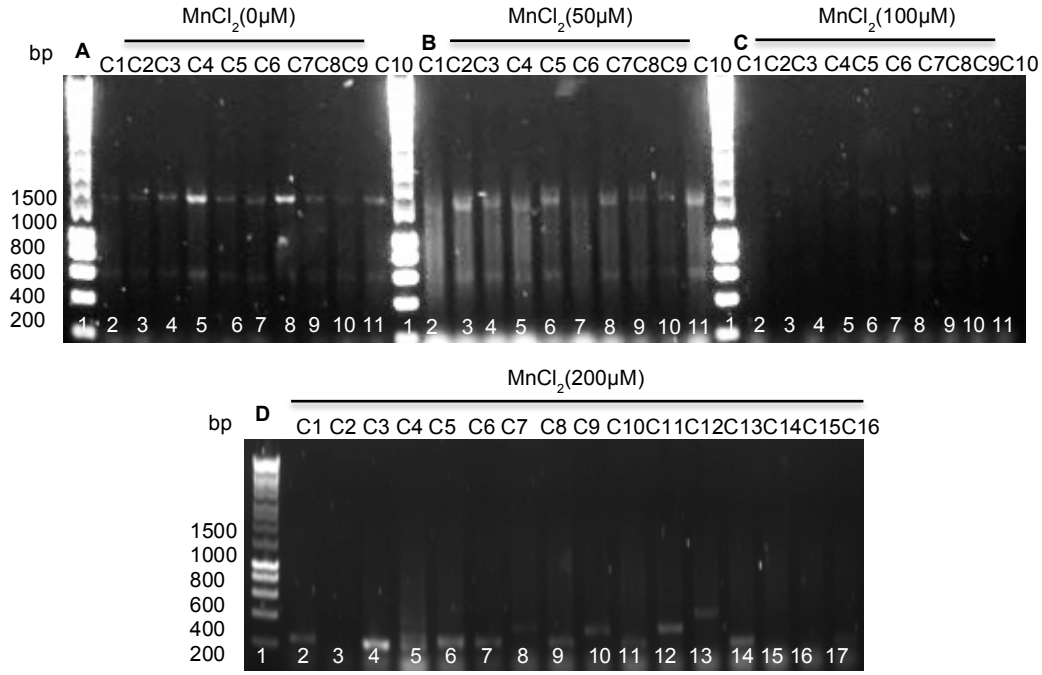

Figure 5.54. Error-prone libraries screened by colony PCR. Randomly selected colonies were screened for the presence of N protein insert in error-prone libraries. **(A)** Lane 1: Marker (1kb) DNA ladder, Lanes 2-11: colonies numbered from 1-10 of 0 µM $MnCl_2$ library (control), **(B)** Lane 1: Marker (1kb) DNA ladder, Lanes 2-11: colonies numbered from 1-10 of 50 µM $MnCl_2$ library , **(C)** Lane 1: Marker (1kb) DNA ladder, Lanes 2-11: colonies numbered from 1-10 of 100 µM $MnCl_2$ library, **(D)** Lane 1: Marker (1kb) DNA ladder, Lanes 2-17: colonies numbered from 1-16 of 200 µM $MnCl_2$ library.

5.2.3 Construction of library variant

To know the mutation frequency of the 50 μM $MnCl_2$ and 100 μM $MnCl_2$ libraries plasmids were re-isolated from 2 colonies that had proved positive for the correct sized insert containing a N protein ORF insert from both libraries and sequenced **Figure 5.55**.

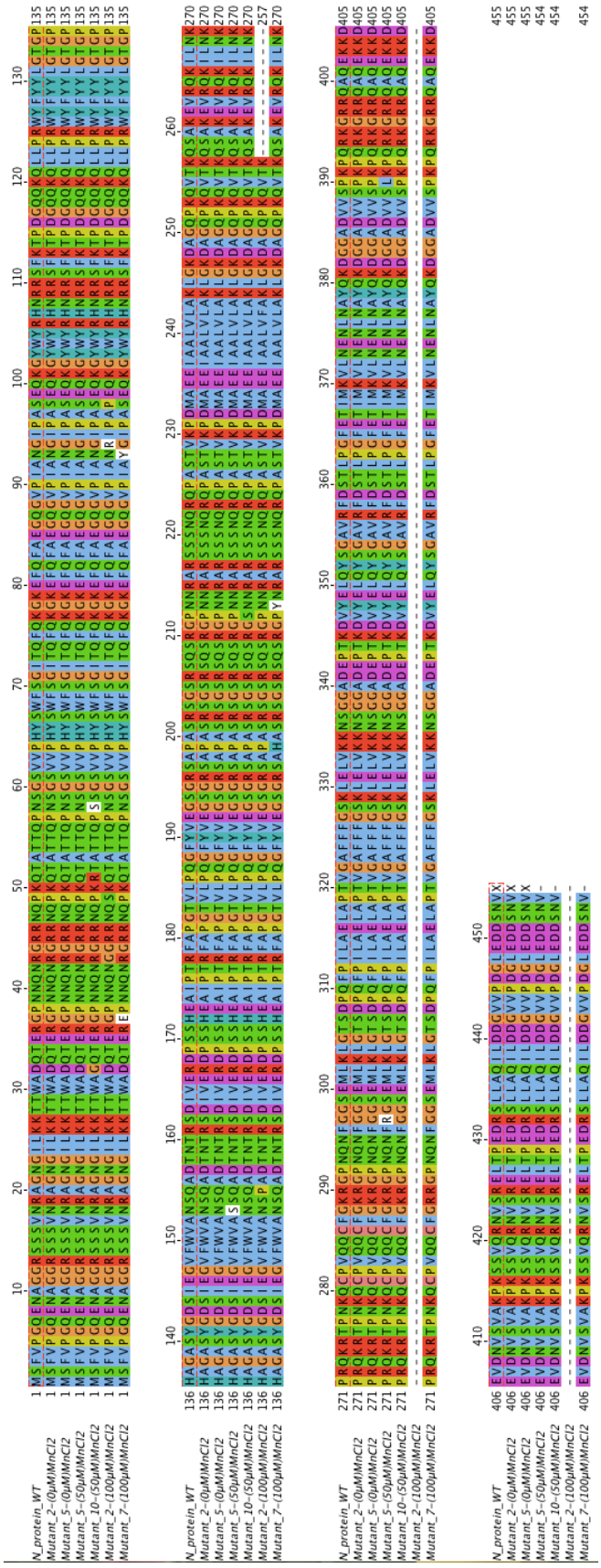

Figure 5.55. Distribution of amino acid mutations in 2 variants from 50μM $MnCl_2$ and 100μM $MnCl_2$ libraries generated with error-prone PCR. Sequenced variants from 50 μM $MnCl_2$ and 100 μM $MnCl_2$ libraries are aligned with wild-type using Jalview software. 0 μM $MnCl_2$ library is used as a control.

The results of the 50 μM $MnCl_2$ library sequencing showed that, at the amino acid level, mutations present in the 50 μM $MnCl_2$ library were less than those introduced into the 100 μM $MnCl_2$ library and that higher $MnCl_2$ concentration increased the percentage of protein truncation through the introduction of stop codons. Two colonies amplified without $MnCl_2$ were used as a control and showed no error. Subsequently a further 10 colonies were sequenced to evaluate the complexity of the library generated by adding 50 μM $MnCl_2$ into the PCR reaction. The results of sequencing showed that no less than 0.2% and no more than 0.6% mutations were introduced in a total of 1365 nucleotides sequenced **Table 5.6**.

Table 5.6. Sequence analysis and mutation rate of 10 mutated colonies generated by error-prone PCR.

Colony	**Mutations**	**Total sequenced**	**%Mut**
Colony 1	6	1365	0.4%
Colony 2	7	1365	0.5%
Colony 3	3	1365	0.2%
Colony 4	3	1365	0.2%
Colony 5	4	1365	0.3%
Colony 6	8	1365	0.6%
Colony 7	8	1365	0.6%
Colony 8	8	1365	0.6%
Colony 9	8	1365	0.6%
Colony 10	6	1365	0.4%
Total	61	13650	0.4%

Similary, the mutations generated by the second ep-PCR protocol were mainly single base substitutions (Appendix **Figure 9.79**) with the level of A to G transitions 36% and T to C 26% of the total number of mutations found since N protein consist more A (400bp) and less C (299bp) in N protein nucleotides distribution **Table 5.7**.

Table 5.7. Point mutations sequence analysis and mutation rates of variant library generated by ep-PCR.

Mutation	Number of the mutation	Found	Expected	Difference
A to C	1	2%	10%	8%
A to G	22	36%	10%	-26%
A to T	5	8%	10%	2%
C to A	0	0%	7%	7%
C to G	0	0%	7%	7%
C to T	6	10%	7%	-3%
G to A	6	10%	9%	-1%
G to C	0	0%	9%	9%
G to T	2	3%	9%	5%
T to A	2	3%	8%	4%
T to C	16	26%	8%	-19%
T to G	1	2%	8%	6%
Total	61			

At the amino acid level, the diversity created in the N protein sequence was an average of approximately 1% and one mutant was truncated **Figure 5.56**. 10 colonies from 100 µM $MnCl_2$ library were sequenced to compare number of truncation with 50 µM $MnCl_2$ library. The results showed 4 mutants were truncated and rest of mutants had 3 to 5 mutations at amino levels (**Appendix 9.80**). Often, truncations of proteins have been found to be a source for dominant negative effect however this is not true in all cases. The higher truncation rates the higher non functional protein. For that reason the 50 µM $MnCl_2$ library were chosen for further investigation for mutants that have a potential dominant negative effect.

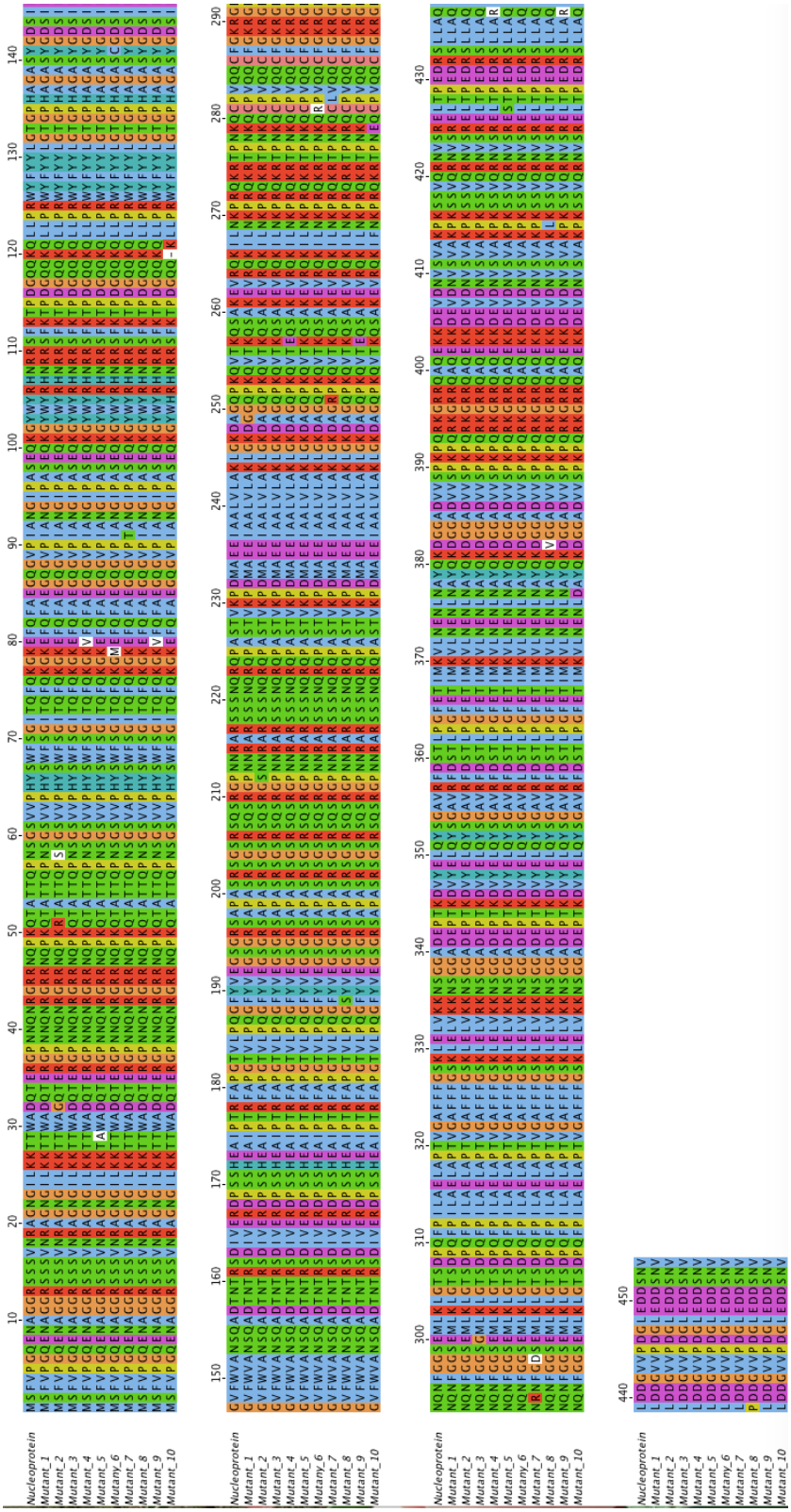

Figure 5.56. Distribution of amino acid mutations in 10 variants generated with error-prone PCR. Sequenced variants from 50μM $MnCl_2$ library are aligned with wild-type using Jalview software.

5.2.4 Phylogenetic tree analysis

Based on results obtained from the library generated in the presence of 50μM $MnCl_2$, 100 more colonies were screened by colony PCR and then sequenced. As shown in **Table 5.8,** transition mutations were more frequent than transversion. The most frequent transition was A to G (33%), whereas GC to CG mutations occurred with a rate of 0.8%.

Table 5.8. Point mutations sequence analysis and mutation rates of 100 variants library generated by error-prone PCR.

Mutation	No. of mutations	Mut%	Average
A to C	10	1.6%	1.8
A to G	211	33.7%	4.6
A to T	43	6.9%	2.2
C to A	6	1.0%	1.7
C to G	3	0.5%	1.5
C to T	45	7.2%	2.5
G to A	60	9.6%	2.7
G to C	2	0.3%	1.3
G to T	24	3.8%	2.1
T to A	24	3.8%	2.3
T to C	169	27.0%	4.3
T to G	9	1.4%	1.8
Deletion	19	3.0%	2.4
Insertion	2	0.3%	1.3
Total	627		

Phylogenetic analysis showed a substantial diversity had been created in the N protein, except for 3 sequences, which were identical to the wild type sequence **Figure 5.57**. The average rate of nucleotides mutation was 1.7% mutations per copy. Transition mutations (A to G and T to C) were commonest. The mutation G18T appeared in 20% of selected mutants, suggesting that this mutation happened in the early rounds of the mutagenic PCR reaction, a “jackpot” mutation that was carried through the subsequent rounds. The mutations A20G, A64G and A148T also appeared in selected mutants.

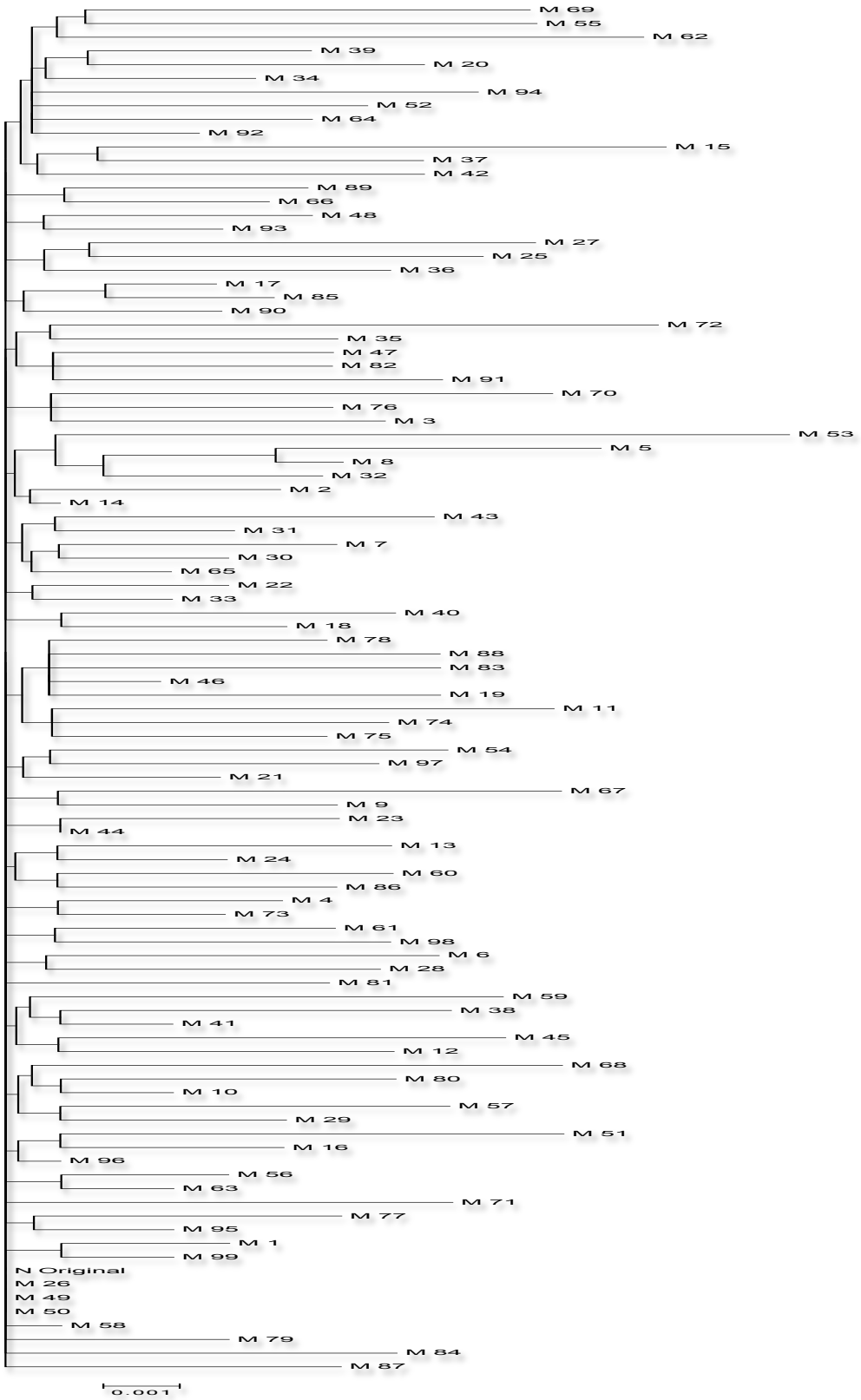

Figure 5.57. Phylogenetic analysis between non-mutated and mutated N protein. The N protein sequences from 100 clones were aligned by CLUSTAL omega alignment for the analysis and the tree was constructed using MEGA6.06 software based on analysis of nucleotides.

At the amino acid level, 33.3% of the mutated N proteins were truncated proteins. Truncated mutants varied from 16 aa (mut33) variant in length to 405 aa (mut82) variant in length while the overall diversity in the complete set of mutants was random and there was no specific type of mutation or site of mutation when the tested library was considered as a whole. This library of 100 variants was therefore considered suitable as a moderately diverse source of N protein variants with which to explore the possibility of selection of a trans-dominant negative inhibitor of MHV replication.

5.2.5 Effect of mutated N proteins on virus replication

To study the effect of mutated N proteins on virus replication in 17clone-1 cells, 24 well plates were seeded with 17clone-1 cells and the next day all plates were transfected with a different transformant selected from the screening of the 50 mM $MnCl_2$ library described, that is 100 individual transfections, and incubated 24 hr at 37°C in 5% CO_2. After incubation, each well was infected with MHV-A59 at a MOI 3. After 16 hours incubation, when the cytopathic effect in the control infections was extensive, the supernatants were removed for storage and the cell monolayers were fixed with formalin and stained with crystal violet. Some wells showed a cell density equal to the density of the control (cells transfected with empty plasmid vector) while other wells showed cell densities that varied when compared to the density of the control.

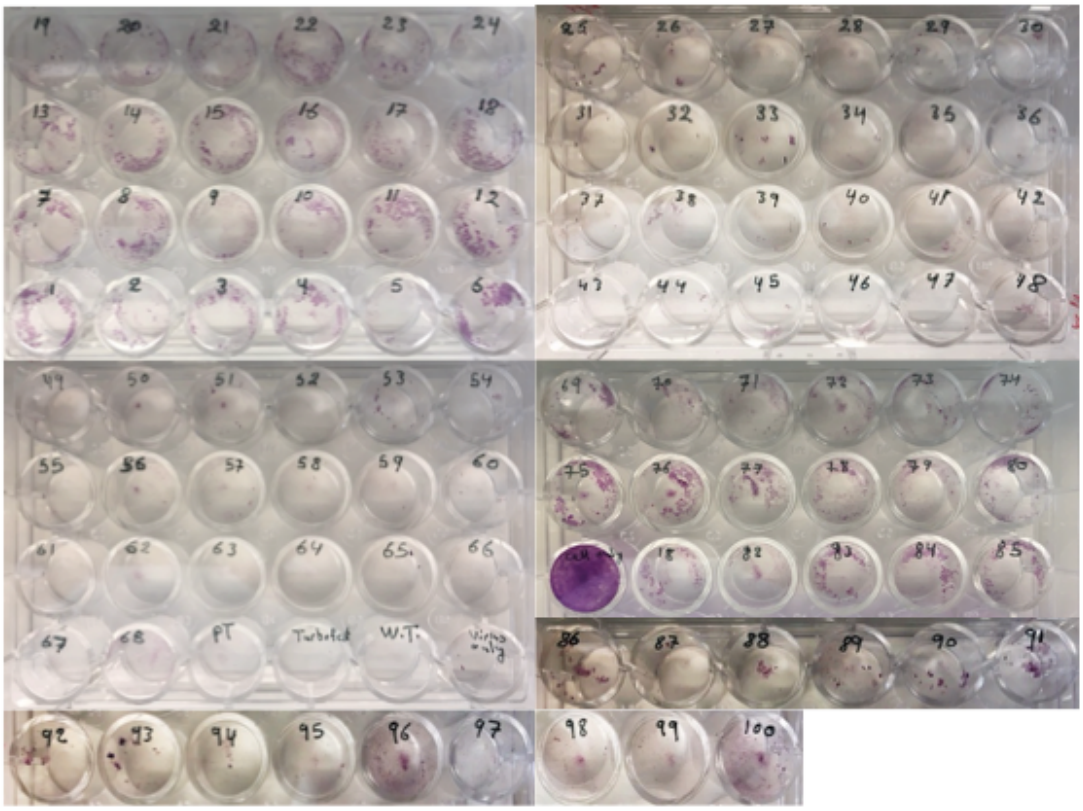

Figure 5.58. Effect of mutated N protein on MHV-A59 virus replication. 17clone-1 cells were cultured, seeded at 1 x 10^4 cells/well using 24 well plates, after 24 hours incubation, each well was transfected with 100 variants, then incubated for 24 hours at 37°C in atmosphere of 5% CO2. Cells were infected with mHV-A59 at MOI=3. After 16 hours incubation, supernatants were collected and each well was stained with crystal violet stain. Virus only, Turbofect and wild type were used as a control.

These monolayers were considered as candidates that had received potential dominant negative variants and the stored supernatants from these wells along with control supernatnats were subjected to plaque assay to measure virus titre as a quantitative measure of any effect. From this visual screen, 37 variants were selected for plaque assay from the original 100 transfections. As before it was observed that the replication of MHV increased when parental N protein was expressed *in trans*, while some mutated N protein variants decreased infectious virion production and were graded for comparison by the extent of the decrease in titre **Figure 5.59**. The distribution of the amino acid mutations in the 37 variants subjected to plaque assay was aligned with the wildtype sequence using Jalview

software **Figure 9.81**. Interestingly, the result indicated that all the mutants that showed a clear inhibitory activity, that is mut74 (222 amino acids), mut66 (120 amino acids), mut1 (290 amino acids), mut2 (149 amino acids) and mut38 (374 amino acid) were truncated except mut94 was not truncated with 3 amino acid mutations (Q251R, Q294R and G298D). Mutants with a similar activity to the wild type protein were point mutations and were not truncated.

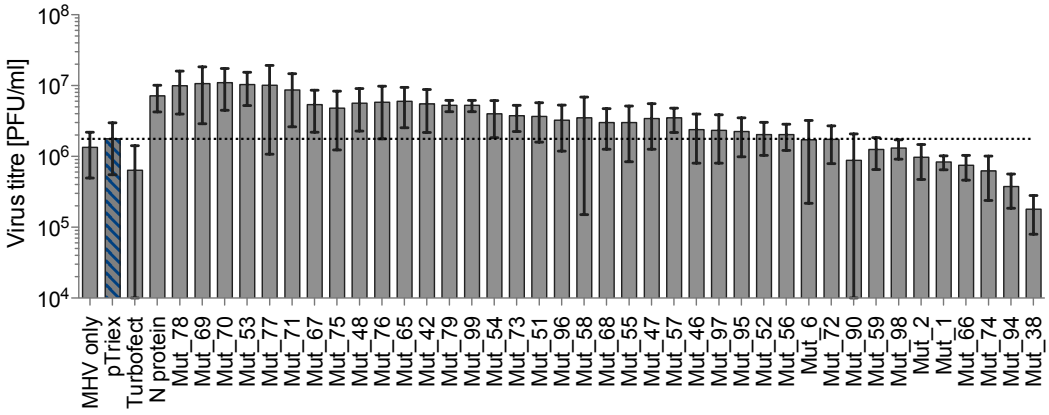

Figure 5.59. Inhibitory activity of different mutated N protein on virus replication tested by plaque assay. Plaque titrations were carried out on 17clone-1; monolayers were transfected with different mutated N protein, 24 hours after transfection, monolayers were infected with MHV-A59 at MOI=3, 16 hours post infection, supernatants were collected and virus titer was measured with plaque assay. Empty plasmid was used a control. The error bars represent mean and SD. The experiment has been repeated once.

The plaque assay was repeated for mut38, the most active inhibitor **Figure 5.60**, confirming the result obtained from the previous experiment that mut38 inhibited virus growth.

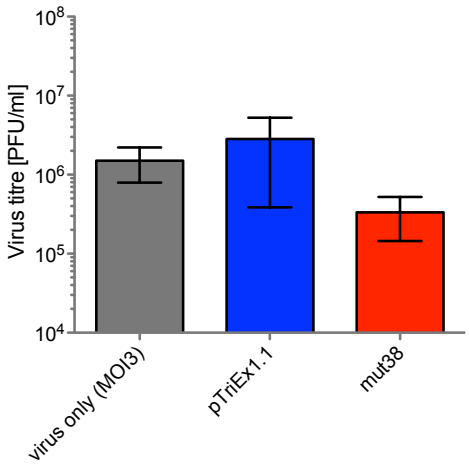

Figure 5.60. Inhibition of MHV-A59 infectivity by mut38. Monolayers of 17clone-1 cells were transfected with mut38, 24 hours after transfection, monolayers were infected with MHV-A59 at MOI=3, 16 hours post infection, supernatants were collected and virus titer was measured with plaque assay. The error bars represent mean and SD. The experiment has been repeated in triplicate

Although the presumed mode of action would be via the truncation protein, the mut38 variant sequence was analyzed at the nucleotide level **Figure 5.61**. The result showed that the mutant had 8 mutations, 3 of the mutations were G to A at sites 232,303 and 1098 while the remaining mutations were T950C, A1006G, G1120T, T1252C and A1310G.

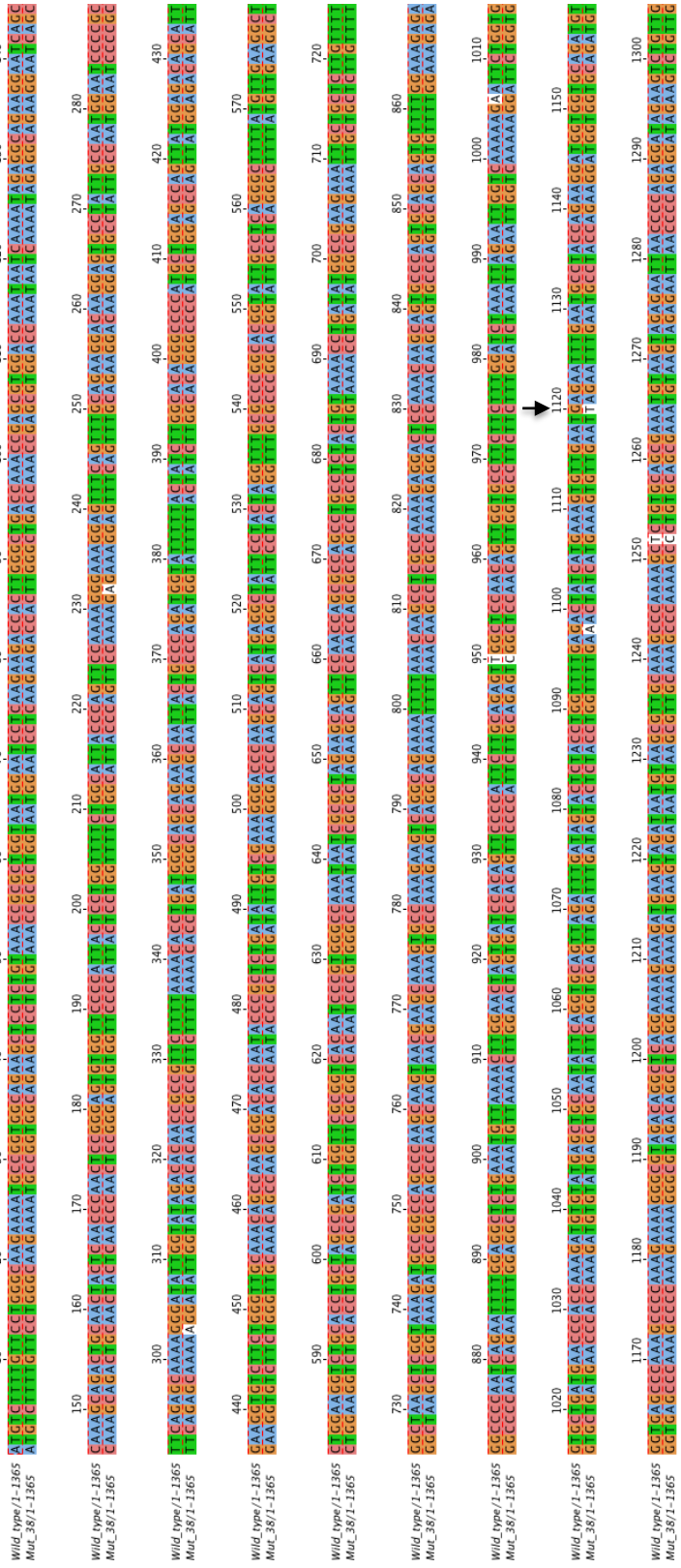

Figure 5.61. Comparison of wild type and mut38 nucleotides sequences. Sequenced mut38 are aligned with wild type using Jalview software. Black arrow represents stop codon sequence.

The mut38 variant was also aligned with the wild type at the protein level. The result revealed that the mut38 protein has 3 amino acid mutations (G78R, L317S and N336D) when compared to the wild type. However the predominant change was mutation of G1120T nucleotide, which resulted in a stop codon introduced at residue 374 instead of glutamate. This mutation therefore results in the production of a truncated N protein of 373 amino acids, compared to the 454 amino acids of the wild type MHV N protein **Figure 5.62**.

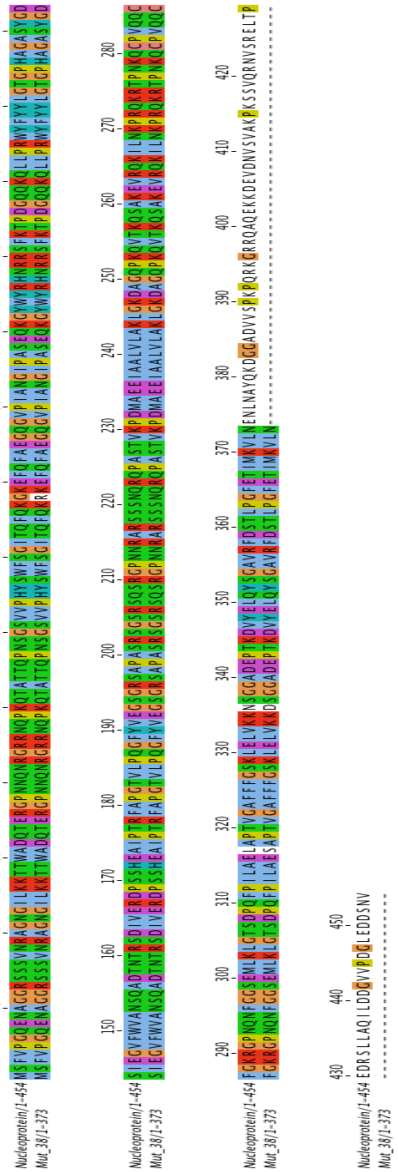

Figure 5.62. Comparison of wild type and mut38 amino acid sequences. Sequenced mut38 are aligned with wild type using Jalview software.

5.2.6 MTT assay

In order to test whether mut38 was exerting a specific effect on MHV replication or was generally cytotoxic through an unknown mechanism in 17clone-1 cells, so reducing virus yield non-specifically, a MTT assay was performed. 17clone-1 cells were seeded in two 24 wells plates (1×10^{4} cells/well) and the adherent cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, NEAA, L-Glutamine, 100 U/ml penicillin and 0.1 µg/ml streptomycin and incubated in the incubator at 37°C and 5% CO_2. When the 17clone-1 cells had reached approximately 75% confluence, wells were transfected in triplicate with plasmids expressing the wild type N protein or mut38 along with empty plasmid as control. 18 hours after transfection, a solution of MTT reagent was added to each well, and incubation continued further for 4 hours with observation every hour until a clear purple precipitate had become visible. The media was removed and DMSO was added to each well. One plate was incubated for one hour at 37°C and the optical density (OD) was measured at wavelength 595nm using a microplate reader. The other plate was incubated for 24hr at 37°C and 5% CO_2 and the optical density measured again at the same wavelength. The assay included wells containing transfection reagent only, DMSO as well as untreated control cells. This experiment was also repeated as described with the cells at approximately 60% confluence. The MTT assay revealed an absence of cytotoxicity for mut38 on 17clone-1 cells **Figure 5.63**.

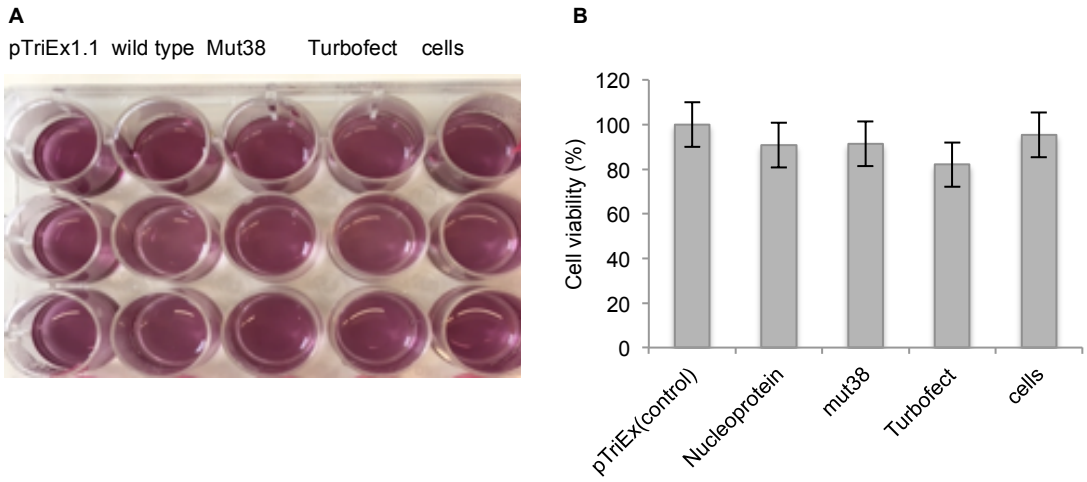

Figure 5.63. Cytotoxicity of wild type and mut38. 17clone-1 cells were cultured, seeded at 1 x 10^4 cells/well using 24 well plates, after 24 hours incubation, each well was transfected with wild type and mut38, then incubated for 18 hours at 37°C in atmosphere of 5% CO2. Cells were treated with MTT solution, incubated for 4 hours, after appearance of purple precipitant, media removed and DMSO added. The optical density measured with a plate reader set at wavelength 595nm after 1 and 24 hours incubation. Three replicates were performed. (A) 17 clone-1 cells were transfected with wild type, mut38, empty plasmid and Turbofect only for 24 h and subjected to MTT assay. (B) Cell viability was measured by MTT assay. The error bars represent mean and SD.

5.3 Discussion

N protein is a key protein involved in coronavirus replication. N protein binds with viral RNA to form the ribonucleoprotein complex (RNP), which protects the RNA in an ordered conformation that is necessary for replication and transcription. In addition, N protein is involved in many viral processes. As a result, the N protein of coronaviruses is a logical target for the development of antiviral compounds because such compounds would have the potential to target many crucial functions during the viral life cycle. A similar study had demonstrated previously that N protein increased SARS replication when provided *in trans* (Pan *et al*., 2008) suggesting that an assay could be established. In another study mutated HCoV-OC43 N protein, when transfected into 293T cells, was shown to significantly decrease the level of viral RNA encoding M protein compared with cells transfected with plasmids encoding the wild type N (Lin *et al*., 2014). Transfected mRNA encoding MHV N protein increases the PFU up to 40-fold but this enhancement was not found when the MHV N protein mRNA contained knockout mutations that inactivated translation of a functional protein (Hurst *et al*., 2010). A further recent study has shown that N protein increases the infectivity of MHV-A59 (Cui *et al*., 2015).

Directed evolution has become a powerful tool both for altering and improving the properties of many proteins, by harnessing Darwinian principles to generate proteins with new or improved properties. The first step in a directed evolution experiment is to introduce mutations randomly into copies of the target gene, resulting in a large and diverse library of variants. Error prone-PCR is a simple and common method for introducing such random mutations into an entire gene. Such

random mutagenesis has been shown to be important for studying the structure, function and evolution of proteins (Tindall and Kunkel, 1988) as well as in the development of novel proteins with desired properties. It is important to decide the level of mutagenesis that is best to create a large library of variants. If the mutation rate is too low it may not be possible to find the variants of interest. If the mutation rate is too high, nearly all the library variants will carry multiple mutations and the protein of interest therefore may be in an inactive form. For the work described here, the first protocol for ep-PCR was based on the protocol of Cadwell and Joyce (1992), which was basically optimized for a 400bp sequence. This protocol introduces errors at a frequency of 0.66% per nucleotide position over the course of the PCR by adding Mn^{2+} ions and unbalanced ratios of dNTPs, nearly all of which are base substitutions. However, the efficiency of cloning was found to be low **Figure 5.51**. The second protocol for library construction took advantage of generating a low frequency of mutations by the addition of only Mn^{2+}. With slight modification the ep-PCR reactions were carried out with 0, 50, 100, 200, 300 and 400μM of $MnCl_2$ with the aim of controlling of the mutation frequency during amplification of N protein. Error prone libriaries generated using 0, 50, 100, 200 μM $MnCl_2$ were gel extracted, but the addition of 300 and 400μM $MnCl_2$ in the PCR reaction produced no clear band and were not used for further investigation **Figure 5.52**. The 0, 50, 100, 200 μM $MnCl_2$ libraries generated were transformed into Stellar competent cells and sequenced to determine the mutation frequency. By sequencing 2 of the isolated clones from the 50 μM $MnCl_2$ and 100 μM $MnCl_2$ reactions, the results indicated that a higher concentration of $MnCl_2$ increased the number of mutations and that the frequency of protein truncations was also higher. Sequencing 10 more clones from the 50 μM $MnCl_2$ library demonstrated that the

mutated sites spanned the entire N protein in keeping with the desired complexity of an N protein library. Substitution of Mg^{2+} by Mn^{2+} affects base pairing specificity and the Mn^{2+} ion may interact with the DNA polymerase, either reducing the accuracy of base selection prior to insertion or inhibiting an "exonuclease-like" proofreading function (Beckman *et al*., 1985). It has been reported that the presence of 200 µM, 300 µM, or 400 µM $MnCl_2$ results in a nucleotide error rate of 2.2, 4.36, and 5.6 per kilobase, retaining protein function by up tp 83% in 200 µM $MnCl_2$ whereas higher $MnCl_2$ concentrations decreased protein functionality (McIsaac *et al*., 2014).

To further demonstrate the diversity of the final mutant library, 100 clones were sequenced and the resulting sequences aligned. A phylogenetic tree was built on the basis of grouping of sequences with similar characteristics with the branches of the tree representing the distance from a common origin. The phylogenetic tree constructed from the 100 sequences using MEGA0.06 software revealed that only 3 mutants were identical to the wild type although several additional mutants showed close homology to the wild type.

At the nucleotide level, 627 mutations were found in 96 mutants, 211 of these 627 mutations were A to G, 169 T to C, 19 deletion mutations and insertion were detected in 2 clones. The data from **Table 5.8** showed that transition mutations (AG and TC) occur much more frequently than other types of mutation and that certain types of mutation like A→G or T→C occur more frequently than others. The N protein consists of 1365bp with a distribution of A 400bp (29%), T 316bp (23%), G 350bp (26%) and C 299bp (22%) and the cumulative total diversity for the 50µM $MnCl_2$ library was a mutation frequency of ~0.4%.

The 100 N protein mutants generated by ep-PCR in this study were transfected in 17clone-1 cells and at 24 hours post-transfection, cells were infected with MHV-A59. Monolayers that appeared to survive the effects of cytopathic effect better than cultures that were not transfected, or were transfected only with empty vector, were selected for quantitative measurement by plaque assay. The results confirmed that replication of MHV increased when the wild type N protein was expressed but that some mutated N protein decreased infectious virion production. An amino acid analysis showed that the majority of mutants that showed an inhibitory effect were truncated and that the length of the truncated mutants varied from 120 aa (mut66), 290 aa (mut1) to 373 aa (mut38). Non-truncated mutants that showed some effect had 3 mutations or more. For example mut2, which contained a 6 amino acids substitution and mut98, a 5 amino acid substitution, but these were not investgated further.

The inhibitory effect of the most effective mutant, mut38, was confirmed by repeated plaque assay with the data confirming that a mutated N protein can interfere with virus replication and that a decrease in infectivity is possible by a mutation that changes relatively few of the residues in the virus N protein. The mut38 is truncated due to introduced stop codon, with only 373 residues and so lacks the N3 and B spacer domains. In SARS CoVs, the C-terminal domain is responsible for oligomeriziation and removal of 40 amino acids apparently decreased the ability of the N protein to oligomerize (Luo *et al*., 2006a). The lack of 13 amino acid (residues 377–389) from the C-terminus of the HCoV-229E N protein has been also found to impair higher-order oligomerization and the virus titre is decreased by the presence of an excess of a peptide based on the

sequence of the C-terminal tail, which interferes with oligomerization of HCoV-229E C-terminal domain (Lo *et al*., 2013). Improvement could be made to this work by conducting further experiments such as confiming which of the mutations in mut38 were responsible for the effect seen on virus infectivity by investigating truncation without mutations, each mutation individually and one mutation with the truncation. A computer algorithm can be applied to examine the role of the mutated proteins' structure. No general cytotoxic effect was noted by MTT assay for mut38 suggesting a specific effect in MHV infected cells. Accordingly, these results confirm that the wild type N and selected mutated N proteins have an effect on MHV-A59 replication when supplied *in trans*.

In summary, this chapter demonstrated that presence of some N protein mutants could decrease virus infectivity and that the most effective variant, mut38 is not generally toxic to cells. Based on the results obtained from this study, there is evidence that coronavirus replication might be inhibited by a mutated N protein if present in cells at the time of infection. Since N protein a genetically stable protein and shows least variation in the gene sequence, therefore indicating it to be an efficient drug target candidate.

6 Directed evolution

6.1 Introduction

Previous chapters indicated that some mutated N proteins derived from MHV N can have potentially inhibitory effects on virus replication. These mutants were found by individual transfection of a large number of singly isolated mutants but it is reasonable to suppose that these mutants, and other potentially more potent variants, would be also present in a library of many thousands of random mutants. To select such mutant(s), which have an inhibitory effect on replication, from various complex libraries a directed evolution approach (also called *in vitro* laboratory evolution) can be useful (Jürgens *et al*., 2000; May *et al*., 2000; Zaccolo and Gherardi, 1999). This approach does not require an understanding of protein structure or function and is predicated on evolutionary paradigms. Directed evolution is a powerful method for generating enzymes and even entire genomes with desired properties. A library of variants is screened for a desired outcome, resulting in an enriched library of perhaps hundreds or thousands of variants from an original much more complex pool. This enriched library is recovered and used in a second round of enrichment, selecting again for a desired trait. The combination of randomness in the initial library and iterative selection results in the selection of a function based purely on phenotype, but when the complexity is reduced to only a few members each can be characterized genotypically and the genetic basis of the new trait determined. The capacity to screen depends on the size of the library for, although limited library diversity can be easily screened, a large library requires a high selectivity in order to be screened effectively, so for each new application, screening and selection are generally re-invented. The outcome of a directed evolution experiment can be also altered by the mutation rate and population size

in addition to the selection strategy. During a directed evolution experiment the frequency of variants in a population changes; the diversity of the variants pool is reduced and the frequency of the desired variants is increased.

With its diverse functionality N protein may be considered an ideal target for the development of an antiviral drug and based on the data presented in previous chapters it can be hypothesized that mutated N may have an inhibitory effect on virus replication and may be considered a dominant negative. As such, the directed evolution of N protein could be considered to develop increasingly effective dominant negative N variant(s), which in turn could lead small molecule development **Figure 6.64**.

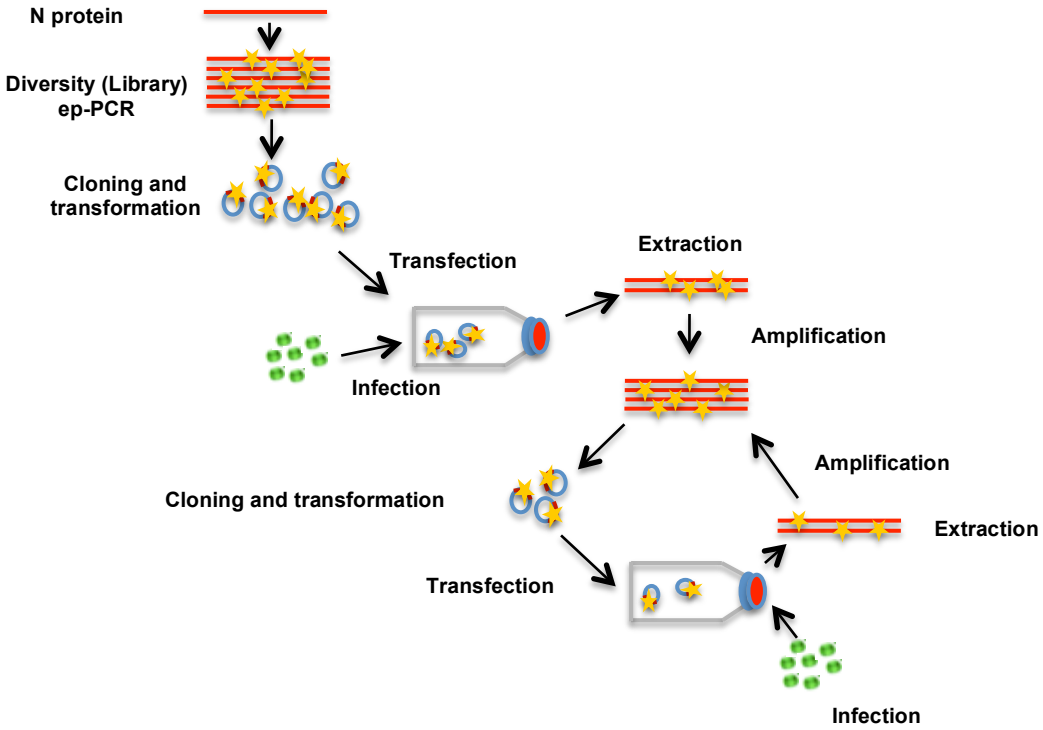

Figure 6.64. Work flow for directed evolution of MHV N protein in 17clone-1 cells. N protein of MHV-A59 is amplified by ep-PCR, library of variants are cloned and transformed into *E. coli*. After transfecting 17clone-1 with N variants, cells will be infected with the virus. Selected clones is extracted from live cells, amplified, cloned, transformed, transfected and infected. Evolution cycle is repeated for several rounds.

In this chapter, a new expression and screening system is developed to select the evolution of mutated N protein variants with the potential to inhibit MHV replication. Libraries were generated by the second protocol ep-PCR protocol previously discussed (50 μM $MnCl_2$) as described in chapter 5, and the resulting library was cloned into the pTriEx expression vector and transformed *en masse* into *E. coli*. Unlike earlier experiments, no colonies were picked, instead the whole library was prepared as DNA and transfected into mammalian cells which were subsequently infected with MHV-A59 virus. Variants that might have exhibited a dominant negative property were isolated from those live cells that survived the infection and subjected to further rounds of selection, purification, re-cloning and transfection. This process was applied for several rounds with the aim of developing a protein inhibitor based on the N protein capable of MHV replication inhibition, but the principle of the method could be applied to any mutated component of the coronavirus genome.

6.2 Results

6.2.1 Directed evolution for N protein and mut38

As a first step to test that selection of potential dominant negative variant(s) by *in vitro* directed evolution, a small mixture containing only the wild type and mut38 was constructed. The wild type and vector containing N mut38 were mixed in a ratio of 1:1 from stored stocks and used as a library that contains only two variants to prove the principle of competition between these two variants. The wild type N, which increased virus production and mut38 variant which decreased virus production.

6.2.1.1 Round 1: wild type and mut38

Mammalian 17 clone-1 cells were grown in a T25 flask, representing a population of about 5 x 10^6 cells. Assuming a transfection efficiency of about 50% then something over 10^6 variants could be sampled using a culture of this size. When the cell confluence reached ~70% the monolayer was transfected with the plasmid mixture which contained equal copies from both wild type and mut38 using Turbofect transfection reagent according to the manufacture's instruction. After incubation for 24 hours the flasks were washed twice with PBS and the monolayers infected with MHV-A59 at an MOI 3 by incubation at 37°C on a low speed platform rotator for 45min. After incubation, the inoculum was discarded and the monolayer again washed with PBS. Fresh media was added and the flask was incubated at 37°C for 16 hours, by which time cytopathic effect was clearly extensive. The media was removed gently by aspiration and the flask was washed carefully with PBS 3 times to eliminate all dead cells. Intact cells were recovered by scraping and low speed centrifugation at 4°C and any plasmid DNA present in the surviving cells

was extracted using a plasmid miniprep kit (Thermo Fisher SCIENTIFIC). The DNA product was amplified with N protein ORF specific forward and reverse primers (see Table 1) as described in section 2.1.5 and the PCR product was analyzed by agarose gel electrophoresis **Figure 6.65**. The result showed a high intensity band with the correct size for the N encoding fragments indicating that the N protein ORF (wild type and/or mut38) was present in rescued cells.

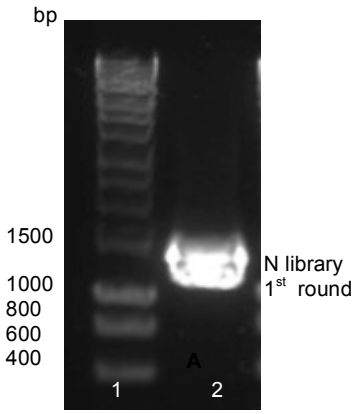

Figure 6.65. Gel electrophoresis of N protein extracted from survived cell. The recovered error-prone library representing N sequence of both wild type and mut38 was amplified using N protein forward and reverse primers. Lane 1: Marker (1kb) DNA ladder, Lane 2: N protein library (1365bp).

The amplified N sequence, which represents an enriched library of variants was purified by a PCR reaction clean up kit and the pool of DNA fragments was cloned into a *Nco*I/ *Xho*I digested pTriEx1.1 plasmid through the In-Fusion protocol. The reaction was transformed into Stellar competent cells (500 µl) and each 100 µl of the transformed mix was inoculated into to 10 ml LB broth with 100 µg/ml of ampicillin and incubated at 37°C for 16hr in the shaker. The last 100 µl from the transformation mix was plated onto LB agar plate supplemented with ampicillin and colonies isolated to provide a snapshot of the mutants present in the rescued

population of N sequences. An enriched library was purified from the overnight culture using a miniprep kit (Thermo Fisher SCIENTIFIC) and used, as before, for a second round of transfection and selection. A sample of the recovered library was also double digested with *Xba*I and *Xho*I to confirm the presence of an insert consistent with the ORF for wild type N protein and/or mut38 **Figure 6.66**.

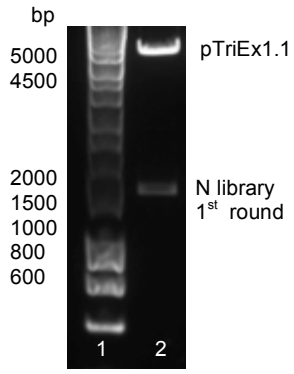

Figure 6.66. Gel electrophoresis of error-prone library double digest containing variants (wild type and mut38) of N protein. The recovered error-prone library representing N sequences following a single round of selection was double digested with *Xba*I and *Xho*I. Lane 1: Marker (1kb) DNA ladder, Lane 2: N protein library (1662bp) (wild type/mut38).

Of the colonies from the plated transformation, 25 were picked randomly and screened by colony PCR for the presence of N sized inserts using primers T7 forward and Rv_N **Table 2.1**. The N insert was found in 15 isolate from a total of 25 isolated that were subjected for colony PCR **Figure 6.67**.

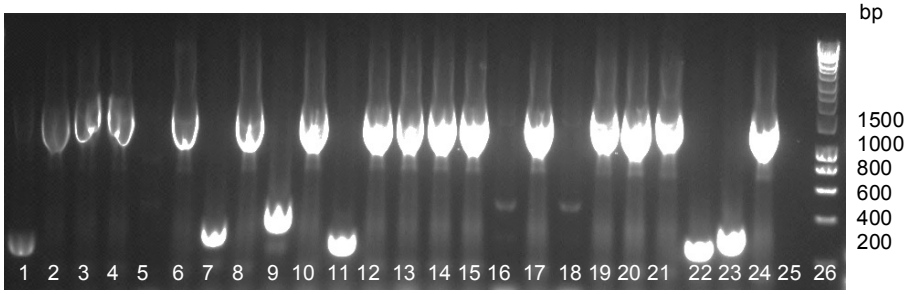

Figure 6.67. N protein (wild type and mut38) screened by PCR. 25 colonies were screened for the presence of N protein insert. Lanes 1-25: colonies numbered from 1-25, Lane 26: Marker (1kb) DNA ladder.

Of the 15 isolates that were positive for the correctly sized insert, ten isolates (2,3,4,6,8,10,12,13,15 and 17 in **Figure 6.67**) were cultured in LB broth with ampicillin and purified in the next day by using miniprep kit (Thermo Fisher SCIENTIFIC). All 10 isolates were sequenced using primers T7 forward and TriExDOWN from SourceBiosinences. The result showed that 6 isolate were identified carrying sequence of the wild type and the remaining 4 isolates carried the mut38 sequence (i.e mut38 was present in 40% of the total sequenced isolates) as shown in **Figure 6.68**. This round was considered 1st round of selection.

Figure 6.68. Sequence alignment of 10 isolates with wild type. Sequence of 10 isolates was aligned with wild type using snapgene software. Red box represent G to A mutation at position 303 which occurred in mut38.

6.2.1.2 Round 2: wild type and mut38

The library products from the 1st round as described above were transfected again in a T25 flask containing a ~70% 17clone-1 cell monolayer using Turbofect reagent and incubated for 24hr at 37°C with 5%CO_2. The flask was washed with PBS twice and the cells were infected with MHV-A59 at MOI 3, incubated for 45 min at low speed platform rotator. After incubation, inoculum was removed, washed twice with PBS, new fresh media was added and the flask incubated at 37°C with 5%CO_2. The next day, the media was aspirated gently, the flask washed with PBS and the plasmids present in surviving cells were rescued by using a plasmid miniprep kit. The DNA product was amplified with Fw_N and Rv_N primers **Table 2.1** and analyzed by gel electrophoresis. The result showed a clear band with of N protein ORF size (1365bp) **Figure 6.69**.

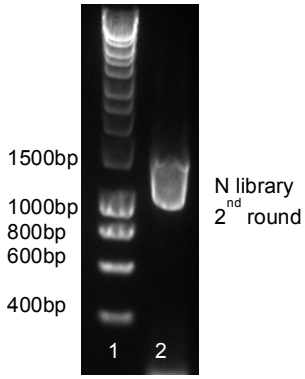

Figure 6.69. Gel electrophoresis of N protein extracted from survived cell (2nd round of selection). The recovered error-prone library representing N sequence of both wild type and mut38 from 2nd rounds of selection that was amplified using N protein forward and reverse primers. Lane 1: Marker (1kb) DNA ladder, Lane 2: N protein library (1365bp).

As before the reaction was cloned and transformed into Stellar competent cells (500 µl) and divided as described in section 6.2.1.1. The last 100 µl from transformed mix was transformed onto LB agar plate containing ampicillin.

Screening was done by colony PCR and 10 isolates grown in LB broth, purified by miniprep kit and sent for sequencing with primers T7 forward and TriExDOWN. The sequencing result showed that 7 of the isolated plasmids had the wild type sequence and the remaining 3 isolates were mut38. This round was considered the 2nd round of selection.

6.2.1.3 Round 3: wild type and mut38

A third round of selection was performed in 17clone-1 cells using the library from 2nd round, also transfected in a T25 flask containing 17clone-1 monolayer and all steps were repeated as described above. The results from 10 sequenced colonies showed that wild type N protein sequence was identified in 6 isolates while the mut38 sequence was present in 4 isolates. The purified library was considered the 3rd round of selection and was ready for a next round of selection.

6.2.1.4 Round 4: wild type and mut38

The purified library from the 3rd round of selection was transfected into T25 flask and all steps were repeated in this round as described previously. The results from 10 sequenced colonies showed that wild type N protein sequence was present in 5 isolates while mut38 sequence was present also in 5 isolates. After the 4th round of selection the result revealed that both WT and mut38 were present in equal numbers. At this stage we stopped the selection experiment since strategy applied in this method can check for selection but the result obtained didn't show any significant in the selection after 4th rounds and we were not sure how many rounds are required for success of this selection method. Another approach maybe developed in the future to gain more accurate results. The next step is to start the selection in a bigger library.

6.2.2 Directed evolution for 10 variants

Despite the lack of selection in the experiments described in section 6.2.1 a directed evolution experiment was designed for a pool of N variants that were more complex, that is the type N protein and 10 of the selected variants as described in Chapter 5. Accordingly, a mixture was generated from plasmids encoding the wild type, mut38, mut1, mut2, mut6, mut59, mut65, mut66, mut90, and mut94 by preparing plasmid stocks from each variant and mixing them equally to form the final sample for transfection, that is, a total of 8 µg for a T25 flask. The ability to select from this pool of variants depended on their effect on virus titre as shown in **Figure 5.59**. The pool included variants that showed a potential inhibitory effect (mut1, mut2, mut66, mut38, and mut94), three variants that showed the same effect of the control (mut90, mut6 and mut59) and mut65, with the same effect as wild type N. In addition the wild type was included in the library. This plasmid mixture was assumed to contain ~40-50% of its members as potential dominant negative variants. According to results in chapter 5 it was assumed that 2^{nd} error prone protocol (50 µM $MnCl_2$) can provide less than 50% dominant negative variants.

A T25 flask containing ~70% 17clone-1 cell monolayer was transfected with the constructed library using Turbofect reagent and incubated for 24hr at 37°C with 5%CO_2. The next day, the flask was washed with PBS twice and cells were infected with MHV-A59 at MOI 3. After 16hr incubation, DNA was extracted from the surviving cells using a miniprep kit, the N coding DNA product was amplified with Fw_N and Rv_N primers **Table 1** and analysed by gel electrophoresis **Figure 6.70**.

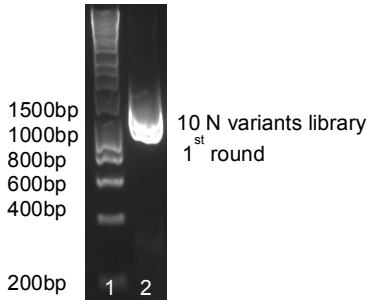

Figure 6.70. Gel electrophoresis of N protein extracted from survived cell. The recovered error-prone library representing N sequence of 10 variant was amplified using N protein forward and reverse primers. Lane 1: Marker (1kb) DNA ladder, Lane 2: N protein library (1365bp).

The amplified N sequence was purified by a PCR reaction clean up kit and the pool of DNA fragments was cloned into pTriEx1.1 plasmid previously digested with *Nco*I/*Xho*I through the In-Fusion protocol. The reaction was transformed into Stellar competent cells (500 µl) and divided as described for the wild type/mut38 library. A sample of the mix was double digested with *Xba*I and *Xho*I for confirmation of suitably sized inserts **Figure 6.71**.

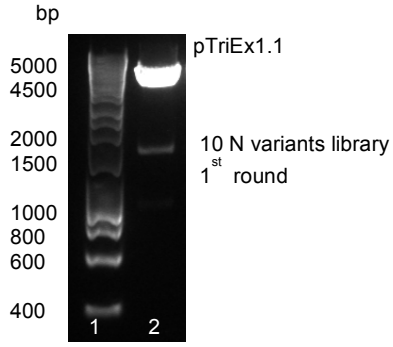

Figure 6.71. Gel electrophoresis of error-prone library (10 variants) double digest. The recovered error-prone library representing N sequences of 10 variants following a single round of selection was double digested with *Xba*I and *Xho*I. Lane 1: Marker (1kb) DNA ladder, Lane 2: N protein/mut38 library (1662bp).

For screening, 24 colonies were selected randomly and screened by colony PCR and N sized insert was found in 9 isolates only. Another 24 colonies were screened by colony PCR **Figure 6.72** and from this screen 10 randomly picked colonies were prepared as plasmid DNA and sent for sequencing with T7 forward and TriExDOWN primers.

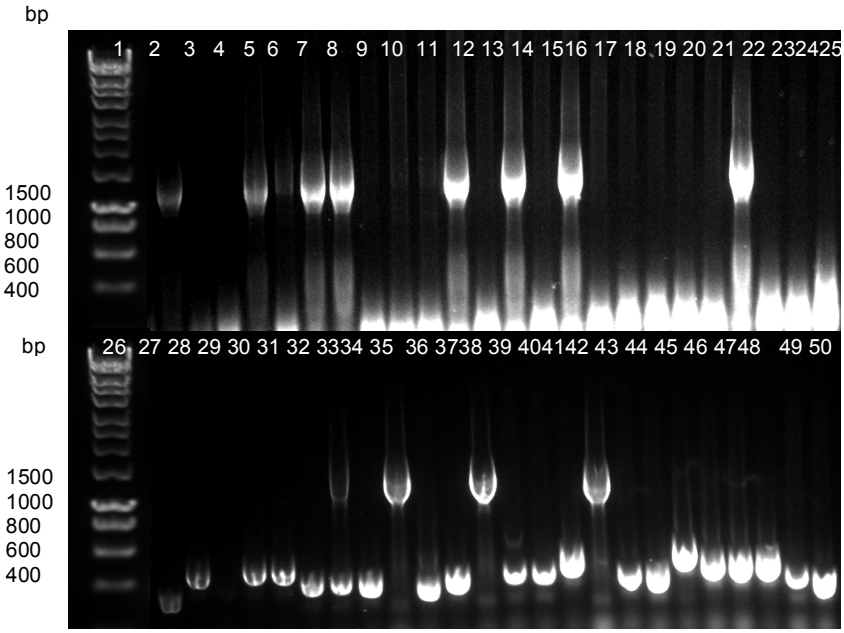

Figure 6.72. N protein (10 variants library) screened by colony PCR. 48 colonies were screened for the presence of N protein insert. Lane 1: Marker (1kb) DNA ladder, Lanes 2-25: colonies numbered from 1-24. Lane 26: Marker (1kb) DNA ladder, Lanes 27-50: colonies numbered from 25-48.

The purified plasmid product from the tissue culture transformation, derived as above, was transfected again to a T25 flask and this process was repeated for 3 rounds. As before, a representative 10 colonies were picked from each round of transfection and selection and the distribution of variants from each of these 10 sequenced isolates from each round is illustrated in **Table 6.9**. The result indicated little overall selection. The data from the 2nd round of selection indicated that the wild type was the predominant species although it was expected that the wild type N would be outcompeted by potentially dominant negatives. In the 3rd round it

appeared that mut65, which showed an increase in virus titre when screened alone, appeared less frequently but this was also true for mut6, which previously had the same inhibitory effect as the control. However, the results across the rounds of selection revealed essentially random recovery and no significant selection for dominant negative variants. Sequencing 10 isolates from each round was helpful in showing the distribution of mutants from the selection procedure, that is that plasmid DNA was successfully re-isolated at each round despite the complexity of the methodology, but it can be concluded that the experiment was too lacking in complexity to demonstrate a clear effect and that many thousands of clones may have to be tested to observe a rising variant among many other equally competing variants.

Table 6.9. Screening of N protein variants after 3 rounds of directed evolution. 10 randomly isolates were selected and sent for sequencing from each round.

Variant	Wild type	Mut1	Mut2	Mut6	Mut38	Mut59	Mut65	Mut66	Mut90	Mut94
Round1	2	1	1	2	0	1	1	1	0	1
Round2	3	2	1	1	1	0	1	1	0	0
Round3	2	0	1	1	1	1	0	2	1	1

6.2.3 Directed evolution of a complex N protein library

In vitro error prone libraries generate a population of many thousands of different variants and screening and selection is a key factor for successful evolution especially from a pool that contains mainly unknown or inactive form of mutants. In directed evolution, the library size to be screened is ultimately limited by the screen itself and a combination of selection approaches can be more useful with big libraries.

Therefore, accepting that the data with limited plasmid mixtures of know N variants did not satisfactorily demonstrate clear selection, all the steps used previously for the wild type/mut38 and the 10 variants libraries were performed for a library containing a random pool of N protein variants. The second error prone protocol (50 µM $MnCl_2$) as described in chapter 5 was used to generate the library of N variants for this experiment. After amplification under error prone conditions, the DNA product was analysed by gel electrophoresis **Figure 6.73**.

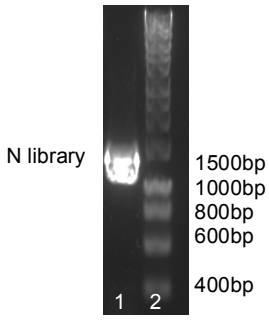

Figure 6.73. Gel electrophoresis of N protein generated by error prone PCR (50 µM $MnCl_2$). N protein was amplified using second error prone PCR. Lane 1: N protein library (1365bp), Lane 2: Marker (1kb) DNA ladder.

The whole error prone N band, representing a random library of N variants, was purified by a clean up kit, cloned by In-Fuison reaction as described and transformed in Stellar competent cell. The transformation mix was cultured *en masse* in LB broth and the next day the culture was extracted using a miniprep kit. A T25 flask was transfected with the library using Turbofect transfection reagent according to manufacturer's instruction and the selection procedure repeated. That is, after 24 hr incubation the 17clone-1 cells monolayer was washed twice with PBS and infected with MHV-A59 at an MOI 3. The inoculum was removed after 45min, monolayer washed with PBS and fresh media added. After 16 hours incubation, media was aspirated gently; the monolayer washed with PBS and DNA

was extracted from the cells that had survived the MHV infection. All these steps were repeated for 4 rounds and the supernatants from each round of selection were stored and titrated for the level of virus replication by plaque assay. The results of plaque assay showed that the virus titre slightly decreased after the first round of transfection and selection but that virus infectivity slightly increased in the 2nd round. In the 3rd and 4th rounds virus titre again decreased slightly **Figure 6.74** although in all cases the error bars of the assays overlapped so no significant overall trend was apparent. It was hoped that the plaque assay result could be used to assess which round might be used to screen for individual clones that could exhibit a dominant negative effect but it is clear that the data do not support such a strategy and other selection methods will be needed to screen for true dominant negatives.

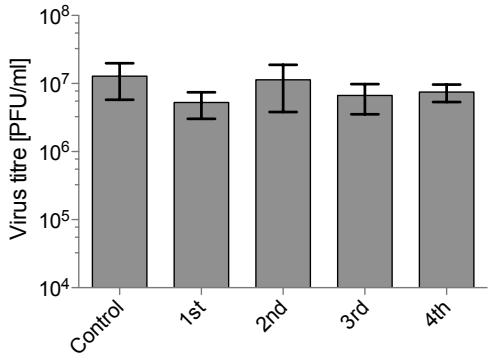

Figure 6.74. Inhibition of virus replication tested by plaque assay. Supernatants from four rounds were subjected for plaque assay to measure virus titre. Cells infected with virus only at MOI=3 is used as a control. The error bar represent mean and SD.

6.3 Discussion

The success of any directed evolution experiment hinges on generating libraries with significant size and diversity and with appropriate methods for screening/selection. As demonstrated in previous chapter, ep-PCR was titrated for its ability to introduce random errors into the N protein sequence and was used and here those libraries were used to try and select dominant negative N variants from different pools of N protein.

In order to screen for potential dominant negative N variants, cell survival was assessed with different pools. For wild type N protein and mut38 screening, mut38 was found in live cells after 4 rounds with the same numbers as the wild type. This stability in numbers after 4 rounds selection suggested that selection was not working efficiently to provide a good starting point for evolution selection/screening. Although disappointing there were many possible explanations for this, which prevented a clear conclusion. For example, WT and mut38 plasmids could both be entering the same cells during transfection and their effects of stimulation (WT) and inhibition (mut38) on MHV replication cancelling each other out. The extensive washing of the surviving cells was assumed to have removed input plasmid DNA but it cannot be ruled out that this too provided a background that obscured any selection taking place, although this would have been expected to decrease with increasing rounds of selection. A technical improvement in future experiments may be to include a DNAse step before the cells are disrupted to remove any loosely bound DNA. Notwithstanding this data an evolution experiment with a bigger pool of variants, the 10 variants isolated in previous experiments, was attempted. Here too, selection was not obvious although it with a more complex mixture the effects at each round might have been expected to be slower. Again, multiple plasmid

entry into single cells may provide an explanation of why selection was not more apparent. In a final and much more complex library selection experiment, a true epPCR library created under conditions known to generate suitable diversity, the virus titre decreased after the 1st round of selection, possibly consistent with the presence of variants that might exhibit dominant negative properties. However, the virus titre recovered in the 2nd round and was variable in the 3rd and 4th rounds indicating no significant trend in selection overall. The possible explanations for the failures above also apply to these results, and are in fact more plausible as the library was highly diverse and high numbers of positive WT-like variants could obscure any dominant negative variants.

In directed evolution experiment, it is important to consider the time and path of evolution that may be required to select a gene of interest with desired properties. Several mutations may be required to switch the N protein from a protein which enhances MHV replication to a dominant negative protein and these may be only a small proportion of the total library used. It may be necessary to perform an incremental approach to N protein evolution by combining low error rate libraries, which include many non-desired variants, with high error rates that are more likely to contain a desired function but will also include many non-functional variants. Therefore, large libraries or mixtures of several thousand mutants should be screened at each round and for many rounds in order to find suitable dominant negative variant(s). Adaptation in the experiments described here were stopped after a few rounds of selection as a result of technical and consumable limitation but the data in earlier chapters suggests such variants do arise although the data in the later chapters suggest their isolation from random libraries is challenging.

In conclusion, this chapter demonstrated that checking for success of directed evolution protocol in selecting dominant negative variant(s) in three different size libraries was inadequate. The precise mechanism for this selection was not clear and more optimization such as sequencing more variants from each round or using different approach may improve selection protocol. Next-generation sequencing (NGS) technologies has been applied for RNA virus detection in recent years (Capobianchi *et al*., 2013; Shan *et al*., 2011; Yu *et al*., 2013; Zhang *et al*., 2014) which help in revealing a huge number of sequences of nucleic acids (DNA or RNA) and can detect various kinds of RNA viruses. The use of next generation sequencing approach could provide more a detailed analysis and scan throughout the library generated by random mutagenesis. A deeper understanding of the mechanism by which dominant negative variant (s) interfere with WT may help in selection stage.

7 Expression and purification of MHV-A59 Mtase protein

7.1 Introduction

The primary goal for any research studying protein characteristic, function, drug development or production of antibodies against these proteins is obtaining a suitbale yield of purified protein. This challenge requires research to select appropriate vectors, purification tags, expression host and many other parameters. Since there is a continued need to study and develop treatments for coronavirus infections including infections by significnat viruses such as SARS and MERS, understanding key coronavirus protein structure, function and interactions will be the cricital to develop novel therapies for current and future coronavirus outbreaks and to improve control strategies. One of the attractive targets for coronavirus drug development is the viral RNA capping enzymes and the recent connection between the requirement for viral RNA capping and evasion of the host cell innate immunity is a promising research field for the development of anti viral interventions. Put simply, an inhibitor of the capping reaction would lead to viral RNA being detected by the innate RNA sensors RIG-I or MDA-5 and the induction of a signalling which could avert a full infection as the success of any viral infection depends on the ability of viruses to evade the host immune response. Viruses have evolved means to deactivate host sensing through either direct antagonism of pathway components, for example by the NS1 protein of influenza virus (Mibayashi *et al*., 2007), or by molecular mimicry of host processes so that detection is avoided. The duplication of capping elements for viral mRNAs is an important example of the molecular mimicry approach (Decroly *et al*., 2012). Higher eukaryotes uses a 5′-terminal capping system to protect mRNA from degradation by 5’

exoribonucleases, ensures efficient mRNA translation, mRNA stability, and helping to distinguish between self/non-self RNA that will lead to initiation of the host immune response (Gu and Lima, 2005; Shuman, 2002, 2000; Yoneyama and Fujita, 2010). This capping system includes methylation of the first one or two 5' nucleotides at the ribose 2'-O position by a distinct host 2'-O methyltransferases (MTases) (Bélanger *et al*., 2010; Werner *et al*., 2011). The vast majority of viruses that replicate in the cytoplasm including coronaviruses (Daffis *et al*., 2010; Szretter *et al*., 2012; Züst *et al*., 2011) have evolved a strategy to cap their RNAs and the coronavirus MTase (nsp16) activity is therefore important for coronavirus virus replication/transcription and an obvious target to develop antivirals for control of coronaviruses infection. In addition, an nsp16/nsp10 interaction is involved in the capping of coronavirus RNA (Bouvet *et al*., 2010; Chen *et al*., 2011; Debarnot *et al*., 2011; Decroly *et al*., 2011, 2008; Lugari *et al*., 2010) and can be used in the future as an additional attractive site for the design of antiviral drugs, rather than targeting the nsp16 active site only.

MTase has become an attractive target to develope inhibitors known to act on viral RNA capping and block the cap formation. Several inhibitors have been reported that bind to either the S-adenosylmethionine (SAM)/ S-adenosylhomocysteine (SAH) binding pocket or the Guanosine-5'-triphosphate (GTP) binding pocket in different viruses such as Dengue and West Nile viruses (Benarroch *et al*., 2004; H. Chen *et al*., 2013; Coutard *et al*., 2014; Lim *et al*., 2011; Stahla-Beek *et al*., 2012). In coronavirus, different screening assays have been used toward the identification of viral cap-methyltransferase inhibitors (Aouadi *et al*., 2017b; Sun *et al*., 2014), which provide a platform toward developing more specific inhibitors against coronavirus methyltransferases.

Protein purity is of key importance for development of much viral biochemistry such as immunodiagnostic assays, viral pathogenesis and structural studies (EM, SAXS, NMR, and crystallography). Expression hosts (mammalian, insect, yeast, bacterial, algal, and cell-free systems), media and expression time may affect any protein expression experiment and optimizing conditions for purification is the initial and major step to obtain a high yield of any protein of interest. The expression and purification of recombinant proteins is valuable for both clinical and investigational purposes. In comparison to other systems, bacterial systems are simple, easy to use and can give good protein yields (Baneyx and Mujacic, 2004). However, well studied expression hosts, such as *E. coli*, lack most of the posttranslational modification apparatus, which may results in poor folding or non-functional proteins, depending on the protein of interest. Recently, several modified strains have been developed that contain additional chaperones to assist the proper folding of proteins (e.g. BL21 (DE3) GroES/L) (Caspers *et al*., 1994; Endo *et al*., 2006; Luo and Hua, 1998). Other Strains contain rare tRNAs to help external eukaryotic proteins to express better (e.g. B834 (DE3) pRARE, BL21 (DE3) CodonPlus, Rosetta series). The type of media used for bacterial growth also has a signficant effect and both Luria Bertani (LB) and Auto induction media (AIM) media are used for bacterial growth for expression purposes. The first media is a simple media in which expression is induced by addition of IPTG while the auto induction system depends basically on a natural turn on of the lac promoter achived through a suitbale balance between the glucose and lactose in the media. Glucose depletion from the media results in cAMP increase in media and allows higher-level expression (Studier 2005). This technique for automated induction of cells does not require IPTG addition. One other main factor is the fusion tag(s), for facilitation of

purification. The polyhistidine (His) tag is the most common tag used for protein purification purposes. This tag is relatively small tag, with minor effect on the overall structure and does not interfere with the protein structure in most of the cases (Mason *et al*., 2001). Proteins are purified by a HisTrap^{TM} column in a process of immobilized metal ion chromatography (IMAC). An immobilsied metal ion in the matrix, usually nickel, preferably binds to the poly histidine residues on the protein surface and then can be eluted competitively with free imidazole, which essentially competes the nickel ion from the imidazole ring of the histidine residues.

In this chapter several parameters affecting the expression and purification of the MHV nsp16 protein expressed in *E. coli* were assessed. The ultimate goal of the work was in keeping with the other chapters of this thesis, that is, to focus on areas that might lead to the development of antiviral strategies. The immediate need however was to increase levels of expression and to develop rapid purification for nsp16 with a high level of purity and yield.

7.2 Results

7.2.1 Optimizing conditions of nsp16 expression and purification

As described in chapter 3, nsp16 was amplified from cDNA of MHV A59 and cloned into the pTriEX expression vector followed by transformation of *E. coli* BL21 (DE3)-pLysS competent cells and induction and screening for product **Figure 3.24**. Nsp16 was also expressed in LOBSTR (low background) competent cells as described in section 2.17.2. Since the nsp16 was successfully expressed in both strains, the next step was to optimize expression conditions for purification. One colony from each strain, that is, *E.coli* BL21 (DE3)-pLysS and LOBSTR cells previously shown to express nsp16 was inoculated into 10 ml LB media containing ampicillin and chloramphenicol and incubated at 37°C with shaking for 16hr. The following day, the starting culture was inoculated in Luria Bertani (LB) media with ampicillin, incubated in the shaker at 37°C until an OD_{600nm} of 0.6 was reached and protein expression induced by the addition of IPTG to 0.1mM final concentration. Cells were harvested by centrifugation after 3 hours induction with IPTG and the cell pellets were resuspended in binding buffer containing a proteinase inhibitor cocktail, sonicated, filtered and applied to a HisTrap™ HP column. The column was washed to background OD_{280} with binding buffer and the bound protein eluted using a linear gradient of imidazole from 0.02 to 0.5 M in binding buffer. The SDS-PAGE gel results showed a higher expression of nsp16 in LOBSTR cells compared to *E. coli* BL21 (DE3)-pLysS and the background was noticeably cleaner **Figure 7.75 and Figure 7.76**. When the LOBSTR extracts were subjected to Western blot using an anti His antibody, nsp16 presence was confirmed in the crude, lysates, the soluble column load and the eluted fractions **Figure 7.76B**.

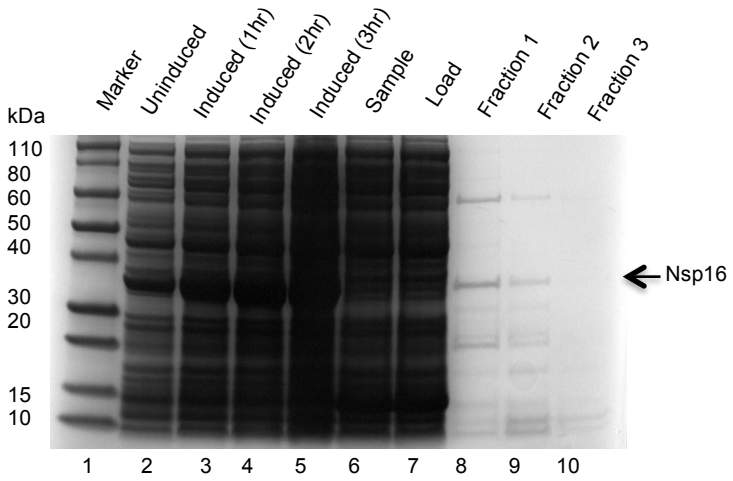

Figure 7.75. Expression and purification of nsp16 protein from BL21 (DE3)-pLysS competent cells. The SDS PAGE gel was stained with coomasie blue™. Nsp16 protein was transformed with vector pTriEX and induced by 1 mM IPTG for 3 hours at 37°C in LB media. Nsp16 expression culture was run through a Ni-affinity column designed for purification of 6x His tag bound recombinant proteins and eluted with an imidazole. Lanes represent different stages of nsp16 purification. 1: Marker; 2:total lysate (uninduced); 3: total lysate (after 1 hour induction); 4: total lysate (after 2 hours induction); 5: total lysate (after 3 hours induction); 6: soluble extract; 7: Ni column flow through; 8: Ni column elute fraction 1; 9: Ni column elute fraction 2; 10: Ni column elute fraction 3. The black arrow represents where a band of 34.5 kDa would run, representing the predicted weight of nsp16.

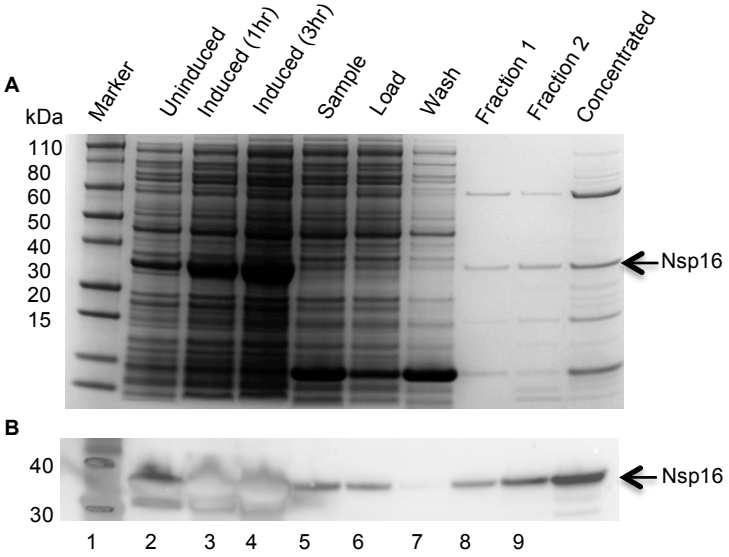

Figure 7.76. Expression and purification of nsp16 protein from LOBSTR competent cells. Nsp16 protein was transformed with vector pTriEX and induced by 1 mM IPTG for 3 hours at 37°C in LB media. Nsp16 expression culture was run through a Ni-affinity column designed for purification of 6-His-tag bound recombinant proteins and eluted with an imidazole. Lanes represent different stages of nsp16 purification. 1: Marker; 2:total lysate (uninduced); 3: total lysate (after 1 hour induction); 4: total lysate (after 3 hours induction); 5:soluble extract; 6:Ni column flow through; 7: Ni column after wash; 8: Ni column eluate fraction 1; 9: Ni column eluate fraction 2; 10: fraction 1 (concentrated). The black arrow represents where a band of 34.5 kDa would run, representing the predicted weight of nsp16. (A: The SDS PAGE gel was stained with coomasie blue™; B: Western blot).

Two media were also compared to choose best one for nsp16 protein purification. As before starting cultures were grown overnight and then inoculated into either Luria Bertani (LB) or Auto induction media (AIM) culture media with ampicilin and incubated with shaking at 37°C until an OD_{600nm} of 0.6 was reached. The autoinduction (AIM) culture was left to grow overnight at 20°C before cells were harvested by centrifugation while cells in LB media were harvested at 3 hours post IPTG induction. Cell pellets were resuspended in binding buffer containing a proteinase inhibitor cocktail, sonicated, filtered and again purified to a HisTrap™ HP column.

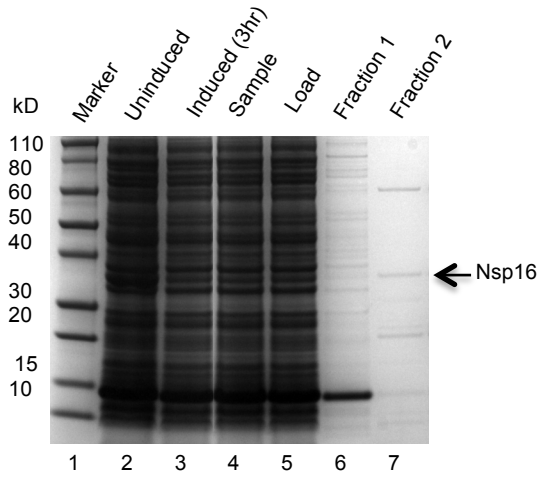

Figure 7.77. Expression and purification of nsp16 protein from LOBSTR competent cells in AIM media. Nsp16 protein was transformed with vector pTriEX and grown at 20°C in AIM media in the shaker. Nsp16 expression culture was run through a Ni-affinity column designed for purification of 6-His-tag bound recombinant proteins and eluted with an imidazole. Lanes represent different stages of nsp16 purification. 1: Marker; 2: total lysate (uninduced); 3: total lysate (after 3 hour induction); 4: soluble extract; 5: Ni column flow through; 6: Ni column eluate fraction 1; 7: Ni column eluate fraction 2. The black arrow represents where a band of 34.5 kDa would run, representing the predicted weight of nsp16.

The protein yield of nsp16 was found to be higher following growth in LB media **Figure 7.76** than in the AIM media **Figure 7.77**. Nsp16 protein concentration was estimated using a Bradford Assay, according to the manufacturer's protocol (Bio-Rad) and Nsp16 protein fractions were concentrated and stored frozen at -80°C for subsequent analysis.

7.3 Discussion

Pure and functionally active forms of coronavirus proteins are critical to the development of anti-viral drugs and vaccines. This chapter presents the work carried out for optimizing conditions required for coronavirus nsp16 purification. The nsp16 protein was shown to be best expressed from LOBSTR cells in LB media after 3 hours induction with IPTG which allowed for successful purification by HisTrap™ HP column chromatorgraphy.

To achieve this, trial experiment using two different bacterial strains were performed, with LOBSTR and *E.coli* BL21 (DE3)-pLysS. The nsp16 from both strains was found in the soluble fraction following sonication so either strain was suitable for the purification procedure. The results showed higher expression when nsp16 was grown in LOBSTR cells. The *E. coli* BL21 (DE3)-pLysS cells carry pLysS for toxic protein expression. It encodes T7 lysozyme which is a natural inhibitor of the T7 RNA polymerase. This gene can inactivate leaky expression of RNA polymerase efficiently but has no direct effect on the levsl of expression following induction. By contrast, LOBSTR cells are designed for the purification of low expressing proteins in *E. coli* by means of lowering the background contamination by two *E. coli* proteins with naturally high levels of His residues, *ArnA* and *SlyD* which bind to the resin during immobilized metal-affinity resins. LOBSTR is engineered to carry genomically modified copies of *arnA* and *SlyD* which encode these proteins and so abolish this histidine rich surface (Andersen *et al*., 2013).

A second experiment was performed to compare the expression level of nsp16 using two different media. Auto induction depends on ability of certain media to

induce protein expression in *E. coli* strains when cells reach saturation (Studier, 2005). Adjusting levels of glucose/lactose in media can regulate auto-induction but one of potential problems is that the expressed protein can be degraded. It is normal to obtain less amount of protein when the protein is expressed in AIM media since the nutrients provided by this media are limited unlike the nutrients provided by LB media. The temperature for expression was also different for both media. Nsp16 was successfully expressed at 37°C in LB media whereas nsp16 was expressed at 20°C in the AIM media. In the optimized protocol, nsp16 protein expression started after one hour of induction at 37°C with maximum soluble amount after 3 hours while harvesting of the culture after an overnight induction (16hr) at 20°C resulted in reduced amount of eluted protein. One reason for this maybe that nsp16 protein is degraded during the extended incubation period. Degradation by *E. coli* indigenous systems or by autolysis could be other reasons for target protein declining.

In conclusion, the optimal expression and purification of the target nsp16 protein was in *E. coli* LOBSTR in LB media with 3hr induction with IPTG at 37°C.

8 General discussion

It has been almost two decades since the first human coronavirus outbreaks were described and, despite progress in our understanding of the coronavirus genome and replication process; there are still no effective antivirals for this group of viruses (Graham *et al*., 2013; Perlman and Netland, 2009). Coronaviruses continue to emerge and cause great losses for both humans and animals (Choudhury *et al*., 2016; Weng *et al*., 2016; Zhang, 2016). To investigate the inhibitory role of coronavirus non mutated and mutated proteins on virus replication, it was necessary to express coronavirus proteins in mammalian cells that were compatible with coronavirus infection. First, In-Fusion cloning was used to produce DNA constructs encoding expressed proteins. Recombinant proteins were expressed in bacterial cells and 17clone-1 mammalian cells. Early attempts to express the coronavirus proteins using the bacterial expression system were successful in obtaining consistent and stable proteins, proving that the expression vectors produced were functional; successful cloning into the pTriEx vector was verified by sequencing, and Western blotting demonstrated bands of appropriate sizes for each protein following the induction of expression. However expression of these proteins in mammalian cells was rather difficult. Despite several attempts, it was not possible to detect MHV-A59 proteins in 17clone-1 mammalian cells with the exception of the N protein. At this stage it was not determined whether these proteins were not expressed, expressed in low level or affected by protease enzymes. Attempts to increase expression signal and reduce the effect of protease enzyme activity were tested but the proteins were not detected, and therefore an alternative strategy was explored (described in chapter 3).

Changing the expression strategy could increase the stability of the selected coronavirus proteins and for that, the SUMOStar tag was used. The covalent attachment of SUMOStar to cellular proteins constitutes a widespread mechanism that rapidly regulates protein function in response to a changing cellular environment. The coronavirus proteins (M protein, nsp9, nsp10, nsp16) were expressed in mammalian cells using this strategy **Figure 4.45** but despite many attempts, it was not possible to express proteins nsp7 and nsp8. Alternative tags or/and vectors may need to be considered to acheive expression of all targeted MHV proteins. Any inhibitory effect of non-mutated coronavirus proteins on virus replication was investagted by plaque assay to measure virus titre. Judging from this **Figure 4.47** the presence of some virus proteins could have been interfering with the virus replication but the mechanism by which they may inhibit virus infectivity is not clear. Several key questions remain to be answered: (i) why some coronavirus proteins are inhibiting virus infectivity and other proteins are increasing infectivity? (ii) why this effect is different between different viruses in the same family? For example, why M protein of SARS-CoV can suppress type I IFN (Lui *et al*., 2016) while M protein of HKU-1 does not (Siu *et al*., 2014) yet both are from the beta group in the coronaviridae family (iii) How can this be taken advantage of in developing antivirals? (described in chapter 4).

Information pertaining to how the presence of coronavirus proteins at the time of infection interfers with virus replication has been limited but in the last few years' studies has described the inhibitory and stimulatory effects of coronavirus proteins. N protein is well studied comparing with other coronavirus proteins in this respect. N protein is a multifunctional protein that can act as a RNA chaperone (Luo *et al*., 2004; Zeng *et al*., 2008; Zhao *et al*., 2008; Zúñiga *et al*., 2007) and interact with

other proteins (Chen *et al*., 2002). Several studies have shown that presence of coronavirus N protein could increase infectivity of the virus (Cui *et al*., 2015; Pan *et al*., 2008). In this study, the hypothesis was based on this data in that if mutated coronavirus N proteins, selected by directed evolution, could have diverse effect and inhibit virus replication they may lead to their consideration as antiviral models. In order to select proposed dominant negative variants a library of variants was required. Directed evolution hasn't proved to be a successful method to solve problems in protein engineering and metabolic engineering only, but also a powerful research tool for problems in protein structure, function and protein folding. *In vitro* directed evolution accelerates the process of natural selection however it requires the development of powerful high-throughput screening or selection methods. One of the most popular random mutagenesis methods is ep-PCR. Ep-PCR is based on increasing of the overall error frequency of *Taq* DNA polymerase by changing some PCR conditions that can increase this error rate, e.g. the addition of Mn^{2+} can reduce the base pairing specificity (Beckman *et al*., 1985). Two different ep-PCR protocols were tested here. The first protocol depended on combining different dNTP concentrations with the presence of Mn^{2+} but had the disadvantage of limited cloning efficiency. The second protocol depended on generating a library of variants by using different concentrations of Mn^{2+} to the PCR reaction buffer. The error rate was estimated for each library by sequencing and the library produced using 50 μM $MnCl_2$ achieved 1-2 mutations per individual amino acid within the whole N sequence and had high cloning efficiency. Thus, a strategy was developed for rapidly generating a library of variants with high cloning efficiency, suitable for studying active sites for any MHV protein. This library was used to screen for trans-dominat mutations of the N

protein but preliminary data was first obtained on the role of singly selected N mutations as proof of principle. Investigation of the inhibitory effect of a set of mutated (including deleted) N variants, supplied in trans, showed a variety of effects on MHV replication from stimulation to inhibition. However, the mechanism by which mutated N protein may inhibit and block virus infectivity, which may be important for designing better screens, remains to be determined (described in chapter 5).

Based on the data obtained by screening single mutations attempts at selection from libraries of N proteins with differing complexities was done. The initial “library” was a mixture of only two members, wild type N and mut38. After four rounds of selection with the two variants library, both the variant mut38 and WT N were still found in more or less equal numbers in 10 randomly selected clones. A second library was constructed from ten variants, all presvioulsy characterised at the individual level and examination of the complexity after 3 rounds of selection still showed a high diversity with little evidence of consistent selection occurring. It may be that effective selection requires many more rounds of selection than was possible in this study. A final attempt to select from a truly diverse ep-PCR library similarly failed to consistently demonstrate an emerging variant from this library after 4 rounds of selection and although changes in virus titre were observed albeit transiently, they were not significant (described in chapter 6).

A last chapter detailed work on the more general topic of MHV protein purification, especially for proteins for which limited structural data exists as purified recombinant proteins with high solubility and stability are a prerequisite for any structure based antivirial strategy. In the field of protein expression several

previous attempts have been made to express reasonable amounts of the nsp16 protein. An evaluation was done of the ability to express nsp16 protein with different expression hosts, media and induction times. The data improved on any previous work described but was still insufficient to progress this protein as a meaningful target for furhter study in this thesis (described in chapter 7).

This thesis study illustrates that significant effect of coronavirus proteins acting in trans on virus infectivity can be observed. The individual steps of applying this finding to the isolation of new trans dominnat mutants were also put in place, that is creating sequence diversity by using ep-PCR, showing transfection into target cells, demonstrating co-infection with MHV and showing recovery of library variants from transfected and infected cells. However combining these parameters failed to achive the desired goal of a randomly selected mutant and the reagents and techniques established will required further work. The goal, if achieved, could lead the way to the desiegn of pan-coronavirus inhbiting agents with great potential for the control of the diseases caused by coronaviruses, both currently circulating and emerging.

9 Appendix

Table 9.10. Coronavirus MHV-A59 protein sizes

Protein	Size (bp)	Size (a.a)	Size (kDa) With his tag	Size (kDa) With SUMO tag
$Nsp3^{Pro}$	774	248	29	43.5
Y-domain	1200	400	45.7	59.8
Nsp5	909	303	34.2	48.7
Nsp6	861	287	30.3	44.8
Nsp7	267	89	10.9	25.4
Nsp8	591	197	23	37.5
Nsp9	330	110	13.3	27.8
Nsp10	393	131	15	29.5
Nsp12	2784	924	106	120
Nsp12N	1902	364	43.3	57.8
Nsp12C	1692	564	68.2	82.7
Nsp16	900	300	34.3	48.8
M protein	687	229	27.1	41.6
N protein	1365	455	50.8	65.3

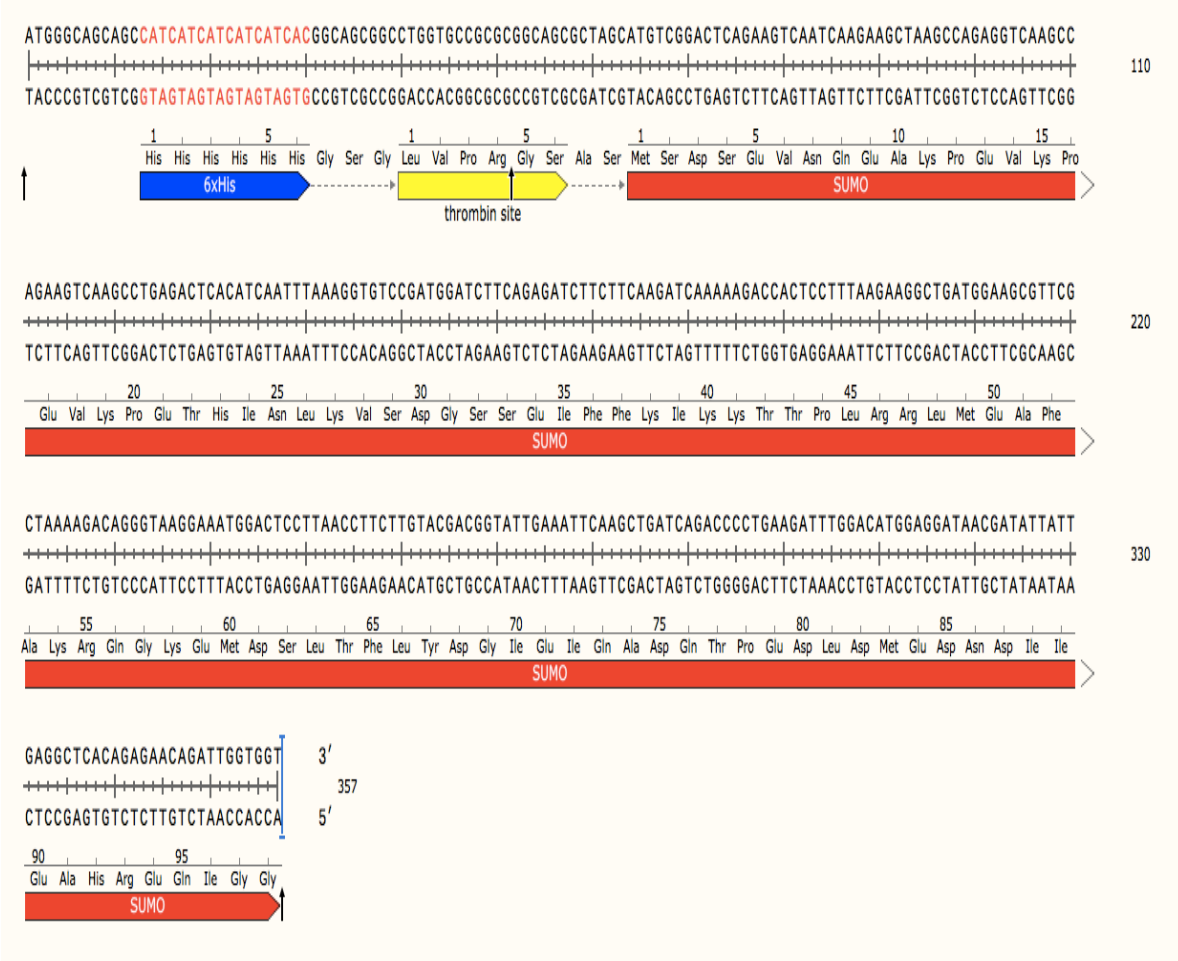

Figure 9.78. The sequence of SUMOStar tag protein. The SUMOStar tag contains a 6x His tag at the N-terminus.

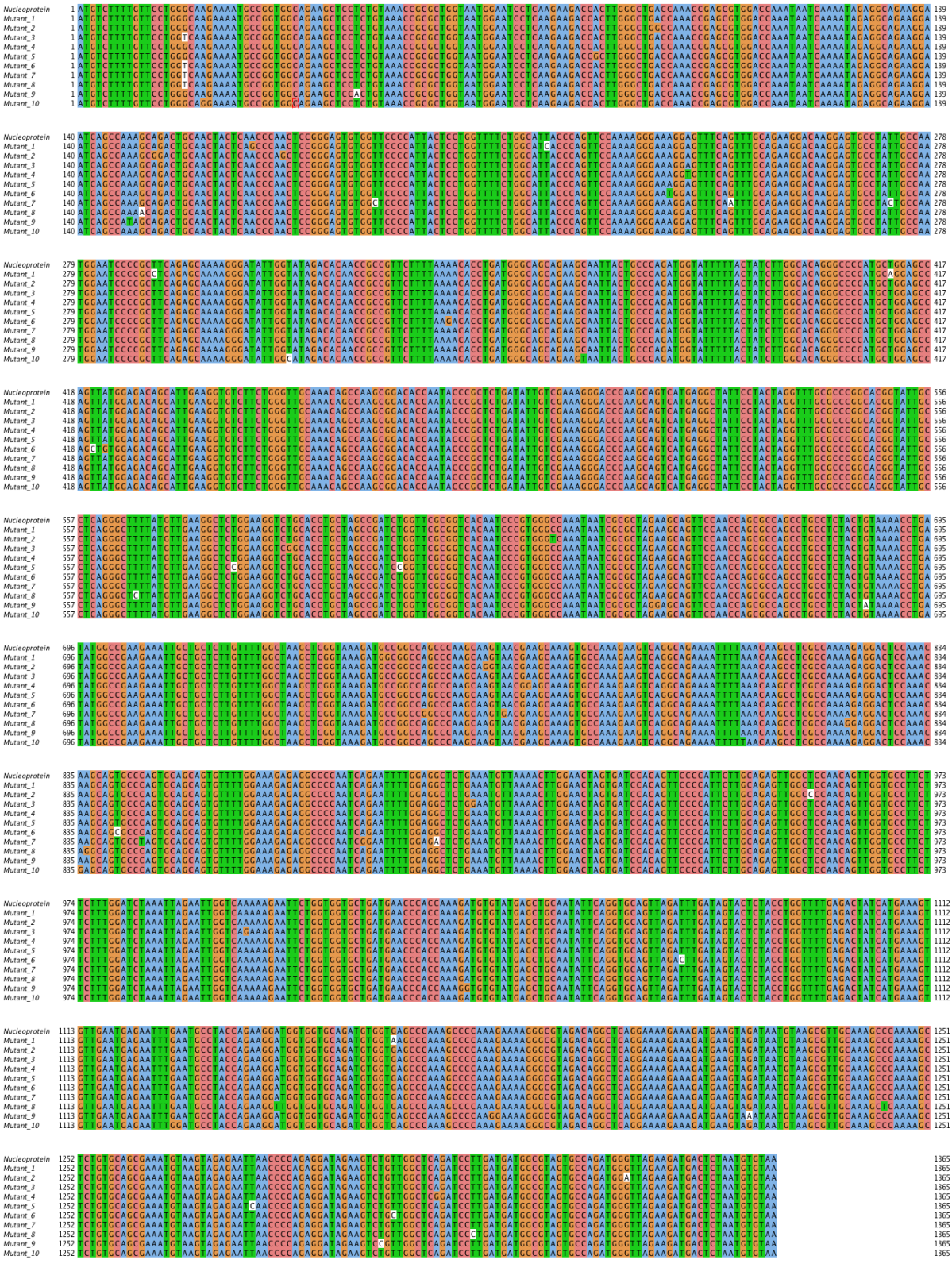

Figure 9.79. Distribution of nucleotide mutations in 10 variants generated by ep-PCR. Sequenced variants from 50μM $MnCl_2$ library are aligned with wild type using Jalview software.

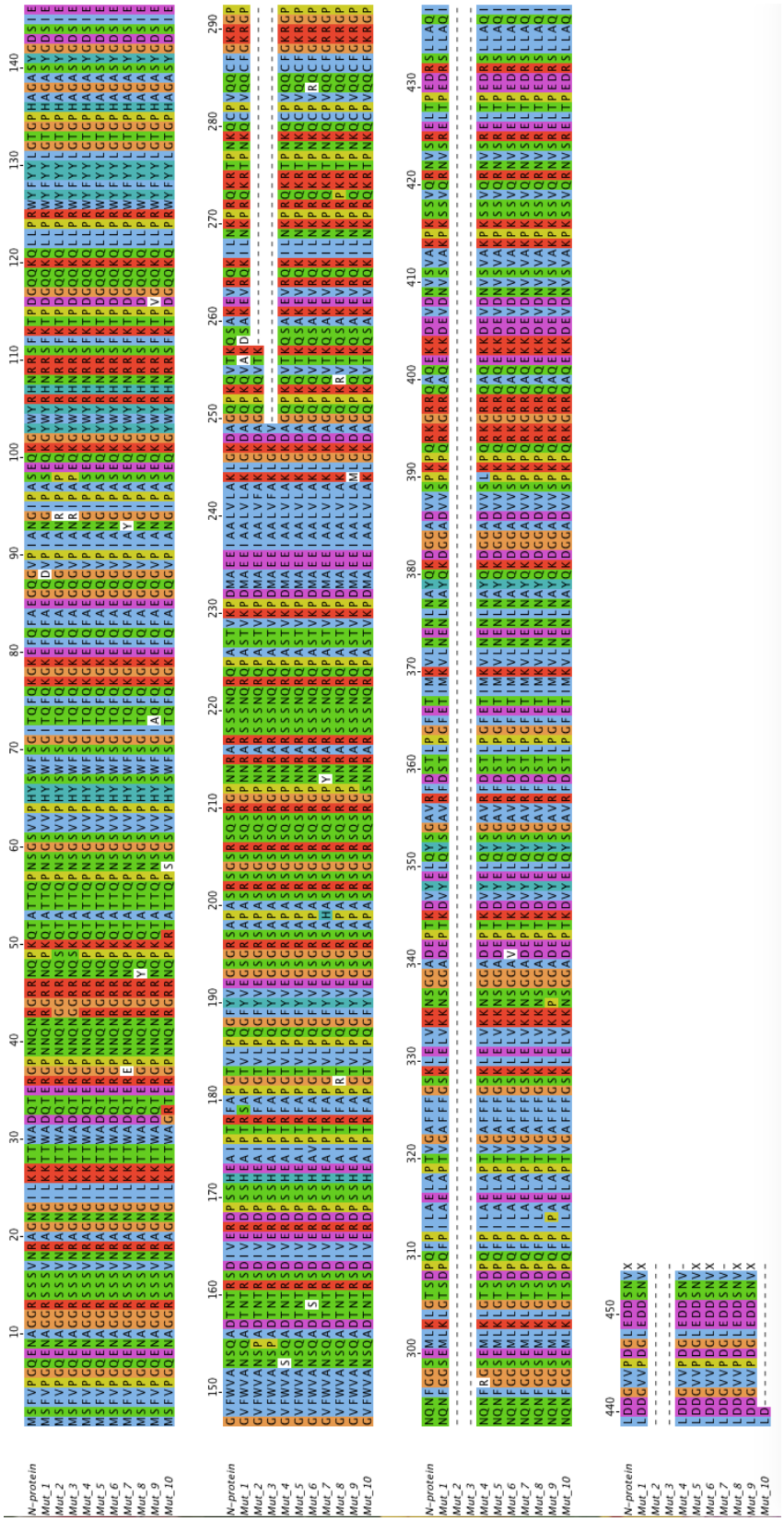

Figure 9.80. Distribution of nucleotide mutations in 10 variants generated by ep-PCR. Sequenced variants from 100μM $MnCl_2$ library are aligned with wild type using Jalview software.

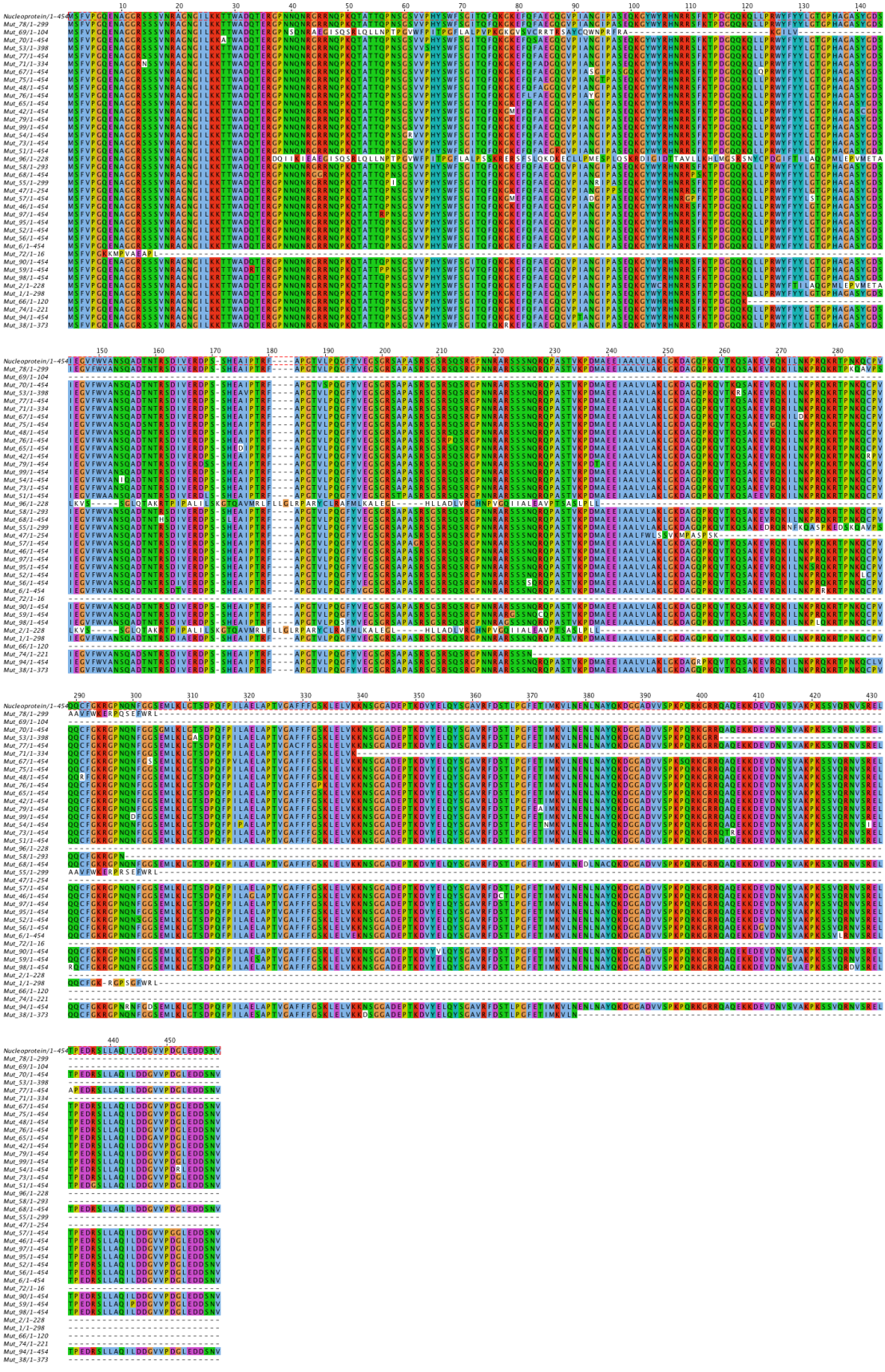

Figure 9.81. Distribution of amino acid mutations in 37 variants generated with ep-PCR and subjected to plaque assay. Sequenced variants are aligned with wild type using Jalview software.

References

AAkerström, S., Mirazimi, A., Tan, Y.-J., 2007. Inhibition of SARS-CoV replication cycle by small interference RNAs silencing specific SARS proteins, 7a/7b, 3a/3b and S. Antiviral Res. 73, 219–227.

Adedeji, A.O., Marchand, B., te Velthuis, A.J., Snijder, E.J., Weiss, S., Eoff, R.L., Singh, K., Sarafianos, S.G., 2012. Mechanism of nucleic acid unwinding by SARS-CoV helicase. PloS One 7, e36521.

Ahlquist, P., Noueiry, A.O., Lee, W.-M., Kushner, D.B., Dye, B.T., 2003. Host factors in positive-strand RNA virus genome replication. J. Virol. 77, 8181–8186.

Almazán, F., Galán, C., Enjuanes, L., 2004. The nucleoprotein is required for efficient coronavirus genome replication. J. Virol. 78, 12683–12688.

Andersen, K.R., Leksa, N.C., Schwartz, T.U., 2013. Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. Proteins Struct. Funct. Bioinforma. 81, 1857–1861.

Angelini, M.M., Akhlaghpour, M., Neuman, B.W., Buchmeier, M.J., 2013. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Proteins 3, 4, and 6 Induce Double-Membrane Vesicles. mBio 4, e00524–13. doi:10.1128/mBio.00524-13

Aouadi, W., Blanjoie, A., Vasseur, J.-J., Debart, F., Canard, B., Decroly, E., 2017a. Binding of the Methyl Donor S-Adenosyl-l-Methionine to Middle East Respiratory Syndrome Coronavirus 2′-O-Methyltransferase nsp16 Promotes Recruitment of the Allosteric Activator nsp10. J. Virol. 91, e02217–16.

Aouadi, W., Eydoux, C., Coutard, B., Martin, B., Debart, F., Vasseur, J.J., Contreras, J.M., Morice, C., Quérat, G., Jung, M.-L., others, 2017b. Toward the identification of viral cap-methyltransferase inhibitors by fluorescence screening assay. Antiviral Res. 144, 330–339.

Arnold, F.H., 2001. Combinatorial and computational challenges for biocatalyst design. Nature 409, 253–257.

Arnold, F.H., 1998. Design by directed evolution. Acc. Chem. Res. 31, 125–131.

Arnold, F.H., 1996. Directed evolution: creating biocatalysts for the future. Chem. Eng. Sci. 51, 5091–5102.

Báez-Santos, Y.M., Barraza, S.J., Wilson, M.W., Agius, M.P., Mielech, A.M., Davis, N.M., Baker, S.C., Larsen, S.D., Mesecar, A.D., 2014. X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases. J. Med. Chem. 57, 2393–2412.

Báez-Santos, Y.M., John, S.E.S., Mesecar, A.D., 2015. The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. Antiviral Res. 115, 21–38.

Baker, S.C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A.E., Koonin, E.V., Lai, M.M., 1993. Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. J. Virol. 67, 6056–6063.

Baneyx, F., Mujacic, M., 2004. Recombinant protein folding and misfolding in Escherichia coli. Nat. Biotechnol. 22, 1399–1408.

Baranowski, E., Ruiz-Jarabo, C.M., Domingo, E., 2001. Evolution of cell recognition by viruses. Science 292, 1102–1105.

Bárcena, M., Oostergetel, G.T., Bartelink, W., Faas, F.G., Verkleij, A., Rottier, P.J., Koster, A.J., Bosch, B.J., 2009. Cryo-electron tomography of mouse hepatitis virus: insights into the structure of the coronavirion. Proc. Natl. Acad. Sci. 106, 582–587.

Baric, Nelson, G.W., Fleming, J.O., Deans, R.J., Keck, J.G., Casteel, N., Stohlman, S.A., 1988. Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. J. Virol. 62, 4280–4287.

Barnard, D.L., Kumaki, Y., 2011. Recent developments in anti-severe acute respiratory syndrome coronavirus chemotherapy. Future Virol. 6, 615–631.

Barretto, N., Jukneliene, D., Ratia, K., Chen, Z., Mesecar, A.D., Baker, S.C., 2005. The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. J. Virol. 79, 15189–15198.

Bayer, P., Arndt, A., Metzger, S., Mahajan, R., Melchior, F., Jaenicke, R., Becker, J., 1998. Structure determination of the small ubiquitin-related modifier SUMO-1. J. Mol. Biol. 280, 275–286.

Beaudette, F.R., Hudson, C.B., 1937. Cultivation of the virus of infectious bronchitis. J Am Vet Med Assoc 90, 51–60.

Beckman, R.A., Mildvan, A.S., Loeb, L.A., 1985. On the fidelity of DNA replication: manganese mutagenesis in vitro. Biochemistry (Mosc.) 24, 5810–5817.

Bélanger, F., Stepinski, J., Darzynkiewicz, E., Pelletier, J., 2010. Characterization of hMTr1, a human Cap1 2′-O-ribose methyltransferase. J. Biol. Chem. 285, 33037–33044.

Belouzard, S., Millet, J.K., Licitra, B.N., Whittaker, G.R., 2012. Mechanisms of Coronavirus Cell Entry Mediated by the Viral Spike Protein. Viruses 4, 1011–1033. doi:10.3390/v4061011

Benarroch, D., Egloff, M.-P., Mulard, L., Guerreiro, C., Romette, J.-L., Canard, B., 2004. A Structural Basis for the Inhibition of the NS5 Dengue Virus mRNA 2′-O-Methyltransferase Domain by Ribavirin 5′-Triphosphate. J. Biol. Chem. 279, 35638–35643. doi:10.1074/jbc.M400460200

Beniac, D.R., Andonov, A., Grudeski, E., Booth, T.F., 2006. Architecture of the SARS coronavirus prefusion spike. Nat. Struct. Mol. Biol. 13, 751.

Bergman, S.J., Ferguson, M.C., Santanello, C., 2011. Interferons as therapeutic agents for infectious diseases. Infect. Dis. Clin. North Am. 25, 819–834.

Bershtein, S., Tawfik, D.S., 2008. Advances in laboratory evolution of enzymes. Curr. Opin. Chem. Biol. 12, 151–158.

Bohren, K.M., Nadkarni, V., Song, J.H., Gabbay, K.H., Owerbach, D., 2004. A M55V polymorphism in a novel SUMO gene (SUMO-4) differentially activates heat shock transcription factors and is associated with susceptibility to type I diabetes mellitus. J. Biol. Chem. 279, 27233–27238.

Bond, C.W., Leibowitz, J.L., Robb, J.A., 1979. Pathogenic murine coronaviruses II. Characterization of virus-specific proteins of murine coronaviruses JHMV and A59V. Virology 94, 371–384.

Bonilla, P.J., Hughes, S.A., Weiss, S.R., 1997. Characterization of a second cleavage site and demonstration of activity in trans by the papain-like proteinase of the murine coronavirus mouse hepatitis virus strain A59. J. Virol. 71, 900–909.

BOS, E.C., Luytjes, W., Van der Meulen, H., KOERTEN, H.K., SPAAN, W.J., 1996. The production of recombinant infectious DI-particles of a murine coronavirus in the absence of helper virus. Virology 218, 52–60.

Bosch, B.J., van der Zee, R., de Haan, C.A., Rottier, P.J., 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J. Virol. 77, 8801–8811.

Boursnell, M.E.G., Binns, M.M., Brown, T.D.K., 1985. Sequencing of coronavirus IBV genomic RNA: three open reading frames in the 5′ "unique" region of mRNA D. J. Gen. Virol. 66, 2253–2258.

Boussau, B., Blanquart, S., Necsulea, A., Lartillot, N., Gouy, M., 2008. Parallel adaptations to high temperatures in the Archaean eon. Nature 456, 942–945.

Bouvet, M., Debarnot, C., Imbert, I., Selisko, B., Snijder, E.J., Canard, B., Decroly, E., 2010. In vitro reconstitution of SARS-coronavirus mRNA cap methylation. PLoS Pathog 6, e1000863.

Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B., Decroly, E., 2012. RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. Proc. Natl. Acad. Sci. 109, 9372–9377.

Bredenbeek, P.J., Pachuk, C.J., Noten, A.F., Charité, J., Luytjes, W., Weiss, S.R., Spaan, W.J., 1990. The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. Nucleic Acids Res. 18, 1825–1832.

Brian, D.A., Baric, R.S., 2005. Coronavirus genome structure and replication, in: Coronavirus Replication and Reverse Genetics. Springer, pp. 1–30.

Brian, D.A., Hogue, B.G., Kienzle, T.E., 1995. The coronavirus hemagglutinin esterase glycoprotein, in: The Coronaviridae. Springer, pp. 165–179.

Brierley, I., Dos Ramos, F.J., 2006. Programmed ribosomal frameshifting in HIV-1 and the SARS–CoV. Virus Res. 119, 29–42. doi:10.1016/j.virusres.2005.10.008

Brockway, S.M., Clay, C.T., Lu, X.T., Denison, M.R., 2003. Characterization of the expression, intracellular localization, and replication complex association of

the putative mouse hepatitis virus RNA-dependent RNA polymerase. J. Virol. 77, 10515–10527.

Budzilowicz, C.J., Weiss, S.R., 1987. < i> In Vitro</i> synthesis of two polypeptides from a nonstructural gene of coronavirus mouse hepatitis virus strain A59. Virology 157, 509–515.

Butt, T.R., Edavettal, S.C., Hall, J.P., Mattern, M.R., 2005. SUMO fusion technology for difficult-to-express proteins. Protein Expr. Purif. 43, 1–9.

Cadwell, R.C., Joyce, G.F., 1992. Randomization of genes by PCR mutagenesis. Genome Res. 2, 28–33. doi:10.1101/gr.2.1.28

Capobianchi, M.R., Giombini, E., Rozera, G., 2013. Next-generation sequencing technology in clinical virology. Clin. Microbiol. Infect. 19, 15–22. doi:10.1111/1469-0691.12056

Caspers, P., Stieger, M., Burn, P., 1994. Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinases in Escherichia coli. Cell. Mol. Biol. Noisy--Gd. Fr. 40, 635–644.

Catanzariti, A.-M., Soboleva, T.A., Jans, D.A., Board, P.G., Baker, R.T., 2004. An efficient system for high-level expression and easy purification of authentic recombinant proteins. Protein Sci. 13, 1331–1339.

Cavanagh, D., 2005. Coronaviridae: a review of coronaviruses and toroviruses, in: Coronaviruses with Special Emphasis on First Insights Concerning SARS. Springer, pp. 1–54.

Chambers, P., Pringle, C.R., Easton, A.J., 1990. Heptad Repeat Sequences are Located Adjacent to Hydrophobic Regions in Several Types of Virus Fusion Glycoproteins. J. Gen. Virol. 71, 3075–3080. doi:10.1099/0022-1317-71-12-3075

Chan, J.F., Chan, K.-H., Kao, R.Y., To, K.K., Zheng, B.-J., Li, C.P., Li, P.T., Dai, J., Mok, F.K., Chen, H., others, 2013. Broad-spectrum antivirals for the emerging Middle East respiratory syndrome coronavirus. J. Infect. 67, 606–616.

Chan, J.F., Lau, S.K., To, K.K., Cheng, V.C., Woo, P.C., Yuen, K.-Y., 2015. Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. Clin. Microbiol. Rev. 28, 465–522.

Chan, J.F.-W., Yao, Y., Yeung, M.-L., Deng, W., Bao, L., Jia, L., Li, F., Xiao, C., Gao, H., Yu, P., others, 2015. Treatment with lopinavir/ritonavir or

interferon-β 1b improves outcome of MERS-CoV infection in a nonhuman primate model of common marmoset. J. Infect. Dis. 212, 1904–1913.

Chan, K.S., Lai, S.T., Chu, C.M., Tsui, E., Tam, C.Y., Wong, M.M.L., Tse, M.W., Que, T.L., Peiris, J.S.M., Sung, J., others, 2003. Treatment of severe acute respiratory syndrome with lopinavir/ritonavir: a multicentre retrospective matched cohort study. Hong Kong Med. J.

Chang, C., Chen, C.-M.M., Chiang, M., Hsu, Y., Huang, T., 2013. Transient oligomerization of the SARS-CoV N protein–implication for virus ribonucleoprotein packaging. PloS One 8, e65045.

Chang, C., Hou, M.-H., Chang, C.-F., Hsiao, C.-D., Huang, T., 2014. The SARS coronavirus nucleocapsid protein–forms and functions. Antiviral Res. 103, 39–50.

Chang, C., Sue, S.-C., Yu, T., Hsieh, C.-M., Tsai, C.-K., Chiang, Y.-C., Lee, S., Hsiao, H., Wu, W.-J., Chang, W.-L., others, 2006. Modular organization of SARS coronavirus nucleocapsid protein. J. Biomed. Sci. 13, 59–72.

Changsen, W., CASTRO, A.F., WILKES, D.M., ALTENBERG, G.A., 1999. Expression and purification of the first nucleotide-binding domain and linker region of human multidrug resistance gene product: comparison of fusions to glutathione S-transferase, thioredoxin and maltose-binding protein. Biochem. J. 338, 77–81.

Chaput, J.C., Woodbury, N.W., Stearns, L.A., Williams, B.A., 2008. Creating protein biocatalysts as tools for future industrial applications. Expert Opin. Biol. Ther. 8, 1087–1098.

Chen, H., Liu, L., Jones, S.A., Banavali, N., Kass, J., Li, Z., Zhang, J., Kramer, L.D., Ghosh, A.K., Li, H., 2013. Selective inhibition of the West Nile virus methyltransferase by nucleoside analogs. Antiviral Res. 97, 232–239. doi:10.1016/j.antiviral.2012.12.012

Chen, H., Wurm, T., Britton, P., Brooks, G., Hiscox, J.A., 2002. Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell. J. Virol. 76, 5233–5250.

Chen, I.-J., Yuann, J.-M.P., Chang, Y.-M., Lin, S.-Y., Zhao, J., Perlman, S., Shen, Y.-Y., Huang, T.-H., Hou, M.-H., 2013. Crystal structure-based exploration of the important role of Arg106 in the RNA-binding domain of human

coronavirus OC43 nucleocapsid protein. Biochim. Biophys. Acta BBA-Proteins Proteomics 1834, 1054–1062.

Chen, Y., Su, C., Ke, M., Jin, X., Xu, L., Zhang, Z., Wu, A., Sun, Y., Yang, Z., Tien, P., others, 2011. Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2′-O-methylation by nsp16/nsp10 protein complex. PLoS Pathog 7, e1002294.

Cheng, V.C., Chan, J.F., To, K.K., Yuen, K.Y., 2013. Clinical management and infection control of SARS: lessons learned. Antiviral Res. 100, 407–419.

Cheng, V.C., Lau, S.K., Woo, P.C., Yuen, K.Y., 2007. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. Clin. Microbiol. Rev. 20, 660–694.

Choudhury, B., Dastjerdi, A., Doyle, N., Frossard, J.-P., Steinbach, F., 2016. From the field to the lab—An European view on the global spread of PEDV. Virus Res. 226, 40–49.

Chu, C.M., Cheng, V.C.C., Hung, I.F.N., Wong, M.M.L., Chan, K.H., Chan, K.S., Kao, R.Y.T., Poon, L.L.M., Wong, C.L.P., Guan, Y., others, 2004. Role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. Thorax 59, 252–256.

Chu, D.K., Leung, C.Y., Gilbert, M., Joyner, P.H., Ng, E.M., Tsemay, M.T., Guan, Y., Peiris, J.S., Poon, L.L., 2011. Avian coronavirus in wild aquatic birds. J. Virol. 85, 12815–12820.

Cinatl, J., Morgenstern, B., Bauer, G., Chandra, P., Rabenau, H., Doerr, H.W., 2003. Treatment of SARS with human interferons. The Lancet 362, 293–294.

Cobb, R.E., Sun, N., Zhao, H., 2013. Directed evolution as a powerful synthetic biology tool. Methods 60, 81–90.

Compton, S.R., Rogers, D.B., Holmes, K.V., Fertsch, D., Remenick, J., McGowan, J.J., 1987. In vitro replication of mouse hepatitis virus strain A59. J. Virol. 61, 1814–1820.

Corse, E., Machamer, C.E., 2003. The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction. Virology 312, 25–34.

Corse, E., Machamer, C.E., 2000. Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles. J. Virol. 74, 4319–4326.

Cottam, E.M., Maier, H.J., Manifava, M., Vaux, L.C., Chandra-Schoenfelder, P., Gerner, W., Britton, P., Ktistakis, N.T., Wileman, T., 2011. Coronavirus nsp6 proteins generate autophagosomes from the endoplasmic reticulum via an omegasome intermediate. Autophagy 7, 1335–1347.

Coughlin, M.M., Prabhakar, B.S., 2012. Neutralizing human monoclonal antibodies to severe acute respiratory syndrome coronavirus: target, mechanism of action, and therapeutic potential. Rev. Med. Virol. 22, 2–17.

Cougot, N., van Dijk, E., Babajko, S., Séraphin, B., 2004. Cap-tabolism. Trends Biochem. Sci. 29, 436–444.

Coutard, B., Decroly, E., Li, C., Sharff, A., Lescar, J., Bricogne, G., Barral, K., 2014. Assessment of Dengue virus helicase and methyltransferase as targets for fragment-based drug discovery. Antiviral Res. 106, 61–70. doi:10.1016/j.antiviral.2014.03.013

Cox, E.C., 1976. Bacterial mutator genes and the control of spontaneous mutation. Annu. Rev. Genet. 10, 135–156.

Cui, L., Wang, H., Ji, Y., Yang, J., Xu, S., Huang, X., Wang, Z., Qin, L., Tien, P., Zhou, X., Guo, D., Chen, Y., 2015. The Nucleocapsid Protein of Coronaviruses Acts as a Viral Suppressor of RNA Silencing in Mammalian Cells. J. Virol. 89, 9029–9043. doi:10.1128/JVI.01331-15

D. Voges, P. Zwickl, Baumeister, and W., 1999. The 26S Proteasome: A Molecular Machine Designed for Controlled Proteolysis. Annu. Rev. Biochem. 68, 1015–1068. doi:10.1146/annurev.biochem.68.1.1015

Daffis, S., Szretter, K.J., Schriewer, J., Li, J., Youn, S., Errett, J., Lin, T.-Y., Schneller, S., Zust, R., Dong, H., 2010. 2 [prime]-O methylation of the viral mRNA cap evades host restriction by IFIT family members. Nature 468, 452–456.

Davies, H.A., Macnaughton, M.R., 1979. Comparison of the morphology of three coronaviruses. Arch. Virol. 59, 25–33.

De Groot,RJ, AMQ, K., Adams, M.J., Cartens, E., Lefkowitz, E.J., 2012. Virus taxonomy, the 9th report of the international committe on taxonomy of viruses. Academic Press,San Diego,CA.

De Haan, C.A., de Wit, M., Kuo, L., Montalto-Morrison, C., Haagmans, B.L., Weiss, S.R., Masters, P.S., Rottier, P.J., 2003. The glycosylation status of the murine hepatitis coronavirus M protein affects the interferogenic capacity of the virus in vitro and its ability to replicate in the liver but not the brain. Virology 312, 395–406.

De Haan, C.A., Smeets, M., Vernooij, F., Vennema, H., Rottier, P.J.M., 1999. Mapping of the coronavirus membrane protein domains involved in interaction with the spike protein. J. Virol. 73, 7441–7452.

De Haan, C.A., Stadler, K., Godeke, G.-J., Bosch, B.J., Rottier, P.J., 2004. Cleavage inhibition of the murine coronavirus spike protein by a furin-like enzyme affects cell-cell but not virus-cell fusion. J. Virol. 78, 6048–6054.

De Haan, C.A., Vennema, H., Rottier, P.J., 2000. Assembly of the coronavirus envelope: homotypic interactions between the M proteins. J. Virol. 74, 4967–4978.

De Marco, V., Stier, G., Blandin, S., De Marco, A., 2004. The solubility and stability of recombinant proteins are increased by their fusion to NusA. Biochem. Biophys. Res. Commun. 322, 766–771.

De Wilde, A.H., Jochmans, D., Posthuma, C.C., Zevenhoven-Dobbe, J.C., van Nieuwkoop, S., Bestebroer, T.M., van den Hoogen, B.G., Neyts, J., Snijder, E.J., 2014. Screening of an FDA-approved compound library identifies four small-molecule inhibitors of Middle East respiratory syndrome coronavirus replication in cell culture. Antimicrob. Agents Chemother. 58, 4875–4884.

De Wilde, A.H., Raj, V.S., Oudshoorn, D., Bestebroer, T.M., van Nieuwkoop, S., Limpens, R.W., Posthuma, C.C., van der Meer, Y., Bárcena, M., Haagmans, B.L., others, 2013. MERS-coronavirus replication induces severe in vitro cytopathology and is strongly inhibited by cyclosporin A or interferon-α treatment. J. Gen. Virol. 94, 1749–1760.

Debarnot, C., Imbert, I., Ferron, F., Gluais, L., Varlet, I., Papageorgiou, N., Bouvet, M., Lescar, J., Decroly, E., Canard, B., 2011. Crystallization and diffraction analysis of the SARS coronavirus nsp10–nsp16 complex. Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. 67, 404–408.

Decroly, E., Debarnot, C., Ferron, F., Bouvet, M., Coutard, B., Imbert, I., Gluais, L., Papageorgiou, N., Sharff, A., Bricogne, G., 2011. Crystal structure and

functional analysis of the SARS-coronavirus RNA cap 2′-O-methyltransferase nsp10/nsp16 complex. PLoS Pathog. 7, e1002059.

Decroly, E., Ferron, F., Lescar, J., Canard, B., 2012. Conventional and unconventional mechanisms for capping viral mRNA. Nat. Rev. Microbiol. 10, 51–65.

Decroly, E., Imbert, I., Coutard, B., Bouvet, M., Selisko, B., Alvarez, K., Gorbalenya, A.E., Snijder, E.J., Canard, B., 2008. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2′O)-methyltransferase activity. J. Virol. 82, 8071–8084.

Den Boon, J.A., Snijder, E.J., Chirnside, E.D., De Vries, A.A., Horzinek, M.C., Spaan, W.J., 1991. Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. J. Virol. 65, 2910–2920.

Denison, M.R., Zoltick, P.W., Hughes, S.A., Giangreco, B., Olson, A.L., Perlman, S., Leibowitz, J.L., Weiss, S.R., 1992. Intracellular processing of the N-terminal ORF 1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events. Virology 189, 274–284.

Denizot, F., Lang, R., 1986. Rapid colorimetric assay for cell growth and survival: modifications to the tetrazolium dye procedure giving improved sensitivity and reliability. J. Immunol. Methods 89, 271–277.

Deshler, J.O., 1992. A simple method for randomly mutating cloned DNA fragments by using chemical mutagens and the polymerase chain reaction. Genet. Anal. Biomol. Eng. 9, 103–106.

Ding, Z., Fang, L., Yuan, S., Zhao, L., Wang, X., Long, S., Wang, M., Wang, D., Foda, M.F., Xiao, S., 2017. The nucleocapsid proteins of mouse hepatitis virus and severe acute respiratory syndrome coronavirus share the same IFN-β antagonizing mechanism: attenuation of PACT-mediated RIG-I/MDA5 activation. Oncotarget 8, 49655–49670. doi:10.18632/oncotarget.17912

Donaldson, E.F., Graham, R.L., Sims, A.C., Denison, M.R., Baric, R.S., 2007. Analysis of murine hepatitis virus strain A59 temperature-sensitive mutant TS-LA6 suggests that nsp10 plays a critical role in polyprotein processing. J. Virol. 81, 7086–7098.

Dougherty, W.G., Semler, B.L., 1993. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. Microbiol. Rev. 57, 781–822.

Doyle, L.P., Hutchings, L.M., 1946. A transmissible gastroenteritis in pigs. J. Am. Vet. Med. Assoc. 108, 257–259.

Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B.-J., Jiang, S., 2009. The spike protein of SARS-CoV—a target for vaccine and therapeutic development. Nat. Rev. Microbiol. 7, 226–236.

Du, L., Zhao, G., Yang, Y., Qiu, H., Wang, L., Kou, Z., Tao, X., Yu, H., Sun, S., Tseng, C.-T.K., others, 2014. A conformation-dependent neutralizing monoclonal antibody specifically targeting receptor-binding domain in Middle East respiratory syndrome coronavirus spike protein. J. Virol. 88, 7045–7053.

Dyall, J., Coleman, C.M., Hart, B.J., Venkataraman, T., Holbrook, M.R., Kindrachuk, J., Johnson, R.F., Olinger, G.G., Jahrling, P.B., Laidlaw, M., others, 2014. Repurposing of clinically developed drugs for treatment of Middle East respiratory syndrome coronavirus infection. Antimicrob. Agents Chemother. 58, 4885–4893.

Dye, C., Siddell, S.G., 2005. Genomic RNA sequence of Feline coronavirus strain FIPV WSU-79/1146. J. Gen. Virol. 86, 2249–2253.

Eckert, D.M., Kim, P.S., 2001. Mechanisms of viral membrane fusion and its inhibition. Annu. Rev. Biochem. 70, 777–810.

Egloff, M.-P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E.J., Gorbalenya, A.E., Cambillau, C., Canard, B., 2004. The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world. Proc. Natl. Acad. Sci. U. S. A. 101, 3792–3796. doi:10.1073/pnas.0307877101

El-Deiry, W.S., Downey, K.M., So, A.G., 1984. Molecular mechanisms of manganese mutagenesis. Proc. Natl. Acad. Sci. 81, 7378–7382.

Elshabrawy, H.A., Fan, J., Haddad, C.S., Ratia, K., Broder, C.C., Caffrey, M., Prabhakar, B.S., 2014. Identification of a broad-spectrum antiviral small molecule against severe acute respiratory syndrome coronavirus and Ebola, Hendra, and Nipah viruses by using a novel high-throughput screening assay. J. Virol. 88, 4353–4365.

Endo, S., Tomimoto, Y., Shimizu, H., Taniguchi, Y., Onizuka, T., 2006. Effects of E. coli chaperones on the solubility of human receptors in an in vitro expression system. Mol. Biotechnol. 33, 199–209.

Escors, D., Ortego, J., Laude, H., Enjuanes, L., 2001. The membrane M protein carboxy terminus binds to transmissible gastroenteritis coronavirus core and contributes to core stability. J. Virol. 75, 1312–1324.

Falzarano, D., De Wit, E., Martellaro, C., Callison, J., Munster, V.J., Feldmann, H., 2013. Inhibition of novel β coronavirus replication by a combination of interferon-α2b and ribavirin. Sci. Rep. 3, 1686.

Fan, H., Ooi, A., Tan, Y.W., Wang, S., Fang, S., Liu, D.X., Lescar, J., 2005. The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties. Structure 13, 1859–1868.

Fang, S., Chen, B., Tay, F.P., Ng, B.S., Liu, D.X., 2007. An arginine-to-proline mutation in a domain with undefined functions within the helicase protein (Nsp13) is lethal to the coronavirus infectious bronchitis virus in cultured cells. Virology 358, 136–147.

Fang, X., Ye, L., Timani, K.A., Li, S., Zen, Y., Zhao, M., Zheng, H., Wu, Z., 2005. Peptide domain involved in the interaction between membrane protein and nucleocapsid protein of SARS-associated coronavirus. BMB Rep. 38, 381–385.

Fang, X., Ye, L., Zhang, Y., Li, B., Li, S., Kong, L., Wang, Y., Zheng, H., Wang, W., Wu, Z., 2006. Nucleocapsid amino acids 211 to 254, in particular, tetrad glutamines, are essential for the interaction between the nucleocapsid and membrane proteins of SARS-associated coronavirus. J. Microbiol.-SEOUL- 44, 577.

Ferrari, M., Fornasiero, M.C., Isetta, A.M., 1990. MTT colorimetric assay for testing macrophage cytotoxic activity in vitro. J. Immunol. Methods 131, 165–172.

Ficken, M.D., Barnes, H.J., Qureshi, M.A., 1991. Acute airsacculitis in turkeys inoculated with cell-free culture filtrate of Pasteurella multocida. Vet. Pathol. Online 28, 46–54.

Flotho, A., Melchior, F., 2013. Sumoylation: a regulatory protein modification in health and disease. Annu. Rev. Biochem. 82, 357–385.

Ford, R.B., 1986. Biological response modifiers in the management of viral infection. Vet. Clin. North Am. Small Anim. Pract. 16, 1191–1204.

Fox, R.J., Huisman, G.W., 2008. Enzyme optimization: moving from blind evolution to statistical exploration of sequence–function space. Trends Biotechnol. 26, 132–138.

Fromant, M., Blanquet, S., Plateau, P., 1995. Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. Anal. Biochem. 224, 347–353.

Furuichi, Y., Shatkin, A.J., 2000. Viral and cellular mRNA capping: past and prospects. Adv. Virus Res. 55, 135–184.

Gadlage, M.J., Denison, M.R., 2010. Exchange of the Coronavirus Replicase Polyprotein Cleavage Sites Alters Protease Specificity and Processing. J. Virol. 84, 6894–6898. doi:10.1128/JVI.00752-10

Gallagher, T.M., Buchmeier, M.J., 2001. Coronavirus spike proteins in viral entry and pathogenesis. Virology 279, 371–374.

Gao, J., Lu, G., Qi, J., Li, Y., Wu, Y., Deng, Y., Geng, H., Li, H., Wang, Q., Xiao, H., others, 2013. Structure of the fusion core and inhibition of fusion by a heptad repeat peptide derived from the S protein of Middle East respiratory syndrome coronavirus. J. Virol. 87, 13134–13140.

Gerlier, D., Thomasset, N., 1986. Use of MTT colorimetric assay to measure cell activation. J. Immunol. Methods 94, 57–63.

Giger, L., Caner, S., Obexer, R., Kast, P., Baker, D., Ban, N., Hilvert, D., 2013. Evolution of a designed retro-aldolase leads to complete active site remodeling. Nat. Chem. Biol. 9, 494–498.

Gill, G., 2004. SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? Genes Dev. 18, 2046–2059.

Gillim-Ross, L., Subbarao, K., 2006. Emerging respiratory viruses: challenges and vaccine strategies. Clin. Microbiol. Rev. 19, 614–636.

Glowacka, I., Bertram, S., Herzog, P., Pfefferle, S., Steffen, I., Muench, M.O., Simmons, G., Hofmann, H., Kuri, T., Weber, F., Eichler, J., Drosten, C., Pöhlmann, S., 2010. Differential Downregulation of ACE2 by the Spike Proteins of Severe Acute Respiratory Syndrome Coronavirus and Human Coronavirus NL63. J. Virol. 84, 1198–1205. doi:10.1128/JVI.01248-09

Gonzalez, J.M., Gomez-Puertas, P., Cavanagh, D., Gorbalenya, A.E., Enjuanes, L., 2003. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. Arch. Virol. 148, 2207–2235.

Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. < i> Nidovirales</i>: Evolving the largest RNA virus genome. Virus Res. 117, 17–37.

Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., Blinov, V.M., 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. 17, 4713–4730.

Gorbalenya, A.E., Pringle, F.M., Zeddam, J.-L., Luke, B.T., Cameron, C.E., Kalmakoff, J., Hanzlik, T.N., Gordon, K.H., Ward, V.K., 2002. The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. J. Mol. Biol. 324, 47–62.

Gosert, R., Kanjanahaluethai, A., Egger, D., Bienz, K., Baker, S.C., 2002. RNA replication of mouse hepatitis virus takes place at double-membrane vesicles. J. Virol. 76, 3697–3708.

Graham, R.L., Denison, M.R., 2006. Replication of Murine Hepatitis Virus Is Regulated by Papain-Like Proteinase 1 Processing of Nonstructural Proteins 1, 2, and 3. J. Virol. 80, 11610–11620. doi:10.1128/JVI.01428-06

Graham, R.L., Donaldson, E.F., Baric, R.S., 2013. A decade after SARS: strategies for controlling emerging coronaviruses. Nat. Rev. Microbiol. 11, 836–848.

Greener, A., Callahan, M., Jerpseth, B., 1996. An efficient random mutagenesis technique using an E. coli mutator strain, in: In Vitro Mutagenesis Protocols. Springer, pp. 375–385.

Grossoehme, N.E., Li, L., Keane, S.C., Liu, P., Dann, C.E., Leibowitz, J.L., Giedroc, D.P., 2009. Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA duplexes. J. Mol. Biol. 394, 544–557.

Gu, M., Lima, C.D., 2005. Processing the message: structural insights into capping and decapping mRNA. Curr. Opin. Struct. Biol. 15, 99–106.

Guarino, L.A., Bhardwaj, K., Dong, W., Sun, J., Holzenburg, A., Kao, C., 2005. Mutational analysis of the SARS virus Nsp15 endoribonuclease:

identification of residues affecting hexamer formation. J. Mol. Biol. 353, 1106–1117.

Haan, C.A.M. de, Kuo, L., Masters, P.S., Vennema, H., Rottier, P.J.M., 1998. Coronavirus Particle Assembly: Primary Structure Requirements of the Membrane Protein. J. Virol. 72, 6838–6850.

Hamilton, M.D., Nuara, A.A., Gammon, D.B., Buller, R.M., Evans, D.H., 2007. Duplex strand joining reactions catalyzed by vaccinia virus DNA polymerase. Nucleic Acids Res. 35, 143–151. doi:10.1093/nar/gkl1015

Hamre, D., Procknow, J.J., 1966. A new virus isolated from the human respiratory tract., in: Proceedings of the Society for Experimental Biology and Medicine. Society for Experimental Biology and Medicine (New York, NY). pp. 190–193.

Harcourt, B.H., Jukneliene, D., Kanjanahaluethai, A., Bechill, J., Severson, K.M., Smith, C.M., Rota, P.A., Baker, S.C., 2004. Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. J. Virol. 78, 13600–13612.

Harger, J.W., Meskauskas, A., Dinman, J.D., 2002. An “integrated model”of programmed ribosomal frameshifting. Trends Biochem. Sci. 27, 448–454.

Harrison, S.C., 2008. Viral membrane fusion. Nat. Struct. Mol. Biol. 15, 690–698.

Hatakeyama, S., Matsuoka, Y., Ueshiba, H., Komatsu, N., Itoh, K., Shichijo, S., Kanai, T., Fukushi, M., Ishida, I., Kirikae, T., others, 2008. Dissection and identification of regions required to form pseudoparticles by the interaction between the nucleocapsid (N) and membrane (M) proteins of SARS coronavirus. Virology 380, 99–108.

He, M.L., Zheng, B.J., Chen, Y., Wong, K.L., Huang, J.D., Lin, M.C., Yuen, K.Y., Sung, J.J., Kung, H.F., 2009. Development of interfering RNA agents to inhibit SARS-associated coronavirus infection and replication. Hong Kong Med. J. Xianggang Yi Xue Za Zhi.

He, R., Leeson, A., Ballantine, M., Andonov, A., Baker, L., Dobie, F., Li, Y., Bastien, N., Feldmann, H., Strocher, U., others, 2004. Characterization of protein–protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus. Virus Res. 105, 121–125.

He, Y., Lu, H., Siddiqui, P., Zhou, Y., Jiang, S., 2005. Receptor-binding domain of severe acute respiratory syndrome coronavirus spike protein contains

multiple conformation-dependent epitopes that induce highly potent neutralizing antibodies. J. Immunol. 174, 4908–4915.

Hilgenfeld, R., 2014. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. FEBS J. 281, 4085–4096.

Hilgenfeld, R., Peiris, M., 2013. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. Antiviral Res. 100, 286–295.

Hochstrasser, M., 2000. Evolution and function of ubiquitin-like protein-conjugation systems. Nat. Cell Biol. 2, E153–E157.

Hogue, B.G., Machamer, C.E., 2008. Coronavirus structural proteins and virus assembly. Nidoviruses ASM Press Wash. DC 179–200.

Hou, L., 2009. Novel methods of genome shuffling in Saccharomyces cerevisiae. Biotechnol. Lett. 31, 671–677.

Hsieh, P.-K., Chang, S.C., Huang, C.-C., Lee, T.-T., Hsiao, C.-W., Kou, Y.-H., Chen, I.-Y., Chang, C.-K., Huang, T.-H., Chang, M.-F., 2005. Assembly of Severe Acute Respiratory Syndrome Coronavirus RNA Packaging Signal into Virus-Like Particles Is Nucleocapsid Dependent. J. Virol. 79, 13848–13855. doi:10.1128/JVI.79.22.13848-13855.2005

Huang, Q., Yu, L., Petros, A.M., Gunasekera, A., Liu, Z., Xu, N., Hajduk, P., Mack, J., Fesik, S.W., Olejniczak, E.T., 2004. Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein. Biochemistry (Mosc.) 43, 6059–6063.

Huang, Y., Yang, Z., Kong, W., Nabel, G.J., 2004. Generation of synthetic severe acute respiratory syndrome coronavirus pseudoparticles: implications for assembly and vaccine production. J. Virol. 78, 12557–12565.

Hung, H.-C., Liu, C.-L., Hsu, J.T.-A., Horng, J.-T., Fang, M.-Y., Wu, S.-Y., Ueng, S.-H., Wang, M.-Y., Yaw, C.-W., Hou, M.-H., 2012. Development of an anti-influenza drug screening assay targeting nucleoproteins with tryptophan fluorescence quenching. Anal. Chem. 84, 6391–6399.

Hurst, K.R., Koetzner, C.A., Masters, P.S., 2013. Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. J. Virol. 87, 9159–9172.

Hurst, K.R., Kuo, L., Koetzner, C.A., Ye, R., Hsue, B., Masters, P.S., 2005. A major determinant for membrane protein interaction localizes to the

carboxy-terminal domain of the mouse coronavirus nucleocapsid protein. J. Virol. 79, 13285–13297.

Hurst, K.R., Ye, R., Goebel, S.J., Jayaraman, P., Masters, P.S., 2010. An interaction between the nucleocapsid protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA. J. Virol. 84, 10276–10288.

Hurst-Hess, K.R., Kuo, L., Masters, P.S., 2015. Dissection of amino-terminal functional domains of murine coronavirus nonstructural protein 3. J. Virol. 89, 6033–6047.

Imbert, I., Guillemot, J.-C., Bourhis, J.-M., Bussetta, C., Coutard, B., Egloff, M.-P., Ferron, F., Gorbalenya, A.E., Canard, B., 2006. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. EMBO J. 25, 4933–4942.

Imbert, I., Snijder, E.J., Dimitrova, M., Guillemot, J.-C., Lécine, P., Canard, B., 2008. The SARS-Coronavirus PLnc domain of nsp3 as a replication/transcription scaffolding protein. Virus Res. 133, 136–148.

Ivanov, K.A., Hertzig, T., Rozanov, M., Bayer, S., Thiel, V., Gorbalenya, A.E., Ziebuhr, J., 2004a. Major genetic marker of nidoviruses encodes a replicative endoribonuclease. Proc. Natl. Acad. Sci. U. S. A. 101, 12694–12699.

Ivanov, K.A., Thiel, V., Dobbe, J.C., van der Meer, Y., Snijder, E.J., Ziebuhr, J., 2004b. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. J. Virol. 78, 5619–5632.

Ivanov, K.A., Ziebuhr, J., 2004. Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5′-triphosphatase activities. J. Virol. 78, 7833–7838.

Jacobse-Geels, H.E.L., Horzinek, M.C., 1983. Expression of feline infectious peritonitis coronavirus antigens on the surface of feline macrophage-like cells. J. Gen. Virol. 64, 1859–1866.

Jayaram, H., Fan, H., Bowman, B.R., Ooi, A., Jayaram, J., Collisson, E.W., Lescar, J., Prasad, B.V., 2006. X-ray structures of the N-and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation. J. Virol. 80, 6612–6620.

JB, D., 1949. A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin. J. Exp. Med. 90, 181–210.

Jentsch, S., Pyrowolakis, G., 2000. Ubiquitin and its kin: how close are the family ties? Trends Cell Biol. 10, 335–342.

Jiang, L., Wang, N., Zuo, T., Shi, X., Poon, K.-M.V., Wu, Y., Gao, F., Li, D., Wang, R., Guo, J., others, 2014. Potent neutralization of MERS-CoV by human neutralizing monoclonal antibodies to the viral spike glycoprotein. Sci. Transl. Med. 6, 234ra59–234ra59.

Jin, Z.H., Xu, B., Lin, S.Z., Jin, Q.C., Cen, P.L., 2009. Enhanced production of spinosad in Saccharopolyspora spinosa by genome shuffling. Appl. Biochem. Biotechnol. 159, 655–663.

Johannes, T.W., Zhao, H., 2006. Directed evolution of enzymes and biosynthetic pathways. Curr. Opin. Microbiol. 9, 261–267.

Jürgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M., Sterner, R., 2000. Directed evolution of a (β α) 8-barrel enzyme to catalyze related reactions in two different metabolic pathways. Proc. Natl. Acad. Sci. 97, 9925–9930.

Kadonaga, J.T., Knowles, J.R., 1985. A simple and efficient method for chemical mutagenesis of DNA. Nucleic Acids Res. 13, 1733–1745. doi:10.1093/nar/13.5.1733

Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., Kubo, H., Makino, S., 2006. Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. Proc. Natl. Acad. Sci. 103, 12885–12890.

Kanjanahaluethai, A., Chen, Z., Jukneliene, D., Baker, S.C., 2007. Membrane topology of murine coronavirus replicase nonstructural protein 3. Virology 361, 391–401.

Kao, R.Y., Yang, D., Lau, L.-S., Tsui, W.H., Hu, L., Dai, J., Chan, M.-P., Chan, C.-M., Wang, P., Zheng, B.-J., others, 2010. Identification of influenza A nucleoprotein as an antiviral target. Nat. Biotechnol. 28, 600–605.

Kazi, L., Lissenberg, A., Watson, R., Groot, R.J. de, Weiss, S.R., 2005. Expression of Hemagglutinin Esterase Protein from Recombinant Mouse Hepatitis Virus Enhances Neurovirulence. J. Virol. 79, 15064–15073. doi:10.1128/JVI.79.24.15064-15073.2005

Keane, S.C., Liu, P., Leibowitz, J.L., Giedroc, D.P., 2012. Functional transcriptional regulatory sequence (TRS) RNA binding and helix destabilizing determinants of murine hepatitis virus (MHV) nucleocapsid (N) protein. J. Biol. Chem. 287, 7063–7073.

Kienzle, T.E., Abraham, S., Hogue, B.G., Brian, D.A., 1990. Structure and orientation of expressed bovine coronavirus hemagglutinin-esterase protein. J. Virol. 64, 1834–1838.

Kindrachuk, J., Ork, B., Hart, B.J., Mazur, S., Holbrook, M.R., Frieman, M.B., Traynor, D., Johnson, R.F., Dyall, J., Kuhn, J.H., others, 2015. Antiviral potential of ERK/MAPK and PI3K/AKT/mTOR signaling modulation for Middle East respiratory syndrome coronavirus infection as identified by temporal kinome analysis. Antimicrob. Agents Chemother. 59, 1088–1099.

King, A.M., Adams, M.J., Lefkowitz, E.J., Carstens, E.B., 2011. Virus taxonomy: IXth report of the International Committee on Taxonomy of Viruses. Access Online via Elsevier.

Komatsu, M., Waguri, S., Ueno, T., Iwata, J., Murata, S., Tanida, I., Ezaki, J., Mizushima, N., Ohsumi, Y., Uchiyama, Y., others, 2005. Impairment of starvation-induced and constitutive autophagy in Atg7-deficient mice. J Cell Biol 169, 425–434.

Kotzia, G.A., Labrou, N.E., 2009. Engineering thermal stability of l-asparaginase by in vitro directed evolution. Febs J. 276, 1750–1761.

Krijnse-Locker, J., Ericsson, M., Rottier, P.J., Griffiths, G., 1994. Characterization of the budding compartment of mouse hepatitis virus: evidence that transport from the RER to the Golgi complex requires only one vesicular transport step. J. Cell Biol. 124, 55–70.

Krishna, S.S., Majumdar, I., Grishin, N.V., 2003. Structural classification of zinc fingers SURVEY AND SUMMARY. Nucleic Acids Res. 31, 532–550.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.

Kumar, V., Jung, Y.-S., Liang, P.-H., 2013. Anti-SARS coronavirus agents: a patent review (2008–present). Expert Opin. Ther. Pat. 23, 1337–1348.

Kuo, C.-J., Liang, P.-H., 2015. Characterization and inhibition of the main protease of severe acute respiratory syndrome coronavirus. ChemBioEng Rev. 2, 118–132.

Kuo, L., Koetzner, C.A., Hurst, K.R., Masters, P.S., 2014. Recognition of the murine coronavirus genomic RNA packaging signal depends on the second RNA-binding domain of the nucleocapsid protein. J. Virol. 88, 4451–4465.

Kuo, L., Koetzner, C.A., Masters, P.S., 2016. A key role for the carboxy-terminal tail of the murine coronavirus nucleocapsid protein in coordination of genome packaging. Virology 494, 100–107.

Kuo, L., Masters, P.S., 2010. Evolved variants of the membrane protein can partially replace the envelope protein in murine coronavirus assembly. J. Virol. 84, 12872–12885.

Kuo, L., Masters, P.S., 2002. Genetic evidence for a structural interaction between the carboxy termini of the membrane and nucleocapsid proteins of mouse hepatitis virus. J. Virol. 76, 4987–4999.

La Monica, N., Yokomori, K., Lai, M., 1992. Coronavirus mRNA synthesis: identification of novel transcription initiation signals which are differentially regulated by different leader sequences. Virology 188, 402–407.

Lai, M.M., Anderson, L.J., 2007. Coronaviridae, Fifth Edition. ed.

Lai, M.M., Cavanagh, D., 1997. The molecular biology of coronaviruses. Adv. Virus Res. 48, 1–100.

Lauber, C., Goeman, J.J., Parquet, M. del C., Thi Nga, P., Snijder, E.J., Morita, K., Gorbalenya, A.E., 2013. The Footprint of Genome Architecture in the Largest Genome Expansion in RNA Viruses. PLoS Pathog 9, e1003500. doi:10.1371/journal.ppat.1003500

Lauber, C., Ziebuhr, J., Junglen, S., Drosten, C., Zirkel, F., Nga, P.T., Morita, K., Snijder, E.J., Gorbalenya, A.E., 2012. Mesoniviridae: a proposed new family in the order Nidovirales formed by a single species of mosquito-borne viruses. Arch. Virol. 157, 1623–1628.

Laviada, M.D., Videgain, S.P., Moreno, L., Alonso, F., Enjuanes, L., Escribano, J.M., 1990. Expression of swine transmissible gastroenteritis virus envelope antigens on the surface of infected cells: epitopes externally exposed. Virus Res. 16, 247–254.

Lee, D.H., Goldberg, A.L., 1998. Proteasome inhibitors: valuable new tools for cell biologists. Trends Cell Biol. 8, 397–403. doi:10.1016/S0962-8924(98)01346-4

Lee, N.-R., Kwon, H.-M., Park, K., Oh, S., Jeong, Y.-J., Kim, D.-E., 2010. Cooperative translocation enhances the unwinding of duplex DNA by SARS coronavirus helicase nsP13. Nucleic Acids Res. 38, 7626–7636.

Leung, D.T.M., Hang, T.F.C., Hung, M.C., Chan, P.K.S., Cheung, J.L.K., Niu, H., Tam, J.S.L., Lim, P.L., 2004. Antibody response of patients with severe acute respiratory syndrome (SARS) targets the viral nucleocapsid. J. Infect. Dis. 190, 379–386.

Leung, D.W., Chen, E., Goeddel, D.V., 1989. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique 1, 11–15.

Levitz, S.M., Diamond, R.D., 1985. A rapid colorimetric assay of fungal viability with the tetrazolium salt MTT. J. Infect. Dis. 152, 938–945.

Lewis, J.D., Izaurflde, E., 1997. The role of the cap structure in RNA processing and nuclear export. Eur. J. Biochem. 247, 461–469.

Li, 2013. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral Res. 100, 246–254.

Li, S.-J., Hochstrasser, M., 2000. The yeast ULP2 (SMT4) gene encodes a novel protease specific for the ubiquitin-like Smt3 protein. Mol. Cell. Biol. 20, 2367–2377.

Li, S.-J., Hochstrasser, M., 1999. A new protease required for cell-cycle progression in yeast. Nature 398, 246–251.

Li, Wong, S.-K., Li, F., Kuhn, J.H., Huang, I.-C., Choe, H., Farzan, M., 2006. Animal origins of the severe acute respiratory syndrome coronavirus: insight from ACE2-S-protein interactions. J. Virol. 80, 4211–4219.

Li, Y., Wan, Y., Liu, P., Zhao, J., Lu, G., Qi, J., Wang, Q., Lu, X., Wu, Y., Liu, W., others, 2015. A humanized neutralizing antibody against MERS-CoV targeting the receptor-binding domain of the spike protein. Cell Res. 25, 1237.

Lim, S.P., Sonntag, L.S., Noble, C., Nilar, S.H., Ng, R.H., Zou, G., Monaghan, P., Chung, K.Y., Dong, H., Liu, B., Bodenreider, C., Lee, G., Ding, M., Chan, W.L., Wang, G., Jian, Y.L., Chao, A.T., Lescar, J., Yin, Z., Vedananda,

T.R., Keller, T.H., Shi, P.-Y., 2011. Small Molecule Inhibitors That Selectively Block Dengue Virus Methyltransferase. J. Biol. Chem. 286, 6233–6240. doi:10.1074/jbc.M110.179184

Lin, S.-Y., Liu, C.-L., Chang, Y.-M., Zhao, J., Perlman, S., Hou, M.-H., 2014. Structural basis for the identification of the N-terminal domain of coronavirus nucleocapsid protein as an antiviral target. J. Med. Chem. 57, 2247–2257.

Lin-Goerke, J.L., Robbins, D.J., Burczak, J.D., 1997. PCR-based random mutagenesis using manganese and reduced dNTP concentration. Biotechniques 23, 409–412.

Liu, Q., Xia, S., Sun, Z., Wang, Q., Du, L., Lu, L., Jiang, S., 2015. Testing of Middle East respiratory syndrome coronavirus replication inhibitors for the ability to block viral entry. Antimicrob. Agents Chemother. 59, 742–744.

Lo, Y.-S., Lin, S.-Y., Wang, S.-M., Wang, C.-T., Chiu, Y.-L., Huang, T.-H., Hou, M.-H., 2013. Oligomerization of the carboxyl terminal domain of the human coronavirus 229E nucleocapsid protein. FEBS Lett. 587, 120–127. doi:10.1016/j.febslet.2012.11.016

Lu, G., Wang, Q., Gao, G.F., 2015. Bat-to-human: spike features determining “host jump”of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends Microbiol. 23, 468–478.

Lu, L., Liu, Q., Zhu, Y., Chan, K.-H., Qin, L., Li, Y., Wang, Q., Chan, J.F.-W., Du, L., Yu, F., others, 2014. Structure-based discovery of Middle East respiratory syndrome coronavirus fusion inhibitor. Nat. Commun. 5, 3067.

Lugari, A., Betzi, S., Decroly, E., Bonnaud, E., Hermant, A., Guillemot, J.-C., Debarnot, C., Borg, J.-P., Bouvet, M., Canard, B., others, 2010. Molecular mapping of the RNA Cap 2′ -O-methyltransferase activation interface between severe acute respiratory syndrome coronavirus nsp10 and nsp16. J. Biol. Chem. 285, 33230–33241.

Lui, P.-Y., Wong, L.-Y.R., Fung, C.-L., Siu, K.-L., Yeung, M.-L., Yuen, K.-S., Chan, C.-P., Woo, P.C.-Y., Yuen, K.-Y., Jin, D.-Y., 2016. Middle East respiratory syndrome coronavirus M protein suppresses type I interferon expression through the inhibition of TBK1-dependent phosphorylation of IRF3. Emerg. Microbes Infect. 5, e39. doi:10.1038/emi.2016.33

Lundberg, K.S., Shoemaker, D.D., Adams, M.W., Short, J.M., Sorge, J.A., Mathur, E.J., 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. Gene 108, 1–6.

Luo, C., Luo, H., Zheng, S., Gui, C., Yue, L., Yu, C., Sun, T., He, P., Chen, J., Shen, J., others, 2004. Nucleocapsid protein of SARS coronavirus tightly binds to human cyclophilin A. Biochem. Biophys. Res. Commun. 321, 557–565.

Luo, H., Chen, J., Chen, K., Shen, X., Jiang, H., 2006a. Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. Biochemistry (Mosc.) 45, 11827–11835.

Luo, H., Wu, D., Shen, C., Chen, K., Shen, X., Jiang, H., 2006b. Severe acute respiratory syndrome coronavirus membrane protein interacts with nucleocapsid protein mostly through their carboxyl termini by electrostatic attraction. Int. J. Biochem. Cell Biol. 38, 589–599.

Luo, H., Ye, F., Chen, K., Shen, X., Jiang, H., 2005. SR-rich motif plays a pivotal role in recombinant SARS coronavirus nucleocapsid protein multimerization. Biochemistry (Mosc.) 44, 15351–15358.

Luo, Z., Weiss, S.R., 1998. Roles in cell-to-cell fusion of two conserved hydrophobic regions in the murine coronavirus spike protein. Virology 244, 483–494.

Luo, Z.-H., Hua, Z.-C., 1998. Increased solubility of glutathione S-transferase-p16 (GST-p16) fusion protein by co-expression of chaperones groes and groel in Escherichia coli. IUBMB Life 46, 471–477.

Ma, Y., Tong, X., Xu, X., Li, X., Lou, Z., Rao, Z., 2010. Structures of the N-and C-terminal domains of MHV-A59 nucleocapsid protein corroborate a conserved RNA-protein binding mechanism in coronavirus. Protein Cell 1, 688–697.

Malakhov, M.P., Mattern, M.R., Malakhova, O.A., Drinker, M., Weeks, S.D., Butt, T.R., 2004. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. J. Struct. Funct. Genomics 5, 75–86.

Mannen, H., Tseng, H.-M., Cho, C., Li, S.S.-L., 1996. Cloning and expression of human homolog HSMT3 to yeast SMT3 suppressor of MIF2 mutations in a centromere protein gene. Biochem. Biophys. Res. Commun. 222, 178–180.

Marra, M.A., Jones, S.J.M., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S.N., Khattra, J., Asano, J.K., Barber, S.A., Chan, S.Y., Cloutier, A., Coughlin, S.M., Freeman, D., Girn, N., Griffith, O.L., Leach, S.R., Mayo, M., McDonald, H., Montgomery, S.B., Pandoh, P.K., Petrescu, A.S., Robertson, A.G., Schein, J.E., Siddiqui, A., Smailus, D.E., Stott, J.M., Yang, G.S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T.F., Bowness, D., Czub, M., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G.A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R.C., Krajden, M., Petric, M., Skowronski, D.M., Upton, C., Roper, R.L., 2003. The Genome Sequence of the SARS-Associated Coronavirus. Science 300, 1399–1404. doi:10.1126/science.1085953

Marsischky, G., LaBaer, J., 2004. Many paths to many clones: a comparative look at high-throughput cloning methods. Genome Res. 14, 2020–2028.

Mason, A.B., He, Q.-Y., Adams, T.E., Gumerov, D.R., Kaltashov, I.A., Nguyen, V., MacGillivray, R.T., 2001. Expression, purification, and characterization of recombinant nonglycosylated human serum transferrin containing a C-terminal hexahistidine tag. Protein Expr. Purif. 23, 142–150.

Masters, P.S., 2006. The molecular biology of coronaviruses. Adv. Virus Res. 66, 193–292.

Masters, P.S., Parker, M.M., Ricard, C.S., Duchala, C., Frana, M.F., Holmes, K.V., Sturman, L.S., 1990. Structure and function studies of the nucleocapsid protein of mouse hepatitis virus, in: Coronaviruses and Their Diseases. Springer, pp. 239–246.

Masters, P.S., Sturman, L.S., 1990. Background Paper Functions of the Coronavirus Nucleocapsid Protein, in: Coronaviruses and Their Diseases. Springer, pp. 235–238.

May, O., Nguyen, P.T., Arnold, F.H., 2000. Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine. Nat. Biotechnol. 18, 317–320.

McBride, C.E., Machamer, C.E., 2010. A single tyrosine in the severe acute respiratory syndrome coronavirus membrane protein cytoplasmic tail is important for efficient interaction with spike protein. J. Virol. 84, 1891–1901.

McBride, R., Fielding, B.C., 2012. The Role of Severe Acute Respiratory Syndrome (SARS)-Coronavirus Accessory Proteins in Virus Pathogenesis. Viruses 4, 2902–2923. doi:10.3390/v4112902

McIsaac, R.S., Engqvist, M.K., Wannier, T., Rosenthal, A.Z., Herwig, L., Flytzanis, N.C., Imasheva, E.S., Lanyi, J.K., Balashov, S.P., Gradinaru, V., others, 2014. Directed evolution of a far-red fluorescent rhodopsin. Proc. Natl. Acad. Sci. 111, 13034–13039.

Melchior, F., 2000. SUMO-nonclassical ubiquitin. Annu. Rev. Cell Dev. Biol. 16, 591–626.

Meluh, P.B., Koshland, D., 1995. Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein CENP-C. Mol. Biol. Cell 6, 793–807.

Menachery, V.D., Debbink, K., Baric, R.S., 2014. Coronavirus non-structural protein 16: Evasion, attenuation, and possible treatments. Virus Res. 194, 191–199.

Mibayashi, M., Martínez-Sobrido, L., Loo, Y.-M., Cárdenas, W.B., Gale, M., García-Sastre, A., 2007. Inhibition of Retinoic Acid-Inducible Gene I-Mediated Induction of Beta Interferon by the NS1 Protein of Influenza A Virus. J. Virol. 81, 514–524. doi:10.1128/JVI.01265-06

Mielech, A.M., Deng, X., Chen, Y., Kindler, E., Wheeler, D.L., Mesecar, A.D., Thiel, V., Perlman, S., Baker, S.C., 2015. Murine Coronavirus Ubiquitin-Like Domain Is Important for Papain-Like Protease Stability and Viral Pathogenesis. J. Virol. 89, 4907–4917. doi:10.1128/JVI.00338-15

Mielech, A.M., Kilianski, A., Baez-Santos, Y.M., Mesecar, A.D., Baker, S.C., 2014. MERS-CoV papain-like protease has deISGylating and deubiquitinating activities. Virology 450, 64–70.

Miller, W.A., Koev, G., 2000. Synthesis of subgenomic RNAs by positive-strand RNA viruses. Virology 273, 1–8.

Minskaia, E., Hertzig, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B., Ziebuhr, J., 2006. Discovery of an RNA virus 3′ → 5′

exoribonuclease that is critically involved in coronavirus RNA synthesis. Proc. Natl. Acad. Sci. U. S. A. 103, 5108–5113.

Mizushima, N., Yamamoto, A., Hatano, M., Kobayashi, Y., Kabeya, Y., Suzuki, K., Tokuhisa, T., Ohsumi, Y., Yoshimori, T., 2001. Dissection of autophagosome formation using Apg5-deficient mouse embryonic stem cells. J. Cell Biol. 152, 657–668.

Mizushima, N., Yamamoto, A., Matsui, M., Yoshimori, T., Ohsumi, Y., 2004. In vivo analysis of autophagy in response to nutrient starvation using transgenic mice expressing a fluorescent autophagosome marker. Mol. Biol. Cell 15, 1101–1111.

Mosmann, T., 1983. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. J. Immunol. Methods 65, 55–63.

Müller, S., Hoege, C., Pyrowolakis, G., Jentsch, S., 2001. SUMO, ubiquitin's mysterious cousin. Nat. Rev. Mol. Cell Biol. 2, 202–213.

Musah, R.A., 2004. The HIV-1 nucleocapsid zinc finger protein as a target of antiretroviral therapy. Curr. Top. Med. Chem. 4, 1605–1622.

Nal, B., Chan, C., Kien, F., Siu, L., Tse, J., Chu, K., Kam, J., Staropoli, I., Crescenzo-Chaigne, B., Escriou, N., van der Werf, S., Yuen, K.-Y., Altmeyer, R., 2005. Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. J. Gen. Virol. 86, 1423–1434. doi:10.1099/vir.0.80671-0

Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J.C., Brierley, I., 2006. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. Nature 441, 244–247. doi:10.1038/nature04735

Narayanan, K., Huang, C., Lokugamage, K., Kamitani, W., Ikegami, T., Tseng, C.-T.K., Makino, S., 2008. Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells. J. Virol. 82, 4471–4479.

Nauwynck, H., Snijder, E.J., Faaberg, Balasuriya, U., Brinton, M., Gorbalenya, A.E., Leung, F.-C., Stadejek, T., Yang, H., Yoo, D., 2012. Family Arteriviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds)Virus Taxonomy:IXth report of the International Committe of Taxonomy.

Neuman, B.W., 2016. Bioinformatics and functional analyses of coronavirus nonstructural proteins involved in the formation of replicative organelles. Antiviral Res. 135, 97–107. doi:10.1016/j.antiviral.2016.10.005

Neuman, B.W., 2008. Supramolecular Architecture of the Coronavirus Particle. Nidoviruses 201.

Neuman, B.W., Adair, B.D., Yoshioka, C., Quispe, J.D., Orca, G., Kuhn, P., Milligan, R.A., Yeager, M., Buchmeier, M.J., 2006. Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. J. Virol. 80, 7918–7928.

Neuman, B.W., Chamberlain, P., Bowden, F., Joseph, J., 2014. Atlas of coronavirus replicase structure. Virus Res., Nidoviruses I 194, 49–66. doi:10.1016/j.virusres.2013.12.004

Neuman, B.W., Joseph, J.S., Saikatendu, K.S., Serrano, P., Chatterjee, A., Johnson, M.A., Liao, L., Klaus, J.P., Yates, J.R., Wüthrich, K., 2008. Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. J. Virol. 82, 5279–5294.

Neuman, B.W., Kiss, G., Kunding, A.H., Bhella, D., Baksh, M.F., Connelly, S., Droese, B., Klaus, J.P., Makino, S., Sawicki, S.G., others, 2011. A structural analysis of M protein in coronavirus assembly and morphology. J. Struct. Biol. 174, 11–22.

Nga, P.T., del Carmen Parquet, M., Lauber, C., Parida, M., Nabeshima, T., Yu, F., Thuy, N.T., Inoue, S., Ito, T., Okamoto, K., 2011. Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. PLoS Pathog. 7, e1002215.

Nieto-Torres, J.L., DeDiego, M.L., Álvarez, E., Jiménez-Guardeño, J.M., Regla-Nava, J.A., Llorente, M., Kremer, L., Shuo, S., Enjuanes, L., 2011. Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. Virology 415, 69–82.

Nilsson, J., St\a ahl, S., Lundeberg, J., Uhlén, M., Nygren, P.-\a ake, 1997. Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. Protein Expr. Purif. 11, 1–16.

Nishiya, Y., Imanaka, T., 1994. Alteration of substrate specificity and optimum pH of sarcosine oxidase by random and site-directed mutagenesis. Appl. Environ. Microbiol. 60, 4213–4215.

O'Flynn, N., 2011. Increasing Baculovirus Transduction of Mammalian Cells for Drug Discovery. University of Reading.

Oba, M., Tanaka, M., 2012. Intracellular internalization mechanism of protein transfection reagents. Biol. Pharm. Bull. 35, 1064–1068.

Oostra, M., Hagemeijer, M.C., van Gent, M., Bekker, C.P., te Lintelo, E.G., Rottier, P.J., de Haan, C.A., 2008. Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. J. Virol. 82, 12392–12405.

Orlowski, R.Z., 1999. The role of the ubiquitin-proteasome pathway in apoptosis. Cell Death Differ. 6.

Otten, L.G., Quax, W.J., 2005. Directed evolution: selecting today's biocatalysts. Biomol. Eng. 22, 1–9.

Pan, J., Peng, X., Gao, Y., Li, Z., Lu, X., Chen, Y., Ishaq, M., Liu, D., DeDiego, M.L., Enjuanes, L., Guo, D., 2008. Genome-Wide Analysis of Protein-Protein Interactions and Involvement of Viral Proteins in SARS-CoV Replication. PLoS ONE 3, e3299. doi:10.1371/journal.pone.0003299

Pascal, K.E., Coleman, C.M., Mujica, A.O., Kamat, V., Badithe, A., Fairhurst, J., Hunt, C., Strein, J., Berrebi, A., Sisk, J.M., others, 2015. Pre-and postexposure efficacy of fully human antibodies against Spike protein in a novel humanized mouse model of MERS-CoV infection. Proc. Natl. Acad. Sci. 112, 8738–8743.

Patrick, W.M., Firth, A.E., 2005. Strategies and computational tools for improving randomized protein libraries. Biomol. Eng. 22, 105–112.

Patten, P.A., Howard, R.J., Stemmer, W.P., 1997. Applications of DNA shuffling to pharmaceuticals and vaccines. Curr. Opin. Biotechnol. 8, 724–733.

Peck, R., 1985. A one-plate assay for macrophage bactericidal activity. J. Immunol. Methods 82, 131–140.

Perlman, S., Netland, J., 2009. Coronaviruses post-SARS: update on replication and pathogenesis. Nat. Rev. Microbiol. 7, 439–450.

Peroutka, R.J., Elshourbagy, N., Piech, T., Butt, T.R., 2008. Enhanced protein expression in mammalian cells using engineered SUMO fusions: secreted phospholipase A2. Protein Sci. 17, 1586–1595.

Pfefferle, S., Schöpf, J., Kögl, M., Friedel, C.C., Müller, M.A., Carbajo-Lozoya, J., Stellberger, T., von Dall'Armi, E., Herzog, P., Kallies, S., others, 2011. The

SARS-coronavirus-host interactome: identification of cyclophilins as target for pan-coronavirus inhibitors. PLoS Pathog. 7, e1002331.

Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y., Jung, S.-H., 2016. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. J. Med. Chem. 59, 6595–6628.

Poon, L.L.M., Chu, D.K.W., Chan, K.H., Wong, O.K., Ellis, T.M., Leung, Y.H.C., Lau, S.K.P., Woo, P.C.Y., Suen, K.Y., Yuen, K.Y., 2005. Identification of a novel coronavirus in bats. J. Virol. 79, 2001–2009.

Popova, R., Zhang, X., 2002. The Spike but Not the Hemagglutinin/Esterase Protein of Bovine Coronavirus Is Necessary and Sufficient for Viral Infection. Virology 294, 222–236. doi:10.1006/viro.2001.1307

Prentice, E., McAuliffe, J., Lu, X., Subbarao, K., Denison, M.R., 2004. Identification and characterization of severe acute respiratory syndrome coronavirus replicase proteins. J. Virol. 78, 9977–9986.

Price, P., McMillan, T.J., 1990. Use of the tetrazolium assay in measuring the response of human tumor cells to ionizing radiation. Cancer Res. 50, 1392–1396.

Pryor, K.D., Leiting, B., 1997. High-Level Expression of Soluble Protein inEscherichia coliUsing a His 6-Tag and Maltose-Binding-Protein Double-Affinity Fusion System. Protein Expr. Purif. 10, 309–319.

Ratia, K., Pegan, S., Takayama, J., Sleeman, K., Coughlin, M., Baliji, S., Chaudhuri, R., Fu, W., Prabhakar, B.S., Johnson, M.E., others, 2008. A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. Proc. Natl. Acad. Sci. 105, 16119–16124.

Ricagno, S., Coutard, B., Grisel, S., Brémond, N., Dalle, K., Tocque, F., Campanacci, V., Lichière, J., Lantez, V., Debarnot, C., 2006. Crystallization and preliminary X-ray diffraction analysis of Nsp15 from SARS coronavirus. Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. 62, 409–411.

Romero, P.A., Arnold, F.H., 2009. Exploring protein fitness landscapes by directed evolution. Nat. Rev. Mol. Cell Biol. 10, 866–876. doi:10.1038/nrm2805

Rossen, J.W., Bekker, C.P., Voorhout, W.F., Strous, G.J., Ende, A. van der, Rottier, P.J., 1994. Entry and release of transmissible gastroenteritis

coronavirus are restricted to apical surfaces of polarized epithelial cells. J. Virol. 68, 7966–7973.

Rossen, J.W.A., Bekker, C.P.J., Strous, G., Horzinek, M.C., Dveksler, G.S., Holmes, K.V., Rottier, P.J.M., 1996. A murine and a porcine coronavirus are released from opposite surfaces of the same epithelial cells. Virology 224, 345–351.

Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Peñaranda, S., Bankamp, B., Maher, K., Chen, M., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J.L., Chen, Q., Wang, D., Erdman, D.D., Peret, T.C.T., Burns, C., Ksiazek, T.G., Rollin, P.E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Günther, S., Osterhaus, A.D.M.E., Drosten, C., Pallansch, M.A., Anderson, L.J., Bellini, W.J., 2003. Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. Science 300, 1394–1399. doi:10.1126/science.1085952

Roth-Cross, J.K., Martínez-Sobrido, L., Scott, E.P., García-Sastre, A., Weiss, S.R., 2007. Inhibition of the alpha/beta interferon response by mouse hepatitis virus at multiple levels. J. Virol. 81, 7189–7199.

Ruch, T.R., Machamer, C.E., 2012. The coronavirus e protein: assembly and beyond. Viruses 4, 363–382.

Saikatendu, K.S., Joseph, J.S., Subramanian, V., Clayton, T., Griffith, M., Moy, K., Velasquez, J., Neuman, B.W., Buchmeier, M.J., Stevens, R.C., Kuhn, P., 2005. Structural Basis of Severe Acute Respiratory Syndrome Coronavirus ADP-Ribose-1″ -Phosphate Dephosphorylation by a Conserved Domain of nsP3. Structure 13, 1665–1675. doi:10.1016/j.str.2005.07.022

Saikatendu, K.S., Joseph, J.S., Subramanian, V., Neuman, B.W., Buchmeier, M.J., Stevens, R.C., Kuhn, P., 2007. Ribonucleocapsid formation of severe acute respiratory syndrome coronavirus through molecular action of the N-terminal domain of N protein. J. Virol. 81, 3913–3921.

Sainz, B., Mossel, E.C., Gallaher, W.R., Wimley, W.C., Peters, C.J., Wilson, R.B., Garry, R.F., 2006. Inhibition of severe acute respiratory syndrome-associated coronavirus (SARS-CoV) infectivity by peptides analogous to the viral spike protein. Virus Res. 120, 146–155.

Saitoh, H., Hinchey, J., 2000. Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. J. Biol. Chem. 275, 6252–6258.

Salvati, A., Ciani, L., Ristori, S., Martini, G., Masi, A., Arcangeli, A., 2006. Physico-chemical characterization and transfection efficacy of cationic liposomes containing the pEGFP plasmid. Biophys. Chem. 121, 21–29.

Sawicki, S.G., Sawicki, D.L., Siddell, S.G., 2007. A contemporary view of coronavirus transcription. J. Virol. 81, 20–29.

Sawicki, S.G., Sawicki, D.L., Younker, D., Meyer, Y., Thiel, V., Stokes, H., Siddell, S.G., 2005. Functional and genetic analysis of coronavirus replicase-transcriptase proteins. PLoS Pathog. 1, e39.

Schalk, A.F., Hawn, M.C., 1931. An apparently new respiratory disease of baby chicks. J Am Vet Med Assoc 78, 19.

Schelle, B., Karl, N., Ludewig, B., Siddell, S.G., Thiel, V., 2005. Selective replication of coronavirus genomes that express nucleocapsid protein. J. Virol. 79, 6620–6630.

Schwer, B., Mao, X., Shuman, S., 1998. Accelerated mRNA decay in conditional mutants of yeast mRNA capping enzyme. Nucleic Acids Res. 26, 2050–2057.

Shafikhani, S., Siegel, R.A., Ferrari, E., Schellenberger, V., 1997. Generation of large libraries of random mutants in Bacillus subtilis by PCR-based plasmid multimerization. Biotechniques 23, 304–311.

Shahwan, K., Hesse, M., Mork, A.-K., Herrler, G., Winter, C., 2013. Sialic acid binding properties of soluble coronavirus spike (S1) proteins: differences between infectious bronchitis virus and transmissible gastroenteritis virus. Viruses 5, 1924–1933.

Shan, T., Li, L., Simmonds, P., Wang, C., Moeser, A., Delwart, E., 2011. The Fecal Virome of Pigs on a High-Density Farm. J. Virol. 85, 11697–11708. doi:10.1128/JVI.05217-11

Shortle, D., Nathans, D., 1978. Local mutagenesis: a method for generating viral mutants with base substitutions in preselected regions of the viral genome. Proc. Natl. Acad. Sci. 75, 2170–2174.

Shuman, S., 2002. What messenger RNA capping tells us about eukaryotic evolution. Nat. Rev. Mol. Cell Biol. 3, 619–625.

Shuman, S., 2000. Structure, mechanism, and evolution of the mRNA capping apparatus. Prog. Nucleic Acid Res. Mol. Biol. 66, 1–40.

Siu, K.-L., Chan, C.-P., Kok, K.-H., Chiu-Yat Woo, P., Jin, D.-Y., 2014. Suppression of innate antiviral response by severe acute respiratory syndrome coronavirus M protein is mediated through the first transmembrane domain. Cell. Mol. Immunol. 11, 141–149. doi:10.1038/cmi.2013.61

Siu, K.-L., Kok, K.-H., Ng, M.-H.J., Poon, V.K., Yuen, K.-Y., Zheng, B.-J., Jin, D.-Y., 2009. Severe acute respiratory syndrome coronavirus M protein inhibits type I interferon production by impeding the formation of TRAF3· TANK· TBK1/IKKε complex. J. Biol. Chem. 284, 16202–16209.

Skandalis, A., Encell, L.P., Loeb, L.A., 1997. Creating novel enzymes by applied molecular evolution. Chem. Biol. 4, 889–898. doi:10.1016/S1074-5521(97)90297-0

Skehel, J.J., Wiley, D.C., 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu. Rev. Biochem. 69, 531–569.

Skilton, A., Ho, J.C., Mercer, B., Outwin, E., Watts, F.Z., 2009. SUMO chain formation is required for response to replication arrest in S. pombe. PLoS One 4, e6750.

Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., Gorbalenya, A.E., 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. 331, 991–1004.

Snijder, E.J., Van Der Meer, Y., Zevenhoven-Dobbe, J., Onderwater, J.J., van der Meulen, J., Koerten, H.K., Mommaas, A.M., 2006. Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex. J. Virol. 80, 5927–5940.

Spaan, W., Cavanagh, D., Horzinek, M.C., 1988. Coronaviruses: structure and genome expression. J. Gen. Virol. 69, 2939–2952.

Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H.-D., Rappuoli, R., 2003. SARS |[mdash]| beginning to understand a new virus. Nat. Rev. Microbiol. 1, 209–218. doi:10.1038/nrmicro775

Stahl, G., McCarty, G.P., Farabaugh, P.J., 2002. Ribosome structure: revisiting the connection between translational accuracy and unconventional decoding. Trends Biochem. Sci. 27, 178–183.

Stahla-Beek, H.J., April, D.G., Saeedi, B.J., Hannah, A.M., Keenan, S.M., Geiss, B.J., 2012. Identification of a Novel Antiviral Inhibitor of the Flavivirus Guanylyltransferase Enzyme. J. Virol. 86, 8730–8739. doi:10.1128/JVI.00384-12

Stohlman, S.A., Lai, M.M., 1979. Phosphoproteins of murine hepatitis viruses. J. Virol. 32, 672–675.

Studier, F.W., 2005. Protein production by auto-induction in high-density shaking cultures. Protein Expr. Purif. 41, 207–234.

Su, D., Lou, Z., Sun, F., Zhai, Y., Yang, H., Zhang, R., Joachimiak, A., Zhang, X.C., Bartlam, M., Rao, Z., 2006. Dodecamer structure of severe acute respiratory syndrome coronavirus nonstructural protein nsp10. J. Virol. 80, 7902–7908.

Subissi, L., Imbert, I., Ferron, F., Collet, A., Coutard, B., Decroly, E., Canard, B., 2014. SARS-CoV ORF1b-encoded nonstructural proteins 12–16: Replicative enzymes as antiviral targets. Antiviral Res. 101, 122–130.

Sui, J., Li, W., Murakami, A., Tamin, A., Matthews, L.J., Wong, S.K., Moore, M.J., Tallarico, A.S.C., Olurinde, M., Choe, H., others, 2004. Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. Proc. Natl. Acad. Sci. U. S. A. 101, 2536–2541.

Sun, Y., Wang, Z., Tao, J., Wang, Y., Wu, A., Yang, Z., Wang, K., Shi, L., Chen, Y., Guo, D., 2014. Yeast-based assays for the high-throughput screening of inhibitors of coronavirus RNA cap guanine-N7-methyltransferase. Antiviral Res. 104, 156–164.

Surjit, M., Lal, S.K., 2008. The SARS-CoV nucleocapsid protein: a protein with multifarious activities. Infect. Genet. Evol. 8, 397–405.

Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., Berrow, N., Owens, R., Gilbert, R., Davidson, A., 2004. The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. Structure 12, 341–353.

Szretter, K.J., Daniels, B.P., Cho, H., Gainey, M.D., Yokoyama, W.M., Gale Jr, M., Virgin, H.W., Klein, R.S., Sen, G.C., Diamond, M.S., 2012. 2′ -O methylation of the viral mRNA cap by West Nile virus evades ifit1-dependent and-independent mechanisms of host restriction in vivo. PLoS Pathog 8, e1002698.

Takeda, M., Chang, C., Ikeya, T., Güntert, P., Chang, Y., Hsu, Y., Huang, T., Kainosho, M., 2008. Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. J. Mol. Biol. 380, 608–622.

Tan, K., Zelus, B.D., Meijers, R., Liu, J., Bergelson, J.M., Duke, N., Zhang, R., Joachimiak, A., Holmes, K.V., Wang, J., 2002. Crystal structure of murine sCEACAM1a [1, 4]: a coronavirus receptor in the CEA family. EMBO J. 21, 2076–2086.

Tan, Y.W., Fang, S., Fan, H., Lescar, J., Liu, D.X., 2006. Amino acid residues critical for RNA-binding in the N-terminal domain of the nucleocapsid protein are essential determinants for the infectivity of coronavirus in cultured cells. Nucleic Acids Res. 34, 4816–4825.

Tanaka, Y., Sato, Y., Sasaki, T., 2013. Suppression of coronavirus replication by cyclophilin inhibitors. Viruses 5, 1250–1260.

Tang, T.-K., Wu, M.P.-J., Chen, S.-T., Hou, M.-H., Hong, M.-H., Pan, F.-M., Yu, H.-M., Chen, J.-H., Yao, C.-W., Wang, A.H.-J., 2005. Biochemical and immunological studies of nucleocapsid proteins of severe acute respiratory syndrome and 229E human coronaviruses. Proteomics 5, 925–937.

Tanner, J.A., Watt, R.M., Chai, Y.-B., Lu, L.-Y., Lin, M.C., Peiris, J.M., Poon, L.L., Kung, H.-F., Huang, J.-D., 2003. The severe acute respiratory syndrome (SARS) coronavirus NTPase/helicase belongs to a distinct class of 5′ to 3′ viral helicases. J. Biol. Chem. 278, 39578–39582.

Te Velthuis, A.J., Arnold, J.J., Cameron, C.E., van den Worm, S.H., Snijder, E.J., 2010. The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. Nucleic Acids Res. 38, 203–214.

Te Velthuis, A.J., van den Worm, S.H., Snijder, E.J., 2012. The SARS-coronavirus nsp7+ nsp8 complex is a unique multimeric RNA polymerase capable of

both de novo initiation and primer extension. Nucleic Acids Res. 40, 1737–1747.

Thiel, V., Herold, J., Schelle, B., Siddell, S.G., 2001. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. J. Gen. Virol. 82, 1273–1281.

Thiel, V., Ivanov, K.A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Wei\s sbrich, B., Snijder, E.J., Rabenau, H., Doerr, H.W., others, 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. J. Gen. Virol. 84, 2305–2315.

Tindall, K.R., Kunkel, T.A., 1988. Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase. Biochemistry (Mosc.) 27, 6008–6013.

To, L.T., Bernard, S., Lantier, I., 1991. Fixed-cell immunoperoxidase technique for the study of surface antigens induced by the coronavirus of transmissible gastroenteritis (TGEV). Vet. Microbiol. 29, 361–368.

Tong, S., Conrardy, C., Ruone, S., Kuzmin, I.V., Guo, X., Tao, Y., Niezgoda, M., Haynes, L., Agwanda, B., Breiman, R.F., 2009. Detection of novel SARS-like and other coronaviruses in bats from Kenya. Emerg. Infect. Dis. 15, 482.

Tong, T.R., 2009. Therapies for coronaviruses. Part 2: Inhibitors of intracellular life cycle. Expert Opin. Ther. Pat. 19, 415–431.

Tooze, J., Tooze, S., Warren, G., 1984. Replication of coronavirus MHV-A59 in sac-cells: determination of the first site of budding of progeny virions. Eur. J. Cell Biol. 33, 281–293.

Tooze, J., Tooze, S.A., 1985. Infection of AtT20 murine pituitary tumour cells by mouse hepatitis virus strain A59: virus budding is restricted to the Golgi region. Eur. J. Cell Biol. 37, 203–212.

Tseng, C.-T.K., Tseng, J., Perrone, L., Worthy, M., Popov, V., Peters, C.J., 2005. Apical Entry and Release of Severe Acute Respiratory Syndrome-Associated Coronavirus in Polarized Calu-3 Lung Epithelial Cells. J. Virol. 79, 9470–9479. doi:10.1128/JVI.79.15.9470-9479.2005

Tseng, Y.-T., Wang, S.-M., Huang, K.-J., Amber, I., Lee, R., Chiang, C.-C., Wang, C.-T., 2010. Self-assembly of severe acute respiratory syndrome coronavirus membrane protein. J. Biol. Chem. 285, 12862–12872.

Tylor, S., Andonov, A., Cutts, T., Cao, J., Grudesky, E., Van Domselaar, G., Li, X., He, R., 2009. The SR-rich motif in SARS-CoV nucleocapsid protein is important for virus replication. Can. J. Microbiol. 55, 254–260.

Tyrell, D.A., Almeida, J.D., Berry, D.M., Cunningham, C.H., Hamre, D., Hofstad, M.S., Malluci, L., Malntosh, K., 1968. Coronaviruses. Nat. Lond. 220, 650.

Van Boheemen, S., de Graaf, M., Lauber, C., Bestebroer, T.M., Raj, V.S., Zaki, A.M., Osterhaus, A.D., Haagmans, B.L., Gorbalenya, A.E., Snijder, E.J., others, 2012. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. MBio 3, e00473–12.

Van der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers, K.C., Wertheim-van Dillen, P.M., Kaandorp, J., Spaargaren, J., Berkhout, B., 2004. Identification of a new human coronavirus. Nat. Med. 10, 368–373.

Vartanian, J.-P., Henry, M., Wain-Hobson, S., 1996. Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. Nucleic Acids Res. 24, 2627–2631.

Venkatagopalan, P., Daskalova, S.M., Lopez, L.A., Dolezal, K.A., Hogue, B.G., 2015. Coronavirus envelope (E) protein remains at the site of assembly. Virology 478, 75–85.

Vennema, H., Godeke, G.J., Rossen, J.W., Voorhout, W.F., Horzinek, M.C., Opstelten, D.J., Rottier, P.J., 1996. Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. EMBO J. 15, 2020.

Verheije, M.H., Hagemeijer, M.C., Ulasli, M., Reggiori, F., Rottier, P.J., Masters, P.S., de Haan, C.A., 2010. The coronavirus nucleocapsid protein is dynamically associated with the replication-transcription complexes. J. Virol. 84, 11575–11579.

Verma, S., Bednar, V., Blount, A., Hogue, B.G., 2006. Identification of functionally important negatively charged residues in the carboxy end of mouse hepatitis coronavirus A59 nucleocapsid protein. J. Virol. 80, 4344–4355.

Verma, S., Lopez, L.A., Bednar, V., Hogue, B.G., 2007. Importance of the penultimate positive charge in mouse hepatitis coronavirus A59 membrane protein. J. Virol. 81, 5339–5348.

Vlasak, R., Luytjes, W., Spaan, W., Palese, P., 1988. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. Proc. Natl. Acad. Sci. 85, 4526–4529.

Von Brunn, A., Teepe, C., Simpson, J.C., Pepperkok, R., Friedel, C.C., Zimmer, R., Roberts, R., Baric, R., Haas, J., 2007. Analysis of Intraviral Protein-Protein Interactions of the SARS Coronavirus ORFeome. PLoS ONE 2, e459. doi:10.1371/journal.pone.0000459

Voß, D., Kern, A., Traggiai, E., Eickmann, M., Stadler, K., Lanzavecchia, A., Becker, S., 2006. Characterization of severe acute respiratory syndrome coronavirus membrane protein. FEBS Lett. 580, 968–973.

Wang, D., Zhao, C., Cheng, R., Sun, F., 2000. Estimation of the mutation rate during error-prone polymerase chain reaction. J. Comput. Biol. 7, 143–158.

Wang, N., Shi, X., Jiang, L., Zhang, S., Wang, D., Tong, P., Guo, D., Fu, L., Cui, Y., Liu, X., Arledge, K.C., Chen, Y.-H., Zhang, L., Wang, X., 2013. Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. Cell Res. 23, 986–993. doi:10.1038/cr.2013.92

Wang, Y., Sun, Y., Wu, A., Xu, S., Pan, R., Zeng, C., Jin, X., Ge, X., Shi, Z., Ahola, T., Chen, Y., Guo, D., 2015. Coronavirus nsp10/nsp16 Methyltransferase Can Be Targeted by nsp10-Derived Peptide In Vitro and In Vivo To Reduce Replication and Pathogenesis. J. Virol. 89, 8416–8427. doi:10.1128/JVI.00948-15

Warrens, A.N., Jones, M.D., Lechler, R.I., 1997. Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. Gene 186, 29–35. doi:10.1016/S0378-1119(96)00674-9

Wathelet, M.G., Orr, M., Frieman, M.B., Baric, R.S., 2007. Severe acute respiratory syndrome coronavirus evades antiviral signaling: role of nsp1 and rational design of an attenuated strain. J. Virol. 81, 11620–11633.

Weiss, S.R., Leibowitz, J.L., 2011. Chapter 4 - Coronavirus Pathogenesis, in: Maramorosch, K., Shatkin, A.J., Murphy, F.A. (Eds.), Advances in Virus Research. Academic Press, pp. 85–164.

Weiss, S.R., Navas-Martin, S., 2005. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiol. Mol. Biol. Rev. 69, 635–664.

Weng, L., Weersink, A., Poljak, Z., de Lange, K., von Massow, M., 2016. An economic evaluation of intervention strategies for Porcine Epidemic Diarrhea (PED). Prev. Vet. Med. 134, 58–68.

Werner, M., Purta, E., Kaminska, K.H., Cymerman, I.A., Campbell, D.A., Mittra, B., Zamudio, J.R., Sturm, N.R., Jaworski, J., Bujnicki, J.M., 2011. 2′ -O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family. Nucleic Acids Res. 39, 4756–4768.

Wilkinson, K.A., Henley, J.M., 2010. Mechanisms, regulation and consequences of protein SUMOylation. Biochem. J. 428, 133–145.

Wilson, I.A., Skehel, J.J., Wiley, D.C., 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 \AA resolution. Nature 289, 366–373.

Woo, P.C., HUANG, Y., LAU, S.K., TSOI, H.-W., YUEN, K.-Y., 2005a. In Silico analysis of ORF1ab in coronavirus HKU1 genome reveals a unique putative cleavage site of coronavirus HKU1 3C-like protease. Microbiol. Immunol. 49, 899–908.

Woo, P.C., Lau, S.K., Chu, C., Chan, K., Tsoi, H., Huang, Y., Wong, B.H., Poon, R.W., Cai, J.J., Luk, W., 2005b. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. 79, 884–895.

Woo, P.C., Lau, S.K., Lam, C.S., Lai, K.K., Huang, Y., Lee, P., Luk, G.S., Dyrting, K.C., Chan, K.-H., Yuen, K.-Y., 2009. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. J. Virol. 83, 908–917.

Woo, P.C., Lau, S.K., Lam, C.S., Lau, C.C., Tsang, A.K., Lau, J.H., Bai, R., Teng, J.L., Tsang, C.C., Wang, M., 2012. Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J. Virol. 86, 3995–4008.

Woo, P.C., Wang, M., Lau, S.K., Xu, H., Poon, R.W., Guo, R., Wong, B.H., Gao, K., Tsoi, H., Huang, Y., 2007. Comparative analysis of twelve genomes of

three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. J. Virol. 81, 1574–1585.

Wootton, S.K., Rowland, R.R., Yoo, D., 2002. Phosphorylation of the porcine reproductive and respiratory syndrome virus nucleocapsid protein. J. Virol. 76, 10569–10576.

Wu, C.-Y., Jan, J.-T., Ma, S.-H., Kuo, C.-J., Juan, H.-F., Cheng, Y.-S.E., Hsu, H.-H., Huang, H.-C., Wu, D., Brik, A., others, 2004. Small molecules targeting severe acute respiratory syndrome human coronavirus. Proc. Natl. Acad. Sci. U. S. A. 101, 10012–10017.

Wu, K., Li, W., Peng, G., Li, F., 2009. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. Proc. Natl. Acad. Sci. 106, 19970–19974. doi:10.1073/pnas.0908837106

Xiao, Y., Ma, Q., Restle, T., Shang, W., Svergun, D.I., Ponnusamy, R., Sczakiel, G., Hilgenfeld, R., 2012. Nonstructural proteins 7 and 8 of feline coronavirus form a 2: 1 heterotrimer that exhibits primer-independent RNA polymerase activity. J. Virol. 86, 4444–4454.

Xiong, S., Wang, Y.-F., Zhang, M.-Y., Liu, X.-J., Zhang, C.-H., Liu, S.-S., Qian, C.-W., Li, J.-X., Lu, J.-H., Wan, Z.-Y., others, 2004. Immunogenicity of SARS inactivated vaccine in BALB/c mice. Immunol. Lett. 95, 139–143.

Xu, B., Jin, Z., Wang, H., Jin, Q., Jin, X., Cen, P., 2008. Evolution of Streptomyces pristinaespiralis for resistance and production of pristinamycin by genome shuffling. Appl. Microbiol. Biotechnol. 80, 261–267.

Xu, Y., Lou, Z., Liu, Y., Pang, H., Tien, P., Gao, G.F., Rao, Z., 2004. Crystal Structure of Severe Acute Respiratory Syndrome Coronavirus Spike Protein Fusion Core. J. Biol. Chem. 279, 49414–49419. doi:10.1074/jbc.M408782200

Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., Li, J., Zhou, Z., Ding, Y., Zhao, Q., others, 2008. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. J. Virol. 82, 2515–2527.

Ying, T., Du, L., Ju, T.W., Prabakaran, P., Lau, C.C., Lu, L., Liu, Q., Wang, L., Feng, Y., Wang, Y., others, 2014. Exceptionally potent neutralization of Middle East respiratory syndrome coronavirus by human monoclonal antibodies. J. Virol. 88, 7796–7805.

Yoneyama, M., Fujita, T., 2010. Recognition of viral nucleic acids in innate immunity. Rev. Med. Virol. 20, 4–22.

Yoshimori, T., Noda, T., 2008. Toward unraveling membrane biogenesis in mammalian autophagy. Curr. Opin. Cell Biol. 20, 401–407.

Yu, I.-M., Oldham, M.L., Zhang, J., Chen, J., 2006. Crystal structure of the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein dimerization domain reveals evolutionary linkage between corona-and arteriviridae. J. Biol. Chem. 281, 17134–17139.

Yu, J., Li, X., Ao, Y., Li, L., Liu, N., Li, J., Duan, Z., 2013. Identification of a Novel Picornavirus in Healthy Piglets and Seroepidemiological Evidence of Its Presence in Humans. PLOS ONE 8, e70137. doi:10.1371/journal.pone.0070137

Yu, L., Pei, X., Lei, T., Wang, Y., Feng, Y., 2008. Genome shuffling enhanced L-lactic acid production by improving glucose tolerance of Lactobacillus rhamnosus. J. Biotechnol. 134, 154–159.

Yuan, L., Kurek, I., English, J., Keenan, R., 2005. Laboratory-Directed Protein Evolution. Microbiol. Mol. Biol. Rev. 69, 373–392. doi:10.1128/MMBR.69.3.373-392.2005

Zaccolo, M., Gherardi, E., 1999. The effect of high-frequency random mutagenesis on in vitro protein evolution: a study on TEM-1 β-lactamase. J. Mol. Biol. 285, 775–783.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D., Fouchier, R.A., 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N. Engl. J. Med. 367, 1814–1820.

Zeng, Y., Ye, L., Zhu, S., Zheng, H., Zhao, P., Cai, W., Su, L., She, Y., Wu, Z., 2008. The nucleocapsid protein of SARS-associated coronavirus inhibits B23 phosphorylation. Biochem. Biophys. Res. Commun. 369, 287–291.

Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M., Rao, Z., 2005. Insights into SARS-CoV transcription and replication from the structure of the nsp7–nsp8 hexadecamer. Nat. Struct. Mol. Biol. 12, 980–986.

Zhang, J., 2016. Porcine deltacoronavirus: Overview of infection dynamics, diagnostic methods, prevalence and genetic evolution. Virus Res. 226, 71–84.

Zhang, W., Li, L., Deng, X., Kapusinszky, B., Delwart, E., 2014. What is for dinner? Viral metagenomics of US store bought beef, pork, and chicken. Virology 468, 303–310. doi:10.1016/j.virol.2014.08.025

Zhao, H., Moore, J.C., Volkov, A.A., Arnold, F.H., 1999. Methods for optimizing industrial enzymes by directed evolution. Man. Ind. Microbiol. Biotechnol. 597–604.

Zhao, H., Zha, W., Svendsen, A., 2004. Evolutionary methods for protein engineering. Enzyme Funct. Des. Eng. Screen. 353–373.

Zhao, P., Cao, J., Zhao, L.-J., Qin, Z.-L., Ke, J.-S., Pan, W., Ren, H., Yu, J.-G., Qi, Z.-T., 2005. Immune responses against SARS-coronavirus nucleocapsid protein induced by DNA vaccine. Virology 331, 128–135.

Zhao, Q., Weber, E., Yang, H., 2013. Recent developments on coronavirus main protease/3C like protease inhibitors. Recent Patents Anti-Infect. Drug Disc. 8, 150–156.

Zhao, X., Nicholls, J.M., Chen, Y.-G., 2008. Severe acute respiratory syndrome-associated coronavirus nucleocapsid protein interacts with Smad3 and modulates transforming growth factor-β signaling. J. Biol. Chem. 283, 3272–3280.

Zheng, B.-J., Guan, Y., Hez, M.L., Sun, H., Du, L., Zheng, Y., Wong, K.-L., Chen, H., Chen, Y., Lu, L., others, 2005. Synthetic peptides outside the spike protein heptad repeat regions as potent inhibitors of SARS-associated coronavirus. Antivir Ther 10, 393–403.

Zhu, B., Cai, G., Hall, E.O., Freeman, G.J., 2007. In-FusionTM assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. Biotechniques 43, 354–359.

Zhu, M.-S., Pan, Y., Chen, H.-Q., Shen, Y., Wang, X.-C., Sun, Y.-J., Tao, K.-H., 2004. Induction of SARS-nucleoprotein-specific immune response by use of DNA vaccine. Immunol. Lett. 92, 237–243.

Ziebuhr, J., 2005. The coronavirus replicase, in: Coronavirus Replication and Reverse Genetics. Springer, pp. 57–94.

Ziebuhr, J., Heusipp, G., Siddell, S.G., 1997. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. J. Virol. 71, 3992–3997.

Ziebuhr, J., Schelle, B., Karl, N., Minskaia, E., Bayer, S., Siddell, S.G., Gorbalenya, A.E., Thiel, V., 2007. Human Coronavirus 229E Papain-Like Proteases Have Overlapping Specificities but Distinct Functions in Viral Replication. J. Virol. 81, 3922–3932. doi:10.1128/JVI.02091-06

Ziebuhr, J., Snijder, E.J., Gorbalenya, A.E., 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J. Gen. Virol. 81, 853–879.

Zúñiga, S., Sola, I., Moreno, J.L., Sabella, P., Plana-Durán, J., Enjuanes, L., 2007. Coronavirus nucleocapsid protein is an RNA chaperone. Virology 357, 215–227.

Zuo, X., Mattern, M.R., Tan, R., Li, S., Hall, J., Sterner, D.E., Shoo, J., Tran, H., Lim, P., Sarafianos, S.G., others, 2005. Expression and purification of SARS coronavirus proteins using SUMO-fusions. Protein Expr. Purif. 42, 100–110.

Züst, R., Cervantes-Barragan, L., Habjan, M., Maier, R., Neuman, B.W., Ziebuhr, J., Szretter, K.J., Baker, S.C., Barchet, W., Diamond, M.S., 2011. Ribose 2 [prime]-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. Nat. Immunol. 12, 137–143.

Züst, R., Cervantes-Barragán, L., Kuri, T., Blakqori, G., Weber, F., Ludewig, B., Thiel, V., 2007. Coronavirus non-structural protein 1 is a major pathogenicity factor: implications for the rational design of coronavirus vaccines. PLoS Pathog 3, e109.