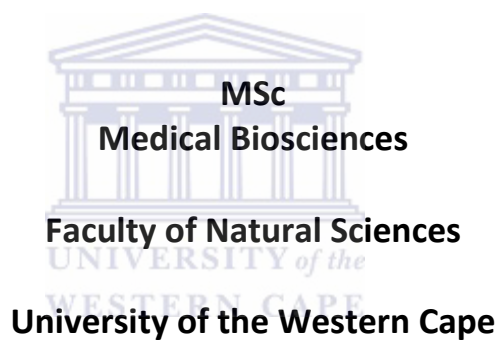


Expression studies of Human Coronavirus NL63- Nucleocapsid, Membrane and Envelope Proteins

**by
Taryn-lee Manasse**

Thesis Submitted in Fulfilment of the Requirements for the Degree



Supervisor: Professor BC Fielding

Department of Medical Biosciences

University of the Western Cape

November 2012

DECLARATION

I, Taryn-lee Manasse, declare that this thesis, "***Expression studies of Human Coronavirus NL63- Nucleocapsid, Membrane and Envelope Proteins***" hereby submitted to the University of the Western Cape for the degree of *Magister Scientiae* (MSc) has not previously been tendered by me for a degree at this or any other university or institution, that it is my own work in design and in execution, and that all materials contained herein have been duly acknowledged.

Taryn-lee Manasse :

Date signed :



DEDICATION

This study is dedicated to my parents Lawrence and Lillian Manasse.



ACKNOWLEDGEMENTS

- I would like to first and foremost thank my Heavenly Father God, for it's through His mercies that I'm able to do this Thesis.
- I would like to express my profound gratitude to my supervisor/ mentor/life coach/friend and enemy at times Prof B.C. Fielding for his immense support and motivation throughout my M.Sc and thank you for allowing me the opportunity to work with you.
- I sincerely would like to thank my senior colleagues, Tasnim Suliman, Michael Berry and Randal Fisher for their knowledge and assistance.
- To my workmates and members of the virology lab Aasiyah Chafekar, Yanga Mnyamana, Tiza Nguni, Thato Motlhalamme, Anele Gela, Marjorie Van Zyl and Shanel Swartz, thank you guys for your constant support, amity and memorable times spent together working in the "Viro Lab".
- Special thanks to my biggest cheerleaders in life this being my parents, siblings, friends and family for their endless encouragement, prayers and supporting me through the most disheartening of times.
- To my fiancé Hylton Barnes thank you for your love, patience, and preposterous amounts of luxuries involved in a fiancé writing a thesis.
- I would like to thank the University of the Western Cape and especially the department of Medical Biosciences, for allowing me the opportunity to complete all my studies from first year of undergrad to my M.Sc.
- Thank you to the NRF for funding my M.Sc.

ABSTRACT

Acute respiratory infections (ARI) continue to be the leading cause of acute illnesses worldwide and remain the most important cause of infant and young children mortality. Many viruses such as rhinoviruses, influenza viruses, parainfluenza viruses, respiratory syncytial viruses, adenoviruses and coronaviruses are deemed to be the etiological agents responsible for ARI's in children. The recently discovered coronaviruses HCoV-HKU1 and HCoV-NL63 contribute significantly to the hospitalization of children with ARI's.

HCoV-NL63 was first identified in 2004, as the pathogen responsible for the hospitalization of a 7 month old child presenting with coryza, conjunctivitis and fever. Since then a significant amount of knowledge has been gained in the clinical spectrum on this virus, however HCoV-NL63 is still not well characterized on the molecular and proteomic level.

This dissertation focuses on bringing about this characterization by cloning the HCoV-NL63 Nucleocapsid gene to be expressed in a bacterial system and transfecting the Nucleocapsid, Membrane and Envelope genes into a Mammalian cell culture system in order for its respective proteins to be expressed. With the use of Bioinformatic analytic tools certain characteristics of HCoV-NL63 Nucleocapsid, Membrane and Envelope proteins are able to be identified, as well as certain motifs and/or regions that are important in the functioning of these proteins. By comparing the results obtained for HCoV-NL63 N,M and E to other well studied coronavirus homologous will enlighten us on the potential role(s) of these proteins in determining HCoV-NL63 pathogenicity and infectivity.

Although certain functions of these proteins can be deduced by the means of bioinformatics analysis, it is still imperative for it to be extensively characterized *In Vitro*. This will therefore form a fundamental step in the development of many other projects, which unfortunately fall outside the scope of this M.Sc thesis.



KEY WORDS

Human Coronavirus NL63, Nucleocapsid Protein, Membrane Protein, Envelope Protein Severe Acute Respiratory Syndrome Coronavirus, Bioinformatic Analysis, Transmembrane regions, Hydropathy Domains, Protein Topology, Molecular Cloning, Bacterial and Mammalian Protein Expression.



TABLE OF CONTENTS

	Declaration	ii
	Dedication	iii
	Acknowledgements	iv
	Abstract	v
	Keywords	vii
	Table of Contents	viii
	List of Abbreviations	xv
	List of Tables	xix
	List of Figures	xx
	List of Appendixes	xxii
	Publications	xxiii
Chapter 1	Literature Review	1
1	Introduction	2
1.1	The Discovery of Coronaviruses	3
1.2	SARS and Coronavirus Research	4

1.3	Human Coronavirus NL63 (HCoV-NL63)	6
1.3.1	The Discovery of HCoV-NL63	6
1.3.2	HCoV-NL63 Infection	9
1.3.3	HCoV-NL63: Possible association with specific diseases:	11
1.3.3.1	Croup	11
1.3.3.2	Kawasaki Disease (KD)	13
1.3.4	Co-infection of HCoV-NL63 with other viruses	14
1.3.5	HCoV-NL63 Pathogenesis	16
1.3.6	HCoV-NL63 Genome Organization	16
1.4	Coronaviruses	18
1.4.1	Coronavirus nucleocapsid (N) protein	19
1.4.2	N-terminal-Domain and C-Terminal Domain involvement in RNA Binding	20
1.4.3	Interaction between the N and M proteins	21
1.4.4	Phosphorylation of the Nucleocapsid protein	22
1.4.5	Localisation of the Nucleocapsid protein	23
1.4.6	Nucleocapsid protein involvement in viral Transcription and replication	24

1.4.7	Coronavirus membrane (M) protein	25
1.4.8	Coronavirus envelope (E) protein	27
	Aims of this thesis	29
Chapter 2	Materials and Methods	30
2.1	Bioinformatic analysis of HCoV-NL63 N, M and E proteins	31
2.1.1	Multiple sequence alignment	31
2.1.2	Identification of transmembrane regions	31
2.1.3	Hydropathy predictions	32
2.1.4	Identification of protein motifs	32
2.2	Cloning of genes for expression in an <i>E.coli</i> cell system	33
2.2.1	Materials and reagents	33
2.2.2	Bacterial strains and plasmids	33
2.2.3	RT of HCoV-NL63 RNA	33
2.2.4	PCR primer design	34
2.2.5	Polymerase chain reaction (PCR) and purification of the amplicons	35
2.2.6	Cloning and verification of purified amplicons	36

2.2.6.1	Ligation into pGEM T-Easy vector	36
2.2.6.2	Transformation of JM109 <i>E.coli</i> and screening for recombinant clones	36
2.2.6.3	Plasmid extraction (mini-prep)	38
2.2.6.4	Restriction endonuclease digest	38
2.2.6.5	Nucleotide sequencing and sequence analysis	39
2.2.7	Ligation into Flexi vector and transformation of KRX competent <i>E.coli</i>	39
2.2.7.1	Restriction with Flexi enzymes (<i>SgfI</i> and <i>PmeI</i>)	39
2.2.7.2	Ligation into Flexi vector	40
2.2.7.3	Transformation of KRX strain with recombinant plasmid	40
2.3	Expression of proteins in a mammalian cell system	42
2.3.1	Bacterial strains and recombinant plasmids	42
2.3.2	Verification of recombinant plasmids	42
2.3.2.1	Restriction endonuclease digest	42
2.3.2.2	Electrophoresis of the restriction enzyme digests of plasmids	44
2.3.3	Plasmid purification	44

2.3.4	Quantification of plasmids by means of fluorometry	47
2.3.5	Verification of plasmid constructs by means of restriction endonuclease digestion	47
2.3.6	Nucleotide sequencing and sequence analysis	47
2.3.7	Mammalian cell culture	47
2.3.8	Transfection of COS7 cells	48
2.3.8.1	Trypan blue dye exclusion	48
2.3.8.2	Cell counting with a hemocytometer (Neubauer chamber)	48
2.3.8.3	Seeding of cells	49
2.3.8.4	Tranfection of COS7 cells with N-, M-, E-HA and the negative control GST-HA	49
2.3.8.5	Cell lysis	50
2.3.9	Protein expression analysis	51
2.3.9.1	Cell lysate preparation	51
2.3.9.2	SDS-PAGE	51
2.3.9.3	Western Blot analysis	52

Chapter 3 Results and Discussion

3.1	Bioinformatic analysis of HCoV-NL63 Nucleocapsid, Matrix and Envelope proteins	54
3.1.1	Multiple sequence alignment	54
3.1.2	Transmembrane domains and hydropathy plots of HCoV- NL63 N, M and E proteins	60
3.2	Molecular cloning of HCoV-NL63 N protein for expression in Bacterial cells	73
3.2.1	PCR	73
3.2.2	<i>EcoRI</i> digest of pGEM-N	74
3.2.3	<i>SgfI</i> and <i>PmeI</i> digest of pGEM-N recombinants	76
3.2.4	Colony PCR of KRX competent <i>E.coli</i>	77
3.3.1	Mammalian expression studies of HCoV-NL63 N, M and E proteins	80
3.3.2	Verification of recombinant plasmids	80
3.3.2.1	Restriction endonuclease digest	80
3.3.3	Plasmid purification (midi-prep)	82
3.3.4	Quantification of plasmids by means of fluorometry	84
3.3.5	Expression of HCoV-NL63 N, M and E proteins in COS7 cells	84

3.3.6	Comparisons of bacterial and mammalian expression of the HCoV-NL63 nucleocapsid protein	88
-------	---	----

Chapter 4	Summary	91
------------------	----------------	----

Chapter 5	References	96
------------------	-------------------	----

Appendix		115
-----------------	--	-----



LIST OF ABBREVIATIONS

°C	degrees Celsius
µg	Micrograms
µl	Microliters
µM	Micromolar
A	adenine
amp	ampicillin
bp	base pair
BSA	bovine serum albumin
C	Cytosine
cfu	colony-forming units
dNTP	deoxyribonucleotide triphosphate
E	envelope
G	guanine
HCl	Hydrochloric acid
HCoV	human coronavirus
IBV	infectious bronchitis virus



IPTG Isopropyl β -D-1- thiogalactopyranoside

K Potassium

kB kilobase pairs

KD Kawasaki disease

LB Luria Bertani

M Molar

M membrane

mg milligrams

MgCl₂ Magnesium Chloride

MHV murine hepatitis virus

ml millileters

mM millimolar

N nucleocapsid

Na Sodium

NaCl Sodium Chloride

ng nanograms



PCR	polymerase chain reaction
RNAasin	ribonuclease inhibitor
S	spike
SARS	severe acute respiratory syndrome
SDS	Sodium dodecyl sulphate
sg	subgenomic
T	thymine
TGEV	porcine transmissible gastroenteritis virus
U	uracil
UV	ultraviolet
V	Volts
VLP	virus-like particle
X-Gal	X-galactose



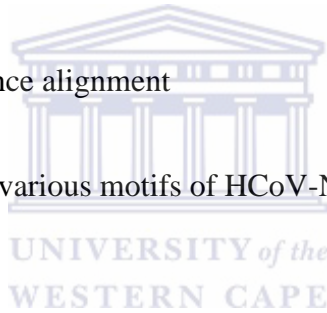
α alpha

β beta



LIST OF TABLES

Table 2.1	Human coronavirus structural proteins used for bioinformatics analysis	31
Table 2.2	Forward and reverse primer sequences for PCR amplification of SARS-CoV N and HCoV-NL63	34
Table 2.3	Constructs used for expression of proteins in COS7 cells	42
Table 2.4	Composition of all the buffers used in this experiment	45
Table 2.5	Functions of components of chosen cell lysis buffer	51
Table 3.1	Multiple sequence alignment	57
Table 3.2	Comparison of various motifs of HCoV-NL63 proteins N, M and E	69
Table 3.3	Restriction enzyme characteristics	81



LIST OF FIGURES

Figure 1.1	The VIDISCA method used to discover HCoV-NL63	7
Figure 1.2	Genomic structure of HCoV-NL63	17
Figure 3.1	Multiple sequence alignment of various human coronavirus nucleocapsid proteins	58
Figure 3.2	Multiple sequence alignment of various human coronavirus membrane (M) proteins	59
Figure 3.3	Multiple sequence alignment of various human coronavirus envelope (E) proteins	60
Figure 3.4	Comparisons of transmembrane regions with hydropathy predictions of HCoV-NL63 N protein	62
Figure 3.5	Comparisons of transmembrane regions with hydropathy predictions of HCoV-NL63 M protein	63
Figure 3.6	Comparisons of transmembrane regions with hydropathy predictions of HCoV-NL63 E protein	65
Figure 3.7	Illustration of the topology of HCoV-NL63 M protein	67
Figure 3.8	Illustration of the topology of HCoV-NL63 E protein	68

Figure 3.9	PCR amplification of viral genes using primers described in Table 2.2 .	74
Figure 3.10	Confirmation of successful ligation by restriction of the pGEM vector with <i>EcoRI</i> enzymes	76
Figure 3.11	Restriction enzyme digest of pGEM vector recombinants with pFLEXI enzymes <i>SgfI</i> and <i>PmeI</i>	77
Figure 3.12	Colony PCR of KRX competent <i>E.coli</i> previously transformed with FLEXI vector ligated with viral genes. Colony PCR was performed with primers described in Table 2.2 and used as confirmation of successful transformation and ligation.	79
Figure 3.13	Restriction enzyme digest of pCAGGS plasmid constructs with <i>EcoRI</i> and <i>NotI</i>	82
Figure 3.14	Restriction enzyme digest of pCAGGS plasmid constructs with <i>EcoRI</i> and <i>NotI</i> for verification	83
Figure 3.15	Western blot analysis of expressed proteins in COS7 cells	86
Figure 3.16	Expression of HCoV-NL63 N protein	88

LIST OF APPENDIXES

Appendix 1	Sequence verification of cloned SARS N-gene in pGEM. Sequencing was done using M13 forward and reverse primers	116
Appendix 2	Sequence verification of cloned HCoV-NL63 N-gene in pGEM	117
Appendix 3	Sequencing vector pGEM-T Easy vector map	118
Appendix 4	Bacterial protein expression vector, pFNA2A Flexi™	119
Appendix 5	pCAGGS vector Map	120
Appendix 6	Sequence verification of pCAGGS-N-HA	121
Appendix 7	Sequence verification of pCAGGS-M-HA	122
Appendix 8	Sequence verification pf pCAGGS-E-HA	123
Appendix 9	Kyte –Doolittle Hydropathy Scores	124

PUBLICATIONS

Abstracts accepted for publication in Journals:

Berry, M. **Manasse**, T. Tan, Y. Fielding, B.C. (2012). "Characterisation of human coronavirus-Nl63 nucleocapsid protein." *African Journal of Biotechnology* **11**(75): 13962-13968.



Chapter 1

Literature Review



1. Introduction

Acute respiratory illnesses (ARIs), also known as acute respiratory infections, are a major international health concern. As a group, ARIs affect all age groups and account for approximately 4 million worldwide deaths annually. These deaths place an economic burden on nations' health care systems, especially those in third world countries. Viruses such as rhinoviruses, influenza viruses, parainfluenza viruses, respiratory syncytial viruses, adenoviruses and coronaviruses have been shown to be most frequently associated with respiratory infections (Bryce et al. 2005).

It has previously been reported that human coronaviruses account for a significant number of hospitalisations for children under 18 years of age, accounting for 4.4% of all admissions for acute respiratory infections (Chiu, Chan et al. 2005).

Human coronaviruses (HCoVs) are considered as among the leading causes of ARIs. HCoVs were first isolated in the 1960s, and since then have not only been found to be mainly associated with respiratory tract illnesses, but can also cause diseases of the enteric and central nervous systems (Wu, Chang et al. 2008).

Until recently, only two human coronaviruses were known to science, i.e. HCoV-229E and HCoV-OC43. With the outbreak of Severe Acute Respiratory Syndrome (SARS) in 2003, a third and previously unknown coronavirus was identified as the aetiological agent. The SARS coronavirus eventually spread from China to more than 30 countries, and resulted in more than 800 fatalities worldwide; this equated to an about 10% mortality rate. Then in 2004, a fourth human coronavirus was isolated from a 7-month-old child with respiratory symptoms. This virus was identified as a close relative of SARS-CoV and was named human coronavirus-NL63 (Pyrce, Jebbink et al. 2004). Other groups independently reported the presence of essentially the same virus in children, calling it HCoV-NL (Fouchier, Hartwig et al. 2004) and HCoV-NH

(Esper, Weibel et al. 2005), respectively; in this dissertation it will be referred to as human coronavirus-NL63 (HCoV-NL63).

1.1. The Discovery of Coronaviruses

The first coronavirus (CoVs), responsible for infectious bronchitis in chickens, was identified in 1937 by Beudette and Hudson by isolating the virus in chick embryos (Beaudette and Hudson, 1937). Corona-virology has advanced significantly since then with the discovery of more coronaviruses, especially human coronaviruses.

In 1965 Tyrell and Bynoe found that they could passage a virus known as B814 (Tyrell, Bynoe et al. 1968; Kahn and McIntosh 2005). It was isolated in an adult with a common cold and then cultured in human embryonic tracheal organ cultures. The presence of an infectious agent was demonstrated by inoculating the medium from these cultures intranasal in human volunteers; in turn, colds were produced in a significant number of subjects, but Tyrell and Bynoe were unable to grow the agent in tissue culture at the time. It was Hamre and Procknow who set it apart as they were able to grow a virus with atypical properties in tissue culture from samples obtained, specifically from medical students with colds (Hamre and Procknow 1966; Pyrc, Berkhout et al. 2007). Both B814 and the virus that Hamre had discovered, called 229E, were ether sensitive and therefore most probably required a lipid-containing coat for infectivity, but these 2 viruses were not related to any known myxo- or paramyxoviruses (Hamre and Procknow 1966; Kahn and McIntosh 2005).

McIntosh et al had discovered multiple strains of ether-sensitive infectious agents from the human respiratory tract, using a technique similar to that of Tyrell and Bynoe, and designated the viruses as ‘OC’ because it was grown in organ cultures (McIntosh, Dees et al. 1967; Kahn and McIntosh 2005). These findings together led to the discovery of the first human coronaviruses, namely 229E and OC43, which were

extensively studied until the outbreak of SARS. As microscopy was not, at the time of Tyrell and Benoe's study, as advanced as it is today, many viruses resembled each other and were grouped together incorrectly. This appeared to be the case with coronaviruses as well. It was only in 1975 that coronaviruses were finally accepted as a new genus (Pyrce, Berkhout et al. 2007).

1.2. SARS and coronavirus research

In 2003, the emergence of Severe Acute Respiratory Syndrome brought the field of 'corona-virology' back into the spotlight (McIntosh 2005). It seemed clear that this new disease, inimitable in its clinical spectrum, had resulted from the movement of an animal CoV across the species barrier; it was found the virus had spread in the human population through a process of adaptation by deletion and mutation (Kuiken, Fouchier et al. 2003; Peiris, Lai et al. 2003; McIntosh 2005).

Towards the end of June 2003, patients from 29 countries had been diagnosed with SARS, with the majority of these cases reported in Asia. Because of the absence of an effective source of treatment, the disease was combated using only quarantine measures and travel restrictions. This, however, proved to be an effective strategy and the SARS epidemic was halted in June 2003. the death toll had climbed to 774 out of 8098 reported infections, and the disease had caused massive economic harm (Stadler, Masignani et al. 2003). The efforts to control SARS were matched by a multi-centred, collaborative effort by scientists to identify and analyse the infectious agent that caused the disease (Hofmann, Pyrc et al. 2005). The collaborative research had culminated in the identification of a novel coronavirus (CoV) in SARS patients (Kuiken, Fouchier et al. 2003; Peiris, Lai et al. 2003); thereafter it was demonstrated that the virus had been the cause of the disease (Fouchier, Kuiken et al. 2003; Kuiken,

Fouchier et al. 2003). Shortly after the discovery of SARS, the virus sequence was made available to the public (Marra, Jones et al. 2003; Rota, Oberste et al. 2003), which led to the analysis and eradication of SARS at a molecular level (Hofmann, Pyrc et al. 2005).

The characterisation of SARS-CoV revealed that highly pathogenic human CoVs (HCoVs) can evolve as result of its ability to mutate. Therefore the identification and characterisation of new HCoV species is a critical objective as it could prevent future epidemics.

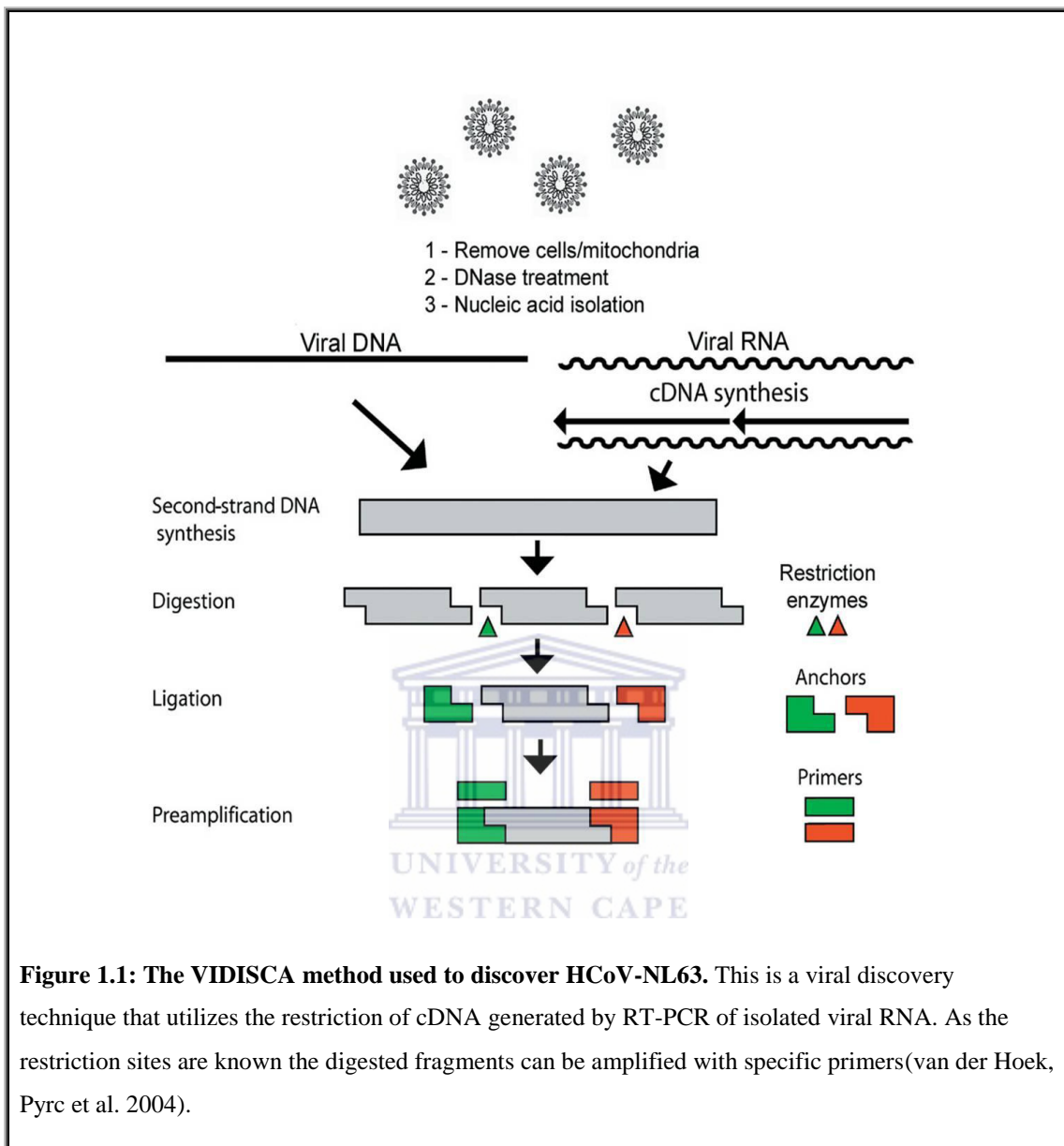
The SARS-CoV is, allegedly, no longer circulating among humans. However, at least four other coronaviruses are continuously making its rounds in the human population, especially among young children. Two of the four coronaviruses were recently discovered: HCoV-NL63 and HCoV-HKU1 (Krzysztof Pyrc 2004; van der Hoek, Pyrc et al. 2004; Woo, Lau et al. 2005). The other two, HCoV-OC43 and HCoV-229E, were identified in the mid-1960s as previously described (Hamre and Procknow 1966; Tyrrell, Bynoe et al. 1968). The latter viruses were tested for pathogenicity in human volunteers, which demonstrated that HCoV-229E and HCoV-OC43 cause the common cold (Bradburne, Bynoe et al. 1967). For now, the new viruses – HCoV-NL63 and HCoV-HKU1 – are lacking human and animal test model systems.

Since the discovery of HCoV-NL63 in 2004, much has been learnt this virus. Several groups have studied the global spread, the association with human disease, and the replication characteristics of HCoV-NL63 (van der Hoek and Berkhout 2005; van der Hoek, Pyrc et al. 2006). These virus characteristics are still to be defined on a molecular level, though.

1.3. Human Coronavirus NL63 (HCoV-NL63)

1.3.1. *The Discovery of HCoV-NL63*

In January of 2003, a 7-month-old child was hospitalised in Amsterdam with signs and symptoms of coryza, conjunctivitis and fever (van der Hoek, Pyrc et al. 2006). The prognosis was as follows: A chest radiograph showed typical features of bronchiolitis; a nasopharyngeal aspirate collected five days after the onset of the disease had tested negative for all known respiratory viruses. The clinical sample collected was then inoculated onto tertiary monkey kidney cells (tMK; Cynomolgus monkey) and a cytopathic effect was observed. The infectious agent could then subsequently be passaged onto LLC-MK2 cells (a monkey kidney cell line). Lia van der Hoek and her colleagues then identified the virus in the supernatant of the LLC-MK2 cell culture by means of the VIDISCA (Virus Discovery cDNA-AFLP) method (van der Hoek, Pyrc et al. 2004). This is a novel viral discovery technique that utilises the restriction of cDNA generated by RT-PCR of isolated viral RNA. As the restriction sites are known, the digested fragments can be amplified with specific primers (Fig 1.1) (van der Hoek, Pyrc et al. 2004). The PCR products obtained from this technique had revealed sequence similarity to the genome sequences of members of the *coronaviridae* family. The analysis of the virus genome had also shown that this virus was not a recombinant but rather a new member of the group I coronaviruses (van der Hoek, Pyrc et al. 2004; Pyrc, Berkhout et al. 2007).



A second research group in the Netherlands described the same virus shortly after the first publication on HCoV-NL63 had been released. Fouchier and colleagues had described a virus (which they named HCoV-NL) in a Vero-E6 cell culture supernatant (Fouchier, Hartwig et al. 2004). But this virus had been isolated in 1988 from a nose swab sample from an 8-month-old boy suffering from pneumonia. It was also cultivated first on tMK cells and then subsequently passaged onto Vero cells. There was a very high similarity (98.8% at the molecular level) with the previously

described HCoV-NL63 strain, and it was concluded that these two virus isolates had represented the same species. This implies that HCoV-NL63 had been circulating in the human population for many years before its first discovery. Most of the differences between the two virus variants are clustered in the amino terminal region of the spike protein (van der Hoek, Pyrc et al. 2004; Pyrc, Berkhout et al. 2007).

Almost one year later, a third group had described the same human coronavirus (Esper, Weibel et al. 2005). This virus was identified (in New Haven, CT, USA) by screening patient samples with the use of universal coronavirus primers. But these did not match at a nucleotide level with HCoV-229E, HCoV-OC43 or SARS-CoV (van der Hoek and Berkhout 2005). Therefore, Esper and colleagues had named their virus the New Haven Coronavirus (HCoV-NH), even though the partial sequences of their isolates are very similar to the isolates from The Netherlands (94–100% identical at the nucleotide level) and therefore represent the same species (van der Hoek and Berkhout 2005).

Literature of the early coronaviruses indicates that HCoV-NL63, or the novel group II virus HCoV-HKU1, may have been previously observed (van der Hoek and Berkhout 2005; Pyrc, Berkhout et al. 2007). In some of the initial work done on human coronaviruses sometime during the mid-1960s, viruses could be propagated in human embryonic tracheal organ cultures that were not, or only distantly, related to HCoV-229E or HCoV-OC43 (McIntosh, Dees et al. 1967; Tyrrell, Bynoe et al. 1968). Unfortunately these strains B814 (Tyrrell, Bynoe et al. 1968), HCoV-OC16, HCoV-OC37, and HCoV-OC48 (McIntosh, Dees et al. 1967) had been lost for follow-up research, so it remains unsure if those viruses had represented HCoV-HKU1 or HCoV-NL63, or perhaps some still-to-be-discovered human coronaviruses (Van der Hoek and Berkhout, 2005).

At the start of the 1980s, studies were conducted on several isolates of human coronaviruses that had been obtained by either cell culture or organ culture (Macnaughton, Flowers et al. 1983; van der Hoek and Berkhout 2005).

A number of those isolates were serologically related to HCoV-OC43, and were therefore termed group II isolates, whereas those serologically related to HCoV-229E were assigned as group I isolates (Van der Hoek and Berkhout 2005; van der Hoek, Pyrc et al. 2006).

A few of the group I viruses assigned as strains – LP, KI, PR and TO – could be cultured in MRC continuous cells (human), but two of the strains known as AD and PA could only be cultured in human foetal tracheal and nasal organ cultures. According to Van der Hoek *et al*, the difference in culture properties reflects the difference in cell tropism of HCoV-229E and HCoV-NL63, and so raises the possibility that some of the earlier group I strains were actually HCoV-NL63 isolates (van der Hoek and Berkhout 2005; van der Hoek, Pyrc et al. 2006). Because Van der Hoek and her colleagues had lost the isolates AD and PA, this hypothesis could not be tested. The NL strain remains the earliest isolate of HCoV-NL63, which had been obtained from a patient in 1988. This was later, in 2004, identified as HCoV-NL63 (Fouchier, Hartwig et al. 2004; van der Hoek, Pyrc et al. 2004).

1.3.2 *HCoV-NL63 Infection*

Literature indicates that infection with HCoV-NL63 happens often (Fouchier, Hartwig et al. 2004; van der Hoek, Pyrc et al. 2004; Fielding 2011; Leung, Chan et al. 2012), with many reports revealing upper and lower respiratory tract illnesses in infected patients.

HCoV-NL63 has also shown a worldwide prevalence, as it has been reported to cause infection in various countries (Fouchier, Hartwig et al. 2004; Esper, Weibel et al. 2005; Vabret, Mourez et al. 2005; van der Hoek, Pyrc et al. 2006; Fielding 2011).

In Australia it had been reported that infection with HCoV-NL63 was detected in 16 out of 766 patients who had presented with acute respiratory disease (van der Hoek, Pyrc et al. 2004; Arden, Nissen et al. 2005).

A study done in Japan, by Ebihara and colleagues on 118 children under the age of two who had been hospitalised for bronchiolitis, had shown that three patients had tested positive for HCoV-NL63 (van der Hoek 2007). A second Japanese study indicated the presence of HCoV-NL63 in 5 out of 419 respiratory tract specimens collected from children with respiratory infections (Suzuki, Okamoto et al. 2005).

Reports in Belgium indicated HCoV-NL63 infections in 7 out of 279 children hospitalised with upper and lower respiratory tract illnesses (Moes, Vijgen et al. 2005). A study done in France had shown 28 out of 300 patients presenting with upper and lower respiratory tract illnesses under the age of 20 tested positive for HCoV-NL63 (Vabret, Mourez et al. 2005).

In Switzerland, research done on neonates had identified HCoV-NL63 in 6 out of 82 children during their first episode of lower respiratory tract illness (Kaiser, Regamey et al. 2005; van der Hoek, Ihorst et al. 2010). Two separate studies in Canada reported 45 patients had tested positive for HCoV-NL63 out of a total of 1765 specimens (Bastien, Anderson et al. 2005; Bastien, Robinson et al. 2005; van der Hoek, Sure et al. 2005).

In Hong Kong a report compiled by Chiu and his colleagues in 2005 indicated that HCoV-NL63 had been detected in 5 out of 587 children. These children had participated in a prospective study on children under the age of 18 with acute

respiratory tract infection. The study group was a good representation of the population in Hong Kong, and proposed that HCoV-NL63 causes infection in more than 200 hospitalised cases per 100 000 children all under the age of 6 (Chiu, Chan et al. 2005; van der Hoek, Sure et al. 2005).

German researchers had reported 49 samples had tested positive for HCoV-NL63 from 949 samples taken from children with lower respiratory tract illness; these children had participated in the PRI.DE (Paediatric Respiratory Infection in Germany) study (van der Hoek, Sure et al. 2005).

Based on these reports, it can be deduced that HCoV-NL63 can cause infection in the age range of 1 month to 100 years, with the highest proportion of positive specimens between the ages of 0-5 years.

This virus has also been implicated as the causative organism in illnesses other than upper and lower respiratory tract infections. Reports show that HCoV-NL63 is frequently observed to cause infection in patients with underlying illnesses, such as those who are immunocompromised because of therapies, the HIV-infected or those with other medical conditions such as chronic obstructive pulmonary disease (Fouchier, Hartwig et al. 2004; Arden, Nissen et al. 2005; Bastien, Anderson et al. 2005; Moes, Vijgen et al. 2005; van der Hoek, Sure et al. 2005).

1.3.3. *HCoV-NL63: Possible association with specific diseases:*

1.3.3.1. Croup

Croup, also known as laryngotracheitis, is indicated as inflammation of the trachea and is associated with a characteristic loud barking cough that may get worse at night. Croup is a common manifestation of lower respiratory tract illness among children. Respiratory viruses such as PIV1 have frequently been implicated as the cause

(Denny, Murphy et al. 1983; van der Hoek, Sure et al. 2005). There is much debate on whether HCoV-NL63 is the causative organism of croup. It hasn't yet, however, been ascertained if it is the causative organism even though research strongly supports this theory. Some researchers believe that there is a clear link between the two. A prospective population-based study done in Germany known as PRI.DE, which looked at lower respiratory tract illness among children younger than 3 years of age, had established a clear link between HCoV-NL63 and respiratory diseases (van der Hoek, Sure et al. 2005). The study had shown that HCoV-NL63 infection is strongly associated with croup. About half of the patients that tested positive for HCoV-NL63 were diagnosed with croup, and an even higher portion of the samples collected from patients with croup contained HCoV-NL63 RNA (17.4%), when compared to patients who did not have croup (4.2%, $P < 0.0001$) (van der Hoek, Sure et al. 2005).

Croup has been reported to occur most frequently in boys and predominantly in the second year of life (Denny, Murphy et al. 1983). A trend is seen to be followed by HCoV-NL63-infected patients: the ratio of boys to girls infected is 10:4, the median age in the outpatient group with HCoV-NL63 is 1.55 years, and it has been shown to mostly occur during the winter season (van der Hoek, Sure et al. 2005).

In a study with 69 samples of patients analysed with croup, croup had been frequently linked to PIV1 (14.5%), but PIV3 (15.9%), RSV-A (13.0%), PIV2 (7.2%), and RSV-B (1.4%) were also detected in a substantial fraction of the samples (van der Hoek, Sure et al. 2005). HCoV-NL63 could be detected in 17.4% of these croup patients and was therefore the most frequently identified respiratory virus for croup (van der Hoek, Sure et al. 2005).

1.3.3.2. Kawasaki Disease (KD)

In the developed world, KD is the most common cause of acquired heart disease in children (Baker, Shimizu et al. 2006; Dominguez, Anderson et al. 2006; van der Hoek, Sure et al. 2006; Pyrc, Berkhout et al. 2007). It presents with prolonged fever and a polymorphic exanthema, oropharyngeal erythema, and bilateral conjunctivitis. A number of epidemiological and clinical observations previously suggested that an infectious agent might be the cause of KD (Burgner and Harnden 2005). But even though many infectious agents have been proposed as the cause of KD, none have been consistently associated with the disease (Esper, Shapiro et al. 2005).

Esper et al., 2005, suggested that there is evidence that Kawasaki disease may be triggered by a response to an infectious agent. From hospital discharge records, they identified 53 children who were diagnosed with Kawasaki disease. From these children they had received 11 respiratory specimens. 8 of the 11 patients all under the age of 4 years tested positive for HCoV-NL63 by RT-PCR. The researchers proposed that their data demonstrate that, in their population during the study period, there was a significant association between HCoV-NL63 infection and KD.

In contrast to Esper et al's results, a study conducted by Shimizu et al suggest that HCoV-NL63 is not associated with Kawasaki disease. Their study tested the possible association between HCoV-NL63 in the respiratory tract and acute KD by RT-PCR and viral culture of respiratory samples from a larger, geographically and ethnically diverse population. They tested 57 samples from 48 patients; only 1 tested positive for HCoV-NL63, which shows no significant association between HCoV-NL63 and KD (Shimizu, Shike et al. 2005).

There is, however, a fascinating link between HCoV-NL63 and Kawasaki disease. According to van der Hoek et al., 2006, multiple groups had screened for the presence of HCoV-NL63 in respiratory material of patients with Kawasaki disease (KD); they would not, however, confirm the findings of Esper et al (Belay, Erdman et al. 2005; Esper, Shapiro et al. 2005; Shimizu, Shike et al. 2005; Ebihara, Endo et al. 2005b; Wu, Chang et al. 2008).

The use of synthetic antibodies by Rowley et al had identified an antigen in respiratory epithelial cells and macrophages from children who were diagnosed with Kawasaki disease (Rowley, Shulman et al. 2005). The origin of this antigen remains unknown.

So whether or not HCoV-NL63 is associated with Kawasaki disease remains an open question, and future studies should continue to investigate the possibility of a respiratory portal of entry as the etiological agent of Kawasaki Disease.

1.3.4. *Co-infection of HCoV-NL63 with other viruses*

Co-infection of HCoV-NL63 with other respiratory viruses is reported to be a common occurrence, especially with other human coronaviruses, Influenza A virus, respiratory syncytial virus (RSV), parainfluenza virus and human metapneumovirus (hMPV) (Chiu, Chan et al. 2005; Kaiser, Regamey et al. 2005; Dare, Fry et al. 2007; Lambert, Allen et al. 2007; Canducci, Debiaggi et al. 2008; Minosse, Selleri et al. 2008; Wu, Chang et al. 2008; Abdul-Rasool and Fielding 2010; van der Hoek, Ihorst et al. 2010). It has been reported that co-infected patients are more likely to be hospitalised, thus indicating how severe a superinfection of this kind could be.

In a German study, co-infection of HCoV-NL63 with RSV-A was shown to be the most common co-infection identified in children under the age of 3 years. This is said to be probably due to the high incidence of RSV-A in winter and the possible overlap

in the seasonality of the viruses (van der Hoek 2007; Abdul-Rasool and Fielding 2010). In Italy it has been reported that HCoV-NL63 circulates as a mixture of variant strains and is also often associated with other viral infections (Minosse, Selleri et al. 2008; Abdul-Rasool and Fielding 2010). Co-infection of patients with HCoV-NL63 and Bacovirus has been reported among hospitalised children in South Africa. Respiratory tract samples from 341 patients were screened for common respiratory viruses, and the co-presence of HCoV-NL63 and bocavirus in at least one sample had been reported (Smuts 2008; Abdul-Rasool and Fielding 2010).

Research has shown that, in co-infected patients the viral load of HCoV-NL63 is lower than in patients infected with only HCoV-NL63. There seems to be a variety of possibilities for this phenomenon (van der Hoek 2007; Abdul-Rasool and Fielding 2010). Infection with HCoV-NL63 could be the initial infection that weakens the immune system to the degree that a second infection may gain a foothold. So by the time the secondary infection shows symptoms, the initial infection with HCoV-NL63 might have already been brought under control by the host immune system. The two viruses could also be in competition for the same target cells or receptors in the respiratory organs. The increase in the activation of the host innate immune response could be triggered by the second respiratory virus, thereby causing inhibition of HCoV-NL63 or lastly. The prolonged persistence of HCoV-NL63 could be at low levels.

The high rate of co-infection of HCoV-NL63 and other respiratory viruses may increase the chances of genetic recombination with these human or zoonotically transmitted viruses. It has been reported by Pyrc et al (2006) that HCoV-NL63 resulted from a recombination event between PEDV and an ancestral HCoV-NL63

strain (Pyrç, Bosch et al. 2006). Hypothetically, these types of recombination may enable the formation of highly pathogenic virus variants (Pyrç, Dijkman et al. 2006; Abdul-Rasool and Fielding 2010).

1.3.5. *HCoV-NL63 Pathogenesis*

HCoV-NL63 infection is initiated by the recognition of a specific receptor on the host cell surface by the Spike protein. This is followed by the internalisation of the virus, which may occur immediately by the direct fusion of the virus with the plasma membrane of the host cell or after endocytosis (Pyrç, Bosch et al. 2006; Pyrc, Berkhout et al. 2007). The fusion of the viral membrane with the cellular membrane activates the release of the viral RNA genome into the host cell cytoplasm. The viral RNA is then copied by the viral replicase in the membrane-associated replication centres (Pyrç, Bosch et al. 2006; Abdul-Rasool and Fielding 2010). During the replication process, copies of full length genomic RNA as well as a nested set of sub-genomic mRNA are generated. These subgenomic mRNAs serve as functional templates for the translation of the structural proteins that are encoded in the 3' one third of the genome. The full-length viral RNA then becomes encapsulated and is released from the host cell as an infectious virus particle. Further investigation of HCoV-NL63 pathogenicity seems warranted, as the virus shares the same cellular receptor as SARS-CoV (Li, Sui et al. 2007; Pyrc, Berkhout et al. 2007).

1.3.6. *HCoV-NL63 Genome Organization*

HCoV-NL63 has a very large, single stranded RNA genome that is made up of 27533 nucleotides; it is also capped and polyadenylated (Pyrç, Berkhout et al. 2007; Abdul-Rasool and Fielding 2010). It has a GC content of 34% and a coronavirus GC content that typically ranges between 32 to 42%; therefore HCoV-NL63 GC content is one of the lowest (van der Hoek, Pyrc et al. 2004; Pyrc, Berkhout et al. 2007). There are

untranslated regions of 286 and 287 nucleotides that are found at the 5' and 3' termini respectively (Pyrce, Berkhout et al. 2007). The genome of HCoV-NL63 follows a specific order, namely 5'-ORF1a-ORF1b-S-ORF3-E-M-N-PolyT-3' (Fig 2).

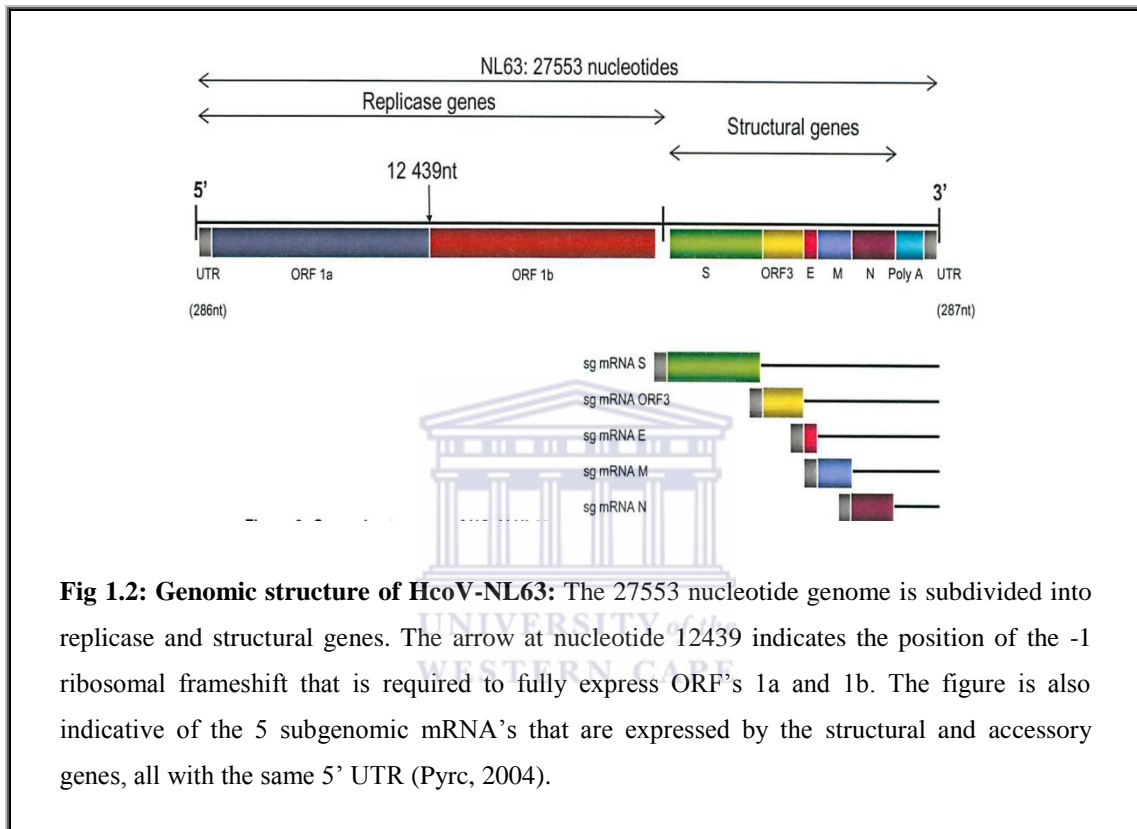


Fig 1.2: Genomic structure of HCoV-NL63: The 27553 nucleotide genome is subdivided into replicase and structural genes. The arrow at nucleotide 12439 indicates the position of the -1 ribosomal frameshift that is required to fully express ORF's 1a and 1b. The figure is also indicative of the 5 subgenomic mRNA's that are expressed by the structural and accessory genes, all with the same 5' UTR (Pyrce, 2004).

There are six mRNAs that produce seven distinct ORFs. This includes the full-length genomic RNA and a nested set of five subgenomic mRNAs (Pyrce, Jebbink et al. 2004). Coronaviruses are said to generate their mRNAs in the membrane-associated replication centres (Woo, Lau et al. 2005). The genes that are predicted to encode the Spike (S), Envelope (E), Membrane (M), and Nucleocapsid (N) proteins are found at the 3' part of the HCoV-NL63 genome (Pyrce, Jebbink et al. 2004; Pyrc, Berkhout et al. 2007). Some group II coronaviruses contain an HE gene which is not present in HCoV-NL63; there is only a single, monocistronic accessory protein ORF (ORF 3)

that is located between the E and S genes (Pyrce, Jebbink et al. 2004; Pyrc, Berkhout et al. 2007).

There are five subgenomic mRNAs that encode for the viral structural and accessory proteins S, ORF 3, E, M and N. A common transcription regulatory sequence (TRS), which contains a core sequence AACUAAA, is located upstream of all the ORFs, with the exception of ORF E. It has been known that this TRS is crucial for the formation of sg mRNA (Pyrce, Berkhout et al. 2007; Abdul-Rasool and Fielding 2010).

During the minus strand synthesis, HCoV-NL63 uses a discontinuous replication strategy to generate sg mRNAs (Pyrce, Jebbink et al. 2004; Pyrc, Berkhout et al. 2007).

This is then copied into plus strand mRNAs. All plus strand mRNAs share a common ~70 nucleotide leader sequence at their 5' end that is identical to the sequence at the 5' end of the genomic RNA (Pyrce, Berkhout et al. 2007; Abdul-Rasool and Fielding 2010).

1.4. Coronaviruses

Coronaviruses (CoVs) belong to the family Coronaviridae in the order *nidovirales*. These coronaviruses are divided into subgroups 1, 2 and 3, based mostly on genetic similarities (Abdul-Rasool and Fielding 2010). Coronaviruses are enveloped viruses with positive-sense, single-stranded, RNA genomes of 27 to 32 kb that are 5' capped and 3' polyadenylated (Pyrce, Jebbink et al. 2004; de Haan, Li et al. 2005). CoVs have a common genome organisation, in which the replicase gene encompasses the 5' two-thirds of the genome and is comprised of two overlapping open reading frames (ORFs), ORF1a and ORF1b. During coronavirus replication, a 3'-coterminal nested

set of subgenomic mRNAs, which encode structural viral proteins such as - spike (S) - envelope (E) - membrane (M) and nucleocapsid (N), are synthesised. Some group II CoVs carry an additional structural protein that encodes a hemagglutinin esterase. The gene is located between the ORF1b and S gene (Wurm, Chen et al. 2001). Expression of non-structural replicase proteins has been shown to be mediated by translation of the genomic RNA that in turn gives rise to the biosynthesis of two large polyproteins, namely pp1a, which is encoded by ORF1a and pp1ab; these are facilitated by a ribosomal frameshift at the ORF1a/1b junction. In contrast, the structural proteins are translated from subgenomic mRNAs. These subgenomic mRNAs are the result of discontinuous transcription, a specific feature of CoV gene expression (Dijkman, Jebbink et al. 2008). The structural gene region harbours several ORFs that are interspersed along the structural protein coding genes. The number and location of these accessory ORFs vary between the CoV species (Stanley G. Sawicki 2007).

1.4.1. Coronavirus Nucleocapsid Protein

During the coronavirus life cycle, the N protein is synthesised in large amounts and is thought to play an important role by specifically packaging the viral genome into a filamentous nucleocapsid of ~10 to 15 nm in diameter and several 100 nm in length, a macromolecular structure that is visible by electron microscopy (Lai 1997; Almazan, Galan et al. 2004). In addition to its structural role, the N protein also participates in viral RNA transcription, replication and in modulating the metabolism of host cells (Baric, Nelson et al. 1988; Hall 2002).

The highly basic N protein has a molecular mass ranging between 45 and 60 kDa in the various groups of coronaviruses and, along with its coding RNA, is synthesised in large amounts during infection (Kuo and Masters 2003; Narayanan, Kim et al. 2003).

The N protein is able to bind ssRNA nonspecifically, but displays an increased affinity for viral genomic RNA (Narayanan, Kim et al. 2003; Hsieh, Chang et al. 2005). Three groups of coronaviruses have been identified to date, although sequence conservation of the N proteins within the genus is low. For instance, the N proteins of coronaviruses IBV (group III) and porcine transmissible gastroenteritis virus (TGEV; group I) have only 29% identity with that of bovine coronavirus (BCoV; group II) and, within the group II coronaviruses, the N proteins of MHV and BCoV have only 70% identity (Lapps, Hogue et al. 1987). Several functions have been postulated for the coronavirus N protein throughout the virus life cycle, such as its various protein-protein interactions as discussed below.



1.4.2. N- Terminal-Domain and C-Terminal-Domains involvement in RNA Binding

The N protein is an essential structural component of the virion; this protein plays important roles in packaging the RNA genome to form a ribonucleoprotein (RNP) complex resulting from assembly of the viral RNA and multiple copies of the nucleocapsid protein. A meticulous dissection and characterisation of the intrinsic RNA-binding properties of these proteins are vital for understanding several important processes in the life cycle of RNA viruses, including assembly of the nucleocapsid, the specific encapsidation of viral RNA and the morphogenesis of virions (Arciola, Campoccia et al. 2006).

According to biochemical studies, the RNA binding site is found in the N-terminal domain (Risco, Anton et al. 1996; Lai 1997; Risco, Anton et al. 1998). The C-

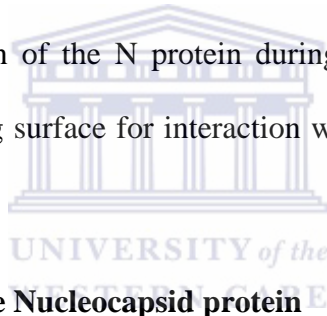
terminal domain is also involved in RNA binding and this was shown by Fan et al (Baric, Nelson et al. 1988). The N-terminal-domain of the IBV Gray strain has shown to exhibit a U-shaped structure, with two arms rich in basic residues at a 1.3-Å resolution, thus providing a specific component for interaction with RNA (Kuo and Masters 2003). The C-terminal-domain has shown to form a tightly intertwined dimer containing a central, intermolecular four-stranded β -sheet platform flanked by α -helices, thus indicating that a dimeric assembly of the N protein forms the basic building block for coronavirus nucleocapsid formation (Kuo and Masters 2003). According to Narayanan et al., the array of quaternary arrangements of the NTD and CTD revealed by the analysis of the different crystal forms could delineate possible interfaces that could be used for the formation of a flexible filamentous ribonucleocapsid (Narayanan, Kim et al. 2003).

1.4.3. Interaction between the N and M proteins

The importance of M–N interaction in viral interaction appears to be an essential process for coronavirus assembly as coronaviruses differ from other enveloped viruses such as paramyxoviruses and rhabdoviruses in that they lack the structural matrix protein that usually links the envelope to the nucleocapsid. Thereby, it can be envisaged that replacing the matrix protein in the role of core stabilisation through nucleocapsid binding appears to be one of the functions of the coronavirus M protein (He, Leeson et al. 2004).

Multiple copies of the nucleocapsid (N) phosphoprotein interact intimately with genomic and subgenomic RNA molecules during the virus life cycle (Fan, Ooi et al. 2005; Jayaram, Fan et al. 2006; Luo, Chen et al. 2006), and together with the M

protein, which is the most abundant envelope protein, participate in genome condensation and packaging. The interaction region between these two proteins has been mapped to their C termini (Masters, Kuo et al. 2006) by means of a reverse genetic-complementation assays (Jayaram, Fan et al. 2006). The C terminus of the M protein is known to be considerably basic, and a recent mutational study on the M protein has demonstrated that interactions between the N protein and M protein are predominantly electrostatic in nature (Baric, Nelson et al. 1988). Jayaram et al has demonstrated that the exposed acidic β -sheet floor, on the opposite side of the proposed RNA-binding region in the C-Terminal-Domain dimer, may promote such an interaction. This shows that the C-terminal-domain may serve a double purpose of mediating the self-association of the N protein during nucleocapsid formation, but also provides a corresponding surface for interaction with the endo-domain of the M protein in the virus envelope.



1.4.4. Phosphorylation of the Nucleocapsid protein

Coronavirus N proteins have a high (7 to 11%) serine content; serines are potential targets for phosphorylation (Laude, Godet et al. 1995). Between the two functional domains (CTD and NTD) is a segment containing several arginine/serine (RS) dipeptides that is structurally flexible. It is proposed that this RS-rich motif is an attribute of cellular precursor mRNA (pre-mRNA) splicing factors known as SR proteins (Fu 1995). The RS domain is phosphorylated with dynamism by several SR protein specific kinases, such as those belonging to the SR protein kinase (SRPK) and Clk families (Stojdl and Bell 1999). It has been shown that the phosphorylation of the RS domain modulates the activity, protein–protein interactions and subcellular localisation of SR proteins (Graveley 2000). Coronavirus N proteins are

phosphorylated in both host cells and in virions (Wootton, Rowland et al. 2002; Chen, Gill et al. 2005; White, Yi et al. 2007) and it has been reported that phosphorylation affects the RNA binding specificity and the nucleo-cytoplasmic shuttling of the N proteins (Chen, Gill et al. 2005; Surjit, Kumar et al. 2005). Phosphorylation can occur within the RS motif of coronavirus N proteins (Calvo, Escors et al. 2005; Surjit, Kumar et al. 2005) and thus may play a role in C-terminal domain dimerisation (Luo, Ye et al. 2005). Whether or not phosphorylation of the RS motif can modulate the functions of N proteins remains to be examined in detail.

1.4.5. Localisation of the Nucleocapsid protein

Coronavirus N proteins are found to be localised to both the cytoplasm and the nucleolus in virus-infected cells (Hiscox, Wurm et al. 2001; Wurm, Chen et al. 2001) and can transfer between the nucleus and the cytoplasm (Timani, Liao et al. 2005). Nucleolar localisation of N protein requires regions in the protein that are rich in arginine residues and is likely cell cycle-dependent (Hiscox, Wurm et al. 2001; Wurm, Chen et al. 2001). The avian infectious bronchitis virus N protein is able to interact and co-localise with the nucleolar proteins such as nucleolin and fibrillarin (Chen, Gill et al. 2005; Dove, You et al. 2006; Reed, Dove et al. 2006).

However, the ability of nucleolar localisation differs between N proteins of various coronaviruses. In the case of SARS-CoVN protein, it is poorly localised to the nucleolus (You, Dove et al. 2005). In a study done by Tsui-Yi Peng et al, it was found that the SARS-CoV N protein appeared in cytoplasmic stress granules (SGs) (Peng, Koetzner et al. 1995; He, Zheng et al. 2003). When eukaryotic cells encounter environmental stress, mRNA metabolism is reprogrammed, which enables it to adapt

to stress-induced damage. Translationally stalled mRNAs, together with a number of translation initiation factors and RNA-binding proteins, are deposited into SGs (Kedersha and Anderson 2002).

Given that amino acid sequence conservation within the N protein of the three different coronavirus groups is low, the fact that N proteins from group I and group II coronaviruses also localise to the nucleolus suggests that nucleolar localisation may be of functional significance.

In another study done on nucleolar localisation by Torsten Wurm et al in 2001, it was investigated if nucleolar localisation is a conserved feature among coronavirus N proteins and determined the consequences of N protein expression on host cell proliferation (Wurm, Chen et al. 2001). Their results show that association with the nucleolus is a common feature of the N proteins from the order *Nidovirales*. In addition, expression of N protein may lead to an inhibition of host cell proliferation and associated polyploidy in some cells, consistent with an inhibition of cytokinesis.

1.4.6. Nucleocapsid protein involvement in viral Transcription and Replication

The coronavirus N protein has shown to be involved in virus RNA synthesis by three lines of evidence. Firstly, the N protein is able to bind specific RNA sequences, this including the leader sequence, the TRS sequence, and sequences located at the 3' end of the virus genome (Shi and Lai 2005). Secondly, a small fraction of the N protein distributed in the host cell cytoplasm co-localises with the RTC during the early stages of infection (van der Meer, Snijder et al. 1999; Bost, Carnahan et al. 2000). And thirdly, the presence of N protein either in *cis* or in *trans* is required for efficient CoV RNA synthesis (Almazan, Galan et al. 2004). A quantitative analysis of TGEV

replicon activity showed an increase of more than 100-fold when N protein was present in *cis*, and an increase of more than 1000-fold when N protein was present in *trans* (Almazan, Galan et al. 2004). The importance of the N protein was also emphasised when it was shown that the rescue of recombinant coronaviruses from cells transfected with infectious RNA was to a great extent enhanced by, if not dependent upon, the expression of N protein in the same cell (Yount, Curtis et al. 2000; Casais, Thiel et al. 2001; Yount, Denison et al. 2002; Yount, Curtis et al. 2003). These annotations were followed up by experiments that showed that, in the absence of the N protein, HCoV-229E replicons expressing GFP from a subgenomic mRNA produced low levels of GFP and little or no amplification of the replicon. However, the expression of N protein in the same cells, in *cis* or in *trans*, significantly increased the number of transduced cells, the levels of GFP expression and thereby, the amplification of the replicon (Schelle, Karl et al. 2005).

1.4.7. Coronavirus Membrane (M) protein

The membrane (M) glycoprotein is known to be the most abundant structural protein (Cornelissen, van Woensel et al. 1998). It spans the membrane bilayer about three times, leaving a shortNH₂-terminal domain outside the virus; this may also be exposed lumenally in intracellular membranes, and a long COOH terminus known as the cytoplasmic domain inside the virion (Rottier, Brandenburg et al. 1984; Baudoux, Besnardeau et al. 1998).

A study done on the M protein of SARS-CoV had showed that it consists of 221 amino acids in length, and the gene sequence shares low homology with M proteins of other members of the coronavirus family. According to previous studies on

coronaviruses, the 25kDa M protein is an important and the most abundant structural protein, it is able to induce an antibody-dependent complement-mediated neutralisation reaction (Fang, Ye et al. 2005). The M protein also plays a predominate role in the envelope and naked virus particles assembly (Kuo and Masters 2003). Research has shown that the M protein contains highly conserved glycosylated sequences, and this may be due to the interaction between virus and host cells (de Haan, Kuo et al. 1998; de Haan, de Wit et al. 2002). In addition to this the M protein is characterised as having three domains; a short N-terminal ectodomain, a triple-spanning transmembrane domain, and a C-terminal endodomain (Armstrong, Niemann et al. 1984; Fang, Ye et al. 2005).

According to Narayanan and colleagues, mouse hepatitis virus assembly takes place at a few sites ie. The budding compartment and the smooth membranes of the intermediate compartment, which is found between the endoplasmic reticulum (ER) and the Golgi apparatus (Tooze, Tooze et al. 1984; Klumperman, Locker et al. 1994). The M protein itself is not able to determine the budding site; when M protein is expressed without the presence of other viral proteins, it migrates further than the budding compartment and localises in the late-Golgi complex (Klumperman, Locker et al. 1994). This is an indication of an unidentified viral factor(s) that restricts the migration of M protein to the budding compartment. A candidate that may cause restriction of the migration of M protein is the viral nucleocapsid. Narayanan and colleagues hypothesise that the binding of the nucleocapsid to M protein causes restriction of migration of the M protein to the budding compartment, and that this M protein-nucleocapsid interaction may facilitate the envelopment of the nucleocapsid at the budding compartment (Narayanan, Maeda et al. 2000).

1.4.8. Coronavirus Envelope Protein

The E protein is small in size and expressed in low quantities *in vivo* (de Haan, Vennema et al. 1998). While only 76 amino acids (aa) or so in length, research revealed that E protein (approximately 9 kDa) played a significant multifunctional role in the coronavirus virion life cycle (Bos, Luytjes et al. 1996; Vennema, Godeke et al. 1996; Shen, Wen et al. 2003).

As a membrane protein, the E protein has a major biological function in the participation of the formation of the viral envelope. It also plays an important role in viral replication in some coronavirus viruses, such as in transmissible gastroenteritis coronavirus (TGEV) and in marine hepatitis virus (MHV) (Godet, L'Haridon et al. 1992; Kuo and Masters 2003; Shen, Wen et al. 2003). In addition, it has been proven that E protein is related to the apoptosis of the E-protein-expressing cells in MHV (An, Chen et al. 1999; Shen, Wen et al. 2003).

When CoV M and E proteins are co-expressed in cell culture it produces virus-like particles (VLPs) (Vennema, Godeke et al. 1996; Baudoux, Carrat et al. 1998; Corse and Machamer 2000; Huang, Yang et al. 2004; Bosch, de Haan et al. 2005; DeDiego, Alvarez et al. 2007).

Several studies on VLPs agree that the formation of coronaviruses is mediated by just the M and the E proteins, and that neither S or N proteins play a vital role in virion morphogenesis (Bos, Luytjes et al. 1996; Baudoux, Carrat et al. 1998; Corse and Machamer 2000; de Haan, Vennema et al. 2000; Mortola and Roy 2004; Hurst, Kuo et al. 2005; Yang, Xiong et al. 2005; DeDiego, Alvarez et al. 2007). The exact role of the E protein in this process has not yet been determined (Hurst, Kuo et al. 2005), but

some evidence suggests an interaction between E and M (Baudoux, Carrat et al. 1998; Corse and Machamer 2000; Hurst, Kuo et al. 2005), while other publications suggest that E acts independently of M during the budding process (Maeda, Maeda et al. 1999; Corse and Machamer 2000).

There are conflicting reports on SARS-CoV about which proteins are required for the formation of VLPs (DeDiego, Alvarez et al. 2007). Some claim that E and M are required for the efficient assembly of these particles in insect cells (Ho, Lin et al. 2004; Mortola and Roy 2004), whereas others suggest that M and N proteins, instead of the E protein, play a key role in the formation of SARS-CoV-like particles in mammalian cells (Huang, Yang et al. 2004).

A study done on a recombinant MHV with a deletion of the E gene had confirmed that the E gene plays a critical role in viral assembly (Fischer, Stegen et al. 1998). The MHV remained viable but had low infectivity and replicated poorly, proving that while it is nonessential for MHV, the E gene is required for the production of an infectious virus (Kuo and Masters 2003). However, in the case of porcine transmissible gastroenteritis virus (TGEV), disruption of the E gene is lethal to the virus, implying that it is a vital protein (Curtis, Yount et al. 2002; Ortego, Escors et al. 2002; Hurst, Kuo et al. 2005). Apart from the described roles in virus assembly, the E protein has additional functions during infection (Weiss and Navas-Martin 2005).

Aims of this thesis:

1. Bioinformatic analysis of HCoV-NL63 M, N and E.
2. Cloning of HCoV-NL63 N gene for expression in a bacterial system.
3. Optimisation of expression of HCoV-NL63 M, N and E in COS-7 cells.
4. Comparison of HCoV-NL63 N expressed in a bacterial and mammalian system.



Chapter 2

Material and Methods



2. MATERIALS AND METHODS

2.1 Bioinformatic analysis of HCoV-NL63 N, M and E proteins

2.1.1 Multiple Sequence Alignment

The protein sequences were selected from the following accession numbers in NCBI: HCoV-NL63 (DQ846901.1), SARS-CoV (AY360146.1), HCoV-OC43 (AY585229.1), HCoV-229E (AAG48597), HKU1 (AAT98585). The selected amino acid sequences (Table 2.1) were aligned with CLUSTAL X version 2.0 (Larkin, Blackshields et al. 2007) and viewed with GeneDoc version 2.6.002 software (Nicholas and Nicholas 1997).

Table 2.1: Human coronavirus structural proteins used for bioinformatics analysis

HCoV	Size of protein (a.a.)		
	Nucleocapsid	Matrix	Envelope
HCoV-NL63	377	226	77
SARS-CoV	422	221	76
HCoV-OC43	448	230	84
HCoV-229E	389	225	77
HKU1	441	223	82

2.1.2 Identification of Transmembrane Regions

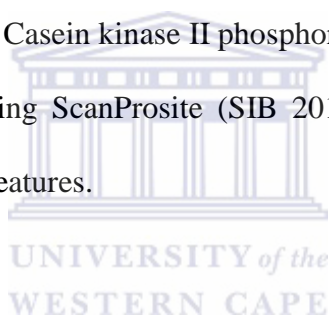
TMHMM Server V.2.0 (Krogh, Larsson et al. 2001) was used to search for transmembrane regions in the N, M and E protein sequences, using default settings for all algorithms.

2.1.3 Hydropathy Predictions

The hydropathy scores for N, M and E were predicted using the Kyte-Doolittle Hydropathy Plots (Kyte and Doolittle 1982); the window size for all amino acids was set at default with a window size of 9. This bioinformatics tool determines hydrophobic and hydrophilic regions within a protein sequence.

2.1.4 Identification of Proteins Motifs

The presence of various motifs including N- and O-linked Glycosylation, Myristoylation, Specific Amino acid rich domains (*e.g.* serine rich domains), Protein kinase C phosphorylation and Casein kinase II phosphorylation sites were predicted in the HCoV-NL63 proteins using ScanProsite (SIB 2012). ProRule (SIB 2012) was used to predict intra-domain features.



2.2 Cloning of genes for expression in an *E. coli* cell system

2.2.1 Materials and Reagents

HCoV-NL63 RNA, used as template to synthesize the 1st strand cDNA, was a kind gift from Prof. L van der Hoek, Holland. The RNA had been extracted from a fifth-passaged virus (strain Amsterdam1), that was obtained from a clinical sample. SARS-CoV N was a kind gift from Prof Y-J Tan, Singapore. SDS gel preparation kit, DMSO and all antibodies were purchased from Sigma Aldrich. All other chemicals were purchased from Promega.

2.2.2 Bacterial strains and plasmids

KRX and JM109 competent cells used in this project were purchased from Promega. JM109 competent cells were used for its high transformation efficiency and α -complementation, thus allowing for blue/white colony selection. On the other hand, KRX competent cells were used for transient expression of the recombinant proteins. Plasmid vectors used included the pGEM-T Easy Vector (Approximately 3.2 kb), which served as a shuttling vector and the pFLEXI vector (Approximately 4.4 kb), which was used as the bacterial expression vector.

2.2.3 RT of HCoV-NL63 RNA

Reverse transcription (RT) was carried out using a mixture of 2 μ l RNA template (10^8 - 10^9 copies/ml), 1 μ l of Oligo (dT)₁₅ primer (100 μ M stock), 1 \times incubation buffer, 2 μ l of dNTP mix (10mM stock), 20U RNasin® Ribonuclease inhibitor, 15U of AMV Reverse Transcriptase and 4 μ l MgCl₂ (25 mM), made up to a final volume of 20 μ l in nuclease-free water, according to the manufacturer's specifications (Promega). The

reaction was then heated at 42°C for 60 min, 95°C for 5 min and then cooled to 0°C for 5 min in order to deactivate the enzyme.

2.2.4 PCR Primer Design

For ligation into the pFlexi™ protein expression vector, restriction sites had to be manually inserted into the gene of interest. For this to be accomplished a *SgfI* (5' GCGATCGC 3') restriction-site was incorporated into the forward primer and a *PmeI* (5' GTTTAAAC 3') restriction site was incorporated into the reverse primer. The addition of a C residue upstream of the ATG in the forward sequence prevented frameshifts from occurring in the protein reading frame. Also, four additional nucleotides were added to the 5'- and 3'-ends to allow for a more efficient restriction digest of the PCR products. Primers for amplification of the N constructs (Table 2.2) were designed using sequences obtained from NCBI (the accession number for each sequence is shown): SARS-CoV N (AY360146.1) and HCoV-NL63 N (DQ846901.1)

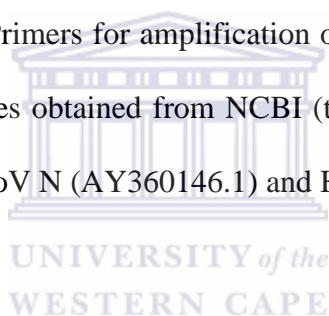


Table 2.2: Forward and Reverse primer sequences for PCR amplification of SARS-CoV N and HCoV-NL63 N

Gene	Primer	Sequence
SARS-CoV N	Forward	AGGA <u>GCGATCGC</u> CAT GTCTGATAATGGACCCCAAT CAAACC
	Reverse	TTGT <u>GTTTAAACT</u> TATGCCTGAGTTGAATCAGCAGA
HCoV-NL63 N	Forward	GGGC <u>GCGATCGC</u> CAT GGCTAGTGTA <u>AAT</u> TGGGCCG
	Reverse	ACAG <u>GTTTAAACT</u> TAATGCAAACCTCGTTGACAAT TTC

SgfI restriction site – striated box; *PmeI* restriction site – dashed line box; ATG start codon – Bold; TAA stop codon underlined

2.2.5 Polymerase Chain Reaction (PCR) and Purification of the Amplicons

The PCR reaction was set-up to a final volume of 25 μ l which included 0.5 μ l dNTP mix (0.2 mM), 2.0 μ l MgCl₂ (1.5 mM), 1 μ l for each forward and reverse primer (10 pmole/ μ l), 5 μ l 5x Green *GoTaq* Flexi™ buffer, 0.5 μ l *GoTaq* polymerase (5U/ μ l) and 2 μ l cDNA template (1:10 dilution of original cDNA sample). The annealing temperature was set to 60°C as it was the lowest temperature at which there was no non-specific amplification. The following PCR conditions were used: 95°C for 3 min, followed by 30 cycles of 95°C for 1 min, 60°C for 1 min and 72°C for 2min. The final elongation was run for 15 min at 72°C. The PCR product was diluted in a ratio of 1:6 with 6x Blue/Orange loading dye [0.4% (v/v) orange G, 0.03% (v/v) bromophenol blue, 0.03% (v/v) xylene cyanol FF, 15% (v/v) Ficoll® 400, 10 mM Tris-HCl (pH 7.5) and 50 mM EDTA (pH 8.0)]. This mixture was separated by electrophoresis on a 1% (w/v) agarose gel (containing 0.001% (v/v) ethidium bromide) in TBE [89 mM tris (hydroxymethyl) aminomethane, 0.089 mM boric acid, 2 mM EDTA (pH 8)] buffer at 90 V for approximately 1 hour. The agarose gel was viewed under short-wave length UV light. Bands of interest (chosen according to expected size) were excised and purified using the Wizard SV Gel and PCR Cleanup System (Promega). Briefly, the excised gel pieces were dissolved in 1 μ l/mg membrane binding solution (4.5 M guanidine isothiocyanate and 0.5 M potassium acetate) at 60°C with frequent vortexing. The dissolved gel mixture was then transferred to a minicolumn assembly and incubated at room temperature for 1 min. The DNA was transferred to the minicolumn membrane by centrifugation at 14000 rpm for 1 min. The column was washed, first with 700 μ l and then with 500 μ l, membrane wash solution (10 mM potassium acetate, 80% (v/v) ethanol and 16.7 μ M EDTA). Next, the columns were

centrifuged for 1 min at 14000 rpm to evaporate the residual ethanol present in the membrane washing solution. DNA was eluted with 30 μ l nuclease free water into a clean 1.5 ml Eppendorf tube by centrifugation at 14000 rpm. The purified PCR product was then quantified using a Qubit™ 1.0 fluorometer quantification system.

2.2.6 Cloning and Verification of Purified Amplicons

2.2.6.1 Ligation into pGEM T-Easy Vector

Next, the purified amplicons were ligated into a pGEM® T-easy shuttling vector according to manufacturer's instructions. This vector allows for easy sequencing of the cloned viral genes. Also, the 3' thymine overhangs present in the cloning site of the pGEM vector creates 'sticky ends' which allow for easy ligation of the amplified PCR product; *Taq* polymerase incorporates a template-independent deoxyadenosine residue to the 3' of the amplified fragment. pGEM® also contains a *β -galactosidase* gene which is used for blue/white colony screening. M13 primer recognition sites are also present up- and down-stream of the insert allowing for easy sequencing of the fragment. The ligation reaction was set up as follows: 5 μ l 2X T4 DNA rapid ligation buffer, 1 μ l pGEM® vector (50 ng), 3 μ l gel-purified PCR product and 1 μ l T₄ DNA Ligase (0.5 U). This reaction was then incubated for one hour at 4°C.

2.2.6.2 Transformation of JM109 *E. coli* and Screening for Recombinant Clones

Luria Bertani (LB) media (10 g pancreatic digest of casein or tryptone powder, 5 g yeast extract powder and 5 g NaCl per 1 L, pH 7.2) is the most common liquid media used in the cultivation of bacteria such as *E. coli*. The media was sterilized by autoclaving, after which 1 μ l/ml ampicillin was added. LB agar was prepared by adding 15 g bacteriological agar per litre of LB media.

JM109 (Promega) competent *E. coli* cells known for their high transformation efficiency was used in the transformation process according to the manufacturer's protocol, however minor modifications were done in order to obtain optimal results. In summary, 50 μ l of the competent cells were thawed on ice for about 5 minutes and thereafter 5 μ l of the ligation reaction was added to the cells, these were then incubated on ice for 20 minutes. The cells were then heat-shocked at 42°C for 45 sec in a water bath and returned to ice for 2 min. Then, 950 μ l of LB medium (without ampicillin), pre-warmed to room temperature, was added and incubated at 37°C for 1 hr 30 min. The cells were then pelleted at 2000 rpm for 10 min and resuspended in 200 μ l of fresh LB media. A spread plate of the cell suspension was done on LB agar plates containing 1 μ g/ml ampicillin, 0.1 M IPTG and 3% (v/v) X-gal and plates were incubated at 37°C for 16hrs. Following the incubation period, the plates were screened for white colonies, which were picked with a sterile toothpick. The toothpick was first used to inoculate a PCR mix for colony PCR-screening and then dropped into a culture-tube containing terrific broth media. The PCR reaction was done as described previously, but the initial denaturation temperature was increased to 5 min to make sure that the cells were completely lysed and the template strand denatured. The cell cultures were incubated at 37°C for 14 hours.

75% glycerol stocks were prepared from overnight cultures confirmed to contain the correct insert size by colony PCR. The glycerol stocks were prepared by adding 750 μ l of the cell suspension to 250 μ l of glycerol; this was thoroughly mixed and stored at -80°C for long-term storage. The remaining cell cultures, used for making glycerol stocks, were used to extract the plasmid DNA using the SV Mini-Prep kit (Promega).

2.2.6.3 Plasmid Extraction (Mini-Prep)

The mini-prep procedure is used to isolate plasmid DNA from bacterial cells using an alkaline lyses method. According to the manufacturer's protocol it had been advised to use 2 ml overnight culture for plasmid isolation however it was found that using a biomass (BM) index was more relevant to ensure the highest yield of plasmid DNA.

Biomass was therefore calculated using the following equation:

$$\text{OD}_{600} \text{ (mg/ml)} \times \text{Vol (ml)} = \text{BM (mg)}$$

A biomass of approximately 4 mg was used for all mini-preps. After the cells were harvested by centrifugation at 14000 rpm the pellet was resuspended in 250 μ l cell suspension buffer (50 mM Tris-HCl, 10 mM EDTA) containing 100 μ g/ml *RNase*. Cells were then lysed in an alkaline (pH 12) lysis buffer (1% (w/v) SDS, 0.2M NaOH) which denatures all DNA to its single stranded form. Plasmid DNA is, however, circular and therefore when the pH is neutralized (to pH 7), with the addition of 350 μ l neutralization buffer (80 mM potassium acetate, 8.3 mM Tris-HCl, 40 μ M EDTA), the plasmid can easily realign, where chromosomal DNA remains in a single stranded form and interacts with cellular debris. On centrifugation at 14000 rpm for 1 min the cellular debris, with bound chromosomal DNA, is pelleted and therefore separated from plasmid DNA. The clear lysate is then transferred to a minicolumn and plasmid DNA is bound to the silica membrane. Plasmid DNA is then washed with 700 μ l wash solution by centrifugation at 14000 rpm for 1 min and eluted in 50 μ l nuclease free water and transferred, by centrifugation at 14000 rpm, to a clean 1.5 ml Eppendorf tube. Plasmid DNA is stored at -20°C.

2.2.6.4 Restriction Endonuclease Digest

As *EcoRI* restriction sites are located up- and down-stream from the cloning site, removal of the ligated product with these restriction enzymes was performed as a

second confirmatory test. The restriction reaction was set up as follows; 2 µl 10x buffer H (50 mM Tris-HCl, 10 mM MgCl₂, 100 mM NaCl), 0.2 µl BSA (bovine serum albumin), 1.0 µl *EcoRI* and 16.8 µl isolated plasmid, which was incubated at 37°C for 3 hours in a water bath, following which, the reaction was viewed on a 1% (w/v) agarose gel under UV light to confirm the presence of the viral gene.

2.2.6.5 Nucleotide Sequencing and Sequence Analysis

Extracted plasmids were sequenced with M13 forward and reverse primers to ensure no base pair mutations occurred (Inqaba Biotec). Results were viewed with FinchTV (www.geospiza.com) and extracted sequences were aligned using the multiple sequence alignment function in ClustalX (Version 2.0.12). Files were subsequently exported to GeneDoc (Version 2.7.000) for viewing.

2.2.7 Ligation into Flexi Vector and Transformation of KRX competent *E. coli*

2.2.7.1 Restriction with Flexi Enzymes (*SgfI* and *PmeI*)

Flexi™ vectors (Appendix 12) contain a *barnase* gene (a lethal gene) between the *SgfI* and *PmeI* restriction sites, which prevents the growth of unsuccessfully ligated plasmids. This gene must first be removed with Flexi™ restriction enzymes before ligation with the viral gene. The *barnase* gene of the Flexi™ vector and viral gene ligated into the pGEM® vector was cut-out according to manufacturer's instruction. Briefly, the restriction reaction to remove the insert from the pGEM® vector was set up as follows, 4 µl 5x Flexi™ digest buffer (10 mM Tris HCl pH 7.4), 150 mM KCl, 50 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 0.25 mg/ml BSA, 0.05% (w/v) Thesit® and 50% (v/v) glycerol), 5 µl (±0.1µg/ul) mini-prep pGEM® vector, 1 µl Flexi™ enzyme blend (*SgfI* and *PmeI*) and 10 µl nuclease free water was incubated at 37°C for 3 hours. The enzyme was then deactivated at 65°C for 20 min. It was noticed that

yield was greatly increased when the nuclease free water was replaced with mini-prep pGEM® vector and therefore this volume was increased to 15 µl and nuclease free water was removed from the reaction. The restriction reaction was run on a 1% (w/v) agarose gel electrophoresis with ethidium bromide and viewed under UV light. The bands of interest were excised from the gel and gel-purified as previously described.

2.2.7.2 Ligation into Flexi Vector

Each insert was ligated into the pFN2A Flexi™ vector according to the manufacturer's instructions. Briefly, 10 µl 2X Flexi™ *ligase* buffer (60 mM Tris-HCl (pH 7.8 at 25°C), 20 mM MgCl₂, 20 mM DTT, 2 mM ATP), 5 µl acceptor Flexi™ vector (cut with *SgfI* and *PmeI* restriction enzymes), 4 µl (±0.1 µg/µl) viral gene product (cut with *SgfI* and *PmeI* restriction enzymes) and 1 µl T4 DNA ligase (5 U/µl) was incubated at 4°C, overnight. The Flexi™ vector appends a GST fusion protein to the N-terminal end of the expressed viral protein.

2.2.7.3 Transformation of KRX strain with Recombinant Plasmid

Ligated pFLEXI constructs were transformed into KRX strain of competent *E. coli* (Promega). Briefly, KRX cells were removed from -80°C freezer and thawed on ice, to which 5 µl ligation product was added and returned to ice for 10 min. The cells were then heat-shocked for 15-20 sec at exactly 42°C and returned to ice for 2 min. 450 µl LB broth, without ampicillin, was added to each transformation reaction and incubated for 60 min at 37°C with shaking at 180 rpm. Following which, cells were harvested at 2000 rpm for 10 min and resuspended in 200 µl LB broth, of which 100 µl was plated on LB plates containing 1 µg/ml ampicillin. Plates were incubated at 37°C for 16-20 hours. Single colonies were then picked and inoculated into terrific broth and corresponding PCR mix. Cultures were incubated at 37°C for 14 hours and colony PCR results were used to select cultures with correct insert. Cultures were

stored in glycerol stocks at -80°C . If colony PCR was not adequate to identify cultures with correct vector construct, the plasmids were isolated using the mini-prep technique and constructs were later identified by either PCR or restriction digest using *SgfI* and *PmeI* restriction enzymes.

Sequence-verified constructs were used for expression using the KRX cells. Even though generating the constructs for expression in a bacterial system formed part of this thesis, expression of the proteins were performed as part of a larger project and did not form part of the work reported here. However, the work reported in this section was included in a publication by Berry and colleagues (Berry 2012).



2.3 Expression of proteins in a mammalian cell system

2.3.1 Bacterial strains and recombinant plasmids

In this thesis, the DNA constructs used for expression of proteins in COS-7 cells were kind gifts from Dr M. Muller, Germany and Prof Y-J Tan, Singapore (Table 2.3); all constructs were received as glycerol stock of transformed *E. coli* cultures. Before proceeding to the expression studies, recombinant plasmids were verified for correct inserts. Briefly, a loop-full of the respective bacterial glycerol stock was streaked onto an LB agar plate and incubated overnight. Thereafter, it was screened for single colonies and inoculated into LB media containing ampicillin for 14 hours. Following incubation, plasmid DNA was extracted as described in 2.2.6.3.

Table 2.3: Constructs used for expression of proteins in COS-7 cells

Constructs	HA-tagged protein expressed	Source
pCAGGS-N-HA	HCoV-NL63 nucleocapsid	M. Muller
pCAGGS-M-HA	HCoV-NL63 membrane	M. Muller
pCAGGS-E-HA	HCoV-NL63 envelope	M. Muller
pXJ40-GST-HA	Glutathione S-transferase (used as positive transfection control)	Y-J Tan

2.3.2. Verification of recombinant plasmid

2.3.2.1 Restriction Endonuclease Digest

Following plasmid DNA extraction (section 2.3.1), restriction endonuclease digests were used to verify correct insert size. In the case of the pCAGGS constructs, the restriction endonucleases used for the digests were *EcoR1*, the recognition site of

which is found upstream from the multiple cloning site, and *Not1*, the recognition sites of which is found downstream from the multiple cloning site. Since *EcoRI* performs optimally in buffer H and *Not1* in buffer D, a double digest was performed to excise the inserts.

For the pXJ40 construct, the restriction endonucleases used for the digests were *XhoI*, the recognition site of which is found upstream from the multiple cloning site, and *BamH1*, the recognition sites of which is found downstream from the multiple cloning site. Since *Xho1* performs optimally in buffer D and *BamH1* performs optimally in Buffer E, a double digest was performed to excise the inserts.

The restriction enzyme reactions were set up as follow:

First digest: 16 µl sterile, deionised water, 2.0 µl 10x Buffer H or D (depending on the enzyme used), 1.0 µl acetylated BSA (10 µg/µl), 1.0µl purified plasmid; mixed by pipetting, then 1.0 µl *EcoRI* or *XhoI* (10 U/µl) was added in a final reaction volume of 21 µl. The reaction was then incubated at 37 °C for 2 hours in a water bath, following deactivation of the enzyme at 65°C for 15 minutes. The enzymatic digest products were purified using the Wizard[®] SV Gel and PCR Clean-Up System (Promega) to remove the small oligonucleotides released by the restriction digest. Briefly, an equal volume of 21 µl/mg membrane binding solution (4.5 M guanidine isothiocyanate and 0.5 M potassium acetate) was added to the reaction tubes and mixed by vortexing. The mixture was then transferred to a minicolumn assembly and incubated at room temperature for 1 min. First digest reaction mix was transferred to the minicolumn membrane by centrifugation at 14000 rpm for 1 min. Electrostatic interactions generated in the presence of chaotropic salts allow DNA to bind to the silica membrane. The DNA was washed with 700 µl and then 500 µl membrane wash solution (10 mM potassium acetate, 80% (v/v) ethanol and 16.7 µM EDTA) by

centrifugation at 14000 rpm for 1 and 5 min, respectively. A further 1 min centrifugation at 14000 rpm was necessary to ensure complete evaporation of residual ethanol. DNA was eluted in 16µl nuclease free water which neutralizes the electrostatic interactions releasing the purified DNA from the membrane. The purified DNA was transferred to a clean 1.5 ml microcentrifuge tube by centrifugation at 14000 rpm.

Second digest: 2.0 µl 10x buffer D or E (depending on the enzyme used), 1.0 µl acetylated BSA (10 µg/µl), 1.0µl isolated plasmid; solutions mixed by pipetting, then 1.0 µl *Not*I or *Bam*HI (10 U/µl) was added in a final reaction volume of 21 µl. The reaction was incubated at 37°C for 2 hours in a water bath, following deactivation of the enzyme at 65°C for 15 minutes.

2.3.2.2 Electrophoresis of the restriction enzyme digests of plasmids

The restriction digest products were diluted in a ratio of 1:6 with 6x Blue/Orange loading dye (0.4% (v/v) orange G, 0.03% (v/v) bromophenol blue, 0.03% (v/v) xylene cyanol FF, 15% (v/v) Ficoll® 400, 10 mM Tris-HCl (pH 7.5) and 50 mM EDTA (pH 8.0)) and separated on a 1% (w/v) agarose gel (containing 0.001% (v/v) ethidium bromide) by electrophoresis in TBE buffer (89 mM tris (hydroxymethyl) aminomethane, 0.089 mM boric acid, 2 mM EDTA (pH 8)) at 90 V for approximately 1 hour. The agarose gel was viewed under short-wave length UV light.

2.3.3 Plasmid purification (Midi-prep)

The midi-prep procedure is used to isolate high copy plasmid DNA from bacterial cells, using a silica-based anion-exchange resin for routine separation of different classes of nucleic acids. The manufacturer's protocol advises using 5-30 ml of overnight culture, which is grown at 37°C with vigorous shaking. However, it was

found that using a biomass (BM) index was more relevant to ensure the highest yield of plasmid DNA. Biomass was therefore calculated using the following equation:

$$\text{OD}_{600} \text{ (mg/ml)} \times \text{Vol (ml)} = \text{BM (mg)}$$

A biomass of approximately 4 mg was used for all midi-preps.

Table 2.4: Composition of all the buffers used in this experiment

Buffer	Composition
Resuspension Buffer S1:	50 mM Tris-HCl, 10 mM EDTA, 100 µg / mL RNase A, pH 8.0
Lysis Buffer S2:	• 200 mM NaOH, 1 % SDS
Neutralization Buffer S3:	• 2.8 M KAc, pH 5.1
Equilibration Buffer N2:	100 mM Tris, 15 % ethanol, 900 mM KCl, 0.15 % Triton X-100, adjusted to pH 6.3 with H ₃ PO ₄
Wash Buffer N3:	• 100 mM Tris, 15 % ethanol, 1.15 M KCl, adjusted to pH 6.3 with H ₃ PO ₄
Elution Buffer N5:	100 mM Tris, 15 % ethanol, 1 M KCl, adjusted to pH 8.5 with H ₃ PO ₄

Cultures were grown overnight (14-16 hrs) in LB broth supplemented with ampicillin at 37°C with constant shaking. The cultures were then centrifuged at 6,000 x g for 15 min at 4°C and the supernatant discarded. The pellet of bacterial cells was resuspended in 4ml of Buffer S1 + *RNase A*. 4mls of Buffer S2 was added to the suspension and the tubes were mixed by gently inverting it six–eight times. The mixture was then incubated at room temperature (18–25 °C) for 2– 3 min. Then, 4 ml

of pre-cooled Buffer S3 (4°C) was added to the suspension. Immediately thereafter, the lysates was mixed by gently inverting the flasks six–eight times until a homogeneous suspension containing an off-white flocculate was formed. The suspension was then incubated on ice for 5 min. A NucleoBond AX 100 (Midi) Column was then equilibrated with 2.5 ml of Buffer N2. The column was allowed to empty by gravity flow. The flow-through was then discarded. To clear the bacterial lysate a folded filter was placed in a funnel of appropriate size. The filter paper was wetted using a few drops of buffer N2 and the bacterial lysate was loaded onto the wet filter and the flow-through was collected. The cleared lysate was now loaded onto the NucleoBond® Column. The column was allowed to empty by gravity flow. Following this, the column was washed with 10 ml of buffer N3 and the flow-through discarded. The plasmid DNA was then eluted with 5 ml of Buffer N5. 3.5 ml of room-temperature isopropanol was then added to precipitate the eluted plasmid DNA. It was mixed carefully and centrifuged at $\geq 15,000 \times g$ for 30 min at 4 °C. The supernatant was then carefully discarded. The pellet was then washed and dried by adding 2 ml of room temperature 70% ethanol to the pellet and resuspended by briefly vortexing. It was then centrifuged at $\geq 15,000 \times g$ for 10 min at room temperature (18–25 °C). All the ethanol was carefully removed from the tube using a pipette tip. The pellet was then allowed to dry at room temperature for 5-10 mins. The DNA was finally reconstituted by dissolving the pellet in 1ml of sterile deionized H₂O under constant spinning for 10-60 min. The resuspended DNA was stored at - 20°C until further use.

2.3.4 Quantification of plasmids by means of fluorometry

The plasmid yield was quantified using the Qubit Fluorometer according to Invitrogen's protocol and application guide for the Qubit™ DNA assay.

2.3.5 Verification of plasmid constructs by means of restriction endonuclease digestion

The plasmids veracity was confirmed by means of restriction enzyme digest using the same enzymes and method as described earlier. Electrophoresis was carried out on a 1% agarose gel at 90 V for 90 minutes to confirm whether the correct plasmid constructs were present.

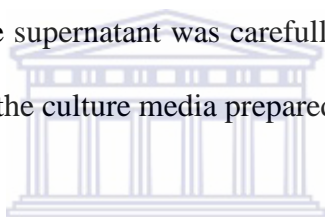
2.3.6 Nucleotide sequencing and sequence analysis

Extracted plasmids were sequenced with M13 reverse primers to ensure no base-pair mutations occurred (Inqaba Biotechnology). Results were viewed with FinchTV (www.geospiza.com) and extracted sequences were aligned using the multiple sequence alignment function in ClustalX (Version 2.0.12). Files were subsequently extracted to GeneDoc (Version 2.7.000) for viewing.

2.3.7 Mammalian Cell Culture

Cos7 cells (African Green Monkey *Cercopithecus aethiops* Fibroblast-like Kidney Cells) were cultured in Dulbecco's modified Eagle's medium (Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum (Gibco), 1% glutamine and 1% penicillin/streptomycin (Sigma-Aldrich). The cells were incubated in a 37°C incubator with 5% CO₂. Adherently growing cells attach themselves to the surface of the culture plate and each other. This is made possible by surface proteins acting as bridges to dissociate cells from the culture flasks or dishes and each other, these

protein bridges must be broken. Trypsin is a digestive enzyme, able to break the protein connections between cells and culture surfaces. Trypsinization is used for passaging cells and to obtain a suspension for further analysis. The cells were passaged every 2-3 days by means of Trypsinization as follows: the media was removed from the flask, the adherent cells were then rinsed 3 times with pre-heated PBS solution. 1-2 mls of Trypsin-Versene EDTA (5%) was added and incubated for 3-5 minutes at 37°C. The cell layer was detached by tapping on the surface of the flask. Then, 2 volumes of serum-containing media were added to the flask to stop trypsinization. The cells were dissociated by pipetting the suspension gently up and down. Next, the cells were transferred into a centrifuge tube and then centrifuged at 500 x g for 5-10 minutes. The supernatant was carefully removed using a pipette and the cells were resuspended in the culture media prepared as above.



2.3.8 Transfection of COS-7 cells

2.3.8.1 Trypan blue dye exclusion

The cells were first trypsinated to dissociate it. Next, 10µl of the cells were diluted in 90 µl of PBS (for the best results 5×10^5 cells/ml were re-suspended in PBS). 100 µl of the cell suspension was transferred to a clean 1.5 ml eppendorf tube and gently mixed in an equal volume of 0.2% trypan blue solution in PBS. It was then incubated for 3 minutes at room temperature.

2.3.8.2 Cell counting with a hemocytometer (Neubauer chamber)

The counting chamber was thoroughly cleaned and covered with a new coverslip. 10 µl Trypan blue stained cell suspension was withdrawn and loaded to the chamber by applying the suspension to the slit between the base and coverslip; the chamber fills by means of capillary action. Under the microscope, the number of viable and non-

viable (stained) cells was counted immediately in 4 large quadrants. The arithmetic mean of the 4 quadrants was calculated and the value was multiplied by 2×10^4 to obtain the concentration (cells/ml) in the unstained cell suspension prepared as above. The cell viability was calculated as: Viable cell count/ total cell count x 100.

2.3.8.3 Seeding of cells

After the cells were counted, it was seeded in a 6 well plate at a concentration of 5×10^5 cells per well. This was then incubated overnight in complete cell culture media without any antibiotics, as this could cause cell death during transfection.

2.3.8.4 Transfection of Cos 7 cells with N-, M-, E-HA and the negative control

GST-HA

When the cells were 80-90% confluent, the transfection experiments were conducted using Lipofectamine 2000 (Invitrogen). One hour before the transfection process, the media was removed from the seeded cells and 2 ml of fresh media was added (DMEM+FBS only) and incubated. Next, 4 μg of the N, M and E constructs and 1 μg of the positive control was used for transfection. The appropriate amount of the viral DNA construct or positive control was added to 100 μl of Opti-M medium (Invitrogen), respectively. In another microcentrifuge tube, 12 μl of Lipofectamine 2000 was added to 100 μl of Opti M (per construct to be transfected). Both mixtures were then incubated for 5 min at room temperature. Following incubation the construct-Opti-M mixture was added to the Lipofectamine-Opti-M mixture and incubated for 20 min at room temperature. The mixtures of DNA and lipofectamine were then added dropwise to the respective labelled wells. The media was changed 4-6 hrs post transfection with fresh complete medium and incubated for 24 hours before lysing the cells.

2.3.8.5 Cell lysis

After the 24 hour incubation period, the cells were harvested by adding ice cold PBS to the wells and scraping them off with a cell scraper. Dislodged cells were transferred to a 15 ml Greiner tube; this process was done three times in order to transfer as much of the cells from the wells to the tubes. The cells were then centrifuged at 3000 x g for 5mins and the supernatant discarded. Cell pellets were re-suspended in 1ml ice cold PBS to wash off the excess culture media and centrifuged at 3000 x g for 5 mins again and the supernatant was discarded. The cells were resuspended once again in 1ml ice cold PBS and transferred to a clean 1.5 ml Eppendorf tube. Following centrifugation, cell pellets were resuspended by vortexing in 2 ml/g lysis buffer (20 mM Tris-Cl pH 8, 200 mM NaCl, 1 mM EDTA pH 8 and 0.5% Triton X- 100) containing a protease inhibitor cocktail (Invitrogen) for mammalian cells (Table 2.4 summarises the various functions of the components of this lysis buffer). The cells were then incubated on ice for 20 mins and thereafter centrifuged at full speed to pellet the membranous part of the cell which does not contain the proteins of interest and the supernatant was collected and stored @ -70°C.

Table 2.5: Functions of components of chosen cell lysis buffer

Compound	Function
TritonX	Non-ionic detergent; improves solubility of GST fusion proteins and prevents aggregation of lysed cells.
NaCl	Provides an osmotic shock to cells
Tris	Interacts with lipopolysaccharides in the outer membrane of the cell and thereby increases permeability.
EDTA	Inhibits divalent cation-dependent proteases

2.3.9 Protein Expression Analysis

2.3.9.1 Cell Lysate Preparation

The lysates were then prepared in order to run on a SDS-PAGE gel. Equal volumes (1:1) of the lysates and Laemmli blue loading dye (4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue and 0.125 M Tris HCl) was mixed in an Eppendorf tube. The samples were then boiled at 98°C for 2 min, except for the M protein which is boiled at 50°C for 10 min.

2.3.9.2 SDS-PAGE

A 3% stacking and a 15% separation SDS-PAGE gel was used to separate proteins in using a Biorad system. Electrophoresis was done at 15 mA/gel for 80 mins in Tris-glycine SDS running buffer (25 mM Tris, 192 mM Glycine, 0.1% (w/v) SDS (pH 8.3)) The SDS-PAGE gel was then transferred to a Western Blot.

2.3.9.3 Western Blot Analysis

For Western Blots, the proteins, separated by SDS-PAGE, were transferred to a nitrocellulose or PVDF membrane, where neither membrane had any noticeable preferential properties. Prior to transfer the nitrocellulose membrane was equilibrated in 20% (v/v) methanol and the PVDF in 100% (v/v) methanol. Proteins were transferred in transfer buffer (27 mM tris(hydroxymethyl)-amino-methane, 191 mM glycine and 20% (v/v) methanol) in a submersion system at 100 V for 90 min, after which, the membrane was blocked with a 5% (w/v) milk and 0.05% (v/v) tween 20 in PBS solution for 30 min on a rocker. The membrane was then incubated at 4°C overnight on a roller in 3% (w/v) milk and 0.05% (v/v) Tween 20 in PBS solution with the primary antibody, rabbit α HA, in a dilution of 1:250. The membrane was then washed in a wash solution (0.05% (v/v) Tween 20 in PBS) for 1 hour. The secondary antibody, peroxidase-labeled goat α rabbit, was added in a dilution of 1:1000 in fresh solution previously described and incubated at room temperature for 1 hour on a roller, following which, the membrane was washed in wash solution with subsequent addition of the peroxidase substrate. The presence of the GST-fusion protein was determined by colorimetric analysis using 3,3',5,5' Tetramethylbenzidine (TMB) liquid substrate system for membranes (Sigma-Aldrich).

Chapter 3

Results and Discussion



3. Results and Discussion

3.1. Bioinformatic Analysis of HCoV-NL63 Nucleocapsid, Matrix and Envelope Proteins

The advancement in genome sequencing technology has brought about a sharp rise in genome databases and the rate at which these genomes are sequenced, by far exceeds our ability to determine the functions of the proteins which they encode. With the help of *in silico* analysis, researchers are able to predict certain functions of a protein based on the biochemical or physiochemical properties of their amino acid sequence. However, these predictions should be verified experimentally.

3.1.1 Multiple sequence alignment

Clustal X was used to compare the amino acid sequences of the N, M and E proteins of HCoV-NL63 and the other human coronaviruses (Table 3.1). The selected HCoV-NL63 structural protein sequences (N, M and E) were shown to be most similar to HCoV-229E N, M and E (44%-63% a.a. identity; 63%-77% a.a. similarity). This is not an unforeseen result, since HCoV-NL63 and HCoV-229E have evolved from a most recent common ancestor in the 11th century (Pyrce, Dijkman et al. 2006). The high a.a. sequence identities and similarities calculated for the structural proteins, provide evidence that their structure and function have been well conserved through evolution.

In Figure 3.1, a comparative analysis of the 5 human coronavirus nucleocapsids, identified a conserved 8 amino acid region between amino acids 128-135, this result is in contrast to results obtained by Xing et al (2007), that had identified the region of amino acids 78-85 within N protein probably was the conserved region for all coronaviruses identified so far including HCoV-NL63. Also, multiple small conserved regions of 4-5 a.a.'s were observed for the M homologues (Figure 3.2). Although the

functions of these conserved regions have not yet been clearly defined, the conserved a.a. region in the HCoV-NL63 N protein (FY YLGTGP) may be involved in RNA binding (Chang, Hsu et al. 2009; Berry 2012) (Chang et al., 2009). Berry et al had seen that at these long stretches of conserved amino acid regions found within the N protein are in fact regions of disorder and according to Chang et al. all disordered regions found within the SARS CoV N protein are also capable of binding to RNA, although the RNA binding domains in the N protein have been previously linked to the N-terminal domain (NTD) and the C-terminal domain (CTD) respectively (Chang, Hsu et al. 2009) .

In contrast to the N and M proteins, the E protein (Figure 3.3) does not show as much sequence identity with some of the viruses in this study group although the percentage similarity with other coronavirus species is high (Table 3.1). When the percentage sequence homology is high, this could be an indication that these viruses share the same topology of the E-protein than others and seeing that the highest percentage homology is with 229E, which forms part of the same group as NL63, this brings about an indication that coronaviruses from the same groups could potentially share the same E protein topology?

This is also an indication that E proteins characteristics are not conserved amongst nidoviruses. This makes sense as different topologies for the E protein has been observed, thus bringing about different functions for the E protein. Potential roles for E in assembly have been previously observed, such as helping to bend membranes or playing a role in membrane scission by its hairpin formation topology (Ruch and Machamer 2012). The efficient trafficking of virions through secretory pathway could be due to a function that may be related to its transmembrane domains forming ion

channels (DeDiego, Alvarez et al. 2007; Ruch and Machamer 2012; Ruch and Machamer 2012).



Table 3.1: Multiple sequence alignment. Comparison of HCoV-NL63 N, M and E amino acid sequences to homologues from selected coronavirus isolates. Identity values (%) are shown in **BOLD** and similarity values (%) are shown in *ITALICS*.

% Amino Acid Homology				
HCoV-NL63 structural proteins	HCoV-229E	HCoV-OC43	HKU1	SARS
N	44	36	33	31
	<i>63</i>	<i>55</i>	<i>49</i>	<i>46</i>
M	63	36	33	32
	<i>77</i>	<i>60</i>	<i>58</i>	<i>55</i>
E	47	44	29	18
	<i>72</i>	<i>48</i>	<i>56</i>	<i>48</i>

To identify homologues, the amino acid sequences of HCoV-NL63 N, M and E was compared to sequences in the GenBank database at the National Centre for Biotechnology by using the Basic Blast Search Server (Altschul et al., 1990). Subsequently, HCoV-NL63 N, M and E was aligned with selected coronavirus N, M and E proteins with CLUSTAL X version 2 (Larkin et al., 2007). The sequences aligned, with Genbank accession numbers in bold, were: HCoV-NL63 N (YP_003771.1), M (YP_003770.1), E (YP_003769.1); HCoV-229E N (NP_073556.1), M (NP_073555.1), E (NP_073554.1); HCoV-OC43 N (NP_937954.1), M (NP_937953.1), E (NP_937952.1); HKU1 N (YP_173242.1), M (YP_173241.1), E (YP_173240.1); SARS-CoV N (AFR58751.1), M (AF58746.1) and E (AFR58745.1).

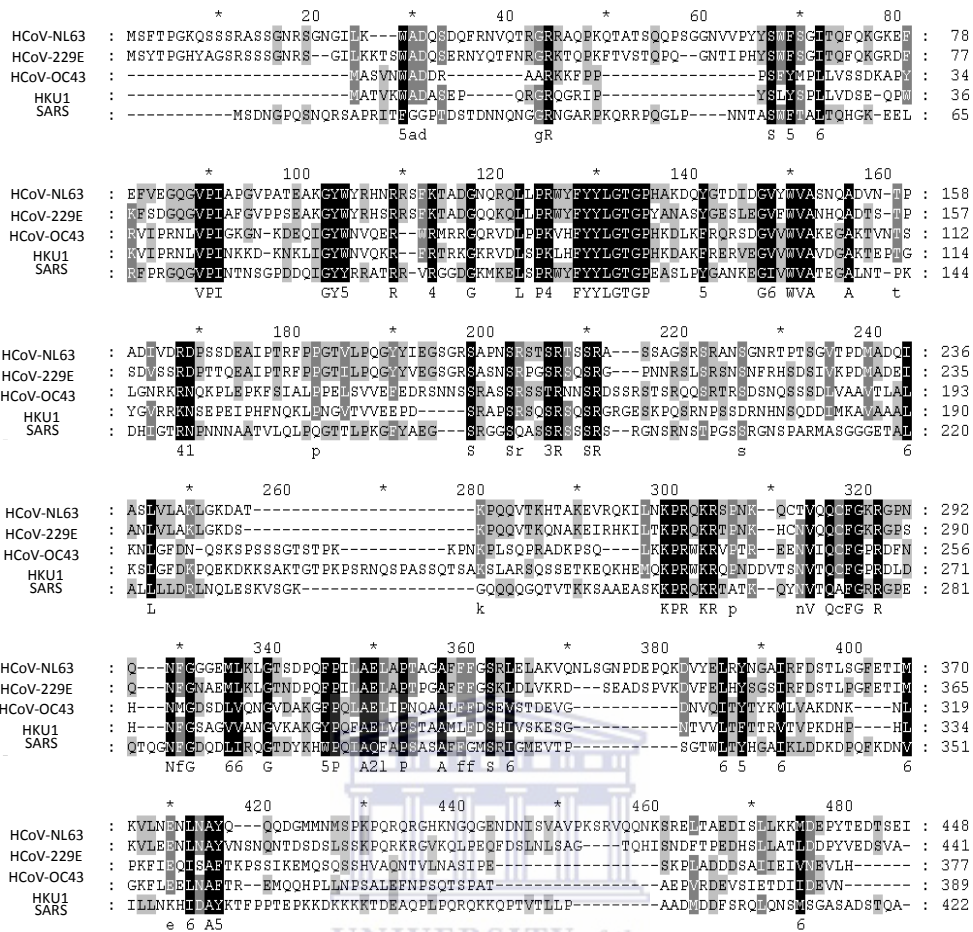


Figure 3.1: Multiple sequence alignment of various human coronavirus nucleocapsid (N) proteins. Selected HCoV-NL63 N-homologues were aligned with CLUSTAL X version 2.0 (Larkin et al. 2007) and viewed with GENEDEC version 2.6.002 software (Nicholas et al; 1997). Shading indicates conserved regions and gaps were introduced to align sequences. Conserved region larger than 5 a.a.'s identified: (i) *FYLLGTGP*. N amino acid sequences were obtained from NCBI (accession numbers in brackets): HCoV-NL63 (YP_003771.1), HCoV-229E (NP_073556.1), HCoV-OC43 (NP_937954.1), HKU1 (YP_173242.1) and SARS-CoV (AFR58751.1).

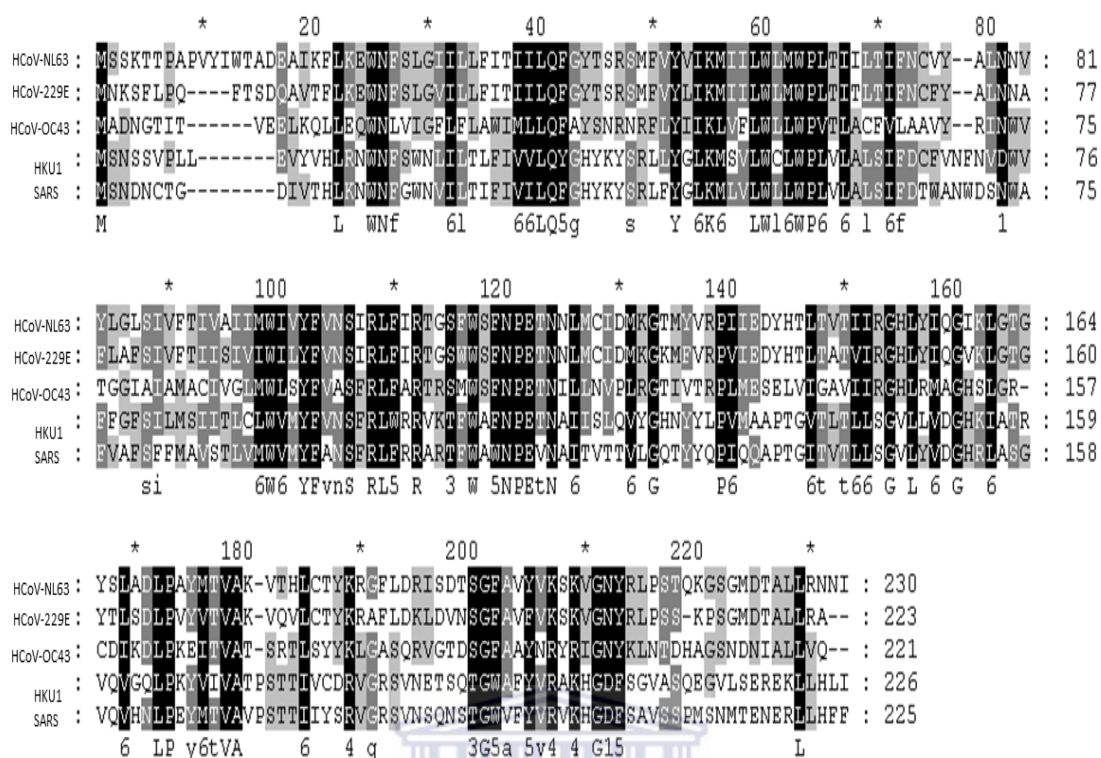
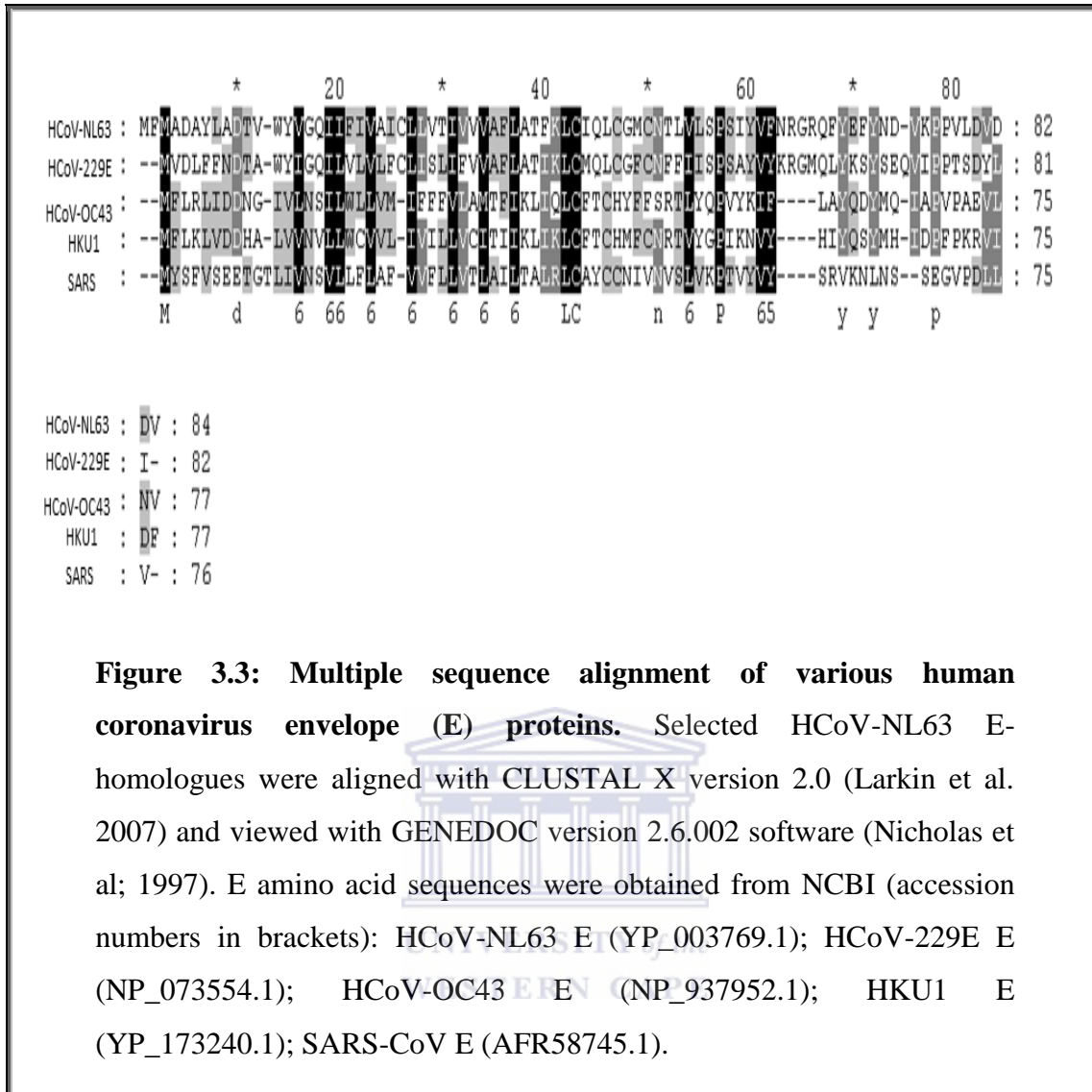


Figure 3.2: Multiple sequence alignment of various human coronavirus membrane (M) proteins. Selected HCoV-NL63 M-homologues were aligned with CLUSTAL X version 2.0 (Larkin et al. 2007) and viewed with GENEDOC version 2.6.002 software (Nicholas et al; 1997). Multiple Conserved regions identified: (i)(ii)(iii)(iv)(v).M amino acid sequences were obtained from NCBI (accession numbers in brackets): HCoV-NL63 M (YP_003770.1), HCoV-229E M (NP_073555.1), HCoV-OC43 M (NP_937953.1), HKU1 M (YP_173241.1), SARS-CoV M (AFR58746.1).



3.1.2 Transmembrane Domains and Hydropathy Plots of HCoV-NL63 N, M and E proteins

Proteins are amino acids that are held together by peptide bonds. Each amino acid has a specific “R” group that is unique to it. These groups determine whether proteins are hydrophilic (water-loving) or hydrophobic (water-fearing). The hydrophobicity of an amino acid would determine where the amino acid would be located in the final structure of the protein (Kyte and Doolittle., 1982). In globular proteins the hydrophobic R group will be located on the inside of the protein, thus causing it not to

be in contact with water in the cytosol. The hydrophilic R groups are located on the outside of the proteins, thereby interacting with the water in the cytosol. This information can give us an indication on the structure of the proteins. One of the basic tenants of biology is that the structure of a protein defines its function. So, being able to make predictions about the structure of the HCoV-NL63 structural proteins, could enable us to find out more about how these proteins function, which could aid in understanding the virus better. A hydropathy plot predicts potential transmembrane or surface regions in proteins (Kyte and Doolittle., 1982).

The Kyte-Doolittle plot works as follows: each amino acid is given a hydrophobicity score (Appendix 6) between 4.6 and -4.6. A score of 4.6 is the most hydrophobic and -4.6 are the most hydrophilic. A window size needs to be determined in order for these scores to show its effect; this window size is determined by the number of amino acids whose hydrophobicity scores will be averaged and assigned to the first amino acid in the window. Then, the computer program moves down one amino acid and calculates the average of all the hydrophobicity scores in that window, thereafter the computer program moves down another amino acid and in doing this the average of all the hydrophobicity scores is calculated in the second window. This pattern will continue until the entire amino acid sequence is screened and the average score for each window is assigned to the first amino acid in that window. The default window size in this bioinformatics tool is 9 amino acids. The averages are then plotted on a graph, with the y-axis

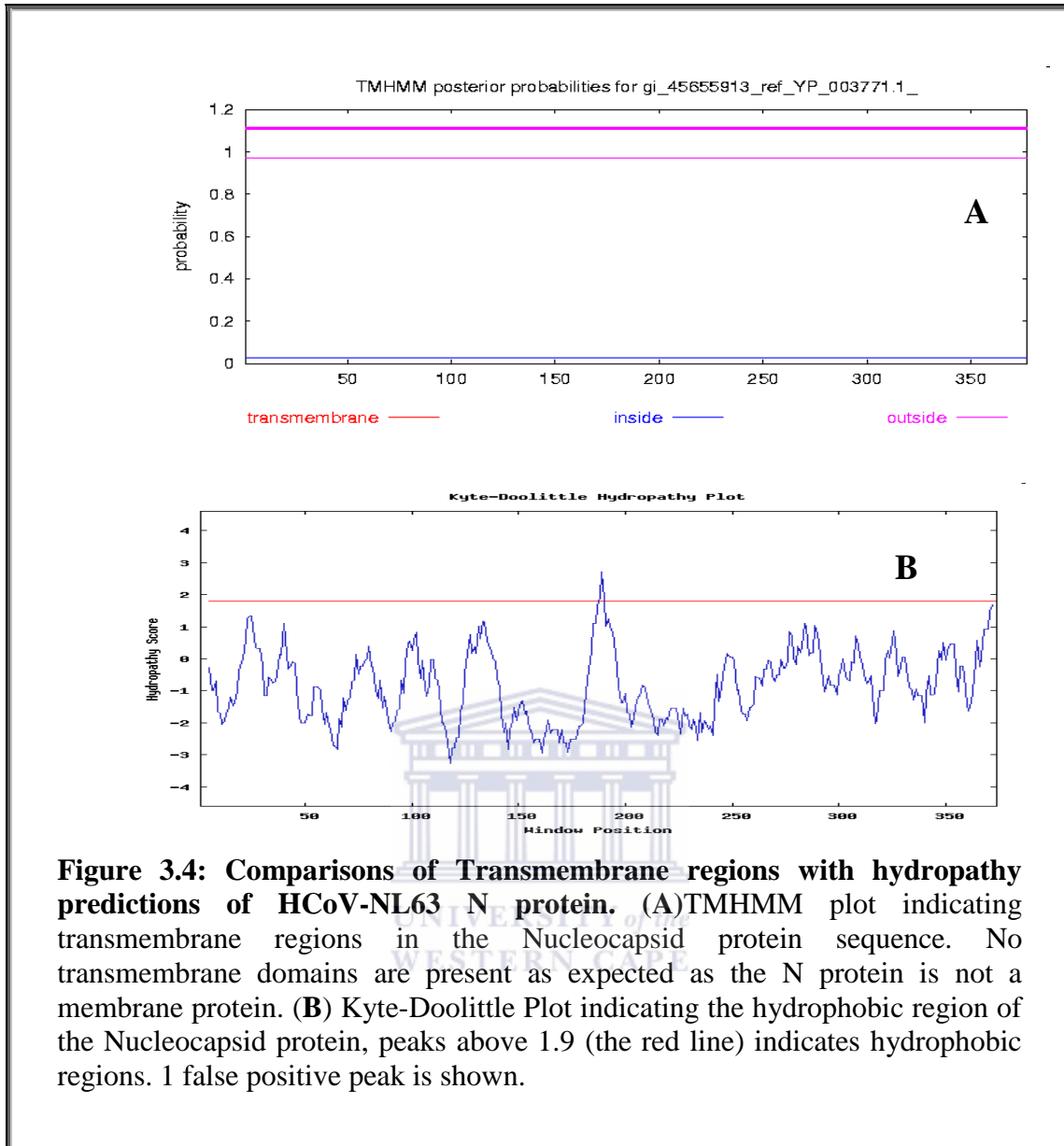


Figure 3.4: Comparisons of Transmembrane regions with hydropathy predictions of HCoV-NL63 N protein. (A) TMHMM plot indicating transmembrane regions in the Nucleocapsid protein sequence. No transmembrane domains are present as expected as the N protein is not a membrane protein. (B) Kyte-Doolittle Plot indicating the hydrophobic region of the Nucleocapsid protein, peaks above 1.9 (the red line) indicates hydrophobic regions. 1 false positive peak is shown.

For HCoV-NL63 N, the hydropathy plot (Figure 3.4. A), indicated that the protein is strongly hydrophilic and not found within the lipid bilayer of the membrane; most of the peaks shown were below 1.6 score. The results obtained for the N protein was used as a “negative control”, when compared to the M and E proteins of HCoV-NL63. The one peak seen on Figure 3.4 (A) between amino acid numbers 150-200, was indicative of a false positive result. TMHMM server V.2.0 (Krogh et al., 2001) was also used to search for transmembrane regions in the N protein sequence (Figure 3.4

B). As expected, since N is not an integral viral protein, no such regions were predicted.

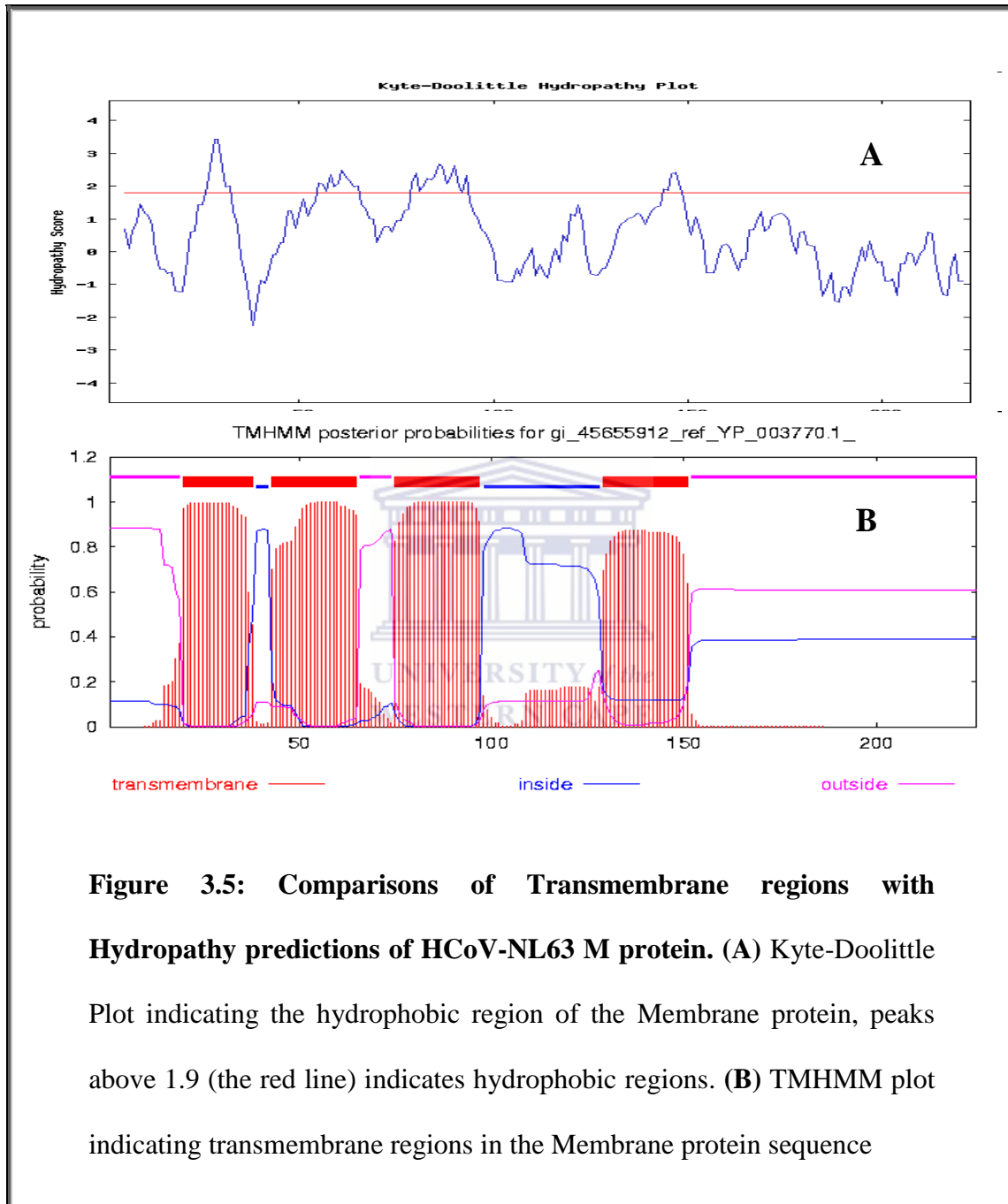
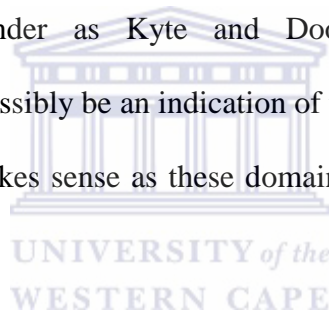


Figure 3.5: Comparisons of Transmembrane regions with Hydropathy predictions of HCoV-NL63 M protein. (A) Kyte-Doolittle Plot indicating the hydrophobic region of the Membrane protein, peaks above 1.9 (the red line) indicates hydrophobic regions. **(B)** TMHMM plot indicating transmembrane regions in the Membrane protein sequence

For the HCoV-NL63 M, four predicted transmembrane regions were seen on the TMHMM plot (Figure 3.5 B), found at a.a. positions 20-38, 43-65, 75-97 and 129-.

These results were confirmed by the hydropathy plot results (Figure 3.5 A) as it corresponded to the hydrophobic peaks on the plot. According to Kyte and Doolittle (1982) these hydrophobic peaks could also indicate transmembrane regions. These results from our analysis corresponded to an earlier report (M. Muller,2009), but contradicted other reports that predicted three transmembrane spanning regions for the HCoV-NL63 M (Rottier, Brandenburg et al. 1984; Rottier, Brandenburg et al. 1984; Masters 2006; Ye and Hogue 2007; Ruch and Machamer 2012). These differences could be due to different Algorithms used in these bioinformatics tools. When I changed the algorithms the amount of transmembrane domains changed aswell, however the Kyte-Doolittle plot confirms the results I had obtained with the Transmembrane domain finder as Kyte and Doolittle (1982) explains how hydrophobic regions could possibly be an indication of transmembrane domains (Kyte and Doolittle 1982). This makes sense as these domains are not exposed to water in the cytosol.



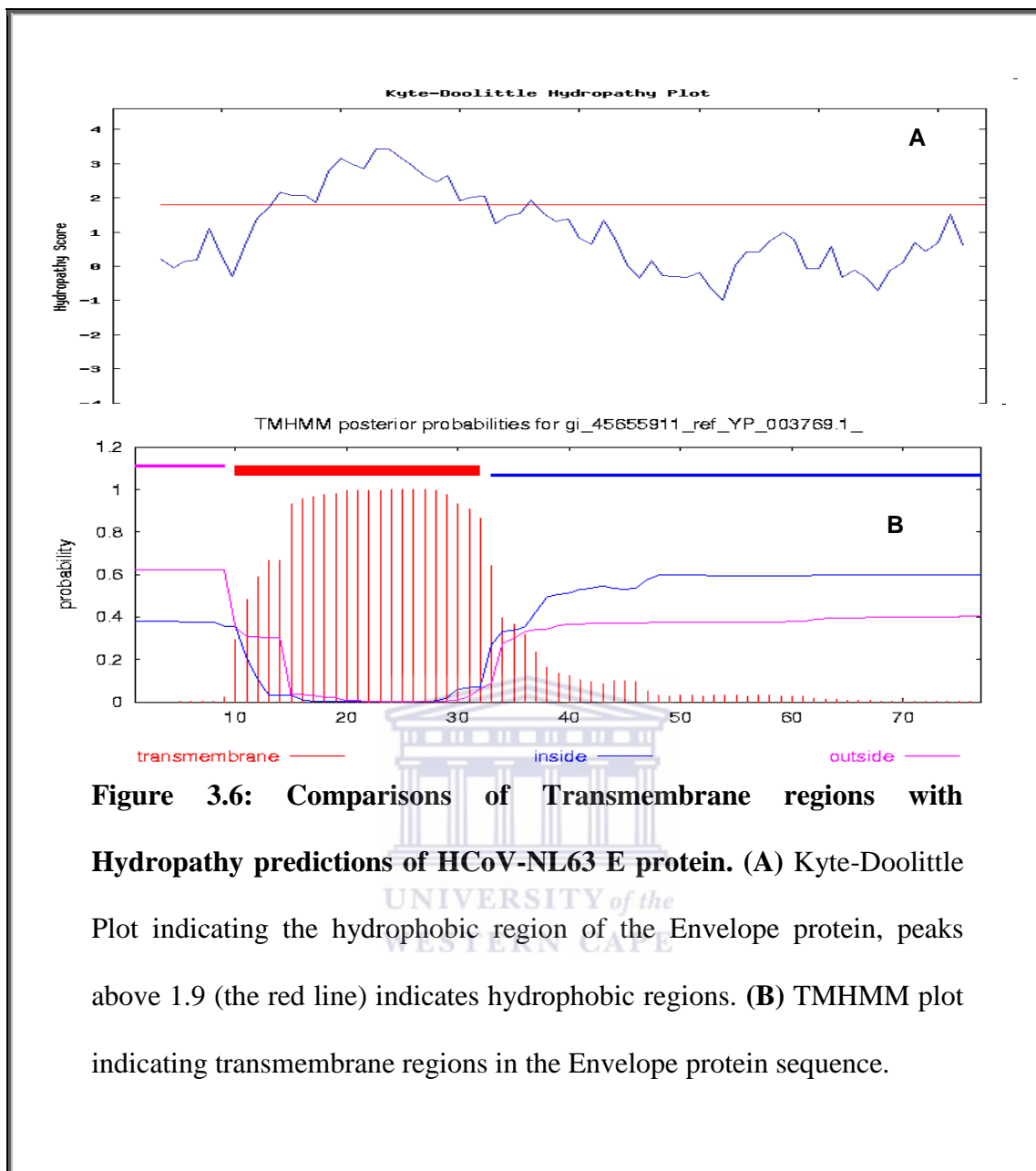
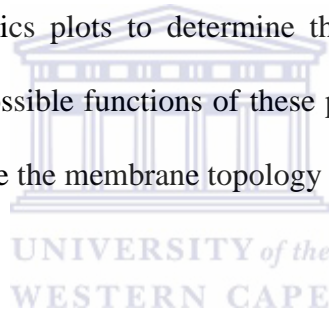


Figure 3.6: Comparisons of Transmembrane regions with Hydropathy predictions of HCoV-NL63 E protein. (A) Kyte-Doolittle Plot indicating the hydrophobic region of the Envelope protein, peaks above 1.9 (the red line) indicates hydrophobic regions. **(B)** TMHMM plot indicating transmembrane regions in the Envelope protein sequence.

In Figure 3.6A, one very large predicted transmembrane region was seen on the plot spanning a.a. position 10-32. This result was confirmed by the hydropathy plot (Figure 3.6B) which also predicted a strong positive peak at position 10-35. This characteristic formation of the E protein is one of the topologies seen in coronavirus E proteins. It is proposed that when the E protein has this topology, it functions as an ion channel (Torres, Parthasarathy et al. 2006; Yuan, Liao et al. 2006; Torres, Maheswari et al. 2007; Ye and Hogue 2007). However, due to the E protein only

having a single hydrophobic domain, it would require oligomerization in order to form an ion channel (Ruch and Machamer 2012). Many other studies suggest an important role for the ion channel activity of the E protein (Wilson, McKinlay et al. 2004; Liao, Yuan et al. 2006; Torres, Parthasarathy et al. 2006; Wilson, Gage et al. 2006; Torres, Maheswari et al. 2007); however the role of this characteristic of the E protein in infection is not entirely clear. Literature even suggests that this putative ion channel activity can be linked with virus replication and release (Ruch and Machamer 2012). The M and E proteins of coronaviruses play a vital role in viral assembly and some literature even indicate that only M and E is needed for viral assembly (Vennema, Godeke et al. 1996; Ho, Lin et al. 2004; Yuan, Liao et al. 2006). Therefore, using bioinformatics plots to determine the structure of these proteins, could aid in predicting the possible functions of these proteins in the virus life cycle. Figures 3.7 and 3.8 summarise the membrane topology of M and E as predicted by the *in silico* analysis.



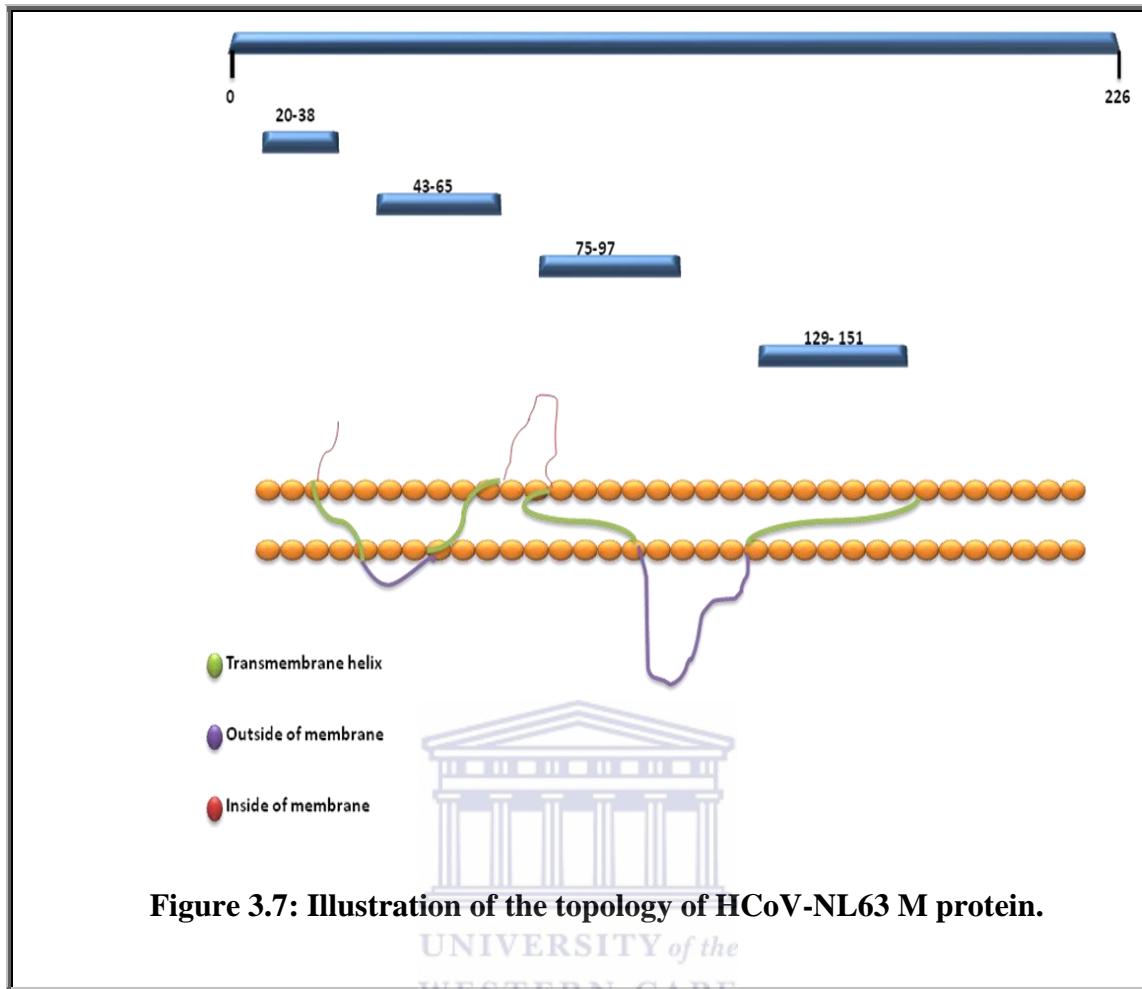


Figure 3.7: Illustration of the topology of HCoV-NL63 M protein.

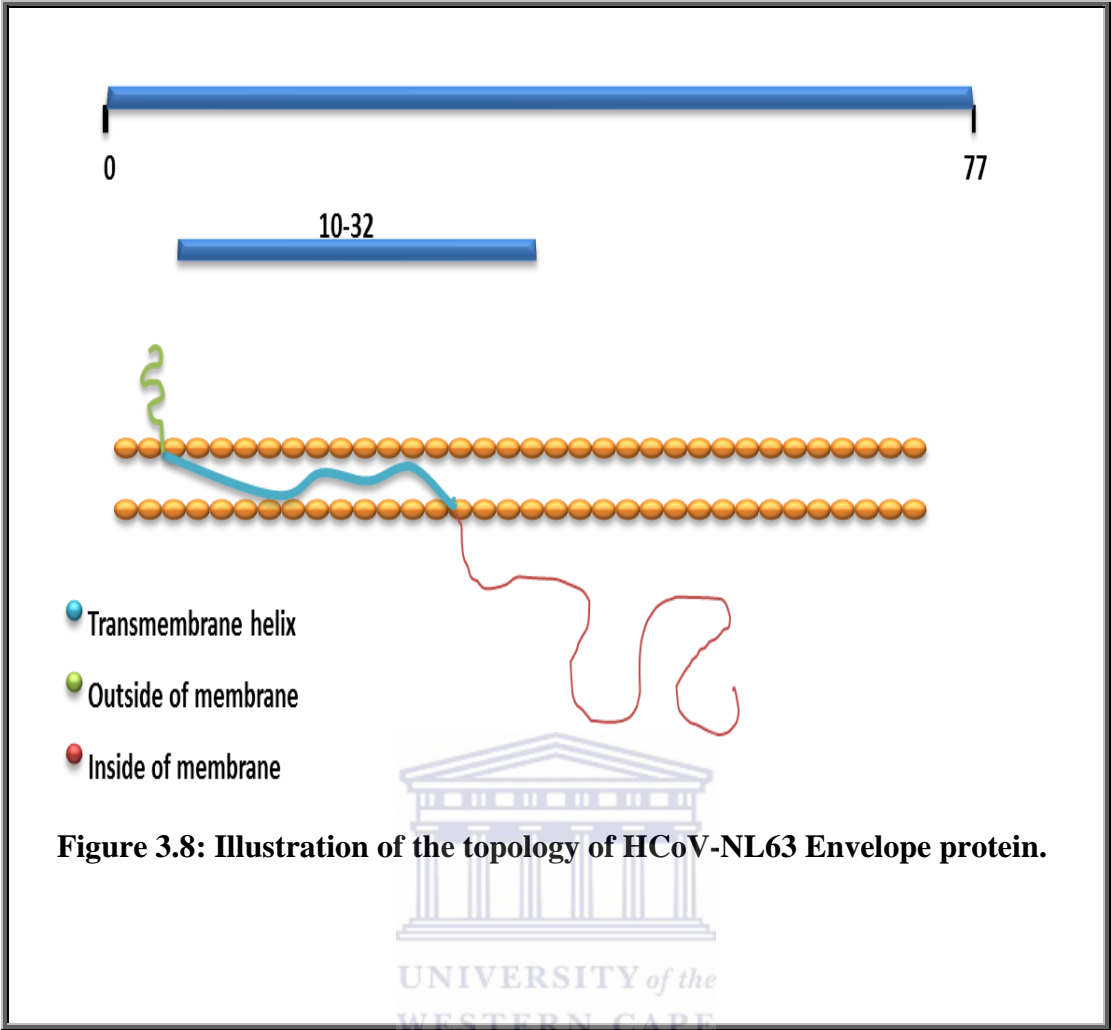


Figure 3.8: Illustration of the topology of HCoV-NL63 Envelope protein.

Table 3.2: Comparison of various motifs of HCoV-NL63 Proteins N, M and E

Viral Proteins	Nucleocapsid	Membrane	Envelope
No. Amino Acids [size in kDa]	377 aa	226 aa	77 aa
No. Transmembrane domains (position)		4	1
No. N-linked glycosylation sites (position)	7 (110-113)(145-148)(146-149)(158-161)(179-182)(200-203)(352-355)	3 (3-6)(19-22)(188-191)	
No. Myristoylation sites (position)		2 (46-51)(139-144)	1 (10-15)
No. Protein kinase C phosphorylation sites (position)	9 (29-31)(147-149)(151-153)(155-157)(163-165)(167-169)(212-214)(307-309)(334-336)	2 (99-101)(217-219)	
No. Casein kinase II phosphorylation sites (position)	7 (136-139)(174-177)(181-184)(241-244)(294-297)(334-337)(354-357)	3 (63-66)(186-189)(217-220)	

Almost all proteins analyzed to date are post-translational modified. The function of a modified protein is often strongly affected by these modifications and therefore, increased knowledge about the potential post-translational modifications of a target protein may increase our understanding of the molecular processes in which the protein plays part (SIB 2012). By using bioinformatic tools, researchers are able to identify domains and motifs found in protein sequences, which can then be used to predict the structure and possible functions of the proteins. Domains are defined as specific combinations of secondary structures that are organized into a characteristic three-dimensional structure or fold. Motifs on the other hand, are short sequences (≤ 20 amino acids) of biological interest (SIB 2012). In Table 3.2, various putative domains and motifs were identified in the HCoV-NL63 N, M and E sequences; some of these are conserved amongst nidoviruses and others are specific to HCoV-NL63. For N, a serine rich domain was predicted at position (144-211). Coronavirus N proteins have a high (7 to 11%) serine content forming a serine rich domain; serines are potential targets for phosphorylation (Peng, Lee et al. 2008). Between the functional C-terminal domain and the N-terminal domain, (CTD and NTD), is a segment containing several arginine/ serine (RS) dipeptides that is structurally flexible. It is proposed that this RS-rich motif is able to affect a multitude of cellular signalling processes, an attribute of cellular precursor mRNA (pre-mRNA) splicing factors known as SR proteins (Graveley 2000). The RS domain is phosphorylated with dynamism by several SR protein specific kinases, such as those belonging to the SR protein kinase (SRPK) and Clk families (Stojdl and Bell 1999; Peng, Lee et al. 2008). It has been shown that the phosphorylation of the RS domain modulates the activity, protein–protein interactions and subcellular localization of SR proteins (Graveley 2000). Coronavirus N proteins are phosphorylated in host cells transiently

expressing N, as well as virions (Wootton, Rowland et al. 2002; Chen, Gill et al. 2005; White, Yi et al. 2007). Phosphorylation affects the RNA binding specificity and the nucleocytoplasmic shuttling of the N proteins (Chen, Gill et al. 2005; Surjit, Kumar et al. 2005).

In our analysis, the N protein was also found to be extensively N- glycosylated on the N terminal. Glycosylation is described as the attachment of a glycan group to a residue of the protein (SIB 2012), thus providing it with a hydrophilic cover on its outer surface. This is evident as the N protein shows strong hydrophilic tendencies when viewed on the Hydropathy plot (Fig 3.4). A previous study had indicated that glycosylation of coronaviral proteins is not required for virus assembly, however the oligosaccharides could potentially be involved in the virus-host interaction (de Haan, de Wit et al. 2003). A study done on HIV-1 had shown that N-linked glycosylation of viral proteins provided new insight into the mechanisms that HIV-1 uses to escape protective immune responses (Huang, Chou et al. 2008). The HCoV-NL63 N protein is N-linked glycosylated at positions (110-113), (145-148), (146-149), (158-161), (179-182), (200-203) and (352-355).

Various putative protein kinase C phosphorylation sites were predicted for N at positions (29-31), (147-149), (151-153), (155-157), (163-165), (167-169), (212-214), (307-309), (334-336) and casein kinase II phosphorylation at positions (136-139), (174-177), (181-184), (241-244), (294-297), (334-337) and (354-357). Phosphorylation is one of the most common post-translation modifications that play a significant regulatory role in modulating protein functions. The coronavirus N protein is highly phosphorylated and it is speculated that this modification has an important role in the regulation of coronavirus N protein (White, Yi et al. 2007). A study done by Wisniewski et al (1999) had indicated that phosphorylation of proteins by casein

kinase II alters their conformation, stability, and DNA binding specificity (Wisniewski, Szewczuk et al. 1999) .

The M protein of HCoV-NL63 is 226 aa in length. In this study, M was shown to be N-glycosylated at amino acids (3-6), (19-22) and (188-191). It also contained putative protein kinase C phosphorylation sites at positions (99-101) and (217-219), as well as casein kinase II phosphorylation sites at amino acid positions (63-66), (186-189) and (217-220). The M and E proteins were predicted to be myristoylated at positions (10-15) for E and (46-51) and (139-144) for M, respectively. The N protein, however, was predicted not to be myristoylated, which suggested that this is be unique to intergral membrane proteins. These myristoylation sites were present more or less at the same position of the transmembrane domains. Myristoylation is the covalent binding of a lipid moiety to the protein, *i.e.* a myristate group attached through an amide bond to the N-terminal glycine residue of the mature form of a protein or to an internal lysine residue (Maurer-Stroh and Eisenhaber 2004; SIB 2012). The myristate can also be attached through a thio-ester bond to an internal cysteine (Maurer-Stroh and Eisenhaber 2004; SIB 2012).

3.2. Molecular Cloning of HCoV-NL63 N Protein for Expression in Bacterial Cells

3.2.1. PCR

The gene of interest (N-gene) was amplified from the 1st copy DNA with gene specific primers for unidirectional cloning into the vectors. Primers were designed to append the *SgfI* and *PmeI* (pFLEXI) restriction enzyme sites to the amplification products. These pFlexi restriction endonucleases infrequently causes restriction of human and other organisms cDNA (98%). The *SgfI* restriction site is upstream of the promoter of the protein coding region. This enables the expression of N-terminally tagged and untagged proteins. The *PmeI* site contains the terminator sequence for the protein coding region, and adds a Valine residue to the carboxy terminus. These enzyme cut sites prevent frameshifts from occurring by maintaining the reading frame and orientation of the inserts. The amplified N protein of SARS-CoV and HCoV-NL63 was separated on a 1% Agarose gel (Fig.3.9) by electrophoresis. The amplification product for HCoV-NL63 N gene was shown to be ± 1100 bp and SARS-CoV N was indicated to be ± 1200 bp in size this corresponds to the gene sizes, determined by the accession number from NCBI, for SARS-CoV N: 1269bp and HCoV-NL63 N: 1134bp.

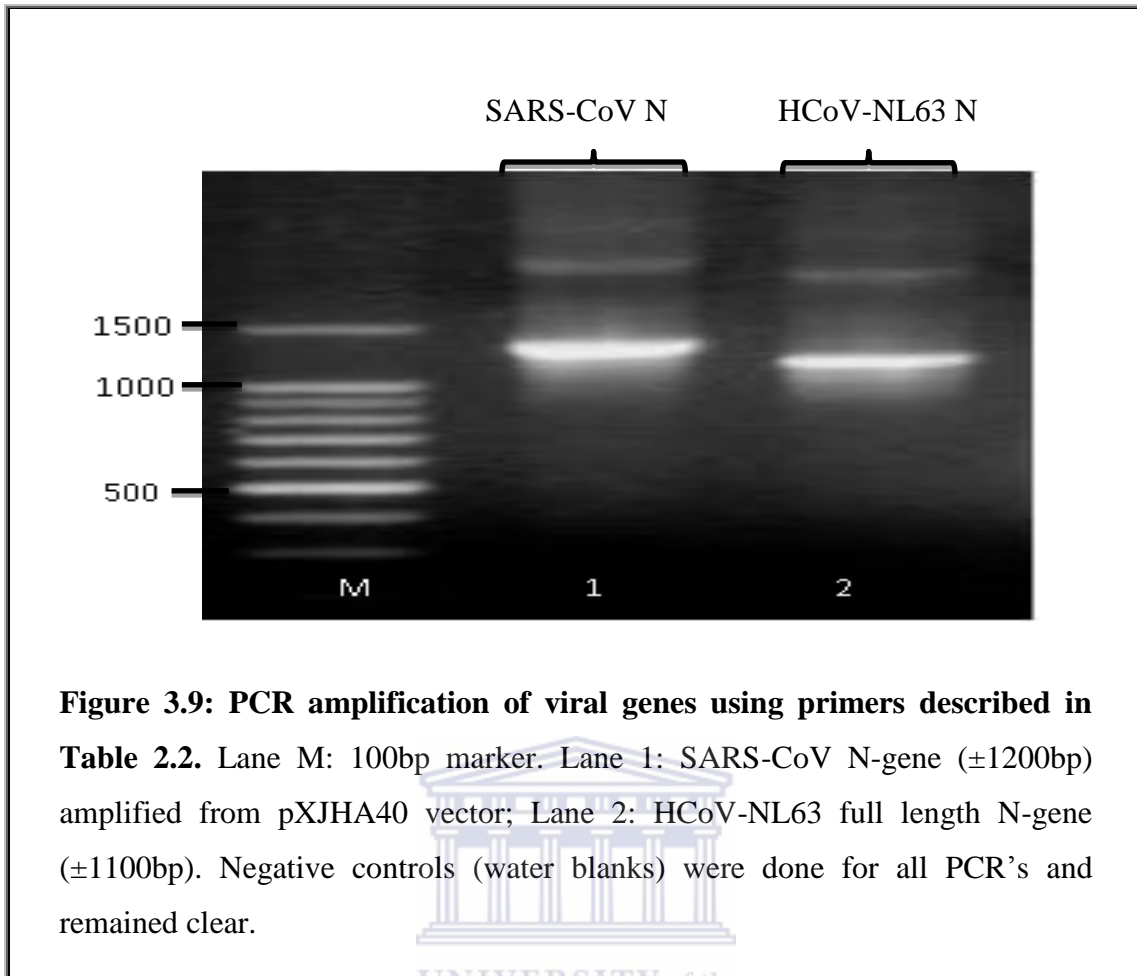


Figure 3.9: PCR amplification of viral genes using primers described in Table 2.2. Lane M: 100bp marker. Lane 1: SARS-CoV N-gene (± 1200 bp) amplified from pXJHA40 vector; Lane 2: HCoV-NL63 full length N-gene (± 1100 bp). Negative controls (water blanks) were done for all PCR's and remained clear.

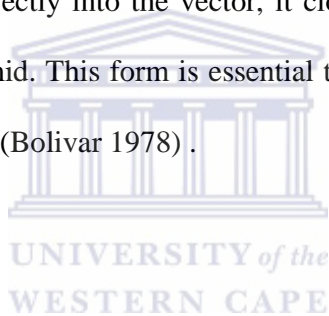
Once the N gene amplification products were identified the bands representing the target genes were excised from the gel for purification with a PCR Gel Clean-Up kit (Promega). The viral genes were then ligated into the pGEM vector and used to transform JM109 competent *E.coli*.

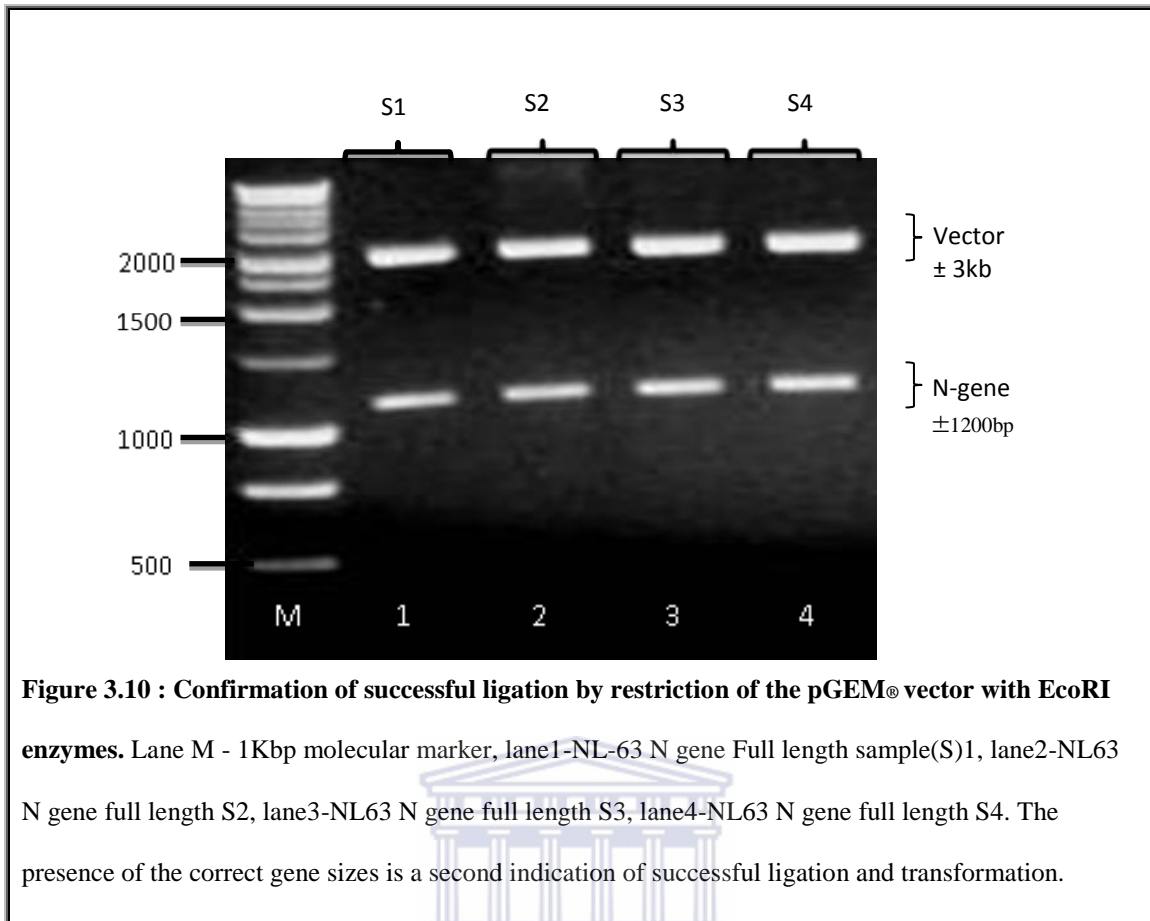
3.2.2. *EcoRI* Digest of pGEM-N

As a second confirmatory step, prior to sequencing, the insert was removed from the vector with the use of *EcoRI* restriction enzyme (Figure 3.10). *EcoRI* recognition sites are located up- and down-stream of the multiple cloning site, allowing for easy removal of the insert. The electrophoresis of the restriction digest of pGEM® vector had confirmed

that the ligation was successful as indicated by the bands of interest present in Fig 3.10 HCoV-NL63 N gene is relatively large in size (± 1100 bp), thus making it quite hard to successfully ligate. The size of the N gene obtained in Fig 3.10 was approximately 1200bp in length for HCoV-NL63 which corresponds to the expected sizes obtained from NCBI (1134bp). The sequencing results (Appendix 1 & 2) had confirmed this successful ligation as the nucleotide and deduced amino acid sequences obtained were compared against the entries in the Genbank database with respectively the BLAST-N and BLAST-P programmes (Altschul, Madden et al. 1997), these bioinformatics programmes are available on the National Centre for Biotechnology Information web page (<http://www.ncbi.nlm.nih.gov/BLAST>).

When the insert is ligated correctly into the vector, it closes the loop of the vector and changes it into a circular plasmid. This form is essential to the introduction of DNA into cells in the transformation step (Bolivar 1978).





3.2.3 SgfI and PmeI digest of pGEM-N recombinants

The plasmid DNA was enzyme digested with *SgfI* and *PmeI* to release the gene of interest (Fig 3.11) from the pGEM vector in order for it to be gel purified using the Wizard SV gel and PCR purification kit (Promega), it was then subsequently ligated into the compatible sites of the pFlexi vector for expression, generating pFlexi-N. These recombinant constructs (pFLEXI-N-GST) was then transformed into KRX *E.coli* cells and plated on LB agar plates containing ampicillin.

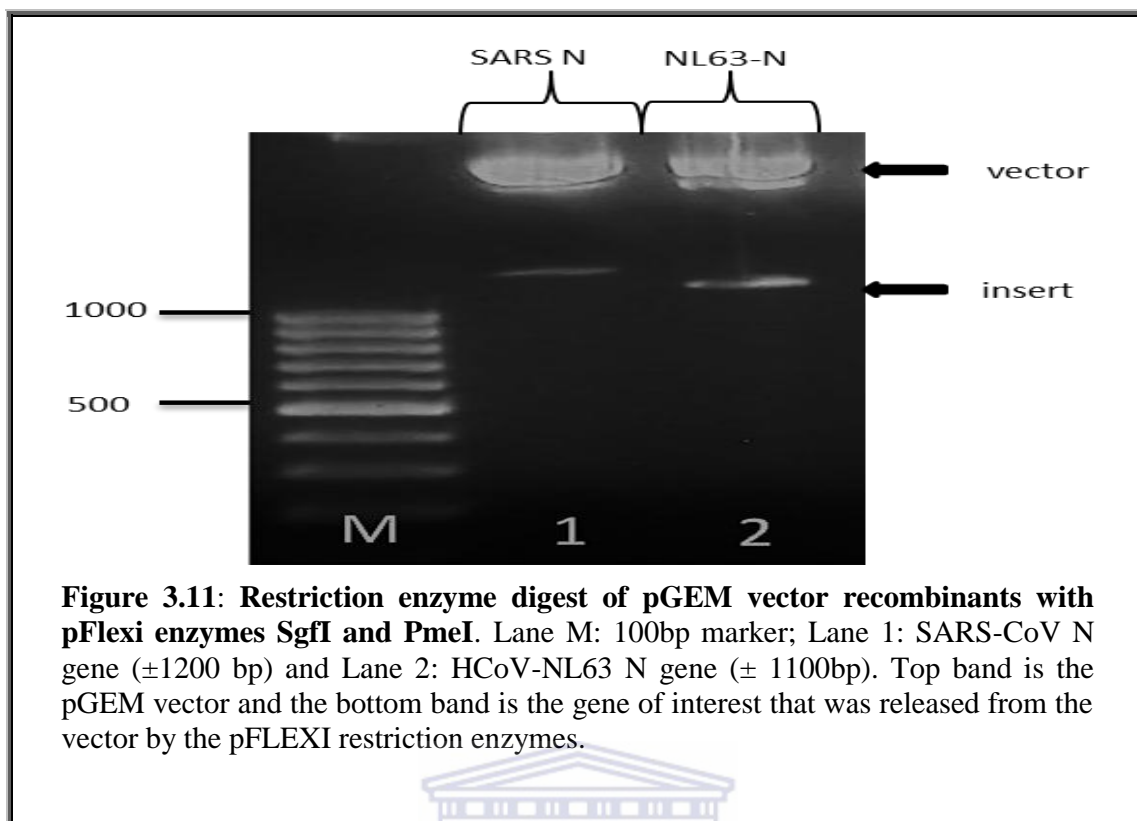


Figure 3.11: Restriction enzyme digest of pGEM vector recombinants with pFlexi enzymes SgfI and PmeI. Lane M: 100bp marker; Lane 1: SARS-CoV N gene (± 1200 bp) and Lane 2: HCoV-NL63 N gene (± 1100 bp). Top band is the pGEM vector and the bottom band is the gene of interest that was released from the vector by the pFLEXI restriction enzymes.

3.2.4. Colony PCR of KRX Competent *E. coli*

All the transformed colonies formed should be an indication of successful ligation of the target gene into the pFLEXI vector. This vector contains a lethal barnase gene which would have been replaced by the N gene preventing its activation, which would have resulted in no colonies being produced. The pFLEXI vector was especially designed to express amino terminal GST-fusion proteins in bacteria. The GST tag can be used for detection and purification of expressed proteins. Cleavage of the GST tag from the expressed proteins is also made possible by the action of TEV proteases. In order for these proteins to be expressed, this vector needs a competent *E.coli* strain that expresses T7 RNA polymerase; therefore the KRX strain was selected.

The successful cloning of the exact protein coding region into the vector (pFLEXI) was verified by means of colony selection and colony PCR subsequent to

transformation into KRX competent cells. Figure 3.12 indicates the successful transformation of the nucleocapsid genes into KRX competent cells. The bands produced in both Fig 3.12 A and B is a clear indication that the genes of interest being HCoV-NL63 N and SARS-CoV N were successfully cloned for the expression of its respective proteins in a bacterial cell system.

Sequence-verified constructs had been used for expression with the KRX cells. Even though the process of generating the constructs for expression in a bacterial system formed part of this thesis, expression of the proteins were performed as part of a larger project and unfortunately did not form part of the work reported here. However, the work reported in this section was included in a publication by Berry and colleagues (Berry 2012).



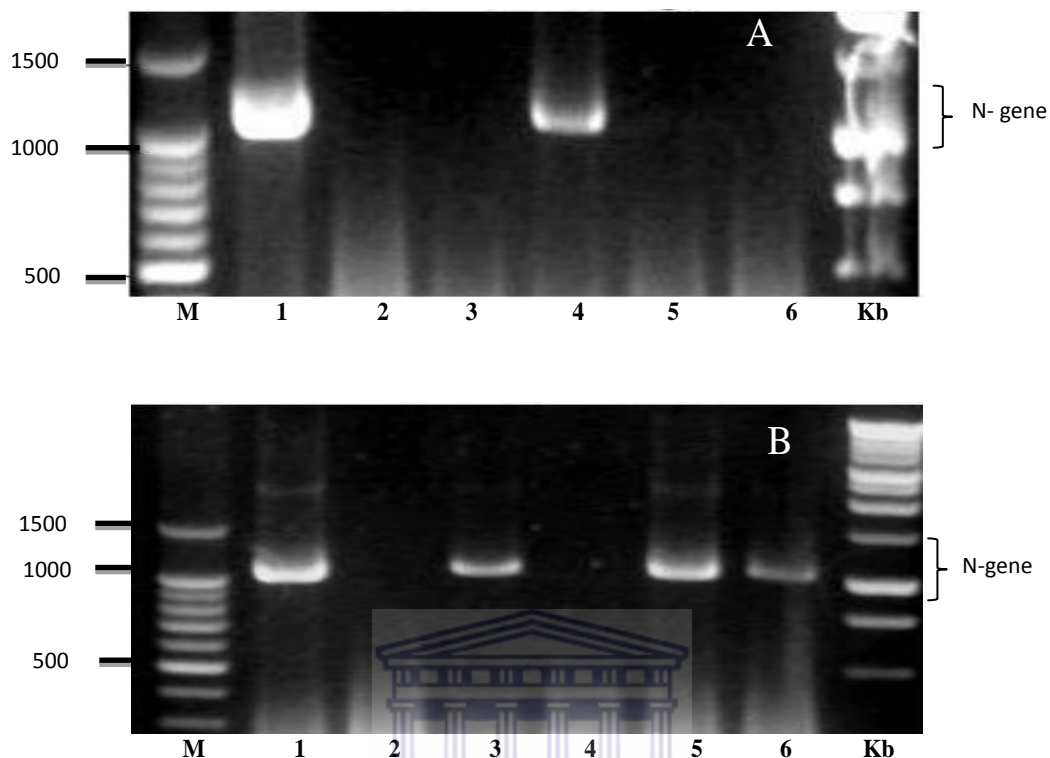


Figure 3.12: Colony PCR of KRX competent *E. coli* previously transformed with Flexi™ vector ligated with viral genes. Colony PCR was performed with primers described in Table 2.2 and used as confirmation of successful transformation and ligation. A: Lane 1- 100bp molecular marker, lane 2- positive control, lane 3- NL63 N gene colony (C)1, lane 4- NL63 C2, lane 5- NL63 N gene C3, lane 6- NL63 N gene C4, lane 7- negative control, lane 8 1Kbp molecular marker. **B:** Lane1- 100bp molecular marker, lane 2- NL63 N gene C5, lane 3- NL63 N gene C6, lane 4- NL63 N gene C7, Lane 5- C8, Lane 6- C9, Lane 7- C10, lane 8- 1Kbp molecular marker. All successful amplifications of viral genes represent colonies, picked from agar plates, which were successfully transformed with the respective Flexi™-viral gene construct.

3.3.1 Mammalian Expression Studies of HCoV-NL63 N, M and E proteins

The DNA constructs that was given as a kind gift and used in this thesis were previously cloned into the pCAGGS vector and the positive control in a pXJ 40 vector for expression in a mammalian cell culture system (Table 2.3).

3.3.2. Verification of recombinant plasmid

3.3.2.1 Restriction Endonuclease Digest

Initial digestion of pCAGGS-N-HA, pCAGGS-M-HA and pCAGGS-E-HA plasmids DNA, to verify the presence of the correct genes, with *EcoRI* and *NotI* excised 1100bp, 700bp and 274bp DNA fragments respectively (Fig.3.13). This is in agreement with the size of HCoV-NL63 N, M and E gene nucleotide sizes found on the NCBI website with accession numbers YP_003771.1, YP_003770.1 and YP_003769.1 respectively. Two bands for each construct were observed, thus confirming that the vectors harbour the respective genes of interest (Fig.13. Lanes 1, 2 and 3).

This is an indication that correct restriction endonucleases were selected and optimal conditions were used for its proper functioning, thereby avoiding “star activity”. The latter is the ability of a restriction enzyme to under non-optimal conditions cleave DNA sequences that are similar, but not identical, to the recognition site of the enzyme. *BamH1* and *EcoR1* are notorious for its star activity. Table 3.3 summarises the characteristics of the restriction enzymes used in this application.

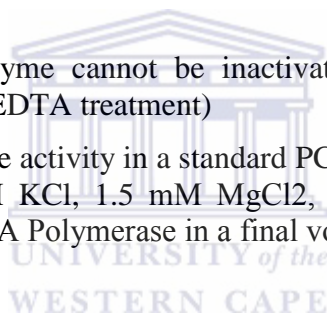
Table 3.3 : Restriction enzymes characteristics

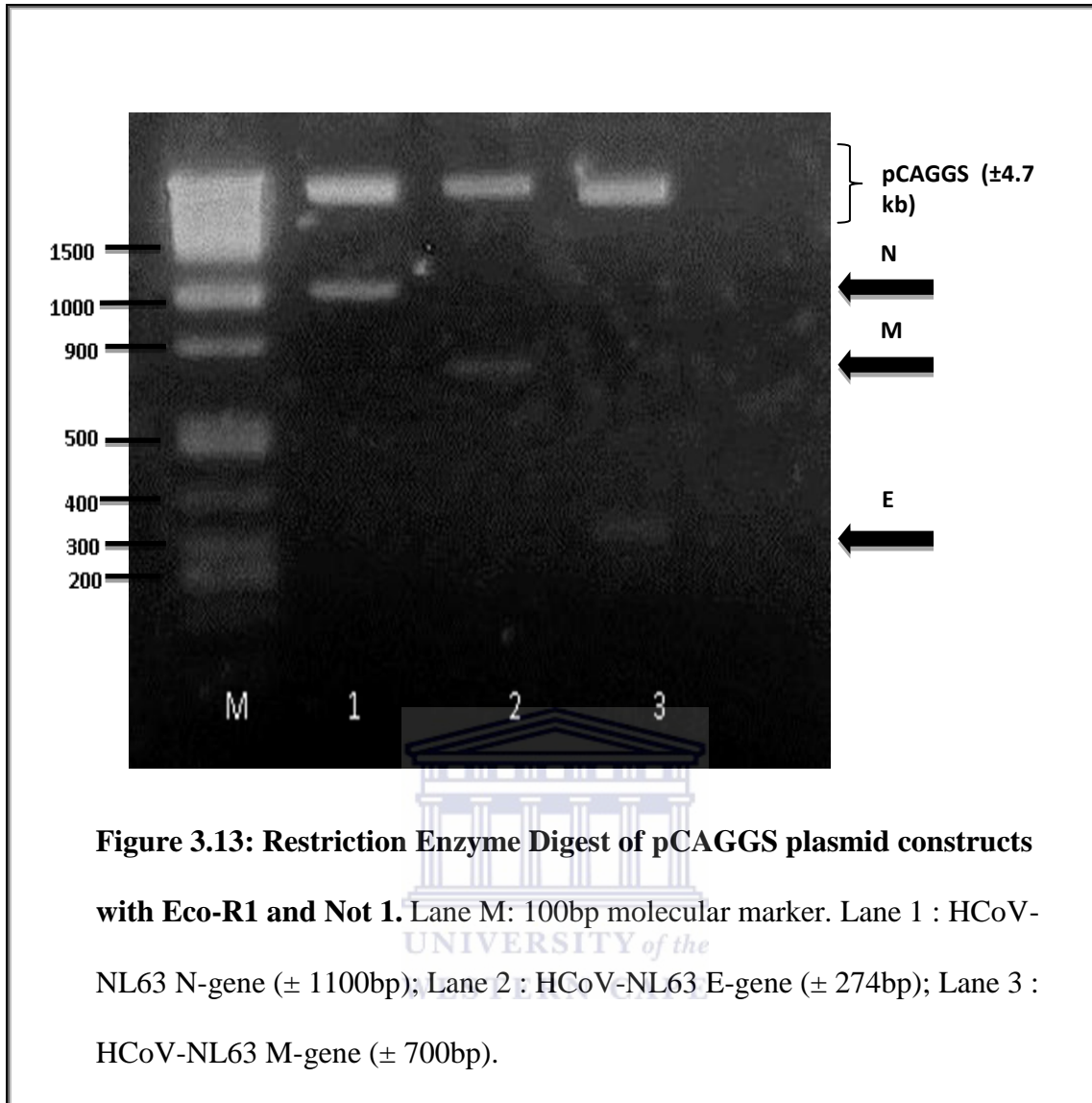
Restriction Enzyme	Sequence	Incubation Temperature	Heat Inactivation	PCR
BamH1	G↓GATCC	37	N	100%
EcoR1	G↓AATTC	37	Y	50%
Not 1	GC↓GGCCGC	37	Y	
Xho1	C↓TCGAG	37	Y	< 5%

Y: Indicates that the enzyme can be inactivated by heat (15 minutes at +65°C unless otherwise stated)

N: Indicates that the enzyme cannot be inactivated by heat and needs an alternative procedure (eg. EDTA treatment)

PCR: Percentage of enzyme activity in a standard PCR Mix (= 10 mM Tris HCl, pH 8.3 at +20°C, 50 mM KCl, 1.5 mM MgCl₂, λ-substrate DNA, 200 μM dNTP's and 2.5 U Taq DNA Polymerase in a final volume of 100 μl)





3.3.3. Plasmid Purification (Midi-Prep)

The purified plasmid DNA obtained from the NucleoBond® AX midi-prep (Macherey-Nagel) system is suitable for use in the most demanding molecular biology applications; including transfection, *in vitro* transcription, automated or manual sequencing, cloning, hybridization and PCR. For this reason, this assay kit was selected to optimise plasmid yield for more efficient transfection results.

Large-scale plasmid extractions (midi-prep) were performed and the recombinant plasmids were characterized by restriction endonuclease digestion using a 1% agarose

gel electrophoresis, as well as by nucleotide sequencing (Inqaba) of the cloned insert using reverse M13 universal primers (See Appendix for sequencing results) . The results obtained in Fig.3.14 corresponded exactly to the previous enzyme digest of pCAGGS-N-HA, pCAGGS-M-HA and pCAGGS-E-HA recombinant plasmids with *EcoRI* and *NotI*.

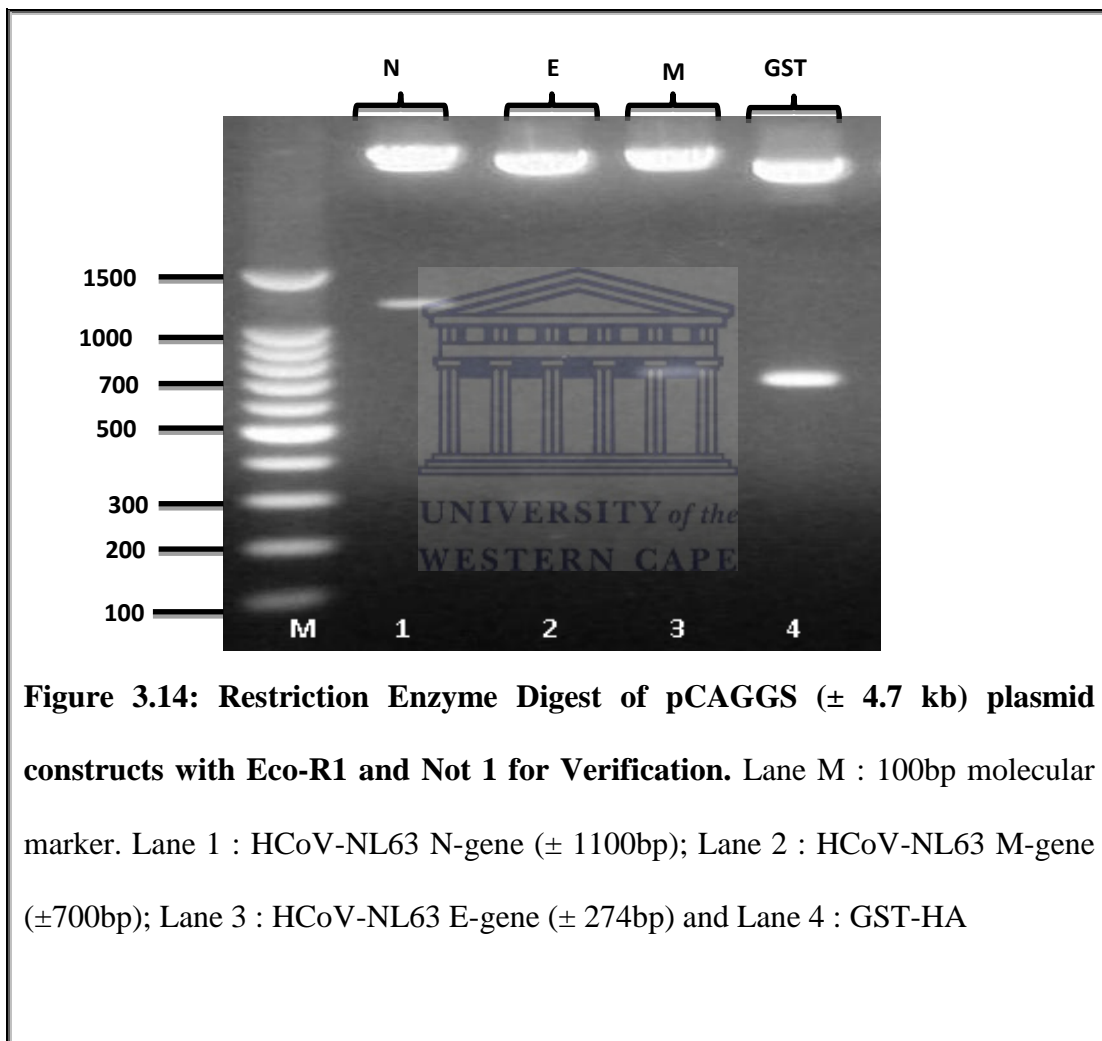


Figure 3.14: Restriction Enzyme Digest of pCAGGS (\pm 4.7 kb) plasmid constructs with *Eco-R1* and *Not 1* for Verification. Lane M : 100bp molecular marker. Lane 1 : HCoV-NL63 N-gene (\pm 1100bp); Lane 2 : HCoV-NL63 M-gene (\pm 700bp); Lane 3 : HCoV-NL63 E-gene (\pm 274bp) and Lane 4 : GST-HA

3.3.4 Quantification of plasmids

According to the manufacturer, the AX 100 (Midi) column should, on average, yield 20–100 µg plasmid DNA. However, a much lower total plasmid yield of $\pm 1.5\text{ug/ul}$ was consistently achieved in this thesis. There could be various possible reasons for this low plasmid yield such as: (i) the pH or salt concentrations of buffers may have been too high; (ii) the sample/lysate was too viscous, thereby causing the filtration of the lysate and flow rate of the column to be insufficient; (iii) the column may have been overloaded with nucleic acids; (iv) the plasmids may not have propagated well in the culture, therefore only a low concentration was present in the cleared lysate; (v) the alkaline lysis was inefficient; this happens when culture volume or pellet weight is too high; (vi) the lysate was incorrectly prepared; (vii) after storage below 20°C, the SDS in Buffer S2 may precipitate, causing inefficient lysis; or (viii) the lysis treatment may have been too harsh (Birnboim and Doly 1979). However, the plasmid yield was sufficient for the transfection process.

3.3.5. Expression of HCoV-NL63 N, M and E proteins in Cos 7 cells

The COS7 cell line was selected as the research tool, as it is known to be an excellent choice for transfection experiments using recombinant plasmids. In order to use these cells as transfection hosts, the cells are imbued with a genetic construct preceded by the SV40 promoter. Several well established methods for delivery of nucleic acids into mammalian cells exists. However no single technique alone is suitable for the multitude of different cellular systems used for transfection experiments. Lipofectamine 200 (Invitrogen) was used as the transfection reagent. This process

involves a lipid with an overall net positive charge. The cationic portion of the lipid molecule associates with the negatively charged nucleic acids, resulting in compaction of the nucleic acid in a liposome/ nucleic acid complex (Kabanov and Kabanov,1995) (Labat-Moleur et al, 1996), presumably from electrostatic interactions between the negatively charged nucleic acids and the positively charged head group of the synthetic lipid. For cultured cells, an overall net positive charge of the liposome/ nucleic acid complex generally results in higher transfer efficiencies, presumably because this allows closer association of the complex with the negatively charged cell membranes. Entry of the liposome complex into the cell may occur by the processes of endocytosis or fusion with the plasma membrane via lipid moieties of the liposome (Gao and Huang, 1995). Opti-MEM® which is a reduced serum media was used with the Lipofectamine 200 (Invitrogen) to aid in the cationic lipid transfection process.

In order to analyse the transient expression of the transfected genes, cells were harvested 24 hours post-transfection. Harvesting the cells before this time resulted in a low yield of expressed proteins; this could be due to the fact that reporter genes are usually expressed between 24-48 hours post-transfection. Harvesting after 24 hours also resulted in a much lower protein yield - this could be due to a reduction in the cell's proliferation rate. COS7 cells grow relatively fast; therefore, after more than a 24 hour incubation period, the cells started to detach. When the cells die, different proteases and lysozymes are released, thereby degrading the expressed proteins in solution.

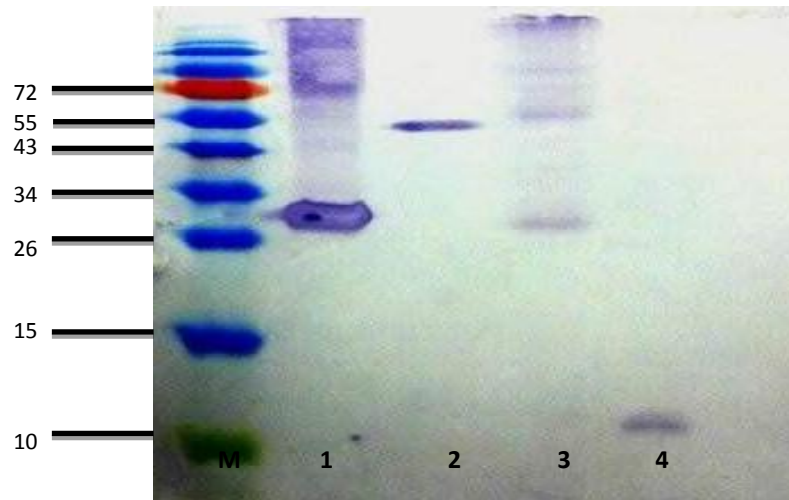
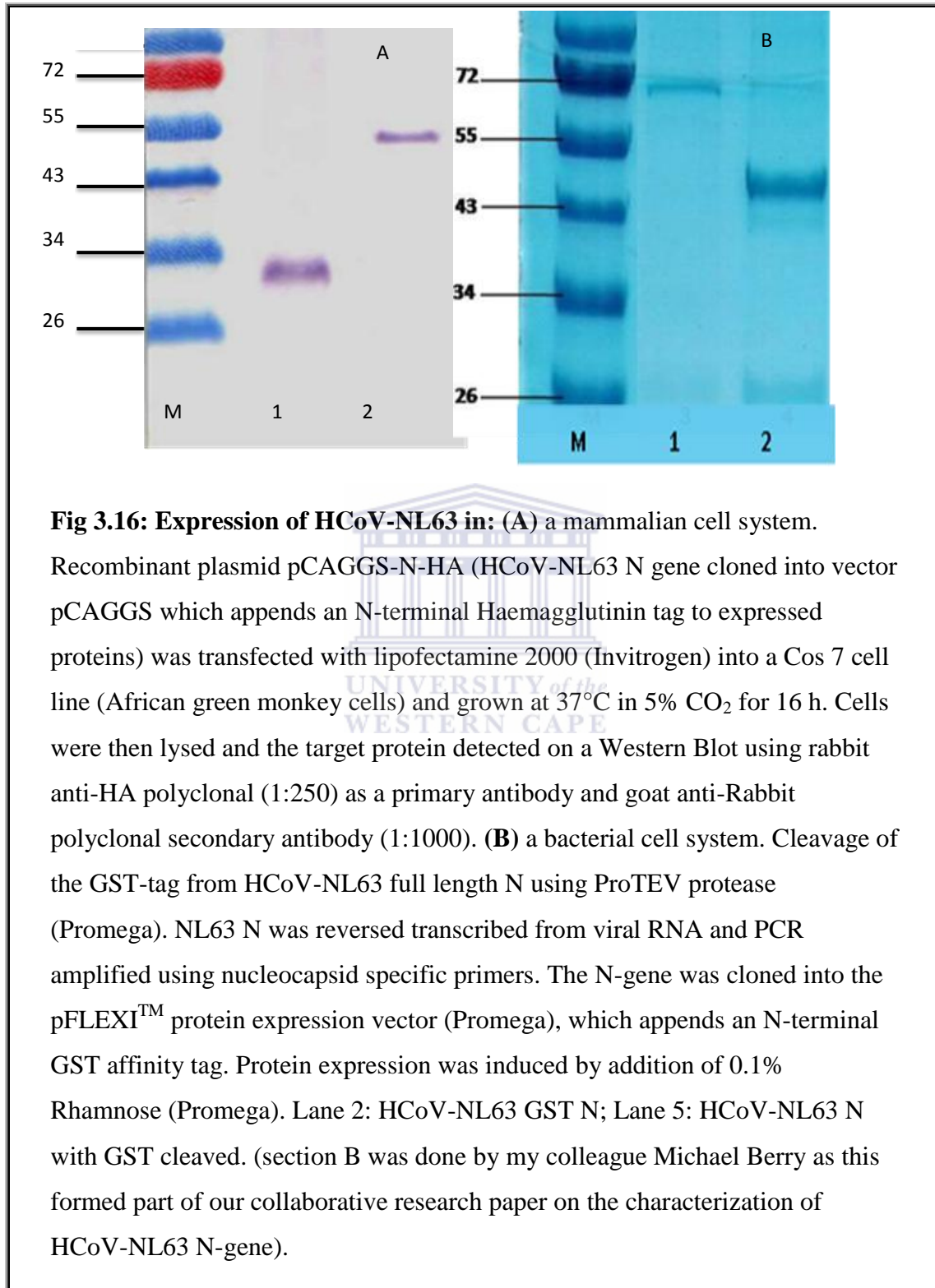


Figure 3.15: Western Blot Analysis of Expressed Proteins in COS7 cells. Recombinant plasmids pCAGGS-N-HA; pCAGGS-M-HA and pCAGGS-E-HA (HCoV-NL63 N, M and E gene cloned into vector pCAGGS which appends an N-terminal Haemagglutinin tag to expressed proteins) was transfected with Lipofectamine 2000 (Invitrogen) into a Cos 7 cell line (African green monkey cells) and grown at 37°C in 5% CO₂ for 16 h. Cells were then lysed and the target protein detected on a Western Blot using rabbit anti-HA polyclonal (1:250) as a primary antibody and goat anti-Rabbit polyclonal secondary antibody (1:1000). The proteins were separated by SDS-PAGE electrophoresis using a 15% polyacrylamide gel @ 15 mAmps per gel. Lane M – pre-stained protein marker (kDa), Lane 1- GST, Lane 2 - NL63 N protein, Lane 3 - NL63 M protein, Lane 4 - NL63 E protein.

pCAGGS-N-HA, pCAGGS-M-HA and pCAGGS-E-HA was selected for expression studies. A 24 hour protein expression time course was run to access the expression characteristics of the N, M and E proteins in the COS7 mammalian cells. Fig.3.15 shows the expression of N-HA, M-HA and E-HA at 24 hours. The cells were harvested; lysed and total proteins were separated on a 15% SDS-PAGE and Western Blotted. The supernatant samples were subjected to immunoblotting with anti-HA antibody (Sigma). A very high concentration of antibodies in the ratio of 1:250 had been used in order to detect the expressed proteins. It is speculated that this is due to very low quality antibodies coupled with a low protein yield, as Dr M.Muller whom I received the recombinants from had also observed a low protein expression yield. The N-HA, M-HA and E-HA protein was detected as 43 kDa, 27kDa and ± 9 kDa proteins respectively (Fig 3.15). The anti-HA antibody specially detected the N, M and E proteins. In figure 3.15 Lane 3 the M protein contains more than one band; this is thought to be due to interaction or aggregations with other viral or cellular proteins. The proteins detected, corresponds to the predicted protein sizes obtained from NCBI (Table 3.2) (Altschul and Koonin 1998).

3.3.6. Comparisons of Bacterial and Mammalian Expression of the HCoV-NL63 Nucleocapsid protein.



HCoV-NL63 genome is about $\pm 27\,553$ nucleotides in size, and has a typical coronavirus genome organization. It produces six separate mRNAs, with all potential ORFs encoding for viral proteins (Abdul-Rasool and Fielding 2010). N is expressed from distinct subgenomic (sg) mRNA 6, the most abundant sg mRNA (Pyrce, Jebbink et al. 2004). In order to determine the size of HCoV-NL63 N, recombinant GST-N plasmid was expressed in a bacterial system. It was then purified by affinity column and treated with TEV to remove the GST-tag, according to the manufacturer's specification (Promega). The recombinant HCoV-NL63 N-HA was transfected and expressed in Cos-7 cells (Khan, Fielding et al. 2006). SDS-PAGE analysis of purified bacterial expressed protein (Figure 3.16A) and Western Blot analysis of total mammalian cell lysate (Figure 3.16B) showed ~50 kDa proteins which is larger than the size of the 42.6 kDa protein that had been predicted by our initial analysis (<http://www.sciencegateway.org/tools/proteinmw.htm>). This difference in size showed that the protein was probably pre-, co- or post-translationally modified; the type of modification is yet to be investigated. Once determined the functions of this potential modifications could shed light on the possible roles of these proteins in HCoV-NL63 pathogenesis.

Chapter 4



Summary

The World Health Organization estimates that about 20% of deaths in children younger than 5 years old are due to acute lower respiratory tract infections. Several viruses have shown to be implicated in these infections, including rhinovirus, influenza viruses, parainfluenza viruses, respiratory syncytial viruses, adenoviruses and coronaviruses. Previous reports had indicated that human coronaviruses account for a significant number of hospitalizations for children under 18 years of age, accounting for 4.4% of all admissions for acute respiratory infections (Chiu, Chan et al. 2005).

Human coronaviruses were first isolated in the 1960s and since then, HCoV have been found to be mainly associated with respiratory tract illness, but can also cause enteric and central nervous system diseases. Until recently, only two human coronaviruses were known to science, i.e. HCoV-229E and HCoV-OC43. With the outbreak of SARS in 2003, a third previously unknown coronavirus was identified as the aetiological agent. SARS human coronavirus eventually spread from China to more than 30 countries and resulted in more than 800 fatalities worldwide, this equated to about a 10% mortality rate. Then in 2004, a fourth human coronavirus was isolated from a 7 month old child with respiratory symptoms. This virus was identified as a close relative of SARS-CoV and was named human coronavirus-NL63 (van der Hoek, Pyrc et al. 2004). The human coronavirus NL63 (HCoV-NL63) is a clinically important virus; it mainly affects children, the elderly and immunocompromised.

HCoV-NL63 infection usually results in mild respiratory tract disease but it has shown to be involved with more severe lower respiratory tract infections such as bronchiolitis, pneumonia, and croup. At a molecular level, this virus is assembled through interactions of its structural proteins, namely the spike (S), membrane (M), envelope (E) and nucleocapsid (N) proteins.

3 out of the 4 major structural proteins are characterized and analysed by means of bioinformatics tools and protein expression strategies. Special attention has been paid to conserved regions of coronavirus N-, M- and E- homologous, identification of transmembrane domains, hydrophobic and hydrophilic regions, putative phosphorylation, N-linked glycosylation and myristoylation sites, which could give insight into the functioning of the proteins. With the use of molecular research tools the N, M and E genes had been transfected into a mammalian cell system in order for it to be transiently expressed. Furthermore comparisons were made of the N protein expression in a bacterial and mammalian cell system.

Bioinformatics tools had revealed 4 transmembrane regions found within the M-protein of HCoV-NL63 that differs slightly with research done on other coronaviruses in which only 3 transmembrane domains are identified (Ruch and Machamer 2012) . This could be due to the use of different algorithms; however these results are confirmed with the use of hydropathy predictions. A single transmembrane domain had been observed in the E protein sequence, this shed s light on its topology as different topologies for the E protein had been proposed (Torres, Wang et al. 2005; Ye and Hogue 2007; Ruch and Machamer 2012). This is an indication that the function of the E protein is not conserved amongst coronaviruses. Many questions have been derived from this. The E protein of HCoV-NL63 shares a high percentage sequence homology with HCoV-229E E protein. Could this be an indication that they share the same topology? And if so does the E protein of coronaviruses from the same group share the same topology? Seeing that the E protein function is not a conserved property amongst coronaviruses, does this protein still remain to be a major structural protein in coronavirus assembly and pathogenesis? All of these questions remain to be answered and can serve as a prelude to further studies.

For the N protein sequence a highly conserved region was found using multiple sequence alignment tools and a SR (serine) rich motif had been identified. The N protein also showed to be highly phosphorylated at multiple serines, and is also highly immunogenic (Berry 2012). This conserved region of ± 8 amino acids may be responsible for the mediating of RNA binding. This region is also found within the disordered region of the N protein sequence. This is evident that these sites of phosphorylation and increased immunogenicity can be directed to its SR-rich region that falls within the disordered state. Could this be a great potential target for antiviral strategies? As disordered regions of a protein have no fixed tertiary structure, thereby allowing access to binding sites.

Even though all of these various characteristic have been identified with the use of bioinformatics tools, it is still necessary to prove these results *in-vitro*. In order to do a comparison of the N protein expression in a bacterial and mammalian system, a recombinant GST-N had been expressed in a bacterial system, purified by an affinity column and treated with TEV to remove the GST-tag. Also, the recombinant HCoV-NL63 N-HA was transfected and expressed in COS7 cells. On analysis of the two protein expression systems a 50kDa protein had been shown, which is larger than the size of 42.6 kDa protein that had been predicted (Berry 2012). This discrepancy in size showed that the protein was probably pre-, co- or post-translationally modified, this type of modification is still to be investigated in future studies.

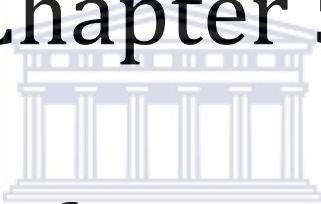
To further characterise the N-, M-, and E-proteins of HCoV-NL63 in an in vitro model, it had been expressed in COS7 cells. Extensive optimization had been done in order for these proteins to be expressed. These include only GST-HA (positive control) had been detected on the blot, which could have been due to a very low concentration of DNA used in transfection, the transfection agent (lipofectamine) was either degraded or lethal to the cells; the cells were not accepting the DNA readily, thereby not expressing the proteins. The

detection method could have been faulted, such as either the primary or secondary antibody had degraded. The problem seemed to have been 1 of 3 things or a collaboration of all 3, as the proteins were detected after being treated with a high concentration of primary antibody (1:250), which means the antibody could have been of low quality or a very low level of proteins were expressed by the COS7 cells, or the lyses buffer was not releasing the proteins well enough from the cell membrane portion. These heterologous expressed proteins can now serve as the basis of several post expression studies.

This dissertation had provided new knowledge of how HCoV-NL63 assembles inside an infected cell. Such knowledge will help us to identify major targets for antiviral drug development, not only for HCoV-NL63 but for other coronaviruses too, as a new coronavirus had been identified in September 2012 and had already resulted in two fatalities.



Chapter 5



References

UNIVERSITY OF THE
WESTERN CAPE

1. Abdul-Rasool, S. and B. C. Fielding (2010). "Understanding Human Coronavirus HCoV-NL63." *Open Virol J* 4: 76-84.
2. Almazan, F., C. Galan, et al. (2004). "The nucleoprotein is required for efficient coronavirus genome replication." *J Virol* 78(22): 12683-12688.
3. Altschul, S. F. and E. V. Koonin (1998). "Iterated profile searches with PSI-BLAST-- a tool for discovery in protein databases." *Trends Biochem Sci* 23(11): 444-447.
4. Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res* 25(17): 3389-3402.
5. An, S., C. J. Chen, et al. (1999). "Induction of apoptosis in murine coronavirus-infected cultured cells and demonstration of E protein as an apoptosis inducer." *J Virol* 73(9): 7853-7859.
6. Arciola, C. R., D. Campoccia, et al. (2006). "Prevalence and antibiotic resistance of 15 minor staphylococcal species colonizing orthopedic implants." *Int J Artif Organs* 29(4): 395-401.
7. Arden, K. E., M. D. Nissen, et al. (2005). "New human coronavirus, HCoV-NL63, associated with severe lower respiratory tract disease in Australia." *J Med Virol* 75(3): 455-462.
8. Armstrong, J., H. Niemann, et al. (1984). "Sequence and topology of a model intracellular membrane protein, E1 glycoprotein, from a coronavirus." *Nature* 308(5961): 751-752.
9. Baker, S. C., C. Shimizu, et al. (2006). "Human coronavirus-NL63 infection is not associated with acute Kawasaki disease." *Adv Exp Med Biol* 581: 523-526.

10. Baric, R. S., G. W. Nelson, et al. (1988). "Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription." *J Virol* 62(11): 4280-4287.
11. Bastien, N., J. L. Robinson, et al. (2005). "Human coronavirus NL-63 infections in children: a 1-year study." *J Clin Microbiol* 43(9): 4567-4573.
12. Bastien, N., K. Anderson, et al. (2005). "Human coronavirus NL63 infection in Canada." *J Infect Dis* 191(4): 503-506.
13. Baudoux, P., C. Carrat, et al. (1998). "Coronavirus pseudoparticles formed with recombinant M and E proteins induce alpha interferon synthesis by leukocytes." *J Virol* 72(11): 8636-8643.
14. Baudoux, P., L. Besnardeau, et al. (1998). "Interferon alpha inducing property of coronavirus particles and pseudoparticles." *Adv Exp Med Biol* 440: 377-386.
15. Belay, E. D., D. D. Erdman, et al. (2005). "Kawasaki disease and human coronavirus." *J Infect Dis* 192(2): 352-353; author reply 353.
16. Berry, M. M., T. Tan, Y. Fielding, B.C. (2012). "Characterisation of human coronavirus-Nl63 nucleocapsid protein." *African Journal of Biotechnology* 11(75): 13962-13968.
17. Birnboim, H. C. and J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." *Nucleic Acids Res* 7(6): 1513-1523.
18. Bolivar, F. (1978). "Construction and characterization of new cloning vehicles. III. Derivatives of plasmid pBR322 carrying unique Eco RI sites for selection of Eco RI generated recombinant DNA molecules." *Gene* 4(2): 121-136.
19. Bos, E. C., W. Luytjes, et al. (1996). "The production of recombinant infectious DI-particles of a murine coronavirus in the absence of helper virus." *Virology* 218(1): 52-60.

20. Bosch, B. J., C. A. de Haan, et al. (2005). "Spike protein assembly into the coronavirus: exploring the limits of its sequence requirements." *Virology* 334(2): 306-318.
21. Bost, A. G., R. H. Carnahan, et al. (2000). "Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly." *J Virol* 74(7): 3379-3387.
22. Bradburne, A. F., M. L. Bynoe, et al. (1967). "Effects of a "new" human respiratory virus in volunteers." *Br Med J* 3(5568): 767-769.
23. Bryce J., Boschi-Pinto C., Shibuya K., Black R. E. (2005) The WHO Child Health Epidemiology Reference Group. "WHO estimates of the causes of death in children." *Lancet* 365: 1147–1152.
24. Burgner, D. and A. Harnden (2005). "Kawasaki disease: what is the epidemiology telling us about the etiology?" *Int J Infect Dis* 9(4): 185-194.
25. Calvo, E., D. Escors, et al. (2005). "Phosphorylation and subcellular localization of transmissible gastroenteritis virus nucleocapsid protein in infected cells." *J Gen Virol* 86(Pt 8): 2255-2267.
26. Canducci, F., M. Debiaggi, et al. (2008). "Two-year prospective study of single infections and co-infections by respiratory syncytial virus and viruses identified recently in infants with acute respiratory disease." *J Med Virol* 80(4): 716-723.
27. Casais, R., V. Thiel, et al. (2001). "Reverse genetics system for the avian coronavirus infectious bronchitis virus." *J Virol* 75(24): 12359-12369.
28. Chang, C. K., Y. L. Hsu, et al. (2009). "Multiple nucleic acid binding sites and intrinsic disorder of severe acute respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging." *J Virol* 83(5): 2255-2264.

29. Chen, H., A. Gill, et al. (2005). "Mass spectroscopic characterization of the coronavirus infectious bronchitis virus nucleoprotein and elucidation of the role of phosphorylation in RNA binding by using surface plasmon resonance." *J Virol* 79(2): 1164-1179.
30. Chiu, S. S., K. H. Chan, et al. (2005). "Human coronavirus NL63 infection and other coronavirus infections in children hospitalized with acute respiratory disease in Hong Kong, China." *Clin Infect Dis* 40(12): 1721-1729.
31. Cornelissen, L. A., P. A. van Woensel, et al. (1998). "Cell culture-grown putative bovine respiratory torovirus identified as a coronavirus." *Vet Rec* 142(25): 683-686.
32. Corse, E. and C. E. Machamer (2000). "Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles." *J Virol* 74(9): 4319-4326.
33. Curtis, K. M., B. Yount, et al. (2002). "Heterologous gene expression from transmissible gastroenteritis virus replicon particles." *J Virol* 76(3): 1422-1434.
34. Dare, R. K., A. M. Fry, et al. (2007). "Human coronavirus infections in rural Thailand: a comprehensive study using real-time reverse-transcription polymerase chain reaction assays." *J Infect Dis* 196(9): 1321-1328.
35. de Haan, C. A., H. Vennema, et al. (1998). "Coronavirus envelope assembly is sensitive to changes in the terminal regions of the viral M protein." *Adv Exp Med Biol* 440: 367-375.
36. de Haan, C. A., H. Vennema, et al. (2000). "Assembly of the coronavirus envelope: homotypic interactions between the M proteins." *J Virol* 74(11): 4967-4978.
37. de Haan, C. A., L. Kuo, et al. (1998). "Coronavirus particle assembly: primary structure requirements of the membrane protein." *J Virol* 72(8): 6838-6850.

38. de Haan, C. A., M. de Wit, et al. (2002). "O-glycosylation of the mouse hepatitis coronavirus membrane protein." *Virus Res* 82(1-2): 77-81.
39. de Haan, C. A., M. de Wit, et al. (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(2): 395-406.
40. de Haan, C. A., Z. Li, et al. (2005). "Murine coronavirus with an extended host range uses heparan sulfate as an entry receptor." *J Virol* 79(22): 14451-14456.
41. DeDiego, M. L., E. Alvarez, et al. (2007). "A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo." *J Virol* 81(4): 1701-1713.
42. Denny, F. W., T. F. Murphy, et al. (1983). "Croup: an 11-year study in a pediatric practice." *Pediatrics* 71(6): 871-876.
43. Dijkman, R., M. F. Jebbink, et al. (2008). "Human coronavirus NL63 and 229E seroconversion in children." *J Clin Microbiol* 46(7): 2368-2373.
44. Dominguez, S. R., M. S. Anderson, et al. (2006). "Blinded case-control study of the relationship between human coronavirus NL63 and Kawasaki syndrome." *J Infect Dis* 194(12): 1697-1701.
45. Dove, B. K., J. H. You, et al. (2006). "Changes in nucleolar morphology and proteins during infection with the coronavirus infectious bronchitis virus." *Cell Microbiol* 8(7): 1147-1157.
46. Ebihara, T., R. Endo, et al. (2005b). "Lack of association between New Haven coronavirus and Kawasaki disease." *J Infect Dis* 192(2): 351-352; author reply 353.
47. Esper, F., C. Weibel, et al. (2005). "Evidence of a novel human coronavirus that is associated with respiratory tract disease in infants and young children." *J Infect Dis* 191(4): 492-498.

48. Esper, F., E. D. Shapiro, et al. (2005). "Association between a novel human coronavirus and Kawasaki disease." *J Infect Dis* 191(4): 499-502.
49. Fan, H., A. Ooi, et al. (2005). "The nucleocapsid protein of coronavirus infectious bronchitis virus: crystal structure of its N-terminal domain and multimerization properties." *Structure* 13(12): 1859-1868.
50. Fang, X., L. Ye, et al. (2005). "Peptide domain involved in the interaction between membrane protein and nucleocapsid protein of SARS-associated coronavirus." *J Biochem Mol Biol* 38(4): 381-385.
51. Fielding, B. C. (2011). "Human coronavirus NL63: a clinically important virus?" *Future Microbiol* 6(2): 153-159.
52. Fischer, F., C. F. Stegen, et al. (1998). "Analysis of constructed E gene mutants of mouse hepatitis virus confirms a pivotal role for E protein in coronavirus assembly." *J Virol* 72(10): 7885-7894.
53. Fouchier, R. A., N. G. Hartwig, et al. (2004). "A previously undescribed coronavirus associated with respiratory disease in humans." *Proc Natl Acad Sci U S A* 101(16): 6212-6216.
54. Fouchier, R. A., T. Kuiken, et al. (2003). "Aetiology: Koch's postulates fulfilled for SARS virus." *Nature* 423(6937): 240.
55. Fu, X. D. (1995). "The superfamily of arginine/serine-rich splicing factors." *RNA* 1(7): 663-680.
56. Godet, M., R. L'Haridon, et al. (1992). "TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions." *Virology* 188(2): 666-675.
57. Graveley, B. R. (2000). "Sorting out the complexity of SR protein functions." *RNA* 6(9): 1197-1211.
58. Hall, K. B. (2002). "RNA-protein interactions." *Curr Opin Struct Biol* 12(3): 283-288.

59. Hamre, D. and J. J. Procknow (1966). "A new virus isolated from the human respiratory tract." *Proc Soc Exp Biol Med* 121(1): 190-193.
60. He, M. L., B. Zheng, et al. (2003). "Inhibition of SARS-associated coronavirus infection and replication by RNA interference." *JAMA* 290(20): 2665-2666.
61. He, R., A. Leeson, et al. (2004). "Characterization of protein-protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus." *Virus Res* 105(2): 121-125.
62. Hiscox, J. A., T. Wurm, et al. (2001). "The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus." *J Virol* 75(1): 506-512.
63. Ho, Y., P. H. Lin, et al. (2004). "Assembly of human severe acute respiratory syndrome coronavirus-like particles." *Biochem Biophys Res Commun* 318(4): 833-838.
64. Hofmann, H., K. Pyrc, et al. (2005). "Human coronavirus NL63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry." *Proc Natl Acad Sci U S A* 102(22): 7988-7993.
65. Hsieh, P. K., S. C. Chang, et al. (2005). "Assembly of severe acute respiratory syndrome coronavirus RNA packaging signal into virus-like particles is nucleocapsid dependent." *J Virol* 79(22): 13848-13855.
66. Huang, Y., Z. Y. Yang, et al. (2004). "Generation of synthetic severe acute respiratory syndrome coronavirus pseudoparticles: implications for assembly and vaccine production." *J Virol* 78(22): 12557-12565.
67. Huang, Z., A. Chou, et al. (2008). "Levels of N-linked glycosylation on the V1 loop of HIV-1 Env proteins and their relationship to the antigenicity of Env from primary viral isolates." *Curr HIV Res* 6(4): 296-305.

68. Hurst, K. R., L. Kuo, et al. (2005). "A major determinant for membrane protein interaction localizes to the carboxy-terminal domain of the mouse coronavirus nucleocapsid protein." *J Virol* 79(21): 13285-13297.
69. Jayaram, H., H. Fan, et al. (2006). "X-ray structures of the N- and C-terminal domains of a coronavirus nucleocapsid protein: implications for nucleocapsid formation." *J Virol* 80(13): 6612-6620.
70. Kahn, J. S. and K. McIntosh (2005). "History and recent advances in coronavirus discovery." *Pediatr Infect Dis J* 24(11 Suppl): S223-227, discussion S226.
71. Kaiser, L., N. Regamey, et al. (2005). "Human coronavirus NL63 associated with lower respiratory tract symptoms in early life." *Pediatr Infect Dis J* 24(11): 1015-1017.
72. Kedersha, N. and P. Anderson (2002). "Stress granules: sites of mRNA triage that regulate mRNA stability and translatability." *Biochem Soc Trans* 30(Pt 6): 963-969.
73. Khan, S., B. C. Fielding, et al. (2006). "Over-expression of severe acute respiratory syndrome coronavirus 3b protein induces both apoptosis and necrosis in Vero E6 cells." *Virus Res* 122(1-2): 20-27.
74. Klumperman, J., J. K. Locker, et al. (1994). "Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding." *J Virol* 68(10): 6523-6534.
75. Krogh, A., B. Larsson, et al. (2001). "Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes." *J Mol Biol* 305(3): 567-580.
76. Krzysztof Pyrc, M. F. J., Ben Berkhout and Lia van der Hoek (2004). "Genome Structure and Transcriptional Regulation of Human Coronavirus NL63."
77. Kuiken, T., R. A. Fouchier, et al. (2003). "Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome." *Lancet* 362(9380): 263-270.

78. Kuo, L. and P. S. Masters (2003). "The small envelope protein E is not essential for murine coronavirus replication." *J Virol* 77(8): 4597-4608.
79. Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." *J Mol Biol* 157(1): 105-132.
80. Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." *J Mol Biol* 157(1): 105-132.
81. Lai, M. M. (1997). "RNA-protein interactions in the regulation of coronavirus RNA replication and transcription." *Biol Chem* 378(6): 477-481.
82. Lambert, S. B., K. M. Allen, et al. (2007). "Community epidemiology of human metapneumovirus, human coronavirus NL63, and other respiratory viruses in healthy preschool-aged children using parent-collected specimens." *Pediatrics* 120(4): e929-937.
83. Lapps, W., B. G. Hogue, et al. (1987). "Deduced amino acid sequence and potential O-glycosylation sites for the bovine coronavirus matrix protein." *Adv Exp Med Biol* 218: 123-129.
84. Larkin, M. A., G. Blackshields, et al. (2007). "Clustal W and Clustal X version 2.0." *Bioinformatics* 23(21): 2947-2948.
85. Laude, H., M. Godet, et al. (1995). "Functional domains in the spike protein of transmissible gastroenteritis virus." *Adv Exp Med Biol* 380: 299-304.
86. Leung, T. F., P. K. Chan, et al. (2012). "Human coronavirus NL63 in children: epidemiology, disease spectrum, and genetic diversity." *Hong Kong Med J* 18 Suppl 2: 27-30.
87. Li, W., J. Sui, et al. (2007). "The S proteins of human coronavirus NL63 and severe acute respiratory syndrome coronavirus bind overlapping regions of ACE2." *Virology* 367(2): 367-374.

88. Liao, Y., Q. Yuan, et al. (2006). "Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein." *Virology* 349(2): 264-275.
89. Luo, H., F. Ye, et al. (2005). "SR-rich motif plays a pivotal role in recombinant SARS coronavirus nucleocapsid protein multimerization." *Biochemistry* 44(46): 15351-15358.
90. Luo, H., J. Chen, et al. (2006). "Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization." *Biochemistry* 45(39): 11827-11835.
91. Macnaughton, M. R., D. Flowers, et al. (1983). "Diagnosis of human coronavirus infections in children using enzyme-linked immunosorbent assay." *J Med Virol* 11(4): 319-325.
92. Maeda, J., A. Maeda, et al. (1999). "Release of coronavirus E protein in membrane vesicles from virus-infected cells and E protein-expressing cells." *Virology* 263(2): 265-272.
93. Marra, M. A., S. J. Jones, et al. (2003). "The Genome sequence of the SARS-associated coronavirus." *Science* 300(5624): 1399-1404.
94. Masters, P. S. (2006). "The molecular biology of coronaviruses." *Adv Virus Res* 66: 193-292.
95. Masters, P. S., L. Kuo, et al. (2006). "Genetic and molecular biological analysis of protein-protein interactions in coronavirus assembly." *Adv Exp Med Biol* 581: 163-173.
96. Maurer-Stroh, S. and F. Eisenhaber (2004). "Myristoylation of viral and bacterial proteins." *Trends Microbiol* 12(4): 178-185.
97. McIntosh, K. (2005). "Coronaviruses in the limelight." *J Infect Dis* 191(4): 489-491.

98. McIntosh, K., J. H. Dees, et al. (1967). "Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease." *Proc Natl Acad Sci U S A* 57(4): 933-940.
99. Minosse, C., M. Selleri, et al. (2008). "Phylogenetic analysis of human coronavirus NL63 circulating in Italy." *J Clin Virol* 43(1): 114-119.
100. Moes, E., L. Vijgen, et al. (2005). "A novel pancoronavirus RT-PCR assay: frequent detection of human coronavirus NL63 in children hospitalized with respiratory tract infections in Belgium." *BMC Infect Dis* 5: 6.
101. Mortola, E. and P. Roy (2004). "Efficient assembly and release of SARS coronavirus-like particles by a heterologous expression system." *FEBS Lett* 576(1-2): 174-178.
102. Narayanan, K., A. Maeda, et al. (2000). "Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells." *J Virol* 74(17): 8127-8134.
103. Narayanan, K., K. H. Kim, et al. (2003). "Characterization of N protein self-association in coronavirus ribonucleoprotein complexes." *Virus Res* 98(2): 131-140.
104. Ortego, J., D. Escors, et al. (2002). "Generation of a replication-competent, propagation-deficient virus vector based on the transmissible gastroenteritis coronavirus genome." *J Virol* 76(22): 11518-11529.
105. Peiris, J. S., S. T. Lai, et al. (2003). "Coronavirus as a possible cause of severe acute respiratory syndrome." *Lancet* 361(9366): 1319-1325.
106. Peng, D., C. A. Koetzner, et al. (1995). "Analysis of second-site revertants of a murine coronavirus nucleocapsid protein deletion mutant and construction of nucleocapsid protein mutants by targeted RNA recombination." *J Virol* 69(6): 3449-3457.

107. Peng, T. Y., K. R. Lee, et al. (2008). "Phosphorylation of the arginine/serine dipeptide-rich motif of the severe acute respiratory syndrome coronavirus nucleocapsid protein modulates its multimerization, translation inhibitory activity and cellular localization." *FEBS J* 275(16): 4152-4163.
108. Pyrc, K., B. Berkhout, et al. (2007). "Antiviral strategies against human coronaviruses." *Infect Disord Drug Targets* 7(1): 59-66.
109. Pyrc, K., B. Berkhout, et al. (2007). "Identification of new human coronaviruses." *Expert Rev Anti Infect Ther* 5(2): 245-253.
110. Pyrc, K., B. Berkhout, et al. (2007). "The novel human coronaviruses NL63 and HKU1." *J Virol* 81(7): 3051-3057.
111. Pyrc, K., B. J. Bosch, et al. (2006). "Inhibition of human coronavirus NL63 infection at early stages of the replication cycle." *Antimicrob Agents Chemother* 50(6): 2000-2008.
112. Pyrc, K., M. F. Jebbink, et al. (2004). "Genome structure and transcriptional regulation of human coronavirus NL63." *Virology* 321(1): 7-14.
113. Pyrc, K., R. Dijkman, et al. (2006). "Mosaic structure of human coronavirus NL63, one thousand years of evolution." *J Mol Biol* 364(5): 964-973.
114. Reed, M. L., B. K. Dove, et al. (2006). "Delineation and modelling of a nucleolar retention signal in the coronavirus nucleocapsid protein." *Traffic* 7(7): 833-848.
115. Risco, C., I. M. Anton, et al. (1996). "The transmissible gastroenteritis coronavirus contains a spherical core shell consisting of M and N proteins." *J Virol* 70(7): 4773-4777.
116. Risco, C., I. M. Anton, et al. (1998). "Structure and intracellular assembly of the transmissible gastroenteritis coronavirus." *Adv Exp Med Biol* 440: 341-346.

117. Rota, P. A., M. S. Oberste, et al. (2003). "Characterization of a novel coronavirus associated with severe acute respiratory syndrome." *Science* 300(5624): 1394-1399.
118. Rottier, P., D. Brandenburg, et al. (1984). "Assembly in vitro of a spanning membrane protein of the endoplasmic reticulum: the E1 glycoprotein of coronavirus mouse hepatitis virus A59." *Proc Natl Acad Sci U S A* 81(5): 1421-1425.
119. Rottier, P., D. Brandenburg, et al. (1984). "In vitro assembly of the murine coronavirus membrane protein E1." *Adv Exp Med Biol* 173: 53-64.
120. Rowley, A. H., S. T. Shulman, et al. (2005). "Cloning the arterial IgA antibody response during acute Kawasaki disease." *J Immunol* 175(12): 8386-8391.
121. Ruch, T. R. and C. E. Machamer (2012). "A single polar residue and distinct membrane topologies impact the function of the infectious bronchitis coronavirus E protein." *PLoS Pathog* 8(5): e1002674.
122. Ruch, T. R. and C. E. Machamer (2012). "The coronavirus E protein: assembly and beyond." *Viruses* 4(3): 363-382.
123. Schelle, B., N. Karl, et al. (2005). "Selective replication of coronavirus genomes that express nucleocapsid protein." *J Virol* 79(11): 6620-6630.
124. Shen, S., Z. L. Wen, et al. (2003). "Emergence of a coronavirus infectious bronchitis virus mutant with a truncated 3b gene: functional characterization of the 3b protein in pathogenesis and replication." *Virology* 311(1): 16-27.
125. Shi, S. T. and M. M. Lai (2005). "Viral and cellular proteins involved in coronavirus replication." *Curr Top Microbiol Immunol* 287: 95-131.
126. Shimizu, C., H. Shike, et al. (2005). "Human coronavirus NL63 is not detected in the respiratory tracts of children with acute Kawasaki disease." *J Infect Dis* 192(10): 1767-1771.

127. SIB, E. B. R. P. (2012). Procite. 2012. Swiss Institute of Bioinformatics, Nucleic Acids Res, 40(W1):W597-W603, 2012.
128. SIB, E. B. R. P. (2012). Procite. 2012. Swiss Institute of Bioinformatics, Nucleic Acids Res, 40(W1):W597-W603, 2012.
129. Smuts, H. (2008). "Human coronavirus NL63 infections in infants hospitalised with acute respiratory tract infections in South Africa." *Influenza Other Respi Viruses* 2(4): 135-138.
130. Stadler, K., V. Masignani, et al. (2003). "SARS--beginning to understand a new virus." *Nat Rev Microbiol* 1(3): 209-218.
131. Stanley G. Sawicki, D. L. S., and Stuart G. Siddell (2007). " A contemporary view of coronavirus transcription."
132. Stojdl, D. F. and J. C. Bell (1999). "SR protein kinases: the splice of life." *Biochem Cell Biol* 77(4): 293-298.
133. Surjit, M., R. Kumar, et al. (2005). "The severe acute respiratory syndrome coronavirus nucleocapsid protein is phosphorylated and localizes in the cytoplasm by 14-3-3-mediated translocation." *J Virol* 79(17): 11476-11486.
134. Suzuki, A., M. Okamoto, et al. (2005). "Detection of human coronavirus-NL63 in children in Japan." *Pediatr Infect Dis J* 24(7): 645-646.
135. Timani, K. A., Q. Liao, et al. (2005). "Nuclear/nucleolar localization properties of C-terminal nucleocapsid protein of SARS coronavirus." *Virus Res* 114(1-2): 23-34.
136. Tooze, J., S. Tooze, et al. (1984). "Replication of coronavirus MHV-A59 in sac- cells: determination of the first site of budding of progeny virions." *Eur J Cell Biol* 33(2): 281-293.

137. Torres, J., K. Parthasarathy, et al. (2006). "Model of a putative pore: the pentameric alpha-helical bundle of SARS coronavirus E protein in lipid bilayers." *Biophys J* 91(3): 938-947.
138. Torres, J., U. Maheswari, et al. (2007). "Conductance and amantadine binding of a pore formed by a lysine-flanked transmembrane domain of SARS coronavirus envelope protein." *Protein Sci* 16(9): 2065-2071.
139. Tyrrell, D. A., M. L. Bynoe, et al. (1968). "Cultivation of "difficult" viruses from patients with common colds." *Br Med J* 1(5592): 606-610.
140. Vabret, A., T. Mourez, et al. (2005). "Human coronavirus NL63, France." *Emerg Infect Dis* 11(8): 1225-1229.
141. van der Hoek, L. (2007). "Human coronaviruses: what do they cause?" *Antivir Ther* 12(4 Pt B): 651-658.
142. van der Hoek, L. and B. Berkhout (2005). "Questions concerning the New Haven coronavirus." *J Infect Dis* 192(2): 350-351; author reply 353-354.
143. van der Hoek, L., G. Ithorst, et al. (2010). "Burden of disease due to human coronavirus NL63 infections and periodicity of infection." *J Clin Virol* 48(2): 104-108.
144. van der Hoek, L., K. Pyrc, et al. (2004). "Identification of a new human coronavirus." *Nat Med* 10(4): 368-373.
145. van der Hoek, L., K. Pyrc, et al. (2006). "Human coronavirus NL63, a new respiratory virus." *FEMS Microbiol Rev* 30(5): 760-773.
146. van der Hoek, L., K. Sure, et al. (2005). "Croup is associated with the novel coronavirus NL63." *PLoS Med* 2(8): e240.
147. van der Hoek, L., K. Sure, et al. (2006). "Human coronavirus NL63 infection is associated with croup." *Adv Exp Med Biol* 581: 485-491.

148. van der Meer, Y., E. J. Snijder, et al. (1999). "Localization of mouse hepatitis virus nonstructural proteins and RNA synthesis indicates a role for late endosomes in viral replication." *J Virol* 73(9): 7641-7657.
149. Vennema, H., G. J. Godeke, et al. (1996). "Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes." *EMBO J* 15(8): 2020-2028.
150. Weiss, S. R. and S. Navas-Martin (2005). "Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus." *Microbiol Mol Biol Rev* 69(4): 635-664.
151. White, T. C., Z. Yi, et al. (2007). "Identification of mouse hepatitis coronavirus A59 nucleocapsid protein phosphorylation sites." *Virus Res* 126(1-2): 139-148.
152. Wilson, L., C. McKinlay, et al. (2004). "SARS coronavirus E protein forms cation-selective ion channels." *Virology* 330(1): 322-331.
153. Wilson, L., P. Gage, et al. (2006). "Validation of coronavirus E proteins ion channels as targets for antiviral drugs." *Adv Exp Med Biol* 581: 573-578.
154. Wisniewski, J. R., Z. Szewczuk, et al. (1999). "Constitutive phosphorylation of the acidic tails of the high mobility group 1 proteins by casein kinase II alters their conformation, stability, and DNA binding specificity." *J Biol Chem* 274(29): 20116-20122.
155. Woo, P. C., S. K. Lau, et al. (2005). "Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia." *J Virol* 79(2): 884-895.

156. Wootton, S. K., R. R. Rowland, et al. (2002). "Phosphorylation of the porcine reproductive and respiratory syndrome virus nucleocapsid protein." *J Virol* 76(20): 10569-10576.
157. Wu, C. H., S. H. Yeh, et al. (2009). "Glycogen synthase kinase-3 regulates the phosphorylation of severe acute respiratory syndrome coronavirus nucleocapsid protein and viral replication." *J Biol Chem* 284(8): 5229-5239.
158. Wu, P. S., L. Y. Chang, et al. (2008). "Clinical manifestations of human coronavirus NL63 infection in children in Taiwan." *Eur J Pediatr* 167(1): 75-80.
159. Wurm, T., H. Chen, et al. (2001). "Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division." *J Virol* 75(19): 9345-9356.
160. Yang, Y., Z. Xiong, et al. (2005). "Bcl-xL inhibits T-cell apoptosis induced by expression of SARS coronavirus E protein in the absence of growth factors." *Biochem J* 392(Pt 1): 135-143.
161. Ye, Y. and B. G. Hogue (2007). "Role of the coronavirus E viroporin protein transmembrane domain in virus assembly." *J Virol* 81(7): 3597-3607.
162. You, J., B. K. Dove, et al. (2005). "Subcellular localization of the severe acute respiratory syndrome coronavirus nucleocapsid protein." *J Gen Virol* 86(Pt 12): 3303-3310.
163. Yount, B., K. M. Curtis, et al. (2000). "Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model." *J Virol* 74(22): 10600-10611.
164. Yount, B., K. M. Curtis, et al. (2003). "Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus." *Proc Natl Acad Sci U S A* 100(22): 12995-13000.

165. Yount, B., M. R. Denison, et al. (2002). "Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59." *J Virol* 76(21): 11065-11078.
166. Yuan, Q., Y. Liao, et al. (2006). "Biochemical evidence for the presence of mixed membrane topologies of the severe acute respiratory syndrome coronavirus envelope protein expressed in mammalian cells." *FEBS Lett* 580(13): 3192-3200.



Appendix



Appendix 1

```

SARS      : -----*-----20-----*-----40-----*-----60-----*-----80-----*-----100----- : -
SARS_FL   : ATGTCTGATAATGGACCCCAATCAAACCAACGTAGTGCCCCCGCATTACATTTGGTGGACCCACAGATTCAACTGACAATAACAGAAATGGAGGACGCAATGGGGCA : 108

SARS      : -----*-----120-----*-----140-----*-----160-----*-----180-----*-----200-----*----- : 16
SARS_FL   : AGGCCAAAACAGGGCGACCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACAGCTCTCACTCAGCATGGCAAGGAGAACTTAGATTCCCTGGAGGCCAGGGC : 216
                                         CC C GG C G

SARS      : -----*-----220-----*-----240-----*-----260-----*-----280-----*-----300-----*-----320----- : 121
SARS_FL   : CCAATTTCCACGCCAATAGTGGTCCARATGACCA--TTGGCTACTACCGAAGAGCTACC--GACGAGTTCGTGGTGGTGACGGCAAAATGGAAGAGCTCAGCCCCAGA : 324
                                         CA C CCAATAGTGGTCCAGATGACCA TTGGCTACTACCGAAGAGCTACC GACGAGTTCGTGGTGGTGACGGCAAAATGGAAGAGCTCAGCCCCAGA

SARS      : -----*-----340-----*-----360-----*-----380-----*-----400-----*-----420-----*----- : 228
SARS_FL   : TGGTACTTCTTATTACTAGGAAGTGGCC--AGAAGCTTCACTTCCCTACGGCGCTAC--AAAGAAGGCATCGTATGGGTTGCAACTGAGGGAGCCCTTGAATACACCCAA : 431
                                         T CTAGGAAGTGGCC AGAAGCTTCACTT CCCTACGGCGCTA AAAGAAGGCATCGTATGGGTTGCAACTGAGGGAGCCCTTGAATACACCCAA

SARS      : -----*-----440-----*-----460-----*-----480-----*-----500-----*-----520-----*-----540----- : 336
SARS_FL   : AGACCACATTTGGCACCCCGAATCCTAATAACAATGCTGCCACCGTCTACAACCTCCCAAGGAACAC--CATTGCCAAAAGGCTTCTACGCAGAGGGGAGCAGAGGGCG : 539
                                         AGACCACATTTGGCACCCCGAATCCTAATAACAATGCTGCCACCGTCTACAACCTCCCAAGGAACAC CATTGCCAAAAGGCTTCTACGCAGAGGGGAGCAGAGGGCG

SARS      : -----*-----560-----*-----580-----*-----600-----*-----620-----*-----640-----*----- : 444
SARS_FL   : CAGTCAAGCCTCTTCTCGCTCCTCATCACGTAGTCCGGTAATTCAGAAGAAATCAACTCCTGGCAGCAGTAGGGGAAATTCCTCTGCTCGAATGGCTAGCGGAGGTGG : 647
                                         CAGTCAAGCCTCTTCTCGCTCCTCATCACGTAGTCCGGTAATTCAGAAGAAATCAACTCCTGGCAGCAGTAGGGGAAATTCCTCTGCTCGAATGGCTAGCGGAGGTGG

SARS      : -----*-----660-----*-----680-----*-----700-----*-----720-----*-----740-----*----- : 552
SARS_FL   : TGAAGTGCCTCGCGCTATTGCTGCTAGACAGATTGAACACGCTTGACAGCAAAGTTTCTGGTAAAGGCCAACACAACAAGGCCAAACTGTCACTAAGAAATCTGG : 755
                                         TGAAGTGCCTCGCGCTATTGCTGCTAGACAGATTGAACACGCTTGACAGCAAAGTTTCTGGTAAAGGCCAACACAACAAGGCCAAACTGTCACTAAGAAATCTGG

SARS      : -----*-----760-----*-----780-----*-----800-----*-----820-----*-----840-----*-----860----- : 660
SARS_FL   : TGCTGAGGCATCTAAAAAGCCTCGCCAAAAAGTACTGCCAAAAACAGTACAACGCTCACTCAAGCATTGGGAGACGTGGTCCAGAACAAACCAAGGAAATTTCCG : 863
                                         TGCTGAGGCATCTAAAAAGCCTCGCCAAAAAGTACTGCCAAAAACAGTACAACGCTCACTCAAGCATTGGGAGACGTGGTCCAGAACAAACCAAGGAAATTTCCG

SARS      : -----*-----880-----*-----900-----*-----920-----*-----940-----*-----960-----*----- : 768
SARS_FL   : GGACCAAGACCTAATCAGACAAGGAAGTATTCAAAACATTTGGCCGCAAAATGGCACAATTTGCTCCAAAGTCCCTCTGCATTCTTTGGAATGTCACGCATTTGGCATGGA : 971
                                         GGACCAAGACCTAATCAGACAAGGAAGTATTCAAAACATTTGGCCGCAAAATGGCACAATTTGCTCCAAAGTCCCTCTGCATTCTTTGGAATGTCACGCATTTGGCATGGA

SARS      : -----*-----980-----*-----1000-----*-----1020-----*-----1040-----*-----1060-----*-----1080----- : 876
SARS_FL   : AGTACACCTTCGGGAACATGGCTGACTTATCATGGAGCCATTAATTTGGATGACAAAAGATCCACAATTCAAAAGACAAGTCACTACTGCTGAACAAGCAGCATTTGACGC : 1079
                                         AGTACACCTTCGGGAACATGGCTGACTTATCATGGAGCCATTAATTTGGATGACAAAAGATCCACAATTCAAAAGACAAGTCACTACTGCTGAACAAGCAGCATTTGACGC

SARS      : -----*-----1100-----*-----1120-----*-----1140-----*-----1160-----*-----1180-----*----- : 984
SARS_FL   : ATACAAAACATTTCCACCAACAGAGCCTAAAAAGGACAAAAAGAAAAGACTGATGAAGCTCAGCCTTTGCCGACAGACAAAAGAAGCAGCCCACTGTGACTCTTCT : 1187
                                         ATACAAAACATTTCCACCAACAGAGCCTAAAAAGGACAAAAAGAAAAGACTGATGAAGCTCAGCCTTTGCCGACAGACAAAAGAAGCAGCCCACTGTGACTCTTCT

SARS      : -----*-----1200-----*-----1220-----*-----1240-----*-----1260-----*-----1280-----*----- : 1092
SARS_FL   : TCCTGGCGGTGACATGGATGATTTCTCCAGACAACCTTCAAATTCATGAGTGGAGCTTCTGCTGATTCAACTCAGGCATAA----- : 1269
                                         TCCTGGCGGTGACATGGATGATTTCTCCAGACAACCTTCAAATTCATGAGTGGAGCTTCTGCTGATTCAACTCAGGCATAA

```

Appendix 1: Sequence verification of the cloned SARS N-gene.

Cloned genes were sequenced with the T7 terminator primer. T7 primer sites lie downstream of the cloning site and upstream of the GST fusion tag sequence. As the polymerase can only read ± 800 bp into the sequence, the forward primer is not suitable as only the GST gene would be sequenced (Data is relevant for appendix 1-5). SARS_FL represents the sequence for the SARS N-gene obtained from ncbi with accession number AY360146.1, where SARS represents the sequence of the cloned SARS N-gene.

Appendix 2

```

N1      : -----*-----20-----*-----40-----*-----60-----*-----80-----*-----100-----*
N1_FL   : ATGGCTAGTGTAAATTTGGGCGATGACAGAGCTGCTAGGAAGAAATTTCCCTCCCTTCATTTTACAGGCCTCTCTTTTGGTTACTTCTGATAAGGCACCATA-TAGGGTC : 108
          CTCT  G T  TTCTGATAAGGCACC ATA TAGGGTC

N1      : ATCC-----AGATCTGTCC-----TATTTGGTAAGGGTAATAAAGATGACGAGATTGGTTATTGGAATGTTCAAGAGCCTTGGCGTATGCGCAGGGGGCAACGTGTTGATTTGCC : 145
N1_FL   : ATCCAGGAAATCTGTCCCTATTTGGTAAGGGTAATAAAGATGACGAGATTGGTTATTGGAATGTTCAAGAGCCTTGGCGTATGCGCAGGGGGCAACGTGTTGATTTGCC : 218
          AT CC A TGTCC TATTTGGTAAGGGTAATAAAGATGACGAGATTGGTTATTGGAATGTTCAAGAGCCTTGGCGTATGCGCAGGGGGCAACGTGTTGATTTGCC

N1      : TCCTAAAGTTCATTTTTATTACCTAGTACTGGACCTCATAAGGACCTTAAATTCAGACAACGTTCTGATGGTGTGTTTGGGTGCTAAGGAAGGTGCTAAAACCTGTTA : 255
N1_FL   : TCCTAAAGTTCATTTTTATTACCTAGTACTGGACCTCATAAGGACCTTAAATTCAGACAACGTTCTGATGGTGTGTTTGGGTGCTAAGGAAGGTGCTAAAACCTGTTA : 328
          TCCTAAAGTTCATTTTTATTACCTAGTACTGGACCTCATAAGGACCTTAAATTCAGACAACGTTCTGATGGTGTGTTTGGGTGCTAAGGAAGGTGCTAAAACCTGTTA

N1      : ATACCAGTCTGGTAATCGCAACGTAATCAGAAACCTTTGGAACCAAAGTTCCTATTGCTTTGCCCTCCAGAGCTCTCTGTTGTTGAGTTTGGAGATCGCTCTAATAAC : 365
N1_FL   : ATACCAGTCTGGTAATCGCAACGTAATCAGAAACCTTTGGAACCAAAGTTCCTATTGCTTTGCCCTCCAGAGCTCTCTGTTGTTGAGTTTGGAGATCGCTCTAATAAC : 438
          ATACCAGTCTGGTAATCGCAACGTAATCAGAAACCTTTGGAACCAAAGTTCCTATTGCTTTGCCCTCCAGAGCTCTCTGTTGTTGAGTTTGGAGATCGCTCTAATAAC

N1      : TCATCTCGGTAGCAGTCTGTTCTTCAACTCGTAACTCAGGAGACTTCTCTGATGCACTTCAAGACAACAGTCTCGCACTCTCTGATTTCTAACCAGTCTTCTTC : 475
N1_FL   : TCATCTCGGTAGCAGTCTGTTCTTCAACTCGTAACTCAGGAGACTTCTCTGATGCACTTCAAGACAACAGTCTCGCACTCTCTGATTTCTAACCAGTCTTCTTC : 548
          TCATCTCGGTAGCAGTCTGTTCTTCAACTCGTAACTCAGGAGACTTCTCTGATGCACTTCAAGACAACAGTCTCGCACTCTCTGATTTCTAACCAGTCTTCTTC

N1      : AGATCTTGTGTGCTGTTACTTTGGCTTTAAAGAAGTCTAGGTTTGTATAACAGTCTGAAAGTCACTAGTTCTTCTGGTACTTCCACTCCTAAGAAACCTAATAAGCCTC : 585
N1_FL   : AGATCTTGTGTGCTGTTACTTTGGCTTTAAAGAAGTCTAGGTTTGTATAACAGTCTGAAAGTCACTAGTTCTTCTGGTACTTCCACTCCTAAGAAACCTAATAAGCCTC : 658
          AGATCTTGTGTGCTGTTACTTTGGCTTTAAAGAAGTCTAGGTTTGTATAACAGTCTGAAAGTCACTAGTTCTTCTGGTACTTCCACTCCTAAGAAACCTAATAAGCCTC

N1      : TTTCTCAACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGGTTCCCTACCAGAGGAAAATGTTTATTCAGTCTTTGGTCCCTCGTGAATTTAA : 695
N1_FL   : TTTCTCAACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGGTTCCCTACCAGAGGAAAATGTTTATTCAGTCTTTGGTCCCTCGTGAATTTAA : 768
          TTTCTCAACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGGTTCCCTACCAGAGGAAAATGTTTATTCAGTCTTTGGTCCCTCGTGAATTTAA

N1      : CACAATATGGGGGATTCAGATCTTGTTCAGAAATGGTGTGATGCCAAAGGTTTCCACAGCTTGTGTAATGATTCCTAATCAGGCTGCGTTATTCTTTGATAGTAGGGT : 805
N1_FL   : CACAATATGGGGGATTCAGATCTTGTTCAGAAATGGTGTGATGCCAAAGGTTTCCACAGCTTGTGTAATGATTCCTAATCAGGCTGCGTTATTCTTTGATAGTAGGGT : 878
          CACAATATGGGGGATTCAGATCTTGTTCAGAAATGGTGTGATGCCAAAGGTTTCCACAGCTTGTGTAATGATTCCTAATCAGGCTGCGTTATTCTTTGATAGTAGGGT

N1      : TAGCACTGATGAAGTGTGATAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGATAATAAAGAACCTTCTAAGTTCAATTGAGCAGATTAGTGTCTTTA : 915
N1_FL   : TAGCACTGATGAAGTGTGATAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGATAATAAAGAACCTTCTAAGTTCAATTGAGCAGATTAGTGTCTTTA : 988
          TAGCACTGATGAAGTGTGATAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGATAATAAAGAACCTTCTAAGTTCAATTGAGCAGATTAGTGTCTTTA

N1      : CTAAACCCAGTTCTATCAAAAGAAATGCAGTCACAAATCAATCTCATGTTGCTCAGAACACAGTACTTAATGCTTCTATCCAGAAATCTAAACCATTTGGCTGATGATGATTC : 1025
N1_FL   : CTAAACCCAGTTCTATCAAAAGAAATGCAGTCACAAATCAATCTCATGTTGCTCAGAACACAGTACTTAATGCTTCTATCCAGAAATCTAAACCATTTGGCTGATGATGATTC : 1098
          CTAAACCCAGTTCTATCAAAAGAAATGCAGTCACAAATCAATCTCATGTTGCTCAGAACACAGTACTTAATGCTTCTATCCAGAAATCTAAACCATTTGGCTGATGATGATTC

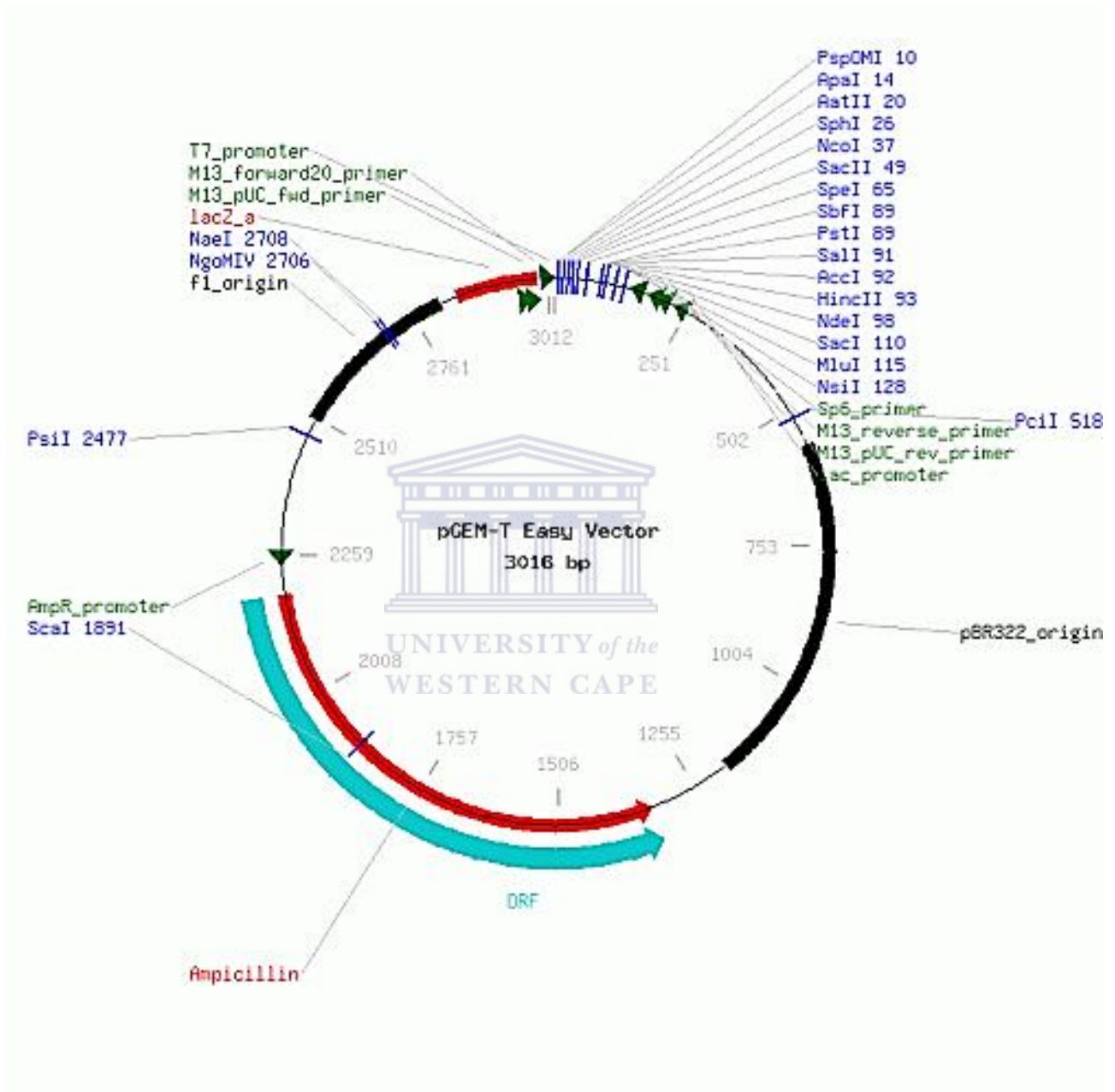
N1      : GCCATTATAGAAATTTGCAACGAGGTTTTGCATTAAGTTTAAACGAATTCGGGTCGGTACCCGGGGATCCCTCTAGAGTCGACCTGCAGGCATGCAAGCTGATCCRKTCTG : 1135
N1_FL   : GCCATTATAGAAATTTGCAACGAGGTTTTGCATTAAGTTTAAACGAATTCGGGTCGGTACCCGGGGATCCCTCTAGAGTCGACCTGCAGGCATGCAAGCTGATCCRKTCTG : 1134
          GCCATTATAGAAATTTGCAACGAGGTTTTGCATTAAGTTTAAACGAATTCGGGTCGGTACCCGGGGATCCCTCTAGAGTCGACCTGCAGGCATGCAAGCTGATCCRKTCTG
    
```

Appendix 2: Sequence verification of the NL63 full length N-gene

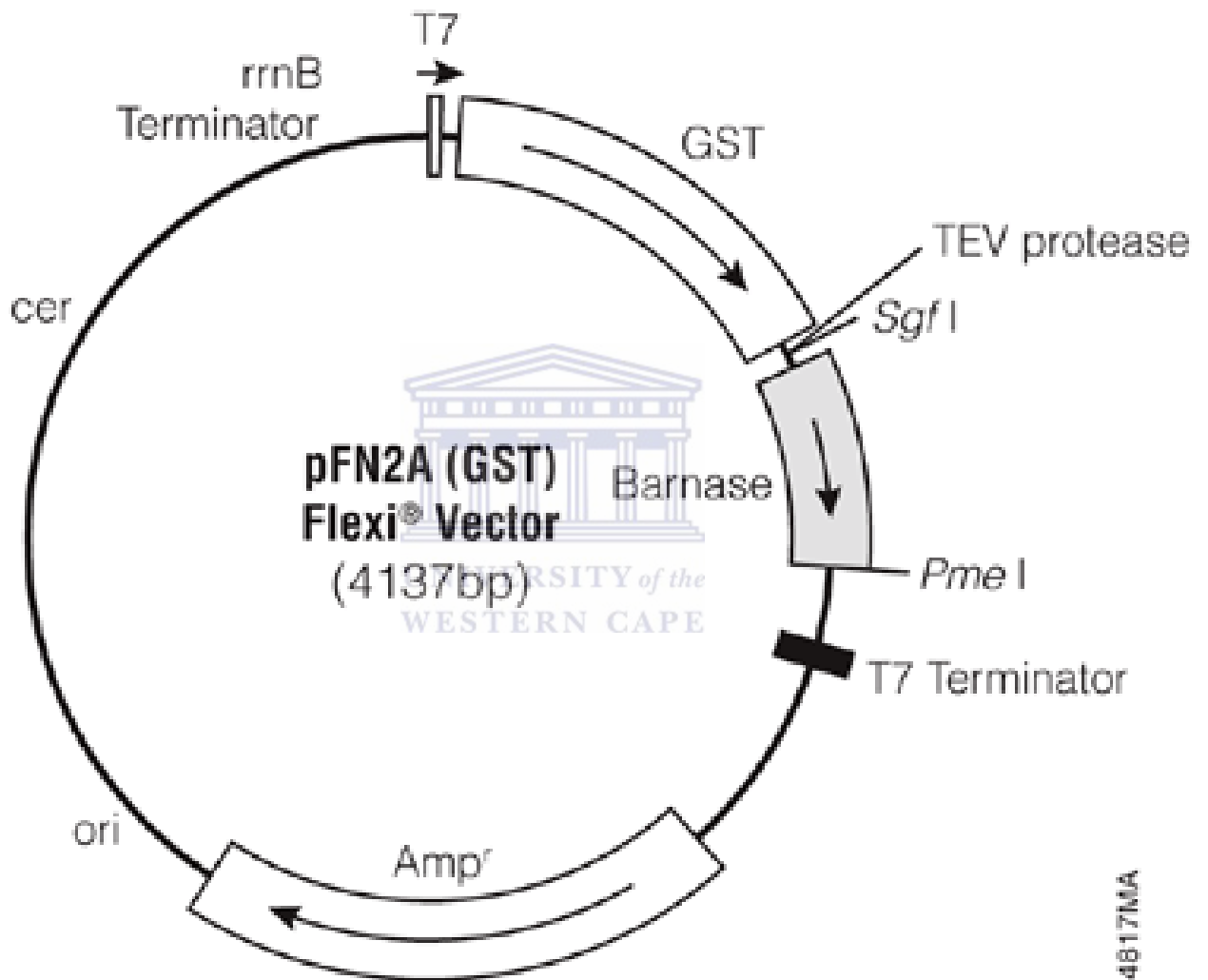
NL63_FL represents the sequence of the NL63 N-gene obtained from NCBI with accession number

DQ846901.1, where N1 represents the sequence of the cloned NL63 N-gene.

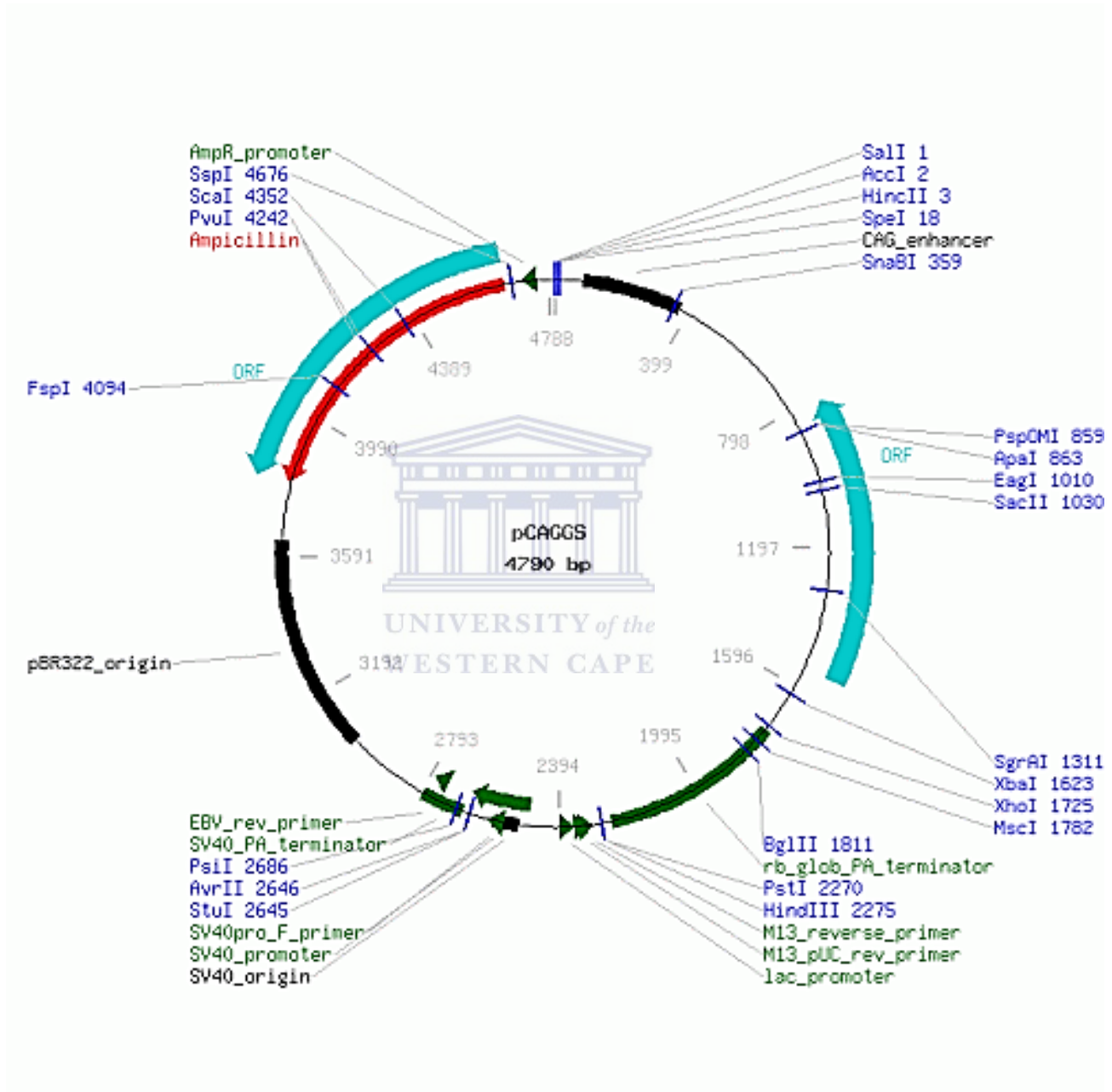
Appendix 3



Appendix 4



Appendix 5



Appendix 6

Amino Acid Name	One Letter Code	Hydropathy Score
Isoleucine	I	4.5
Valine	V	4.2
Leucine	L	3.8
Phenylalanine	F	2.8
Cysteine	C	2.5
Methionine	M	1.9
Alanine	A	1.8
Glycine	G	-0.4
Threonine	T	-0.7
Tryptophan	W	-0.9
Serine	S	-0.8
Tyrosine	Y	-1.3
Proline	P	-1.6
Histidine	H	-3.2
Glutamic acid	E	-3.5
Glutamine	Q	-3.5
Aspartic acid	D	-3.5
Asparagine	N	-3.5
Lysine	K	-3.9
Arginine	R	-4.5

Appendix 9: Kyte-Doolittle Hydropathy scores. These scores are based on the values given by the original Kyte-Doolittle paper (Kyte and Doolittle 1982).