

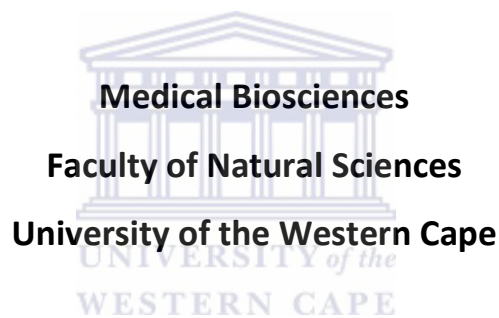
Cloning and Characterization of the Human Coronavirus NL63 Nucleocapsid Protein

by

Michael Berry

Thesis Submitted in Fulfilment of the Requirements for the Degree

MSc



Supervisor: Professor BC Fielding

Department of Medical Biosciences
University of the Western Cape

Co-Supervisor: Prof YJ Tan

Department of Microbiology
Yong Loo Lin School of Medicine
National University of Singapore

November 2011

DECLARATION

I, Michael Berry, declare that this thesis, "***Cloning and Characterization of the Human Coronavirus NL63 Nucleocapsid Protein***" 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.

Michael Berry :

Date Signed :



ACKNOWLEDGEMENTS

- I would like to express my gratitude to Prof B.C. Fielding for his supervision during my MSc and his constant insight and motivation.
- I would like to thank Prof. Y-J. Tan for her co-supervision and Prof. E. Pool for his collaboration on research conducted.
- I sincerely thank the members of Molecular Virology lab, with especial recognition to the senior students, namely Randal Fisher and Tasnim Suliman, for their mentoring and support. To the remaining students, namely Tarryn-Lee Manasse, Aasiyah Chafekar, Anele Gela, Yanga Mnyamana, Thato Motlhalamme, Tiza Nguni and Chanel Swartz for your help rendered, sharing of equipment and reagents and the interesting discussions raised about my work.
- Thank you to the NRF for personal and research funds which made this MSc possible.
- To the University of the Western Cape and especially to the Department of Medical Bioscience, thank you for allowing me the opportunity to complete all my studies, from first year of undergrad to my MSc.
- I would like to thank all the staff and students of the Department of Medical Bioscience for creating an enjoyable working environment.
- Finally, to my parents, girlfriend, Erin Fielding, and friends, a heart-felt thank you for your encouragement and support!

ABSTRACT

The human coronavirus NL63 was discovered in 2004 by a team of researchers in Amsterdam. Since its discovery it has been shown to have worldwide spread and affects mainly children, aged 0-5 years old, the immunocompromised and the elderly. Infection with HCoV-NL63 commonly results in mild upper respiratory tract infections and presents as the common cold, with symptoms including fever, cough, sore throat and rhinorrhoea. Lower respiratory tract findings are less common but may develop into more serious complications including bronchiolitis, pneumonia and croup.

The primary function of the HCoV-NL63 nucleocapsid (N) protein is the formation of the protective ribonucleocapsid core. For this particle to assemble, the N-protein undergoes N-N dimerization and then interacts with viral RNA. Besides the primary structural role of the N-protein, it is also understood to be involved in viral RNA transcription, translation and replication, including several other physiological functions. The N-protein is also highly antigenic and elicits a strong immune response in infected patients. For this reason the N-protein may serve as a target for the development of diagnostic assays.

We have used bioinformatic analysis to analyze the HCoV-NL63 N-protein and compared it to coronavirus N-homologues. This bioinformatic analysis provided the data to generate recombinant clones for expression in a bacterial system. We constructed recombinant clones of the N-protein of SARS-CoV and HCoV-NL63 and synthesized truncated clones corresponding to the N- and C-terminal of the HCoV-NL63 N-protein. These heterologously expressed proteins will serve the basis for several post-expression studies including characterizing the immunogenic epitope of the N-protein as well identifying any antibody cross-reactivity between coronavirus species.

KEY WORDS

Human Coronavirus NL63, Nucleocapsid Protein, Severe Acute Respiratory Syndrome Coronavirus, Bioinformatic Analysis, Disordered Regions, Cloning and Bacterial Expression, Antigenic Epitopes, Antibody Cross-reactivity



TABLE OF CONTENTS

	Declaration	ii
	Acknowledgements	iii
	Abstract	iv
	Keywords	v
	Table of Contents	vi
	List of Abbreviations	xi
	List of Tables	xvii
	List of Figures	xviii
	List of Appendixes	xx
	Presentations and Publications	xxi
Chapter 1	The Human Coronavirus NL63: A Review	1
	Introduction	2
1	Outbreak of SARS-CoV: Stimulating Coronavirus Research	4
2	HCoV-NL63: A Novel Respiratory Virus	6
2.1	Discovery of HCoV-NL63	6
2.2	Clinical Findings and Prevalence of HCoV-NL63	9
2.3	Association of HCoV-NL63 with Croup	11
2.4	Association of HCoV-NL63 with Non-respiratory Findings	12
2.4.1	Kawasaki Disease	12
2.4.2	Gastrointestinal Disease	14

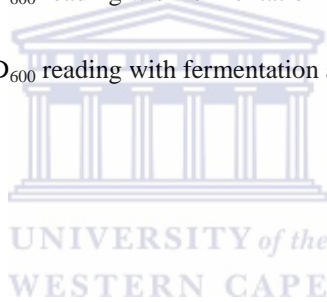


2.5	Co-infection of HCoV-NL63 and other respiratory viruses	15
3	Coronavirus Genome, Transcription and Replication	16
4	HCoV-NL63 Cellular Entry by ACEII	20
5	The Coronavirus Nucleocapsid Protein	22
5.1	Multimerization of Coronavirus N	23
5.2	Intracellular Localization of the Nucleocapsid Protein	24
5.3	Phosphorylation of the N-Protein	25
5.4	Perturbation of Host Cell Processes by the N-protein	26
5.4.1	Deregulation of host cell-cycle	27
5.4.2	Inhibition of interferon production	27
5.4.3	Up-regulation of COX-2	28
5.4.4	Up-regulation of the activator protein (AP) 1 pathway	29
5.4.5	Induction of apoptosis in serum starved COS1-cells	29
5.4.6	Binding of N-protein to cyclophilin A	30
	Aims of this thesis	30
Chapter 2	Bioinformatic Analysis of the HCoV-NL63 Nucleocapsid Protein	33
1	Abstract	34
2	Introduction	35
3	Materials and Method	39
3.1	Multiple Sequence Alignment	39
3.2	Order-Disorder Prediction	39
3.3	Phosphorylation State Prediction	39

3.4	Characterization of Antigenic State	40
4	Result and Discussion	41
4.1	Multiple Sequence Alignment	41
4.1.1	Nucleocapsid Dimerization	45
4.1.2	N-M Heterotypic Interactions	46
4.2	Order/Disorder	46
4.3	Prediction of Phosphorylation Sites	49
4.4	Characterization of the antigenic state	52
5	Conclusion	53
Chapter 3	Cloning and Expression of the HCoV-NL63 Nucleocapsid Protein	56
1	Abstract	57
2	Introduction	58
3	Materials and Methods	63
3.1	Amplification of Full Length and Deletion Mutant Nucleocapsid Genes	63
3.1.1	RT-PCR	63
3.1.2	Design of PCR Primers	63
3.1.3	Polymerase Chain Reaction (PCR) and Purification of Amplified Genes	64
3.2	Cloning and Verification of Amplification Products	66
3.2.1	Ligation into pGEM [®] T-easy Vector	66
3.2.2	Preparation of culture media	67
3.2.3	Transformation of JM109 <i>E.coli</i> and Screening for Recombinant Clones	67
3.2.4	Plasmid Extraction	68



3.2.5	Restriction Endonuclease Digestion	69
3.2.6	Nucleotide Sequencing and Sequence Analysis	70
3.3	Ligation into Flexi™ Vector and Transformation of KRX Competent <i>E.coli</i>	70
3.3.1	Restriction with Flexi™ Enzymes (<i>SgfI</i> and <i>PmeI</i>)	70
3.3.2	Ligation into Flexi™ Vector	71
3.3.3	Preparation of Chemically Competent <i>E. coli</i> , KRX Strain, and Transformation with Recombinant Plasmid	71
3.4	Expression of Full-Length SARS-CoV and HCoV-NL63 N and Deletion Mutants in <i>E. coli</i>	73
3.4.1	Autoinduction	73
3.4.2	Induction at specific OD ₆₀₀ reading with fermentation at 37°C	73
3.4.3	Induction at specific OD ₆₀₀ reading with fermentation at 25°C	74
3.4.4	Time course expression	74
3.4.5	Large scale expression	75
3.5	Protein Analysis	75
3.5.1	Protein extraction and Preparation of Cell Lysates	75
3.5.2	SDS-Polyacrylamide Gel Electrophoresis and Coomassie	77
3.5.3	Western Blotting of Total Proteins	78
3.6	Purification of Recombinant Protein	78
3.7	Cleavage of Fusion Tag and Purification of Native Protein	79
4	Results	81
4.1	PCR	81
4.2	Colony PCR of JM109 Competent <i>E. coli</i>	82
4.3	<i>EcoRI</i> Digest of pGEM®-N	83



4.4	<i>SgfI</i> and <i>PmeI</i> Restriction	84
4.5	Colony PCR of KRX Competent <i>E. coli</i>	85
4.6	Bacterial Expression and Purification of Recombinant Viral Proteins	87
4.7	Cleavage of Fusion Tag	92
5	Discussion and Conclusion	93
Chapter 4	Using an ELISA to Map the Antigenic Epitopes of HCoV-NL63 Nucleocapsid Protein: A Pilot Study	97
1	Abstract	98
2	Introduction	99
3	Materials and Methods	102
3.1	Patient serum samples	102
3.2	ELISA	102
3.3	Statistical Analysis	103
4	Results	105
5	Discussion and Conclusion	110
Chapter 5	Synopsis	114
	References	120
	Appendix	136



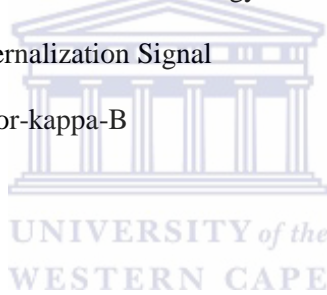
LIST OF ABBREVIATIONS

~	Approximately
±	plus/minus
®	Registered
°C	degrees celsius
µM	micro molars
14-3-3	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein
14-3-3θ	Theta isoform of tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein
3'	Three Prime
5'	Five Prime
A	Adenine
aa	amino acid
ACE	Angiotensin Converting Enzyme
AFLP	Amplified Fragment Length Polymorphism
AmpR	Ampicillin Resistance Gene
AP1	activator protein 1
ATF2	Activating Transcription Factor 2
ATP	Adenosine Triphosphate
Bcl2	B-cell lymphoma 2
BM	Biomass
bp	base pairs
BPer	Bacterial protein extraction reagents
BSA	Bovine Serum Albumin
C	Cytosine
cAMP	Cyclic adenosine monophosphate
CCAAT	cytidine-cytidine-adenosine-adenosine-thymidine

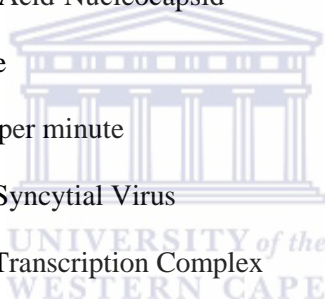
CD13	Aminopeptidase N
CDK	Cyclin dependant kinase
cDNA	Complementary Deoxyribonucleic Acid
cDNA-AFLP	cDNA amplified restriction fragment-length polymorphism
cGMP	cyclic guanosine monophosphate
CK	Casein kinase
CMV	Cytomegalovirus
CoV	Coronavirus
COX-2	Cyclo oxygenase-2
CREB-1	cAMP Responsive Element Binding Protein 1
CRM-1	Chromosome Region Maintenance 1
C-terminal	Carboxyl Terminal
DAS	Double Antibody Sandwich
DEPP	Disorder Enhanced Phosphorylation Predictor
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphate
DTT	Dithiothreitol
E	Envelope
<i>E. coli</i>	<i>Escherichia coli</i>
<i>EcoRI</i>	<i>Escherichia coli</i> Restriction Enzyme 1
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme Linked Immunosorbent Assay
ER	Endoplasmic reticulum
Erk	Extracellular signal-regulated kinases
<i>et al.</i>	<i>Et alii</i> (and others)
FIPV	Feline infectious peritonitis virus
g	grams
G	Guanine

GSK	Glycogen Synthase Kinase
GST	Glutathione S Transferase
GuHCl	Guanidine Hydrochloride
HCoV	Human Coronavirus
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human Immunodeficiency Virus
hnRNP	Heterogeneous Nuclear Ribonucleoproteins
IBV	Infectious Bronchitis Virus
IFN	Interferon
<i>In silico</i>	Performed via computer simulation
<i>In vitro</i>	In an artificial environment
<i>In vivo</i>	Within the living
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IRF-3	Interferon regulatory factor 3
ISRE	Interferon-stimulated response element
JAK/STAT	Janus Kinase/Signal Transducer and Activator of Transcription
Jnk	c-Jun N-terminal kinases
K ₂ HPO ₄	Dipotassium phosphate
kb	Kilobases
KCl	Potassium Chloride
kDa	Kilo Daltons
KH ₂ PO ₄	Potassium Dihydrogen Phosphate
L	Liter
LB	Luria Bertani
M	Membrane
M	Molar
mA	mili-Amps
MAPK	Mitogen-activated protein kinase

mg	milligrams
MgCl ₂	Magnesium Chloride
MHV	Mouse Hepatitis Virus
min	minutes
mM	mili molar
M-N	Membrane-Nucleocapsid
N	Nucleocapsid
N/A	Not Applicable
Na ₂ HPO ₄	disodium hydrogen orthophosphate
NaCl	Sodium Chloride
NaOH	Sodium Hydroxide
NCBI	National Center for Biotechnology Information
NES	Nuclear Externalization Signal
NF-κB	Nuclear factor-kappa-B
ng	nanograms
NH	New Haven
NLS	Nuclear Localization Signal
NMR	Nuclear Magnetic Resonance
N-M-RNA	nucleocapsid-membrane-ribonucleic acid
N-N	Nucleocapsid-Nucleocapsid
nsp	Non-structural Protein
N-terminal	Amino Terminal
NΔC	Nucleocapsid delta (without) carboxy terminal
NΔN	Nucleocapsid delta (without) amino terminal
NΔNΔC	Nucleocapsid delta (without) amino and carboxy terminal
O/N	Over night
OD ₆₀₀	Optical Density at a wavelength of 600nm
ORF	Open Reading Frame



p38 MAPK	p38 Mitogen Activated Protein Kinase
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEDV	Porcine Epidemic Diarrhea Virus
pFN2A	pFlexi™ Amino Terminal 2 Ampicillin
pH	potential Hydrogen
PIV1	Parainfluenza Virus Type 1
PONDR	Predictors of Natural Disordered Regions
PVDF	Polyvinylidifluoride
Rep	RNA polymerase
RNA	Ribonucleic Acid
RNA-N	Ribonucleic Acid-Nucleocapsid
RNase	Ribonuclease
rpm	Revolutions per minute
RSV	Respiratory Syncytial Virus
RTC	Replication Transcription Complex
RT-PCR	Reverse Transcription Polymerase Chain Reaction
S	Spike
SARS	Severe Acute Respiratory Syndrome
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
Sec	Seconds
SPR	Surface Plasmon Resonance
SR	Serine-Arginine
ssDNA	Single stranded Deoxyribonucleic Acid
STAT	Signal Transducer and Activator of Transcription
T	Thymine
TBE	Tris/Borate/EDTA



TEV	Tobacco Etch Virus
TGEV	Transmissible Gastroenteritis Virus
T _m	Melting Temperature
Tris-HCl	Tris Hydrochloride
™	Trademark
u	units
USA	United States of America
UV	Ultraviolet
V	Volts
v/v	volume/volume
VIDISCA	Virus-Discovery-cDNA-AFLP
VL3	Variously characterized Long disordered regions 3
VL-XT	Variously characterized Long disordered regions - X-ray characterized Terminal disordered regions.
Vol	Volume
VSL1	Variously characterized Short and Long disordered regions
w/v	weight/volume
Xgal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside
XL1-XT	X-ray characterized Long disordered regions 1 - X-ray characterized Terminal disordered regions
α	alpha
β	beta
μl	micro liter

LIST OF TABLES

Table 2.1	Multiple sequence alignment of 10 coronavirus nucleocapsid homologues.	42
Table 2.2	Regions of interaction within the coronavirus nucleocapsid proteins.	45
Table 3.1	Description of heterologously expressed proteins.	59
Table 3.2	Forward and reverse primer sequences constructed for PCR amplification of the SARS-CoV N and HCoV-NL63 N-genes as well as truncated mutants of the HCoV-NL63 N-gene of N2, N3 and N4.	64
Table 3.3	Function of components of chosen cell lysis buffer.	76
Table 3.4	Buffers used to increase protein solubility from the membrane-bound fraction.	77
Table 3.5	OD ₆₀₀ (mg/ml) of cultures after time specified.	87
Table 3.6	Concentration of purified proteins in mg/ml.	91
Table 4.1	Recombinant proteins included in the study.	103
Table 4.2	Patient antibody response to nucleocapsid constructs.	109

LIST OF FIGURES

Figure 1.1	Phylogenetic analysis of coronaviruses.	4
Figure 1.2	VIDISCA technique used to identify the new HCoV-NL63.	7
Figure 1.3	Organization of the HCoV-NL63 genome.	17
Figure 1.4	Structure of an RNA pseudoknot.	18
Figure 2.1	Alignment of the amino acid sequences of HCoV-NL63 N and selected coronavirus N-homologues.	43
Figure 2.2	A Conserved region (blocked) within the N-terminal of the nucleocapsid protein of 4 coronaviruses.	43
Figure 2.3	Multiple sequence alignment of 4 Group II N-homologues.	44
Figure 2.4	Conserved SR rich domain (blocked) within coronavirus nucleocapsid proteins.	47
Figure 2.5	Predicted disordered regions of the SARS N-protein using PONDR algorithms.	47
Figure 2.6	Predicted disordered regions of the HCoV-NL63 N-protein using PONDR algorithm.	46
Figure 2.7	Predicted phosphorylation sites of HCoV-NL63 N-protein.	50
Figure 2.8	Predicted phosphorylation sites of SARS-CoV N-protein.	51
Figure 2.9	Antigenic profile of the SARS-CoV (top) and the HCoV-NL63 (bottom) nucleocapsid protein.	52
Figure 3.1	PCR amplification of viral genes using primers described in table 3.1.	81
Figure 3.2	Detection of viral genes in JM109 competent E. coli transformed with pGEM® vector previously ligated with viral genes. Genes were detected using PCR with primers described in Table 3.1.	83

Figure 3.3	Confirmation of successful ligation by restriction of the pGEM® vector with EcoRI enzymes.	84
Figure 3.4	Restriction of the pGEM® vector, with SgfI and PmeI, isolated from JM109 competent E. coli to remove viral genes.	85
Figure 3.5	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 3.1 and used as confirmation of successful transformation and ligation.	86
Figure 3.6	Schematic description of SARS-CoV and HCoV-NL63 N-proteins and recombinant mutants of N2, N3 and N4. Numbers indicate amino acid positions.	88
Figure 3.7	Detection of SARS-CoV and HCoV-NL63 N-proteins by Western Blot with peroxidase labeled antibodies against the GST fusion protein. (A) Proteins expressed in membrane-bound, insoluble fraction of the cell. (B) Cytoplasmic expression of soluble fusion proteins.	89
Figure 3.8	Digestion of cellular membranes using lysozymes to release membrane-bound, insoluble proteins. (A) Western Blot representing proteins which remained in the insoluble fraction following digestion with lysozymes. (B) Proteins made soluble following digestion with lysozymes.	90
Figure 4.1	Detection of rabbit anti-N-IgG antibodies using heterologously expressed N-proteins. Results are represented as serial dilutions.	105
Figure 4.2	ELISA using a confirmed negative patient serum.	106
Figure 4.3	Reaction of SARS positive serum with the heterologous expressed N-proteins.	107
Figure 4.4	ROC curve analysis indicating the sensitivity and specificity of the N-protein in detecting serum antibodies.	108

LIST OF APPENDIXES

Appendix 1	Sequence verification of the cloned SARS N-gene in pGEM. Cloned genes were sequenced with M13 forward and reverse primers.	137
Appendix 2	Sequence verification of the cloned full length HCoV-NL63 N-gene in pGEM.	138
Appendix 3	Sequence verification of the cloned 5' half (N2: 1-567bp) of the HCoV-NL63 N-gene in pGEM.	139
Appendix 4	Sequence verification of the cloned 3' half (N2: 568-1134bp) of the HCoV-NL63 N-gene in pGEM.	140
Appendix 5	Sequence verification of the cloned middle region (N4: 379-756bp) of the HCoV-NL63 N-gene in pGEM.	140
Appendix 6	Sequence verification of the cloned SARS N-gene.	141
Appendix 7	Sequence verification of the HCoV-NL63 full length N-gene.	142
Appendix 8	Sequence verification of the 5' (1-567bp) half of the HCoV-NL63 N-gene coding for the N-terminal half (N2) of the HCoV-NL63 N-protein.	143
Appendix 9	Sequence verification of the 3' (568-1134bp) half of the HCoV-NL63 N-gene coding for the C-terminal half (N3) of the HCoV-NL63 N-protein.	144
Appendix 10	Sequence verification of the middle region (379-756bp) of the HCoV-NL63 N-gene coding for the middle region (N4) of the HCoV-NL63 N-protein.	144
Appendix 11	Sequencing vector pGEM-T Easy vector.	145
Appendix 12	Bacterial protein expression vector, pFN2A Flexi™.	145

PRESENTATIONS AND PUBLICATIONS

Manuscripts in preparation from this Thesis:

1. **M. Berry** and B. C. Fielding. Analysis of human coronavirus-NL63 nucleocapsid protein.
2. **M. Berry**, Y. Tan, T. Suliman, C. Drosden, M. Muller and B. C. Fielding. Identifying the immunogenic region and antibody cross-reactivity of the NL63 nucleocapsid protein.
3. A. Chafekar, **M. Berry**, E. Pool and B. Fielding. Human coronavirus-NL63 nucleocapsid induced immune response in human whole blood cultures.

Conference Presentations from this Thesis:

1. **M. Berry** and B.C. Fielding. Characterising the domains involved in the dimerization of human coronavirus NL63 nucleocapsid protein. 194th General Meeting of the Experimental Biology Group, University of the Western Cape, 14 October 2010.
2. **M. Berry**, T. Suliman, M. Muller, C. Drosden, E. Pool, Y. J. Tan and B.C. Fielding. Detection of serum antibodies to the human coronavirus-NL63 Nucleocapsid protein. 39th Annual Conference of the Physiology Society of Southern Africa, UWC, Cape Town, 19-31 August 2011.
3. A. Chafekar, **M. Berry**, E. Pool and B.C. Fielding. Human coronavirus-NL63 nucleocapsid protein-induced immune response in human whole blood cultures. 39th Annual Conference of the Physiology Society of Southern Africa, UWC, Cape Town, 19-31 August 2011.
4. **M. Berry**, A. Chafekar, E. Pool and B.C. Fielding. The NL63 Nucleocapsid is highly antigenic and induces an IL-6 response. Annual UWC Research Open for postgraduate students, 25-25 October 2011.

Abstracts accepted for publication in Journals:

1. A. Chafekar, **M. Berry**, E. Pool and B.C. Fielding. Human coronavirus-NL63 nucleocapsid protein-induced immune response in human whole blood cultures. Scientific Research and Essays, Special Issue December 2011.
2. **M. Berry**, T. Suliman, M. Muller, C. Drosden, E. Pool, Y.J. Tan and B.C. Fielding. Detection of serum antibodies to the human coronavirus-NL63 Nucleocapsid protein. Scientific Research and Essays, Special Issue December 2011.

Awards:

1. 3rd Prize for Oral Presentation, MSc category, Annual UWC Research Open Day for postgraduate students, 25-25 October 2011.

UNIVERSITY OF THE WESTERN CAPE

Chapter 1

The Human Coronavirus NL63: A Review



Department of Medical Bioscience, University of the Western Cape, Cape Town, South Africa

Introduction

Coronaviruses (CoVs) were first identified in 1937, when the virus responsible for infectious bronchitis in chickens was isolated from chick embryos (Beaudette and Hudson 1937). Several other animal coronaviruses were later identified in the 1940s with the discovery of mouse hepatitis virus (MHV) (Cheever, Daniels *et al.* 1949) and transmissible gastroenteritis virus (TGEV) in pigs (Doyle and Hutchings 1946). The first human coronaviruses were identified in the 1960s by Tyrell and Bynoe who passaged a virus, named B814, in human embryonic tracheal organ cultures. When this virus was inoculated intranasally into human volunteers, common cold-like symptoms were produced (Tyrrell and Bynoe 1965; Tyrrell and Bynoe 1966). Hamre and Procknow (1966) later isolated a virus, which they were able to grow in tissue culture, from medical students also presenting with symptoms of the common cold. The virus was later named human coronavirus 229E (HCoV-229E) (Hamre and Procknow 1966).

It was shown that all of these viruses were ether-sensitive, indicating the presence of a lipid covering. McIntosh *et al.* (1967) were then able to recover several ether-sensitive viruses in organ culture which were termed “OC” viruses to indicate passage in organ culture. Under electron microscopy all of these viruses had morphologies similar to several other animal viruses previously identified, namely infectious bronchitis virus (IBV), MHV and TGEV (McIntosh, Dees *et al.* 1967). These viruses were later classified as coronaviruses, which was accepted as a new genus in 1975. Unfortunately, many of the clinical samples collected and stored in the 1960s, which were positive for coronavirus-like particles, were subsequently lost. Therefore, until the discovery of SARS-CoV in 2003, the study of human coronaviruses was restricted to research of HCoV-229E and HCoV-OC43 (McIntosh 2005).

CoVs are enveloped viruses and are now placed in the order *Nidovirales*, along with the arteriviruses, toroviruses and roniviruses. The *Coronaviridae* family is composed of single-stranded, positive sense RNA viruses, with the largest known RNA viral genomes, ranging in size from 25 to 30 kb (He, Dobie *et al.* 2004). Traditionally, coronaviruses were divided into Group I, II and III based on serological relationships (Figure 1.1). Viruses in Groups I and II are known to typically infect mammalian hosts, whereas members of Group III infect birds (van der Hoek, Pyrc *et al.* 2006; Pyrc, Berkhout *et al.* 2007). The International Committee for Taxonomy of Viruses (ICTV) recently proposed that these three groups be replaced by three new genera, namely *Alphacoronavirus*, *Betacoronavirus* and *Gammacoronavirus* (Woo, Huang *et al.* 2010).

Coronaviruses are responsible for a variety of diseases in animals resulting in respiratory, enteric, nephritic, hepatic and neurological conditions with a varying degree of severity (McIntosh 2005; Lau, Li *et al.* 2010). In humans, coronaviruses typically result in approximately 10-30% of upper respiratory tract infections, which are known to lead to the common cold. Until the discovery of the severe acute respiratory syndrome coronavirus (SARS-CoV) it was thought that human coronavirus infections were not severe. Infection with SARS-CoV however resulted in approximately 10% mortality of infected patients, with mortality as high as 40-55% in certain population (Surjit, Kumar *et al.* 2005) and age (Marra, Jones *et al.* 2003) groups. SARS presents as an atypical pneumonia (He, Dobie *et al.* 2004; Lee, Lee *et al.* 2008), with pneumocytes being the primary target of infection. Infection results in hemorrhagic inflammation in most pulmonary alveoli with alveolar thickening, diffuse alveolar damage, desquamation of pneumocytes, formation of hyaline membranes and

multinucleated pneumocytes with capillary engorgement and microthrombosis (Surjit, Liu *et al.* 2004).

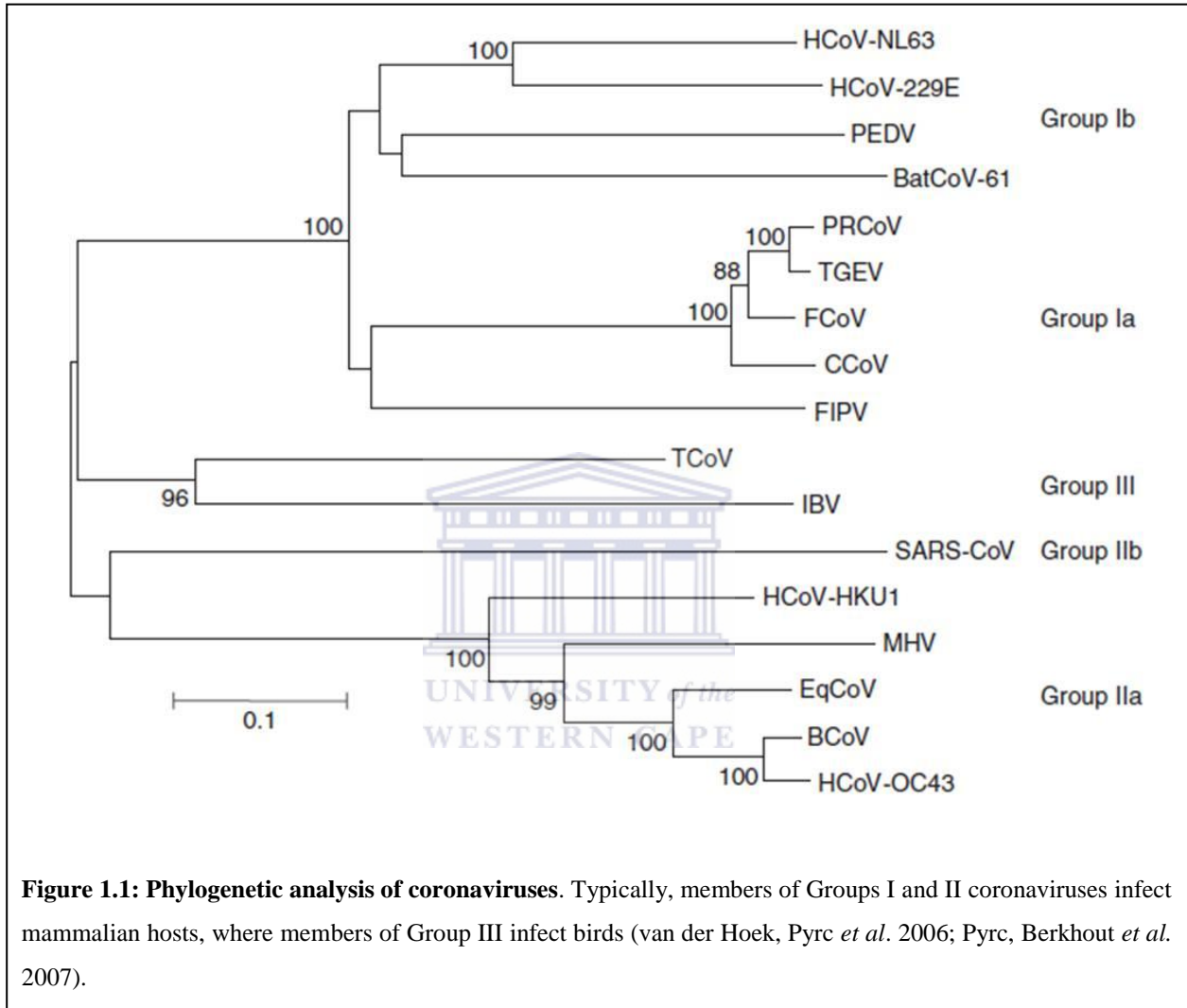


Figure 1.1: Phylogenetic analysis of coronaviruses. Typically, members of Groups I and II coronaviruses infect mammalian hosts, where members of Group III infect birds (van der Hoek, Pyrc *et al.* 2006; Pyrc, Berkhout *et al.* 2007).

1. Outbreak of SARS-CoV: Stimulating Coronavirus Research

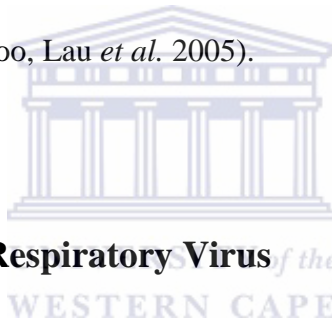
In November 2002 an outbreak of atypical pneumonia occurred in Guangdong Province, China. From November 2002 to July 2003 a total of 8098 patients were affected by the atypical pneumonia resulting in 774 deaths globally. The disease was termed severe acute respiratory syndrome and the causative agent was identified as the severe acute respiratory

syndrome coronavirus (SARS-CoV) (Drosten, Gunther *et al.* 2003; Ksiazek, Erdman *et al.* 2003; Peiris, Lai *et al.* 2003). The SARS epidemic was halted in 2003 by a highly effective global public health response and SARS-CoV is currently not circulating in the human population (Pyrç, Berkhout *et al.* 2007). SARS-CoV probably originated from a wild animal reservoir, most likely bats, and was transmitted in a zoonotic event to humans *via* infected civet cats, which were sold as food in live animal markets in China (van der Hoek, Pyrc *et al.* 2006). Interestingly, earlier seroepidemiological studies have shown that 40% of wild animal traders and 20% of people responsible for the slaughtering of animals in the region where human SARS was thought to originate, were seropositive for SARS-CoV, although all cases were asymptomatic. This indicates that these people were previously exposed to a SARS-like virus, which resulted in asymptomatic infection (Guan, Zheng *et al.* 2003).

Lau *et al.* (2010) have shown that 10 strains of SARS-related coronaviruses (SARSr-CoV) in bats undergo frequent recombination. This recombination possibly gave rise to SARSr-CoV in civets, which was then transmitted to humans resulting in the outbreak of SARS. SARSr-CoVs are currently circulating in animal hosts and reservoirs (Lau, Li *et al.* 2010). Up to 39% of all sequenced Group I and II coronaviruses are found in bats, with nearly all Group III viruses found in birds. The huge diversity of coronaviruses found in bats and birds has indicated that bats possibly provided the gene pools for Group I and II coronaviruses, where birds provided the gene pools for Group III viruses (Woo, Lau *et al.* 2009). The huge diversity of coronaviruses found in bats and birds could be due, in part, to the diversity of these animals themselves with bats accounting for up to 20% of all mammalian species and birds being the most diverse tetrapod vertebrates. The ability to fly and social habits of these animals may also promote intra- and interspecies viral spread (Woo, Lau *et al.* 2009). The

coronavirus diversity found in these animals may lead to recombination and generation of new coronavirus species, as seen with SARS-CoV.

Following the outbreak of SARS it was shown that coronaviruses are capable of undergoing recombination events resulting in a more lethal form of the virus. This ignited a new interest in the detection and characterization of previously unidentified respiratory viruses. There are now 5 known human coronaviruses with four currently circulating in the human population, namely HCoV-229E, HCoV-OC43, HCoV-HKU1 and HCoV-NL63. HCoV-HKU1 was isolated from a 71 year old man presenting with pneumonia. It was identified by RT-PCR, however it has posed difficult to characterize due to its inability to be cultured in various cell lines (Woo, Lau *et al.* 2005).

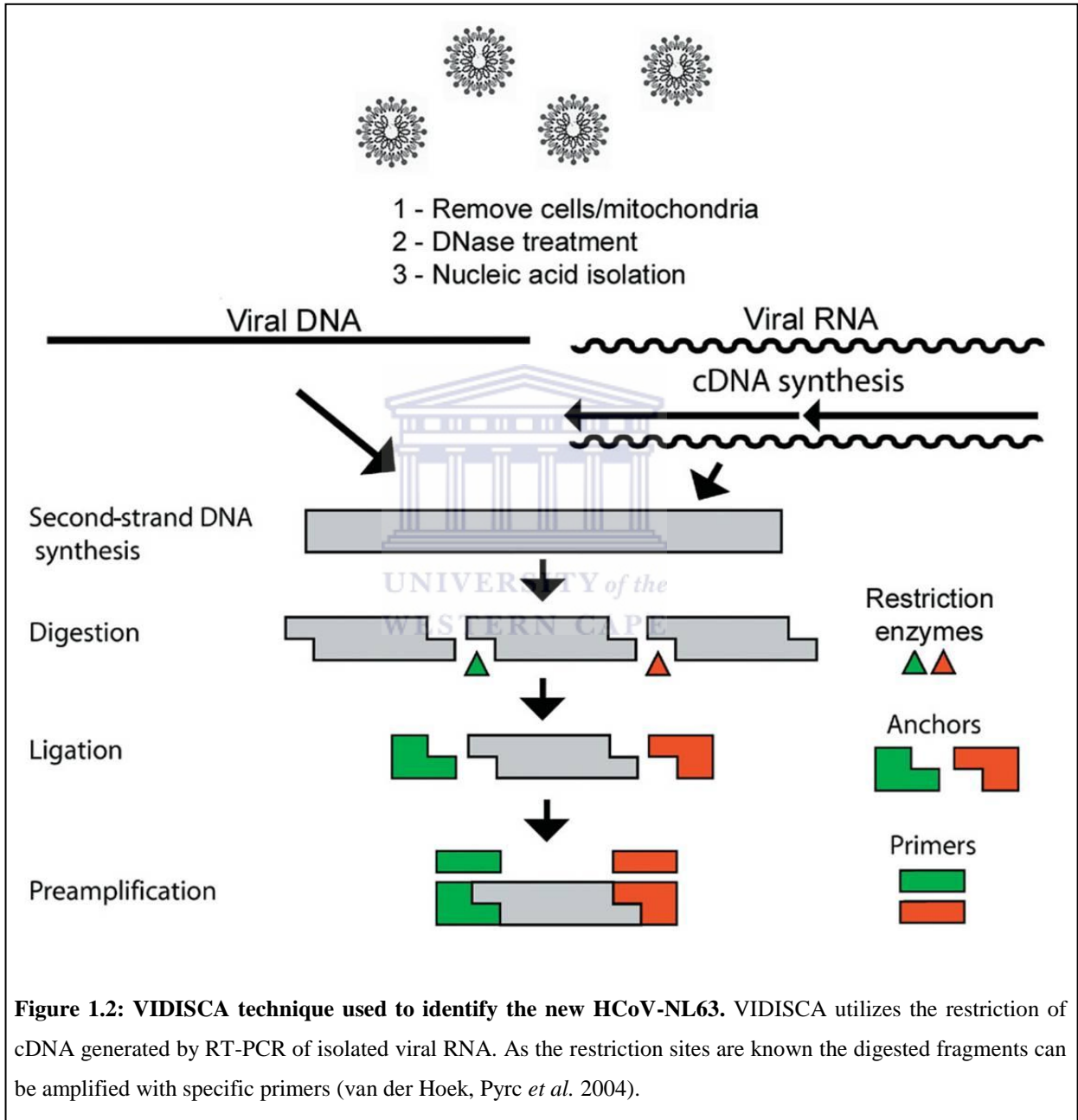


2. HCoV-NL63: A Novel Respiratory Virus

2.1 Discovery of HCoV-NL63

In 2003 a 7 month old child presenting with bronchiolitis and conjunctivitis was screened for several respiratory viruses to identify the causative agent, with all diagnostics yielding negative results. The group led by Lia van der Hoek then used a modified cDNA amplified restriction fragment-length polymorphism (cDNA-AFLP) technique (Virus-Discovery-cDNA-AFLP or VIDISCA), to identify the causative agent. Briefly, the technique utilizes reverse transcription-PCR of viral RNA with subsequent restriction digest of the cDNA using frequently cutting restriction enzymes. Since the restriction sites are selected and therefore known, the resultant “sticky ends” can be ligated into anchors for amplification and sequencing with specific primers (Figure 1.2). The results showed highest sequence

similarities with known coronaviruses, but with significant sequence divergence indicating the discovery of a new coronavirus species, later named human coronavirus NL63 (van der Hoek, Pyrc *et al.* 2004).



At about the same time, two other independent groups identified essentially the same virus (Fouchier, Hartwig *et al.* 2004; Esper, Weibel *et al.* 2005). Shortly after the van der Hoek paper (van der Hoek, Pyrc *et al.* 2004), a novel coronavirus that replicated efficiently in tertiary monkey kidney and Vero cells, was retrospectively isolated from a nose swab sample collected in 1988 from an 8 month-old boy presenting with pneumonia. This virus was reported to be similar to HCoV-NL63 and named HCoV-NL (Fouchier, Hartwig *et al.* 2004). Then in 2005, Esper *et al.* (2005) also reported the identification of a novel coronavirus isolated in New Haven, Connecticut, which was named NH. This novel virus was identified by PCR which was adapted to amplify a conserved region within the replicase 1a gene (Esper, Weibel *et al.* 2005). Subsequent sequence analysis showed that these three viruses were essentially all the same virus or variants thereof (Enserink 2005; Esper, Weibel *et al.* 2005).

It is obvious that HCoV-NL63 has been circulating in the human population since before 1988 (Fouchier, Hartwig *et al.* 2004). In fact, molecular clock analyses have shown that HCoV-NL63 and HCoV-229E diverged from a most common ancestor approximately 900 years ago, with HCoV-NL63 later diverging into two lineages with subsequent recombination of the two lineages during co-infection. This frequent recombination has given HCoV-NL63 a mosaic structured genome (Pyrc, Dijkman *et al.* 2006). HCoV-NL63 has since been isolated and identified in Australia (Arden, Nissen *et al.* 2005), Japan (Ebihara, Endo *et al.* 2005), Belgium (Moes, Vijgen *et al.* 2005), France (Vabret, Mourez *et al.* 2005), Switzerland (Kaiser, Regamey *et al.* 2005), Canada (Bastien, Anderson *et al.* 2005), Hong Kong (Chiu, Chan *et al.* 2005), Italy (Minosse, Selleri *et al.* 2008), Germany and South Africa (Smuts 2008; Venter, Lassauniere *et al.* 2011), to name a few, which indicates a global spread of this virus (van der Hoek, Pyrc *et al.* 2006).

2.2 Clinical Findings and Prevalence of HCoV-NL63

HCoV-NL63 has been shown to affect both the upper and lower respiratory tract and often presents clinically as the common cold. Symptoms have been found to be similar to those of infection with HCoV-229E and HCoV-OC43 (Abdul-Rasool and Fielding 2010). Mild symptoms affecting the upper respiratory tract include fever, cough, sore throat and rhinorrhoea, with more serious lower respiratory tract complications including bronchiolitis, pneumonia and croup (van der Hoek, Sure *et al.* 2005; van der Hoek, Pyrc *et al.* 2006).

In an isolated case, a patient who underwent allo-haematopoietic cell transplant surgery presented with a respiratory prodrome five months after surgery. He presented with a dry cough, fever and intercostal pain due to coughing with influenza-like symptoms. The disease was most likely contracted from his five year old son who presented with similar symptoms a week earlier. Bronchoalveolar lavage, tracheal and nasopharyngeal fluid was shown to be negative for fungi, acid-fast bacilli, gram stain and culture, cytology, CMV and respiratory viruses including RSV, influenza A and B, parainfluenza virus, metapneumovirus, adenovirus and rhinovirus, however, was PCR positive for HCoV-NL63 in bronchoalveolar lavage fluid. The patient's condition worsened and on day four he was intubated and by day seven steroid treatment was initiated, without improvement. Five weeks following admission to hospital, the patient died as a result of progressive respiratory failure. Post-mortem autopsy showed severe alveolar damage with no pathogens, including HCoV-NL63, present, indicating that the patient had cleared the virus. These results indicate that routine screening of coronaviruses, in particular for HCoV-NL63, should be conducted prior to and following transplant surgery. In this case, HCoV-NL63 was shown to result in severe lower respiratory

tract infection and disease which, subsequent to clearing of the virus, led to fatality (Oosterhof, Christensen *et al.* 2010).

Prevalence of infection with HCoV-NL63 is highest in the age group 0-5 years old and is also frequently observed in patients who are immunocompromised, as well as the elderly (van der Hoek, Pyrc *et al.* 2006). HCoV-NL63 has been detected in 1.0-9.3% of respiratory tract infections in children (Fielding 2011), with 75% of children aged 2.5-3.5 years being seropositive for HCoV-NL63 (Dijkman, Jebbink *et al.* 2008). In the Esper study, 11 patients who tested positive for HCoV-NH had been hospitalized since birth, indicating that HCoV-NH and HCoV-NL63 may represent a nosocomial or congenital infection (Esper, Weibel *et al.* 2005). In fact, maternally acquired antibodies are present in newborns and probably provide protection against infection. These maternally acquired antibodies usually diminished within 4-5 months after birth (Shao, Guo *et al.* 2007; Dijkman, Jebbink *et al.* 2008). In a comprehensive population based study conducted in Germany, it was found that HCoV-NL63 resulted in infection of 7 per 1000 children per year, which can be extrapolated to 16 929 visits to the doctor in Germany alone. Hospitalization rates were much lower with a rate 22 per 100 000 children, indicating a total of 522 hospitalizations in Germany each year (van der Hoek, Ihorst *et al.* 2010).

Although HCoV-NL63 infection has been found to occur most frequently during the winter months (van der Hoek, Ihorst *et al.* 2010; Fielding 2011), several reports indicate seasonal variations. In China, HCoV-NL63 infections peaked in spring and autumn (Chiu, Chan *et al.* 2005), and in Taiwan, infections peaked during summer (Wu, Chang *et al.* 2008). In Thailand, however, no seasonal variation was identified as HCoV-NL63 could be detected throughout the year (Dare, Fry *et al.* 2007). This may be due, in part, to the tropical climate of

Thailand where seasonal changes are not as significant as within temperate climates. The peak in seasonal incidence during the winter months may also promote co-infection with other respiratory viruses which spike during the same period. This may also promote recombination of respiratory viruses which co-infect the same cell.

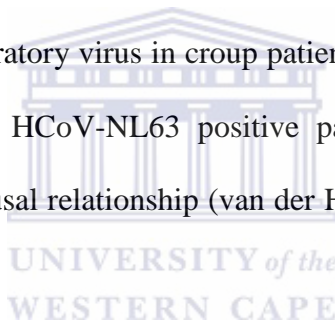
The epidemiology of coronaviruses, including HCoV-NL63, remains poorly defined due largely to the fact that these viruses are not routinely screened for. Human CoVs are rather clinically diagnosed leading to a potentially great underestimation of the role of HCoVs in respiratory infections (van Elden, van Loon *et al.* 2004). The respiratory site chosen for sampling also plays a large role in misdiagnosis of respiratory viruses as different results have been obtained from nasopharyngeal aspirates and nose/throat swabs (Kleines, Scheithauer *et al.* 2007). Different diagnostic assays used in detection have also been shown to have varying degrees of sensitivity leading to further underestimation of these viruses in disease (Fielding 2011). Therefore, the development of reliable and routine screening methods for HCoVs in respiratory infections is of increasing importance in understanding the epidemiology of these viruses and the role(s) they play in other diseases.

2.3 Association of HCoV-NL63 with Croup

Croup is a common manifestation of lower respiratory tract disease and is characterized as an inflammation of the trachea which obstructs normal breathing, producing the characteristic “barking” cough that usually worsens at night (van der Hoek, Sure *et al.* 2005). Croup patients present with pharyngitis and hospitalization is often required (Fielding 2011). The causal agent is generally assumed to be a respiratory virus, with parainfluenza virus 1 (PIV1) and influenza A frequently being implicated. Other respiratory viruses associated with croup

include rhinovirus, PIV3, respiratory syncytial virus, human bocavirus and several others in a small minority of subjects, with co-infections found in 11% of patients (Sung, Lee *et al.* 2010).

An association between HCoV-NL63 and croup has been reported by several researchers (Chiu, Chan *et al.* 2005; van der Hoek, Sure *et al.* 2005; Wu, Chang *et al.* 2008; Sung, Lee *et al.* 2010). van der Hoek *et al.* (2005) reported a strong association between HCoV-NL63 and croup, with 43% of HCoV-NL63-positive patients - with single infection and high viral load - presenting with croup. Of croup patients, 17.4% tested positive for HCoV-NL63, compared to 4.2% of non-croup patients. HCoV-NL63 is therefore the second most commonly identified respiratory virus in croup patients. The chance of developing croup is in fact 6.6 times higher in HCoV-NL63 positive patients than HCoV-NL63 negative patients, suggesting a strong causal relationship (van der Hoek, Sure *et al.* 2005; Sung, Lee *et al.* 2010).



2.4 Association of HCoV-NL63 with Non-respiratory Findings

2.4.1 Kawasaki Disease

Kawasaki disease is one of the most common forms of childhood vasculitis and presents with prolonged fever and a polymorphic exanthema, oropharyngeal erythema and bilateral conjunctivitis (van der Hoek, Pyrc *et al.* 2006) and may result in aneurysms of coronary arteries (Esper, Shapiro *et al.* 2005). Kawasaki disease has been recognized as the most common cause of chronic heart disease in children (Esper, Shapiro *et al.* 2005). An infective agent has long been thought to be the cause of Kawasaki disease because of seasonal peaks, focal epidemics and the acute onset of symptoms such as fever, rash, conjunctival infection and

cervical adenitis, which resemble other infectious diseases. Since a respiratory syndrome precedes symptoms of Kawasaki disease, a respiratory virus has been proposed as the causative agent (Bell, Brink *et al.* 1981). A strong IgA response has also been detected in the respiratory tract and medium & large arteries during Kawasaki disease, indicating possible entry of a microbe *via* the respiratory tract (Rowley, Shulman *et al.* 2000). Furthermore, the highest incidence is found in toddlers, with only rare cases in adults and infants, which suggest a causative agent to which adults are immune to and from which infants are protected by passive immunity. Bacterial and viral cultures and serological investigations, however, have yet to identify the causative agent (Shimizu, Shike *et al.* 2005). The diagnosis of Kawasaki disease is based entirely on clinical features and currently no diagnostic tests are available.

An association between Kawasaki disease and HCoV-NL63 was recently reported by Esper *et al.* (2005). In their study, viral sequences of HCoV-NL63 were detected from respiratory secretions for 8 of 11 patients presenting with Kawasaki disease; only 1 of 22 control subjects were positive for HCoV-NL63 genetic material. These results are indicative of a strong association between Kawasaki disease and HCoV-NL63 infection (Esper, Shapiro *et al.* 2005). However, in a separate study by Shimizu *et al.* (2005) only 1 of 48 respiratory secretions from patients suffering from Kawasaki disease was positive for HCoV-NL63. The contradiction in these findings may be due to differences in geographical region, definition of cases, detection method used or respiratory site sampled. Also, since the Esper study only used samples from patients presenting with respiratory symptoms, it was sample biased for a high rate of HCoV-NL63 infection (Shimizu, Shike *et al.* 2005). Several other groups have since reported no correlation between HCoV-NL63 infection and Kawasaki disease (Belay, Erdman

et al. 2005; Ebihara, Endo *et al.* 2005b; Chang, Chiang *et al.* 2006; Lehmann, Klar *et al.* 2009) and therefore this association should be approached with a level of caution.

2.4.2 Gastrointestinal Disease

HCoV-NL63 has been reported to be present in the stool samples of patients with diarrhea, indicating a causal role in gastroenteritis (Vabret, Mourez *et al.* 2005; Leung, Li *et al.* 2009; Risku, Lappalainen *et al.* 2010). The involvement in gastrointestinal disease is not unique to HCoV-NL63, as all non-SARS CoVs have been found in stool samples of children presenting with acute gastroenteritis (Risku, Lappalainen *et al.* 2010). In addition, in several studies a relationship was found between SARS-CoV and gastrointestinal disease (Leung, To *et al.* 2003; Clark and McKendrick 2004). Involvement of HCoVs in gastrointestinal disease is a result of viral invasion of the intestinal mucosa (Principi, Bosis *et al.* 2010). The majority of patients with HCoVs detectable in stool samples, also tested positive for other gastroenteritis viruses, including norovirus and rotavirus (Risku, Lappalainen *et al.* 2010). HCoVs are more frequently reported in outpatients than hospitalized patients presenting with gastroenteritis (Kheyami, Nakagomi *et al.* 2010), indicating that gastroenteritis related to HCoV infection may possibly be less severe and therefore underrepresented.

A recent report showed that HCoV-NL63 was detectable in 4 of 878 stool samples from patients presenting with gastroenteritis (Risku, Lappalainen *et al.* 2010). Another report however showed that gastroenteritis developed in 33.3% of patients infected with HCoV-NL63 (Leung, Li *et al.* 2009). These and other results could indicate that HCoV-NL63, and other HCoVs, may have a causal role in gastroenteritis. However, since many HCoV-NL63

positive patients present with co-infections with other gastrointestinal viruses, the exact role of HCoV-NL63 cannot be identified at this time (Fielding 2011).

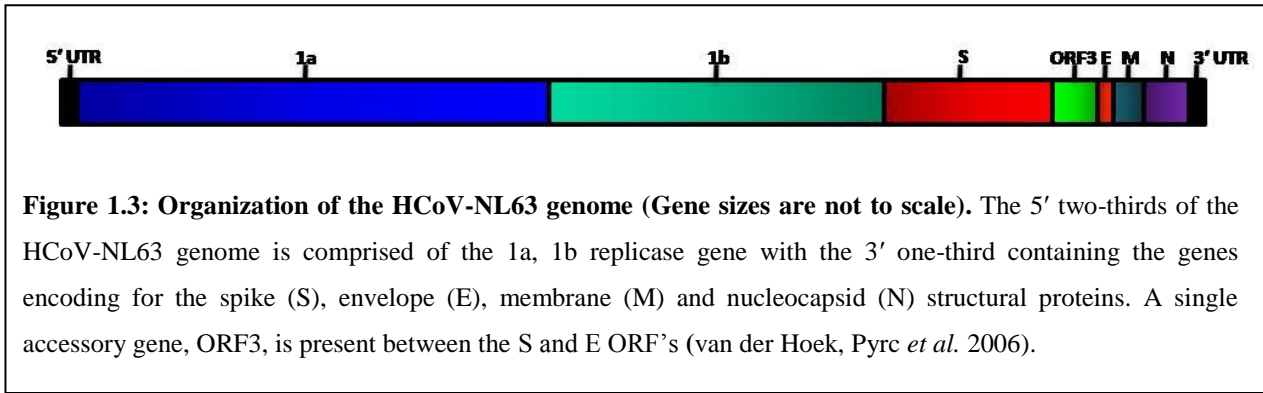
2.5 Co-infection of HCoV-NL63 and other respiratory viruses

As many as half of HCoV-NL63 infections may present as co-infections with another respiratory virus including influenza A (Chiu, Chan *et al.* 2005), respiratory syncytial virus, parainfluenza-3 or human metapneumovirus (Arden, Nissen *et al.* 2005; van der Hoek, Sure *et al.* 2005). HCoV-NL63 presenting as a co-infection is more likely to lead to hospitalization than in patients presenting with HCoV-NL63 alone, indicating the increased severity of co-infections. In patients co-infected with HCoV-NL63 and another respiratory virus, the total viral load of HCoV-NL63 is lower. This could indicate that HCoV-NL63 is capable of lowering the immune response to allow for a secondary infection, by which time the HCoV-NL63 infection has subsided or that the secondary infection activates the innate immune response thereby inhibiting HCoV-NL63 progression. Other possible reasons for this decreased HCoV-NL63 viral load could be that the co-infected viruses compete for the same target cell or receptor or that HCoV-NL63 has a prolonged persistence at low levels (van der Hoek, Pyrc *et al.* 2006). The decreased viral load could also be attributed to HCoV-NL63 being less well adapted to the host compared to the co-infecting respiratory virus. HCoV-NL63 has also been shown to promote bacterial co-infection by increasing the adherence of *Streptococcus pneumoniae* to epithelial cells (Golda, Malek *et al.* 2011).

3. Coronavirus Genome, Transcription and Replication

Coronaviruses have a positive sense, non-segmented, single stranded RNA genome of approximately 30kb, which is typically organized in the conserved order: 5'-replicase (1a, 1b), spike (S), envelope (E), membrane (M) and nucleocapsid (N)-3' (He, Dobie *et al.* 2004) (Figure 1.3). The genome has a 5'cap and is 3'polyadenylated (Pyrce, Berkhout *et al.* 2007). Whereas HCoV-NL63 has one additional open reading frame (ORF) - ORF3 - between the S and E ORFs (Pyrce, Jebbink *et al.* 2004), SARS-CoV has 9 ORFs distributed amongst the structural genes. The 5' two-thirds of the coronavirus genome contain the large 1a and 1b ORFs. These ORFs encode proteins necessary for RNA replication and are termed the non-structural proteins (nsp) (van der Hoek, Pyrc *et al.* 2006).

Coronaviruses utilize a discontinuous replication strategy to synthesize its subgenomic (sg) mRNA during synthesis of the negative strand from the positive sense genome template (Pyrce, Jebbink *et al.* 2004). The sg mRNA is then copied into a positive sense strand. Transcription of the HCoV-NL63 genome produces six distinct mRNAs (Pyrce, Berkhout *et al.* 2007), which include the entire RNA genome and a nested set of five sg mRNAs. The six mRNAs encode for the 7 ORFs of HCoV-NL63, with ORF3 and S, M, E and N encoded by the five subgenomic mRNAs (Pyrce, Jebbink *et al.* 2004). A transcription regulatory sequence of AACUAAA, which is essential for sg mRNA synthesis, is located at the 5'-end of each ORF with the exception of ORF-E (Pyrce, Jebbink *et al.* 2004; Pyrc, Berkhout *et al.* 2007; Sawicki, Sawicki *et al.* 2007). All positive sense mRNAs also share a leader sequence of approximately 70 nucleotides which is identical to the leader sequence of the genomic RNA (Pyrce, Jebbink *et al.* 2004; Sawicki, Sawicki *et al.* 2007).



The expression of the entire 1ab gene produces two large polyproteins (1a and 1ab), where the synthesis of the 1ab polyprotein requires a -1 ribosomal frameshift during translation of the 1a gene (Pyrc, Jebbink *et al.* 2004; Sawicki, Sawicki *et al.* 2007). With a -1 ribosomal frame shift, the ribosome is triggered to move one nucleotide upstream to encode a completely new protein (Namy, Moran *et al.* 2006). The process by which a ribosomal frameshift occurs is not completely understood, but is triggered by a heptanucleotide “slippery” sequence lying upstream of a RNA pseudoknot structure. A RNA pseudoknot is comprised of two stem sequences folding back on themselves to form a two loop structure and a brief complementary sequence of approximately 17 base pairs (Figure 1.4). The frameshift is initiated when the elongating ribosome is stalled by the heptanucleotide sequence followed by a frameshift stimulated by the pseudoknot. This most likely occurs by a difference in free energy between the folded and partially folded pseudoknot structure which promotes frameshifting by increasing the kinetic barrier (Giedroc, Theimer *et al.* 2000).

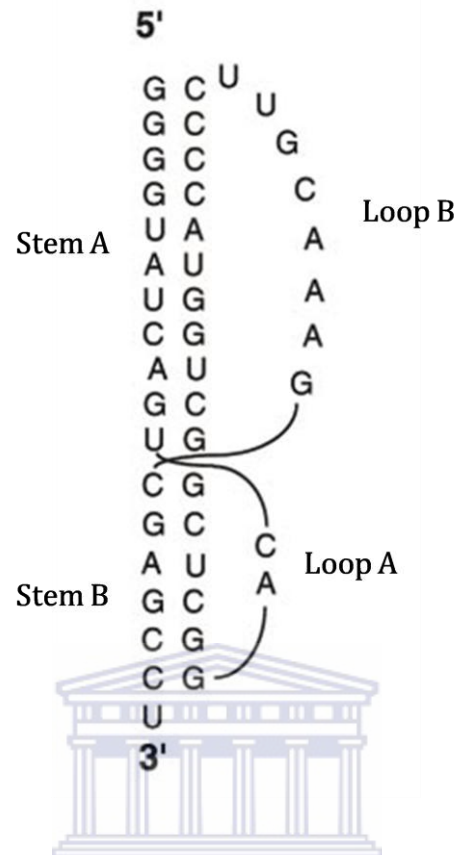


Figure 1.4: Structure of an RNA pseudoknot. An RNA pseudoknot is formed by stem A forming loop A which folds back on stem B to form a brief complimentary sequence with stem B and A. The complimentary sequence folds again to form loop B which rejoins to stem B (Giedroc, Theimer *et al.* 2000).

The 1a and 1ab polyproteins are subsequently cleaved by viral encoded proteases to produce 16nsps. Polyprotein 1a is cleaved to form nsp 1-11, where polyprotein 1ab forms nsp 12-16. These nsps form the replicase-transcriptase proteins of coronaviruses and assemble, along with other viral and host proteins, into a membrane bound replication-transcription complex. Previous studies have shown that these proteins have at least 8 enzymatic functions including protease, RNA-dependant RNA polymerase and 5'-to-3' helicase activity (Marra, Jones *et al.* 2003; Sawicki, Sawicki *et al.* 2007).

The 3' one-third of the genome encodes the structural proteins S, E, M and N, with haemagglutinin esterase found in Group II viruses. The accessory proteins, which differ in number and position for different coronaviruses, are found in between the structural proteins (van der Hoek, Pyrc *et al.* 2006). The function of these accessory proteins is important in determining pathogenesis of coronaviruses as accessory proteins of certain coronaviruses have shown to contribute to the progression of the disease (Kopecky-Bromberg, Martinez-Sobrido *et al.* 2007).

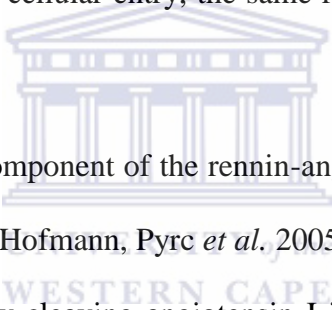
RNA virus genomes have a high mutation rate and therefore the probability of generating genetic diversity is high. This mutation rate is estimated to be approximately one mutation per replication cycle, and as a consequence, RNA viruses exist as a dynamic distribution of variants which closely resemble a master sequence. This high mutation rate allows the virus to adapt quickly to environmental changes and pressures (van der Hoek, Pyrc *et al.* 2006). Coronaviruses are also capable of recombination, as with SARS-CoV, if 2 or more viruses infect the same cell (Kahn and McIntosh 2005) and as HCoV-NL63 presents frequently as a co-infection with other respiratory viruses, this may - although highly unlikely - increase the possibility of recombination (Pyrc, Dijkman *et al.* 2006).

Nidovirus replication is restricted to the cytoplasm (Rowland, Chauhan *et al.* 2005) and viral attachment to the host cell is mediated by the S protein. The viral membrane proteins, including the M- and S-proteins, insert into the endoplasmic reticulum (ER) and Golgi intermediate compartment. During this process the full length RNA genome associates with the N-protein. The resulting ribonucleoprotein complex then associates with the M-protein in the membrane of the ER. Viral particles are formed as the nucleocapsid complex

buds into the lumen of the ER. The virus then migrates through the Golgi complex and exits the cell, most likely *via* exocytosis (Marra, Jones *et al.* 2003).

4. HCoV-NL63 Cellular Entry by ACEII

HCoV-229E, the most closely related human coronavirus to HCoV-NL63, has been shown to utilize CD13 (aminopeptidase N) for cellular entry and before the discovery of HCoV-NL63 it was thought that all Group I coronaviruses utilize CD13 as a receptor for cellular entry (van der Hoek, Pyrc *et al.* 2006). HCoV-NL63 has, however, been shown to use angiotensin-converting enzyme (ACE) II for cellular entry, the same receptor for cellular entry as SARS-CoV.



ACEII is an important component of the rennin-angiotensin system, which is involved in the control of blood pressure (Hofmann, Pyrc *et al.* 2005). ACE is responsible for activating the rennin-angiotensin system by cleaving angiotensin I into angiotensin II, whereas ACEII negatively regulates this system by inactivating angiotensin II (Pyrc, Berkhout *et al.* 2007). ACEII molecules are found to be localized to arterial and venous endothelial cells, arterial smooth muscle cells and epithelia of small intestines as well as the ciliated cells of human nasal and tracheobronchial airway epithelia, thereby explaining the infection of cells of the upper airways (van der Hoek, Pyrc *et al.* 2006; Pyrc, Berkhout *et al.* 2007).

The S-glycoproteins of coronaviruses provide the virion with a corona (crown) like appearance under electron microscopy and mediates infection of host cells. Entry of coronaviruses into the host cell is facilitated either by direct fusion with the plasma membrane or by endocytosis and subsequent fusion with the endosomal membrane (Pyrc, Berkhout *et al.*

2007). The amino-terminal of the S-protein contains the receptor-binding domain and the C-terminal domain contains the membrane anchoring domain with fusogenic function (Pyrç, Berkhout *et al.* 2007). Binding of the S protein to certain receptors on the host cell determines, to a large extent, the type of host cell which the virus can infect, which has important implications for viral pathogenicity (Hofmann, Pyrc *et al.* 2005). Hofmann *et al.* (2005) have shown that HCoV-NL63 and SARS-CoV use the same receptor for cellular entry even though there is no similarity between the SARS-CoV and HCoV-NL63 S-protein that could explain this shared receptor usage. HCoV-NL63 and SARS-CoV must therefore have evolved their ability to bind to ACEII separately due to a common “virus-binding hotspot” on ACEII (Chatterjee, Johnson *et al.* 2009).

This common receptor usage indicates that HCoV-NL63 is able to infect the same host cell as SARS-CoV. HCoV-NL63 has however, only been shown to result in mild or moderate respiratory symptoms and it is therefore conceivable that HCoV-NL63 lacks a pathogenicity factor present in SARS-CoV. It is assumed that this pathogenicity factor could be encoded by 1 or more of the 9 accessory genes present in the SARS-CoV genome. Another possible explanation for the difference in pathogenicity of HCoV-NL63 and SARS-CoV is a difference in interaction of S with ACEII or a different binding site on ACEII. Hoffman *et al.* have shown that HCoV-NL63 has a much lower binding affinity to ACEII compared to SARS-CoV. The S-protein may therefore partially account for the difference in pathogenicity of HCoV-NL63 and SARS-CoV (Hofmann, Pyrc *et al.* 2005).

5. The Coronavirus Nucleocapsid Protein

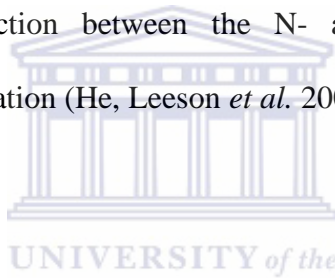
The coronavirus nucleocapsid (N) protein is one of the most abundantly produced viral proteins within an infected cell (Krokhin, Li *et al.* 2003) and has several functions. The primary function is the formation of the ribonucleocapsid core, which requires homotypic N dimerization to form higher order oligomers (He, Dobie *et al.* 2004; Surjit, Liu *et al.* 2004; Luo, Ye *et al.* 2005; Yu, Gustafson *et al.* 2005; Chang, Sue *et al.* 2006; Luo, Chen *et al.* 2006) with subsequent interaction with viral RNA (Stohlman, Baric *et al.* 1988; Cologna, Spagnolo *et al.* 2000; Nelson, Stohlman *et al.* 2000; Fan, Ooi *et al.* 2005; Spencer and Hiscox 2006; Tan, Fang *et al.* 2006; Chen, Chang *et al.* 2007; Chang, Hsu *et al.* 2009; Huang, Hsu *et al.* 2009; Ma, Tong *et al.* 2010). Once this particle has assembled it interacts with the membrane glycoprotein, which in turn interacts with the envelope protein (Lim and Liu 2001) during viral assembly (He, Leeson *et al.* 2004; Hurst, Koetzner *et al.* 2009). The N-protein has also been shown to interact with cellular heterogenous nuclear ribonuclear protein A1 (hnRNP-A1), which is thought to play a role in RNA synthesis (Wang and Zhang 1999; Luo, Chen *et al.* 2005; Chang, Sue *et al.* 2006) and with the human cyclophilin A protein (Luo, Luo *et al.* 2004). Besides the structural functions of the N-protein, it is also involved in the regulation of transcription, translation and replication (Baric, Nelson *et al.* 1988; Almazan, Galan *et al.* 2004; Pyrc, Jebbink *et al.* 2004; Verheije, Hagemeijer *et al.* 2010) of viral RNA as it has been shown to interact with replication-transcription complexes (Verheije, Hagemeijer *et al.* 2010). It has also been shown to be involved in perturbation of host cell physiological processes (He, Leeson *et al.* 2003; Surjit, Liu *et al.* 2004; Surjit, Liu *et al.* 2006; Yan, Hao *et al.* 2006; Kopecky-Bromberg, Martinez-Sobrido *et al.* 2007). N is also highly immunogenic and has therefore served as a target for the development of diagnostic and sero-prevalence assays (Che, Qiu *et al.* 2004; Lau, Woo *et al.* 2004; Lehmann, Wolf *et al.* 2008; Das, Kammila *et*

*al.*2010) and for the possible production of various vaccines (Kim, Lee *et al.* 2004; Girard, Cherian *et al.* 2005). This high level of immunogenicity will be discussed in more detail in subsequent chapters. Posttranslational modifications of the N-protein include sumoylation (Li, Xiao *et al.* 2005), acetylation (Krokhin, Li *et al.* 2003) and phosphorylation (Calvo, Escors *et al.* 2005; Chen, Gill *et al.* 2005; Surjit, Kumar *et al.* 2005; White, Yi *et al.* 2007).

5.1 Multimerization of Coronavirus N

The primary function of the N-protein is to associate with the viral RNA to form a protective covering. To do this the N-protein must be able to recognize the viral RNA and it must be able to undergo N-N homotypic interactions and self-associate into an oligomer to form the nucleocapsid core (He, Dobie *et al.* 2004; Surjit and Lal 2008). A serine-arginine (SR) rich region is essential for this multimerization and self-interaction of the N-protein in SARS-CoV, as deletion of these amino acids completely abolish N multimerization (He, Dobie *et al.* 2004). He and colleagues also reported that a serine deletion mutant not only affected multimerization of the N-protein, but also altered the cellular distribution of the protein (He, Dobie *et al.* 2004). Another study also identified a 209 amino acid region at the C-terminal domain of the SARS-CoV N-protein that contains the interaction site important for trimer, tetramer or hexamer formation (Surjit, Liu *et al.* 2004; Surjit and Lal 2008). These results however contradict the results of He *et al.*(2004) as the SR rich region lies outside the 209 amino acid region proposed by Surjit *et al.*(2004). The dimerization process of the N-protein is therefore rather complicated and will be covered in more detail, along with RNA-N interactions, in subsequent chapters.

The major heterotypic protein interactions involved in assembly of coronaviruses is that between the N- and M-protein. The M-protein is the most abundantly expressed envelope protein and has a triple spanning transmembrane region with a large carboxy terminal in the interior of the virion which is responsible for interaction with the N-protein. The M-protein inserts into the membranes of the ER-GIC complex and combines with the N-protein during cellular budding (Hurst, Koetzner *et al.* 2009). A region of the N-protein (a.a. 168-208), which is also responsible for N-protein dimerization, is responsible for the interaction with M in SARS-CoV (He, Leeson *et al.* 2004). Since coronaviruses differ from other enveloped viruses (in that they lack the matrix protein that commonly links the envelope of the virus to the ribonucleocapsid core), interaction between the N- and M-proteins are essential for coronaviral assembly and maturation (He, Leeson *et al.* 2004).



5.2 Intracellular Localization of the Nucleocapsid Protein

Even though several studies have shown that the N-protein of several members of *Nidovirales* localize to the nucleus and/or nucleolus (Hiscox, Wurm *et al.* 2001; Wurm, Chen *et al.* 2001; Timani, Liao *et al.* 2005; Reed, Dove *et al.* 2006), others have shown that N localizes predominantly or exclusively to the cytoplasm (Rowland, Chauhan *et al.* 2005; Surjit, Kumar *et al.* 2005; You, Dove *et al.* 2005). This indicates that the localization of N to the nucleus or nucleolus is not a conserved property of nidoviruses (Rowland, Chauhan *et al.* 2005). There is currently no data concerning the localization of the N-protein of HCoV-NL63.

Movement of N from the cytoplasm through the nuclear pore complex and into the nucleoplasm is dependent and initiated by a nuclear localization signals (NLSs) on the N-

protein, which binds to a NLS-binding site on importin- α (Rowland, Chauhan *et al.* 2005). NLSs are enriched with basic amino acids (such as lysines and arginines) and generally conform to one of three types, *i.e.* monopartite (pat4 and pat7) and bipartite motifs (Rowland, Chauhan *et al.* 2005). On the other hand, export of proteins from the nucleus is mediated by nuclear export signals (NESs); which are predominantly CRM-1-dependant signals (You, Dove *et al.* 2005). Nuclear pore complexes also allow for the passive transport of ions, small molecules and proteins with mass of 40 to 60 kDa in both directions between the cytoplasm and nucleoplasm. Coronavirus N-proteins from all three groups meet both criteria in that they have NLSs and are less than 50kDa in size (Wurm, Chen *et al.* 2001).

5.3 Phosphorylation of the N-Protein

Phosphorylation is a common post-translational modification and plays a role in the physiological functioning of proteins. N is heavily phosphorylated at multiple residues with the majority of phosphorylations occurring at serine residues. Surjit *et al.* (2005) showed that the N-protein of SARS-CoV is phosphorylated in the nucleus and cytoplasm. The stable phospho-protein then localizes to the cytoplasm and associates with the M-protein. In addition to the multiple phosphorylation sites on the N-protein for glycogen synthase kinase (GSK), protein kinase A and casein kinase (CK) I and II, there are also single sites for cyclin-dependent kinases (CDK), protein kinase B, mitogen-activated protein kinase (MAPK) and polo-like kinase, indicating that the N-protein is phosphorylated at multiple sites by multiple enzymes. The coronavirus N-protein is phosphorylated immediately after synthesis and remains stably phosphorylated (Surjit, Kumar *et al.* 2005; Surjit and Lal 2008; Wu, Yeh *et*

*al.*2009). Cytosolic enzymes are responsible for phosphorylating the N-protein. It is, however, also possible that multiple kinases phosphorylate the N-protein in the cytoplasm and nucleus (Surjit, Kumar *et al.* 2005). N-proteins of many other coronaviruses are similarly phosphorylated, but the mechanism by which this occurs and the functional relevance thereof, if any, remains to be determined. Possible reasons for the phosphorylation of the N-protein may include: 1) increasing protein stability, 2) aiding in self-association of the protein into multimers for interactions with RNA, 3) playing a role in localization of the protein, 4) playing a role in nucleocytoplasmic shuttling (Chang, Sue *et al.* 2006) or 5) creating a more favorable environment for viral replication by modulating activity of kinases (Surjit, Kumar *et al.* 2005). Chen *et al.* (2005) have identified phosphorylation sites that are proximal to RNA binding domains. They noted that phosphorylated N-proteins bind to viral RNA with the same affinity as non-phosphorylated N-proteins. Importantly though, the phosphorylated N-proteins bind with a higher affinity to viral RNA than to non-viral RNA. This may indicate that phosphorylation of the N-protein may increase its ability to recognize viral RNA (Chen, Gill *et al.* 2005).

5.4 Perturbation of Host Cell Processes by the N-protein

The N-protein of coronaviruses has been known to act as a regulatory protein and thereby perturb certain cellular processes. To date, coronavirus N proteins have been shown to cause deregulation of the host cell-cycle (Wurm, Chen *et al.* 2001; Surjit, Liu *et al.* 2006), inhibit the production of interferon (Kopecky-Bromberg, Martinez-Sobrido *et al.* 2007), up-regulate the production of COX2 (Yan, Hao *et al.* 2006), up-regulate AP1 activity (He, Leeson *et al.* 2003),

induce apoptosis in serum starved COS1-cells (Surjit, Liu *et al.* 2004) and interact with various cellular proteins including cyclophilin A (Luo, Luo *et al.* 2004) and hnRNP-A1 (Wang and Zhang 1999; Luo, Chen *et al.* 2005), which may have possible pathological consequences.

5.4.1 Deregulation of host cell-cycle

Deregulation of the host cell cycle is a common strategy employed by certain viruses which enable them to exploit the host cell machinery for their own benefit. Different steps in the host cell cycle typically include the G₁ phase (preparation for DNA synthesis), S phase (replication of the genome), G₂ phase (preparation for cell division) and the M phase (mitosis).

Expression of the N-protein of TGEV and MHV leads to the inhibition of host cell proliferation induced by a cell cycle delay or arrest, most likely in the G₂/M phase (Wurm, Chen *et al.* 2001). On the other hand, the N-protein of SARS-CoV inhibits the S-phase in mammalian cell lines by inhibiting the activity of the cyclin-CDK complex, which results in hypophosphorylation of retinoblastoma protein and subsequent down-regulation of the S phase gene products (Surjit, Liu *et al.* 2006).

5.4.2 Inhibition of interferon production

Interferon (IFN) production by the host cell is a key defense mechanism against infection. SARS-CoV infection, however, does not result in the production of interferon even though pretreatment of cells with interferon inhibits SARS-CoV infection (Surjit and Lal 2008). This indicates that SARS-CoV has developed a response to overcome the IFN defense. When cells are virally infected, several transcription factors are activated, including IRF-3 and NF- κ B.

These transcription factors induce the expression of the IFN gene. Interferon is then synthesized and released from cells and binds to its receptors inducing the JAK/STAT signaling cascade pathway. This cascade pathway results in the activation of transcription factors which translocate to the nucleus and activate genes containing an interferon-stimulated response element (ISRE). Activation of the ISRE allows cells to combat viral infection and prevent viral replication (Kopecky-Bromberg, Martinez-Sobrido *et al.* 2007).

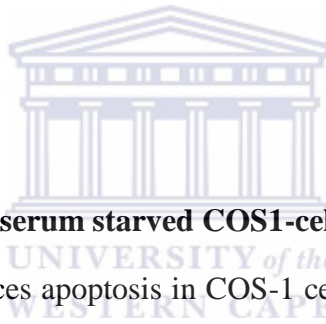
In a previous study, Kopecky-Bromberg *et al.* (2006) showed that SARS-CoV ORF 3b, ORF 6 and N antagonize the action of interferon. They showed that while N inhibits the production of IFN, ORF 3b and ORF 6 inhibit both IFN production and IFN signaling. Whereas IRF-3 was inhibited in cells expressing the N, ORF 3b and ORF 6, NF- κ B was only inhibited in cells expressing N. ORF 3b and ORF 6 was also shown to block ISRE, with ORF 6 shown to inhibit the translocation of STAT 1. These results show that SARS-CoV has at least 3 proteins which are able to inhibit certain aspects of the IFN response (Kopecky-Bromberg, Martinez-Sobrido *et al.* 2007).

5.4.3 Up-regulation of COX-2

Cyclo oxygenase-2 (COX-2) protein is another pro-inflammatory molecule which, when stimulated in SARS-CoV infection, results in inflammation of the lungs. The SARS-CoV N-protein stimulates the activation of COX-2 by binding to the COX-2 promoter. This activation of COX-2 requires binding to a NF- κ B binding site and a CCAAT/enhancer binding protein binding site. SARS-CoV N-protein was able to bind directly to both of these regulatory sequences (Yan, Hao *et al.* 2006).

5.4.4 Up-regulation of the activator protein (AP) 1 pathway

The AP-1 pathway is involved in cellular processes including cell proliferation, differentiation and apoptosis. The levels of transcription factors binding to promoter sequences of c-Fos, ATF2, CREB-1 and FosB was increased during the expression of SARS-CoV N-protein. Since the AP-1 signal transduction pathway is composed of homodimers and heterodimers of these molecules, He *et al.* (2003) proposed that the AP-1 signal transduction pathway is activated by the SARS-CoV N-protein. This was confirmed using the PathDetect system and the N-protein was shown to activate the AP-1 pathway in a dose dependent manner. The mechanism involved in the activation of this pathway, however, remains unknown (He, Leeson *et al.* 2003).



5.4.5 Induction of apoptosis in serum starved COS1-cells

The SARS-CoV N-protein induces apoptosis in COS-1 cells in the absence of growth factors. The induction of apoptosis is induced by 1) downregulation of Erk, 2) upregulation of p38 MAPK (MAPK/Erk pathway is involved in cell division), 3) upregulation of Jnk (c-Jun N-terminal kinases: involved in apoptosis, cell differentiation and proliferation, inflammation and cytokine production), 4) downregulation of P-Akt and Bcl-2 (anti-apoptotic) and 5) induction of caspase 3 and 7 (pro-apoptotic). The activation of the p38 MAPK pathway induces actin reorganization in the absence of growth factors. Both the induction of apoptosis and actin reorganization induced by the activation of p38 MAPK are independent of each other as blocking the p38 MAPK pathway did not lead to inhibit activation of caspase 3, indicating 2 independent functional properties of N. The induction of apoptosis was also shown to be independent of the Fas, Fas ligand mediated pathway. This cell death was shown

to be tissue type specific as no significant cell death was observed in human hepatoma cells of epithelial cell lineage (Surjit, Liu *et al.* 2004).

5.4.6 Binding of N-protein to cyclophilin A

Cyclophilins have a high affinity against cyclosporine A, which is an immunosuppressive drug used to prevent allograft rejection. Cyclosporine A also blocks HIV-1 viral infection where cyclophilin A has recently been shown to have important implications in HIV infection by having an essential role in replication. Cyclophilin A interacts with the capsid domain of HIV-1. Although the SARS-CoV genome does not encode a HIV-like capsid protein, sequence alignment shows homology between the SARS-CoV N-protein and HIV-1 capsid protein. This homology indicates that the N-protein of SARS may function in the same way as HIV-1 capsid protein. Luo *et al.* (2004) verified interaction between the SARS-CoV N-protein and cyclophilin A using surface plasmon resonance and mutagenesis experiments. They showed that SARS-CoV N-protein binds to cyclophilin A with a much higher affinity than that of HIV-1 capsid protein (100-2700 times higher). This observation may provide some evidence of a possible pathogenicity factor of SARS-CoV (Luo, Luo *et al.* 2004).

Aims of this thesis

Human coronaviruses have recently been shown to cause both upper and lower respiratory tract infections which often leads to hospitalization. With the exception of SARS-CoV, the other four known human coronaviruses - HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HKU - continually circulate in the human population. HCoV-NL63 was identified in

2004 from the nasal aspirate of a child with respiratory symptoms. Although many reports of clinical data for this virus have since been reported, the virus is not well characterized at the molecular and protein levels. Therefore, this dissertation focuses on the *in silico* and *in vitro* characterization of the HCoV-NL63 structural nucleocapsid protein.

Bioinformatic tools will be used to characterize the HCoV-NL63 N-protein, as well as identify selected motifs and/or regions important for protein functionality. Comparing HCoV-NL63 N to other coronavirus homologues will shed light on the possible role(s) of this protein in determining HCoV-NL63 pathogenicity and infectivity.

To identify antigenic epitopes on the N-protein, patient serum samples will be screened. For this reason HCoV-NL63 N-gene will be amplified and cloned for expression in a bacterial system. Purified recombinant HCoV-NL63 N-protein will be used to develop an ELISA that will be used to screen the patient samples.

Not much is known about the proteins of HCoV-NL63, including the N-protein. The work that will be reported in this thesis, will serve as a prelude to extensive laboratory characterization of this protein. It will therefore form a crucial “first-step” in many other projects - which will fall outside the scope of this MSc project -, including in the generation of antibodies for the development of an antigen-capture ELISA that will be used to screen patient samples collected in the Western Cape.

Major Outcomes will include

- Bioinformatic characterization of the HCoV-NL63 N-protein.

- Cloning, expression and purification of the HCoV-NL63 N-protein and generation of truncated clones.
- Characterizing the N-specific antibody cross-reactivity between coronavirus species and identify the antigenic epitope of the HCoV-NL63 N-protein.



Chapter 2

Bioinformatic Analysis of the HCoV-NL63 Nucleocapsid Protein



This chapter is an expanded version of the manuscript in preparation for submission as: M. Berry, Y.J. Tan and B.C. Fielding, Analysis of the human coronavirus-NL63 nucleocapsid protein, African Journal of Biotechnology. For this manuscript the Results and Discussion sections are combined.

1. Abstract

The advent of genome sequencing technologies has resulted in large databases of genomes and the rate of genome sequencing has surpassed our ability to determine the function of the proteins they encode. *In silico* analysis has given us the ability to predict the function of these proteins, which can then be verified with the use of *in vitro* and *in vivo* studies.

Bioinformatic tools were used for the comparative analysis of HCoV-NL63 N and selected coronaviruses N-homologues. Multiple sequence alignment tools were used to identify conserved regions and/or motifs within the coronavirus N-proteins. Next, the order/disorder state (which may indicate regions of interaction) of N was identified with PONDR. NetPhos 2.0 and DEPP were used to identify the putative phosphorylation state of the N-protein. Lastly, the peptide select tool of Invitrogen was used to characterize the antigenicity of the N-protein and identify putative antigenic epitopes.

The HCoV-NL63 N-gene was predicted to encode a ~42.6 kDa protein most similar to HCoV-229E N (43% amino acid identity; 63% amino acid similarity). PONDR identified one disordered region in the HCoV-NL63 N-protein. N was shown to be highly phosphorylated by multiple kinases at multiple serine residues. HCoV-NL63 N was shown to be highly antigenic with a spike in antigenicity from residues ~140-180 for HCoV-NL63. This manuscript reports the computational analysis of the human coronavirus NL63 N-protein by comparing the amino acid sequences of coronavirus N-homologues.

2. Introduction

The recent advent of genome sequencing technologies has provided large genome sequence databases and the rate at which genomes are sequenced has surpassed our ability to determine the function of the proteins these genomes encode. *In silico* analysis provides us with a tool to predict possible functions of these proteins by comparing novel proteins of unknown function with proteins of known function, or by analyzing the unknowns using an algorithm based on proteins of known function. These results can then be verified with the use of *in vitro* and *in vivo* studies.

All N-proteins have the same primary function, to form the ribonucleoprotein core. However, using multiple protein sequence alignment with ClustalX the coronavirus N-homologues were shown to have low sequence homology. Formation of this structure requires self-association of the N-protein into dimers and higher oligomeric structures, as well as interaction with viral RNA. This may indicate that the regions of interaction within the N-protein share a higher sequence homology than that of the overall N-protein.

SARS-CoV N is the most highly characterized of the coronavirus N-proteins, with the first evidence of dimerization shown using a mammalian two-hybrid system. A serine/arginine (SR) rich motif between residues 184 and 196 was shown to be essential for the dimerization of N (He, Dobie *et al.* 2004). Another study showed that dimerization of N occurs *via* a 209 amino acid region of the C-terminal, which lay upstream of the SR rich motif (Surjit, Liu *et al.* 2004). These results show that the process of N dimerization is complicated. In two later studies this C-terminal region was, however, reduced to residues 283-422 (Luo, Ye *et al.* 2005; Yu, Gustafson *et al.* 2005). The SR rich motif was found to be responsible for dimerization by interacting with the middle region (a.a. 211-290) and dimerization of this

region inhibits the formation of multimers in the C-terminal (a.a. 283-422) domain (Luo, Ye *et al.* 2005). A C-terminal fragment between 343 and 402 (Luo, Chen *et al.* 2006) was later found to be sufficient for oligomerization of the N-protein. As this region has been supported by several researchers we have assumed it to be the probable region responsible for dimerization with the function of the SR rich motif yet to be properly elucidated.

Other Coronavirus N-proteins are less well characterized with the dimerization domains of only MHV, IBV and OC43 being mapped. The region of dimerization of the MHV N-protein was shown to occur *via* the N-terminal residues, 1-292, using a GST-binding assay (Wang and Zhang 1999). The IBV N-protein was found to dimerize at the C-terminal, spanning residues 218-329 (Fan, Ooi *et al.* 2005). Finally, the dimerization region of OC43 was identified to be between residues 301 and 448 of the C-terminal (Huang, Hsu *et al.* 2009).

Another protein-protein interaction pivotal in viral particle assembly is that between the N- and M-protein. The interaction links the ribonucleocapsid core to the envelope of the virus as the M-protein is the most highly expressed envelope protein (Kuo and Masters 2002). This region has only been characterized for SARS-CoV, where it was identified by a mammalian two-hybrid system and located between residues 168-208 (He, Leeson *et al.* 2004), and for MHV which was characterized as a negatively charged domain in the C-terminal, spanning residues 406-454 (Verma, Bednar *et al.* 2006).

The N-protein has been shown to interact with the leader RNA sequence (Baric, Nelson *et al.* 1988), the 3'-end of the genomic RNA (Nelson, Stohlman *et al.* 2000; Zhou and Collisson 2000), the transcriptional regulatory sequence (Grossoehme, Li *et al.* 2009) and the packaging signal (Chen, Gill *et al.* 2005). Interactions between the CoV RNA and N

predominantly occur *via* the N-terminal domain. However, several binding sites, spanning both the N- and C-terminal domains of N as well as the 3 disordered regions linking these domains, are involved in RNA binding (Chang, Hsu *et al.* 2009). For SARS-CoV N these binding sites were mapped to the N-terminal amino acid (a.a.) residues 45-181 (Huang, Yu *et al.* 2004; Chang, Sue *et al.* 2006; Chen, Chang *et al.* 2007) and to the C-terminal residues 248-365 (Chen, Chang *et al.* 2007), which has a stronger nucleic acid binding affinity than the N-terminal. The RNA binding domain of MHV N was mapped to residues 177-231 (Nelson, Stohlman *et al.* 2000). HCoV-OC43 N has 3 possible RNA binding domains, spanning the N-terminus (a.a. 1-173) and middle region (a.a. 174-232 and a.a. 233-300), with the N-terminal portion having the highest affinity for RNA (Huang, Hsu *et al.* 2009). Finally, the RNA binding domain of IBV N was mapped between residues 91-171 of the N-terminal and residues 203-409 of the C-terminal, where binding affinity was greater at the C-terminal domain (Zhou and Collisson 2000). As the N-terminal of CoV N-proteins is more highly conserved (Sastre, Dijkman *et al.* 2011) we have aligned these regions to identify sequence homology.

Multiple sequence alignment of these experimentally proven regions within the N-protein of the previously mentioned coronaviruses showed very low sequence homologies with few conserved regions. These results possibly indicate that other properties play a bigger role in interactions of the N-protein. For this reason we have chosen to identify the order/disorder state of the coronavirus N-proteins.

A disordered protein, or region of a protein, refers to the lack of a tertiary structure and these proteins have no fixed 3 dimensional shape. A disordered protein has various functions including DNA, RNA and protein binding as the disordered region allows access to these

binding sites. On binding, the protein is understood to go through a disorder-to-order transition (Dunker, Garner *et al.* 1998; Garner, Romero *et al.* 1999; Dunker, Lawson *et al.* 2001). Interestingly, the amino composition, sequence complexity, hydrophobicity and charge *etc.* adjacent to phosphorylation sites is similar in disordered regions and therefore indicates that disorder is important for phosphorylation. We can therefore assume that the majority of predicted phosphorylation sites occur within the disordered regions of the N-protein (Iakoucheva, Radivojac *et al.* 2004).

Given that the order/disorder state of a protein may play a large role in homologous and heterologous protein interactions, it would also be interesting to identify the antigenic state of the protein and attempt to identify whether there is an increase in antigenicity over these disordered regions. Two methods, namely the Welling (1985) and Parker (1994) methods, were used to identify the antigenic state of the SARS-CoV and HCoV-NL63 N-protein (Welling, Weijer *et al.* 1985; Parker, Bednarek *et al.* 1994).

We have used various bioinformatic tools to compare the N-proteins of the novel coronavirus, HCoV-NL63 (Amsterdam1 strain) and other coronaviruses. The HCoV-NL63 N protein sequence was shown most similar to HCoV-229E N (43% amino acid identity; 63% amino acid similarity). PONDR identified one disordered region in HCoV-NL63 N. N was shown to be highly phosphorylated by multiple kinases at multiple serine residues. HCoV-NL63 N was predicted to be highly antigenic with a spike in antigenicity from residues ~140-180 for HCoV-NL63.

3. Materials and Methods

3.1 Multiple Sequence Alignment

Nucleotide sequences were obtained from NCBI (accession numbers in brackets): HCoV-NL63 (DQ846901.1), SARS-CoV (AY360146.1), MHV (AAU06361), IBV (AAA46214), HCoV-OC43 (AY585229.1), HCoV-229E (AAG48597), TGEV (AAA47915), PEDV (AF353511), HKU1 (AAT98585) and HKU7 (DQ666343). Selected N-homologues 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).

3.2 Order-Disorder Prediction

PONDR (Predictors of Natural Disordered Regions) was used to predict possible disordered regions within the SARS-CoV and HCoV-NL63 N-protein. PONDR (<http://www.pondr.com>) uses several algorithms, including VL-XT, XL1-XT, VL3 and VSL1 to predict possible disordered regions. We have chosen to use the VL-XT and VSL1 algorithms. The VL-XT incorporates 3 feedforward neural networks which are based on long disordered regions characterized by x-ray crystallographic data (Romero, Obradovic *et al.* 1997; Li, Romero *et al.* 1999; Romero, Obradovic *et al.* 2001). The VSL1 utilizes two algorithms for short (<30aa) and long (>30aa) disordered regions and is trained on 1335 protein sequences (Obradovic, Peng *et al.* 2005).

3.3 Phosphorylation State Prediction

NetPhos 2.0 and the DEPP (Disorder Enhanced Phosphorylation Predictor) programs of PONDR were used to predict possible phosphorylation sites in the N-protein of HCoV-NL63 and SARS-CoV (Blom, Gammeltoft *et al.* 1999). NetPhos 2.0 predicted serine, threonine and

tyrosine phosphorylation sites when threshold values (ranging from 0.0000 to 1.0000) were above 0.5. However values closer to 0.5 are less likely to be functional phosphorylation sites compared to motifs closer to 1.0. DEPP is based on 1500 experimentally confirmed phosphorylation sites. 3Dinsight motif finder was used to predict possible kinases responsible for the phosphorylation of these residues by identifying specific patterns in the amino acid sequence.

3.4 Characterization of Antigenic State

The peptide select tool of Invitrogen, which utilizes methods described by Welling (1985) and Parker (1994), was used to characterize the antigenic state of the SARS-CoV and HCoV-NL63 N-protein. The Welling method describes how an immunogenic peptide has an approximate length of 6-7 amino acids, and based on previously identified antigenic epitopes from 20 proteins, an average composition of an antigenic region can be identified. The Welling method therefore scans each 7 amino acids to identify the antigenicity of that region (Welling, Weijer *et al.* 1985). The Parker method utilizes a similar approach, however, it also incorporates amino acid side chains (Parker, Bednarek *et al.* 1994).

4. Result and Discussion

4.1 Multiple Sequence Alignment

The N-protein of HCoV-NL63 was compared to several other coronavirus nucleocapsid homologues using multiple sequence alignment with ClustalX to determine degree of sequence homology, as well as the presence of any possible conserved regions within the selected coronavirus N sequences. The amino acid sequence of HCoV-NL63 N was compared to other coronavirus N proteins (Table 2.1). HCoV-NL63 was shown to be most similar to the human coronavirus HCoV-229E N (43% identity; 63% similarity). This was not totally unexpected, since previous molecular clock analyses estimate that HCoV-NL63 and HCoV-229E evolved from a most recent common ancestor in the 11th century (Pyrç, Dijkman *et al.* 2006). Interestingly, HCoV-NL63 N also showed high amino acid similarity to the bat coronavirus HKU7 N (40% identity; 57% similarity). As speculated previously (Fielding and Suliman 2009), this high amino acid similarity raises interesting questions about the relationship between this human and bat coronavirus: Could a bat be the original host of HCoV-NL63? In fact, present evidence shows that bat coronaviruses are the gene pools of the Group 1 and Group 2 coronaviruses (Woo, Lau *et al.* 2009).

Table 2.1: Multiple sequence alignment of 10 coronavirus nucleocapsid homologues. Sequence identity (top value) and similarity (bottom value) is shown.

	NL63	SARS	MHV	IBV	OC43	229E	TGEV	PEDV	HKU1	HKU7
NL63	377	20% 33%	17% 30%	15% 30%	16% 31%	43% 63%	31% 51%	36% 50%	18% 33%	40% 57%
SARS		422	28% 42%	20% 33%	26% 42%	19% 33%	22% 38%	17% 30%	27% 42%	20% 33%
MHV			454	17% 31%	70% 81%	18% 33%	18% 35%	15% 30%	65% 80%	17% 31%
IBV				409	16% 30%	13% 25%	16% 30%	14% 27%	17% 30%	15% 28%
OC43					448	18% 32%	18% 33%	14% 30%	62% 76%	14% 31%
229E						389	32% 49%	31% 48%	19% 33%	38% 55%
TGEV							382	25% 43%	19% 36%	31% 48%
PEDV								441	14% 29%	37% 52%
HKU1									441	17% 34%
HKU7										391



UNIVERSITY of the
WESTERN CAPE

Comparative analysis of 10 coronavirus species identified a 9 amino acid region conserved between the N homologues, *i.e.* FYY(L/S)(G/A)TGPH (amino acid 78-86 for HCoV-NL63, Figure 2.1). Conserved regions within a family of proteins are commonly indicative of structural and functional conservation, as well as an evolutionary relationship. The function of this sequence, if any, is yet to be elucidated. However, this conserved amino acid region was shown to be present in previously identified RNA binding domains of other coronavirus N-proteins (Figure 2.2), indicating a possible role in RNA binding. Site-directed mutagenesis of residues in this region may be necessary to elucidate any definitive role it plays in RNA binding.

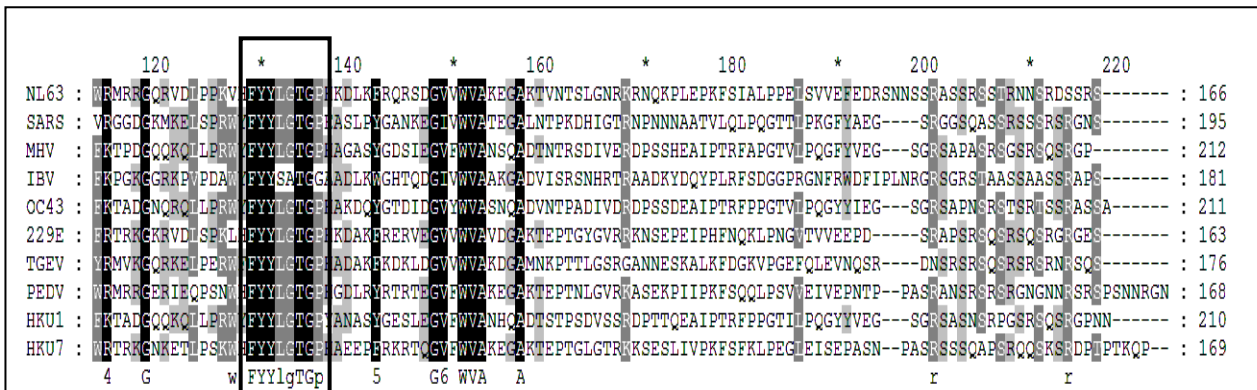


Figure 2.1: Alignment of the amino acid sequences of HCoV-NL63 N and selected coronavirus N-homologues. Shading indicates conserved regions (black: 100% homology; dark grey: 80% homology; light grey: 60% homology) and gaps were introduced to align sequences. The conserved motif is boxed.

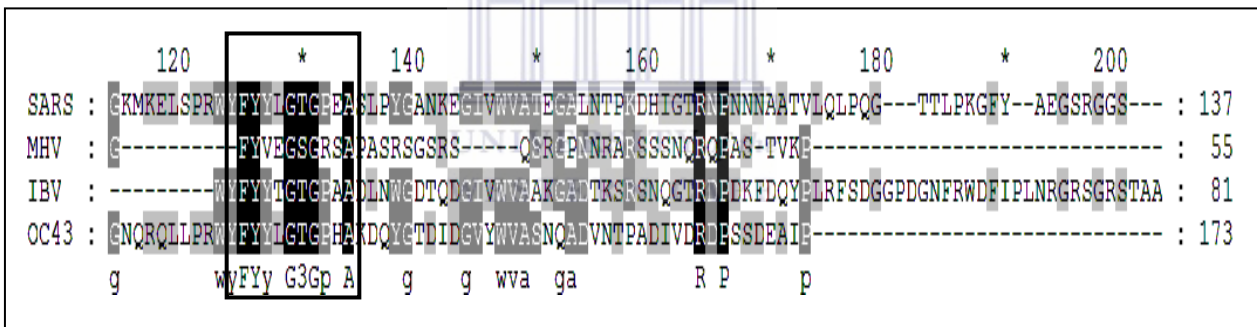
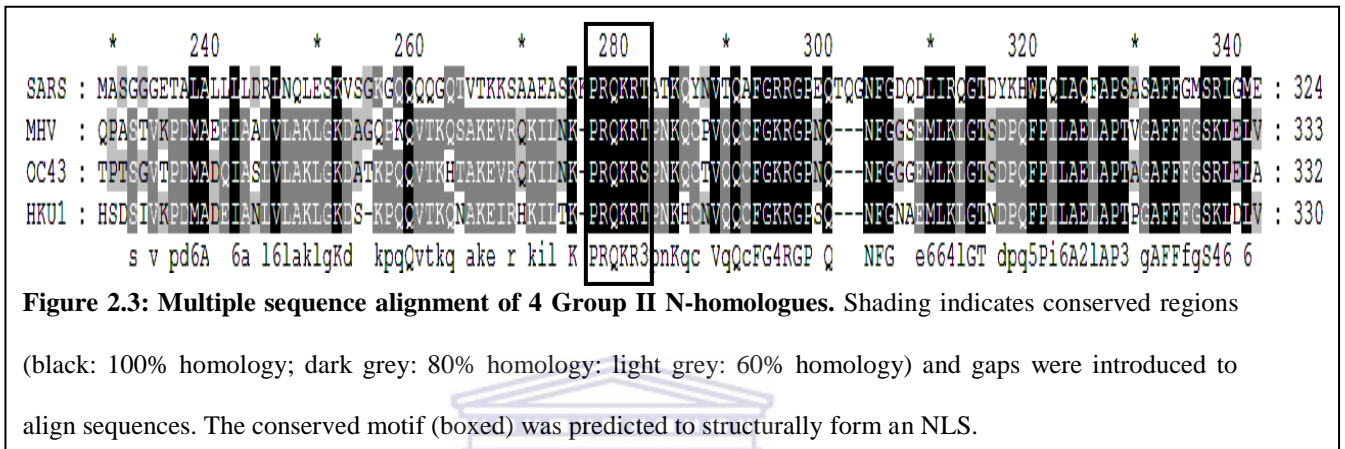


Figure 2.2: A Conserved region (boxed) within the N-terminal of the nucleocapsid protein of 4 coronaviruses. Regions aligned have been identified from literature and are shown to be responsible for RNA binding (references can be found in Table 2.2). Shading indicates conserved regions (black: 100% homology; dark grey: 80% homology; light grey: 60% homology) and gaps were introduced to align sequences.

The N-protein of Group II coronaviruses show a higher sequence homology compared to that of Group I coronaviruses, with several more conserved regions found within the viruses of Group II. The 4 coronaviruses analyzed from Group II (SARS-CoV, HCoV-OC43, MHV and HKU1) contained the conserved motif of PRQKR (amino acid: 259-264 for SARS-CoV:

Figure 2.3), which was predicted by PSORTII to structurally form a nuclear localization signal (NLS). This indicates that conserved regions within a protein likely form a functional unit. This conserved region is absent in Group I viruses.



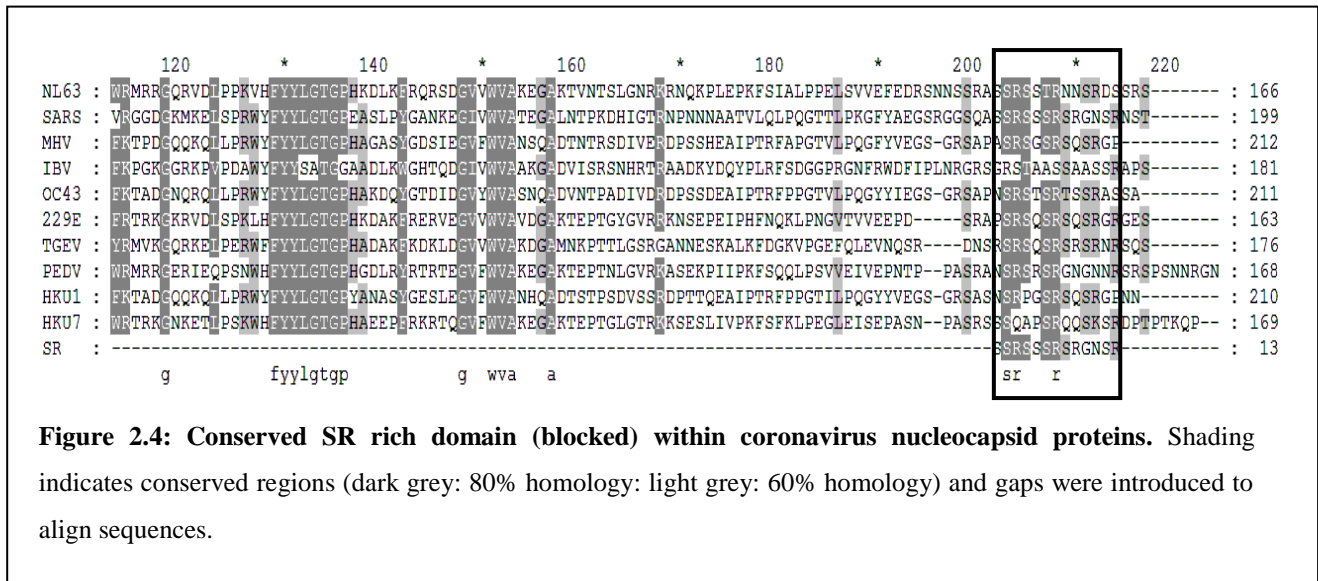
Multiple sequence alignment has shown an overall low sequence homology between the N homologues, however all N-proteins have the same primary function *i.e.* the formation of the ribonucleocapsid core. This process involves N-N homotypic interactions and RNA-N interactions. These have been experimentally plotted by various researchers and the domains have been summarized in Table 2.2, which also summarizes the region of N-M interaction for SARS and MHV N. M-N heterotypic interactions have been shown to be important in viral assembly. We have used multiple sequence alignment to determine whether these experimentally determined regions have a higher sequence homology compared to that of the overall homology.

Table 2.2: Regions of interaction within the coronavirus nucleocapsid proteins.

	Region of Multimerization	M-N Interaction	Region of RNA Interaction
SARS-CoV	343-402 (Luo, Chen <i>et al.</i> 2006)	168-208 (He, Leeson <i>et al.</i> 2004)	45-181 (Huang, Yu <i>et al.</i> 2004)
MHV	1-292 (Wang and Zhang 1999)	406-454 (Verma, Bednar <i>et al.</i> 2006)	177-231 (Nelson, Stohlman <i>et al.</i> 2000)
IBV	218-329 (Fan, Ooi <i>et al.</i> 2005)		91-171 (Zhou and Collisson 2000)
HCoV-OC43	301-448 (Huang, Hsu <i>et al.</i> 2009)		1-173 (Huang, Hsu <i>et al.</i> 2009)

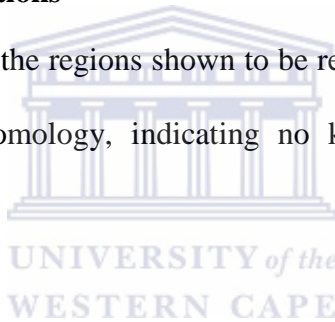
4.1.1 Nucleocapsid Dimerization

Multiple sequence alignment of the fragments responsible for N-N interactions showed very low sequence homology with the highest being between SARS-CoV and HCoV-OC43 (10% identity and 23% similarity). These results show even lower homology than compared to that of the full length N homologues; no conserved regions were identifiable between the 4 fragments. This indicated that there is probably no consensus sequence for N-protein dimerization. However, the SR rich motif (responsible for dimerization) was relatively well conserved between the 10 coronavirus species aligned (Figure 2.4). There is a lot of contradictory information regarding the dimerization process of the N-protein - especially with regards to SARS-CoV N - and therefore the role of this SR-rich motif may be under-presented.



4.1.2 N-M Heterotypic Interactions

Multiple sequence alignment of the regions shown to be responsible for N-M interactions also showed very low sequence homology, indicating no known conserved region for N-M interactions.



These results indicated that sequence homologies play little or no role in homologous or heterologous protein-protein interactions as few conserved regions were identifiable within these experimentally determined regions. It may also indicate that structural properties, such as order/disordered state, electrostatic forces and disulphide bond formation, within the protein may be more significant in protein-protein interactions compared to sequence homology.

4.2 Order/Disorder

Using the VL-XT and VSL1 algorithms the SARS-CoV N-protein has been shown to have 3 possible disordered regions located at the N-terminus (residue 1-69), middle (residue 152-292) and C-terminus (residue 358-422) (Figure 2.5).

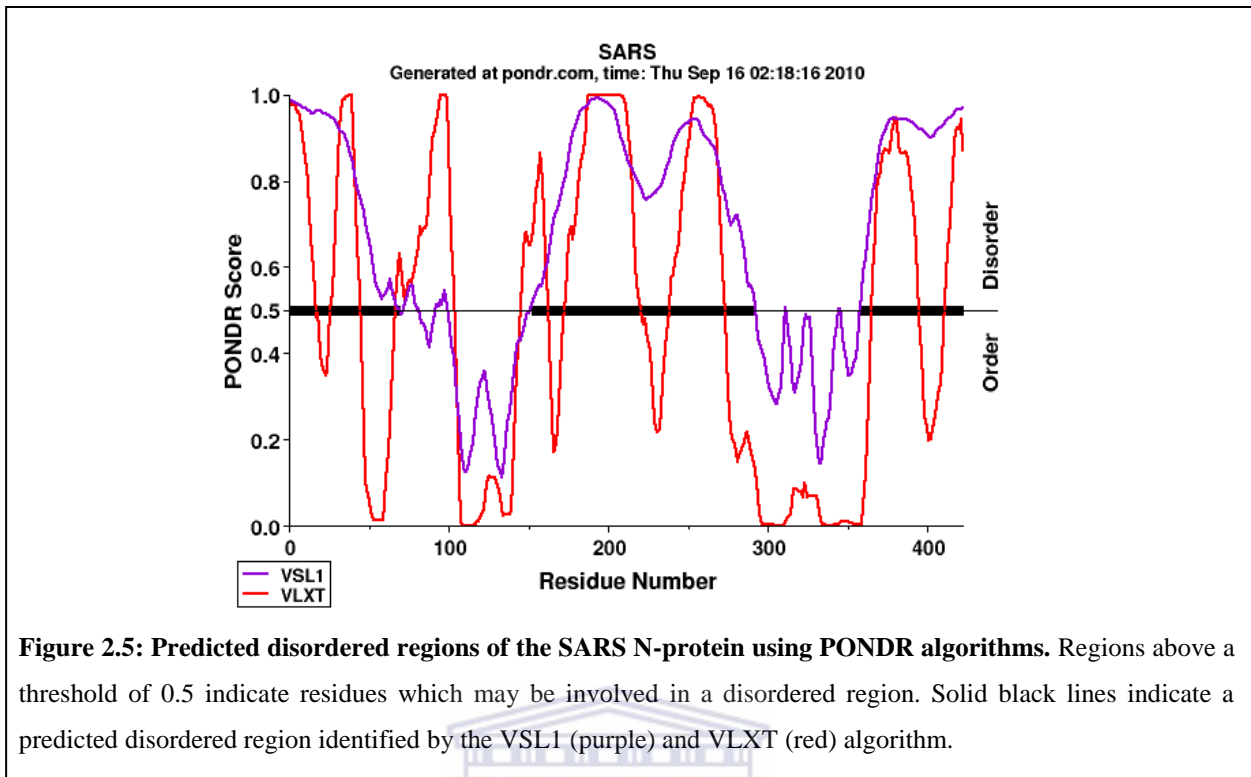


Figure 2.5: Predicted disordered regions of the SARS N-protein using PONDR algorithms. Regions above a threshold of 0.5 indicate residues which may be involved in a disordered region. Solid black lines indicate a predicted disordered region identified by the VSL1 (purple) and VLXT (red) algorithm.

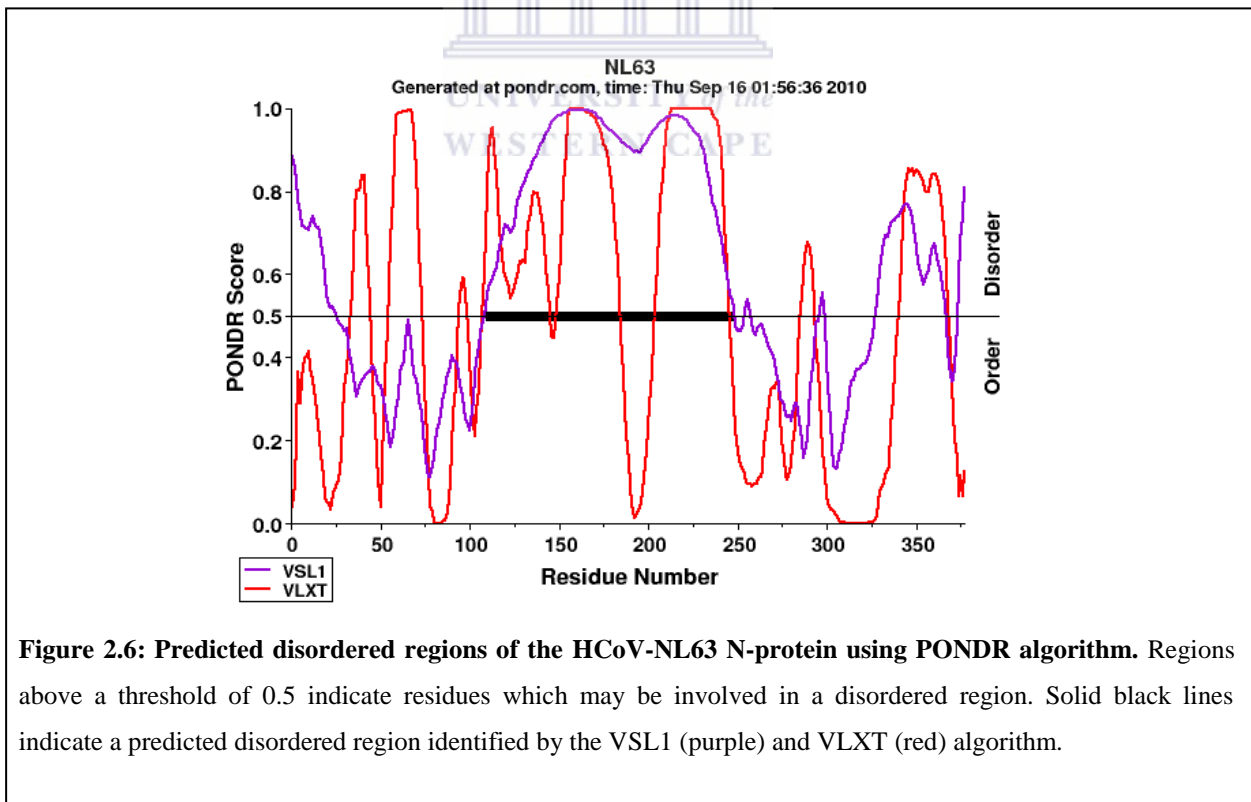


Figure 2.6: Predicted disordered regions of the HCoV-NL63 N-protein using PONDR algorithm. Regions above a threshold of 0.5 indicate residues which may be involved in a disordered region. Solid black lines indicate a predicted disordered region identified by the VSL1 (purple) and VLXT (red) algorithm.

HCoV-NL63 N has only one putative disordered region spanning residues 109-248 (Figure 2.6). Chang *et al.* (2005) used NMR to verify these claims and found that fragments 1-44, 182-247 and 366-422 of SARS-CoV N are, in fact, disordered. They also showed that several other coronavirus N homologues share the same pattern as that of SARS-CoV N (Chang, Sue *et al.* 2006), raising questions as to why HCoV-NL63 is different.

SARS-CoV N interacts with the M-glycoprotein through residues 168-208 (He, Leeson *et al.* 2004). In this study, we used PONDR to show that this region falls within the middle disordered region of the SARS-CoV N indicating a possible necessity for state-of-disorder in interactions. In fact, the previously identified regions responsible for dimerization are located in disordered regions within the SARS-CoV N. (Surjit, Liu *et al.* 2004). Similarly, the SR-rich 13 residue segment identified by He and colleagues (He, Dobie *et al.* 2004) and the RNA binding domain of SARS-CoV N (Huang, Yu *et al.* 2004), falls within disordered regions. If disorder is essential for RNA binding, the spanning of 2 disordered regions may explain the helical formation of the ribonucleoprotein.

The presence of 3 disordered regions in SARS-CoV N compared to the 1 disordered region in HCoV-NL63 N may indicate that SARS-CoV N has a higher binding affinity to viral, as well as host cellular proteins. For example SARS-CoV N has been shown to tightly bind human cyclophilin-A through residues 235-369 (Luo, Luo *et al.* 2004), of which approximately 60 residues fall within a disordered region. The binding of cyclophilin-A to the N-protein has a great impact on pathogenicity of the SARS-CoV virus. This may indicate a probable basis for the increased pathogenicity of SARS-CoV compared to HCoV-NL63.

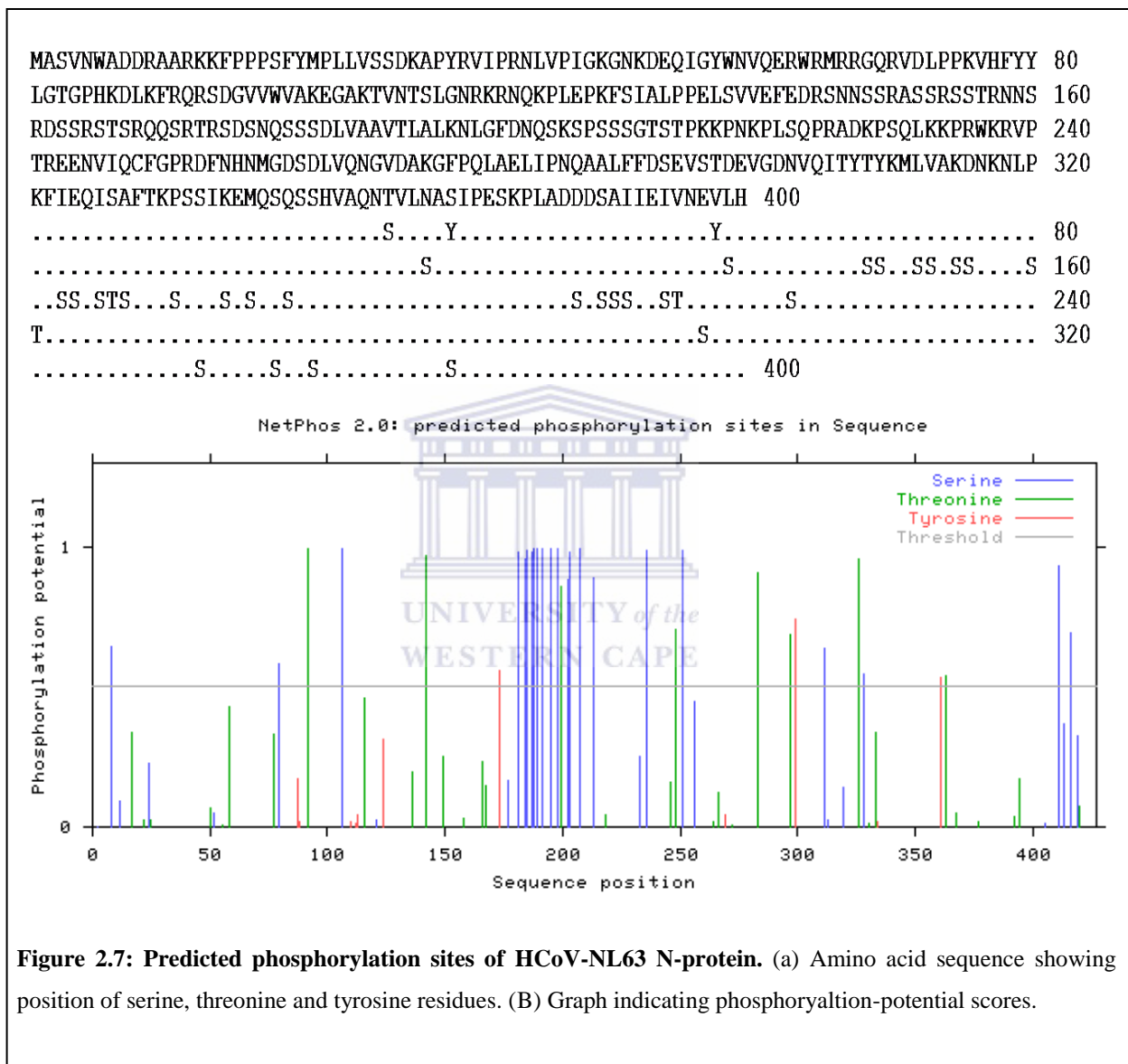
These results possibly indicate that the order-disorder structure of the N-protein plays a bigger role in protein interaction than that of sequence homology. If disorder is essential for these interactions, these results indicate that these interactions should take place between amino acids 109-248 of the HCoV-NL63 N-protein. Further experimental analysis, however, is needed to confirm these results.

4.3 Prediction of Phosphorylation Sites

For the HCoV-NL63 N a total of 34 possible phosphorylation sites were identified with NetPhos 2.0 including 29 serine, 3 threonine and 2 tyrosine residues; DEPP predicted 42 possible phosphorylation sites with 32 serines, 6 threonines and 4 tyrosines. These results indicate that HCoV-NL63 N-protein is heavily serine phosphorylated at multiple residues (Figure 2.7). 3Dinsight motif finder identified multiple residues for phosphorylation by protein kinase C and casein kinase II. Protein kinase C exhibits a preference for the phosphorylation of serine or threonine residues found near to a C-terminal basic residue (Kishimoto, Nishiyama *et al.* 1985; Woodgett, Gould *et al.* 1986). Casein kinase II is another serine and threonine kinase, but has a preference for acidic residues (Pinna 1990).

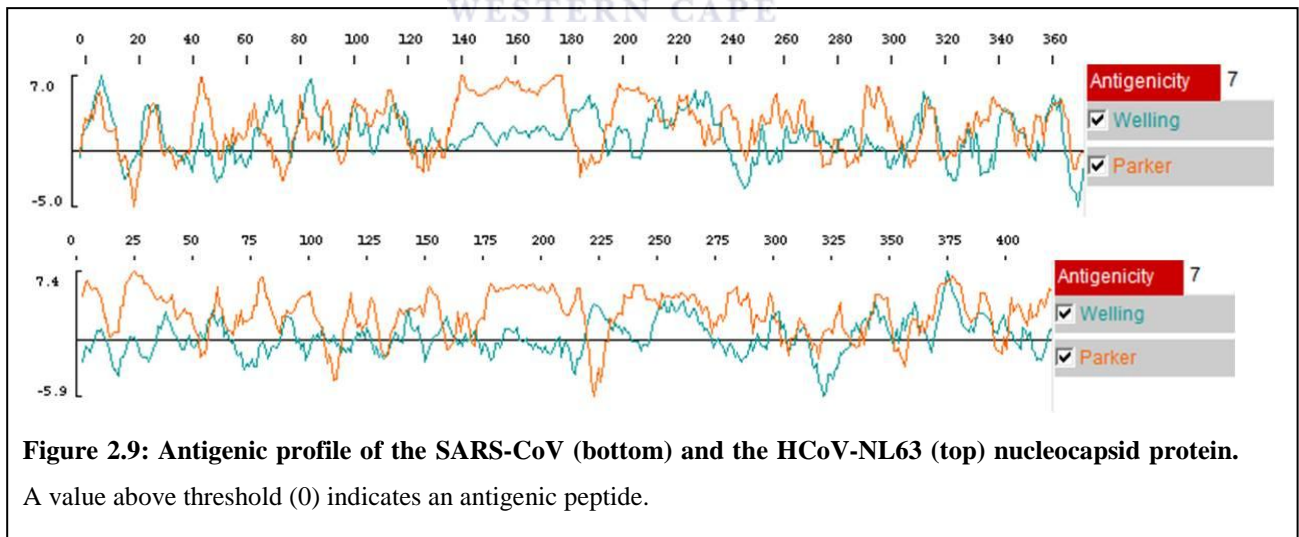
SARS-CoV N showed similar results with NetPhos 2.0 indicating a total of 33 possible phosphorylation sites of which 22 were serine, 8 threonine and 3 tyrosine. DEPP showed a total of 43 phosphorylation sites with 24 serine, 13 threonine and 6 tyrosines. This indicated that SARS-CoV N was also heavily serine phosphorylated at multiple residues (Figure 2.8). The SARS-CoV N also has predicted protein kinase C and casein kinase II phosphorylation sites, as well as 2 cAMP- and cGMP-dependant phosphorylation sites, which have been shown to

phosphorylate serine and threonine residues in close proximity to at least two consecutive basic residues close to the N-terminal (Glass and Smith 1983; Glass, el-Maghrabi *et al.* 1986).



4.4 Characterization of the antigenic state

The methods described by Welling and Parker showed that both the SARS-CoV and HCoV-NL63 N-protein were highly antigenic. The results indicated that the antigenic profiles of both N-proteins were similar with the Parker method showing a spike in antigenicity from residues ~140-180 for HCoV-NL63 and residue ~175-220 for SARS-CoV (Figure 2.9). The method described by Welling could have missed this region of antigenicity as it does not incorporate amino acid side chains. Since both these regions fall within the disordered regions of these proteins, it could highlight another role of disorder in protein physiology. As both properties are independent of each other, it could also indicate an evolutionary relationship that the immune response has developed against amino acids which form the accessible disordered region and raises the question, “are immunogenic epitopes preferentially disordered?”



5. Conclusion

The HCoV-NL63 genome is approximately 27 553 nucleotides in size, with the typical coronavirus genome organization (Pyrce, Dijkman *et al.* 2006). It produces six separate mRNAs, with all potential ORFs encoding for viral proteins. N has been shown to be expressed from distinct subgenomic (sg) mRNA 6, the most abundant sg mRNA (Pyrce, Jebbink *et al.* 2004). Expression of recombinant HCoV-NL63 N in a bacterial system yielded a ~44 kDa protein (result not shown) which corresponded to the size of the putative 42.6 kDa predicted by our initial analysis (<http://www.sciencegateway.org/tools/proteinmw.htm>).

Besides the relatively high conservation between the N-protein of NL63 and HCoV-229E, other coronavirus homologues showed a significantly lower sequence homology. All coronavirus N-proteins, however, have the same primary function, *i.e.* the formation of the ribonucleocapsid core. This process requires self-association of the N-protein into dimers and higher oligomeric structures as well as interaction with viral RNA. Once this particle has assembled, the N-protein has been shown to interact with the M-protein during virion assembly. Is it then possible that these regions of interaction share a higher degree of homology compared to the full length N-protein?

The conflicting literature on the dimerization status of the coronavirus N-protein raises a lot of questions as to the correct process involved in this interaction. Two possible regions have been identified to be responsible for the dimerization process; that being 1) the C-terminal domain and 2) a SR-rich region. Multiple sequence alignment has shown that the C-terminal region is weakly conserved between coronavirus N-homologues; however, the SR-rich region has been shown to be relatively highly conserved. Do the conflicting results indicate that when truncated mutants of full-length N are cloned, several regions within the N-

protein are capable of forming dimers, with only one region preferentially forming a dimer *in vivo*? SR rich domains have also previously been shown to mediate RNA binding, with the RS domains being required to mediate protein-protein interactions (Graveley and Maniatis 1998). Various SR rich proteins have been shown to have a major involvement in pre-mRNA splicing, however several researchers have also shown these domains to be responsible for mediating protein-protein interactions (Yuryev, Patturajan *et al.* 1996; Bourquin, Stagljar *et al.* 1997; Yang, Embree *et al.* 1998). These results show that both the SR and RS domains can mediate protein-protein interactions and therefore may be responsible for mediating the dimerization of the N-proteins.

With exception of the SR-motif, multiple sequence alignment reveals very little homology between regions shown to take part in dimerization and N-M heterotypic interactions. These regions do however fall within a disordered region, indicating that the protein's order/disorder structure plays an important role in mediating these interactions as well as the phosphorylation state of a protein, as the majority of predicted phosphorylation sites occur within the middle disordered region. *In silico* analysis of proteins can therefore play an important function in determining which regions to construct as recombinant clones prior to the initiation of *in vivo* and *in vitro* studies.

In silico analysis is a good indication of the structure of a protein, however, errors can occur and therefore claims must be substantiated with experimental evidence. The order/disorder state on the HCoV-NL63 N-protein has been reported to be different to that of other coronavirus N-proteins, with HCoV-NL63 N only having one disordered region compared to the 2-3 disordered regions within other coronavirus N-proteins (Chen, Gill *et al.* 2005). As majority of protein-protein interactions occur within these disordered regions, it

indicates that residues 109-248 of the HCoV-NL63 N will form the hub of these interactions. This region is also understood to be highly phosphorylated and involved in eliciting an immune response as it contains highly antigenic peptides. This information will be used in the construction of recombinant clones to verify these claims *in vitro*.



UNIVERSITY OF THE WESTERN CAPE

Chapter 3

Cloning and Expression of the HCoV-NL63 Nucleocapsid Protein



**Department of Medical Bioscience, University of the Western Cape, Cape Town, South
Africa**

1. Abstract

Bacterial expression of proteins remains to be a popular method for the production of heterologous, recombinant proteins, which do not require post-translational modifications such as phosphorylation and glycosylation for activity. The nucleocapsid protein of HCoV-NL63 and SARS-CoV has been cloned and expressed in a bacterial system for later use in an antibody capture ELISA for the detection of serum antibodies against HCoV-NL63.

RT-PCR was used to generate cDNA using viral RNA as a template. The N and truncated N-genes of HCoV-NL63 were amplified from this cDNA, where the SARS-CoV N-gene was amplified from a pXJHA40 vector. Amplified genes were then cloned into a pGEM[®] T-easy sequencing vector and used to transform JM109 *E. coli*. Once sequencing was confirmed, plasmids were isolated by alkaline lysis and the amplified gene was restricted using enzymes specific for ligation into a Flexi[™] protein expression vector. This gene-vector construct was used to transform KRX *E. coli* which served as our protein expression model. Protein expression was confirmed by Western Blot analysis against the GST fusion protein, which was also used to facilitate purification. Once proteins were purified, the fusion protein was removed using ProTEV protease.

The N-proteins of SARS-CoV and HCoV-NL63 were successfully expressed using the Flexi[™]/KRX expression system; however, a large degree of troubleshooting was necessary to optimize expression levels where several problems, including insolubility and background expression, were encountered. The system was also unable to adequately express the middle disordered region of the HCoV-NL63 N-protein.

2. Introduction

The nucleocapsid (N) protein of coronaviruses is abundantly expressed during infection and is also highly immunogenic (Timani, Ye *et al.* 2004). For this reason it poses a good target for the development of a diagnostic assay. However, as a large majority of people have been exposed to HCoV-NL63 and other human coronaviruses (HCoV), with most HCoVs being able to cause re-infection throughout life (Sastre, Dijkman *et al.* 2011), the development of a diagnostic assay to screen for seroconversion would pose irrelevant as antibodies are almost always present. The N-protein has also been shown to be capable of cross-reaction with serum antibodies against other HCoVs (Sun and Meng 2004; Woo, Lau *et al.* 2004; Chan, Cheng *et al.* 2005; Che, Qiu *et al.* 2005; Vlasova, Zhang *et al.* 2007; Venter, Lassauniere *et al.* 2011). Understanding how these antigens interact with serum antibodies against the virus is therefore of importance prior to the production of antibodies for the generation of an antigen capture enzyme linked immunosorbent assay (ELISA).

Bacterial expression of proteins would allow for the production of the N-protein to characterize this interaction. Bacterial expression remains to be a popular method for the production of heterologous, recombinant proteins, which do not require post-translational modifications such as phosphorylation and glycosylation for activity. This method of protein production is used by both researchers and in the large scale manufacture of pharmaceuticals such as insulin (Masters 1992) due to the ability of bacterial cultures to rapidly reach a high optical density on relatively inexpensive substrates (François 1999). Bacterial expression is even capable of heteromultimeric expression of heterologous proteins as seen with the expression of the α and β subunits of human hemoglobin to form a fully functional tetrameric structure (Hoffman, Looker *et al.* 1990).

This chapter will detail the cloning and bacterial expression of recombinant clones of the SARS-CoV and HCoV-NL63 N-genes corresponding to full-length SARS-CoV N and HCoV-NL63 N (N1), the N-terminal domain of HCoV-NL63 (N2), the C-terminal domain of HCoV-NL63 (N3) and the middle region of HCoV-NL63 (N4). The SARS-CoV N-gene is 1269 base pairs (bp) and codes a putative 422 amino acid (aa) protein with a molecular weight of 46.05kDa. Full length SARS-CoV N will allow us to identify any antibody cross-reactivity between SARS-CoV and HCoV-NL63 antibodies. N1 is an 1134 bp sequence which codes a 377 aa protein with a molecular weight of 42.6 kDa. N2 corresponds to the N-terminal half of N1 and spans residues 1-189 with a molecular weight of 21.51 kDa. N3 corresponds to the C-terminal half of N1 and codes for residues 190-377 with a molecular weight of 20.9 kDa. Finally, N4 codes for the middle third region and overlaps both N2 and N3 and codes for aa's 127-252 with a molecular weight of 14.01 kDa (Table 3.1).

Table 3.1: Description of heterologously expressed proteins.

Gene Name	Nucleotide Region	Protein Region	Molecular Weight (kDa)	Fusion Tag
SARS-N	1-1269	1-422	46.05	GST
N1	1-1134	1-377	42.6	GST
N2	1-567	1-189	21.51	GST
N3	568-1134	190-377	20.9	GST
N4	381-756	127-252	14.01	GST

All clones were expressed downstream from a glutathione S transferase (GST) fusion protein, which is used to facilitate detection and purification as well as increase correct folding and solubility. Fusion proteins have been assumed to increase correct folding of their downstream partner as they reach a native conformation efficiently and rapidly, thereby promoting correct folding of the entire protein (François 1999). This fusion protein can be cleaved from the expressed protein using Tobacco Etch Virus (TEV) protease.

KRX (Promega) strain of competent *E. coli* was used for the expression of all constructs utilizing a Flexi™ (pFN2A) vector system. The Flexi™ vector utilizes two unique restriction enzymes, namely *SgfI* and *PmeI*, which ensures high fidelity, unidirectional cloning and also allows for easy transfer of the coding region between a variety of Flexi™ vectors. A lethal barnase gene is also located between the cloning sites, which upon successful ligation is replaced by the cloned gene. As the 5' and 3' ends cannot ligate together, this gene prevents the growth of cells which are not successfully ligated. KRX has a high transformation efficiency and is also nuclease and protease deficient, making it a good strain for cloning and protein expression (Hartnett, Gracyalny *et al.* 2006). The absence of these proteases and nucleases is crucial for efficient production of heterologous proteins as cytoplasmic degradation is employed by *E. coli* to conserve cellular resources (François 1999). KRX utilizes a T7 RNA polymerase gene, which is controlled by a rhamnose inducer, to provide tight control over protein expression *via* a T7 promoter, provided by the Flexi™ vector. The rhamnose is not metabolized during growth due to deletion of the isomerase, kinase and aldolase genes (Hartnett, Gracyalny *et al.* 2006). To inhibit the growth of plasmid free cells an antibiotic resistant gene is inserted into the plasmid, in our case an AmpR gene, which is used to incorporate ampicillin resistance. These antibiotics can however be degraded, inactivated or detoxified by leakage of periplasmic enzymes into the growth media (François 1999). Plasmids can be lost at a low frequency during cell division, however this loss can be increased if the ligated gene it carries is toxic to the cell (François 1999). It is therefore essential that selective pressures are installed to prevent the growth of plasmid free cells. The KRX and Flexi™ vector system has been proven to be a good model for the large scale expression of heterologous, recombinant proteins.

In this study, expression of the SARS-CoV and HCoV-NL63 N recombinant proteins could be detected in both the soluble and insoluble fraction of the cell. This has not been reported in any other literature and is peculiar as the GST fusion tag is known to increase solubility of the expressed protein. Insolubility could arise through one of two mechanisms, that being; 1) the formation of inclusion bodies or 2) interaction with the cellular membranes of the host cell. Inclusion body formation occurs in cytoplasmically expressed proteins (Hannig and Makrides 1998) and are highly expressed, denatured, homologous proteins which aggregate together to form a high molecular weight structure (Singh and Panda 2005; Rinas, Hoffmann *et al.* 2007). Inclusion bodies are generally difficult to solubilize and refold, however, as these proteins are homologous it aids in isolation and purification. Inclusion body formation also generally protects proteins from degradation (François 1999). There are several methods to reduce the formation of inclusion bodies, but by far the simplest is reduction of the fermentation temperature. This reduction in temperature slows protein synthesis and decreases protein self-association (Georgiou and Valax 1996). Expression at low temperature and in the presence of ethanol has also been reported to decrease inclusion body formation as ethanol is a strong inducer of heat shock proteins which act as chaperones for the correct folding of expressed proteins (Steczko, Donoho *et al.* 1991).

Several attempts to decrease the formation of inclusion bodies were not shown to increase the proteins solubility, indicating that insolubility did not arise through formation of inclusion bodies. Lysozymes were subsequently used as they have the ability to digest the peptidoglycan layer of the bacterial cell wall (Peti and Page 2007). This digestion dramatically increased solubility and as this step was carried out following cell lysis it indicates that the N-proteins interact with the peptidoglycan layer or another component of the bacterial cell wall.

By interacting these previously detailed, heterologously expressed proteins with serum antibodies against the protein we will be able to identify the antigenic epitope as well as any cross-reactivity between SARS-CoV and HCoV-NL63. These clones will also later be able to answer other questions regarding the dimerization process, regions involved in RNA binding, heterotypic protein interactions, *etc.*

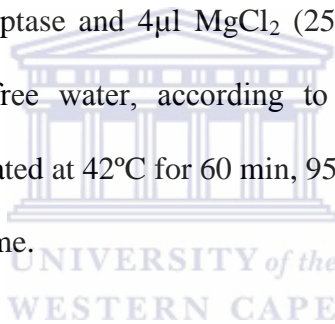


3. Materials and Methods

3.1 Amplification of Full Length and Deletion Mutant Nucleocapsid Genes

3.1.1 RT-PCR

HCoV-NL63 RNA (a kind gift from Prof. L van der Hoek, Holland) was used as template to synthesize 1st strand cDNA. RNA was extracted from a fifth-passage virus (strain Amsterdam 1) obtained from a clinical sample. Reverse transcription was carried out using a mixture of 2µl RNA template (108-109 copies/ml), 1µl of Oligo (dT)15 primer (100µM stock), 1×incubation buffer, 2µl of dNTP mix (10mM stock), 20U RNasin[®] Ribonuclease inhibitor, 15U of AMV Reverse Transcriptase and 4µl MgCl₂ (25 mM) and was made up to a total volume of 20µl in nuclease-free water, according to the manufacturer's specifications (Promega). The reaction was heated at 42°C for 60 min, 95°C for 5 min and then cooled to 0°C for 5 min to deactivate the enzyme.



3.1.2 Design of PCR Primers

For ligation into the Flexi[™] protein expression vector it is necessary to artificially insert Flexi[™] restriction sites into the gene of interest. To accomplish this, 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. An additional C residue was appended upstream of the ATG in the forward sequence to prevent a frameshift in the reading frame. Primers were designed according to the NCBI accession number for SARS-CoV N, AY360146.1, and HCoV-NL63 N, DQ846901.1. Recombinant deletion mutants of N2 (NΔN), N3 (NΔC) and N4 (NΔNΔC) were also designed. For the N3 and N4 sequences an ATG was

incorporated to the 5' as there was no existing start codon. The *PmeI* restriction site in the reverse primer will incorporate a stop codon in the 3' if there is none present. An additional 4 nucleotides were added to the 5' and 3' which promote a more complete digestion of the PCR product by *SgfI* and *PmeI* restriction enzymes. Genes of interest were amplified with primers described in Table 3.1.

Table 3.2: Forward and reverse primer sequences constructed for PCR amplification of the SARS-CoV N and HCoV-NL63 N-genes as well as truncated mutants of the HCoV-NL63 N-gene of N2, N3 and N4.

Gene	Primer	Primer Sequence
SARS-CoV N	Forward	AGGAGCGATCGCCATGTCTGATAATGGACCCCAATCAAACC
	Reverse	TTGTGTTTAAACTTATGCCTGAGTTGAATCAGCAGA
HCoV-NL63 N1 (Full length N)	Forward	GGGCGCGATCGCCATGGCTAGTGTAATTGGGCCG
	Reverse	ACAGGTTTAAACTTAATGCAAACCTCGTTGACAATTC
HCoV-NL63 N2: N-terminal	Forward	GGGCGCGATCGCCATGGCTAGTGTAATTGGGCCG
	Reverse	TTGTGTTTAAACTTAAACAGCAGCAACAAGATCTGAA
HCoV-NL63 N3:C-terminal	Forward	TGTCGCGATCGCCATGACTTTGGCTTTAAAGAACTTAGGTTT TGA
	Reverse	ACAGGTTTAAACTTAAATGCAAACCTCGTTGACAATTC
HCoV-NL63 N4: Middle Region	Forward	TCTCGCGATCGCCATGTTCTCTATTGCTTTGCCTCCA
	Reverse	GTCGGTTTAAACTTAAGGACCAAAGCACTGAATAACA

SgfI cut site; *PmeI* cut site

3.1.3 Polymerase Chain Reaction (PCR) and Purification of Amplified Genes

The N-gene of HCoV-NL63 was amplified from cDNA of the virus, previously generated, where the N-gene of SARS-CoV was amplified from a pXJHA40 vector designed and provided by Prof. Tan (Singapore). The PCR reaction was set-up in a final volume of 25µl

with 0.2mM dNTP mix, 1.5mM MgCl₂, 4μM forward and reverse primers (10pmole/ul), 5μl 5x Green GoTaq Flexi™ buffer, 0.5μl GoTaq polymerase (5U/μl) and 2μl cDNA template. Of the 10 primers the lowest T_m was 64.68°C for the HCoV-NL63 N2 reverse primer and therefore an annealing temperature of 60°C was chosen to ensure high specificity of binding of base pairs to the template strand. PCR amplification was therefore run under the following conditions: 95°C for 3 min followed by 30 cycles of 95°C for 1 min, 60°C for 1min and 72°C for 2min. The final elongation was run for 15min 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, 10mM Tris-HCl (pH 7.5) and 50mM EDTA (pH 8.0)) and run on a 1% (w/v) agarose gel, containing 0.001% (v/v) ethidium bromide, by electrophoresis in TBE (89mM Tris (hydroxymethyl) aminomethane, 0.089mM boric acid, 2mM EDTA (pH 8)) buffer at 90V for approximately 1 hour. The agarose gel was viewed under short-wave length UV light and bands of interest were excised and dissolved in 1μl/mg membrane binding solution (4.5M guanidine isothiocyanate and 0.5M 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 14000rpm for 1 min. Electrostatic interactions generated in the presence of chaotropic salts allow DNA to bind to the silica membrane. The DNA was repeatedly washed with 700μl and 500μl membrane wash solution (10mM potassium acetate, 80% (v/v) ethanol and 16.7μM EDTA) by centrifugation at 14000rpm for 1 and 5 min respectively. A further 1 min centrifugation at 14000rpm was necessary to ensure complete evaporation of residual ethanol in the washing solution. DNA was eluted in 30μl nuclease free water which

neutralizes the electrostatic interactions releasing the purified DNA from the membrane. The purified DNA was transferred to a clean 1.5ml eppendorf by centrifugation at 14000rpm. The purified PCR product was then quantified using a Qubit™ 1.0 fluorometer quantification system.

3.2 Cloning and Verification of Amplification Products

3.2.1 Ligation into pGEM® T-easy Vector

The PCR amplified viral gene was ligated into the pGEM® T-easy sequencing and shuttling vector (Appendix 11) as per manufacturer's instructions, which were set up as follows; 5µl 2X T4 DNA rapid ligation buffer, 1µl (50ng) pGEM® vector, 3µl PCR product and 1µl T4 DNA Ligase. The reaction was incubated at 4°C overnight to obtain the maximum number of transformants. The pGEM® vector contains 3' thymine overhangs in the cloning site which allows for easy ligation of the amplified product as the Taq polymerase incorporates a template-independent deoxyadenosine residue to the 3' of the amplified fragment. pGEM® also allows for blue/white colony selection through the presence of a β-galactosidase gene, which, when disrupted by the successful ligation of the amplified fragment produces the formation of a white colony. On unsuccessful ligation the β-galactosidase gene remains active and in the presence of its inducer (IPTG) and substrate (Xgal) a blue colony is formed allowing for easy selection of successfully transformed cells. M13 primer recognition sites are also present up- and down-stream of the insert allowing for easy sequencing of the fragment. Several recognition sites for restriction enzymes are also present up- and down-stream from the cloning site allowing for a single-digestion to release the cloned gene.

3.2.2 Preparation of culture media

Luria Bertani: LB media is the most common liquid media used in the cultivation of bacteria such as *E. coli*. It consists of 10g pancreatic digest of casein or tryptone powder, 5g yeast extract powder and 5g NaCl per 1L. A pH of 7.2 was established with the titration of 1M NaOH. Media was sterilized by autoclave, after which, 1µg/ml ampicillin was added, unless otherwise stated.

LB agar: LB agar was prepared as per LB media with the addition of 15g agar bacteriologica. Agar was thoroughly dissolved by convection heat prior to autoclaving. For blue/white colony selection, seen with pGEM[®], 30µl 3% (v/v) Xgal and 100µl 0.1M IPTG was spread over the surface of the solidified agar.

Terrific broth: Rich media such as terrific broth can be used to support cultures requiring a higher cell density. For protein expression cultures no significant difference was noted in final protein yield between LB and terrific broth, however, a higher cell density was beneficial for cultures used in plasmid extraction. Terrific broth was prepared with 12g tryptone powder, 24g yeast extract powder, 4ml glycerol, 17mM KH₂PO₄ and 72mM K₂HPO₄ which was thoroughly dissolved in 1L distilled water.

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

High transformation efficiency ($>1 \times 10^8$ cfu/µg DNA) JM109 (Promega) competent *E. coli* cells were used for transformation. Transformations were followed as per manufacturer's protocols with minor modifications. Briefly, 50µl competent cells were thawed on ice for approximately 5 minutes and 5µl ligation reaction was added to the cells, which were returned

to ice for approximately 20 min. The cells were then heat-shocked in a water bath at exactly 42°C for 45 sec and returned to ice for 2 min. 950µl room temperature LB media, without ampicillin, was then added and incubated at 37°C for 1.5 hours. Cells were then pelleted at 2000rpm for 10 min and resuspended in 200µl LB media. This cell suspension was spread-plated on LB agar plates containing 1µg/ml ampicillin, 0.1M IPTG and 3% (v/v) Xgal and incubated at 37°C for 16-20 hours.

Recombinant plasmids were screened by colony-PCR where white colonies were selected and inoculated into terrific media and a corresponding PCR mix. The cultures were incubated at 37°C for 14 hours. PCR reaction was run as previously described, where initial denaturation was increased to 5 min to ensure complete lysis of the cells and denaturation of template strand. Results were used to identify colonies with the correct gene construct and glycerol stocks were made of corresponding cultures. Glycerol stocks were prepared with 750µl media and 250µl glycerol, which was thoroughly resuspended and stored at -80°C. The remaining culture was used for plasmid DNA extraction using the SV Mini-Prep kit (Promega).

3.2.4 Plasmid Extraction

The mini-prep (SV Mini-Prep, Promega) procedure is used to isolate plasmid DNA from bacterial cells using an alkaline lyses method. Manufacturer's protocols advised using 2ml 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: $OD_{600}(\text{mg/ml}) \times \text{Vol}(\text{ml}) = \text{BM}(\text{mg})$. A biomass of approximately 4mg was used for all mini-preps. After the cells were harvested by centrifugation at 14000rpm the pellet was resuspended in 250 μl cell suspension buffer (50mM Tris-HCl, 10mM EDTA) containing 100 $\mu\text{g/ml}$ 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 μl neutralization buffer (80mM potassium acetate, 8.3mM Tris-HCl, 40 μM EDTA), the plasmid can easily realign, where chromosomal DNA remains in a single stranded form and interacts with cellular debris. On centrifugation at 14000rpm for 1min 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 with the same principle as in 3.1.3. Plasmid DNA is then washed with 700 μl wash solution by centrifugation at 14000rpm for 1min and eluted in 50 μl nuclease free water and transferred, by centrifugation at 14000rpm, to a clean 1.5ml eppendorf. Plasmid DNA is stored at -20°C .

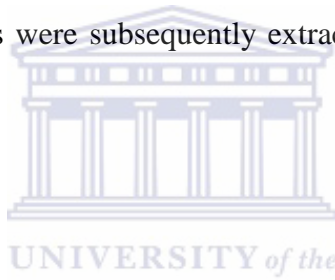
3.2.5 Restriction Endonuclease Digestion

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 (50mM Tris-HCl, 10mM MgCl_2 , 100mM 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.

3.2.6 Nucleotide Sequencing and Sequence Analysis

Extracted plasmids were sequenced with M13 forward and reverse primers to ensure no base-pair mutations occurred. Sequencing reactions were conducted by Inqaba Biotechnical Industries (Ltd) Pty, South Africa. 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.



3.3 Ligation into Flexi™ Vector and Transformation of KRX Competent *E. coli*

3.3.1 Restriction with Flexi™ Enzymes (*SgfI* and *PmeI*)

Flexi™ vectors (Appendix 12) contain a lethal, or barnase, 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, 4µl 5x Flexi™ digest buffer (10mM TrisHCl (pH 7.4), 150mM KCl, 50mM NaCl, 0.1mM EDTA, 1mM DTT, 0.25mg/ml BSA, 0.05% (w/v) Thesit® and 50% (v/v) glycerol), 5µl (±0.1µg/ul) of the various recombinant plasmids,

1µl Flexi™ enzyme blend (*SgfI* and *PmeI*) and 10µl nuclease free water was incubated at 37°C for 3 hours. The enzyme was then deactivated at 65°C for 20min. It was later noticed that yield was greatly increased when the nuclease free water was replaced with recombinant pGEM® vector and therefore this volume was increased to 15µl 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. The barnase gene was removed from the Flexi™ vector under the same conditions as previously described.

3.3.2 Ligation into Flexi™ Vector

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

3.3.3 Preparation of Chemically Competent *E. coli*, KRX Strain, and Transformation with Recombinant Plasmid

KRX strain of competent *E. coli* was purchased from the manufacturers (Promega) and used to prepare subsequent cultures of competent cells. To prepare competent cells, cells were heat-shocked at 42°C for 15 sec and allowed to recover in 450µl LB media at 37°C for 1 hour with shaking at 180rpm. Cells were subsequently plated on LB agar plates and grown for 16-20

hours at 37°C. Single colonies were inoculated into 100ml, ampicillin free LB media and incubated at 37°C, with agitation, for approximately 3 hours until the OD₆₀₀ reached 0.35. Cultures were cooled to 0°C prior to harvesting by centrifugation at 4100rpm for 10 min. All subsequent work was done at 4°C and on ice. Cells were resuspended in 30ml ice cold solution of 80mM MgCl₂ and 20mM CaCl₂ and subsequently harvested for a second time by centrifugation. Pellets were again resuspended in 2ml ice cold 0.1M CaCl₂ solution with 70µl DMSO and stored on ice for 15 min. An additional 70µl DMSO was added and gently mixed. 50µl aliquots were prepared and instantly frozen in liquid nitrogen prior to storage at -80°C.

Transformation reactions were set up as per manufacturer's instructions. Briefly, KRX cells were removed from -80°C freezer and thawed on ice, to which 5µl ligation product was added and returned to ice for 5-10 min. The cells were then heat-shocked for 15-20 sec at exactly 42°C and returned to ice for 2 min. 450µl LB broth, without ampicillin, was added to each transformation reaction and incubated for 60 min at 37°C with shaking at 180rpm. Following which, cells were harvested at 2000rpm for 10 min and resuspended in 200µl LB broth, of which 100µl was plated on LB plates containing 1µg/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 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, using *SgfI* and *PmeI* restriction enzymes, of the vector.

3.4 Expression of Full-Length SARS-CoV and HCoV-NL63 N and Deletion Mutants in *E. coli*

3.4.1 Autoinduction

The autoinduction protocol as described by Schagat *et al.* (2008) utilizes glucose and rhamnose to tightly control the T7 RNA polymerase gene *via* a rhamnose promoter (*rhaP_{BAD}*) (Schagat, Ohana *et al.* 2008), which is inhibited by the metabolites of glucose *via* a cyclic AMP (cAMP) pathway (Holcroft and Egan 2000). Glucose may therefore be used to inhibit pre-induction protein expression, where rhamnose will induce protein expression but only once the glucose is completely metabolized from the culture media.

An early autoinduction protocol was utilized where glycerol stocks of SARS-CoV N-Flexi, N1-Flexi, N2-Flexi, N3-Flexi and N4-Flexi were inoculated into a starter culture of 10ml LB media and incubated at 37°C for 14 hours with shaking at 180rpm. Starter cultures were diluted 1:100 into LB media supplemented with 0.05% (w/v) glucose and 0.1% (w/v) rhamnose and incubated at 37°C with shaking at 180rpm. Samples of 50ml were taken at 8 and 24 hours and harvested by centrifugation at 6000rpm for 15min at 4°C to prevent protein degradation. Dry cell pellet was frozen at -80°C until further use. We were unable to detect any protein expression and therefore chose to modify the expression protocols as follows.

3.4.2 Induction at specific OD₆₀₀ reading with fermentation at 37°C

Glycerol stocks were inoculated into 10ml of LB media and incubated for 14 hours at 37°C with shaking at 180rpm. Starter cultures were diluted 1:100 into 100ml LB media and

incubated for approximately 4 hours when OD₆₀₀ readings were taken. Once the culture reached an OD₆₀₀ of 0.3-0.4 expression was induced by the addition of 0.1% (w/v) rhamnose. Cultures were incubated as previously described for 24 hours when cells were harvested by centrifugation at 6000rpm for 15min. Expression was detectable but however at a low concentration and therefore we further chose to modify the expression protocol by reducing the fermentation temperature to 25°C.

3.4.3 Induction at specific OD₆₀₀ reading with fermentation at 25°C

The previous method was maintained until after induction of expression where the fermentation temperature was reduced to 25°C. Low yields of expressed proteins may indicate toxicity to the cell by the heterologous protein and therefore decreasing the fermentation temperature will slow the initial expression of the protein allowing the cell to acclimatize and thereby increasing final yields. We, however, were not able to detect any protein expression and therefore returned all fermentation temperatures to 37°C.

3.4.4 Time course expression

Expression protocols described in 3.4.2 were maintained where starter cultures were inoculated into 400ml LB media. Samples of 100ml were then taken at 5, 11, 18 and 24 hours post-induction to ascertain at which point expression is at its highest. Protein catabolism was noted following 5 hours post-induction and therefore for all subsequent expressions cells were harvested 5 hours following induction.

3.4.5 Large scale expression

Due to the low levels of protein expression a large scale culture was necessary to obtain an ideal concentration for protein purification. The expression protocols described in 3.4.2 and 3.4.4 were maintained where starter cultures were inoculated in 500ml LB media. Following 5 hours post-induction cells were harvested as previously described.

3.5 Protein Analysis

3.5.1 Protein extraction and Preparation of Cell Lysates

Following centrifugation each gram of cell pellet was resuspended by vortexing in 2ml of lysis buffer containing 1% (v/v) TritonX, 150mM NaCl, 10mM Tris(hydroxymethyl)-amino-methane (Tris), 5mM EDTA and a protease inhibitor cocktail (Sigma). Table 3.2 describes the function of all constituents in the lysis buffer. Cells were subsequently lysed by sonication in an ice bath to prevent protein degradation. All subsequent work was done on ice or at 4°C. Following sonication, insoluble cell fraction was separated from soluble fraction by centrifugation at 14000rpm for 15 min. The soluble fraction was removed and stored at -20°C. The insoluble fraction was resuspended in 1ml lysis buffer by sonication as vortexing posed insufficient.

Table 3.3: Function 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

Protein expression was detectable in both the soluble and insoluble fraction of the cell and therefore several buffers containing chaotropic compounds, which were identified from the literature and known to increase protein solubility were used to increase solubility of the membrane bound fraction. All buffers are described in Table 3.3. None of the buffers used showed any ability to increase protein solubility and therefore lysozymes were introduced to the lysis buffer previously described in a final concentration of 0.1mg/ml. The insoluble fraction was incubated for 3 hours at room temperature on a rocker in the buffer containing lysozymes (Thermo Scientific). Subsequently samples were centrifuged at 14000rpm for 15 min and the supernatant was removed and stored at -20°C. The pellet was further resuspended as previously described to ascertain whether any expressed proteins remained in the insoluble fraction.

Table 3.4: Buffers used to increase protein solubility from the membrane-bound fraction.

	Description
Buffer 1	2% (v/v) TritonX
Buffer 2	BPer (20mMTris buffer (pH 7.5) and proprietary additives)
Buffer 3	20mM HEPES, 420mM NaCl, 2mM EDTA, 25mM MgCl ₂ and 0.1M DTT
Buffer 4	1% (v/v) TritonX, 20mM HEPES, 420mM NaCl, 0.2mM EDTA, 1.5mM MgCl ₂ , 1mM DTT, 5% (w/v) SDS, 33% (v/v) glycerol
Buffer 5	1% (v/v) TritonX, 4M NaCl, 160mM Tris, 40mM Imidazole, 6M GuHCl
Buffer 6	1% (v/v) TritonX, 150mM NaCl ₂ , 5mM EDTA, 10mM Tris, 1.5% (w/v) N-Lauroylsarcosine,

3.5.2 SDS-Polyacrylamide Gel Electrophoresis and Coomassie Staining

A 3% stacking and a 12.5% separation SDS-PAGE gel was used to separate proteins in a bio-rad system. Electrophoresis was run at 15mA/gel for 80min in a Tris-glycine SDS running buffer (25mMTris, 192mM Glycine, 0.1% (w/v) SDS (pH 8.3)) The SDS-PAGE gel was then either stained in Coomassie Brilliant Blue stain (40% (v/v) methanol, 10% (v/v) acetic acid and 0.025% (w/v) Coomassie Brilliant Blue R250) overnight, to identify all proteins expressed in the cell, or transferred to a nitrocellulose or PVDF membrane for Western Blot analysis, to identify the GST tag with the appended N-protein. Coomassie stain was removed using a coomassie destain solution (50% (v/v) methanol and 10% (v/v) acetic acid)and preserved in cellophane sheeting after treatment with gel drying solution (50% (v/v) methanol, 10% (v/v) acetic acid and 10% (v/v) glycerol).

3.5.3 Western Blotting of Total Proteins

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 (27mM tris(hydroxymethyl)-amino-methane, 191mM glycine and 20% (v/v) methanol) in a submersion system at 100V for 90min, after which, the membrane was blocked with a 5% (w/v) milk and 0.05% (v/v) tween 20 in PBS solution for 30min 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 GST, in a dilution of 1:5000. 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 1:5000 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 colourimetric analysis.

3.6 Purification of Recombinant Protein

The MagneGST™ purification system utilizes the ability of reduced Glutathione to bind GST fusion proteins. Glutathione is therefore covalently linked to a paramagnetic particle to allow for isolation of the Glutathione-GST complex in a magnetic separation stand.

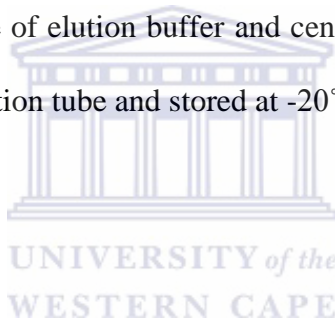
The MagneGST™ particles were equilibrated by washing 3 times in the MagneGST™ binding/wash buffer (4.2mM Na₂HPO₄, 2mM K₂HPO₄, 140mM NaCl, 10mM KCl). Particles

were separated from the binding/wash buffer by placing tubes in the magnetic stand. Glutathione particles were then resuspended in 250µl binding/wash buffer and 25µl lysozyme treated pellet and 25µl soluble fraction was added and incubated at room temperature on an orbital shaker for 30min. Glutathione particles (now bound to GST fusion proteins) were isolated in the magnetic stand and washed 3 times in the binding/wash buffer. The particles were then resuspended in 100µl elution buffer (50mM Glutathione, 50mM Tris-HCl) and incubated at room temperature for 30min on an orbital shaker. Particles were again separated in the magnetic stand and the supernatant (containing the purified GST fusion proteins) was removed and stored at -20°C. Purified proteins were subsequently quantified using the Quant-IT™ protein assay kit with the Qubit™ Fluorometer.

3.7 Cleavage of Fusion Tag and Purification of Native Protein

For certain post-expression studies the native protein is required and therefore the full length GST fusion protein of SARS-CoV and HCoV-NL63 N was cleaved using ProTEV protease. ProTEV is a highly specific proteolytic enzyme that cleaves between glutamine and glycine or serine of the amino acid sequence ENLYFQG (Dougherty and Parks 1991; Carrington, Haldeman *et al.* 1993). The Flexi™ vector is engineered to contain this cut site upstream of the cloning site and downstream of the GST sequence. Cleavage of the GST affinity tag was done according to manufacturer's instruction; briefly reactions were prepared in 100µl with 5µl 20xProTEV buffer, 1µl 100mM DTT and 1µl 10UProTEV protease. Nuclease free water was replaced with purified fusion protein which was increased to a total volume of 93µl. The reaction was incubated for 3 hours at 30°C.

Following incubation the sample was separated by SDS-PAGE as previously described and viewed using Coomassie Brilliant Blue R250. The GST affinity fusion tag and native N-proteins were excised from the SDS gel and purified using an EzWay™ PAG protein elution kit (Koma Biotech). Gel slices were homogenized in the 2ml masher tubes provided and centrifuged at 14 000rpm for 2min. The homogenized gel slice was subsequently resuspended in equal volumes of elution buffer following which, centrifugation was repeated. The collection tube, containing the cleaved proteins, was then vortexed to resuspend the sample, which was transferred to a spin filter and centrifuged at 14 000rpm for 5min for the first recovery of cleaved proteins. For the second recovery the sample in the spin filter was resuspended in an equal volume of elution buffer and centrifuged again. The purified protein was transferred to a clean collection tube and stored at -20°C.



4. Results

4.1 PCR

The N-gene of SARS-CoV was PCR amplified from a pXJHA40 vector and HCoV-NL63 was amplified from cDNA generated by RT-PCR of the viral RNA. The gene sizes, as determined by the accession number from NCBI, were as follows: SARS-CoV N: 1269bp; N1: 1134bp; N2: 567bp; N3: 567bp and N4: 378bp. Following PCR, the products were run on a 1% (w/v) agarose gel and viewed under UV light (Figure 3.1). The N-genes were identified as the correct size and gel excised for purification with a PCR gel Clean-Up kit (Promega). The viral genes were subsequently ligated into a pGEM[®] vector and used to transform JM109 strain competent *E. coli*.

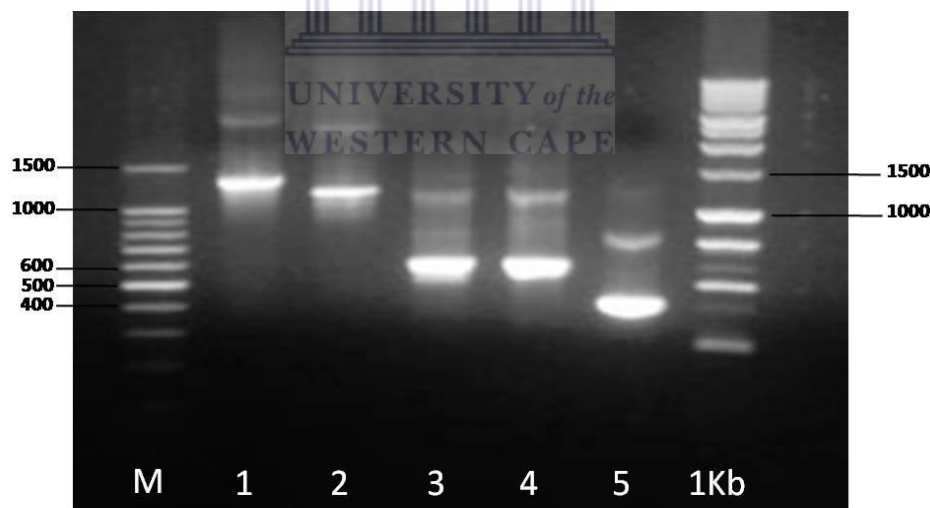


Figure 3.1: PCR amplification of viral genes using primers described in table 3.1. Lane M: 100bp marker. Lane 1: SARS-CoV N-gene (± 1200 bp) amplified from pXJHA40 vector; Lane 2: HCoV-NL63 full length N-gene (± 1100 bp) (N1); Lane 3: 5' half of the HCoV-NL63 N-gene (± 550 bp) coding for the N-terminal fragment of the N-protein (N2); Lane 4: 3' half of the HCoV-NL63 N-gene (± 550 bp) coding for the C-terminal fragment of the N-protein (N3); Lane 5: Middle $\frac{1}{3}$ of the HCoV-NL63 N-gene (± 400 bp) coding for the middle fragment of the N-protein. Lane 1Kb: 1Kb marker. All genes represented in lanes 2-5 were PCR amplified from cDNA previously generated from reverse transcription of viral RNA.

4.2 Colony PCR of JM109 Competent *E. coli*

As identification of the amplified gene by size chromatography is not completely accurate, the viral genes were subsequently ligated into the pGEM[®] T-easy vector for sequencing with M13 primers, which will also allow for identification of any base pair mutations. Following ligation into pGEM[®] T-easy vector, JM109 competent *E. coli* was transformed with the plasmid. The transformed cells were inoculated onto LB-agar plates and incubated O/N at 37°C after which, single colonies were selected and inoculated into LB-media and corresponding PCR mix.

The colony PCR allowed for identification of successfully ligated vectors and transformed cells (Figure 3.2). PCR reactions were viewed on a 1% (w/v) agarose gel under UV light and results were used to select cultures with correct gene constructs. Successfully transformed cells were stored in glycerol stocks at -80°C. Plasmids were extracted from JM109 cells and used for sequencing which was conducted by Inqaba Biotech. Forward and reverse sequences were aligned using multiple sequence alignment with ClustalX (Version 2.0.12) and viewed in GeneDoc (Version 2.7.000). For results please refer to appendix 1-5.

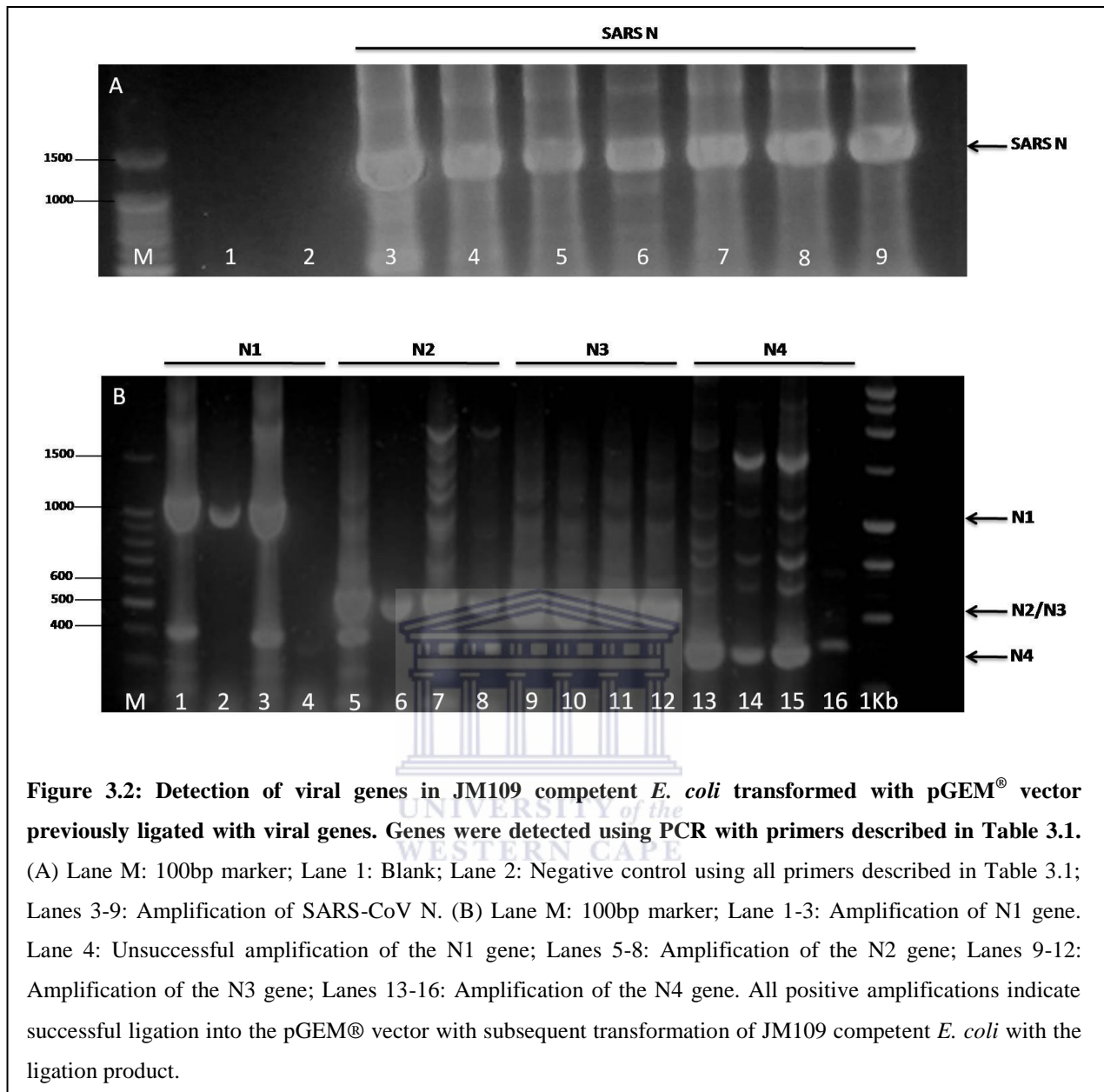
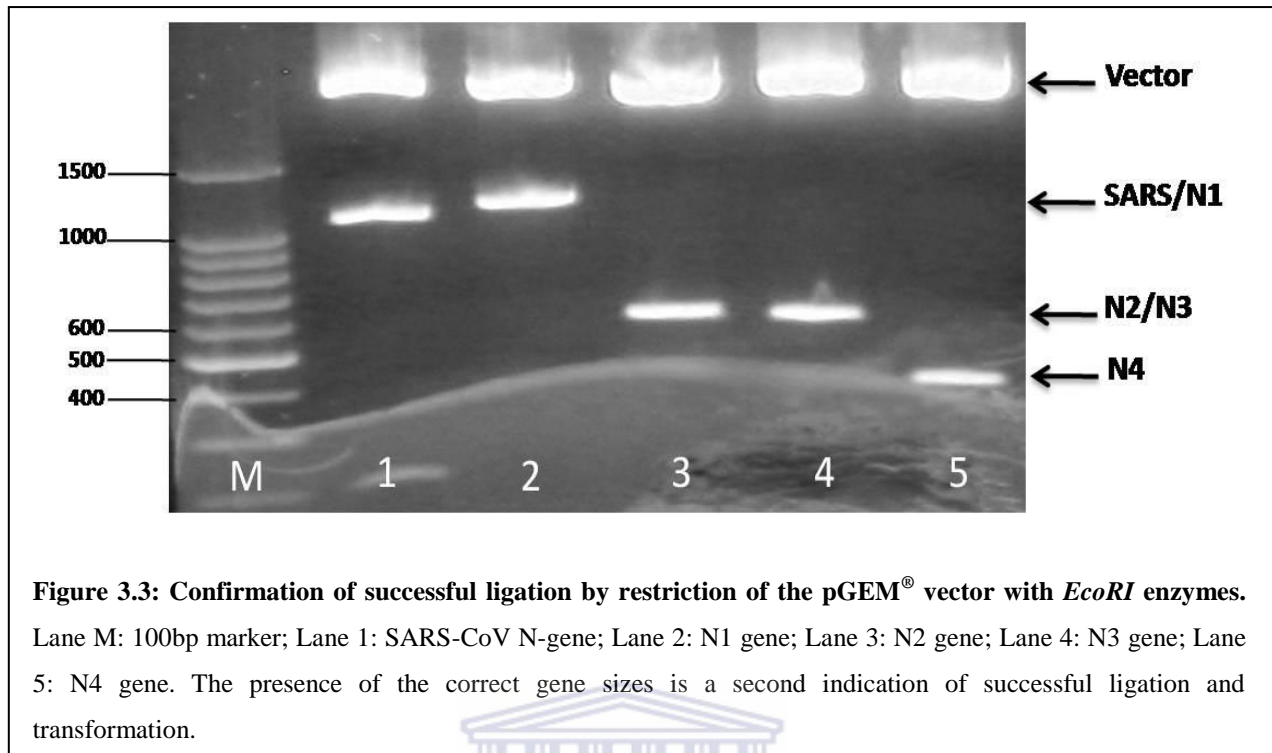


Figure 3.2: Detection of viral genes in JM109 competent *E. coli* transformed with pGEM[®] vector previously ligated with viral genes. Genes were detected using PCR with primers described in Table 3.1. (A) Lane M: 100bp marker; Lane 1: Blank; Lane 2: Negative control using all primers described in Table 3.1; Lanes 3-9: Amplification of SARS-CoV N. (B) Lane M: 100bp marker; Lane 1-3: Amplification of N1 gene. Lane 4: Unsuccessful amplification of the N1 gene; Lanes 5-8: Amplification of the N2 gene; Lanes 9-12: Amplification of the N3 gene; Lanes 13-16: Amplification of the N4 gene. All positive amplifications indicate successful ligation into the pGEM[®] vector with subsequent transformation of JM109 competent *E. coli* with the ligation product.

4.3 *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 enzymes (Figure 3.3). *EcoRI* recognition sites are located up- and down-stream of the cloning site allowing for easy removal of the insert.



4.4 *SgfI* and *PmeI* Restriction

Once the presence of viral genes was confirmed by sequencing, the inserts were removed by restriction with the Flexi[™] restriction enzymes, *SgfI* and *PmeI*. To ensure vectors were completely cut by both enzymes, the reaction was viewed on a 1% (w/v) agarose gel under UV light (Figure 3.4) and viral genes were identified by their correct sizes. Bands were then excised from the gel, purified and ligated into the Flexi[™] protein expression vector. KRX competent cells were then transformed using the newly synthesized viral gene-vector construct.

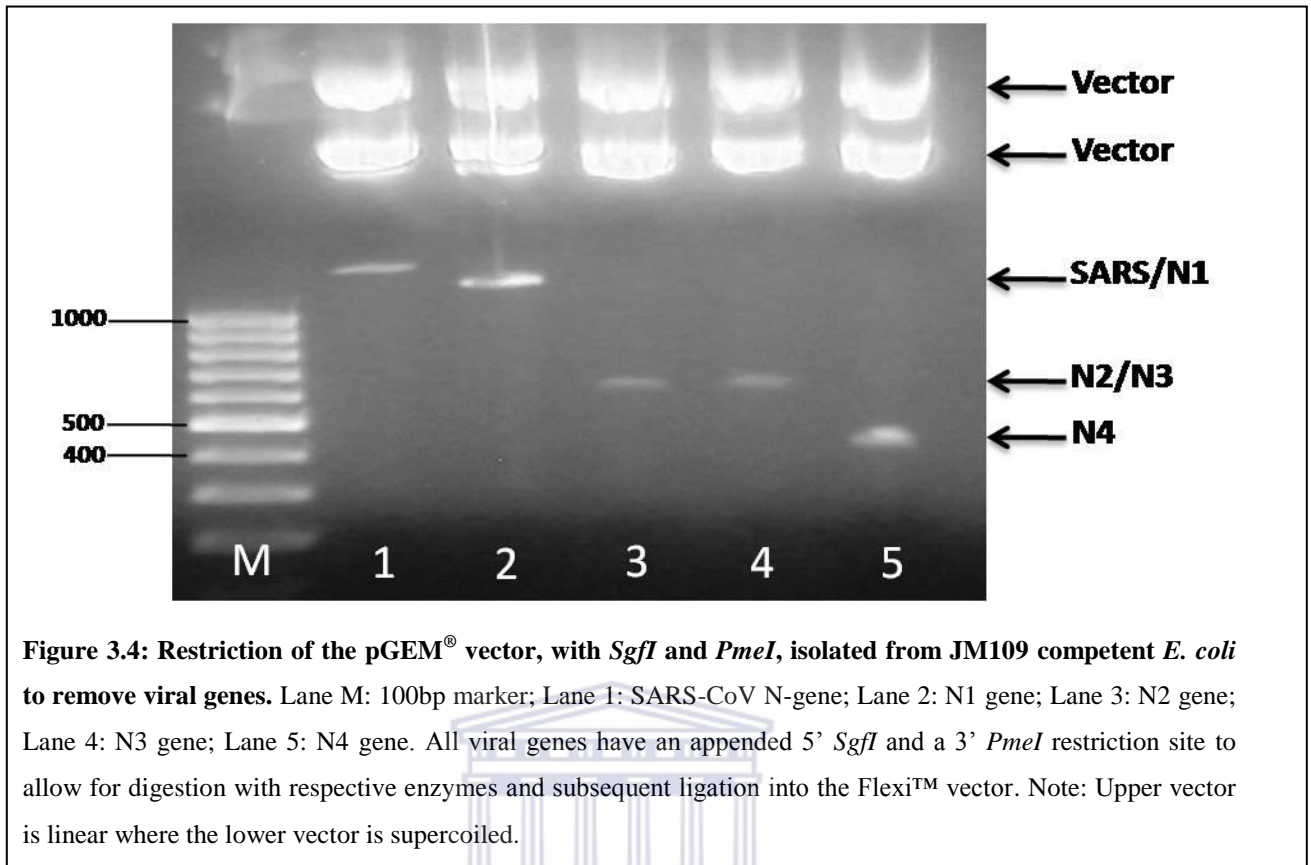


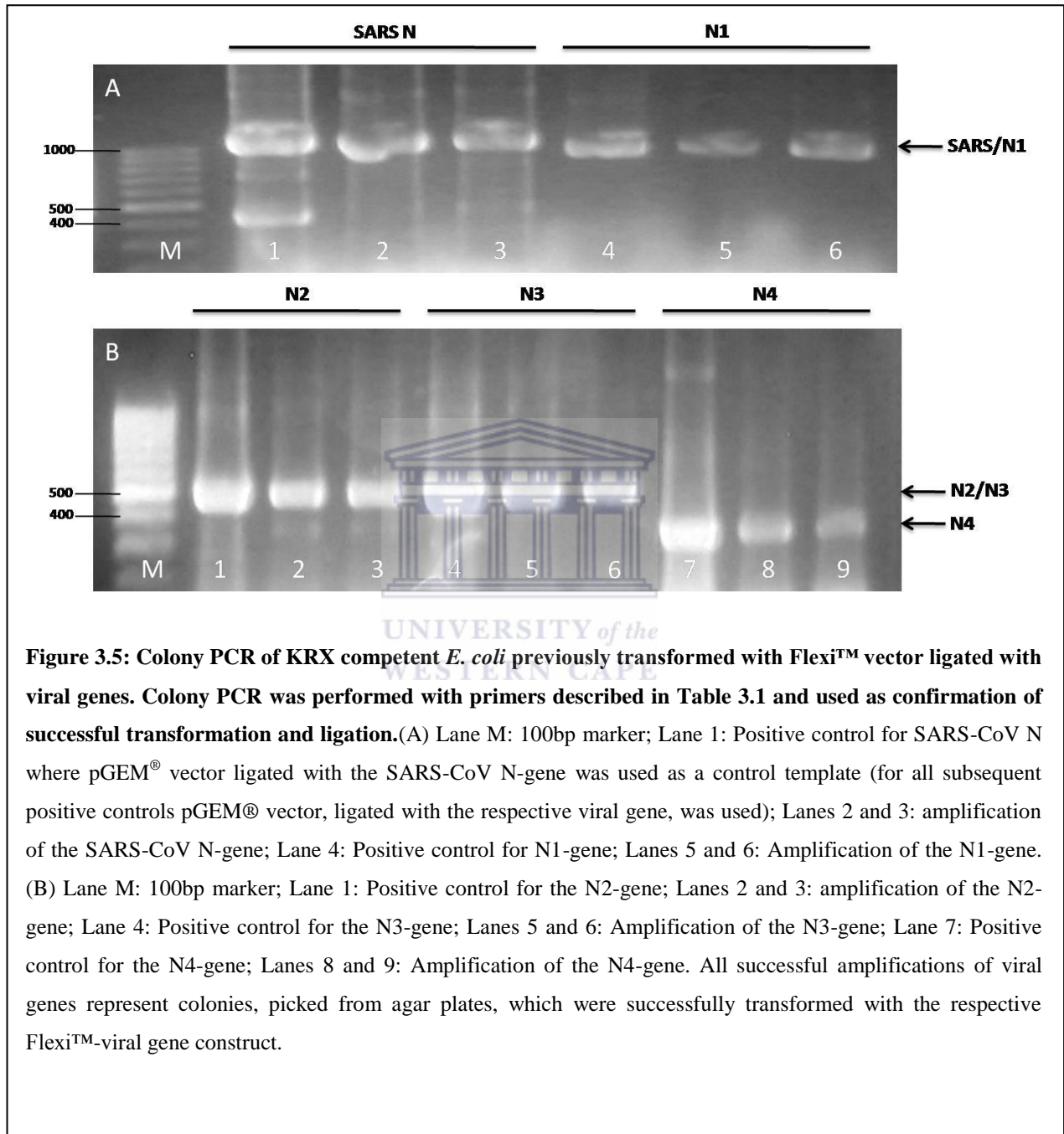
Figure 3.4: Restriction of the pGEM[®] vector, with *SgfI* and *PmeI*, isolated from JM109 competent *E. coli* to remove viral genes. Lane M: 100bp marker; Lane 1: SARS-CoV N-gene; Lane 2: N1 gene; Lane 3: N2 gene; Lane 4: N3 gene; Lane 5: N4 gene. All viral genes have an appended 5' *SgfI* and a 3' *PmeI* restriction site to allow for digestion with respective enzymes and subsequent ligation into the Flexi[™] vector. Note: Upper vector is linear where the lower vector is supercoiled.

UNIVERSITY of the
WESTERN CAPE

4.5 Colony PCR of KRX Competent *E. coli*

To confirm ligation of the Flexi[™] vector with the viral genes and the transformation of the KRX cells, single colonies were inoculated into a culture and corresponding PCR reaction and viral genes were amplified as previously described. Results were again used to select successfully transformed colonies (Figure 3.5). As colony PCR can result in false positives and negatives a second sequencing step is critical where the T7 terminator (reverse) primer was used. The T7 promoter (forward) primer was not utilized as the T7 recognition site is upstream of the GST tag sequence resulting in the polymerase not reading sufficiently into the ligated open reading frame. For these constructs reverse sequences were seen as sufficient to

confirm ligation and reference can be made to pGEM[®] sequences to confirm entire gene sequence. Please refer to appendix 6-10 for sequencing results using the T7 terminator primer.



4.6 Bacterial Expression and Purification of Recombinant Viral Proteins

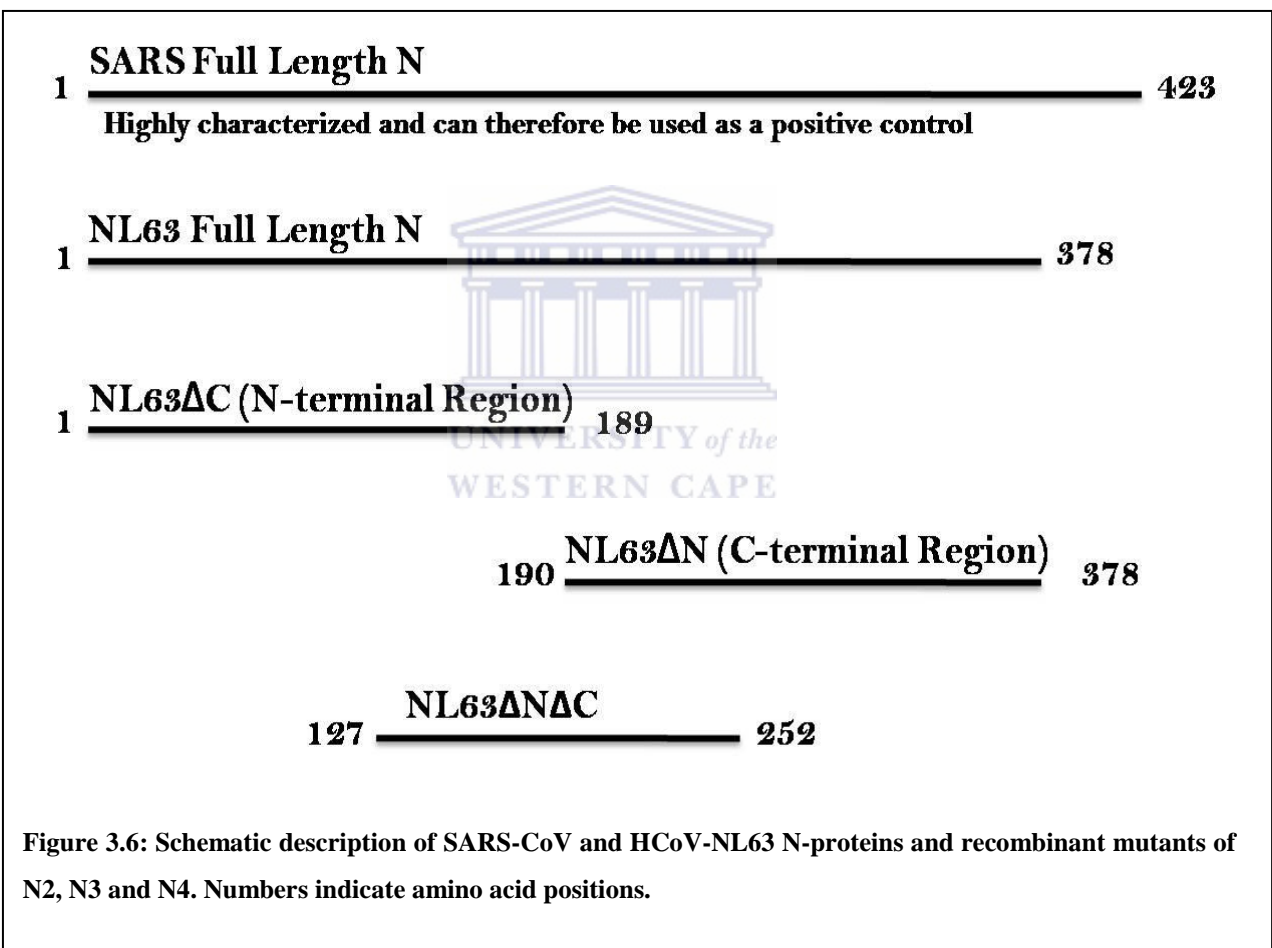
Figure 3.6 depicts the different regions of the N-protein which were expressed. The expression protocol which was adopted included subculturing of starter cultures into an expression culture in a dilution of 1:100. Cultures were grown for approximately 4 hours until an OD_{600} of 0.3-0.4 was reached when expression was induced with the addition of 0.1% (w/v) rhamnose. OD_{600} of cultures are shown in table 3.4. Cells were harvested 5 hours post-induction and lysed using protocols previously described.

Table 3.5: OD_{600} (mg/ml) of cultures after time specified.

Construct	Starter Culture	5 hours post-induction
Uninduced	0.63	0.71
SARS-CoV N	0.57	0.40
N1	0.62	0.44
N2	0.57	0.56
N3	0.58	0.61
N4	0.62	0.75

As shown in Table 3.4 the expression of the viral-GST fusion proteins decreased the OD_{600} readings in comparison to the uninduced control. This indicates possible toxicity of the proteins to the bacterial cell, which would explain the difficulty in modifying protocols for successful expression of the viral proteins. Expression of the heterologous proteins was detected by Western Blot in both the cytoplasmic and membrane bound fractions of the cell (Figure 3.7). The presence of expressed proteins in the insoluble fraction indicates either

formation of inclusion bodies or interaction of the protein with the bacterial membrane. The pellet was therefore treated with chaotropic agents known to disassemble inclusion bodies allowing for detection in the soluble fraction. Following treatment there was, however, no increase in the proportion of soluble proteins, which indicates possible interactions with the bacterial cell membrane.



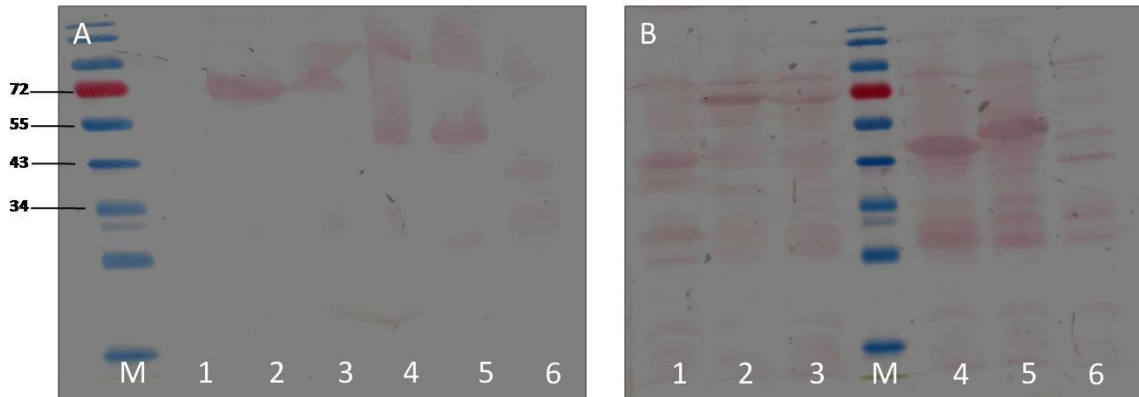
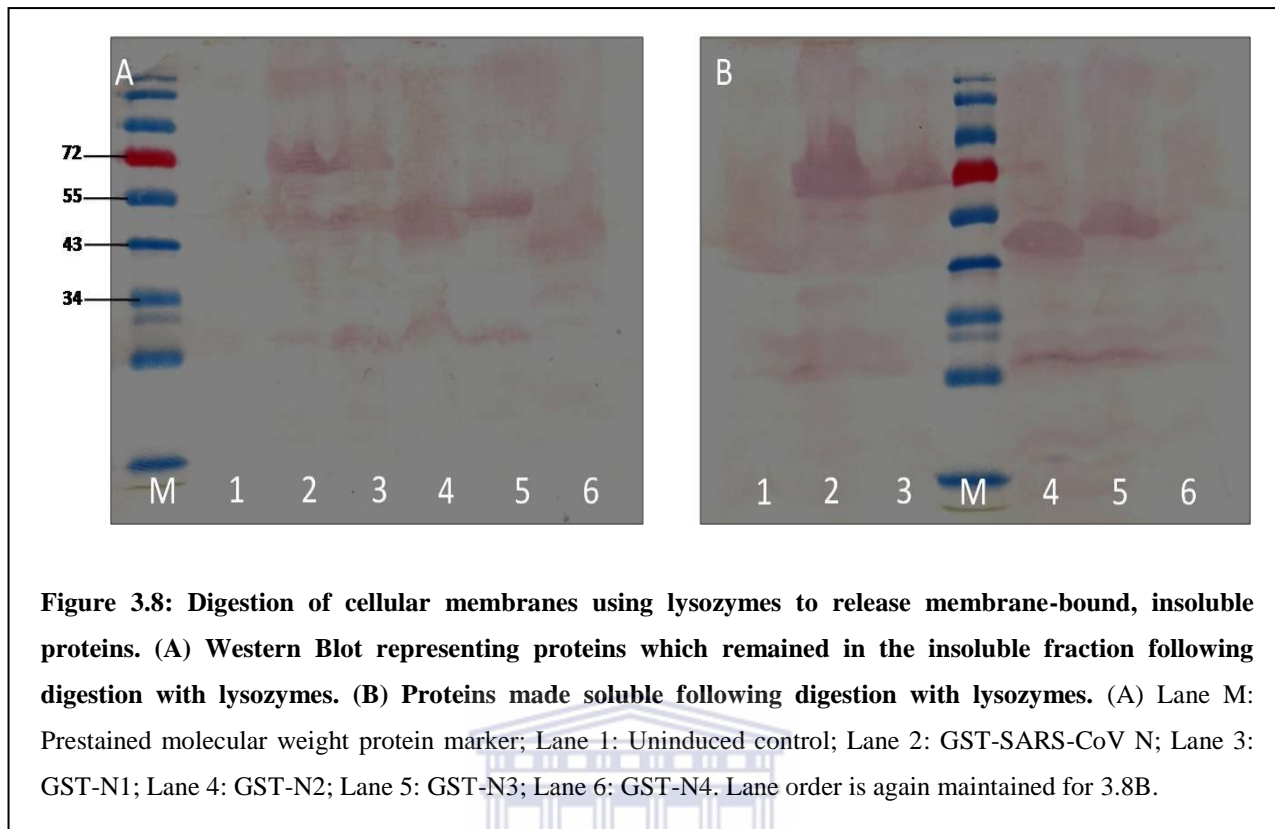


Figure 3.7: Detection of SARS-CoV and HCoV-NL63 N-proteins by Western Blot with peroxidase labeled antibodies against the GST fusion protein. (A) Proteins expressed in membrane-bound, insoluble fraction of the cell. (B) Cytoplasmic expression of soluble fusion proteins.(A) Lane M: Prestained molecular weight protein marker; Lane 1: Uninduced control to determine any background or cellular protein expression where cultures of N2 were used; Lane 2: Expression of the 73.02kDa GST-SARS-CoV N fusion protein; Lane 3: Expression of the 69.6kDa GST-N1 fusion protein; Lane 4: Expression of the 48.51kDa GST-N2 fusion protein; Lane 5: Expression of the 47.9kDa GST-N3 fusion protein; Lane 6: Expression of the 41.01kDa GST-N4 fusion protein. Lane order is maintained for Figure 3.7B.

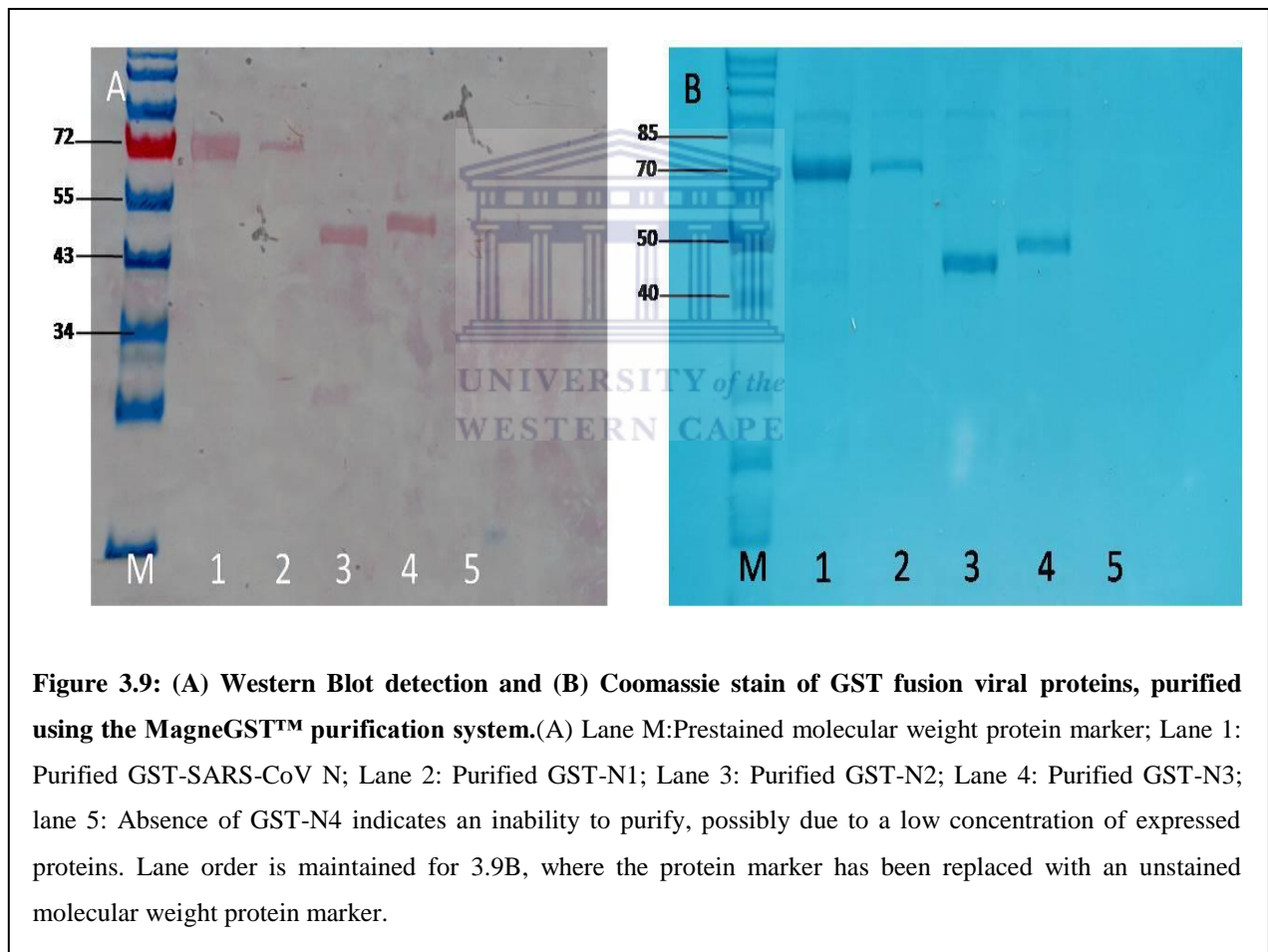
To test the hypothesis that the heterologously expressed viral-fusion proteins interact with bacterial cell wall, the cell pellet was treated with lysozymes in the same buffer used for lysis. Following 3 hours of treatment with lysozymes a large proportion of the heterologous proteins were released to the soluble portion (Figure 3.8). As lysozymes are known to digest the peptidoglycan layer of the bacterial cell wall, and as treatment was carried out following lysis by sonication, it indicates that the N-proteins interact with the bacterial cell wall during synthesis. This may provide a possible mechanism for toxicity as shown with the envelope protein of SARS-CoV, which is known to increase membrane permeability and subsequently lead to cell lysis (Liao, Lescar *et al.* 2004).



The presence of a band in Figure 3.7B and 3.8B, in the uninduced lane 1, with the same molecular weight as N2 indicates that the Flexi™ vector system is susceptible to leaky background expression *via* the T7 promoter, which is in strong contradiction that the T7 RNA polymerase is tightly controlled by rhamnose induction *via* the T7 promoter. Following isolation of the heterologously expressed, viral-GST fusion proteins purification of the cytoplasmically expressed and lysozyme treated membrane fraction were done with the use of the MagneGST™ purification system (Figure 3.9). The inability to purify N4 may be due to the low yield of expression and the low proportion released by lysozyme digestion. Final concentration of purified proteins is represented in Table 3.5.

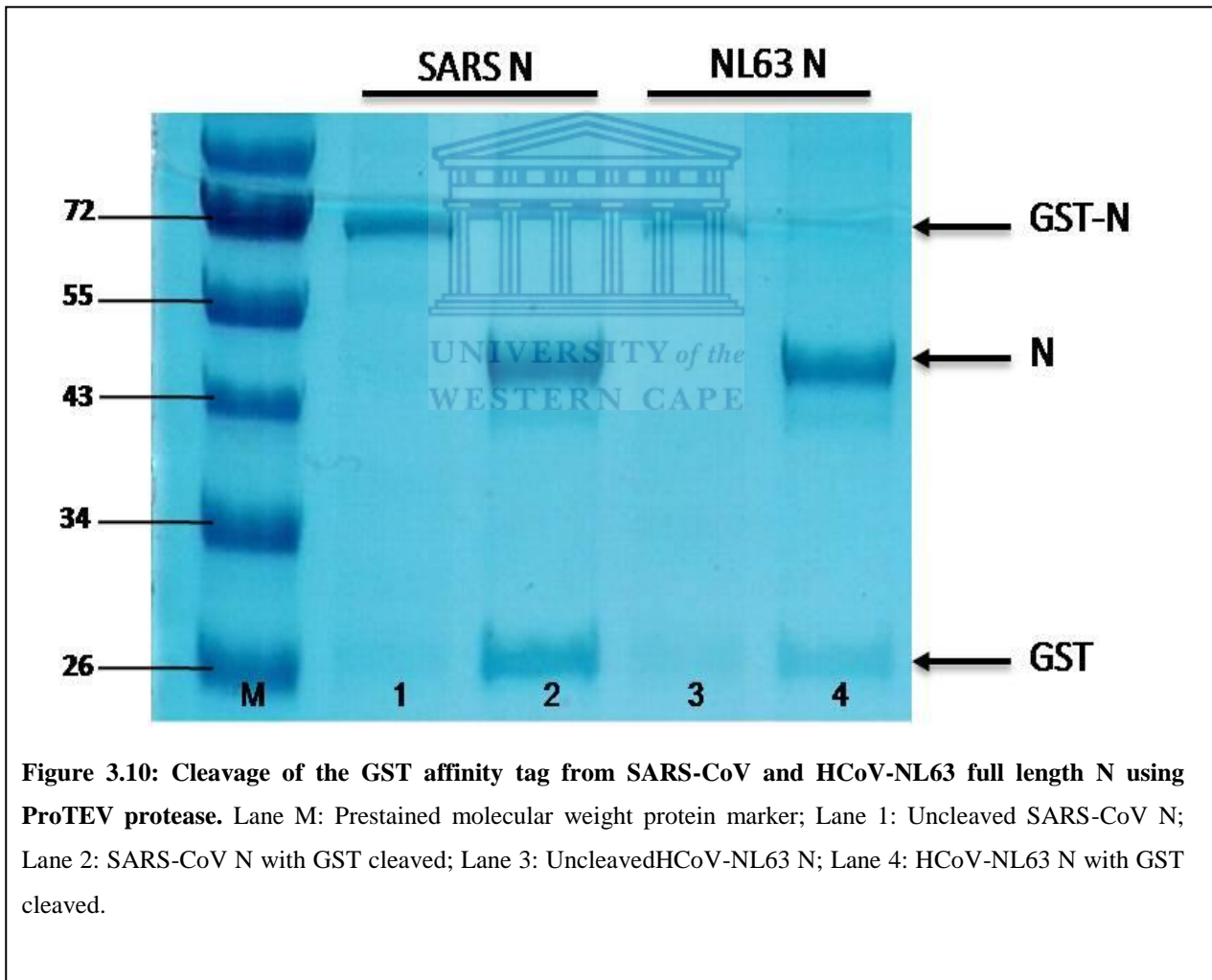
Table 3.6: Concentration of purified proteins in mg/ml.

Construct	Concentration (mg/ml)
SARS-CoV N-GST	0.44
N1-GST	0.11
N2-GST	0.24
N3-GST	0.17
N4-GST	0.00



4.7 Cleavage of Fusion Tag

The GST affinity tag of the heterologously expressed SARS-CoV and HCoV-NL63 full length N-proteins was successfully cleaved using ProTEV protease. The cleavage yielded a ~27kDa GST protein and the ~46kDa SARS-CoV N-protein, with the HCoV-NL63 N-protein presenting at approximately the same size. The size of the HCoV-NL63 N-protein is slightly bigger than the predicted size of 42.26kDa (Figure 3.10).



5. Discussion and Conclusion

The Flexi™ vector system has proven to be a suitable method for the heterologous expression of the SARS-CoV and HCoV-NL63 nucleocapsid proteins, however, not without several drawbacks including the inability to purify the expressed N4 protein, possibly through low levels of synthesis. The inability to detect purified N4 through quantification, Western Blot and coomassie may be due to the disordered nature of this region within the protein, and therefore a lack of a tertiary structure, leading to possible catalytic breakdown through the exposure of all active sites. This can be supported in that the full length N-protein, which also contains the full disordered region, yielded the lowest concentration of expressed proteins in comparison to the N- and C-terminals, which only contained their respective halves. This region will therefore have to be expressed in a cell-free system to alleviate the possibility of cellular degradation.

The instability of the coronavirus N-proteins has also been reported for SARS-CoV (Wang, Wu *et al.* 2004), 229E (Tang, Wu *et al.* 2005) and IBV (Fan, Ooi *et al.* 2005). Intracellular degradation has been reported by Fan *et al.* (2005), where bacterial proteases were found to be responsible. Fan *et al.* (2005) however utilized a BL21 protein expression system, which, when compared to KRX, is much less protease deficient and therefore a lower degree of protein degradation should be seen in the KRX system. Autolytic degradation has also been reported to occur at 4°C where the purified N-protein of IBV was found to degrade rapidly (Fan, Ooi *et al.* 2005). With our findings it may indicate that this degradation is directed *via* the middle, disordered region, corroborated by the fact that degradation yields the C-terminal spanning region from amino acid 169-422 of the 229E N-protein (Tang, Wu *et al.* 2005), indicating that cleavage does occur in the middle, disordered region. With these

findings it should be advised that all protein storage should be between -80 and -20°C, as we could not detect any autolytic degradation at -20°C.

Expression of a considerable fraction of the heterologous proteins in the insoluble fraction is another cause for concern as this will result in a lower yield of purified proteins. Solubilizing this fraction then becomes crucial to obtain a high enough protein concentration to perform post-expression studies. Research into the development of solubilizing techniques has increased over the years with several approaches being adopted, however, each method is dependent on the protein expressed and trial and error is the only approach available to determine which technique is effective, making solubilization a time consuming, labour intensive and often expensive procedure. Following solubilization, the protein is often incorrectly folded due to chaotropic agents used to denature the protein or inappropriate environment provided by the cell and therefore refolding strategies, which are also dependant on the expressed protein, becomes the next time-consuming cause for concern. Various techniques have been proven to be effective in refolding denatured proteins including controlled air oxidation (Menzella, Gramajo *et al.* 2002), co-expression of molecular chaperones (Hannig and Makrides 1998) and altering fermentation conditions with the addition of 3% ethanol, which increases the expression of heat shock proteins known to act as chaperones (Steczko, Donoho *et al.* 1991). These approaches are however trial and error and can therefore pose very time consuming.

We have improved the proportion of soluble proteins by treating the insoluble fraction with lysozymes, and as this technique has no impact on the expressed protein, as it purely digests the peptidoglycan layer of the bacterial cell wall, there is no need for *in vitro* refolding of the protein for any bio-activity. As treatment with lysozymes is a relatively inexpensive and

rapid approach to solubilizing insoluble proteins it should be adopted as the first technique prior to treatment with chaotropic agents and other time consuming solubilizing methods. However, lysozymes are not capable of solubilizing inclusion bodies and therefore should only be used to ascertain interaction with the cell wall.

Through OD_{600} readings it was identified that both the SARS-CoV and HCoV-NL63 N-proteins cause a reduction in cell density, in comparison to an uninduced control, with SARS-CoV N resulting in the greatest reduction and N4 resulting in no reduction. These results are expected as SARS-CoV N is the highest expressed of the 5 heterologous proteins and N4 being barely detectable. The reduction in cell density indicates possible toxicity of the N-protein to the bacterial expression system, most likely through its interaction with cellular membranes as has been shown for the envelope (E) protein of SARS-CoV. The E-protein is known to slow *E. coli* (BL21) growth by increasing membrane permeability and subsequently leading to cell lysis. This lysis would allow for new progeny virus to spread to neighboring cells, thereby continuing viral lifecycle (Liao, Lescar *et al.* 2004). The SARS-CoV N-protein also induces apoptosis in COS-1 cells in the absence of growth factors (Surjit, Liu *et al.* 2004). These results may therefore indicate a possible pathological mechanism coronaviruses use to induce disease. As *E. coli* does not undergo any form of apoptosis it would be impossible to confirm toxicity of the N-protein but these results do lead to more interesting research in a mammalian system.

Toxicity induced by heterologous proteins to the host cell, resulting in a decrease in growth rate, has also been shown to increase plasmid loss and therefore a later reduction in overall protein yield (François 1999). For these reasons it is important to conduct expressions

in a large scale culture to ensure a sufficient final protein yield to conduct post-expression studies.

When the uninduced N2-Flexi construct was used as a control to ascertain any ambiguity with cellular expression it was noted that there was a significant level of background expression of GST-N2, as bands of similar size were only seen in the test lane for GST-N2. When a foreign gene is expressed in a host cell it produces many metabolic stresses to the cell by utilizing the cell's resources and possibly producing metabolites that are toxic to the cell. This must especially be avoided in the logarithmic phase to allow for stabilization of the host and the introduced plasmid (Glick 1995). Background expression should be avoidable with the use of strong promoters to enable the ability to turn "on" and "off" gene expression, however, the T7 promoter, used in this study, has been shown to be a leaky promoter (Weickert, Doherty *et al.* 1996) thereby producing a metabolic burden to the host cell and subsequently reducing cell growth and final yield of expressed protein.

Given the several problems posed by the Flexi™ vector system it still remains a suitable method for the cloning and expression of the SARS-CoV and HCoV-NL63 N-proteins with the generation of truncated clones of the N- and C-terminals of HCoV-NL63 N. These heterologously expressed proteins will now provide the basis for several post-expression studies including characterizing the regions responsible for eliciting an immune response and identifying antibody cross-reactivity between coronavirus species. Various other studies, outside the scope of this thesis, including identifying the regions involved in RNA binding and dimerization of the HCoV-NL63 N-protein and antibody production in a murine model can also be conducted.

Chapter 4

Using an ELISA to Map the Antigenic Epitopes of
HCoV-NL63 Nucleocapsid Protein: A Pilot Study



1. Abstract

Shortly after the SARS-CoV outbreak, another human coronavirus (HCoV-NL63) was identified. This virus has since been shown to have a worldwide distribution and is continuously circulating in the human population. The primary function of the coronavirus nucleocapsid (N) protein is the formation of the ribonucleocapsid core by interacting with viral RNA during virus assembly. The nucleocapsid protein is highly antigenic and highly expressed during viral infection making it suitable for the development of diagnostic assays.

In this study, an antibody-capture ELISA assay was developed using recombinant clones of the full length HCoV-NL63 (N1: aa 1-378) N protein, as well as truncated clones corresponding to the N- (N2: aa 1-189) and C-terminals (N3: aa 190-378) of the HCoV-NL63 N-protein. The ELISA was then used to screen human serum samples (15 HCoV-NL63 seropositive) to identify the antigenic epitopes of N.

Serum antibodies detected HCoV-NL63 N, but cross-reactivity between SARS-CoV N and the serum HCoV-NL63-specific antibodies was also observed. We could not identify the antigenic epitope as both the N- and C-terminal reacted with equal intensity. This could be an indication of: 1) the presence of several immunogenic sites on both the N- and C-terminal domains, or 2) the division of one immunogenic site between the two fragments. Some cross-reactivity to another coronavirus N protein was observed and the assay needs to be optimized.

2. Introduction

Coronaviruses (CoVs) are enveloped viruses in the order *Nidovirales*. The *Coronaviridae* family is composed of single-stranded, positive sense RNA viruses, with genomes ranging in size from 25 to 30 kb (He, Dobie *et al.* 2004). CoVs are responsible for a variety of diseases in animals resulting in respiratory, enteric, nephritic, hepatic and neurological conditions with a varying degree of severity (McIntosh 2005; Lau, Li *et al.* 2010). In humans, coronaviruses result in approximately 10-30% of upper respiratory tract infections, which commonly manifests as common cold-like symptoms. Until the discovery of the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV), it was thought that CoV infections were not severe. SARS-CoV however, resulted in ~10% mortality of infected patients globally. In fact, mortality was as high as 40% in certain population groups (Surjit, Kumar *et al.* 2005) and between 43-55% in people older than 60 years (Marra, Jones *et al.* 2003).

Following the outbreak of SARS it was shown that CoVs are capable of undergoing recombination events resulting in a more lethal form of the virus. This ignited a new interest in the detection and characterization of previously unidentified respiratory viruses, which led to the later discovery of HKU1 (Woo, Lau *et al.* 2005) and HCoV-NL63 (van der Hoek, Pyrc *et al.* 2004). HCoV-NL63 presents clinically as the common cold with symptoms including fever, cough, sore throat and rhinitis (Chiu, Chan *et al.* 2005; Ebihara, Endo *et al.* 2005; Kaiser, Regamey *et al.* 2005; Suzuki, Okamoto *et al.* 2005; Vabret, Mourez *et al.* 2005; van der Hoek, Pyrc *et al.* 2006; Han, Chung *et al.* 2007; Kuypers, Martin *et al.* 2007; Smuts 2008; Wu, Chang *et al.* 2008; Talbot, Crowe *et al.* 2009; Fielding 2011). Prevalence of infection with HCoV-NL63 is highest in the age group 0-5 years old and is also frequently observed in patients who are immunocompromised and the elderly (van der Hoek, Pyrc *et al.* 2006).

As all human coronaviruses (HCoV) currently in circulation present with similar clinical symptoms, the diagnosis of individual HCoVs based on symptomatic presentation is virtually impossible. Currently the diagnosis of HCoV-NL63 is largely based on genetic markers identified by PCR and real time PCR (Bastien, Robinson *et al.* 2005; Koetz, Nilsson *et al.* 2006; Brittain-Long, Nord *et al.* 2008). Real time PCR is, however, expensive and requires a high level of competence from the user and is therefore not suitable for use in poorly equipped diagnostic facilities, which are common in third world countries. The development of an affordable and reliable diagnostic antibody capture enzyme linked immunosorbent assay (ELISA) - for the direct detection of viral antigens - would therefore be beneficial in such an environment.

Since N is highly antigenic, abundantly expressed during viral infection (Timani, Ye *et al.* 2004) and detectable as early as the 1st day after infection (Che, Hao *et al.* 2004), it is a suitable target for the development of such a diagnostic assay (Lehmann, Wolf *et al.* 2008). Early detection of infection using serodiagnosis, however, is unreliable as seroconversion takes a median time of 17-19 days (Che, Qiu *et al.* 2004; Woo, Lau *et al.* 2004). Also, the use of serodiagnosis in the detection of HCoVs is even less reliable as HCoVs often cause reinfection throughout life (Sastre, Dijkman *et al.* 2011), making detection of serum antibodies to current infections, without detecting prior infections, impossible. For these reasons the direct detection of viral proteins would serve more reliable.

In this pilot study, we have developed an ELISA using recombinant N-proteins of HCoV-NL63 and SARS-CoV expressed in a bacterial system. Truncated clones corresponding to the N- and C-terminals of the HCoV-NL63 N were also synthesized to identify the antigenic epitope of the protein. The inclusion of SARS-CoV N as a control would allow us to

identify any cross-reactivity between coronavirus species, which is important for determining the specificity of the assay. HCoV-NL63-positive patient sera (n=15) were screened to determine the sensitivity and specificity of our assay in detecting anti-N-IgG serum antibodies against HCoV-NL63 N. Serum from all patients screened cross-reacted with the SARS-CoV N-protein. There was also no identifiable antigenic epitope as both the N- and C-terminals of the N-protein reacted equally with patient serum.



3. Materials and Methods

3.1 Patient serum samples

A panel of 17 serum samples collected from young adults in Germany was used in this study. The samples included: 1) 15 samples previously confirmed to be HCoV-NL63 seropositive; 2) one confirmed HCoV-NL63 seronegative sample; and 3) one SARS-CoV seropositive sample (this serum sample was not previously confirmed HCoV-NL63 seronegative). All ELISA screens were done at University of Bonn, Germany.

3.2 ELISA

Expression of recombinant proteins has been described in detail in **Chapter 3**. Proteins were quantified and standardized to 3 μ g/ml in coating buffer (0.1M NaHCO₃). Each well of a Maxisorb microtitre plate (Thermo Scientific) was coated with 50 μ l of recombinant protein (Table 4.1) and incubated overnight at 4°C.

Plates were washed 5 times with 300 μ l washing buffer (PBS containing 0,1% (v/v) Tween) and blocked with 5% (w/v) fat-free milk in washing buffer for 1 hour at room temperature. Plates were subsequently washed 5 times as previously mentioned. Serial dilutions (1:100 to 1:6400) of the appropriate primary antibody (rabbit-anti-N used in the positive control; human sera when screening patient samples or rabbit-anti-GST for coating controls) in 1% (w/v) fat-free milk in PBS-Tween was incubated for 2 hours at room temperature. Plates were again washed 5 times, following which peroxidase-labeled goat-anti-rabbit or human-anti-IgG in a dilution of 1:4000 with PBS-Tween and 1% (w/v) milk, was

added to respective samples and incubated for 1 hour at room temperature. Plates were again washed 5 times and the reaction was viewed after addition of TMB, peroxidase substrate. 2M sulphuric acid was used to stop the reaction after 3-5 min. Absorbance readings were done at 450/630nm with a Glomax (Promega)

Table 4.1: Recombinant proteins included in the study.

Construct Name	Description
N/A [#]	GST tagged protein control to confirm plates were adequately coated
GST	GST only to confirm antigenicity detected was not directed against the fusion protein
SARS-N	SARS-CoV full length N (aa 1-423)
N1	HCoV-NL63 full length N (aa 1-378)
N2	N-terminal half of the HCoV-NL63 N (aa 1-189)
N3	C-terminal half of the HCoV-NL63 N (aa190-378)

[#]N/A – Not applicable

3.3 Statistical Analysis

ELISA results were taken at a wavelength of 450nm and 630nm. The readings at 630nm were subtracted from readings at 450nm to give results without background. An average of the blank or no-coat lane was taken and this was subtracted from the results without background. These readings were then used for statistical analysis.

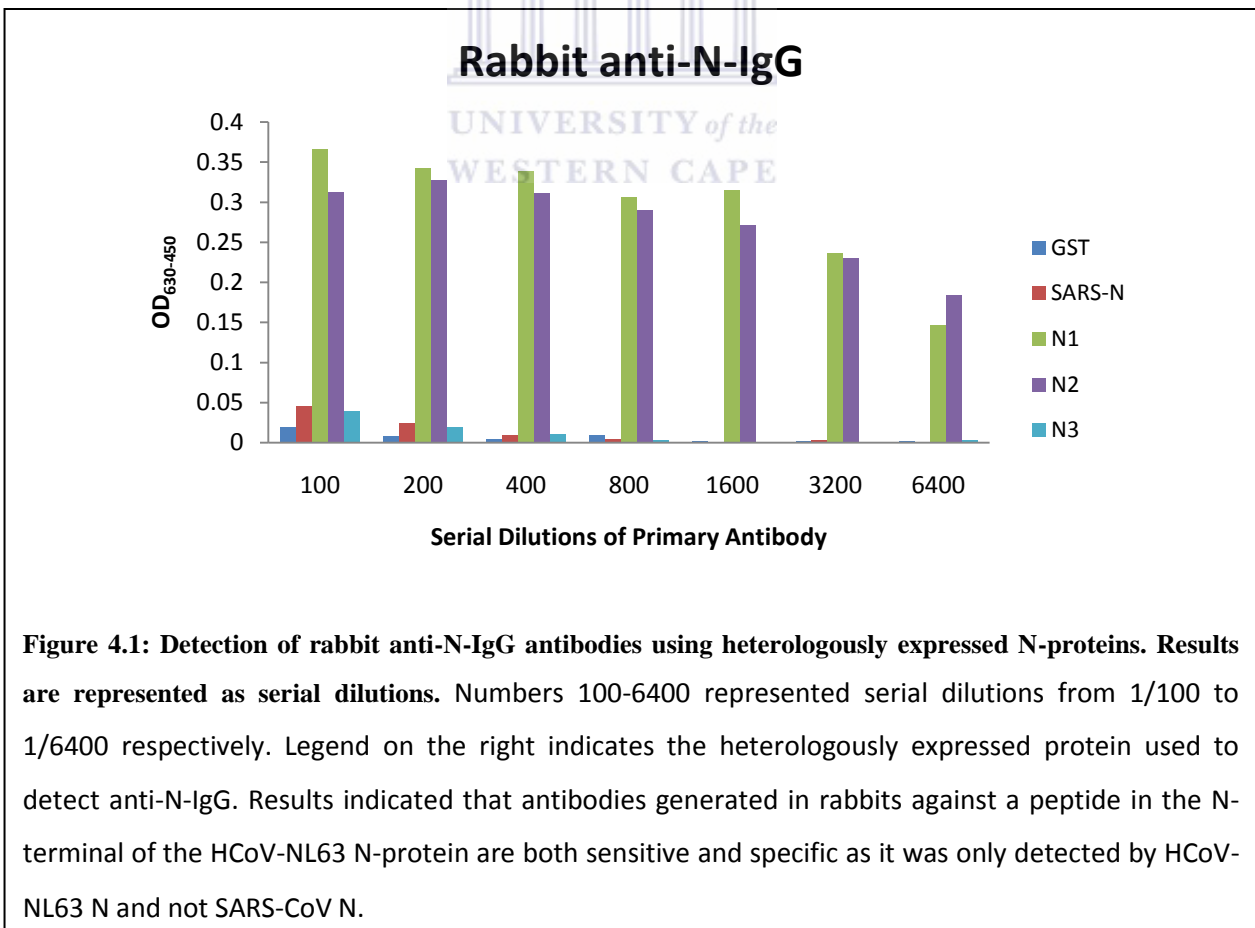
The sensitivity and specificity of the N-protein in detecting anti-N-IgG antibodies was determined with the use of receiver operating characteristic (ROC) curve analysis from

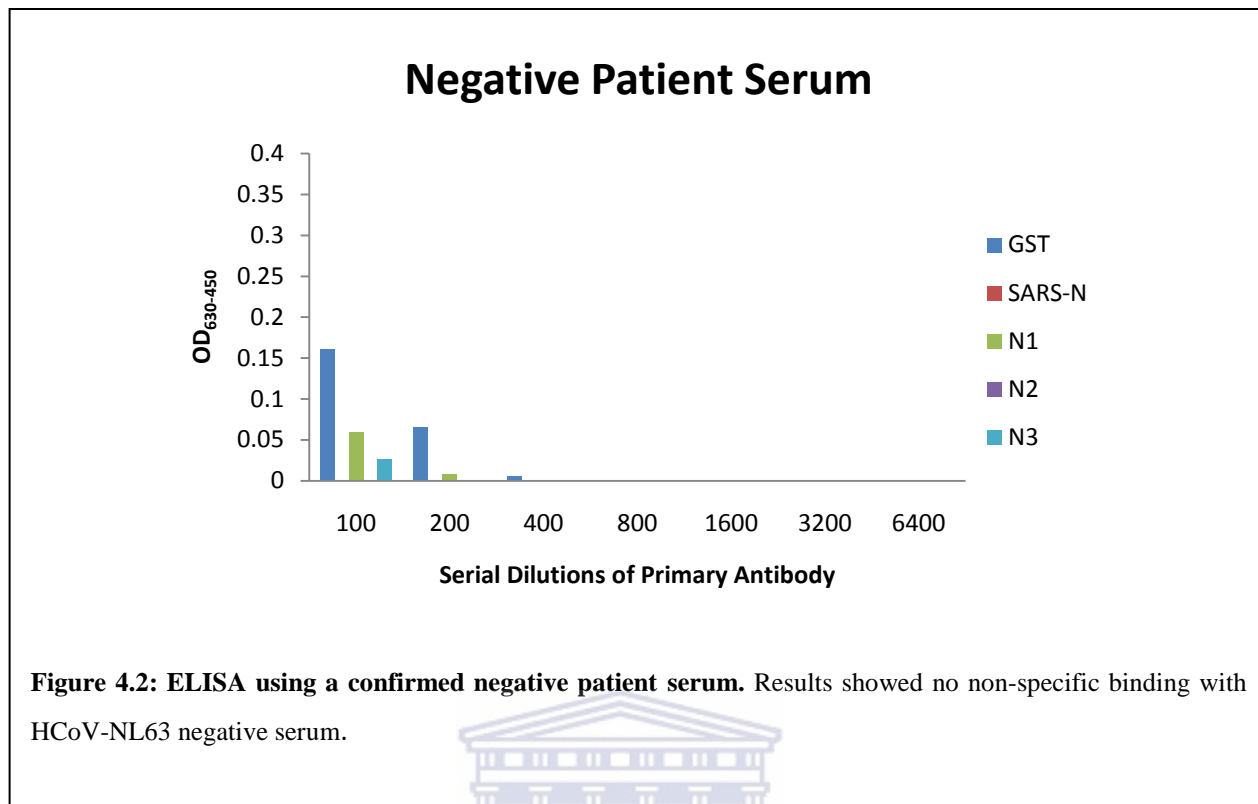
MedCalc Ver 12.0.0.0 (Zweig and Campbell 1993). When screening the HCoV-NL63 seropositive patients, the HCoV-NL63 N-protein was used as the positive control/experiment, and the SARS N-protein was used as the negative control. A statistical difference ($P < 0.05$) between the N2 and N3 construct would indicate a preferential immunogenic epitope. We therefore used a paired sample ANOVA, or T-test, from MedCalc Ver 12.0.0.0 to calculate the P-value.



4. Results

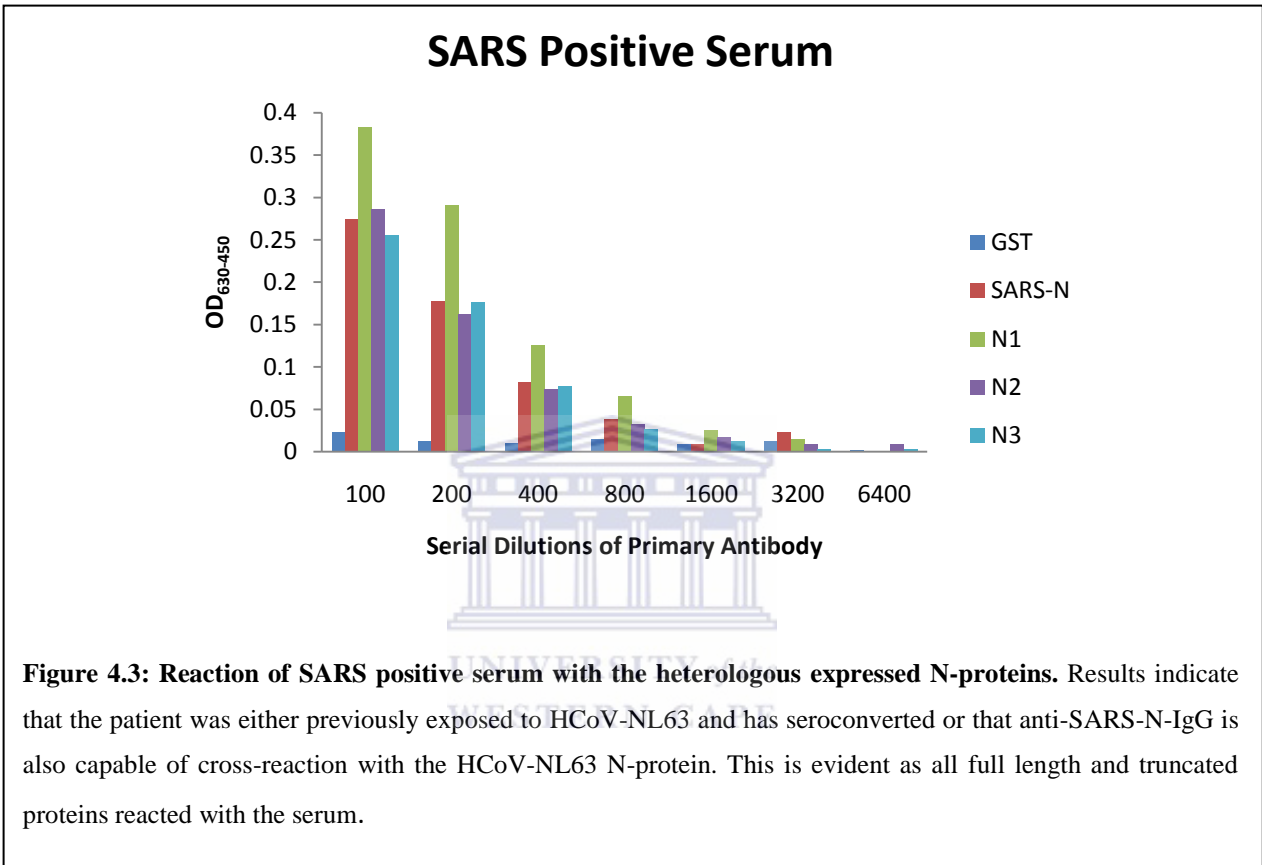
A positive control using rabbit anti-N-IgG against the HCoV-NL63 N-protein showed, as expected, reaction with the HCoV-NL63 N-protein, indicating sensitivity of the assay. The rabbit IgG also failed to detect any significant levels of the SARS-CoV N-homologue (Figure 4.1). These results indicated that the HCoV-NL63 N is both sensitive and specific in reacting with anti-N serum antibodies generated in rabbits and can therefore be used to screen patient serum. Rabbit anti-N-IgG was generated against a peptide in the N-terminal domain and therefore these results cannot be used to identify the antigenic epitope. A negative control, obtained from a confirmed HCoV-NL63 seronegative serum sample, also revealed no reaction, further indicating specificity of the assay (Figure 4.2).





When patient serum was screened all HCoV-NL63 positive patient sera reacted positively, as expected, with the full length N1 protein. Patient serum also, however, showed cross-reaction with the SARS-CoV N-protein but with a lower intensity in majority of patient sera. We therefore included a SARS-CoV positive serum control to differentiate between cross-reactivity and detection of SARS-CoV N-antibodies. The results revealed that both SARS-CoV and HCoV-NL63 N-proteins reacted with similar intensity at a dilution of 1:100 with SARS N decreasing with the concentration of primary antibody. In our previously confirmed SARS-CoV positive serum, the results indicated that the patient has previously also been exposed to HCoV-NL63, with HCoV-NL63 N eliciting a stronger immune response in comparison to SARS-CoV N, or it indicated that the SARS-positive patient serum was also capable of cross reaction with the HCoV-NL63 N-protein. These findings can, however, not be confirmed in this study and therefore requires additional research (Figure 4.3). ROC curve

analysis indicated that the N-protein is 100% sensitive and 46.7% specific at a criterion of >0.207 (Figure 4.4).



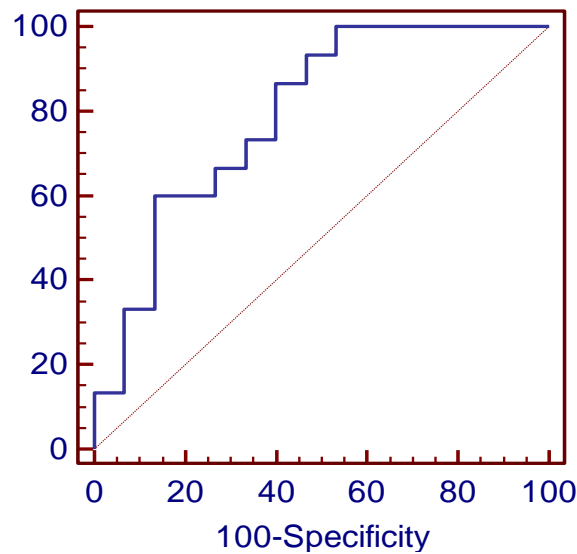


Figure 4.4: ROC curve analysis indicating the sensitivity and specificity of the N-protein in detecting serum antibodies. An ELISA with a perfect 100% sensitivity and specificity will have a diagonal line from the bottom left to top right corner, represented by the dashed line. Each sensitivity reading has a corresponding specificity reading. At the most valid criterion of >0.207 the ROC curve shows that the N-protein is 100% sensitive but only 46.7% specific in detecting anti-N-IgG antibodies.

With the exception of patient 1, all patients showed low levels of cross-reaction with the GST tag indicating that the antigenicity detected was not against the GST fusion protein. To identify the antigenic epitope of the HCoV-NL63 N-protein, the N2 and N3 constructs were analyzed using a paired sample ANOVA, or T-test, to identify any significant difference between the regions. A P-value of 0.37 indicated that the two constructs are not significantly different and therefore neither the N- or C-terminal was preferentially immunogenic. We were, therefore, unable to detect the antigenic epitope of the HCoV-NL63 N-protein as both the N- and C-terminals reacted with equal intensity for majority of patient serums, showing that both the N- and C-terminal of the HCoV-NL63 N-protein are equally immunogenic. Some patients did, however, show a preference for either the N- or C-terminal, which may indicate infection with various isolates of HCoV-NL63 or a patient specific immune response (Table 4.2).

Table 4.2: Patient antibody response to nucleocapsid constructs. The full length N-protein of HCoV-NL63 (N1) was taken as 100% response and therefore the remaining 3 constructs were represented as a percentage of N1. Values were taken from the standard curve where the response to GST was closest to 0.

Patient	HCoV-NL63 Full Length N (N1)	HCoV-NL63 N-Terminal (N2)	HCoV-NL63 C-Terminal (N3)	SARS-CoV Full Length N
1	100	51.64	55.96	83.59
2	100	67.30	57.62	32.93
3	100	56.32	68.52	54.39
4	100	72.22	72.22	63.77
5 ⁺	100	45.15	107.38	47.26
6	100	52.45	78.39	52.45
7	100	66.15	81.77	47.92
8	100	55.56	60.00	55.56
9	100	68.32	62.87	45.54
10	100	64.79	74.18	57.75
11	100	38.56	44.56	37.06
12 ⁺	100	64.15	100.00	40.39
13	100	94.29	83.37	89.82
14	100	16.22	33.33	25.23
15 ⁺	100	106.02	38.07	46.02
Average	100	61.27	67.88	51.98
St. Dev.	0	21.34	20.97	17.27

⁺Patients sera that showed a preference for either the N- or C-terminal of the HCoV-NL63 N

5. Discussion and Conclusion

The diagnosis of individual HCoV-NL63 would provide fundamental information on the distribution of these viruses and their health related implications. HCoV-NL63 is currently understood to be an aetiological agent of the “common cold”. However, it has recently been shown that this virus commonly presents as co-infection with more serious “flu-viruses” (Esper, Weibel *et al.* 2005; Fielding 2011). This indicates that HCoV-NL63 could pose as a predisposing factor to allow infection by more lethal respiratory viruses. As acute lower respiratory tract infections have been shown to result in approximately 1.9 million deaths in children younger than 5 (Venter, Lassauniere *et al.* 2011), the diagnosis of these predisposing factors would prove pivotal in the initiation of adequate and correct treatment.

The development of an ELISA for the detection of HCoV-NL63 antigens would provide an affordable, reproducible and reliable diagnostic assay. As the N-protein is highly antigenic and abundantly expressed during viral infection it would serve as a good marker in such an assay. In our study, however, the N-protein of SARS-CoV was detected by serum antibodies against HCoV-NL63. Although the N-protein showed 100% sensitivity in reacting with serum antibodies against HCoV-NL63, the high degree of cross-reactivity and low level of specificity raised questions as to the viability of using the N-protein as a diagnostic marker.

Previous studies have also reported cross-reactivity between HCoV-OC43 and HCoV-229E but failed to show cross-reaction between SARS-CoV and other known HCoVs (Lehmann, Wolf *et al.* 2008). Other reports however showed that two-way cross-reactivity between SARS-CoV and HCoV-229E, HCoV-NL63 and HCoV-OC43 is possible (Woo, Lau *et al.* 2004; Chan, Cheng *et al.* 2005; Che, Qiu *et al.* 2005) and that polyclonal antibodies from Group I coronaviruses including HCoV-229E, FIPV, Canine CoV and TGEV are capable of

reacting with SARS-CoV infected Vero cells (Sun and Meng 2004). The same report however failed to show any cross-reaction between SARS-CoV and the Group II coronavirus to which HCoV-OC43 belongs (Sun and Meng 2004; Venter, Lassauniere *et al.* 2011). Using recombinantly expressed antigenic epitopes, the region of this cross-reaction was localized between amino acids 120-208 of the SARS-CoV N-protein and cross reaction mediated by the N-protein of HCoV-NL63 was localized to amino acids 1-39 of the N-terminal (Vlasova, Zhang *et al.* 2007). This cross-reactivity has been proposed to be either at an antigenic level, with cross-reaction occurring through cross-reactive epitopes, or at an antibody level, with the production of polyclonal antibodies (Chan, Cheng *et al.* 2005). In our previous chapter it was shown that the N-proteins of CoVs lack a high degree of homology and therefore it can be assumed that this cross-reaction is mediated by the production of polyclonal antibodies. We have also shown that polyclonal antibodies generated in rabbits did not have any degree of cross-reaction, which further indicates that human antibodies have a broader degree of antigen recognition. The spike protein is not capable of cross-reaction and has therefore been used as a confirmatory step (Woo, Lau *et al.* 2004; Vlasova, Zhang *et al.* 2007).

Although the N-protein of coronaviruses show a high level of cross-reactivity, our study has shown that antibodies developed against the N-protein in rabbits has a lower degree of cross-reactivity. This is evident as our positive control using polyclonal rabbit-anti-N IgG was only able to detect the N-protein of HCoV-NL63 and not of SARS-CoV. A report by Che *et al.* (2005) also failed to detect any substantial level of cross-reactivity between HCoV-229E, HCoV-OC43 and SARS-CoV when immune rabbit serum was used (Che, Qiu *et al.* 2005). Identifying and directing detection against immunogenic epitopes will also increase specificity as the use of synthetic peptides, opposed to the full length protein, may show a

reduction in cross-reactivity (Sun and Meng 2004; Woo, Lau *et al.* 2004; Sastre, Dijkman *et al.* 2011). Several studies have also shown that the development of double antibody sandwich (DAS) ELISAs using a combination of monoclonal antibodies against the N-protein are highly sensitive and specific for the detection of HCoV-NL63, HCoV-229E (Sastre, Dijkman *et al.* 2011) and SARS-CoV (Lau, Woo *et al.* 2004).

We used truncated mutants corresponding to the N- and C-terminals of the HCoV-NL63 N-protein to attempt to identify the antigenic epitope. The results, however, showed that both terminals reacted equally but to a lesser degree than the full length N-protein with sera from HCoV-NL63 positive patients. This indicated either the presence of several immunogenic peptides, or the division of a single peptide between both fragments. The middle and C-terminal (110-422 aa) of the N-protein are responsible for eliciting an immune response in SARS-CoV infection (Lee, Lee *et al.* 2008). If the N-protein of both HCoV-NL63 and SARS-CoV behave similarly, it may substantiate our latter claim that the immunogenic epitope of HCoV-NL63N is located in the middle fragment. Both SARS-CoV and HCoV-NL63 have a long disordered region in the middle of their N-proteins which may expose this region to an immune response, explaining these results. This was also noted in a report showing that the disordered region of SARS-CoV and porcine CoV is responsible for mediating cross-reactivity (Vlasova, Zhang *et al.* 2007). The region of 122-422 of the N-protein of SARS-CoV was more specific than the full length N-protein. Therefore narrowing down of the antigenic epitope may increase sensitivity and specificity of a diagnostic assay and therefore inclusion of this middle region might yield valuable information (Yu, Le *et al.* 2007). However, due to the disordered nature of this region it poses difficult to express in a

bacterial system. Expression in cell-free system may be necessary to synthesize this region to determine its viability in a diagnostic assay.

The N-protein of HCoV-NL63 has been shown to be sensitive in the detection of serum antibodies. It is also capable of cross-reaction with other coronaviruses and therefore its use as a diagnostic marker is questionable. However, the production of monoclonal antibodies in rabbits using immunogenic peptides, possibly found within the middle, disordered region, rather than the use of the full length proteins, can greatly increase specificity.



Chapter 5

Synopsis



The human coronavirus NL63 has been shown to be a clinically important virus, affecting mainly children, the elderly and immunocompromised. Infection mainly results in a mild, upper respiratory tract disease with symptoms including fever, cough, sore throat and rhinorrhoea. HCoV-NL63 also presents with more severe lower respiratory tract findings with infection being associated with bronchiolitis, pneumonia and croup.

At a molecular level, the virus is assembled through the interaction of four structural proteins, namely the spike (S), membrane (M), envelope (E) and nucleocapsid (N) proteins. The N-protein of HCoV-NL63 is poorly characterized; however coronavirus homologues provide a good platform to base research on. The primary function of this protein is to interact with viral RNA to form the protective ribonucleocapsid core. For this particle to assemble, the N-protein has to undergo N-N homotypic interactions with subsequent interaction with viral RNA. The N-protein has also been shown to interact with the M-protein in the surface of the virion during viral assembly. Besides the primary structural role of the N-protein, it is also known to regulate viral transcription, translation and replication and certain cellular pathways. The N-protein is also highly phosphorylated and highly antigenic and has been shown to elicit a strong immune response.

We have used various bioinformatic tools and molecular techniques to characterize the properties of this protein. Special attention has been paid to conserved regions of coronavirus N-homologues, identification of disordered regions and putative phosphorylation sites and, with the use of molecular tools, we have cloned and expressed the N-protein of HCoV-NL63 and SARS-CoV. These heterologously expressed proteins have allowed us to identify antibody cross-reactivity between coronavirus species and immunogenic epitopes responsible for eliciting an antibody response.

Using multiple sequence alignment, coronavirus N-protein homologues have been shown to be poorly conserved, with the highest homology between HCoV-NL63 and HCoV-229E. However, these N-proteins all share the same function *i.e.* formation of the ribonucleocapsid core. We therefore aligned the regions of the N-protein homologues which have been experimentally proven to mediate these interactions to identify if these regions were more highly conserved. We, however, found that these regions are even less conserved than that of the full length N-protein homologues. Two highly conserved regions between all ten coronavirus homologues- FYYLGTGPH (at the N-terminal) and a SR rich motif – were identified. The region FYYLGTGPH may be responsible for mediating RNA binding, as the N-terminal has no other identified, specific function. The SR rich motif may be responsible for mediating N-N homotypic interactions, as shown by other reports.

Since the sequence homology between the coronavirus N-homologues was shown to be relatively low, even at regions known to mediate certain interactions, the order/disorder state of the N-protein was studied. Disordered regions of a protein have no fixed tertiary structure, thereby allowing access to binding sites. These regions have previously been shown to mediate protein-protein and protein-RNA/DNA interactions. Using the VSL1 and VLXT algorithms of PONDR, the N-protein of HCoV-NL63 was shown to be different to other coronavirus N-proteins, with regard to the order/disordered state. Majority of coronavirus N-proteins have been shown to have a structurally disordered N- and C-terminal with a third disordered region located in the middle. The N-protein of HCoV-NL63, however, has only one disordered region located in the middle of the protein. Using the SARS-CoV N-protein as a model, we showed with bioinformatic analysis that the majority of interactions the protein is understood to be involved in occur within these disordered regions. If disorder is essential for

mediating these interactions, the middle disordered region located between residues 109-248 of the HCoV-NL63 N-protein should form the hub of these interactions for HCoV-NL63. In this study, it was shown that the N-protein is highly phosphorylated, particularly at serine residues, and is also highly immunogenic. These sites of phosphorylation and increased immunogenicity have also been shown to preferentially occur within disordered regions.

To further characterize the N-protein, in an *in vitro* model, we have cloned and expressed the N-protein of SARS-CoV and HCoV-NL63. Using the viral RNA as a template, RT-PCR was used to generate cDNA of the HCoV-NL63 genome. The N-gene of HCoV-NL63 (N1) and SARS-CoV, as well as truncated mutants corresponding to the N- (N2) and C-terminal (N3) and the middle disordered region (N4) of the HCoV-NL63 N-protein, were PCR amplified and ligated into the pGEM[®] T-easy vector for sequencing. Once sequences were confirmed using M13 primers the genes were removed by *SgfI* and *PmeI* enzyme restriction. These enzymes allow for cloning into the Flexi[™] protein expression vector. The Flexi[™] vector appends a GST affinity tag to the N-terminal of the protein to facilitate detection and purification and was used to genetically transform the KRX strain of competent *E. coli*, which served as the protein expression model.

Transformed cells were grown in an LB culture and protein expression was induced with the addition of 0.1% rhamnose after an OD₆₀₀ of 0.4 was reached. Cultures were subsequently harvested by centrifugation and heterologously expressed proteins were extracted by sonication. Protein detection by western blot analysis against the GST affinity tag proved that the Flexi[™]/KRX system is a good model for the expression of the N-protein and generation of truncated clones. Several drawbacks were, however, encountered. These drawbacks include (1) the inability to express the N4 construct, most likely due to the

disordered nature of this region, (2) a high level of uninduced, background expression, which puts the cell under unnecessary stress in the logarithmic phase of growth and thereby reduces final yield, and (3) a high level of expressed proteins inserting into the insoluble cell fraction, making purification difficult. Utilizing several techniques known to increase protein solubility, we found the introduction of lysozymes into the lysis buffer to be the most effective in increasing protein solubility from the membrane bound cell fraction.

These heterologously expressed proteins were subsequently purified using a glutathione resin or MagneGSTTM purification system and can now serve the basis of several post-expression studies. Future studies out of the scope of this dissertation will include identifying the regions involved in dimerization and RNA binding, production of antibodies in a murine model and identifying cytokines produced in response to infection. We have developed an antibody capture ELISA using the recombinantly expressed proteins to identify antibody cross-reactivity between coronavirus species and to identify the region of the N-protein responsible for eliciting an immune response.

Using a cohort of 15 patients who had previously been infected with HCoV-NL63 we identified that the N-protein is 100% sensitive in detecting anti-N-IgG serum antibodies but only 46.7% specific as the N-protein of SARS-CoV was also able to detect these antibodies. We found a positive control in the form of rabbit-anti-N-IgG generated against a peptide in the N-terminal of the N-protein to be highly sensitive and specific. It has been shown that antibodies generated in animals against smaller immunogenic peptides to be more sensitive and specific compared to the full length protein. Identifying the immunogenic peptide of the N-protein would therefore pose beneficial prior to the production of antibodies in a murine model. Both the N- and C-terminal of the N-protein were, however, equally sensitive in

detecting these serum antibodies, therefore indicating the presence of several immunogenic peptides dispersed along the entire length of the HCoV-NL63 N-protein, or the division of one immunogenic peptide between the N- and C-terminal. Inclusion of the middle disordered region would therefore serve highly beneficial in characterizing the immunogenicity of the N-protein.



References



- Abdul-Rasool, S. and B. C. Fielding (2010). "Understanding Human Coronavirus HCoV-NL63." Open Virol J 4: 76-84.
- Almazan, F., C. Galan, *et al.* (2004). "The nucleoprotein is required for efficient coronavirus genome replication." J Virol 78(22): 12683-12688.
- 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.
- Baric, R. S., G. W. Nelson, *et al.* (1988). "Interactions between Coronavirus Nucleocapsid Protein and viral RNA - Implication for Viral Transcription." Journal of Virology 62(11): 4280-4287.
- Bastien, N., K. Anderson, *et al.* (2005). "Human coronavirus NL63 infection in Canada." J Infect Dis 191(4): 503-506.
- 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.
- Beaudette, F. R. and C. B. Hudson (1937). "Cultivation of the virus of infectious bronchitis." J Am Vet Med Assoc 90: 51-58.
- Belay, E. D., D. D. Erdman, *et al.* (2005). "Kawasaki disease and human coronavirus." J Infect Dis 192(2): 352-353; author reply 353.
- Bell, D. M., E. W. Brink, *et al.* (1981). "Kawasaki syndrome: description of two outbreaks in the United States." N Engl J Med 304(26): 1568-1575.
- Blom, N., S. Gammeltoft, *et al.* (1999). "Sequence and structure-based prediction of eukaryotic protein phosphorylation sites." J Mol Biol 294(5): 1351-1362.
- Bourquin, J. P., I. Stagljar, *et al.* (1997). "A serine/arginine-rich nuclear matrix cyclophilin interacts with the C-terminal domain of RNA polymerase II." Nucleic Acids Res 25(11): 2055-2061.
- Brittain-Long, R., S. Nord, *et al.* (2008). "Multiplex real-time PCR for detection of respiratory tract infections." J Clin Virol 41(1): 53-56.
- 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.
- Carrington, J. C., R. Haldeman, *et al.* (1993). "Internal cleavage and trans-proteolytic activities of the VPg-proteinase (NIa) of tobacco etch potyvirus in vivo." J Virol 67(12): 6995-7000.

- Chan, K. H., V. C. Cheng, *et al.* (2005). "Serological responses in patients with severe acute respiratory syndrome coronavirus infection and cross-reactivity with human coronaviruses 229E, OC43, and NL63." Clin Diagn Lab Immunol 12(11): 1317-1321.
- 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.
- Chang, C. K., S. C. Sue, *et al.* (2006). "Modular organization of SARS coronavirus nucleocapsid protein." J Biomed Sci 13(1): 59-72.
- Chang, L. Y., B. L. Chiang, *et al.* (2006). "Lack of association between infection with a novel human coronavirus (HCoV), HCoV-NH, and Kawasaki disease in Taiwan." J Infect Dis 193(2): 283-286.
- Chatterjee, A., M. A. Johnson, *et al.* (2009). "Nuclear magnetic resonance structure shows that the severe acute respiratory syndrome coronavirus-unique domain contains a macrodomain fold." J Virol 83(4): 1823-1836.
- Che, X. Y., W. Hao, *et al.* (2004). "Nucleocapsid protein as early diagnostic marker for SARS." Emerg Infect Dis 10(11): 1947-1949.
- Che, X. Y., L. W. Qiu, *et al.* (2005). "Antigenic cross-reactivity between severe acute respiratory syndrome-associated coronavirus and human coronaviruses 229E and OC43." J Infect Dis 191(12): 2033-2037.
- Che, X. Y., L. W. Qiu, *et al.* (2004). "Sensitive and specific monoclonal antibody-based capture enzyme immunoassay for detection of nucleocapsid antigen in sera from patients with severe acute respiratory syndrome." J Clin Microbiol 42(6): 2629-2635.
- Cheever, F. S., J. B. Daniels, *et al.* (1949). "A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin." J Exp Med 90(3): 181-210.
- Chen, C. Y., C. K. Chang, *et al.* (2007). "Structure of the SARS coronavirus nucleocapsid protein RNA-binding dimerization domain suggests a mechanism for helical packaging of viral RNA." J Mol Biol 368(4): 1075-1086.
- 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.
- 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.

- Clark, B. and M. McKendrick (2004). "A review of viral gastroenteritis." Curr Opin Infect Dis 17(5): 461-469.
- Cologna, R., J. F. Spagnolo, *et al.* (2000). "Identification of nucleocapsid binding sites within coronavirus-defective genomes." Virology 277(2): 235-249.
- 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.
- Das, D., S. Kammila, *et al.* (2010). "Development, characterization, and application of monoclonal antibodies against severe acute respiratory syndrome coronavirus nucleocapsid protein." Clin Vaccine Immunol 17(12): 2033-2036.
- Dijkman, R., M. F. Jebbink, *et al.* (2008). "Human coronavirus NL63 and 229E seroconversion in children." J Clin Microbiol 46(7): 2368-2373.
- Dougherty, W. G. and T. D. Parks (1991). "Post-translational processing of the tobacco etch virus 49-kDa small nuclear inclusion polyprotein: identification of an internal cleavage site and delimitation of VPg and proteinase domains." Virology 183(2): 449-456.
- Doyle, L. P. and L. M. Hutchings (1946). "A transmissible gastroenteritis in pigs." J Am Vet Med Assoc 108: 257-259.
- Drosten, C., S. Gunther, *et al.* (2003). "Identification of a novel coronavirus in patients with severe acute respiratory syndrome." N Engl J Med 348(20): 1967-1976.
- Dunker, A. K., E. Garner, *et al.* (1998). "Protein disorder and the evolution of molecular recognition: theory, predictions and observations." Pac Symp Biocomput: 473-484.
- Dunker, A. K., J. D. Lawson, *et al.* (2001). "Intrinsically disordered protein." Journal of Molecular Graphics and Modelling 19(1): 26-59.
- Ebihara, T., R. Endo, *et al.* (2005). "Detection of human coronavirus NL63 in young children with bronchiolitis." J Med Virol 75(3): 463-465.
- 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.
- Enserink, M. (2005). "Virology. One virus, three names, three claims." Science 307(5709): 493.
- Esper, F., E. D. Shapiro, *et al.* (2005). "Association between a novel human coronavirus and Kawasaki disease." J Infect Dis 191(4): 499-502.

- 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.
- 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.
- Fielding, B. C. (2011). "Human coronavirus NL63: a clinically important virus?" Future Microbiol 6(2): 153-159.
- Fielding, B. C. and T. Suliman (2009). "Comparative analysis of human coronavirus-NL63 ORF3 protein homologues." African Journal of Biotechnology 8(14): 3175-3178.
- 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.
- François, B. (1999). "Recombinant protein expression in Escherichia coli." Current Opinion in Biotechnology 10(5): 411-421.
- Garner, E., P. Romero, *et al.* (1999). "Predicting Binding Regions within Disordered Proteins." Genome Inform Ser Workshop Genome Inform 10: 41-50.
- Georgiou, G. and P. Valax (1996). "Expression of correctly folded proteins in Escherichia coli." Curr Opin Biotechnol 7(2): 190-197.
- Giedroc, D. P., C. A. Theimer, *et al.* (2000). "Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting." J Mol Biol 298(2): 167-185.
- Girard, M. P., T. Cherian, *et al.* (2005). "A review of vaccine research and development: human acute respiratory infections." Vaccine 23(50): 5708-5724.
- Glass, D. B., M. R. el-Maghrabi, *et al.* (1986). "Synthetic peptides corresponding to the site phosphorylated in 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase as substrates of cyclic nucleotide-dependent protein kinases." J Biol Chem 261(6): 2987-2993.
- Glass, D. B. and S. B. Smith (1983). "Phosphorylation by cyclic GMP-dependent protein kinase of a synthetic peptide corresponding to the autophosphorylation site in the enzyme." J Biol Chem 258(24): 14797-14803.
- Glick, B. R. (1995). "Metabolic load and heterologous gene expression." Biotechnol Adv 13(2): 247-261.

- Golda, A., N. Malek, *et al.* (2011). "Infection with human coronavirus NL63 enhances streptococcal adherence to epithelial cells." J Gen Virol 92(Pt 6): 1358-1368.
- Graveley, B. R. and T. Maniatis (1998). "Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing." Mol Cell 1(5): 765-771.
- Grossoehme, N. E., L. Li, *et al.* (2009). "Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA duplexes." J Mol Biol 394(3): 544-557.
- Guan, Y., B. J. Zheng, *et al.* (2003). "Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China." Science 302(5643): 276-278.
- 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.
- Han, T. H., J. Y. Chung, *et al.* (2007). "Human Coronavirus-NL63 infections in Korean children, 2004-2006." J Clin Virol 38(1): 27-31.
- Hannig, G. and S. C. Makrides (1998). "Strategies for optimizing heterologous protein expression in *Escherichia coli*." Trends Biotechnol 16(2): 54-60.
- Hartnett, J., J. Gracyalny, *et al.* (2006). "The Single Step (KRX) Competent Cells: Efficient Cloning and High Protein Yields." Promega Notes 94: 27-30.
- He, R., F. Dobie, *et al.* (2004). "Analysis of multimerization of the SARS coronavirus nucleocapsid protein." Biochem Biophys Res Commun 316(2): 476-483.
- He, R., A. Leeson, *et al.* (2003). "Activation of AP-1 signal transduction pathway by SARS coronavirus nucleocapsid protein." Biochem Biophys Res Commun 311(4): 870-876.
- He, R., A. Leeson, *et al.* (2003). "Activation of AP-1 signal transduction pathway by SARS coronavirus nucleocapsid protein." Biochem Biophys Res Commun 311(4): 870-876.
- 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.
- Hiscox, J. A., T. Wurm, *et al.* (2001). "The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus." J Virol 75(1): 506-512.
- Hoffman, S. J., D. L. Looker, *et al.* (1990). "Expression of fully functional tetrameric human hemoglobin in *Escherichia coli*." Proc Natl Acad Sci U S A 87(21): 8521-8525.

- 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.
- Holcroft, C. C. and S. M. Egan (2000). "Roles of cyclic AMP receptor protein and the carboxyl-terminal domain of the alpha subunit in transcription activation of the *Escherichia coli* rhaBAD operon." J Bacteriol 182(12): 3529-3535.
- Huang, C. Y., Y. L. Hsu, *et al.* (2009). "Elucidation of the stability and functional regions of the human coronavirus OC43 nucleocapsid protein." Protein Sci 18(11): 2209-2218.
- Huang, Q., L. Yu, *et al.* (2004). "Structure of the N-terminal RNA-binding domain of the SARS CoV nucleocapsid protein." Biochemistry 43(20): 6059-6063.
- Hurst, K. R., C. A. Koetzner, *et al.* (2009). "Identification of in vivo-interacting domains of the murine coronavirus nucleocapsid protein." J Virol 83(14): 7221-7234.
- Iakoucheva, L. M., P. Radivojac, *et al.* (2004). "The importance of intrinsic disorder for protein phosphorylation." Nucleic Acids Res 32(3): 1037-1049.
- Kahn, J. S. and K. McIntosh (2005). "History and Recent Advances in Coronavirus Discovery." The Pediatric Infectious Disease Journal 24(Supplement): S223-S227.
- 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.
- Kheyami, A. M., T. Nakagomi, *et al.* (2010). "Detection of coronaviruses in children with acute gastroenteritis in Maddina, Saudi Arabia." Ann Trop Paediatr 30(1): 45-50.
- Kim, T. W., J. H. Lee, *et al.* (2004). "Generation and Characterization of DNA Vaccines Targeting the Nucleocapsid Protein of Severe Acute Respiratory Syndrome Coronavirus." J Virol 78(9): 4638-4645.
- Kishimoto, A., K. Nishiyama, *et al.* (1985). "Studies on the phosphorylation of myelin basic protein by protein kinase C and adenosine 3':5'-monophosphate-dependent protein kinase." J Biol Chem 260(23): 12492-12499.
- Kleines, M., S. Scheithauer, *et al.* (2007). "High prevalence of human bocavirus detected in young children with severe acute lower respiratory tract disease by use of a standard PCR protocol and a novel real-time PCR protocol." J Clin Microbiol 45(3): 1032-1034.
- Koetz, A., P. Nilsson, *et al.* (2006). "Detection of human coronavirus NL63, human metapneumovirus and respiratory syncytial virus in children with respiratory tract infections in south-west Sweden." Clin Microbiol Infect 12(11): 1089-1096.

- Kopecky-Bromberg, S. A., L. Martinez-Sobrido, *et al.* (2007). "Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists." J Virol 81(2): 548-557.
- Krokhin, O., Y. Li, *et al.* (2003). "Mass spectrometric characterization of proteins from the SARS virus: a preliminary report." Mol Cell Proteomics 2(5): 346-356.
- Ksiazek, T. G., D. Erdman, *et al.* (2003). "A novel coronavirus associated with severe acute respiratory syndrome." N Engl J Med 348(20): 1953-1966.
- Kuo, L. and P. S. Masters (2002). "Genetic Evidence for a Structural Interaction between the Carboxy Termini of the Membrane and Nucleocapsid Proteins of Mouse Hepatitis Virus." J Virol 76(10): 4987-4999.
- Kuypers, J., E. T. Martin, *et al.* (2007). "Clinical disease in children associated with newly described coronavirus subtypes." Pediatrics 119(1): e70-76.
- Larkin, M. A., G. Blackshields, *et al.* (2007). "Clustal W and Clustal X version 2.0." Bioinformatics 23(21): 2947-2948.
- Lau, S. K., K. S. Li, *et al.* (2010). "Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events." J Virol 84(6): 2808-2819.
- Lau, S. K., P. C. Woo, *et al.* (2004). "Detection of severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein in sars patients by enzyme-linked immunosorbent assay." J Clin Microbiol 42(7): 2884-2889.
- Lee, H. K., B. H. Lee, *et al.* (2008). "Detection of antibodies against SARS-Coronavirus using recombinant truncated nucleocapsid proteins by ELISA." J Microbiol Biotechnol 18(10): 1717-1721.
- Lehmann, C., R. Klar, *et al.* (2009). "Kawasaki disease lacks association with human coronavirus NL63 and human bocavirus." Pediatr Infect Dis J 28(6): 553-554.
- Lehmann, C., H. Wolf, *et al.* (2008). "A line immunoassay utilizing recombinant nucleocapsid proteins for detection of antibodies to human coronaviruses." Diagn Microbiol Infect Dis 61(1): 40-48.
- Leung, T. F., C. Y. Li, *et al.* (2009). "Epidemiology and clinical presentations of human coronavirus NL63 infections in hong kong children." J Clin Microbiol 47(11): 3486-3492.
- Leung, W. K., K. F. To, *et al.* (2003). "Enteric involvement of severe acute respiratory syndrome-associated coronavirus infection." Gastroenterology 125(4): 1011-1017.

- Li, F. Q., H. Xiao, *et al.* (2005). "Sumoylation of the nucleocapsid protein of severe acute respiratory syndrome coronavirus." FEBS Lett 579(11): 2387-2396.
- Li, X., P. Romero, *et al.* (1999). "Predicting Protein Disorder for N-, C-, and Internal Regions." Genome Inform Ser Workshop Genome Inform 10: 30-40.
- Liao, Y., J. Lescar, *et al.* (2004). "Expression of SARS-coronavirus envelope protein in *Escherichia coli* cells alters membrane permeability." Biochemical and Biophysical Research Communications 325(1): 374-380.
- Lim, K. P. and D. X. Liu (2001). "The missing link in coronavirus assembly. Retention of the avian coronavirus infectious bronchitis virus envelope protein in the pre-Golgi compartments and physical interaction between the envelope and membrane proteins." J Biol Chem 276(20): 17515-17523.
- Luo, C., H. Luo, *et al.* (2004). "Nucleocapsid protein of SARS coronavirus tightly binds to human cyclophilin A." Biochem Biophys Res Commun 321(3): 557-565.
- 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.
- Luo, H., Q. Chen, *et al.* (2005). "The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1." FEBS Lett 579(12): 2623-2628.
- 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.
- Ma, Y., X. Tong, *et al.* (2010). "Structures of the N- and C-terminal domains of MHV-A59 nucleocapsid protein corroborate a conserved RNA-protein binding mechanism in coronavirus." Protein Cell 1(7): 688-697.
- Marra, M. A., S. J. Jones, *et al.* (2003). "The Genome sequence of the SARS-associated coronavirus." Science 300(5624): 1399-1404.
- Masters, P. S. (1992). "Localization of an RNA-binding domain in the nucleocapsid protein of the coronavirus mouse hepatitis virus." Arch Virol 125(1-4): 141-160.
- McIntosh, K. (2005). "Coronaviruses in the limelight." J Infect Dis 191(4): 489-491.
- 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.

- Menzella, H. G., H. C. Gramajo, *et al.* (2002). "High recovery of prochymosin from inclusion bodies using controlled air oxidation." Protein Expr Purif 25(2): 248-255.
- Minosse, C., M. Selleri, *et al.* (2008). "Phylogenetic analysis of human coronavirus NL63 circulating in Italy." J Clin Virol 43(1): 114-119.
- 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.
- Namy, O., S. J. Moran, *et al.* (2006). "A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting." Nature 441(7090): 244-247.
- Nelson, G. W., S. A. Stohlman, *et al.* (2000). "High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA." J Gen Virol 81(Pt 1): 181-188.
- Nicholas, K. B. and H. B. Nicholas (1997). "Genedoc: a tool for editing and annotating multiple sequence alignments." Distributed by author.
- Obradovic, Z., K. Peng, *et al.* (2005). "Exploiting heterogeneous sequence properties improves prediction of protein disorder." Proteins 61 Suppl 7: 176-182.
- Oosterhof, L., C. B. Christensen, *et al.* (2010). "Fatal lower respiratory tract disease with human corona virus NL63 in an adult haematopoietic cell transplant recipient." Bone Marrow Transplant 45(6): 1115-1116.
- Parker, K. C., M. A. Bednarek, *et al.* (1994). "Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains." J Immunol 152(1): 163-175.
- Peiris, J. S., S. T. Lai, *et al.* (2003). "Coronavirus as a possible cause of severe acute respiratory syndrome." Lancet 361(9366): 1319-1325.
- Peti, W. and R. Page (2007). "Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost." Protein Expr Purif 51(1): 1-10.
- Pinna, L. A. (1990). "Casein kinase 2: an 'eminence grise' in cellular regulation?" Biochim Biophys Acta 1054(3): 267-284.
- Principi, N., S. Bosis, *et al.* (2010). "Effects of coronavirus infections in children." Emerg Infect Dis 16(2): 183-188.
- Pyrk, K., B. Berkhout, *et al.* (2007). "The novel human coronaviruses NL63 and HKU1." J Virol 81(7): 3051-3057.

- Pyrk, K., R. Dijkman, *et al.* (2006). "Mosaic structure of human coronavirus NL63, one thousand years of evolution." J Mol Biol. 364(5): 964-973
- Pyrk, K., M. F. Jebbink, *et al.* (2004). "Genome structure and transcriptional regulation of human coronavirus NL63." Virology 1: 7.
- 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.
- Rinas, U., F. Hoffmann, *et al.* (2007). "Inclusion body anatomy and functioning of chaperone-mediated in vivo inclusion body disassembly during high-level recombinant protein production in Escherichia coli." J Biotechnol 127(2): 244-257.
- Risku, M., S. Lappalainen, *et al.* (2010). "Detection of human coronaviruses in children with acute gastroenteritis." Journal of Clinical Virology 48(1): 27-30.
- Romero, Obradovic, *et al.* (1997). "Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family." Genome Inform Ser Workshop Genome Inform 8: 110-124.
- Romero, P., Z. Obradovic, *et al.* (2001). "Sequence complexity of disordered protein." Proteins 42(1): 38-48.
- Rowland, R. R., V. Chauhan, *et al.* (2005). "Intracellular localization of the severe acute respiratory syndrome coronavirus nucleocapsid protein: absence of nucleolar accumulation during infection and after expression as a recombinant protein in vero cells." J Virol 79(17): 11507-11512.
- Rowley, A. H., S. T. Shulman, *et al.* (2000). "IgA plasma cell infiltration of proximal respiratory tract, pancreas, kidney, and coronary artery in acute Kawasaki disease." J Infect Dis 182(4): 1183-1191.
- Sastre, P., R. Dijkman, *et al.* (2011). "Differentiation between human coronaviruses NL63 and 229E using a novel double-antibody sandwich enzyme-linked immunosorbent assay based on specific monoclonal antibodies." Clin Vaccine Immunol 18(1): 113-118.
- Sawicki, S. G., D. L. Sawicki, *et al.* (2007). "A contemporary view of coronavirus transcription." J Virol 81(1): 20-29.
- Schagat, T., R. Ohana, *et al.* (2008). "KRX Autoinduction Protocol: A Convenient Method for Protein Expression." Promega Notes 98: 16-18.
- Shao, X., X. Guo, *et al.* (2007). "Seroepidemiology of group I human coronaviruses in children." J Clin Virol 40(3): 207-213.

- 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.
- Singh, S. M. and A. K. Panda (2005). "Solubilization and refolding of bacterial inclusion body proteins." J Biosci Bioeng 99(4): 303-310.
- 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.
- Spencer, K. A. and J. A. Hiscox (2006). "Characterisation of the RNA binding properties of the coronavirus infectious bronchitis virus nucleocapsid protein amino-terminal region." FEBS Lett 580(25): 5993-5998.
- Steczko, J., G. A. Donoho, *et al.* (1991). "Effect of ethanol and low-temperature culture on expression of soybean lipoxygenase L-1 in *Escherichia coli*." Protein Expr Purif 2(2-3): 221-227.
- Stohlman, S. A., R. S. Baric, *et al.* (1988). "Specific interaction between coronavirus leader RNA and nucleocapsid protein." J Virol 62(11): 4288-4295.
- Sun, Z. F. and X. J. Meng (2004). "Antigenic cross-reactivity between the nucleocapsid protein of severe acute respiratory syndrome (SARS) coronavirus and polyclonal antisera of antigenic group I animal coronaviruses: implication for SARS diagnosis." J Clin Microbiol 42(5): 2351-2352.
- Sung, J. Y., H. J. Lee, *et al.* (2010). "Role of human coronavirus NL63 in hospitalized children with croup." Pediatr Infect Dis J 29(9): 822-826.
- 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.
- Surjit, M. and S. K. Lal (2008). "The SARS-CoV nucleocapsid protein: a protein with multifarious activities." Infect Genet Evol 8(4): 397-405.
- Surjit, M., B. Liu, *et al.* (2006). "The nucleocapsid protein of severe acute respiratory syndrome-coronavirus inhibits the activity of cyclin-cyclin-dependent kinase complex and blocks S phase progression in mammalian cells." J Biol Chem 281(16): 10669-10681.
- Surjit, M., B. Liu, *et al.* (2004). "The SARS coronavirus nucleocapsid protein induces actin reorganization and apoptosis in COS-1 cells in the absence of growth factors." Biochem J 383(Pt 1): 13-18.

- Surjit, M., B. Liu, *et al.* (2004). "The nucleocapsid protein of the SARS coronavirus is capable of self-association through a C-terminal 209 amino acid interaction domain." Biochem Biophys Res Commun 317(4): 1030-1036.
- Suzuki, A., M. Okamoto, *et al.* (2005). "Detection of human coronavirus-NL63 in children in Japan." Pediatr Infect Dis J 24(7): 645-646.
- Talbot, H. K., J. E. Crowe, Jr., *et al.* (2009). "Coronavirus infection and hospitalizations for acute respiratory illness in young children." J Med Virol 81(5): 853-856.
- Tan, Y. W., S. Fang, *et al.* (2006). "Amino acid residues critical for RNA-binding in the N-terminal domain of the nucleocapsid protein are essential determinants for the infectivity of coronavirus in cultured cells." Nucleic Acids Res 34(17): 4816-4825.
- Tang, T. K., M. P. Wu, *et al.* (2005). "Biochemical and immunological studies of nucleocapsid proteins of severe acute respiratory syndrome and 229E human coronaviruses." Proteomics 5(4): 925-937.
- 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.
- Timani, K. A., L. Ye, *et al.* (2004). "Cloning, sequencing, expression, and purification of SARS-associated coronavirus nucleocapsid protein for serodiagnosis of SARS." J Clin Virol 30(4): 309-312.
- Tyrrell, D. A. and M. L. Bynoe (1965). "Cultivation of a Novel Type of Common-Cold Virus in Organ Cultures." Br Med J 1(5448): 1467-1470.
- Tyrrell, D. A. and M. L. Bynoe (1966). "Cultivation of viruses from a high proportion of patients with colds." Lancet 1(7428): 76-77.
- Vabret, A., T. Mourez, *et al.* (2005). "Human coronavirus NL63, France." Emerg Infect Dis 11(8): 1225-1229.
- van der Hoek, L., G. Ithorst, *et al.* (2010). "Burden of disease due to human coronavirus NL63 infections and periodicity of infection." Journal of Clinical Virology 48(2): 104-108.
- van der Hoek, L., K. Pyrc, *et al.* (2006). "Human coronavirus NL63, a new respiratory virus." FEMS Microbiol Rev 30(5): 760-773.
- van der Hoek, L., K. Pyrc, *et al.* (2004). "Identification of a new human coronavirus." Nat Med 10(4): 368-373.
- van der Hoek, L., K. Sure, *et al.* (2005). "Croup is associated with the novel coronavirus NL63." PLoS Med 2(8): e240.

- van Elden, L. J., A. M. van Loon, *et al.* (2004). "Frequent detection of human coronaviruses in clinical specimens from patients with respiratory tract infection by use of a novel real-time reverse-transcriptase polymerase chain reaction." J Infect Dis 189(4): 652-657.
- Venter, M., R. Lassauniere, *et al.* (2011). "Contribution of common and recently described respiratory viruses to annual hospitalizations in children in South Africa." J Med Virol 83(8): 1458-1468.
- Verheije, M. H., M. C. Hagemeyer, *et al.* (2010). "The coronavirus nucleocapsid protein is dynamically associated with the replication-transcription complexes." J Virol 84(21): 11575-11579.
- Verma, S., V. Bednar, *et al.* (2006). "Identification of Functionally Important Negatively Charged Residues in the Carboxy End of Mouse Hepatitis Coronavirus A59 Nucleocapsid Protein." J Virol 80(9): 4344-4355.
- Vlasova, A. N., X. Zhang, *et al.* (2007). "Two-way antigenic cross-reactivity between severe acute respiratory syndrome coronavirus (SARS-CoV) and group 1 animal CoVs is mediated through an antigenic site in the N-terminal region of the SARS-CoV nucleoprotein." J Virol 81(24): 13365-13377.
- Wang, Y., X. Wu, *et al.* (2004). "Low Stability of Nucleocapsid Protein in SARS Virus."
- Wang, Y. and X. Zhang (1999). "The nucleocapsid protein of coronavirus mouse hepatitis virus interacts with the cellular heterogeneous nuclear ribonucleoprotein A1 in vitro and in vivo." Virology 265(1): 96-109.
- Weickert, M. J., D. H. Doherty, *et al.* (1996). "Optimization of heterologous protein production in *Escherichia coli*." Curr Opin Biotechnol 7(5): 494-499.
- Welling, G. W., W. J. Weijer, *et al.* (1985). "Prediction of sequential antigenic regions in proteins." FEBS Lett 188(2): 215-218.
- 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.
- Woo, P. C., Y. Huang, *et al.* (2010). "Coronavirus genomics and bioinformatics analysis." Viruses 2(8): 1804-1820.
- 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.
- Woo, P. C., S. K. Lau, *et al.* (2009). "Coronavirus diversity, phylogeny and interspecies jumping." Exp Biol Med (Maywood) 234(10): 1117-1127.

- Woo, P. C., S. K. Lau, *et al.* (2004). "Longitudinal profile of immunoglobulin G (IgG), IgM, and IgA antibodies against the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein in patients with pneumonia due to the SARS coronavirus." Clin Diagn Lab Immunol 11(4): 665-668.
- Woo, P. C., S. K. Lau, *et al.* (2004). "False-positive results in a recombinant severe acute respiratory syndrome-associated coronavirus (SARS-CoV) nucleocapsid enzyme-linked immunosorbent assay due to HCoV-OC43 and HCoV-229E rectified by Western blotting with recombinant SARS-CoV spike polypeptide." J Clin Microbiol 42(12): 5885-5888.
- Woodgett, J. R., K. L. Gould, *et al.* (1986). "Substrate specificity of protein kinase C." European Journal of Biochemistry 161(1): 177-184.
- 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.
- 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.
- 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.
- Yan, X., Q. Hao, *et al.* (2006). "Nucleocapsid protein of SARS-CoV activates the expression of cyclooxygenase-2 by binding directly to regulatory elements for nuclear factor-kappa B and CCAAT/enhancer binding protein." Int J Biochem Cell Biol 38(8): 1417-1428.
- Yang, L., L. J. Embree, *et al.* (1998). "Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing." J Biol Chem 273(43): 27761-27764.
- 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.
- Yu, F., M. Q. Le, *et al.* (2007). "Recombinant truncated nucleocapsid protein as antigen in a novel immunoglobulin M capture enzyme-linked immunosorbent assay for diagnosis of severe acute respiratory syndrome coronavirus infection." Clin Vaccine Immunol 14(2): 146-149.
- Yu, I. M., C. L. Gustafson, *et al.* (2005). "Recombinant severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein forms a dimer through its C-terminal domain." J Biol Chem 280(24): 23280-23286.

- Yuryev, A., M. Patturajan, *et al.* (1996). "The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins." Proc Natl Acad Sci U S A 93(14): 6975-6980.
- Zhou, M. and E. W. Collisson (2000). "The amino and carboxyl domains of the infectious bronchitis virus nucleocapsid protein interact with 3' genomic RNA." Virus Res 67(1): 31-39.
- Zweig, M. H. and G. Campbell (1993). "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine." Clin Chem 39(4): 561-577.



UNIVERSITY OF THE WESTERN CAPE

Appendix



Department of Medical Bioscience, University of the Western Cape, Cape Town, South Africa

```

SARS_N_F : AWRKGGGGCGAATTGGGGCCGACGTGCGATGCTCCCGGCCCCATGGCGGGCCGGGAATTCGATTAGGAGCGATCGCCATGCTGATAATGGACCCCAATCAAAAC : 106
SARS_N_FL : -----ATGTCGTGATAATGGACCCCAATCAAAAC----- : 27
SARS_N_R : -----atgtctgataatggaccccaatcaaac----- : -

SARS_N_F : CAACGTAAGTGGCCCCGGCATTACATTTGGTGGACCCACAGATTCACTGACAATAACCCAGAATGGAGGACGCAATGGGGCAAGGCCAAAACAGCCGGACCCCAAG : 212
SARS_N_FL : CAACGTAAGTGGCCCCGGCATTACATTTGGTGGACCCACAGATTCACTGACAATAACCCAGAATGGAGGACGCAATGGGGCAAGGCCAAAACAGCCGGACCCCAAG : 133
SARS_N_R : -----caacgtagtgccccgcgattacatttggtagccacacagattcaactgacaataaccagaatggaggagcgaatggggcaaggccaaaacagcggcggaccccaag : -

SARS_N_F : GTTTACCCAATAATAATGGGTCITGGTTACAGGCTCACTCAGCATGGCAAGGAGGAACCTTAGATTCCCTCGAGCCAGGGCGTTCCATCAACACCCATATGTTGG : 318
SARS_N_FL : GTTTACCCAATAATAATGGGTCITGGTTACAGGCTCACTCAGCATGGCAAGGAGGAACCTTAGATTCCCTCGAGCCAGGGCGTTCCATCAACACCCATATGTTGG : 239
SARS_N_R : -----gtttaccaataata tgcgtcttggttcacagc ctcactcagcatggcaaggaggaacttagattccctcgaggCcAGGGCGTTCCATcaACACCATaTgTgg : 26

SARS_N_F : TCCAGATGACCAATTTGGCTACTACCGAAGAGTACCAGCAGATTCGTGGTGGTACGGGCAAAATGAAAGAGTACAGCCACAGATGTTACTTCTTACCTAGGA : 424
SARS_N_FL : TCCAGATGACCAATTTGGCTACTACCGAAGAGTACCAGCAGATTCGTGGTGGTACGGGCAAAATGAAAGAGTACAGCCACAGATGTTACTTCTTACCTAGGA : 345
SARS_N_R : -----tcCAGatgaccaaatTggcTactACCGAAGAGTACCAGCAGGtTcGTGGTgTgtagcggcaAAATGAAAGAGTcAGCCcCAGATGgtacTtCTaTTacCTAGga : 101

SARS_N_F : ACTGCCCCAGAGCTCACTTCCCTACCGCGCTAACAAAGAGCATCGTATGGGTGCACTGAGGGAGCCTTGAATACACCRAAGACACATGGCCCCCGCA : 530
SARS_N_FL : ACTGCCCCAGAGCTCACTTCCCTACCGCGCTAACAAAGAGCATCGTATGGGTGCACTGAGGGAGCCTTGAATACACCRAAGACACATGGCCCCCGCA : 451
SARS_N_R : -----ACTG_cCCAGAG--cTcAcT--cTAcC_cCGTc_cAAAGRAc_cATCGTATGG--tGCAACTGAGGAGc--cGAATACAc_cRAAGAc_cACAT_cGCACCCCGCA : 189

SARS_N_F : ATCCTAATFACAATGCTGCCACCGTGTACAACCTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGAGAGGGAAGCAGAGGGCGGCACTCAAGCCCTTCTCCG : 636
SARS_N_FL : ATCCTAATFACAATGCTGCCACCGTGTACAACCTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGAGAGGGAAGCAGAGGGCGGCACTCAAGCCCTTCTCCG : 557
SARS_N_R : -----ATCCTA--TACAATGCTGCCACCGTGTACAACCTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGAGAGGGAAGCAGAGGGCGGCACTCAAGCCCTTCTCCG : 293

SARS_N_F : CTCTCATCAGCTAGTCGCGGTAATFCAAGAAATTCAACTCCTGGCAGCAGTAGGGGAAATTCCTCTCGAATGGCTAGCGGAGGTGGTGAACCTGCCCTCGCC : 742
SARS_N_FL : CTCTCATCAGCTAGTCGCGGTAATFCAAGAAATTCAACTCCTGGCAGCAGTAGGGGAAATTCCTCTCGAATGGCTAGCGGAGGTGGTGAACCTGCCCTCGCC : 663
SARS_N_R : -----CTCTCATCAGCTAGTCGCGGTAATFCAAGAAATTCAACTCCTGGCAGCAGTAGGGGAAATTCCTCTCGAATGGCTAGCGGAGGTGGTGAACCTGCCCTCGCC : 399

SARS_N_F : CTATTTCGCTAGCAGATGAACCAGCTTGAGAGCAAAGTTCTGGTAAAGGCCAACACACAAAGGCCAACTGTCACTAAGAAATCTGCTGCTGAGGCATCTA : 848
SARS_N_FL : CTATTTCGCTAGCAGATGAACCAGCTTGAGAGCAAAGTTCTGGTAAAGGCCAACACACAAAGGCCAACTGTCACTAAGAAATCTGCTGCTGAGGCATCTA : 769
SARS_N_R : -----CTATTTCGCTAGCAGATGAACCAGCTTGAGAGCAAAGTTCTGGTAAAGGCCAACACACAAAGGCCAACTGTCACTAAGAAATCTGCTGCTGAGGCATCTA : 505

SARS_N_F : AAA_GCCTCGCCAAAAC_GCTACTGCCAAAAC_CAGTACAACGTCACCTCA_GCATTTGGGAGACGTC_TCCAGAC--AAOCCAAG--AAATTCGGGAC--AGACCT : 943
SARS_N_FL : AAA_GCCTCGCCAAAAC_GCTACTGCCAAAAC_CAGTACAACGTCACCTCA_GCATTTGGGAGACGTC_TCCAGACAAOCCAAGAAATTCGGGACAAAGACCT : 875
SARS_N_R : -----AAA_GCCTCGCCAAAAC_GCTACTGCCAAAAC_CAGTACAACGTCACCTCA_GCATTTGGGAGACGTC_TCCAGAACAAOCCAAGAAATTCGGGACAAAGACCT : 611

SARS_N_F : A_TcAGAcAG--ACTGAT_ACAA_CATTGGCCGCA--TGCACAAT--GCTC--AGTGC_TCTGCATTCCT--CATGTACGCAT--GCATC--AGTCACACCT : 1030
SARS_N_FL : A_TcAGAcAGAGACTGAT_ACAA_CATTGGCCGCAAT_TGCACAAT_TGCTCAGTGCCTCTGCATTCCTTGGATGTcACGCAT_TGGCATGGAGTCACACCT : 981
SARS_N_R : -----A_TcAGAcAGAGACTGAT_ACAA_CATTGGCCGCAAT_TGCACAAT_TGCTCAGTGCCTCTGCATTCCTTGGATGTcACGCAT_TGGCATGGAGTCACACCT : 717

SARS_N_F : CCG--ACMTc--CTGACTTAcT--GGAGCCATAc--TGGATGAc--GATTC_cGCTTCGAc--CGCTTcTCTGTGAcCTc_cATTGAc_cGGAcAcCTc : 1119
SARS_N_FL : CCGGACATcCTGACTTAcTGGAGCCATAcTGGATGAcGATTCcGCTTCGAcCGCTTcTCTGTGAcAcCGc_cATTGAc_cGGAcAcAcTc : 1086
SARS_N_R : -----cCGGAcATcCTGACTTAcTGGAGCCATAcTGGATGAcGATTCcGCTTCGAcCGCTTcTCTGTGAcAcCGc_cATTGAc_cGGAcAcAcTc : 822

SARS_N_F : WCA----- : 1122
SARS_N_FL : ACAATTCcCAcCAcAGAcGcCTAAAAAGGAcCAAAAAAGAAAAAGAcTcGATGAAGCTcAGCCcTTTcCGcCAGAcGAcAAAAAGAcGcCCAcCTcGTGAcTcTCTcTcCTc : 1192
SARS_N_R : -----ACAATTCcCAcCAcAGAcGcCTAAAAAGGAcCAAAAAAGAAAAAGAcTcGATGAAGCTcAGCCcTTTcCGcCAGAcGAcAAAAAGAcGcCCAcCTcGTGAcTcTCTcTcCTc : 928

SARS_N_F : ----- : 1280
SARS_N_FL : cGGcTcAcATGGATGATTTcTCCAGAcAcACTcCAAAATTCcATGAcTGGAGcTTCcTcGcTcGATTCcAACTcAGGCATAcA----- : 1269
SARS_N_R : -----cGGcTcAcATGGATGATTTcTCCAGAcAcACTcCAAAATTCcATGAcTGGAGcTTCcTcGcTcGATTCcAACTcAGGCATAcAGTTAcCAcAAAAAcTcACTcAGTGAATTcCGc : 1034

```

Appendix 1: Sequence verification of the cloned SARS N-gene in pGEM. Cloned genes were sequenced with M13 forward and reverse primers. M13 primer binding sites are located up- and down-stream of the multiple cloning site allowing for easy sequencing (Data applies to appendix 1-5). The sequencing polymerase is only capable of sequencing ± 800 bp and therefore sequencing of the entire fragment is not possible, which accounts for the large amount of gaps and mismatch at respective ends (Data applies to appendix 1 and 2). SARS_N_F represents the gene strand sequenced with M13 forward primers, where the SARS_N_R represents the gene strand sequenced with M13 reverse primers. SARS_N_FL is the sequence of the SARS N-gene obtained from NCBI with accession number AY360146.1.


```

NL63_N_R : -----*-----20-----*-----40-----*-----60-----*-----80-----*-----100-----: -
NL63_N_FL : -----ATGGCTAGTGTAAATTTGGGCGGATGAC-----: 27
NL63_N_F : TTGWGGGCGAATTTGGGCCCGAGCTGCGATGCTCCCGGCCCCATGGCGGCCGGGAATTCGATTTGGGCGGATCGCCATGGCTAGTGTAAATTTGGGCGGATGAC : 106
                                           atggcctagtgtaaatttgggccgatgac

NL63_N_R : -----*-----120-----*-----140-----*-----160-----*-----180-----*-----200-----*-----: 26
NL63_N_FL : AGAGCTGCTAGGAAGAAATTTCTCTCTCCTTCATTTTACATGCCTCTTTTGGTTA-----TTCTGATAAGGCACCATATAGGGTCATCCAGGAATCTTGTCCCTATTTG : 133
NL63_N_F : AGAGCTGCTAGGAAGAAATTTCTCTCTCCTTCATTTTACATGCCTCTTTTGGTTA-----TTCTGATAAGGCACCATATAGGGTCATCCAGGAATCTTGTCCCTATTTG : 212
                                           agagcctagtgtaagaatctctcctctcattttacatgcctcttttggtta  ttctgataaggccacatattagggTCATCCAGGAATCTTgtCCcTatTTG

NL63_N_R : -----*-----220-----*-----240-----*-----260-----*-----280-----*-----300-----*-----3-----: 99
NL63_N_FL : GTAAAGCGTAATAAAGATTTGAGCCGATTTGCTATGA-----TTTTCAGAGCGTTGGCCTATGCCAGGGGCCACGCTTTGATTTGGCTCTTAAAGTCAATTTG : 234
NL63_N_F : GTAAAGCGTAATAAAGATTTGAGCCGATTTGCTATGA-----TTTTCAGAGCGTTGGCCTATGCCAGGGGCCACGCTTTGATTTGGCTCTTAAAGTCAATTTG : 318
                                           GTAAAGGtaataaaga  TGAGCagattggTtatTggrA  TgTTCaAgAGCGTTGgctTATGcGcAGGGggCraCgtg  TTGATTTgCctCctaaAGTCAATTTT

NL63_N_R : -----*-----340-----*-----360-----*-----380-----*-----400-----*-----420-----*-----: 189
NL63_N_FL : TACTA-----GTAAGCGTCTTAAAGCACTTAAGGACCTTAAATTCAGACAACCGTTC-----GATGGTCTGTTTGGTGGCTAAGACG-----CTTAAACTGTTAATACCATCTTC : 340
NL63_N_F : TACTA-----GTAAGCGTCTTAAAGCACTTAAGGACCTTAAATTCAGACAACCGTTC-----GATGGTCTGTTTGGTGGCTAAGACG-----CTTAAACTGTTAATACCATCTTC : 424
                                           TAtTAcctagTGTgaCctcaTAAGGACccttaAATTCAGACAACGTTCTGATGgtGtTgTTGGGTtGcTAAGGgAaggtGtCTAAACTGTtAAATACCATCTTCG

NL63_N_R : -----*-----440-----*-----460-----*-----480-----*-----500-----*-----520-----*-----: 286
NL63_N_FL : GTA-----TCGCA-----CGTA-----TCAGAACTTTTA-----CAAAGTTTYTATGCTTTGG-----TCCAGAGCTYTCGTGTGTGAGTT-----GAGGATCGCTTAATAACCATCTTCG : 286
NL63_N_F : GTA-----TCGCA-----CGTA-----TCAGAACTTTTA-----CAAAGTTTYTATGCTTTGG-----TCCAGAGCTYTCGTGTGTGAGTT-----GAGGATCGCTTAATAACCATCTTCG : 446
                                           GTAatTCGCAaAcGTAaTCAGAAAcYTTkqgaacCAAGTTCTCTATGCTTTGcctCCAGAGCTCTCTGTGTGAGTTtGAGGATCGCTTAATAACCATCTTCG

NL63_N_R : -----*-----540-----*-----560-----*-----580-----*-----600-----*-----620-----*-----: 391
NL63_N_FL : RGCTAGCAGTCGTTCTCACTCGTAACAACCTACGGRGACITCTTCCTGAGCACTTCAAGACAACAGTCTCGCACTCGT-----TGATTCTAACCGTCTTCTTCAGAT : 391
NL63_N_F : RGCTAGCAGTCGTTCTCACTCGTAACAACCTACGGRGACITCTTCCTGAGCACTTCAAGACAACAGTCTCGCACTCGT-----TGATTCTAACCGTCTTCTTCAGAT : 552
                                           TGCTAGCAGTCGTTCTCACTCGTAACAACCTACGGRGACITCTTCCTGAGCACTTCAAGACAACAGTCTCGCACTCGT-----TGATTCTAACCGTCTTCTTCAGAT : 636
                                           TGCTAGCAGTCGTTCTCACTCGTAACAACCTACGGRGACITCTTCCTGAGCACTTCAAGACAACAGTCTCGCACTCGT-----TGATTCTAACCGTCTTCTTCAGAT

NL63_N_R : -----*-----640-----*-----660-----*-----680-----*-----700-----*-----720-----*-----740-----: 496
NL63_N_FL : CTGTGCTGCTTACTTTGGCTTTAAAGAACTTAGGTTT-----GATAACCACTCGAAGTCACTTAGTCTCTTCTGTAATCCACCTAAGAAACCTAATAAGGCTC : 496
NL63_N_F : CTGTGCTGCTTACTTTGGCTTTAAAGAACTTAGGTTT-----GATAACCACTCGAAGTCACTTAGTCTCTTCTGTAATCCACCTAAGAAACCTAATAAGGCTC : 658
                                           CTGTGCTGCTTACTTTGGCTTTAAAGAACTTAGGTTT-----GATAACCACTCGAAGTCACTTAGTCTCTTCTGTAATCCACCTAAGAAACCTAATAAGGCTC : 742
                                           CTGTGCTGCTTACTTTGGCTTTAAAGAACTTAGGTTT-----GATAACCACTCGAAGTCACTTAGTCTCTTCTGTAATCCACCTAAGAAACCTAATAAGGCTC

NL63_N_R : -----*-----760-----*-----780-----*-----800-----*-----820-----*-----840-----*-----: 601
NL63_N_FL : TTTCTCAACCCAGGGGTGATAGCCCTTCTCAGTTGAAGAA-----CCTCGTTGGAGCGGTTCCTACCAGAGAGGAAATGTTATTCAGTGCCTTTG-----TCCTCGTGATT : 601
NL63_N_F : TTTCTCAACCCAGGGGTGATAGCCCTTCTCAGTTGAAGAA-----CCTCGTTGGAGCGGTTCCTACCAGAGAGGAAATGTTATTCAGTGCCTTTG-----TCCTCGTGATT : 763
                                           TTTCTCAACCCAGGGGTGATAGCCCTTCTCAGTTGAAGAA-----CCTCGTTGGAGCGGTTCCTACCAGAGAGGAAATGTTATTCAGTGCCTTTG-----TCCTCGTGATT : 845
                                           TTTCTCAACCCAGGGGTGATAGCCCTTCTCAGTTGAAGAA-----CCTCGTTGGAGCGGTTCCTACCAGAGAGGAAATGTTATTCAGTGCCTTTG-----TCCTCGTGATT

NL63_N_R : -----*-----860-----*-----880-----*-----900-----*-----920-----*-----940-----*-----: 707
NL63_N_FL : TTAATCACAAATATGGGGATTCAGATCTTGTTCAGAA-----TGGTGTGATGCCAAAGGTTTCCACAGCTTGTGAAATGATTCCTAATCAGCTGGCTTATTCTTTGA : 707
NL63_N_F : TTAATCACAAATATGGGGATTCAGATCTTGTTCAGAA-----TGGTGTGATGCCAAAGGTTTCCACAGCTTGTGAAATGATTCCTAATCAGCTGGCTTATTCTTTGA : 869
                                           TTAATCACAAATATGGGGATTCAGATCTTGTTCAGAA-----TGGTGTGATGCCAAAGGTTTCCACAGCTTGTGAAATGATTCCTAATCAGCTGGCTTATTCTTTGA : 935
                                           TTAATCACAAATATGGGGATTCAGATCTTGTTCAGAA-----TGGTGTGATGCCAAAGGTTTCCACAGCTTGTGAAATGATTCCTAATCAGCTGGCTTATTCTTTGA

NL63_N_R : -----*-----960-----*-----980-----*-----1000-----*-----1020-----*-----1040-----*-----1060-----: 813
NL63_N_FL : TGTGTGGGTAGCACTGATGAGGTGATGATGTTGTCAGATACCTACACACTACAAATGCTTGTAGCTAAGGATATAAGAAACCTTCTAAGTTCATTGAGCAG : 813
NL63_N_F : TGTGTGGGTAGCACTGATGAGGTGATGATGTTGTCAGATACCTACACACTACAAATGCTTGTAGCTAAGGATATAAGAAACCTTCTAAGTTCATTGAGCAG : 975
                                           TGTGTGGGTAGCACTGATGAGGTGATGATGTTGTCAGATACCTACACACTACAAATGCTTGTAGCTAAGGATATAAGAAACCTTCTAAGTTCATTGAGCAG : 1018
                                           TaGTGaGGtTAGCaCTGATGAAGTgATGATAaTGTTCAgatTACcTAcAcTACAaaTgCTTGTgtagCTAAGgataATAAgAacCtCTTAAGTcattgagCag

NL63_N_R : -----*-----1080-----*-----1100-----*-----1120-----*-----1140-----*-----1160-----*-----: 919
NL63_N_FL : ATTAAGCTTTTACTAARCCAGTTTAAAGAAATGCAAGTCACTATCATCTCTCTGTTGTTGAAACACAGTACTTAATGCTCTAATCCAGAAATTAACCAAT : 919
NL63_N_F : ATTAAGCTTTTACTAARCCAGTTTAAAGAAATGCAAGTCACTATCATCTCTCTGTTGTTGAAACACAGTACTTAATGCTCTAATCCAGAAATTAACCAAT : 1081
                                           GACAAAGGTTTACTAA-----CACTGCTCAAG-----TGGCTGATCTCTGKTG-----TCGAAAGG-----ACTAGCTCAGGATC-----ATTCAGG-----AT : 1104
                                           attAgTgctTTTACTAAaccAGTTCATcaAaGaaatgcagtcCaCAATcATCTcAtgTtGctCmGAAAcCagTActTaatgctTcAtTccagAAATcTaaaccAT

NL63_N_R : -----*-----1180-----*-----1200-----*-----1220-----*-----1240-----*-----1260-----*-----: 1024
NL63_N_FL : TGCTGATGAT-----GATTCAAGCCATTATAGAAATTTGCAACGAGGTTTGTCAATTAAGTTTAAACMWGTAAMCWMKAGWAATTGCGGCCCGCTGCAGGTCGACCA : 1024
NL63_N_F : TGCTGATGAT-----GATTCAAGCCATTATAGAAATTTGCAACGAGGTTTGTCAATTAAGTTTAAACMWGTAAMCWMKAGWAATTGCGGCCCGCTGCAGGTCGACCA : 1134
                                           TGCTGATgAT  GAttCagccattatagaaattgtcaacgaggtttgtcattaa : 1122

```

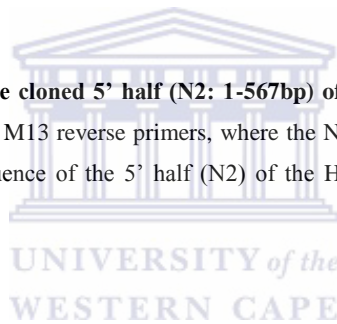
Appendix 2: Sequence verification of the cloned full length HCoV-NL63 N-gene in pGEM®. NL63_N_R represents the gene strand sequenced with M13 reverse primers, where the NL63_N_F represents the gene strand sequenced with M13 forward primers. NL63_N_FL is the sequence of the NL63 N-gene obtained from NCBI with accession number DQ846901.1.

```

120      *      140      *      160      *      180      *      200      *      220
N2_R : TCGCCATGGCTAGTGTAAATGGGCCGATGACAGAGCTGCTAGGAAGAAATTCCTCCTCCTCATTTCACATGCCTCTTTGGTTAATTCTGATAAGGCACCATATAGGG : 222
N2_FL : -----ATGGCTAGTGTAAATGGGCCGATGACAGAGCTGCTAGGAAGAAATTCCTCCTCCTCATTTCACATGCCTCTTTGGTTAATTCTGATAAGGCACCATATAGGG : 106
N2_F : TCGCCATGGCTAGTGTAAATGGGCCGATGACAGAGCTGCTAGGAAGAAATTCCTCCTCCTCATTTCACATGCCTCTTTGGTTAATTCTGATAAGGCACCATATAGGG : 189
      *      240      *      260      *      280      *      300      *      320      *
N2_R : TCATCCAGGAATCTTGTCCCTATTGGTAAGGTAATAAAGATGAGCAGATTGGTTATTGGAATGTTCAAGAGCGTTGGCGTATGCGCAGGGGGCAACGTTGATTTCG : 333
N2_FL : TCATCCAGGAATCTTGTCCCTATTGGTAAGGTAATAAAGATGAGCAGATTGGTTATTGGAATGTTCAAGAGCGTTGGCGTATGCGCAGGGGGCAACGTTGATTTCG : 217
N2_F : TCATCCAGGAATCTTGTCCCTATTGGTAAGGTAATAAAGATGAGCAGATTGGTTATTGGAATGTTCAAGAGCGTTGGCGTATGCGCAGGGGGCAACGTTGATTTCG : 300
      *      340      *      360      *      380      *      400      *      420      *      440
N2_R : CTCCTAAAGTTCATTTTTATTACCTAGGTACTGGACCTCATAAGGACCTTAAATTCAGACAACGTTCTGATGGTGTGTTTGGGTTGCTAAGGAAGTTCGCTAAAAGTCTTA : 444
N2_FL : CTCCTAAAGTTCATTTTTATTACCTAGGTACTGGACCTCATAAGGACCTTAAATTCAGACAACGTTCTGATGGTGTGTTTGGGTTGCTAAGGAAGTTCGCTAAAAGTCTTA : 328
N2_F : CTCCTAAAGTTCATTTTTATTACCTAGGTACTGGACCTCATAAGGACCTTAAATTCAGACAACGTTCTGATGGTGTGTTTGGGTTGCTAAGGAAGTTCGCTAAAAGTCTTA : 411
      *      460      *      480      *      500      *      520      *      540      *
N2_R : ATACCAGTCTTGGTAATCGCAAACGTAATCAGAAACCTTTGGAACCAAAGTTCTCTATTGCTTTGCCCTCCAGAGCTCCTCTGTTGAGTTTGAGGATCGCTCTAATAACT : 555
N2_FL : ATACCAGTCTTGGTAATCGCAAACGTAATCAGAAACCTTTGGAACCAAAGTTCTCTATTGCTTTGCCCTCCAGAGCTCCTCTGTTGAGTTTGAGGATCGCTCTAATAACT : 439
N2_F : ATACCAGTCTTGGTAATCGCAAACGTAATCAGAAACCTTTGGAACCAAAGTTCTCTATTGCTTTGCCCTCCAGAGCTCCTCTGTTGAGTTTGAGGATCGCTCTAATAACT : 522
      *      560      *      580      *      600      *      620      *      640      *      660
N2_R : CATCTCGTGTAGCAGTCGTTCTTCAACTCGTAACAACCTCAGGAGACTCTTCTCGTAGCACTTCAAGACAACAGTCTCGCAGCTGTTCTGATTTCAACCAGTCTTCTTCAG : 666
N2_FL : CATCTCGTGTAGCAGTCGTTCTTCAACTCGTAACAACCTCAGGAGACTCTTCTCGTAGCACTTCAAGACAACAGTCTCGCAGCTGTTCTGATTTCAACCAGTCTTCTTCAG : 550
N2_F : CATCTCGTGTAGCAGTCGTTCTTCAACTCGTAACAACCTCAGGAGACTCTTCTCGTAGCACTTCAAGACAACAGTCTCGCAGCTGTTCTGATTTCAACCAGTCTTCTTCAG : 633
      *      680      *      700      *      720      *      740      *      760      *
N2_R : ATCTTGTGTGCTGCTGTTAAAGTTTAAACACAAAATCACTAGTGAATTCGGGGCCGCTGCAGGTGCAACCATATGGGAGAGCTCCCAACCGGTTGGATGCATAGCTTGAGTA : 777
N2_FL : ATCTTGTGTGCTGCTGTTAAAGTTTAAACACAAAATCACTAGTGAATTCGGGGCCGCTGCAGGTGCAACCATATGGGAGAGCTCCCAACCGGTTGGATGCATAGCTTGAGTA : 570
N2_F : ATCTTGTGTGCTGCTGTTAAAGTTTAAACACAAAATCACTAGTGAATTCGGGGCCGCTGCAGGTGCAACCATATGGGAGAGCTCCCAACCGGTTGGATGCATAGCTTGAGTA : 744
      ATCTTGTGTGCTGCTGTTAAAGTTTAAACACAAAATCACTAGTGAATTCGGGGCCGCTGCAGGTGCAACCATATGGGAGAGCTCCCAACCGGTTGGATGCATAGCTTGAGTA

```

Appendix 3: Sequence verification of the cloned 5' half (N2: 1-567bp) of the HCoV-NL63 N-gene in pGEM@. N2_R represents the gene strand sequenced with M13 reverse primers, where the N2_F represents the gene strand sequenced with M13 forward primers. N2_FL is the sequence of the 5' half (N2) of the HCoV-NL63 N-gene obtained from NCBI with accession number DQ846901.1.




```

120      *      140      *      160      *      180      *      200      *      220
N3_R : GCGATCGCCATGACTTTGGCTTAAAGAAGCTTAGGTTTTGATAACCAAGTCGAAGTCACCTAGTCTCTTGGTACTTCCACTCCTAAGAACCCTAATAAGCCTCTTCTCAA : 221
N3_FL : -----ATGACTTTGGCTTAAAGAAGCTTAGGTTTTGATAACCAAGTCGAAGTCACCTAGTCTCTTGGTACTTCCACTCCTAAGAACCCTAATAAGCCTCTTCTCAA : 102
N3_F : GCGATCGCCATGACTTTGGCTTAAAGAAGCTTAGGTTTTGATAACCAAGTCGAAGTCACCTAGTCTCTTGGTACTTCCACTCCTAAGAACCCTAATAAGCCTCTTCTCAA : 183
      gcgatcgccatgactttggctttaaagaagcttaggttttgataaccagtcgaagtcacctagttctcttggtaacttccactcctaagaaccctaataagcctcttctcaa

      *      240      *      260      *      280      *      300      *      320      *
N3_R : CCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGTGTCTCTACCAGAGAGGAAAATGTTATTCAGTGCTTTGGTCTCGTGATTTAATCAAAATAG : 332
N3_FL : CCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGTGTCTCTACCAGAGAGGAAAATGTTATTCAGTGCTTTGGTCTCGTGATTTAATCAAAATAG : 213
N3_F : CCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGTGTCTCTACCAGAGAGGAAAATGTTATTCAGTGCTTTGGTCTCGTGATTTAATCAAAATAG : 294
      CCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGTGTCTCTACCAGAGAGGAAAATGTTATTCAGTGCTTTGGTCTCGTGATTTAATCAAAATAG

      *      340      *      360      *      380      *      400      *      420      *      440
N3_R : GGGGATTCAGACTTGTTCAGAAATGGTGTTCATGCCAAAGGTTTTCCACAGCTTGTCTGAATTGATTCCTAATCAGGCTGCGTTATTCCTTGAAGTGAAGTTAGCACTGAT : 443
N3_FL : GGGGATTCAGACTTGTTCAGAAATGGTGTTCATGCCAAAGGTTTTCCACAGCTTGTCTGAATTGATTCCTAATCAGGCTGCGTTATTCCTTGAAGTGAAGTTAGCACTGAT : 324
N3_F : GGGGATTCAGACTTGTTCAGAAATGGTGTTCATGCCAAAGGTTTTCCACAGCTTGTCTGAATTGATTCCTAATCAGGCTGCGTTATTCCTTGAAGTGAAGTTAGCACTGAT : 405
      GGGGATTCAGACTTGTTCAGAAATGGTGTTCATGCCAAAGGTTTTCCACAGCTTGTCTGAATTGATTCCTAATCAGGCTGCGTTATTCCTTGAAGTGAAGTTAGCACTGAT

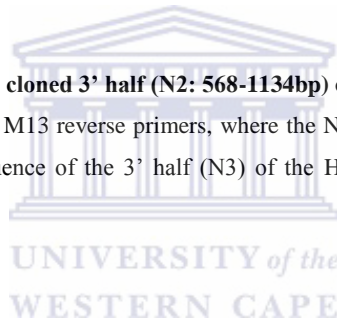
      *      460      *      480      *      500      *      520      *      540      *
N3_R : GAAGTGGTGATAAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGTCAAGGATAATAAGAACCTTCTTAAGTTCATTGAGCAGATTAGTGCTTTTACTAAACCCAGT : 554
N3_FL : GAAGTGGTGATAAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGTCAAGGATAATAAGAACCTTCTTAAGTTCATTGAGCAGATTAGTGCTTTTACTAAACCCAGT : 435
N3_F : GAAGTGGTGATAAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGTCAAGGATAATAAGAACCTTCTTAAGTTCATTGAGCAGATTAGTGCTTTTACTAAACCCAGT : 516
      GAAGTGGTGATAAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGTCAAGGATAATAAGAACCTTCTTAAGTTCATTGAGCAGATTAGTGCTTTTACTAAACCCAGT

      *      560      *      580      *      600      *      620      *      640      *      660
N3_R : TCTATCAAGAAGAAATGCAGTCACAATCATCTCATGTTGCTCAGAACACAGTACTTAATGCTTCTATTCCAGAACTAAACCATTGGCTGATGATGATTCAGCCATTATAGAA : 665
N3_FL : TCTATCAAGAAGAAATGCAGTCACAATCATCTCATGTTGCTCAGAACACAGTACTTAATGCTTCTATTCCAGAACTAAACCATTGGCTGATGATGATTCAGCCATTATAGAA : 546
N3_F : TCTATCAAGAAGAAATGCAGTCACAATCATCTCATGTTGCTCAGAACACAGTACTTAATGCTTCTATTCCAGAACTAAACCATTGGCTGATGATGATTCAGCCATTATAGAA : 627
      TCTATCAAGAAGAAATGCAGTCACAATCATCTCATGTTGCTCAGAACACAGTACTTAATGCTTCTATTCCAGAACTAAACCATTGGCTGATGATGATTCAGCCATTATAGAA

      *      680      *      700      *      720      *      740      *      760      *
N3_R : ATTGTCACACGAGGTTTTCGATTAAGTTTAAACCTGTAATCACTAGTGAATTCGCGGCCGCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTG : 776
N3_FL : ATTGTCACACGAGGTTTTCGATTAAGTTTAAACCTGTAATCACTAGTGAATTCGCGGCCGCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTG : 570
N3_F : ATTGTCACACGAGGTTTTCGATTAAGTTTAAACCTGTAATCACTAGTGAATTCGCGGCCGCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTG : 738
      ATTGTCACACGAGGTTTTCGATTAAGTTTAAACCTGTAATCACTAGTGAATTCGCGGCCGCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTG

```

Appendix 4: Sequence verification of the cloned 3' half (N2: 568-1134bp) of the HCoV-NL63 N-gene in pGEM®. N3_R represents the gene strand sequenced with M13 reverse primers, where the N3_F represents the gene strand sequenced with M13 forward primers. N3_FL is the sequence of the 3' half (N3) of the HCoV-NL63 N-gene obtained from NCBI with accession number DQ846901.1.



```

120      *      140      *      160      *      180      *      200      *      220
N4_R : GCGATCGCCATGTTCTCTATTGCTTTGCCGCCAGAGCTCTCTGTTGTTGAGTTTGAAGTCCGCTCTAATAACATCATCTCGTGTAGCAGTCGTTCTTCAACTCGTAAACAAC : 222
N4_FL : -----ATGTTCTCTATTGCTTTGCCGCCAGAGCTCTCTGTTGTTGAGTTTGAAGTCCGCTCTAATAACATCATCTCGTGTAGCAGTCGTTCTTCAACTCGTAAACAAC : 102
N4_F : GCGATCGCCATGTTCTCTATTGCTTTGCCGCCAGAGCTCTCTGTTGTTGAGTTTGAAGTCCGCTCTAATAACATCATCTCGTGTAGCAGTCGTTCTTCAACTCGTAAACAAC : 176
      gcgatcgccatgttctctattgctttgccgccagagctctctgttgttgagtttgaagtccgctctaataaacatcatctcgtgtagcagtcgttcttcaactcgtaacaac

      *      240      *      260      *      280      *      300      *      320      *
N4_R : TCACGAGACTCTTCGCTAGCACTTCAAGACAACAGTCTCGCACCTGTTCTGATCTCAACCACTCTTTCAGATCTTGTGTGCTGTTACTTTGGCTTAAAGAAGCTTA : 333
N4_FL : TCACGAGACTCTTCGCTAGCACTTCAAGACAACAGTCTCGCACCTGTTCTGATCTCAACCACTCTTTCAGATCTTGTGTGCTGTTACTTTGGCTTAAAGAAGCTTA : 213
N4_F : TCACGAGACTCTTCGCTAGCACTTCAAGACAACAGTCTCGCACCTGTTCTGATCTCAACCACTCTTTCAGATCTTGTGTGCTGTTACTTTGGCTTAAAGAAGCTTA : 287
      TCACGAGACTCTTCGCTAGCACTTCAAGACAACAGTCTCGCACCTGTTCTGATCTCAACCACTCTTTCAGATCTTGTGTGCTGTTACTTTGGCTTAAAGAAGCTTA

      *      340      *      360      *      380      *      400      *      420      *      440
N4_R : GGTTTTGATAAACCAGTGAAGTCACCTAGTCTTCTGGTACTTCCACTCCTAAGAACCCTAATAAGCCTCTTCTCAACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAA : 444
N4_FL : GGTTTTGATAAACCAGTGAAGTCACCTAGTCTTCTGGTACTTCCACTCCTAAGAACCCTAATAAGCCTCTTCTCAACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAA : 324
N4_F : GGTTTTGATAAACCAGTGAAGTCACCTAGTCTTCTGGTACTTCCACTCCTAAGAACCCTAATAAGCCTCTTCTCAACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAA : 398
      GGTTTTGATAAACCAGTGAAGTCACCTAGTCTTCTGGTACTTCCACTCCTAAGAACCCTAATAAGCCTCTTCTCAACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAA

      *      460      *      480      *      500      *      520      *      540      *
N4_R : CCTCGTTGGAAGCGTGTCTCTACCAGAGAGGAAAATGTTATTCAGTGCTTTGGTCTTAAAGTTTAAACCGACAATCACTAGTGAATTCGCGGCCGCTGCAGGTCGACCAT : 555
N4_FL : CCTCGTTGGAAGCGTGTCTCTACCAGAGAGGAAAATGTTATTCAGTGCTTTGGTCTTAAAGTTTAAACCGACAATCACTAGTGAATTCGCGGCCGCTGCAGGTCGACCAT : 384
N4_F : CCTCGTTGGAAGCGTGTCTCTACCAGAGAGGAAAATGTTATTCAGTGCTTTGGTCTTAAAGTTTAAACCGACAATCACTAGTGAATTCGCGGCCGCTGCAGGTCGACCAT : 509
      CCTCGTTGGAAGCGTGTCTCTACCAGAGAGGAAAATGTTATTCAGTGCTTTGGTCTTAAAGTTTAAACCGACAATCACTAGTGAATTCGCGGCCGCTGCAGGTCGACCAT

```

Appendix 5: Sequence verification of the cloned middle region (N4: 379-756bp) of the HCoV-NL63 N-gene in pGEM®. N4_R represents the gene strand sequenced with M13 reverse primers, where the N4_F represents the gene strand sequenced with M13 forward primers. N4_FL is the sequence of the middle region (N4) of the HCoV-NL63 N-gene obtained from NCBI with accession number DQ846901.1.

```

SARS      : ----- : -
SARS_FL   : ATGTCCTGATAATGGACCCCAATCAAACCAACGTAGTGCCCCCGCATTACATTGGTGGACCCACAGATTCAACTGACAATAACCGAATGGAGGACGCCAATGGGGCA : 108

          *      20      *      40      *      60      *      80      *      100

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

SARS      : CCAATTTCCAGMCCAATAGTGGTCCARATGACCA--TTGGTACTACCGAAGAGTACC--GACGAGTTCGTGGTGGTACGGCAAATGAAAGAGAGCTCAGCCCCGA : 121
SARS_FL   : GTTCCAAATCAACCCAATAGTGGTCCAGATGACCAATTTGGTACTACCGAAGAGTACC--GACGAGTTCGTGGTGGTACGGCAAATGAAAGAGAGCTCAGCCCCGA : 324
          CA C CCAATAGTGGTCCAGATGACCA TTGGTACTACCGAAGAGTACC GACGAGTTCGTGGTGGTACGGCAA AA AGTTCAGCCCCGA

          *      220     *      240     *      260     *      280     *      300     *      320

SARS      : TGGTACTTCTTATCTAGGAACCTGGCC--AGAAGCTTCACCTTCCCTACGGCGCTAACAAGAAGGCATCGTATGGGTTGCAACTGAGGAGCCTTGAATACACCCAA : 228
SARS_FL   : TGGTACTTCTTATCTAGGAACCTGGCC--AGAAGCTTCACCTTCCCTACGGCGCTAACAAGAAGGCATCGTATGGGTTGCAACTGAGGAGCCTTGAATACACCCAA : 431
          TGGTACTTCT T CTAGGAACCTGGCC AGAAGCTTCACCTT CCCTACGGCGCTA AAGAAGGCATCGTATGGGTTGCAACTGAGGAGCCTTGAATACACCCAA

          *      340     *      360     *      380     *      400     *      420     *

SARS      : AGACCACATTTGGCACCCCAATCCTAATAACAATGTGCCACCGTCTACAACCTCCTCAAGGAACAGCATTGGCCAAAAGGCTTCTACGCAGAGGGAGCAGAGGCGG : 336
SARS_FL   : AGACCACATTTGGCACCCCAATCCTAATAACAATGTGCCACCGTCTACAACCTCCTCAAGGAACAGCATTGGCCAAAAGGCTTCTACGCAGAGGGAGCAGAGGCGG : 539
          AGACCACATTTGGCACCCCAATCCTAATAACAATGTGCCACCGTCTACAACCTCCTCAAGGAACA CATTGGCCAAAAGGCTTCTACGCAGAGGGAGCAGAGGCGG

          *      440     *      460     *      480     *      500     *      520     *      540

SARS      : CAGTCAAGCCTCTTCTCGCTCCTCAATCAGCTAGTCGCGGTAATTCAGAATAATCAACTCCTGGCAGCAGTAGGGGAAATTCCTCGTCTGCAATGGCGAGGAGTGG : 444
SARS_FL   : CAGTCAAGCCTCTTCTCGCTCCTCAATCAGCTAGTCGCGGTAATTCAGAATAATCAACTCCTGGCAGCAGTAGGGGAAATTCCTCGTCTGCAATGGCGAGGAGTGG : 647
          CAGTCAAGCCTCTTCTCGCTCCTCAATCAGCTAGTCGCGGTAATTCAGAATAATCAACTCCTGGCAGCAGTAGGGGAAATTCCTCGTCTGCAATGGCGAGGAGTGG

          *      560     *      580     *      600     *      620     *      640

SARS      : TGAAACTGCCCTCGCGCTATTGCTGTAGACAGATTGAACCCAGCTTGAGAGCAAAAGTTTCTGGTAAAGGCCAACACAACAAAGGCCAACTGTCACTAAGAAATCTGC : 552
SARS_FL   : TGAAACTGCCCTCGCGCTATTGCTGTAGACAGATTGAACCCAGCTTGAGAGCAAAAGTTTCTGGTAAAGGCCAACACAACAAAGGCCAACTGTCACTAAGAAATCTGC : 755
          TGAAACTGCCCTCGCGCTATTGCTGTAGACAGATTGAACCCAGCTTGAGAGCAAAAGTTTCTGGTAAAGGCCAACACAACAAAGGCCAACTGTCACTAAGAAATCTGC

          *      660     *      680     *      700     *      720     *      740     *

SARS      : TGCTGAGGCATCTAAAAAGCCTCGCCAAAACGTAAGTCCCAAAAACAGTACAACTCAAGCTCACTCAAGCATTTGGGAGACGTGGTCCAGACAACAAAGGAAATTTCCG : 660
SARS_FL   : TGCTGAGGCATCTAAAAAGCCTCGCCAAAACGTAAGTCCCAAAAACAGTACAACTCAAGCTCACTCAAGCATTTGGGAGACGTGGTCCAGACAACAAAGGAAATTTCCG : 863
          TGCTGAGGCATCTAAAAAGCCTCGCCAAAACGTAAGTCCCAAAAACAGTACAACTCAAGCTCACTCAAGCATTTGGGAGACGTGGTCCAGACAACAAAGGAAATTTCCG

          *      760     *      780     *      800     *      820     *      840     *      860

SARS      : GGACCAAGACCTAATCAGACAAGGAAGTATTACAAAACATTTGGCCGCAAAATTTGCACAAATTTGCTCCAAAGTGCCCTCTGCATTTCTTGAATGTACAGCATTTGGCATGGA : 768
SARS_FL   : GGACCAAGACCTAATCAGACAAGGAAGTATTACAAAACATTTGGCCGCAAAATTTGCACAAATTTGCTCCAAAGTGCCCTCTGCATTTCTTGAATGTACAGCATTTGGCATGGA : 971
          GGACCAAGACCTAATCAGACAAGGAAGTATTACAAAACATTTGGCCGCAAAATTTGCACAAATTTGCTCCAAAGTGCCCTCTGCATTTCTTGAATGTACAGCATTTGGCATGGA

          *      880     *      900     *      920     *      940     *      960     *

SARS      : AGTCACACCTTCGGGAACATGGCTGACTTATCATGGAGCCATTAATTTGGATGACAAGATCCCAATTTCAAAGACAACGTCATACCTGCTGAACAAGCACATTTGACCG : 876
SARS_FL   : AGTCACACCTTCGGGAACATGGCTGACTTATCATGGAGCCATTAATTTGGATGACAAGATCCCAATTTCAAAGACAACGTCATACCTGCTGAACAAGCACATTTGACCG : 1079
          AGTCACACCTTCGGGAACATGGCTGACTTATCATGGAGCCATTAATTTGGATGACAAGATCCCAATTTCAAAGACAACGTCATACCTGCTGAACAAGCACATTTGACCG

          *      1000    *      1020    *      1040    *      1060    *      1080

SARS      : ATACAAAACATTTCCACCAACAGAGCCTAAAAAGGACAAAAGAAAAGAAAGACTGATGAAGCTCAGCCCTTGGCCGAGAGACAAAAGAAAGCAGCCACTGTGACTCTTCT : 984
SARS_FL   : ATACAAAACATTTCCACCAACAGAGCCTAAAAAGGACAAAAGAAAAGAAAGACTGATGAAGCTCAGCCCTTGGCCGAGAGACAAAAGAAAGCAGCCACTGTGACTCTTCT : 1187
          ATACAAAACATTTCCACCAACAGAGCCTAAAAAGGACAAAAGAAAAGAAAGACTGATGAAGCTCAGCCCTTGGCCGAGAGACAAAAGAAAGCAGCCACTGTGACTCTTCT

          *      1100    *      1120    *      1140    *      1160    *      1180

SARS      : TCCTGCGGCTGACATGGATGATTTCCAGACAACCTCAAATTTCCATGAGTGGAGCTTCTGCTGATTCAACTCAGGCATAAGTTTAAACGAATTCGGGCTCGGTACC : 1092
SARS_FL   : TCCTGCGGCTGACATGGATGATTTCCAGACAACCTCAAATTTCCATGAGTGGAGCTTCTGCTGATTCAACTCAGGCATAA----- : 1269
          TCCTGCGGCTGACATGGATGATTTCCAGACAACCTCAAATTTCCATGAGTGGAGCTTCTGCTGATTCAACTCAGGCATAA

```

Appendix 6: 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.

```

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

N1 : 120-----*-----140-----*-----160-----*-----180-----*-----200-----*-----220
N1_FL : ATCCC---AGATCTGTCC---TATTGGTAAGGTAATAAAGATGAGCAGATTGGTTATTGGAATGTTCAAGAGCGTTGGCGTATGCCAGGGGGCAACGTGTGATTGGC : 145
N1_FL : ATGCCAGGTAATCTGTCCGATTGGTAAGGTAATAAAGATGAGCAGATTGGTTATTGGAATGTTCAAGAGCGTTGGCGTATGCCAGGGGGCAACGTGTGATTGGC : 218
AT CC A TGCC TATTGGTAAGGTAATAAAGATGAGCAGATTGGTTATTGGAATGTTCAAGAGCGTTGGCGTATGCCAGGGGGCAACGTGTGATTGGC

N1 : 240-----*-----260-----*-----280-----*-----300-----*-----320-----*
N1_FL : TCCTAAAGTTCATTTTATTACCTAGTACTGGACCTCATAAGGACCTTAAATTCAGACAACGTTCTGATGGTGTGTTGGGTTGCTAAGGAAGGTCTAAAACGTTA : 255
N1_FL : TCCTAAAGTTCATTTTATTACCTAGTACTGGACCTCATAAGGACCTTAAATTCAGACAACGTTCTGATGGTGTGTTGGGTTGCTAAGGAAGGTCTAAAACGTTA : 328
TCCTAAAGTTCATTTTATTACCTAGTACTGGACCTCATAAGGACCTTAAATTCAGACAACGTTCTGATGGTGTGTTGGGTTGCTAAGGAAGGTCTAAAACGTTA

N1 : 340-----*-----360-----*-----380-----*-----400-----*-----420-----*-----440
N1_FL : ATACCAGTCTTGGTAATCGCAACGTAATCAGAAACCTTTGGAACCAAAGTTCTCTATTGCTTTGCCCTCCAGAGCTCCTCTGTTCTGATGAGTTTGGATCCGCTTAATAAC : 365
N1_FL : ATACCAGTCTTGGTAATCGCAACGTAATCAGAAACCTTTGGAACCAAAGTTCTCTATTGCTTTGCCCTCCAGAGCTCCTCTGTTCTGATGAGTTTGGATCCGCTTAATAAC : 438
ATACCAGTCTTGGTAATCGCAACGTAATCAGAAACCTTTGGAACCAAAGTTCTCTATTGCTTTGCCCTCCAGAGCTCCTCTGTTCTGATGAGTTTGGATCCGCTTAATAAC

N1 : 460-----*-----480-----*-----500-----*-----520-----*-----540-----*
N1_FL : TCATCTGCTAGCAGTCTTCTCAACTCGTAACAACTCACGAGACTCTTCTCGTAGCATTCAAGACAACAGTCTCGCACTCGTCTGATTCTAACAGTCTTCTTC : 475
N1_FL : TCATCTGCTAGCAGTCTTCTCAACTCGTAACAACTCACGAGACTCTTCTCGTAGCATTCAAGACAACAGTCTCGCACTCGTCTGATTCTAACAGTCTTCTTC : 548
TCATCTGCTAGCAGTCTTCTCAACTCGTAACAACTCACGAGACTCTTCTCGTAGCATTCAAGACAACAGTCTCGCACTCGTCTGATTCTAACAGTCTTCTTC

N1 : 560-----*-----580-----*-----600-----*-----620-----*-----640-----*-----660
N1_FL : AGATCTTGTGCTGCTTACTTTGGCTTTAAAGAACTTAGGTTTGGATAACCAGTCGAAGTCACCTAGTTCCTCTGTTACTCCACTCCTAAGAAACCTAATAAGCCTC : 585
N1_FL : AGATCTTGTGCTGCTTACTTTGGCTTTAAAGAACTTAGGTTTGGATAACCAGTCGAAGTCACCTAGTTCCTCTGTTACTCCACTCCTAAGAAACCTAATAAGCCTC : 658
AGATCTTGTGCTGCTTACTTTGGCTTTAAAGAACTTAGGTTTGGATAACCAGTCGAAGTCACCTAGTTCCTCTGTTACTCCACTCCTAAGAAACCTAATAAGCCTC

N1 : 680-----*-----700-----*-----720-----*-----740-----*-----760-----*
N1_FL : TTTTCACAAACCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGTGTCTTACCAGAGAGGAAAATGTTATTCAGTCTTTGGTCCCTCGTATTTTAAAT : 695
N1_FL : TTTTCACAAACCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGTGTCTTACCAGAGAGGAAAATGTTATTCAGTCTTTGGTCCCTCGTATTTTAAAT : 768
TTTTCACAAACCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGTGTCTTACCAGAGAGGAAAATGTTATTCAGTCTTTGGTCCCTCGTATTTTAAAT

N1 : 780-----*-----800-----*-----820-----*-----840-----*-----860-----*-----880
N1_FL : CACAATATGGGGGATTCAGATCTTGTTCAGAATGGTGTGATGCCAAAGGTTTCCACAGCTTGCTGAATTGATTCCTAATCAGGCTCGCTTATCTTTGATAGTGAGGT : 805
N1_FL : CACAATATGGGGGATTCAGATCTTGTTCAGAATGGTGTGATGCCAAAGGTTTCCACAGCTTGCTGAATTGATTCCTAATCAGGCTCGCTTATCTTTGATAGTGAGGT : 878
CACAATATGGGGGATTCAGATCTTGTTCAGAATGGTGTGATGCCAAAGGTTTCCACAGCTTGCTGAATTGATTCCTAATCAGGCTCGCTTATCTTTGATAGTGAGGT

N1 : 900-----*-----920-----*-----940-----*-----960-----*-----980-----*
N1_FL : TAGCACTGATGAAGTGTGATAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGATAATAAGAACCTTCTAAGTTCATTGAGCAGATTAGTGCCTTTA : 915
N1_FL : TAGCACTGATGAAGTGTGATAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGATAATAAGAACCTTCTAAGTTCATTGAGCAGATTAGTGCCTTTA : 988
TAGCACTGATGAAGTGTGATAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGATAATAAGAACCTTCTAAGTTCATTGAGCAGATTAGTGCCTTTA

N1 : 1000-----*-----1020-----*-----1040-----*-----1060-----*-----1080-----*-----1100
N1_FL : CTAACCCAGTTCATCAAAGAAATGCAGTCACAATCATCTCATGTTGCCACAGAACAGTACTTAATGCTTCTATTCCAGAACTAAACCAATTGGCTGATGATGATCA : 1025
N1_FL : CTAACCCAGTTCATCAAAGAAATGCAGTCACAATCATCTCATGTTGCCACAGAACAGTACTTAATGCTTCTATTCCAGAACTAAACCAATTGGCTGATGATGATCA : 1098
CTAACCCAGTTCATCAAAGAAATGCAGTCACAATCATCTCATGTTGCCACAGAACAGTACTTAATGCTTCTATTCCAGAACTAAACCAATTGGCTGATGATGATCA

N1 : 1120-----*-----1140-----*-----1160-----*-----1180-----*-----1200-----*
N1_FL : GCCATTATAGAAATGTCACACGAGGTTTGCATTAAATTTAAACGAATCGGGCTCGGTACCGGGGATCCTCTAGAGTCGACCTGCAGGCTGCACAGCTGATCCCKCTG : 1135
N1_FL : GCCATTATAGAAATGTCACACGAGGTTTGCATTAAATTTAAACGAATCGGGCTCGGTACCGGGGATCCTCTAGAGTCGACCTGCAGGCTGCACAGCTGATCCCKCTG : 1134
GCCATTATAGAAATGTCACACGAGGTTTGCATTAA

```

WESTERN CAPE

Appendix 7: Sequence verification of the HCoV-NL63 full length N-gene. NL63_FL represents the sequence for the HCoV-NL63 N-gene obtained from NCBI with accession number DQ846901.1, where N1 represents the sequence of the cloned HCoV-NL63 N-gene.

```

      *      460      *      480      *      500      *      520      *      540      *
N2   : ATTCACCATGAAACATCAACATCAGCACCAAGCGATCGCCATGGCTAGTGTAATTTGGCCGATGACAGAGCTGCTAGGAAGAAATTCCTCCCTTCATTTTACATGC : 550
N2_FL : -----ATGGCTAGTGTAATTTGGCCGATGACAGAGCTGCTAGGAAGAAATTCCTCCCTTCATTTTACATGC : 70
      ATGGCTAGTGTAATTTGGCCGATGACAGAGCTGCTAGGAAGAAATTCCTCCCTTCATTTTACATGC

      560      *      580      *      600      *      620      *      640      *      660
N2   : CTCCTTTGGTTATTTCTGATAAGGCACCATAFAGGGTCATTCOCAGGAATCTTGTCOCCTATTGGTAAGGGTAATAAAGATGAGCAGATTGGTTATTGGAATGTTCAAGAG : 660
N2_FL : CTCCTTTGGTTATTTCTGATAAGGCACCATAFAGGGTCATTCOCAGGAATCTTGTCOCCTATTGGTAAGGGTAATAAAGATGAGCAGATTGGTTATTGGAATGTTCAAGAG : 180
      CTCCTTTGGTTA TTTCTGATAAGGCACCATAFAGGGTCATTCOCAGGAATCTTGTCOCCTATTGGTAAGGGTAATAAAGATGAGCAGATTGGTTATTGGAATGTTCAAGAG

      *      680      *      700      *      720      *      740      *      760      *
N2   : CGTTGGCGTATGCCGAGGGGCAACGTTGATTGGCTCCTAAAGTTCATTTTTATTACCTAGGTACTGGACCTATAAGGACCTTAATTCAGACAACTTCTGATGG : 770
N2_FL : CGTTGGCGTATGCCGAGGGGCAACGTTGATTGGCTCCTAAAGTTCATTTTTATTACCTAGGTACTGGACCTATAAGGACCTTAATTCAGACAACTTCTGATGG : 290
      CGTTGGCGTATGCCGAGGGGCAACGTTGATTGGCTCCTAAAGTTCATTTTTATTACCTAGGTACTGGACCTATAAGGACCTTAATTCAGACAACTTCTGATGG

      780      *      800      *      820      *      840      *      860      *      880
N2   : TGTGTTTGGGTTGCTAAGGAAGGTGCTAAAACGTGTAATACCAGTCTTGGTAATCGCAAACGTAATCAGAAACCTTTGGAACCAAAGTTCCTATTGCTTTGCCCTCCAG : 880
N2_FL : TGTGTTTGGGTTGCTAAGGAAGGTGCTAAAACGTGTAATACCAGTCTTGGTAATCGCAAACGTAATCAGAAACCTTTGGAACCAAAGTTCCTATTGCTTTGCCCTCCAG : 400
      TGTGTTTGGGTTGCTAAGGAAGGTGCTAAAACGTGTAATACCAGTCTTGGTAATCGCAAACGTAATCAGAAACCTTTGGAACCAAAGTTCCTATTGCTTTGCCCTCCAG

      *      900      *      920      *      940      *      960      *      980      *
N2   : AGCTCTCTGTTGTTGAGTTTGAGGATCGCTCTAATAACTCATCTCGTGTAGCAGTCTTCAACTCGTAACAACCTCAGGACTCTTCTCGTAGCACTTCAAGACAA : 990
N2_FL : AGCTCTCTGTTGTTGAGTTTGAGGATCGCTCTAATAACTCATCTCGTGTAGCAGTCTTCAACTCGTAACAACCTCAGGACTCTTCTCGTAGCACTTCAAGACAA : 510
      AGCTCTCTGTTGTTGAGTTTGAGGATCGCTCTAATAACTCATCTCGTGTAGCAGTCTTCAACTCGTAACAACCTCAGGACTCTTCTCGTAGCACTTCAAGACAA

      1000      *      1020      *      1040      *      1060      *      1080      *      1100
N2   : CAGTCTCGCACCTGTTCTGATTCCTAACCAGTCTTCTTCAGATCTTGTGCTGCTGTTTAAAGTTTAAACGAAATTCGGGCTCGGTACCCGGGGATCCCTTAGAGTCGACCTG : 1100
N2_FL : CAGTCTCGCACCTGTTCTGATTCCTAACCAGTCTTCTTCAGATCTTGTGCTGCTGTTTAA----- : 570
      CAGTCTCGCACCTGTTCTGATTCCTAACCAGTCTTCTTCAGATCTTGTGCTGCTGTTTAA

```

Appendix 8: Sequence verification of the 5' (1-567bp) half of the HCoV-NL63 N-gene coding for the N-terminal half (N2) of the HCoV-NL63 N-protein. N2_FL represents the sequence for the 5' half of the HCoV-NL63 N-gene obtained from NCBI with accession number DQ846901.1, where N2 represents the sequence of the cloned N2-gene.

```

340      *      360      *      380      *      400      *      420      *      440
N3  : GCGATCGCCATGACTTTGGCTTTAAGAAGCTTAGGTTTGGATAACCAAGTCGAAGTACACCTAGTCTTCTGGTACTTCCACTCCTAAGAAACCTAATAAGCCTCTTCTCA : 440
N3_FL : -----ATGACTTTGGCTTTAAGAAGCTTAGGTTTGGATAACCAAGTCGAAGTACACCTAGTCTTCTGGTACTTCCACTCCTAAGAAACCTAATAAGCCTCTTCTCA : 101
          ATGACTTTGGCTTTAAGAAGCTTAGGTTTGGATAACCAAGTCGAAGTACACCTAGTCTTCTGGTACTTCCACTCCTAAGAAACCTAATAAGCCTCTTCTCA

*      460      *      480      *      500      *      520      *      540      *
N3  : ACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAAGCGTGTCTCCAGAGAGGAAAATGTTATTCAGTGCCTTGGTCCCTCGTATTTAATCACAATA : 550
N3_FL : ACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAAGCGTGTCTCCAGAGAGGAAAATGTTATTCAGTGCCTTGGTCCCTCGTATTTAATCACAATA : 211
          ACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAAGCGTGTCTCCAGAGAGGAAAATGTTATTCAGTGCCTTGGTCCCTCGTATTTAATCACAATA

560      *      580      *      600      *      620      *      640      *      660
N3  : TGGGGATTTCAGATCTGTTCAGAAATGGTGTGATGCCAAAGTTTTCCACAGCTTGCCTGAATTGATTCCTAATCAGGCTCGCTTATCTTTGATAGTGAGGTTAGCACT : 660
N3_FL : TGGGGATTTCAGATCTGTTCAGAAATGGTGTGATGCCAAAGTTTTCCACAGCTTGCCTGAATTGATTCCTAATCAGGCTCGCTTATCTTTGATAGTGAGGTTAGCACT : 321
          TGGGGATTTCAGATCTGTTCAGAAATGGTGTGATGCCAAAGTTTTCCACAGCTTGCCTGAATTGATTCCTAATCAGGCTCGCTTATCTTTGATAGTGAGGTTAGCACT

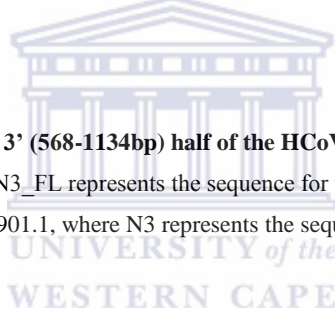
*      680      *      700      *      720      *      740      *      760      *
N3  : GATGAAGTGGTGATAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGATAATAAGAAACCTTCTAAGTTCATTGAGCAGATTAGTGCTTTTACTAAACC : 770
N3_FL : GATGAAGTGGTGATAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGATAATAAGAAACCTTCTAAGTTCATTGAGCAGATTAGTGCTTTTACTAAACC : 431
          GATGAAGTGGTGATAATGTTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGATAATAAGAAACCTTCTAAGTTCATTGAGCAGATTAGTGCTTTTACTAAACC

780      *      800      *      820      *      840      *      860      *      880
N3  : CAGTTCATCAAAGAAATGCAGTCAAAATCATCTCATGTTGCTCAGAACAAGTACTTAATGCTTCTATTCAGAAATCAAACCATTGGCTGATGATGATTCAGCCATTA : 880
N3_FL : CAGTTCATCAAAGAAATGCAGTCAAAATCATCTCATGTTGCTCAGAACAAGTACTTAATGCTTCTATTCAGAAATCAAACCATTGGCTGATGATGATTCAGCCATTA : 541
          CAGTTCATCAAAGAAATGCAGTCAAAATCATCTCATGTTGCTCAGAACAAGTACTTAATGCTTCTATTCAGAAATCAAACCATTGGCTGATGATGATTCAGCCATTA

*      900      *      920      *      940      *      960      *      980      *
N3  : TAGAAATTGTCAACGAGGTTTTGCATTAAAGTTTAAACGAAATTCGGGCTCGGTACCCSGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTGATCCCKGCTKYTAWWWW : 990
N3_FL : TAGAAATTGTCAACGAGGTTTTGCATTAA-----ATGTTTAAACGAAATTCGGGCTCGGTACCCSGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTGATCCCKGCTKYTAWWWW : 570
          TAGAAATTGTCAACGAGGTTTTGCATTAA

```

Appendix 9: Sequence verification of the 3' (568-1134bp) half of the HCoV-NL63 N-gene coding for the C-terminal half (N3) of the HCoV-NL63 N-protein. N3_FL represents the sequence for the 3' half of the HCoV-NL63 N-gene obtained from NCBI with accession number DQ846901.1, where N3 represents the sequence of the cloned N3-gene.



```

120      *      140      *      160      *      180      *      200      *      220
N4  : ATCCCGCGAAATTAATACGACTCACTATAGGCAATTTTAACTTTACTAAGGAGAATTCACCATGAAACATCAACATCAGCACCAAGCGATCGGCATGTTCTCTATTGCTTTGG : 226
N4_FL : -----ATGTTCTCTATTGCTTTGG : 19
          ATGTTCTCTATTGCTTTGG

*      240      *      260      *      280      *      300      *      320      *      34
N4  : CTCCAGAGCTCTCTGTGTTGAGTTTGAGGATCGCTCTAATAACTCATCTCGTGTAGCAGTCTTCTCAACTCGTAAACACTCACGAGACTCTTCTCTAGTCACTTCAAGA : 339
N4_FL : CTCCAGAGCTCTCTGTGTTGAGTTTGAGGATCGCTCTAATAACTCATCTCGTGTAGCAGTCTTCTCAACTCGTAAACACTCACGAGACTCTTCTCTAGTCACTTCAAGA : 132
          CTCCAGAGCTCTCTGTGTTGAGTTTGAGGATCGCTCTAATAACTCATCTCGTGTAGCAGTCTTCTCAACTCGTAAACACTCACGAGACTCTTCTCTAGTCACTTCAAGA

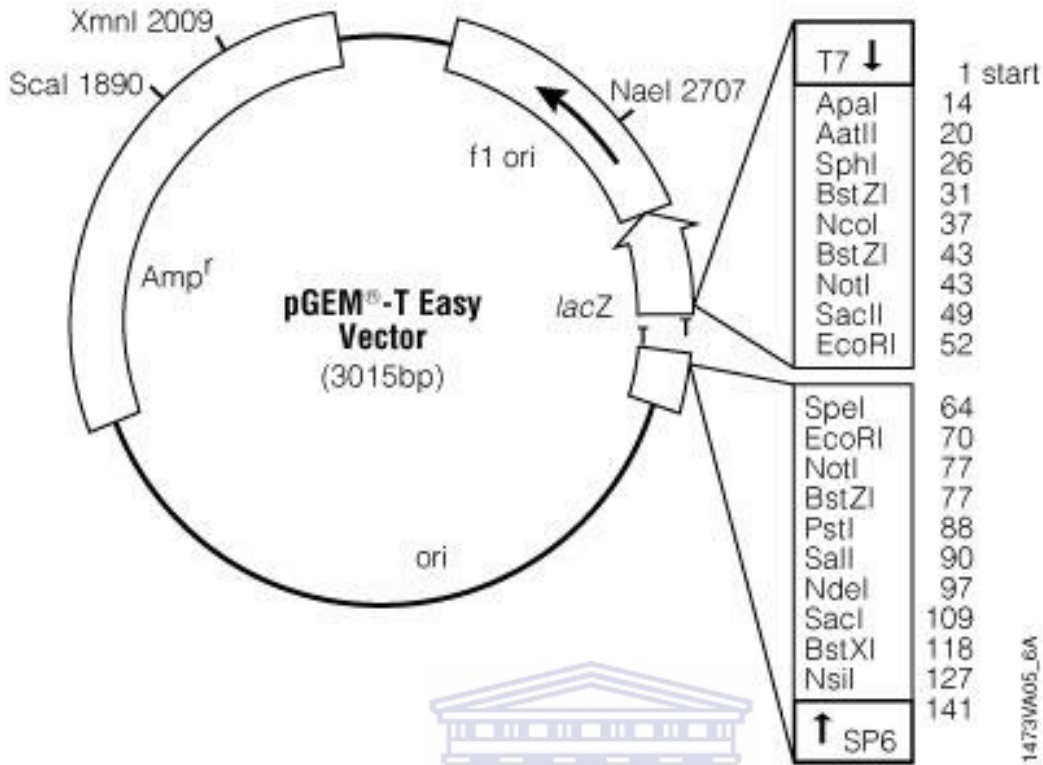
0      *      360      *      380      *      400      *      420      *      440
N4  : CAACAGCTCTCGCACTCGTCTGATTCTAACCAGCTTCTCAGATCTTGTGCTGCTGTACTTTGGCTTTAAAGAAGCTTAGGTTTGGATAACCAAGTCGAAGTCACCTAGTTC : 452
N4_FL : CAACAGCTCTCGCACTCGTCTGATTCTAACCAGCTTCTCAGATCTTGTGCTGCTGTACTTTGGCTTTAAAGAAGCTTAGGTTTGGATAACCAAGTCGAAGTCACCTAGTTC : 245
          CAACAGCTCTCGCACTCGTCTGATTCTAACCAGCTTCTCAGATCTTGTGCTGCTGTACTTTGGCTTTAAAGAAGCTTAGGTTTGGATAACCAAGTCGAAGTCACCTAGTTC

460      *      480      *      500      *      520      *      540      *      560
N4  : TTCTGGTACTTCCACTCCTAAGAAACCTAATAAGCCTTTTCTCAACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGTGTCTCTACCAGAGAGGAAA : 565
N4_FL : TTCTGGTACTTCCACTCCTAAGAAACCTAATAAGCCTTTTCTCAACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGTGTCTCTACCAGAGAGGAAA : 358
          TTCTGGTACTTCCACTCCTAAGAAACCTAATAAGCCTTTTCTCAACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGAAGCGTGTCTCTACCAGAGAGGAAA

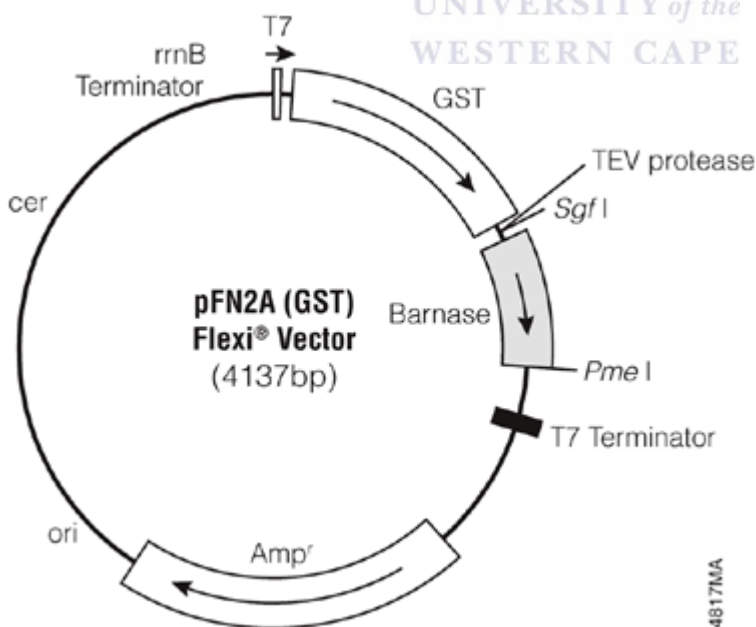
*      580      *      600      *      620      *      640      *      660      *      6
N4  : ATGTTATTTCAGTGCTTTGGTCTTAAAGTTTAAACGAAATTCGGGCTCGGTACCCSGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTGATCCCKGCTGYTAAACAAKCCCG : 678
N4_FL : ATGTTATTTCAGTGCTTTGGTCTTAA-----ATGTTTAAACGAAATTCGGGCTCGGTACCCSGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTGATCCCKGCTGYTAAACAAKCCCG : 384
          ATGTTATTTCAGTGCTTTGGTCTTAA

```

Appendix 10: Sequence verification of the middle region (379-756bp) of the HCoV-NL63 N-gene coding for the middle region (N4) of the HCoV-NL63 N-protein. N4_FL represents the sequence for the middle region of the HCoV-NL63 N-gene obtained from NCBI with accession number DQ846901.1, where N4 represents the sequence of the cloned N4-gene.



Appendix 11: Sequencing vector pGEM®-T Easy vector.



Appendix 12: Bacterial protein expression vector, pFN2A Flexi™.