EXPRESSION OF HUMAN CORONAVIRUS NL63 AND SARS-CoV NUCLEOCAPSID PROTEINS FOR ANTIBODY PRODUCTION

By

Yanga Eddie Mnyamana

Thesis submitted in fulfilment of the requirements for the Degree of Magister Scientiae (MSc)

Medical Bio-Sciences
Faculty of Natural Sciences
University of the Western Cape

Supervisor: Professor B.C. Fielding
Department of Medical Biosciences
University of the Western Cape

Co-supervisor: Doctor Ruth McBride
Department of Medical Biosciences
University of the Western Cape

November 2012
Expression of Human Coronavirus NL63 and SARS-CoV Nucleocapsid Proteins for Antibody Production

Yanga E. Mnyamana

KEYWORDS

Human Coronavirus

Human Coronavirus NL63

Severe Acute Respiratory Syndrome Coronavirus

Nucleocapsid proteins

Antibodies

Protein expression

Antibody validation

Enzyme-Linked Immunosorbent Assay

Viral antigens

MagneGST Purification System
ABSTRACT

Expression of Human Coronavirus NL63 and SARS-COV Nucleocapsid Proteins for Antibody Production

Y. E. Mnyamana, Degree of Magister Scientiae (MSc) Thesis, Department of Medical Biosciences, University of the Western Cape

Human Coronaviruses (HCoVs) are found within the family Coronaviridae (genus, Coronavirus) and are enveloped, single-stranded, positive-sense RNA viruses. Infections of humans by coronaviruses are not normally associated with severe diseases. However, the identification of the coronavirus responsible for the outbreak of severe acute respiratory syndrome (SARS-CoV) showed that highly pathogenic coronaviruses can enter the human population. The SARS-CoV epidemic resulted in 8 422 cases with 916 deaths globally (case fatality rate: 10.9%). In 2004 a group 1 Coronavirus, designated Human Coronavirus NL63 (HCoV-NL63), was isolated from a 7 month old Dutch child suffering from bronchiolitis.

In addition, HCoV-NL63 causes disease in children (detected in approximately 10% of respiratory tract infections), the elderly and the immunocompromised. This study was designed to express the full length nucleocapsid (N) proteins of HCoV-NL63 and SARS-CoV for antibody production in an animal model. The NL63-N/pFN2A and SARS-N/pFN2A plasmid constructs were used for this study. The presence of the insert on the Flexi ® vector was confirmed by restriction endonuclease digest and sequence verification. The sequenced chromatographs obtained from Inqaba Biotec were consistent with sequences from the NCBI Gen_Bank. Proteins were expressed in a KRX Escherichia coli bacterial system and analysed using 15% SDS-PAGE and Western Blotting. Thereafter, GST-tagged proteins were purified with an affinity column purification system. Purified
fusion proteins were subsequently cleaved with Pro-TEV Plus protease, separated on 15% SDS-PAGE gel and stained with Coomassie Brilliant Blue R250. The viral fusion proteins were subsequently used to immunize Balbc mice in order to produce polyclonal antibodies. A direct ELISA was used to analyze and validate the production of polyclonal antibodies by the individual mice. This is a preliminary study for development of diagnostic tools for the detection of HCoV-NL63 from patient samples collected in the Western Cape. 

November 2012
PUBLICATION OF RESEARCH

Oral presentation

YE Mnyamana, BC Fielding and R McBride, Development of diagnostic tools for Human Coronavirus NL63 (HCoV-NL63), Medical Biosciences Seminar Series, University of the Western Cape, 31st of March 2011. (Oral presentation)
DECLARATION

I, Yanga Eddie Mnyamana, declare that the thesis entitled “Expression of Human Coronavirus NL63 and SARS-CoV Nucleocapsid Proteins for Antibody Production” 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.

Yanga Eddie Mnyamana  

Date Signed
DEDICATION

This dissertation is dedicated to my parents, Zandisile Lucas Mnyamana and Noluntu Virginia Mnyamana who have granted me an opportunity to obtain an education and continue to support me in my academic endeavours. To my siblings, Yangisa Shirley Mnyamana and Lwazi Jonathan Mnyamana, who have always helped me and believed that I could obtain all I desire. To my late friend, Zolile Norman Hobo who was a guide, philosopher and most importantly a brother in the faith. I thank you all for your prayers, love and support throughout the years of my academics.
ACKNOWLEDGEMENTS

The printed pages of this dissertation hold far more than the culmination of years in completing this thesis. Without the support, patience and guidance of the following people, this study would not have been accomplished. It is to them that I owe my genuine gratitude.

- To my Lord and saviour Jesus Christ, for granting me health, strength and wisdom in order for me to complete this study.

- To my supervisors Professor Burtram C. Fielding and Doctor Ruth McBride, for your guidance, patience, and providing me with an excellent atmosphere for doing research.

- To Professor Edmund Pool, for his assistance in antibody production and validation.

- To my laboratory colleagues thank you all for your encouraging words, constructive criticism, time and attention during this study.

- To the Department of Science and Technology (DST) and National Research Foundation (NRF) and the Medical Biosciences Department for financial support during this study.
LIST OF ABBREVIATIONS

aa  Amino acids
A  Adenine
ACE  Angiotensin Converting Enzyme
Amp  Ampicillin
μg  Micrograms
μl  Microliters
mM  Micromolar
bp  Base pairs
BCV  Bovine Coronavirus
BSA  Bovine Serum Albumin
C  Cytosine
°C  Degrees Celsius
cDNA  Complementary Deoxyribonucleic Acid
COPD  Chronic Obstructive Pulmonary Diseases
DNA  Deoxyribonucleic Acid
dNTP  Deoxyribonucleotide triphosphate
DTT  Dithiothreitol
E  Envelope
E.coli  Escherichia coli
EDTA  Ethylenediaminetetraacetic acid
ELISA  Enzyme-Linked Immunosorbent Assay
g  Grams
G  Guanine
GC  Guanine Cytosine
GST  Glutathione-S-Transferase
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCoV</td>
<td>Human Coronavirus</td>
</tr>
<tr>
<td>HE</td>
<td>Hemagglutinin Esterase</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish Peroxidase</td>
</tr>
<tr>
<td>IBV</td>
<td>Infectious Bronchitis Virus</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-d-thiogalactopyranoside</td>
</tr>
<tr>
<td>K</td>
<td>Potassium</td>
</tr>
<tr>
<td>Kbs</td>
<td>Kilo bases</td>
</tr>
<tr>
<td>KD</td>
<td>Kawasaki Disease</td>
</tr>
<tr>
<td>kDA</td>
<td>Kilo Daltons</td>
</tr>
<tr>
<td>L</td>
<td>Litre</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Bertani</td>
</tr>
<tr>
<td>mg/ml</td>
<td>Milligrams per millilitre</td>
</tr>
<tr>
<td>M</td>
<td>Membrane</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>Mg</td>
<td>Milligrams</td>
</tr>
<tr>
<td>MHV</td>
<td>Mouse Hepatitis Virus</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>Magnesium chloride</td>
</tr>
<tr>
<td>ml</td>
<td>Millilitre</td>
</tr>
<tr>
<td>mM</td>
<td>Millimolar</td>
</tr>
<tr>
<td>NaCl</td>
<td>Sodium Chloride</td>
</tr>
<tr>
<td>NA</td>
<td>Sodium</td>
</tr>
<tr>
<td>ng</td>
<td>Nanograms</td>
</tr>
<tr>
<td>OD</td>
<td>Optical Density</td>
</tr>
<tr>
<td>ORF 3</td>
<td>Open reading frame 3</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide Gel Electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PEDV</td>
<td>Porcine Epidemic Diarrhea Virus</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td>TGEV</td>
<td>Transmissible Gastroenteritis Virus</td>
</tr>
<tr>
<td>SARS-CoV</td>
<td>Severe Acute Respiratory Syndrome Coronavirus</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulphate</td>
</tr>
<tr>
<td>T</td>
<td>Thymine</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco Etch Virus</td>
</tr>
<tr>
<td>U</td>
<td>Uracil</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>vs</td>
<td>Versus</td>
</tr>
<tr>
<td>V</td>
<td>Volts</td>
</tr>
<tr>
<td>v/v</td>
<td>volume per volume</td>
</tr>
<tr>
<td>VLP</td>
<td>Virus-like Particle</td>
</tr>
<tr>
<td>w/v</td>
<td>Weight per volume</td>
</tr>
<tr>
<td>w/w</td>
<td>Weight per weight</td>
</tr>
<tr>
<td>X-GAL</td>
<td>5-bromo-4-chloro-3-indolyl-β-d-galactopyranoside</td>
</tr>
<tr>
<td>%</td>
<td>Percentage</td>
</tr>
<tr>
<td>&lt;</td>
<td>Less-than</td>
</tr>
<tr>
<td>&gt;</td>
<td>Greater-than</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Prokaryotic vectors used in this study.</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Represents the functions of the components in a cell lysis buffer.</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Concentration of purified proteins in mg/ml</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Represents the viral fusion proteins that were used in this study</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figures</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>7</td>
</tr>
<tr>
<td>Negatively stained electron micrograph of HCoV-NL63</td>
<td></td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>8</td>
</tr>
<tr>
<td>Genome organization of HCoV-NL63 and other group I coronaviruses</td>
<td></td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>12</td>
</tr>
<tr>
<td>Phylogenetic analysis of partial 1a and partial spike gene sequences of HCoV-NL63 isolates</td>
<td></td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>16</td>
</tr>
<tr>
<td>A typical coronavirus virion morphology</td>
<td></td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>19</td>
</tr>
<tr>
<td>The spike (S) protein</td>
<td></td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>22</td>
</tr>
<tr>
<td>The membrane (M), envelope (E), and nucleocapsid (N) proteins</td>
<td></td>
</tr>
<tr>
<td>Figure 1.7</td>
<td>30</td>
</tr>
<tr>
<td>The Coronavirus life cycle</td>
<td></td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>101</td>
</tr>
<tr>
<td>Verification of insert sizes using restriction endonuclease digest</td>
<td></td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>103</td>
</tr>
<tr>
<td>Sequence verification of the cloned HCoV-NL63 Nucleocapsid gene in pGEM® vector</td>
<td></td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>104</td>
</tr>
<tr>
<td>Sequence verification of the cloned SARS-CoV Nucleocapsid gene in pGEM® vector</td>
<td></td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>106</td>
</tr>
<tr>
<td>Expression of unpurified total viral fusion proteins detected using (A) Coomassie stained SDS-PAGE gel and (B) Western Blot</td>
<td></td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>108</td>
</tr>
<tr>
<td>Detection of purified total proteins using (A) Coomassie stained SDS-PAGE gel and (B) Western Blot</td>
<td></td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>110</td>
</tr>
<tr>
<td>Digest of the GST affinity tag from HCoV-NL63 full length N protein using ProTEV protease enzyme</td>
<td></td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>110</td>
</tr>
<tr>
<td>Digest of the GST affinity tag from SARS-CoV full length N protein using ProTEV protease</td>
<td></td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>133</td>
</tr>
<tr>
<td>Detection of viral antibodies against an HCoV-NL63 N coated plate</td>
<td></td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>133</td>
</tr>
<tr>
<td>Detection of viral antibodies against a SARS-CoV N coated plate</td>
<td></td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>134</td>
</tr>
<tr>
<td>Detection of viral antibodies treated with SDS against an HCoV-NL63 N coated plate</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.4  Detection of viral antibodies treated with SDS against an SARS-CoV N coated plate
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title page</td>
<td></td>
</tr>
<tr>
<td>Keywords</td>
<td>ii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iii-iv</td>
</tr>
<tr>
<td>Publication of Research</td>
<td>v</td>
</tr>
<tr>
<td>Declaration</td>
<td>vi</td>
</tr>
<tr>
<td>Dedication</td>
<td>vii</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>viii</td>
</tr>
<tr>
<td>List of abbreviations</td>
<td>ix-xi</td>
</tr>
<tr>
<td>List of tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of figures</td>
<td>xii-xiv</td>
</tr>
</tbody>
</table>

## CHAPTER 1: Human Coronavirus: Literature Review

1.1 Background 2-3

1.2 Genome of Coronaviruses 3-12

1.2.1 Taxonomy of Coronaviruses 4-8

1.2.2 Discovery of Group 1 Coronaviruses 8-10

1.2.3 Genetic differences between isolates 10-12

1.3 Virion morphology and structural proteins 13-28
1.3.1  The Virus and Nucleocapsid protein  13-16  
1.3.2  Spike Protein  16-19  
1.3.3  Membrane Protein  19-22  
1.3.4  Envelope Protein  22-24  
1.3.5  Nucleocapsid Protein  24-28  

1.4  Viral replication cycle, Assembly and Protein Interactions  28-39  

1.4.1  Virion Assembly and Protein Interactions  30-39  

1.4.1.1  M Protein–M Protein Interactions  32  
1.4.1.2  S Protein–M Protein Interactions  32-36  
1.4.1.3  N Protein–M Protein Interactions  36-37  
1.4.1.4  Role of E Protein  38-39  

1.5  Discovery of HCoV NL63  40-53  

1.5.1  Infection with HCoV-NL63  41-53  

1.5.1.1  Worldwide spread  41-42  
1.5.1.2  Epidemiology of HCoV Infections in Children  42-45  
1.5.1.3  Seasonality of infection  46  
1.5.1.4  Disease(s) associated with HCoV-NL63  46-48  
1.5.1.5  Transcription and replication of HCoV-NL63  49-51  
1.5.1.6  Antiviral agents  51-53  

1.6  Aims of this study  53
CHAPTER 2: Verification of constructs and expression of proteins

2.1 Abstract

2.2 Introduction

2.3 Methodology

2.3.1 Bacterial strains and plasmids

2.3.2 The recombinant HCoV-NL63 N and SARS-CoV N full length constructs

2.3.3 Large-scale preparation of recombinant bacterial expression plasmids

2.3.3.1 Autoinduction

2.3.3.2 Transformation of competent cells

2.3.3.3 Plasmid DNA Extraction

2.3.4 Characterization of recombinant bacterial expression plasmids

2.3.4.1 Restriction endonucleases digestions

2.3.4.2 Agarose gel electrophoresis

2.3.4.3 Nucleotide sequencing and analysis

2.3.5 Protein analysis
CHAPTER 3: Generation and Validation of Antibodies

3.1 Abstract 120-121

3.2 Introduction 122-124

3.3 Methodology 125-128

3.3.1 Balb/c mice strain 125

3.3.2 Antigen preparation 125-126

3.3.3 Immunization of the mice 126-127

3.3.4 Direct ELISA on antigen coated plates 127-128

3.3.5 Statistical Analysis 128

3.4 Results and Discussion 129-135

3.5 Conclusion 136

3.6 References 137-142
<table>
<thead>
<tr>
<th>Section</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER 4: Summary</td>
<td>143-146</td>
</tr>
<tr>
<td>Appendix 1</td>
<td>148</td>
</tr>
<tr>
<td>Appendix 2</td>
<td>149</td>
</tr>
<tr>
<td>Appendix 3</td>
<td>149</td>
</tr>
<tr>
<td>Appendix 4</td>
<td>150</td>
</tr>
<tr>
<td>Appendix 5</td>
<td>151-152</td>
</tr>
<tr>
<td>Appendix 6</td>
<td>152</td>
</tr>
<tr>
<td>Appendix 7</td>
<td>152</td>
</tr>
</tbody>
</table>
CHAPTER ONE

Human Coronavirus NL63: Literature Review

Department of Medical Biosciences, Faculty of Natural Sciences, University of the Western Cape, South Africa.
1.1 BACKGROUND

Coronaviruses, a genus of the *Coronaviridae* family, are enveloped viruses with a positive strand RNA genome. The genomic RNA is 27-32 kb in size, capped and polyadenylated. Coronaviruses have been identified in bats, mice, rats, chickens, turkeys, swine, dogs, cats, rabbits, horses, cattle, humans and cause highly prevalent diseases such as respiratory, enteric, cardiovascular and neurological disorders (Hofmann, Pyrc et al. 2005; Dijkman, Jebbink et al. 2008). Originally, coronaviruses were classified on the basis of antigenic cross-reactivity, and three antigenic groups were recognized. When coronavirus genome sequences began to accumulate, the original antigenic groups were converted into genetic groups based on similarity of the nucleotide sequences (Hofmann, Pyrc et al. 2005; Dijkman, Jebbink et al. 2008). The coronavirus genome contains at least seven open reading frames (ORFs) and untranslated regions (UTRs) at the 5’ and 3’ ends. The two large 5’ terminal ORFs 1a and 1ab encode non-structural proteins that are required for viral replication (Moes, Vijgen et al. 2005).

The remaining ORFs at the 3’ end of the genome encode the four structural proteins spike (S), envelope (E), membrane (M) and nucleocapsid (N) for group I coronaviruses and an additional haemaglutinin esterase (HE) protein for group II coronaviruses. Several accessory ORFs are located between the structural protein genes. The position and order of these ORFs varies between species (Fouchier, Hartwig et al. 2005). Recent studies have shown that some of the accessory proteins of severe acute respiratory syndrome coronavirus (SARS-CoV) are structural proteins that are incorporated in the virion (Pyrc, Bosch et al. 2006). The earliest-described HCoV strains, HCoV-229E and HCoV-OC43, which are group I and group II coronaviruses, respectively, have now been joined by the more recently described group I and II strains HCoV-NL63.
and HCoV-HKU1, respectively (Fouchier, Hartwig et al. 2004; van der Hoek, Pyrc et al. 2004; Woo, Lau et al. 2004; Perlman and Netland 2009). The SARS-CoV constitutes a fifth HCoV, which was in circulation for a limited time during 2002 and 2003, when a novel virus appeared in humans and caused an outbreak affecting at least 8,000 people (Drosten, Gunther et al. 2003; Peiris, Yuen et al. 2003). Mortality rate was high, at approximately 10% (Peiris, Yuen et al. 2003). Furthermore, in 2012, a sixth novel CoV was discovered and sequenced at Erasmus Medical Centre (EMC) in the Netherlands (Corman, Eckerle et al. 2012). The new virus, designated HCoV-EMC, has been responsible for two severe cases in England and Saudi Arabia, one of which was fatal (Corman, Eckerle et al. 2012).

1.2 GENOME OF CORONAVIRUSES (CoVs)

The genomes of coronaviruses are non-segmented, single-stranded RNA molecules of positive sense, that is, the same sense as mRNA; and structurally they resemble most eukaryotic mRNAs in having both 50 caps and 30 poly (A) tails (Lomnicz and Kennedy 1977; Schochetman, Stevens et al. 1977; Lai and Stohlman 1978; Wege, Muller et al. 1978; Lai and Stohlman 1981). In contrast with other eukaryotic mRNAs coronavirus genomes are extremely large, with lengths ranging from 27.3 kb (HCoV-229E) to 31.3 kb (MHV), which are three times the size of alphavirus and flavivirus genomes and four times the size of picornavirus genomes (Masters 2006). The genes for the four canonical structural proteins will be discussed below and they account for less than one-third of the coding capacity of the genome and are clustered at the 3´ end. A single gene, which encodes the viral replicase, occupies the 5´which is mostly two-thirds of the genome. The invariant gene order in all members of the coronavirus family is 5´-replicase- S-E-M-N-3´.
(Masters 2006). However, engineered rearrangement of the gene order of MHV was found to be completely tolerated by the virus (de Haan, Volders et al. 2002b).

This implies that the native order, although it became fixed early in the evolution of the family, is not functionally essential. At the termini of the genome are a 5′ UTR, ranging from 210 to 530 nucleotides, and a 3′ UTR, ranging from 270 to 500 nucleotides. The non-coding regions between the ORFs are generally quite small; in some cases, there is a small overlap between adjacent ORFs. Additionally, one or a number of accessory genes are intercalated among the structural protein genes (Masters 2006).

Subsequently, with almost all other positive-sense RNA viruses, the genomic RNA of coronaviruses is infectious when transfected into permissive host cells, as was originally shown for TGEV (Norman, McClurkin et al. 1968), IBV (Lomniczi and Kennedy 1977; Schochetman, Stevens et al. 1977), and MHV (Wege, Muller et al. 1978). The genome has multiple functions during infection. It acts initially as an mRNA that is translated into the huge replicase polyproteins, the complete synthesis of which requires a ribosomal frameshifting event. The replicase is the only translation product derived from the genome; all downstream ORFs are expressed from sub-genomic RNAs. The genome next serves as the template for replication and transcription. Finally, the genome plays a role in assembly, as progeny genomes are incorporated into progeny virions (Masters 2006).

1.2.1 Taxonomy of Coronaviruses

Coronaviruses are currently classified as one of the two genera in the family Coronaviridae (Enjuanes, Brian et al. 2000b). However, it is likely that the coronaviruses, as well as the other genus within the Coronaviridae, to which the toroviruses belong
(Snijder and Horzinek 1993), will each be assigned the taxonomic status of family in the near future (Gonzalez, Gomez-Puertas et al. 2003). Both the coronaviruses and the toroviruses, together with, the Arteriviridae (Snijder and Meulenberg 1998) and the Roniviridae families (Cowley, Dimmock et al. 2000; Dhar, Cowley et al. 2004), have been grouped together in the order Nidovirales. This higher level of organization recognizes a relatedness among these families that sets them apart from other non-segmented positive-strand RNA viruses. The most significant features that all nidoviruses have in common are: (i) gene expression through transcription of a set of multiple 3’-nested sub-genomic RNAs, (ii) expression of the replicase polyproteins via ribosomal frame-shifting, (iii) unique enzymatic activities among the replicase protein products, (iv) a virion membrane envelope and (v) a multi-spanning integral membrane protein in the virion. The first of these qualities provides the name for the order, which derives from the Latin nido for nest (Enjuanes, Spaan et al. 2000a). Despite these commonalities, nidovirus families differ from one another in distinct ways, most evidently in the numbers, types, and sizes of the structural proteins in their virions and in the morphologies of their nucleocapsids. A more detailed comparison of characteristics of these virus families has been given by Lai and Cavanagh (1997) and Enjuanes et al. (2000b) (Lai and Cavanagh 1997; Enjuanes, Brian et al. 2000b).

Classification into groups was originally based on antigenic relationships. However, such a criterion reflects the properties of a limited subset of viral proteins, and cases have arisen where clearly related viruses in group 1 were found not to be serologically cross-reactive (Sanchez, Jimenez et al. 1990). Subsequently, sequence comparisons of entire viral genomes have come to be the basis for group classification (Gorbalenya, Snijder et al. 2004). Approximately all group 1 and group 2 viruses have mammalian hosts, with human coronaviruses falling into each of these groups. Viruses of
group 3, by contrast, have been isolated solely from avian hosts. Consequently, it came as quite a surprise, in 2003, when the causative agent of SARS was found to be a coronavirus (SARS-CoV) (Masters 2006). Equally astonishing have been the outcomes of renewed efforts, following the SARS epidemic, to detect previously unknown viruses; these investigations have led to the discovery of two more human respiratory coronaviruses, HCoV-NL63 (van der Hoek, Pyrc et al. 2004) and HCoV-HKU1 (Woo, Lau et al. 2005a). Three distinct bat coronaviruses have also been isolated: two are members of group 1, and the third, in group 2, is a likely precursor of the human SARS-CoV (Lau, Woo et al. 2005; Poon, Chu et al. 2005; Li, Shi et al. 2005c). In addition, new IBV-like viruses have been found that infect geese, pigeons, and ducks (Jonassen, Kofstad et al. 2005).

In almost all cases, the assignment of a coronavirus species to a given group has been explicit. However, the classification of SARS-CoV has provoked considerable controversy. The original, unrooted, phylogenetic characterizations of the SARS-CoV genome sequence posited this virus to be roughly distinctive from each of the three previously established groups (Masters 2006). It was thus proposed to be the first recognized member of a fourth group of coronaviruses (Marra, Jones et al. 2003; Rota, Oberste et al. 2003). However, a subsequently constructed phylogeny based on gene 1b, which contains the viral RNA-dependent RNA polymerase and which was rooted in the toroviruses as an out-group, concluded that SARS-CoV is most closely related to the group 2 coronaviruses (Snijder, Bredenbeek et al. 2003).

Furthermore, it was noted that regions of gene 1a of SARS-CoV contain domains that are unique to the group 2 coronaviruses (Gorbalenya, Snijder et al. 2004). Additional analysis of a subset of structural gene sequences (Eickmann, Becker et al. 2003) and of RNA secondary structures in the 3´ untranslated regions (3´ UTR) of the genome (Goebel,
Taylor et al. 2004b) also supported a group 2 assignment. By contrast, some authors have argued, based on bioinformatics methods, that the ancestor of SARS-CoV was derived from multiple recombination events among progenitors from all three groups (Rest and Mindell 2003; Stanhope, Brown et al. 2004). While these latter studies assume that historically there has been limitless opportunity for intergroup recombination, there is no well-documented example of recombination between extant coronaviruses of different groups. Moreover, it is not clear that intergroup recombination is even possible, owing to replicative incompatibilities among the three coronavirus groups (Goebel, Taylor et al. 2004b). Therefore, although SARS-CoV does indeed have unique features, the currently available evidence best supports the conclusion that it is more closely allied with the group 2 coronaviruses and that it has not sufficiently diverged to constitute a fourth group (Gorbalenya, Snijder et al. 2004).

![Image of HCoV-NL63](image)

**Figure 1.1: Negatively stained electron micrograph of HCoV-NL63.** (960 x 720 – 37.9 KB – jpeg) (van der Hoek, Pyrc et al. 2006)
1.2.2 Discovery of Group 1 Coronaviruses (CoVs)

The first described coronavirus (CoV) of group 1 was porcine transmissible gastroenteritis virus (TGEV), which was isolated in 1946 from pigs suffering from gastroenteritis (Toyoda, Tanaka et al. 2006). Almost two decades later, a United Kingdom based research group identified a human respiratory tract pathogen from nasal washings of persons with the common cold (Uehara, Miyamoto et al. 2005). This novel pathogen, HCoV-229E, was later characterized morphologically by electron microscopy and compared with the already-well-known avian Infectious Bronchitis Virus (IBV) (Yamagami, Usui et al. 2005). Both these viruses exhibited a typical crown-like appearance (from Latin corona). Subsequently, another group 1 member, canine coronavirus, was isolated from sentry dogs with diarrhea and mild gastroenteritis (Minogue, Liu et al. 2005). Thereafter, similar clinical symptoms were observed in pigs on four separate swine breeding farms during a diarrheal outbreak in 1978. (Yang, Hayashi et al. 2005). The recovered pathogen, now known as porcine epidemic diarrhea virus
(PEDV), was first mistyped as a member of the rotavirus family, yet it soon became clear that the virus shared the morphological characteristics of CoV but was serologically distinct from TGEV (Yasuo, Nakao et al. 2006). Two cat-associated CoV species were identified in 1981. Feline enteric coronavirus (FEC) and feline infectious peritonitis virus (FIPV) shared serological characteristics but differed in clinical outcome (Ebihara, Kondoh et al. 2006). In 1986, another porcine CoV was isolated, porcine respiratory coronavirus, a close relative of TGEV (Yang, Hayashi et al. 2005). Hereafter, no new group 1 members were discovered for more than 15 years. In 2004, a group in Netherlands isolated HCoV-NL63 from a 7-month-old child with bronchiolitis (van der Hoek, Pyrc et al. 2004). Shortly thereafter, Fouchier et al independently described the same virus from a clinical sample collected in 1988 (Fouchier, Hartwig et al. 2005). In 2005, it became clear that several bat species can harbor CoVs that belong to group 1 (Beppu, Sasaki et al. 2005; Ebihara, Ishikawa et al. 2005; Kimata, Sakuraba et al. 2005). Most of these viruses cluster with HCoV-229E, HCoV NL63 and PEDV, although none of them is a very close relative to any of these viruses. There are notable differences in the genome composition that divides the group 1 viruses into two separate branches, named 1A and 1B.

All group 1A members contain several short accessory protein-coding genes between the S and E genes and one or two accessory protein genes on the 3’ side of the N gene. In contrast, all group 1B members carry only one accessory protein gene, between the S and E genes, with the exception of some bat CoVs. Three bat CoVs carry an additional ORF at the 3’-side of the N gene. The function of the accessory proteins from the group 1 CoVs is unknown. Reverse genetic analyses of FIPV and extensive cell culture adaptation of PEDV, TGEV and HCoV-229E suggest that they are not required for in vitro virus replication. Moreover, deletion of FIPV, PEDV and TGEV accessory genes results in attenuation of the virus, which indicates that the group 1 accessory proteins represent
pathogenicity factors replication. The discovery timeline of the group 1 CoVs illustrates that this group has grown only recently into a more mature form in which its members can infect a diversity of mammalian hosts. It is not likely that additional members will be identified in the near future. (Asada, Yamaya et al. 2005; Ebihara, Groseth et al. 2005; Ma, Endo et al. 2005; Nakamura, Sugaya et al. 2005; Dijkman, Jeubbink et al. 2006).

1.2.3 Genetic differences between isolates

The RNA viruses have an enormous potential to generate genetic diversity due to their high mutation rates, which are estimated at approximately one mutation per genome per replication cycle (McIntosh, Chao et al. 1974b; Drake and Holland 1999). As a result of this, RNA viruses are known to exist as a dynamic distribution of variants with a closely related, yet non-identical genome centered on a master sequence, the so called quasispecies concept (Domingo 2002). These high mutational rates provide RNA viruses with the capacity to increase their level of genetic variability and thereby their ability to adapt quickly to a change in the environment. In addition to the high mutation rate, recombination events increase the level of genetic diversity. Coronavirus recombination is assumed to be mediated by a ‘copy choice mechanism’ with template switching by the polymerase during RNA-dependent RNA synthesis (Kuijpers, Wiegman et al. 1999b). In-vivo recombination has been described between group I feline and canine coronaviruses, between strains of Mouse Hepatitis Virus (MHV) and between strains of Infectious Bronchitis Virus (IBV) (Kuba, Imai et al. 2005).

Phylogenetic analysis of a small part of the 1a gene of HCoV-NL63 shows diversity of the natural isolates and subsequent clustering in NL63 subgroups (Arden, Nissen et al. 2005; Bastien, Anderson et al. 2005a). The close clustering of isolates from
the Netherlands, Belgium, Canada, Hong Kong, Korea and Australia indicates that this clustering does not correlate with the geographic origin (Figure 1.3). Phylogenetic analysis of the S gene also reveals the presence of subgroups (McIntosh, Becker et al. 1967; McIntosh, Dees et al. 1967). Moreover, but not unexpectedly for coronaviruses, the clustering based on the 1a gene does not match with that of the S gene, indicating that recombination has occurred between different HCoV-NL63 isolates (McIntosh 1996). Sequencing of full length genomes of field isolates verifies the inharmonious clustering of 1a-gene sequences and S sequences and confirms that recombination is also common for HCoVNL63. Besides the variable regions observed, other parts of the HCoV-NL63 genome are highly conserved among isolates. This is especially true of the 50 region of the 1b gene and the nucleocapsid protein gene, which both show a low level of heterogeneity among natural isolates of HCoVNL63. These regions therefore present the best targets for the design of diagnostic primers and probes. Clinical samples will contain the virus with its full-length genome, but may also contain sub-genomic mRNAs from infected cells. Considering that the sequence of the nucleocapsid (N) gene is present in the genomic RNA and also in all sub-genomic mRNAs (Pinon, Teng et al. 1999), this region of the HCoV-NL63 genome may represent the ideal target for the design of molecular-based detection assays.
Figure 1.3: Phylogenetic analysis of partial 1a and partial spike gene sequences of HCoV-NL63 isolates. HCoV-229E was used to root the tree for the 1a gene sequences. The neighbour-joining method of the MEGA program was used with bootstrap re-sampling (1000 replicates) to place approximate confidence limits on individual branches; only bootstrap values above 80 are shown. GenBank numbers: NC_005831–AY567488, AY567494–AY675541, AY675553–AY746451, AY746458–AY758283, AY758287–AY758301, DQ093116–DQ093123, DQ106888–DQ106901. CAN, Canada; Q, Australia; NL, the Netherlands; BE, Belgium; KR, Korea. NL63-Amst1 is the prototype sequence of HCoV-NL63. Red and blue indicate the strains that cluster for the 1a region and for which spike sequences are also available (van der Hoek, Pyrc et al. 2006).
1.3 VIRION MORPHOLOGY AND STRUCTURAL PROTEINS

1.3.1 The Virus and Nucleocapsid

Coronaviruses are roughly spherical and moderately pleomorphic (Figure 1.4). Virions have typically been reported to have average diameters of 80–120 nm, but extreme sizes as small as 50 nm and as large as 200 nm are occasionally given in the older literature (Oshiro 1973; McIntosh, Chao et al. 1974b).

The surface spikes or peplomers of these viruses, described as either club-like, pear-shaped, or petal-shaped, project some 17–20 nm from the virion surface, having a thin base that swells to a width of about 10 nm at the distal extremity (McIntosh, Chao et al. 1974b; Sugiyama and Amano 1981). For some coronaviruses a second set of projections, 5–10-nm long, forms undergrowth beneath the major spikes (Sugiyama and Amano 1981; Patel, Davies et al. 1982; Guy, Breslin et al. 2000). These shorter structures are now known to be the haemaglutinin-esterase (HE) proteins that are found in a subset of group 2 coronaviruses.

In essence, some of the heterogeneity in coronavirus particle morphology can be attributed to the distorting effects of negative-staining procedures. Freeze-dried (Roseto, Bobulesco et al. 1982) and cryo-electron microscopic (Risco, Anton et al. 1996) preparations of BCoV and TGEV, respectively, showed much more homogeneous populations of virions, with diameters 10–30 nm greater than virions in comparable samples prepared by negative staining. Extraordinary three-dimensional images have been obtained for SARS-CoV virions emerging from infected Vero cells (Ng, Lee et al. 2004). These scanning electron micrographs and atomic force micrographs reveal knobby, rosette-like viral particles resembling tiny cauliflowers.
The internal component of the coronavirus virion is obscure in electron micrographs of whole virions. In negative-stained images the core appears as an indistinct mass with a densely staining centre, giving the virion a “punched-in” spherical appearance. Imaging of virions that have burst spontaneously, expelling their contents, or that have been treated with non-ionic detergents has allowed visualization of the coronavirus core (Masters 2006). This analysis led to the discovery of another unique characteristic of the coronavirus family: that its members possess helically symmetric nucleocapsids. Such nucleocapsid symmetry is distinctive for negative-strand RNA viruses, but almost all positive-strand RNA animal viruses have icosahedral ribonucleoprotein cupids. However, although it is fairly well accepted that coronaviruses have helical nucleocapsids, there are surprisingly few published data that bear on this issue. Additionally, the reported results vary considerably with both the viral species and the method of preparation (Masters 2006).

The earliest study of nucleocapsids from spontaneously disrupted HCoV-229E virions found tangled, threadlike structures 8–9 nm in diameter; these were unravelled or clustered to various degrees and, in rare cases, retained some of the shape of the parent virion (Kennedy and Johnson-Lussenburg 1975). A subsequent analysis of spontaneously disrupted virions of HCoV-229E and MHV observed more clearly helical nucleocapsids, with diameters of 14–16 nm and hollow cores of 3–4 nm (Macneughton and Davies 1978). The most highly resolved images of any coronavirus nucleocapsid were obtained with NP-40-disrupted HCoV-229E virions (Caul, Ashley et al. 1979).

These preparations showed filamentous structures 9–11 or 11–13 nm in diameter, depending on the method of staining, with a 3–4 nm central canal. The coronavirus nucleocapsid was noted to be thinner in cross-section than those of paramyxoviruses and
also to lack the sharply segmented “herringbone” appearance characteristic of paramyxovirus nucleocapsids (Masters 2006). By contrast, in early studies, IBV and TGEV nucleocapsids were undetectable with the techniques that had been successful with other viruses. Visualization of IBV nucleocapsids, which seemed to be very sensitive to degradation (Macneughton and Davies 1978), was finally achieved by electron microscopy of viral samples prepared by carbon-platinum shadowing (Davies, Dourmashkin et al. 1981). This revealed linear strands, some as long as 6–7 µm, which were only 1.5-nm thick, suggesting that they represented unwound helices. TGEV, on the other hand, was found to be more resistant to non-ionic detergents. Treatment of virions of this species with NP-40 resulted in spherical sub-viral particles with no threadlike substructure visible (Garwes, Pocock et al. 1976). The TGEV core was later seen as a spherically symmetric, possibly icosahedral, superstructure that only dissociated further into a helical nucleocapsid following Triton X-100 treatment of virions (Risco, Anton et al. 1996). Such a collection of incomplete and often discrepant results makes it clear that much further examination of the internal structure of coronavirus virions is warranted. This information would substantially aid our understanding of coronavirus structure and assembly if we had available a detailed description of nucleocapsid shape, length, diameter, helical repeat distance, and protein: RNA stoichiometry (Masters 2006).
1.3.2 Spike Protein (S)

There are three protein components of the viral envelope (Figure 1.5). The most prominent of these is the S glycoprotein (formerly called E2) (Cavanagh 1995), which mediates receptor attachment and viral and host cell membrane fusion (Collins, Knobler et al. 1982). The S protein is a very large, N-exo, C-endo transmembrane protein that assembles into trimers (Delmas and Laude 1990a; Song, Seo et al. 2004) to form the distinctive surface spikes of coronaviruses. S protein is inserted into the endoplasmic reticulum (ER) via a cleaved, amino-terminal signal peptide (Cavanagh, Davis et al.)
Human Coronavirus NL63: Literature Review

1986b). The ectodomain makes up most of the molecule, with only a small carboxyl-terminal segment (of 71 or fewer of the total 1162–1452 residues) constituting the transmembrane domain and endodomain. Monomers of S protein, prior to glycosylation, are 128–160 kDa, but molecular masses of the glycosylated forms of full-length monomers fall in the range of 150–200 kDa. The S molecule is thus highly glycosylated, and this modification is exclusively N-linked (Holmes, Doller et al. 1981; Rottier, Horzinek et al. 1981). S protein ectodomains have from 19 to 39 potential consensus glycosylation sites, but a comprehensive mapping of actual glycosylation has not yet been reported for any coronavirus. A mass spectrometric analysis of the SARS-CoV S protein has shown that at least 12 of the 23 candidate sites are glycosylated in this molecule (Krokhin, Li et al. 2003). For the TGEV S protein, it has been demonstrated that the early steps of glycosylation occur co-translationally, but that terminal glycosylation is preceded by trimerization, which can be rate limiting in S protein maturation (Delmas and Laude 1990a). In addition, glycosylation of TGEV S may assist monomer folding, given that tunicamycin inhibition of high-mannose transfer was found to also block trimerization. The S protein ectodomain has between 30 and 50 cysteine residues, and within each coronavirus group the positions of cysteines are well conserved (Abraham, Kienzle et al. 1990; Eickmann, Becker et al. 2003). However, as with glycosylation, a comprehensive mapping of disulfide linkages has not yet been achieved for any coronavirus S protein (Masters 2006).

Predominantly, in group 2 and all group 3 coronaviruses, the S protein is cleaved by a trypsin-like host protease into two polypeptides, S1 and S2, of roughly equal sizes. Even for uncleaved S proteins, that is, those of the group 1 coronaviruses and SARS-CoV, the designations S1 and S2 are used for the amino-terminal and carboxyl-terminal halves of the S protein, respectively (Masters 2006). Peptide sequencing has shown that cleavage
occurs following the last residue in a highly basic motif: RRFRR in IBV S protein (Cavanagh, Davis et al. 1986b), RRAHR in MHV strain A59 S protein (Luytjes, Sturman et al. 1987), and KRRSRR in BCoV S protein (Abraham, Kienzle et al. 1990). Similar cleavage sites are predicted from the sequences of other group 2 S proteins, except that of SARS-CoV. It has been noted that the S protein of MHV strain JHM has a cleavage motif (RRARR) more basic than that found in MHV strain A59 (RRAHR) (Masters 2006). An expression study has shown that this difference accounts for the almost total extent of cleavage of the JHM S protein that is seen in cell lines in which the A59 S protein undergoes only partial cleavage (Bos, Heijnen et al. 1995).

The S1 domain is the most divergent region of the molecule, both across and within the three coronavirus groups. Even among strains and isolates of a single coronavirus species, the sequence of S1 can vary extensively (Parker, Gallagher et al. 1989; Gallagher, Parker et al. 1990; Wang, Junker et al. 1994). By contrast, the most conserved part of the molecule across the three coronavirus groups is a region that encompasses the S2 portion of the ectodomain, plus the start of the transmembrane domain (de Groot, Luytjes et al. 1987). An early model for the coronavirus spike, which has held up well in light of subsequent work, proposed that the S1 domains of the S protein oligomer constituted the bulb portion of the spike. The stalk portion of the spike, on the other hand, was envisioned to be a coiled-coil structure, analogous to that in influenza HA protein, formed by association of heptad repeat regions of the S2 domains of monomers (de Groot, Luytjes et al. 1987).
Figure 1.5: The spike (S) protein. At the right is a linear map of the protein, denoting the amino-terminal S1 and the carboxyl-terminal S2 portions of the molecule. The arrowhead marks the site of cleavage for those S proteins that become cleaved by cellular protease(s). The signal peptide and regions of mapped receptor-binding domains (RBDs) are shown in S1. The heptad repeat regions (HR1 and HR2), putative fusion peptide (F), transmembrane domain, and endodomain are indicated in S2. At the left is a model for the S protein trimer (Masters 2006).

1.3.3 Membrane Protein (M)

The M glycoprotein (formerly called E1) is the most abundant constituent of coronaviruses (Sturman 1977a; Sturman, Holmes et al. 1980) and gives the virion envelope its shape. The preglycosylated M polypeptide ranges in size from 25 to 30 kDa (221–262 amino acids), but multiple higher-molecular-mass glycosylated forms are often observed by SDS-PAGE (Locker, Griffiths et al. 1992a). The M protein of MHV has also been noted to multimerize under standard conditions of SDS-PAGE (Sturman 1977a). This is a multispanning membrane protein with a small, amino terminal domain located on the exterior of the virion, or, intracellularly, in the lumen of the ER (Figure 1.6) (Masters
2006). The ectodomain is followed by three transmembrane segments and thereafter a large carboxyl terminus comprising the major part of the molecule. This latter domain is situated in the interior of the virion or on the cytoplasmic face of intracellular membranes (Rottier, Nakamura et al. 2005). M proteins within each coronavirus group are moderately well conserved, but they are quite divergent across the three groups (Masters 2006). The region of M protein showing the most conservation among all coronaviruses is a segment of some 25 residues encompassing the end of the third transmembrane domain and the start of the endodomain; a portion of this segment even retains homology to its torovirus counterpart (Den Boon, Snijder et al. 1991). The ectodomain, which is the least conserved part of the M molecule, is glycosylated. For most group 2 coronaviruses, glycosylation is O-linked, although two exceptions to this pattern are MHV strain 2 (Yamada, Yabe et al. 2000) and SARS CoV (Nal, Chan et al. 2005), both of which have M proteins with N-linked carbohydrate. Group 1 and group 3 coronavirus M proteins, by contrast, exhibit N-linked glycosylation exclusively (Stern and Sefton 1982; Garwes, Bountiff et al. 1984; Jacobs, van der Zeijst et al. 1986; Cavanagh and Davis 1988a).

At the time of its discovery in the MHV M protein, O-linked glycosylation had not previously been seen to occur in a viral protein (Holmes, Doller et al. 1981), and MHV M has since been used as a model to study the sites and mechanism of this type of posttranslational modification (Niemann, Boschek et al. 1982; Locker, Griffiths et al. 1992a; de Haan, Roestenberg et al. 1998b). Although the functions of M protein glycosylation are not fully understood, the glycosylation status of M can influence both organ tropism in vivo and the capacity of some coronaviruses to induce alpha interferon in vitro (Charley and Laude 1988; Laude, Gelfi et al. 1992; de Haan, de Wit et al. 2003a).
The coronavirus M protein was the first polytopic viral membrane protein to be described (Armstrong, Niemann et al. 1984; Rottier, Brandenburg et al. 1984a), and the atypical topology of the MHV and IBV M proteins was examined in considerable depth in cell-free translation and cellular expression studies. For both of these M proteins, the entire ectodomain was found to be protease sensitive. However, at the other end of the molecule, no more than 20–25 amino acids could be removed from the carboxyl terminus by protease treatment (Rottier, Brandenburg et al. 1984a; Rottier, Welling et al. 1986; Cavanagh, Davis et al. 1986a; Mayer, Tamura et al. 1988). This pattern suggested that almost all of the endodomain of M is firmly associated with the surface of the membrane or that it has an unusually compact structure that is refractory to proteolysis (Rottier 1995). Most M proteins do not possess a cleaved amino-terminal signal peptide (Rottier, Brandenburg et al. 1984a; Cavanagh, Davis et al. 1986b), and for both IBV and MHV it was demonstrated that either the first or the third transmembrane domain alone is sufficient to function as the signal for insertion and anchoring of the protein in its native orientation in the membrane (Machamer and Rose 1987; Mayer, Tamura et al. 1988; Locker, Rose et al. 1992b). The M proteins of a subset of group 1 coronaviruses (TGEV, FIPV, and CCoV) each contain a cleavable amino-terminal signal sequence (Laude, Rasschaert et al. 1987), although this element may not be required for membrane insertion (Kapke, Tung et al. 1988; Vennema, Rijnbrand et al. 1991). Another anomalous feature of at least one group 1 coronavirus, TGEV, is that roughly one-third of its M protein assumes a topology in which part of the endodomain constitutes a fourth transmembrane segment, thereby positioning the carboxyl terminus of the molecule on the exterior of the virion (Risco, Anton et al. 1995).
1.3.4 Envelope Protein (E)

The E protein (formerly called sM) is a small polypeptide, ranging from 8.4 to 12 kDa (76–109 amino acids), and therefore is only a minor constituent of virions (Fig. 3). Owing to its tiny size and limited quantity, E was recognized as a virion component much later than other structural proteins were, first in IBV (Liu and Inglis 1991) and then in TGEV (Godet, L'Haridon et al. 1992) and MHV (Yu, Bi et al. 1994). Its significance was also obscured by the fact that in some coronaviruses, the coding region for E protein occurs as the furthest-downstream open reading frame (ORF) in a bi- or tricistronic mRNA and must therefore be expressed by a nonstandard translational mechanism (Budzilowicz and Weiss 1987). The E protein sequences are extremely divergent across the three coronavirus groups and in some cases, among members of a single group. Nevertheless, the same general architecture can be discerned in all E proteins: a short hydrophilic amino terminus (8–12 residues), followed by a large hydrophobic region (21–29 residues).
containing two to four cysteines, and a then hydrophilic carboxyl-terminal tail (39–76 residues); the latter constituting most of the molecule. Furthermore, the E protein is an integral membrane protein, as has been shown for both the MHV and IBV E proteins by the criterion of resistance to alkaline extraction (Vennema, Godeke et al. 1996; Corse and Machamer 2000), and membrane insertion occurs without cleavage of a signal sequence (Raamsman, Locker et al. 2000). The E protein of IBV has been shown to be palmitoylated on one or both of its two cysteine residues (Corse and Machamer 2002), but it is not currently clear whether this modification is a general characteristic. One study of MHV E showed a gel mobility shift of E caused by hydroxylamine treatment, which cleaves thioester linkages (Yu, Bi et al. 1994), but attempts to incorporate labelled palmitic acid into either the TGEV or MHV E protein have been unsuccessful (Godet, L’Haridon et al. 1992; Raamsman, Locker et al. 2000). The topology of E in the membrane is at least partially resolved. Although one early report suggested a C-exo, N-endo membrane orientation for the TGEV E protein (Godet, L’Haridon et al. 1992), more extensive investigations of the MHV and IBV E proteins both concluded that the carboxyterminal tail of the molecule is cytoplasmic (or, correspondingly, is situated in the interior of the virion) (Corse and Machamer 2000; Raamsman, Locker et al. 2000). Moreover, for IBV E, it was shown that the carboxyl-terminal tail, in the absence of the membrane-bound domain, specifies targeting to the budding compartment (Corse and Machamer 2002).

The IBV E protein amino terminus was inaccessible to antibodies at the cytoplasmic face of the Golgi membrane, suggesting that this end of the molecule is situated in the lumen (corresponding to the exterior of the virion) (Corse and Machamer 2000). Such a single transit, placing the termini of the protein on opposite faces of the membrane, is consistent with a prediction performed by molecular dynamics simulations; that a broad set of E proteins occur as transmembrane oligomers (Torres, Wang et al.
Conflicting results were obtained with MHV E, though. Based on the cytoplasmic reactivity of an engineered amino-terminal epitope tag, it was proposed that the MHV E protein amino terminus is buried within the membrane near the cytoplasmic face (Maeda, Repass et al. 2001).

These results are consistent with the findings that no part of the MHV E protein in purified virions is accessible to protease treatment (Raamsman, Locker et al. 2000). Such an orientation would mean that the hydrophobic domain of E protein forms a hairpin, looping back through the membrane. This topology agrees with the outcome of a biophysical analysis of the SARS-CoV E protein transmembrane domain (Arbely, Khattari et al. 2004). However, in the latter study it was asserted that the palindromic hairpin configuration of the transmembrane segment is unique to the SARS-CoV E protein, which begs the question of how the other coronavirus E proteins are situated in the membrane and why the E protein of SARS-CoV should differ.

### 1.3.5 Nucleocapsid Protein (N)

The N protein, which ranges from 43 to 50 kDa, is the protein component of the helical nucleocapsid and is thought to bind the genomic RNA in a beads-on-a-string fashion (Laude and Masters 1995a). Based on a comparison of sequences of multiple strains, it has been proposed that the MHV N protein is divided into three conserved domains which are separated by two highly variable spacer regions (Parker and Masters 1990). Domains 1 and 2, which constitute most of the molecule, are rich in arginines and lysines, as is typical of many viral RNA-binding proteins. In contrast, the short, carboxyl-terminal domain 3 has a net negative charge resulting from an excess of acidic over basic residues. While there is now considerable evidence to support the notion that domain 3
truly constitutes a separate domain (Koetzner, Parker et al. 1992; Hurst, Kuo et al. 2005), little is known about the structure of the other two putative domains. The overall features of the three-domain model appear to extend to N proteins of coronaviruses in groups 1 and 3, although the boundaries between domains appear to be less clearly defined for these latter N proteins. There is not a high degree of intergroup sequence homology among N proteins, with the exception of a strongly conserved stretch of 30 amino acids, near the junction of domains 1 and 2, which contains many aromatic hydrophobic residues (Laude and Masters 1995a).

The main activity of N protein is to bind to the viral RNA. Unlike the helical nucleocapsids of non-segmented negative-strand RNA viruses, coronavirus ribonucleoprotein complexes are quite sensitive to the action of ribonucleases (Macneughton and Davies 1978). A significant portion of the stability of the nucleocapsid may derive from N–N monomer interactions (Narayanan, Kim et al. 2003b). Both sequence-specific and non-specific modes of RNA binding by N have been assayed in vitro (Robbins, Frana et al. 1986; Nelson and Stohlman 1993; Zhou, Williams et al. 1996; Chen, Gill et al. 2005). Specific RNA substrates that have been identified for N protein include the positive sense transcription regulating sequence (Stohlman, Baric et al. 1988a; Nelson, Stohlman et al. 2000; Chen, Gill et al. 2005), regions of the 30 UTR (Zhou, Williams et al. 1996) and the N gene (Cologna, Spagnolo et al. 2000a), and the genomic RNA packaging signal (Molenkamp and Spaan 1997; Cologna, Spagnolo et al. 2000a). The RNA-binding capability of the MHV N protein has been mapped to domain 2 of this molecule (Masters 1992a; Nelson and Stohlman 1993). However, for IBV, two separate RNA-binding sites have been found to map, respectively, to amino- and carboxyl-terminal fragments of N protein (Zhou and Collisson 2000), and RNA-binding activity has been reported for a fragment of the SARS-CoV N protein containing parts of domains 1 and 2.
(Huang, Yu et al. 2004b). This phosphoprotein, has been shown for MHV, IBV, BCoV, TGEV, and SARS-CoV (Stohlman and Lai 1979; Lomniczi and Morser 1981; King and Brian 1982; Calvo, Escors et al. 2005; Zakhartchouk, Viswanathan et al. 2005).

Phosphorylation of MHV N occurs exclusively on serine residues (Stohlman and Lai 1979; Siddell, Barthel et al. 1981), but in IBV nucleocapsid, a phosphothreonine residue was also found (Chen, Gill et al. 2005). Kinetic analysis has shown that MHV N protein rapidly acquires phosphates following its synthesis (Siddell, Barthel et al. 1981; Stohlman, Fleming et al. 1983) and phosphorylation may lead to the association of N with intracellular membranes (Stohlman, Fleming et al. 1983; Calvo, Escors et al. 2005). Although some 15% of the amino acids of coronavirus N proteins are candidate phosphoacceptor serines and threonines, phosphorylation appears to be targeted to a small subset of residues. For MHV, this was concluded both from the degree of charge heterogeneity of N protein observed in two-dimensional gel electrophoresis and from the limited number of tryptic phosphopeptides of N that could be separated by HPLC (Bond, Leibowitz et al. 1979; Wilbur, Nelson et al. 1986). Mass spectrometry has been employed to map the sites of phosphorylation of the IBV and TGEV N proteins. For IBV N, this was accomplished by comparison of unphosphorylated N protein expressed in bacteria with phosphorylated N protein expressed in insect cells (Chen, Gill et al. 2005). Four sites of phosphorylation were found, two each in domains 2 and 3: Ser190, Ser192, Thr378, and Ser379. For TGEV N, purified virions and multiple fractions from infected cells were analysed (Calvo, Escors et al. 2005). Here also, four sites of phosphorylation were found, one in domain 1 and three in domain 2: Ser9, Ser156, Ser254, and Ser256. In both of these analyses, the degree of sequence coverage achieved did not entirely rule out the possibility of additional, undetected phosphorylated residues in each of these N proteins. The role of
N protein phosphorylation is currently unresolved, but this modification has long been speculated to have regulatory significance.

In vitro binding evidence has been presented that phosphorylated IBV N is better able to distinguish between viral and non-viral RNA substrates than is non-phosphorylated N (Chen, Gill et al. 2005). Possibly related to this result is the early conclusion, inferred from the differential accessibilities of some monoclonal antibodies, that phosphorylation induces a conformational change in the MHV N protein (Stohlman, Fleming et al. 1983). It has also been found that only a subset of the intracellular phosphorylated forms of BCoV N protein are incorporated into virions, suggesting that phosphorylation is linked to virion assembly and maturation (Hogue 1995). The recent mapping of at least some of the N phosphorylation sites in some coronaviruses has now laid the groundwork for testing of the hypothetical functions of phosphorylation by reverse genetic methods.

A number of potential activities, other than its structural role in the virion, have been put forward for N protein. Based on the specific binding of N protein to the transcription-regulating sequence within the leader RNA, it has been proposed that N participates in viral transcription (Baric, Nelson et al. 1988; Stohlman, Baric et al. 1988a; Choi, Huang et al. 2002). However, an engineered HCoV-229E replica RNA that was devoid of the N gene and all other structural protein genes retained the capability to synthesize sub-genomic RNA (Thiel, Karl et al. 2003b).

Therefore, if the N protein does function in transcription, it must be in a modulator, but not essential, capacity. Likewise, the binding of N protein to leader RNA has been implicated as a means for preferential translation of viral mRNAs (Tahara, Dietlin et al. 1998), although data supporting this attractive hypothesis are, as yet, incomplete. N protein has also been found to enhance the efficiency of replication of replica or genomic RNA in
reverse genetic systems in which infections are initiated from engineered viral RNA (Thiel, Herold et al. 2001a; Yount, Denison et al. 2002; Almazan, Galan et al. 2004; Schelle, Karl et al. 2005). This may be indicative of a direct role of N in RNA replication, but it remains possible that the enhancement actually results from the sustained translation of a limiting replicase component.

Finally, studies have shown that, in addition to its presence in the cytoplasm, IBV N protein localized to the nucleoli of about 10% of cells that were infected with IBV or were independently expressing N protein (Hiscox, Mawditt et al. 1995). This observation was extended to the N proteins of MHV and TGEV, suggesting that nucleolar localization is a general feature of all three coronavirus groups. Such localization was proposed to correlate with the arrest of cell division (Wurm, Chen et al. 2001). Additionally, both MHV and IBV N proteins were found to bind to two nucleolar proteins, fibrillarin and nucleolin (Chen, Gill et al. 2005). However, the nucleolar localization of N was not observed in TGEV-infected or SARS-CoV-infected cells by other groups of investigators (Calvo, Escors et al. 2005; Rowland, Chauhan et al. 2005). All steps of coronavirus replication are thought to occur outside of the nucleus. For MHV, several studies have shown that viral replication could occur in enucleated cells or in cells treated with actinomycin D or α-amanitin, host RNA polymerase inhibitors (Brayton, Ganges et al. 1981; Wilhelmsen, Leibowitz et al. 1981). By contrast, other studies reported that similar conditions reduced the growth yield of IBV, HCoV-229E, or FCoV (Kennedy and Johnson-Lussenburg 1979; Evans and Simpson 1980a; Lewis, Harbour et al. 1992). Even though coronavirus replication does not have an absolute dependence on the nucleus, the possibility remains that some viruses can alter host nuclear functions so as to create an environment more favourable for viral infection. Such a modification might be brought about through the nuclear trafficking of one or more viral components (Masters 2006).
1.4 VIRAL REPLICATION CYCLE, ASSEMBLY AND PROTEIN INTERACTIONS

Coronavirus infections are initiated by the binding of virions to cellular receptors (Figure 1.7). This sets off a series of events culminating in the deposition of the nucleocapsid into the cytoplasm, where the viral genome becomes available for translation. The positive-sense genome, which also serves as the first mRNA of infection, is translated into the enormous replicase polyproteins. The replicase then uses the genome as the template for the synthesis, via negative-strand intermediates, of both progeny genomes and a set of sub-genomic mRNAs. The latter are translated into structural proteins and accessory proteins. The membrane-bound structural proteins, M, S, and E, are inserted into the ER, from where they transit to the endoplasmic reticulum-Golgi intermediate compartment (ERGIC). Nucleocapsids are formed from the encapsidation of progeny genomes by N protein, and these coalesce with the membrane-bound components, forming virions by budding into the ERGIC. Finally, progeny virions are exported from infected cells by transport to the plasma membrane in smooth-walled vesicles, or Golgi sacs, that remain to be more clearly defined (Masters 2006).

During infection by some coronaviruses, but not others, a fraction of S protein that has not been assembled into virions ultimately reaches the plasma membrane. At the cell surface S protein can cause the fusion of an infected cell with adjacent, uninfected cells, leading to the formation of large, multinucleate syncytia. This enables the spread of infection independent of the action of extracellular virus, thereby providing some measure of escape from immune surveillance (Masters 2006).
1.4.1 Virion Assembly and Protein Interactions

Once the full program of viral gene expression is underway, through transcription, translation, and genome replication, progeny viruses can begin to assemble. Coronavirus virion assembly occurs through a series of cooperative interactions that occur in the ER and the ERGIC among the canonical set of structural proteins, S, M, E, and N. The M protein participates in most, if not all, of these interactions and has come to be recognized as the central organizer of the assembly process.

Despite its dominant role, however, M protein alone is not sufficient for virion formation. Independent expression of M protein does not result in its assembly into virion-like structures. Under these circumstances, M was shown to traverse the secretory pathway as far as the trans-Golgi (Klumperman, Locker et al. 1994), where it forms large,
detergent-insoluble complexes (Weisz, Swift et al. 1993; Locker, Opstelten et al. 1995). By contrast, MHV, IBV, TGEV, and FIPV, which are representative species from each of the three coronavirus groups were found to bud into a proximal compartment, the ERGIC (Tooze, Tooze et al. 1988; Klumperman, Locker et al. 1994; Krijnse-Locker, Ericsson et al. 1994). These observations suggested that some factor, in addition to M, must determine the site of virion assembly and budding.

The identification of the unknown factor came from the development of virus-like particle (VLP) systems for coronaviruses. Such studies showed that, for MHV, co-expression of both M protein and the minor virion component, E protein, was necessary and sufficient for the formation of particles (Bos, Luytjes et al. 1996; Vennema, Godeke et al. 1996). The resulting VLPs were morphologically identical to virions (minus spikes) and were released from cells by a pathway similar to that used by virions. Notably, neither the S protein nor the nucleocapsid was found to be required for VLP formation. These results were subsequently generalized for coronaviruses from all three groups: BCoV and TGEV (Baudoux, Carrat et al. 1998a), IBV (Corse and Machamer 2000; Corse and Machamer 2003), and SARS-CoV (Mortola and Roy 2004a). Currently, there is one known exception to this trend: in a separate study of SARS-CoV, M and N proteins were reported to be necessary and sufficient for VLP formation, whereas E protein was dispensable (Huang, Yang et al. 2004a). This latter contradiction remains to be resolved. It may reflect a unique aspect of SARS-CoV virion assembly, or, alternatively, it may indicate that VLP requirements can vary with different expression systems (Masters 2006).
1.4.1.1 M Protein–M Protein Interactions

Since VLPs contain very little E protein, it is assumed that lateral interactions between M protein monomers are the driving force for virion envelope formation. These interactions have been explored through examination of the ability of constructed M protein mutants to support or to interfere with VLP formation. A study that tested the structural requirements of the M protein found that mutations either in the ectodomain, any of the three transmembrane domains, or in the carboxy-endo domain/endodomain, could inhibit or abolish VLP formation (de Haan, Kuo et al. 1998a). In particular, the carboxyl terminus of M was extremely sensitive to small deletions or even to point mutations of the final residue of the molecule. Construction of many of these latter mutations in the viral genome revealed a consistent set of effects on viral viability. Yet, virions were better able than VLPs to tolerate carboxyl-terminal alterations in M protein, presumably because virions were stabilized by additional intermolecular interactions not present in VLPs. In experiments in which both wild-type and mutant M proteins were co-expressed with E protein, wild-type M protein was able to rescue low concentrations of assembly-defective mutant M proteins into VLPs (de Haan, Kuo et al. 1998a). This finding, coupled with results from co-immunoprecipitation analyses, provided the basis for further work, which concluded that monomers of M interact via multiple contacts throughout the molecule and particularly in the transmembrane domains (de Haan, Vennema et al. 2000).

1.4.1.2 S Protein–M Protein Interactions

That VLPs could be formed in the absence of S protein (Bos, Luytjes et al. 1996; Vennema, Godeke et al. 1996) confirmed the much earlier discovery that treatment of MHV-infected cells with the glycosylation inhibitor tunicamycin led to the assembly and
release of spike less (and consequently, non-infectious) virions (Holmes, Doller et al. 1981; Rottier, Horzinek et al. 1981). These findings were also consistent with the properties of certain classical temperature-sensitive mutants of MHV and IBV, which, owing to S gene lesions, failed to incorporate spikes into virions at the non-permissive temperature (Ricard, Koetzner et al. 1995; Lai and Cavanagh 1997; Luytjes, Gerritsma et al. 1997; Shen, Law et al. 2004). Independently expressed MHV, FIPV, or IBV S proteins enter the default secretory pathway and ultimately reach the plasma membrane (Vennema, Heijnen et al. 1990a).

In the presence of M protein, however, a major fraction of S is retained in intracellular membranes, as was shown by coimmunoprecipitation of S and M proteins from MHV infected cells (Opstelten, Raamsman et al. 1995a). Moreover, the interaction of M with S was demonstrated to be specific; complexes of M did not impede the progress of a heterologous glycoprotein (the VSV G protein) to the plasma membrane. Additionally, kinetic experiments revealed that the folding and oligomerization of S protein in the ER is rate limiting in the M–S interaction, in which nascent M protein immediately participates (Opstelten, Raamsman et al. 1995a). Complexes of the M and S proteins were similarly observed in BCoV-infected cells, for which it was found that M also determines the selection of HE protein for incorporation into virions (Nguyen and Hogue 1997). The simplest conclusion that can be drawn from all this evidence, then, is that S protein is entirely passive in assembly but becomes trapped by M protein upon passage through the ER (Masters 2006). Nevertheless, there are indications that, in some cases, S cooperates in its own capture. The possession of endo H resistance, in independently expressed S protein was found to be transported to the cell surface with much slower kinetics than S protein that was incorporated into virions (Masters 2006). This led to the proposal that free S protein harbours intracellular retention signals that become hidden during virion assembly.
(Vennema, Heijnen et al. 1990a). Such signals have been found in the (group 3) IBV S protein cytoplasmic endodomain, which contains both a dilysine motif that was shown to specify retention in the ERGIC and a tyrosine-based motif that causes retrieval by endocytosis from the plasma membrane (Lontok, Corse et al. 2004).

Additionally, a novel dibasic ERGIC retention signal was identified in the S protein endodomains of group 1 coronaviruses (TGEV, FIPV, and HCoV-229E) and SARS-CoV, but not other group 2 coronaviruses, such as MHV and BCoV (Masters 2006). Although the S protein is not required for VLP formation, it does become incorporated into VLPs if it is co-expressed with the M and E proteins (Bos, Luytjes et al. 1996; Vennema, Godeke et al. 1996). VLP manipulations thus made it possible to begin to dissect the molecular basis for the specific selection of S protein by M protein. As for M–M homotypic interactions, the sites within M protein that bind to S protein have not yet been pinpointed. On a broader scale, deletion mapping has indicated that the ectodomain of M protein and the carboxyl-terminal 25 residues of the endodomain do not participate in interactions with S, even though both of these regions are critical for VLP formation (de Haan, Smeets et al. 1999). The residues of S protein that interact with M protein, on the other hand, have been much more precisely localized. This mapping began with the swapping of ectodomains between the very divergent S proteins of MHV (Kuo, Godeke et al. 2000) and FIPV (Godeke, de Haan et al. 2000). This type of exchange showed that the incorporation of S protein into VLPs of a given species was determined by the presence of merely the transmembrane domain and endodomain of S protein from the same species. The source of the S ectodomain did not matter. The assembly competence of the 1324-residue MHV S protein or the 1452-residue FIPV S protein was therefore restricted to just the 61-amino-acid, carboxyl-terminal region of each of these molecules. That the domain switched S molecules were completely functional was demonstrated by the construction of
an MHV mutant, designated fMHV, in which the ectodomain of the MHV S protein was replaced by that of the FIPV S protein (Kuo, Godeke et al. 2000). As predicted, this mutant gained the ability to grow in feline cells, while losing the ability to grow in mouse cells.

The fMHV chimera provided the basis for powerful selections, based on host cell species restriction, that have been used with the reverse genetic system of targeted RNA recombination (Masters 1999; Kuo and Masters 2002; Masters and Rottier 2005). The converse construct, an FIPV mutant designated mFIPV, in which the ectodomain of the FIPV S protein was replaced by that of the MHV S protein, had properties exactly complementary to those of fMHV (Haijema, Volders et al. 2003).

More detailed dissection of the transmembrane domain and endodomain of the MHV S protein has been carried out to further localize the determinants of S incorporation into virions (Ye, Montalto-Morrison et al. 2004; Bosch, Martina et al. 2004a). In one study, the S protein transmembrane domain, or the endodomain, or both, were swapped with the corresponding region(s) of a heterologous transmembrane protein, which was expressed as an extra viral gene product (Ye, Montalto-Morrison et al. 2004). Mutations were constructed in this surrogate virion structural protein, or, alternatively, directly in the S protein. From this work, the virion assembly property of S was found to map solely to the 38-residue endodomain, with a major role assigned to the charge-rich, carboxyl-terminal region of the endodomain. Additionally, it was observed that the adjacent, membrane-proximal, cysteine-rich region of the endodomain was critical for cell–cell fusion during infection, consistent with results previously reported from investigations using S protein expression systems (Bos, Heijnen et al. 1995; Chang, Sheng et al. 2000). A second study, based on analysis of a progressive series of carboxyterminal truncations of the S protein in VLPs and in viral mutants, also mapped the virion assembly competence
of S to the endodomain (Bosch, de Haan et al. 2005). In this work, however, the major role in assembly was attributed to the cysteine-rich region of the endodomain, and the overall size, rather than the sequence of the endodomain, was seen to be critical. Therefore, the precise nature of the interaction between the S protein endodomain and the M protein remains to be resolved (Masters 2006).

1.4.1.3 N Protein–M Protein Interactions

The interaction of the viral nucleocapsid with M protein was originally examined by the fractionation of purified MHV virions (Sturman, Holmes et al. 1980). At 4°C, the M protein was separated from other components on density gradient centrifugation of NP-40-solubilized virion preparations, but M reassociated with the nucleocapsid when the temperature was elevated to 37°C. Further analysis suggested that, contrary to expectations, this temperature-dependent association was mediated by M binding to viral RNA, rather than to N protein (Masters 2006).

The notion of M protein as an RNA-binding protein has been revived in light of recent results on the mechanism of genome packaging (Narayanan, Chen et al. 2003a). For TGEV virions, the use of particular low-ionic-strength conditions of NP-40 treatment similarly demonstrated that a fraction of M protein was persistently integrated with subviral cores (Risco, Anton et al. 1996). For assay of this association, in vitro translated M protein was bound to immobilize nucleocapsid purified from virions (Escors, Ortego et al. 2001a). Through the combined approaches of deletion mapping, inhibition by antibodies of defined specificity, and peptide competition, the M-nucleocapsid interaction was localized to a segment of 16 residues adjacent to the carboxyl terminus of the 262-residue TGEV M protein (Masters 2006).
Studies of MHV have taken genetic avenues to explore the N protein–M protein interaction. In one report, a viral mutant was constructed in which the carboxyl-terminal two amino acids of the 228-residue MHV M protein were deleted (Kuo and Masters 2002), a lesion previously known to abolish VLP formation (de Haan, Kuo et al. 1998a). The resulting highly impaired virus, designated MΔ2, formed tiny plaques and grew to maximal titers many orders of magnitude lower than those of the wild type. Multiple independent second-site revertants of the MΔ2 mutant were isolated and mapped to either the carboxyl terminus of M or that of N. Reconstruction of some of these compensating mutations, in the presence of the original MΔ2 mutation, provided evidence for a structural interaction between the carboxyl termini of the M and the N proteins. In a complementary analysis, a set of viral mutants were created containing all possible clustered charged-to-alanine mutations in the carboxyl-terminal domain 3 of the N protein (Hurst, Kuo et al. 2005). One of the members of this set, designated N-CCA4 was extremely defective, having a phenotype similar to that of the MΔ2 mutant. Multiple independent second-site suppressors of N-CCA4 were found to map in the carboxyl-terminal region of either the N or the M protein, thereby reciprocating the genetic cross-talk uncovered with the MΔ2 mutant. Additionally, it was shown that the transfer of N protein domain 3 to a heterologous protein allowed incorporation of that protein into MHV virions (Masters 2006).

1.4.1.4 Role of E Protein

In contrast to the more overt structural roles of the M, S, and N proteins, the role that is played by E protein in assembly is enigmatic. On discovery of the essential nature of E in VLP formation, it was speculated that the low amount of E protein in virions and
VLPs indicated a catalytic, rather than structural, function for this factor. E protein might serve to induce membrane curvature in the ERGIC, or it might act to pinch off the neck of the viral particle in the final stage of the budding process (Vennema, Godeke et al. 1996). In pursuit for evidence correlating the VLP findings to the situation in whole virions, a set of clustered charged-to-alanine mutations were constructed in the E gene of MHV. One of the resulting mutants was markedly thermo-labile, and its assembled virions had striking morphologic defects, exhibiting pinched and elongated shapes that were rarely seen among wild-type virions (Fischer, Stegen et al. 1998a). This phenotype clearly supported a critical role for E protein in virion assembly. Surprisingly, however, it was later found to be possible to entirely delete the E gene from the MHV genome, although the resulting ΔE mutant virus was only minimally viable compared to the wild type (Kuo and Masters 2003). This indicated that, for MHV, the E protein is important, but not absolutely essential, to virion assembly. By contrast, for TGEV, two independent reverse genetic studies showed that knockout of the E gene was lethal (Kuo and Masters 2003).

Viable virus could be recovered only if E protein was provided in trans (Curtis, Yount et al. 2002a; Ortego, Escors et al. 2002a). This discordance may point to basic morphogenic differences between group 2 coronaviruses (such as MHV) and group 1 coronaviruses (such as TGEV). Alternatively, it is possible that E protein has multiple activities, one of which is essential for group 1 coronaviruses but is largely dispensable for group 2 coronaviruses. The information available about E protein at this time is not sufficiently complete to allow us to understand the function of this tiny molecule. One of the most intriguing questions is whether it is necessary for E protein to directly physically interact with M protein, or whether E acts at a distance. If E protein has multiple roles, then perhaps both of these possibilities are applicable. Direct interaction between the E and M proteins is implied by the observation that, at least in some cases, co-expression of E
and M proteins from different species does not support VLP formation (Baudoux, Carrat et al. 1998a). The demonstration that IBV E and M can be cross-linked to one another also has established that the two proteins are in close physical proximity in infected or transfected cells (Corse and Machamer 2003). Contrary to this, some data appear to argue that E acts independently of M. The specific expression of MHV or IBV E protein resulted in membrane vesicles that are exported from cells (Maeda, Maeda et al. 1999; Corse and Machamer 2000).

Additionally, Raamsman et al. (2000) has shown that the expression of MHV E protein alone leads to the formation of clusters of convoluted membranous structures highly similar to those seen in coronavirus-infected cells (Raamsman, Locker et al. 2000). This suggests that the E protein, without other viral proteins, acts to induce membrane curvature in the ERGC. Some indirect evidence also seems to indicate that E does not directly contact other viral proteins. Multiple revertant searches with E gene mutants failed to identify any suppressor mutations that map in M or in any gene other than E (Fischer, Stegen et al. 1998a). Similarly, none of the intergenic suppressors of the MΔ2 mutant mapped to the E gene (Kuo and Masters 2002). It has been found that the SARS-CoV E protein forms cation-selective ion channels in a model membrane system (Wilson, McKinlay et al. 2004). Moreover, this channel-forming property was contained in the amino-terminal 40 residues of the 76-residue SARS-CoV E molecule. Such an activity made be the basis for an independent mode of action of E protein.
1.5 DISCOVERY OF HCoV-NL63

In January 2003, a 7–month old child was brought to an Amsterdam hospital with bronchiolitis. Diagnostic tests for all known respiratory viruses were done and they were negative, but a cytopathic effect on the Rhesus monkey kidney epithelial cells (LLC-MK2) was evident. Using the virus discovery cDNA-amplified fragment length polymorphism (AFLP) (VIDISCA) method, a novel virus was discovered and the complete genome of the virus (named HCoV-NL63) showed that it was a novel group 1 human coronavirus (van der Hoek, Pyrc et al. 2004).

Shortly thereafter, a second research group in the Netherlands reported detection of essentially the same virus. Fouchier et al. (2005), described a virus (which they named HCoV-NL) in a Vero-E6 cell culture supernatant that was originally obtained in 1988 from an 8 month old boy suffering from pneumonia (Fouchier, Hartwig et al. 2005). The similarity to the previously described HCoV-NL63 strain was very high (98.8% at the nucleotide level) and it could be concluded that these two virus isolates represented the same species. Most of the differences between the two virus variants are clustered in the amino terminal region of the spike protein. HCoV-NL63 forms a sub-cluster with HCoV-229E, PEDV (porcine epidemic diarrhea virus), and Bat-CoV (Fujimoto, Masuzaki et al. 2005), which was assigned group Ib.

Almost a year later, a third group described the same human coronavirus (Esper, Weibel et al. 2005b). Using universal coronavirus primers, patient samples were identified with coronaviruses that did not match at the nucleotide level with HCoV-229E, HCoV-OC43 or SARS-CoV. These authors gave their virus the name ‘New Haven coronavirus’ (HCoV-NH), although the partial sequences of their isolates clearly showed that the novel coronaviruses identified in New Haven, Connecticut, United States of America (USA),
were very similar to the isolates from the Netherlands (94–100% identical at nucleotide level) and thus represented the same species (Esper, Weibel et al. 2005b).

1.5.1 Infection with HCoV-NL63

1.5.1.1 Worldwide Occurrence

The first two major studies on the worldwide occurrence of HCoV-NL63 clearly indicated that infection by this virus was not a rare event (Fouchier, Hartwig et al. 2005; Gondo, Hirose et al. 2005). Both reports presented several HCoV-NL63 infected patients that suffered from upper or lower respiratory tract illnesses. Since then a number of reports have shown that the virus is not unique to the Netherlands, and that it could be found worldwide. In Australia, HCoV-NL63 was detected in 16 of 766 patients with acute respiratory disease (2.1%) (Arden, Nissen et al. 2005). A Japanese study on 118 children under the ages of 2 that suffered from bronchiolitis and who were admitted to the hospital revealed three positive patients (2.5%) (Ebihara, Endo et al. 2005a), and a second Japanese study presented five HCoV-NL63 infections among 419 specimens collected from children with respiratory infections (1.2%) (Wu, Ohsaga et al. 2005). Amongst 279 hospitalized children with upper and lower respiratory tract illness in Belgium, seven children were identified with HCoV-NL63 (2.3%) (Horm, Gutierrez et al. 2012).

In France, 28 of 300 patients with upper and lower respiratory tract illness under the age of 20 harbored HCoV-NL63 RNA (9.3%) and in a Swiss cohort of neonates, HCoV-NL63 was identified in six of 82 cases (7%) during their first episode of lower respiratory tract illness (Imai, Kuba et al. 2005; Suzuki, Okamoto et al. 2005). Two studies by a Canadian group reported 45 NL63-positive patients of a total of 1765 specimens that
were tested (2.5%) (Bastien, Anderson et al. 2005a; Bastien, Robinson et al. 2005b). Seven of these Canadian patients were involved in an outbreak of acute respiratory-tract infection in a personal care home for the elderly (Bastien, Anderson et al. 2005a).

HCoV-NL63 was also detected in Hong Kong in 15 of 587 (2.6%) children (Chiu, Chan et al. 2005). The children that participated in this prospective clinical and virological study were under the age of 18 with acute respiratory tract infection. This study represented the Hong Kong population, so it could be estimated that HCoV-NL63 infection causes more than 200 hospital admissions each year per 100 000 children under the age of 6. In a PRI.DE study (Pediatric Respiratory Infection in Germany) of 949 children with lower respiratory tract illness, HCoV-NL63 was detected in 49 children (5.2%) (van der Hoek, Pyrc et al. 2004). Seemingly, patients with HCoV-NL63 range in age from 1 month to 100 years, however in all studies the proportion of positive specimens is the highest in the 0–5-year group. HCoV-NL63 infection is frequently observed in patients with an underlying disease, whether they are immune-compromised due to chemotherapy or HIV infection or have another medical condition such as chronic obstructive pulmonary disease (COPD) (Arden, Nissen et al. 2005; Fouchier, Hartwig et al. 2005; Bastien, Robinson et al. 2005b; Horm, Gutierrez et al. 2012). An infection of children, the elderly or immuno-compromised individuals, regularly requires hospitalization due to the severity of respiratory symptoms.

1.5.1.2 Epidemiology of HCoV Infections in Children

There is a profound difference between the epidemiology of the infections caused by SARS-CoV and that of all other HCoV infections. SARS-CoV emerged in November 2002 and was successfully eradicated in April 2004 (Blakqori, Lowen et al. 2012). During
those 18 months, SARS-CoV was isolated in several countries, some of which were very
distant from each other. However, the total of 8,098 cases of SARS diagnosed worldwide
(Blakqori, Lowen et al. 2012) is substantially fewer than the number usually found during
epidemics of the most common respiratory viruses, such as respiratory syncytial virus
(RSV) and influenza viruses (Esposito, Bosis et al. 2006). Furthermore, sero-
epidemiologic studies of high-risk and low-risk residential areas have clearly revealed that
the prevalence of immunoglobulin G against SARS-CoV was low in children and adults;
this result indicates that SARS-CoV not only had a restricted period of circulation but also
that it had limited spread (Maliogka, Calvo et al. 2012).

Proportionally fewer children were involved: <5% of all cases were diagnosed in
patients <18 years of age (Blakqori, Lowen et al. 2012). The properties of SARS-CoV and
its low level of transmissibility seem to be the main reasons for the low risk for contagion
in children. In most of the areas in which outbreaks occurred, healthcare workers and adult
patients were mainly involved and, because they were immediately hospitalized, the risk
for the infection spreading to children was greatly reduced because they are not usually
permitted to visit hospitals (Blakqori, Lowen et al. 2012). This hypothesis seems to be
further supported by the fact that the early detection and isolation of symptomatic patients
were the most important measures in controlling the SARS epidemic. Unlike SARS-CoV,
HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1 have been in continuous
circulation since their first isolation and every year cause a large number of infections that
more frequently involve children than adults (Monto 1974; van der Hoek, Pyrc et al. 2004;
Ying, Hao et al. 2004; van der Hoek, Sure et al. 2005; Wu, Chang et al. 2008; van der
Hoek, Ihorst et al. 2010). In particular, the data concerning the earlier-appearance HCoVs
indicates that they are distributed throughout the world and mainly circulate during the
winter and early spring, with outbreaks occurring every 2–4 years (Monto and Lim 1974).
McIntosh et al. (1967) found that the global incidence of LRTIs due to HCoV-229 and HCoV-OC43 in hospitalized children was no higher than 3.8% (McIntosh, Becker et al. 1967; McIntosh, Dees et al. 1967; McIntosh, Chao et al. 1974b). Later studies have shown that when all respiratory infections are considered, the etiologic prevalence of these pathogens in pediatric patients can be significantly higher, varying from ~5% to >30% (Fouchier, Hartwig et al. 2005; Vabret, Mourez et al. 2005).

An overall evaluation of the available data suggests that these distinctions can be attributed to various research methods. The infections caused by these viruses are more prevalent in November through March, frequently affect children <5 years of age those examined in the community, and those without underlying risk factors; and are more often identified with serologic methods (Shao, Guo et al. 2007; Dijkman, Jebbink et al. 2008). In this regard, two recent surveys that used serologic methods alone found previous HCoV-229E infection in 42.9%–50.0% of children 6–12 months of age (Shao, Guo et al. 2007) and in 65% of those 2.5–3.5 years of age (Dijkman, Jebbink et al. 2008). Similarly, in 75% of the cases, children 2.5–3.5 years of age had antibodies against HCoV-NL63 (Dijkman, Jebbink et al. 2008). All these factors can explain why the lowest incidences of HCoV-229E and HCoV-OC43 infections are usually found when the study population includes older children or adolescents, patients with underlying severe chronic diseases or hospitalized patients, when only highly symptomatic infections are considered, and when the study is conducted during the whole year. Moreover, what has been clearly shown is that the original HCoVs are commonly detected in childhood and frequently isolated in the nasopharyngeal secretions of children with respiratory infection. In some cases, co-infections with other respiratory viruses, mainly RSV, influenza viruses, and human metapneumovirus, have also been found. However, the real incidence of HCoV-229E and HCoV-OC43 co-infections with other respiratory pathogens has not yet been defined.
because only a few of the published studies were planned to identify all the main respiratory viruses. Similar conclusions can be drawn in relation to the more recently identified HCoVs. HCoV-NL63, which can be found in 1.0%–9.3% of nasopharyngeal aspirates from patients with RTIs, circulates throughout the world (predominantly during the winter in temperate regions), infects mainly younger children and subjects with underlying severe chronic diseases, and is more frequently found in non-hospitalized children (Arden, Nissen et al. 2005; Choi, Lee et al. 2006; Esposito, Bosis et al. 2006; de Souza Luna, Heiser et al. 2007; Kuypers, Martin et al. 2007; Smuts, Workman et al. 2008; Wu, Chang et al. 2008). Although it was not isolated until 2000, HCoV-NL63 has probably been circulating for some time because one of the first detections occurred in a sample of nasopharyngeal secretions collected from a child with pneumonia that had been kept in a freezer since January 2003 before evaluation (van der Hoek, Pyrc et al. 2004). HCoV-HKU1 was identified in 2005 and has once again been found in nasopharyngeal secretions of children and adults with respiratory infections in countries that are very distant from each other. Its incidence varies from <1% to 6% and sero-epidemiologic surveys based on antibodies reacting with the recombinant HKU1 nucleocapsid protein suggest that infection may be relatively common in humans, although generally asymptomatic (Woo, Lau et al. 2005b; Daubin, Parienti et al. 2006; Shao, Guo et al. 2007; Dijkman, Jebbink et al. 2008). HCoV-NL63 and HCoV-HKU1 are often associated with co-infections with other respiratory viruses, mainly RSV and influenza viruses (van der Hoek, Sure et al. 2005; Lau, Woo et al. 2006; de Souza Luna, Heiser et al. 2007; Wu, Chang et al. 2008).
1.5.1.3 Seasonality of infection

The HCoV-OC43 and HCoV-229E viruses tend to be transmitted predominantly in the winter season in temperate climate countries (Hemming, Rodriguez et al. 1987). The seasonality of HCoV-NL63 in the Netherlands shows the same preference for the winter season. Studies in Belgium, Australia, Japan, Canada, Germany and France also report a winter predominance of this virus (McIntosh, Becker et al. 1967; McIntosh, Dees et al. 1967; Arden, Nissen et al. 2005; Suzuki, Okamoto et al. 2005; Bastien, Anderson et al. 2005a; Ebihara, Endo et al. 2005a; Ebihara, Endo et al. 2005b; Sloots, McErlean et al. 2006). Only in Hong Kong did HCoV-NL63 show a spring-summer peak of activity, whereas in the same study HCoV-OC43 was mostly found in the winter season (Chiu, Chan et al. 2005). This indicates that the seasonality of HCoV-NL63 in tropical and subtropical regions may not be restricted to the winter season.

1.5.1.4 Disease(s) associated with HCoV NL63

The first described cases of HCoV NL63 infection were young children with severe lower respiratory tract illnesses in hospital settings (Fouchier, Hartwig et al. 2004; van der Hoek, Pyrc et al. 2004; Vabret, Mourez et al. 2005). One of the HCoV-NL63 infected elderly Canadian patients died 5 days after the onset of disease (Bastien, Anderson et al. 2005a), demonstrating how lethal it be. All studies on the incidence of HCoV-NL63 infection have used respiratory clinical material from patients suffering from respiratory tract illnesses, mostly lower respiratory tract illnesses. It is therefore not surprising that the symptoms associated with NL63-infection represent respiratory diseases. Nevertheless, one can provide an initial overview of the symptoms related to HCoV-NL63 infection by focusing on patients without a secondary infection. Van der Hoek et al (2004) summarizes
the symptoms in patients with an HCoV-NL63 infection (van der Hoek, Pyrc et al. 2004). These results were obtained from various countries and are derived only from studies that did not select for certain clinical symptoms, because selection for a respiratory disease (e.g. bronchiolitis or lower respiratory tract illness) would bias the overview towards these clinical symptoms. As mentioned above, a substantial number of patients suffered from lower respiratory tract illnesses that required hospitalization. On the other hand, some of the patients showed relatively mild symptoms like fever, cough, sore throat and rhinitis (Bastien, Anderson et al. 2005a). For instance, only 16 of the 45 patients with HCoV-NL63 in Canada had been hospitalized because of the respiratory illness, demonstrating that infection with HCoV-NL63 does not always require hospitalization (Bastien, Anderson et al. 2005a; Bastien, Robinson et al. 2005b).

In a PRI.DE (Pediatric Respiratory Infection in Germany) study, which investigated children with lower respiratory tract illness, HCoV-NL63 infection was also more often detected in outpatients than hospitalized patients, confirming that HCoV-NL63 generally causes less severe problems (Vabret, Dina et al. 2006a). This suggests strongly that HCoV-NL63, similar to HCoV-229E and HCoV-OC43, is a common cold virus (Lim and Liu 1998) and that HCoV-NL63, like HCoV- 229E and HCoV-OC 43 can cause a more severe clinical picture in young children, the elderly and immuno-compromised individuals (van der Hoek, Pyrc et al. 2004; McIntosh 2005). A link between HCoV-NL63 and respiratory diseases was established in the (PRI.DE study) on lower respiratory tract illness in children less than 3 years of age (Kaiser, Regamey et al. 2005; Esper, Weibel et al. 2005b). That study demonstrated a strong association between laryngotracheitis (croup) and HCoV-NL63 infection. Croup is an inflammation of the trachea and is characterized by a loud barking cough that usually worsens at night. Croup patients are not often hospitalized, and even less frequently sampled for respiratory virus diagnostics. For this
reason, croup patients are under-represented in most studies that analyzed hospitalized RTI patients. The PRI.DE study and the Hong Kong study were the only population based studies, thus excluding a bias in patient selection. Both studies observed the high frequency of croup among the HCoV-NL63 positive patients (43% and 27%, respectively) (van der Hoek, Pyrc et al. 2004; Chiu, Chan et al. 2005). Furthermore, HCoV-NL63 has also been associated with Kawasaki disease which is one of the most common forms of childhood vasculitis (Burns 2001; Esper, Shapiro et al. 2005a). Respiratory secretions from 11 children with Kawasaki disease were tested and eight samples were positive for HCoV-NL63 (73%). HCoVNL63 was detected in only one subject of a control group of 22 patients matched for age with the Kawasaki disease patients who visited the hospital in the same period (Esper, Shapiro et al. 2005a). It presents with prolonged fever and a polymorphic exanthem, oropharyngeal erythema, and bilateral conjunctivitis. Scarlet fever and measles are the closest clinical mimics. A number of epidemiological and clinical observations suggested previously that an infectious agent might be the cause of Kawasaki disease (Esper, Shapiro et al. 2005a).

These observations include the acute self-limiting nature of the febrile illness, the peak of incidence in patients under 18 months, the geographic clustering of outbreaks and the seasonal predominance of the illness in late winter/early spring (Ksiazek, Erdman et al. 2003). Coronary complications can be reduced significantly by the use of intravenous immunoglobulin therapy combined with oral aspirin (Burns and Glode 2004). The link between The association of HCoV-NL63 and Kawasaki disease is questionable due to the fact that several research groups have been unable to confirm the findings of Esper et al (2005) (Reed 1984; Belay, Erdman et al. 2005; Esper, Shapiro et al. 2005a; Ebihara, Endo et al. 2005b; Chang, Chiang et al. 2006).
1.5.1.5 Transcription and replication of HCoV-NL63

Recently, an infectious cDNA clone of HCoV-NL63 was discovered by Donaldson et al, 2008 (Donaldson, Yount et al. 2008). This study entailed the utilization of a reverse genetics system for NL63, whereby five contiguous cDNAs that span the entire genome were used to generate a full-length cDNA. Recombinant NL63 viruses which contained the expected marker mutations were efficiently replicated as the wild-type NL63 (wtNL63) virus. Furthermore, they engineered a heterologous green fluorescent protein gene in place of ORF3 of the NL63 clone, simultaneously creating a unique marker for NL63 infection and demonstrating that the ORF3 protein product is nonessential for the replication of NL63 in cell culture (Donaldson, Yount et al. 2008). The genome of HCoV-NL63 contains important signals that are needed for efficient replication. Translation of the genomic RNA generates one large polyproteins encoded by the 1a gene, and a 1 ribosomal frame shift results in the production of the 1ab polyproteins (Pyrc, Jebbink et al. 2004; van der Hoek, Sure et al. 2005; van der Hoek, Pyrc et al. 2006). A potential pseudoknot RNA structure that could facilitate the frame shift for HCoV-NL63 is located within the RNA dependent RNA polymerase gene. The predicted pseudoknot structure consists of a hairpin with a highly conserved 11-base pair stem, and a loop of eight nucleotides that forms five base pairs with a sequence 167 nucleotides further downstream. The heptanucleotide UUUAAAC sequence is present just upstream of the pseudoknot and probably acts as the slippery sequence where the actual ribosomal frame shift occurs (Pyrc, Jebbink et al. 2004).

Coronaviruses utilize posttranslational proteolytic processing as a key regulatory mechanism in the expression of their replicative proteins (van der Hoek, Pyrc et al. 2006). The HCoV-NL63 1a and 1ab polyproteins are potentially cleaved by viral proteases to
facilitate the assembly of a multi-subunit protein complex that is responsible for viral replication and transcription. The genome of HCoV-NL63 is predicted to encode two proteinases in the 50 region of the 1a polyproteins (Pyrc, Jebbink et al. 2004). First, a papain-like proteinase (PL\textsuperscript{pro}) is expressed by the non-structural protein (nsp) 3 gene situated near the 50 end of the genome. This putative papain-like proteinase of HCoV-NL63 consists of two domains – PL\textsubscript{1}\textsuperscript{pro} and PL\textsubscript{2}\textsuperscript{pro} – and both are expected to have catalytic activity (Pyrc, Jebbink et al. 2004).

The enzyme is predicted to cleave the 1a/1b protein at three sites between nsp1|nsp2, nsp2|nsp3, and nsp3|nsp4, releasing the functional papain-like proteinase protein (nsp3) molecule by auto-cleavage. Analysis of the predicted cleavage sites of PL\textsuperscript{pro} indicates that this HCoV-NL63 enzyme has the specificity to cut between two small amino acids with short uncharged side chains, similar to homologous enzymes in other coronaviruses (Lim and Liu 1998). The nsp5 gene is predicted to encode the second proteinase of HCoV-NL63. It contains a serine-like proteinase domain and is designated the main proteinase (M\textsuperscript{pro}) to stress its important function. The putative NL63-M\textsuperscript{pro} processes the majority of cleavage sites between non-structural proteins in the 1a/1b polyproteins (Pinon, Teng et al. 1999). Analysis of the M\textsuperscript{pro} cleavage sites suggests that the substrate specificity of NL63-M\textsuperscript{pro} is different from that of the other coronaviruses. The cleavage site between nsp13 and nsp14 contains a histidine residue at position P1 instead of the standard glutamine that is present in other coronavirus species. The histidine at P1 is present in all HCoV-NL63 isolates sequenced thus far (Amsterdam 1, Amsterdam 57, Amsterdam 496 and NL). Additionally, the recently discovered HCoV-HKU1 virus seems to share this unusual cleavage site (van der Hoek, Sure et al. 2005). HCoV-NL63 genome replication generates several sub-genomic mRNAs for the spike, ORF3, envelope, membrane and nucleocapsid protein genes (Pinon, Teng et al. 1999). All sub-genomic
mRNAs have an identical 50 part; the untranslated leader of 72 nt. Creation of sub-genomic mRNA requires pausing at a transcription regulatory sequence immediately upstream of the ORF during negative strand RNA synthesis. Subsequently, the nascent RNA chain switches in a strand transfer event to a transcription regulatory sequence located in the 50 untranslated region of the genome such that a common leader sequence is added to each sub-genomic mRNA (Pinon, Teng et al. 1999).

The transcription regulatory sequence motif of HCoV-NL63 is AACUAAA (Pinon, Teng et al. 1999). This transcription regulatory sequence core sequence is conserved in all sub-genomic mRNA, except for the envelope protein gene that has the suboptimal transcription regulatory sequence core AACUAUA (Pinon, Teng et al. 1999). The sub-genomic mRNA for the nucleocapsid protein is the most abundant and there is an inverse correlation between the distance of a gene to the 30 untranslated regions and its RNA expression level. The envelope protein gene forms a notable exception for which the amount of the sub-genomic mRNA is lower than expected, which may be due to the suboptimal transcription regulatory sequence motif (Pinon, Teng et al. 1999).

1.5.1.6 Antiviral agents

An effective antiviral treatment may be required for HCoV-NL63 infected patients that are admitted to the intensive care unit due to acute respiratory disease (van der Hoek, Pyrc et al. 2006). Several inhibitors are known to reduce replication of at least some coronaviruses including HCoV-NL63 (Pyrc, Jebbink et al. 2004; Sims, Baric et al. 2005). These inhibitors act at various steps of the coronavirus replication cycle, such as receptor binding, membrane fusion, transcription and posttranslational processing. An interesting HCoV-NL63 inhibitor is intravenous immunoglobulin (Pyrc, Jebbink et al. 2004); which is
already approved as an intravenously delivered drug by the Food and Drug Administration. Intravenous immunoglobulin has been used successfully to treat several diseases, mostly primary immune deficiencies and autoimmune neuromuscular disorders, but also respiratory diseases (e.g. RSV) (Hamre and Procknow 1966) and Kawasaki disease (Sims, Baric et al. 2005). Inhibition of viral replication can also occur at the level of fusion of the viral and cellular membranes. The spike protein of HCoV-NL63 contains two heptad repeat regions, HR1 and HR2, situated in the S2 part of the spike protein close to the transmembrane domain. After binding of virus to the receptor, a conformational change leads to the formation of a six-helix bundle containing three HR1s and three HR2s and subsequent exposure of the fusion peptide mediates membrane fusion between the virus and the host cell. For retroviruses, paramyxoviruses and coronaviruses (Bosch, Martina et al. 2004; Pyrc, Jebbink et al. 2004), peptides derived from the HR2 domain can inhibit virus infection, most likely by interacting with HR1. The peptide thus blocks formation of the natural HR1–HR2 interaction, prevents membrane fusion and as a consequence reduces infection. Another novel means to inhibit replication is RNA interference (RNAi) (Groskreutz and Schenborn 1997). The antiviral potential of small interfering RNA (siRNA) targeting HCoV-NL63 was explored by Pyrc and colleagues (Pyrc, Jebbink et al. 2004). The inhibitory properties of two siRNAs targeting conserved sequences within the spike protein gene were analyzed in cell culture infections. Transfection of a relatively low amount of siRNA into HCoV-NL63-susceptible cells made them resistant to virus infection. HCoV-NL63 can also be inhibited at the transcriptional level by pyrimidine nucleoside analogues: b-D-N4-hydroxycytidine and 6-azauridine (Pyrc, Jebbink et al. 2004). As yet, however, the exact mechanism by which these agents inhibit HCoV-NL63 transcription is unclear.
Finally, protease inhibitors act at the level of posttranslational processing. The M\textsuperscript{pro} of coronaviruses has a highly conserved substrate recognition pocket, thus providing the opportunity to design broad-spectrum antiviral drugs against several coronavirus species (Yang, Xie et al. 2005). Yang et al designed main protease inhibitors and measured the inhibitory capacity in an in vitro protease activity assay. One potent inhibitor, N3, showed wide-spectrum inhibition of various M\textsuperscript{pro} enzymes, including the one encoded by HCoV-NL63 (Yang, Xie et al. 2005).

1.7 AIMS OF THIS STUDY

Six human coronaviruses (HCoVs) strains have been described, which are associated with a spectrum of diseases, from mild, febrile upper respiratory tract illnesses to more severe illnesses, croup, bronchiolitis, and pneumonia. With the exception of SARS-CoV, there are other five known human coronaviruses - HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU and HCoV-EMC; continually circulate in the human population.

HCoV-NL63 was identified in 2003 from a nasopharyngeal aspirate of a 7- month-old child suffering from bronchiolitis, conjunctivitis and fever. Many studies of clinical data for this virus have been reported. However, this virus is not well characterized at molecular and protein levels.

Therefore, this study was designed to express the full length nucleocapsid (N) proteins of HCoV-NL63 and SARS-CoV for antibody production in an animal model. Molecular, Proteomic and bioinformatic tools were used characterize the two (HCoV-NL63 and SARS-CoV) N proteins. Polyclonal antibodies were produced and validated using a direct ELISA. This is a preliminary study for development of diagnostic tools for the detection of HCoV-NL63 from patient samples collected in the Western Cape.
1.8 REFERENCES


Cavanagh, D., P. J. Davis, et al. (1986b). "Coronavirus IBV: partial amino terminal sequencing of spike polypeptide S2 identifies the sequence Arg-Arg-Phe-Arg-Arg


CHAPTER TWO

Verification of constructs and expression of proteins

Department of Medical Biosciences, Faculty of Natural Sciences, University of the Western Cape, South Africa.
2.1 ABSTRACT

Historically, coronaviruses were predominantly associated with mild upper respiratory disease in humans. More recently, three novel coronaviruses related to severe human respiratory disease were found, including (i) the severe acute respiratory syndrome coronavirus, responsible for significant atypical pneumonia and 10% mortality; (ii) HKU-1, which is linked to chronic pulmonary disease; and (iii) NL63, associated with both upper and lower respiratory tract disease. Human coronavirus NL63 was first isolated from the aspirate of a 7-month-old baby in early 2004. Infection with HCoV-NL63 has since been shown to be a common occurrence worldwide and has been associated with many clinical symptoms and diagnoses, including severe lower respiratory tract infection, croup and bronchiolitis. HCoV-NL63 causes disease in children, the elderly and the immunocompromised, and has been detected in 1.0–9.3% of respiratory tract infections in children. These crucial discoveries established coronaviruses as important human pathogens and emphasize the need for continued research toward the development of techniques that will enable genetic manipulation of the viral genome, allowing for rapid, rational development and testing of candidate vaccines, vaccine vectors, and therapeutics.

In this study, the NL63-N/pFN2A and SARS-N/pFN2A plasmid constructs were confirmed by restriction endonuclease digest and sequence verification. Proteins were expressed in a KRX _Escherichia coli_ bacterial system and analysed using 15% SDS-PAGE and Western Blotting. Next, GST-tagged proteins were purified with an affinity column purification system. Purified fusion proteins were subsequently cleaved with Pro-TEV Plus protease, separated on 15% SDS-PAGE gel and stained with Coomassie Brilliant Blue R250. The GST (~26kDa) affinity tags was successfully cleaved from the fusion proteins SARS-CoV N (~45kDa) and HCoV-NL63 N (~44kDa) and confirmed to be the correct estimated sizes.
2.2 INTRODUCTION

Coronaviruses (CoVs) are the largest known single-stranded positive-sense RNA viruses; encoding 5'-capped, polyadenylated genomes ranging in size from 27 to 32 kb. Until recently, CoVs were predominantly associated with severe disease in domestic animals, including bovines (bovine CoV), swine (porcine epidemic diarrhea virus and transmissible gastroenteritis virus [TGEV]), avian (infectious bronchitis virus [IBV]) (Lai and Cavanagh 1997; Baker 2004; Cavanagh 2005), and mice (mouse hepatitis virus [MHV]) (Perlman 1998), while infections in humans were primarily associated with mild upper respiratory tract diseases caused by Human CoVs (HCoVs) HCoV-229E and HCoV-OC43 (Lai and Cavanagh 1997). However, the identification of a novel CoV as the etiological agent responsible for severe acute respiratory syndrome (SARS), an atypical pneumonia with a 10% mortality rate (Stadler, Masignani et al. 2003), indicated that HCoVs are capable of causing severe disease in humans and that unidentified HCoVs continue to exist in natural world. More recent discoveries have led to the identification of two additional HCoVs; HKU-1 and HCoV NL63 (van der Hoek, Pyrc et al. 2004; Lau, Woo et al. 2006). Moreover, NL63 has been associated with croup in infants and young children (van der Hoek, Sure et al. 2005; van der Hoek, Sure et al. 2006; Vabret, Dina et al. 2009).

Croup is a disease caused by many different viruses and is characterized by the sudden onset of a distinctive barky cough, stridor, hoarse voice, and respiratory distress resulting from upper-airway obstruction (Bjornson and Johnson 2008). Croup accounts for roughly 250,000 hospitalizations each year in the United States, and cases severe enough to require hospitalization can be fatal (Henrickson, Kuhn et al. 1994). In addition, although understudied, HCoV infection can result in a particularly severe pneumonia in the elderly,
as evidenced by an outbreak of HCoV-OC43 in a retirement community that was associated with an ~10% mortality rate (Patrick, Petric et al. 2006).

Taxonomically, CoVs are classified as members of the order *Nidovirales*, family *Coronaviridae*, and genus *Coronavirus* (McIntosh 1974; Lai and Cavanagh 1997; Enjuanes, Brian et al. 2000b; Ebel, Fitzpatrick et al. 2011). Currently, the *Coronavirus* genus is further divided into three primary groups based upon serological and phylogenetic data. Among the HCoVs, group 1 contains HCoV-NL63 and HCoV-229E, while group 2 strains include HCoV-OC43, HCoV-HKU-1, and SARS-CoV (Enjuanes, Brian et al. 2000b). The CoVs are roughly 100 nm in diameter, are enveloped, and contain three core structural spikes, including a 180- to 190-kDa spike glycoprotein (S), a 26-kDa membrane glycoprotein (M), and an envelope protein (E) of ~9 kDa. The genomic RNA is surrounded by a helical nucleocapsid composed of the ~50 to 60 kDa nucleocapsid protein (N) (Pyrc, Dijkman et al. 2006).

Additionally, despite large differences in S glycoprotein sequences between SARS-CoV and HCoV-NL63 (less than 50% identity at the nucleotide level), both viral S glycoproteins have been reported to interact with human angiotensin-converting enzyme-2 (ACE2) as a receptor for docking and entry into cells (Hofmann, Pyrc et al. 2005; Pohlmann, Gramberg et al. 2006; Smith, Tusell et al. 2006; Li, Sui et al. 2007). Upon entry into the host cell, the genomic RNA is uncoated and immediately translated into two large polyproteins (Lai and Cavanagh 1997; Masters 2006). The first two-thirds of the CoV genome encodes non-structural replicase proteins in two overlapping open reading frames (ORFs). The final one-third of the genome consists of the structural proteins S, E, M, and N, as well as accessory proteins specific to different strains that are translated from a nested set of 3´ coterminal sub-genomic mRNAs (Lai and Cavanagh 1997; Masters 2006). For HCoV-NL63, there are six genes with a gene order of 5´-replicase-S-ORF3-E-M-N-3´,
wherein gene 1 encodes the non-structural replicase proteins, gene 2 encodes S, gene 3 encodes an accessory protein of unknown function known as ORF3, gene 4 encodes E, gene 5 encodes M, gene 6 encodes N, and an overlapping ORF6b has been predicted to encode an additional accessory protein of unknown function (Pyrc, Jebbink et al. 2004; van der Hoek, Pyrc et al. 2004). All CoV genomes contain group-specific genes in the final one-third of the genome, and many of these genes encode group specific accessory proteins of undetermined function that are dispensable for replication. In addition, ORF3 of HCoV-NL63 encodes a 225-amino-acid protein that is homologous to ORF4 of HCoV-229E (53% similarity) and to ORF3A of SARS-CoV (23% similarity) (Muller, van der Hoek et al. 2010), and both of these proteins have unknown functions.

The structural protein genes of SARS-CoV contain four major ORFs; including the surface S glycoprotein, the small M protein, the E glycoprotein, the N protein, and a set of accessory proteins whose number and sequence vary among different coronaviruses (Siddell 1995). The N proteins have been shown to be strong immunogens in several coronaviruses, such as murine coronavirus (Wege, Schliephake et al. 1993), turkey coronavirus (Akin, Lin et al. 2001), porcine reproductive and respiratory syndrome coronavirus (Casal, Rodriguez et al. 1998). Collisson et al (2000) has reported that N protein accumulates intra-cellularly even before it is packed in the mature virus (Collisson, Pei et al. 2000) and is the most abundant virus derived-protein throughout the infection, probably because its template mRNA is the most abundant sub-genomic RNA (Keck, Hogue et al. 1988; Hiscox, Cavanagh et al. 1995). These features make it a suitable candidate for raising antibodies for diagnostic applications.

In this study, the NL63-N/pFN2A and SARS-N/pFN2A plasmid constructs were confirmed by restriction endonuclease digest and sequence verification. Proteins were expressed in a KRX *Escherichia coli* bacterial system and analysed using 15% SDS-PAGE.
and Western Blotting. Thereafter, GST-tagged proteins were purified with an affinity column purification system. Purified fusion proteins were subsequently cleaved with Pro-TEV Plus protease, separated on 15% SDS-PAGE gel and stained with Coomassie Brilliant Blue R250. The cleaved purified proteins were subsequently used as antigens for HCoV-NL63 and SARS-CoV N antibody production (Chapter 3).
2.3 METHODOLOGY

2.3.1 Bacterial strains and plasmids

The bacterial strains and plasmids that were utilized in this study were purchased from Promega and included *Escherichia coli* (*E. coli*) KRX and JM109 competent cells. KRX is an *E. coli* K12 derivative that has important features associated with cloning and screening strains. For initial transformation of the recombinant vector, JM109 competent cells were used, whereas KRX competent cells were used for optimization and final expression of recombinant protein. The plasmid vectors used were pGEM®-T Easy Vector system II and pFN2A (GST) Flexi® Vector system. The Flexi vector contains a T7 promoter for bacterial or *in vitro* protein expression of a protein-coding region. The vector also has an N-terminal glutathione-S-transferase (GST) coding region that can be used to detect and purify the expressed protein. The GST tag appends a TEV protease site for removal of the tag after protein purification. In addition, the vector also has a lethal barnase gene for positive selection of the insert, an ampicillin-resistance gene for selection of the plasmid and unique *SgfI* and *PmeI* enzyme sites that allow easy insertion or transfer of the sequence of interest. Inserts containing a protein-coding region can easily be transferred from the pFN2A (GST) Flexi® Vector to other Flexi® Vectors with different expression protocols. Therefore, the Flexi vector was used because of its unique ability to express fusion or native proteins in order to study protein structure and function as well as interactions. In contrast, pGEM®-T Easy vector was utilized because of its convenience for cloning of PCR products. The vectors are prepared by cutting with a restriction endonuclease to leave a blunt end and adding a 3’terminal thymidine to both ends. These single 3’ T overhangs at the insertion site greatly improve ligation efficiency of a PCR product into the plasmid by preventing recircularization of the vector and providing a compatible overhang for PCR products with 5’ A overhangs.
2.3.2 The recombinant NL63-N and SARS-N full length constructs

The NL63-N/pFN2A and SARS-N/pFN2A constructs used in transformation for protein expression of NL63-N and SARS-N were supplied by M. Berry (Department of Medical Bio-Sciences, University of the Western Cape). The constructs were maintained and propagated in KRX *E. coli* competent cells. Subsequently, the constructs were cultured in a 15 ml tubes containing Luria Bertani (LB) medium which permits fast growth; good growth yields and is the most common liquid media used in the cultivation of bacteria such as *E. coli*. These cultures were incubated at 37°C with agitation at 180 rpm and maintained at 4 °C on LB agar (LB broth containing 15 g bacteriological agar or at -80°C as glycerol stock cultures). For plasmid DNA selection and maintenance in *E. coli*, the medium was supplemented with 1µl/ml of ampicillin.

The LB medium 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 autoclaving, after which, 1µl/ml ampicillin was added, unless otherwise stated. Additionally, LB agar was prepared as per LB media with the inclusion of 15g bacteriological agar. Agar was thoroughly dissolved by convection heat prior to autoclaving.

**Table 2.1: Prokaryotic vectors used in this study. The Flexi® vector was used as a protein expression vector.**

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEM®-T Easy</td>
<td>Prokaryotic vector</td>
<td>Promega, South Africa</td>
</tr>
<tr>
<td>pFN2A (GST) Flexi®</td>
<td>Prokaryotic vector</td>
<td>Promega, South Africa</td>
</tr>
</tbody>
</table>
2.3.3. Large-scale preparation of recombinant bacterial expression plasmids

2.3.3.1 Autoinduction

KRX is an *E. coli* K strain that contains a chromosomal copy of the T7 RNA polymerase gene that is tightly regulated by a Rhamnose promoter (rhaP_{BAD}). These competent cells allow efficient transformation, cloning, screening and protein expression (Trista Schagat, Rachel Friedman et al. 2008). The rhaP_{BAD} promoter is subject to catabolite repression by glucose via cyclic AMP (cAMP) and the cAMP receptor protein. The presence of Rhamnose activates the promoter, but only once glucose is consumed from the culture medium. The latter provides dramatic control of proteins expressed via a T7 promoter. Pre-induction protein expression levels are exceptionally low. Proteins critical for Rhamnose metabolism [isomerase (RhaA), kinase (RhaB) and aldolase (RhaD)] are deleted in KRX. Therefore, Rhamnose is not metabolized by the cells or consumed during growth, allowing long culture times in the presence of Rhamnose (Trista Schagat, Rachel Friedman et al. 2008).

An early auto-induction protocol adapted from Promega was utilized where glycerol stocks of SARS-CoV N in Flexi and HCoV-NL63 N in Flexi where inoculated into a starter culture of 10ml LB media and incubated at 37°C for 14 hours with shaking at 150rpm.

Low levels of protein expression were previously observed therefore large scale culture was conducted in order to obtain ideal concentrations for protein purification. Subsequently, starter cultures were diluted 1:100 and inoculated in 500ml LB media supplemented with 0.05% (w/v) glucose and 0.1% (w/v) Rhamnose and incubated at 37°C for 24 hours with shaking at 150rpm. Cells were subsequently harvested by centrifugation at 6000 rpm for 15 min in a cold room.
2.3.3.2 Transformation of competent cells

Transformation was conducted in order to verify the presence of the constructs in E. coli (KRX) competent cells. This was performed using the pFN2A (GST) Flexi® vector system as per manufacturer’s instructions (Promega Corporation).

The ligation products were centrifuged and 5 μl of the ligation reactions were dispensed into sterile 1.5 ml micro-centrifuge tubes 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. LB agar was prepared (1% Bacto –Tryptone, 0.5% Bacto – Yeast extract, 0.5% NaCl and 1.5% agar) and sterilized by autoclaving. Subsequently, LB agar plates were prepared from the LB agar. Working aseptically, and with the aid of a glass loop in 70% ethanol, 100μl of the transformation products were spread plated on LB plates containing 1μg/ml ampicillin. Subsequently, the plates were incubated overnight at 37°C for 16-20 hours. Single colonies were then picked and inoculated into LB broth and corresponding PCR mix. Cultures were incubated at 37°C for 14 hours and PCR results were used to select cultures with correct inserts. Cultures were stored in glycerol stocks at -80°C. The plasmids were isolated using the mini-prep technique and constructs were later identified by restriction endonucleases, using SgfI and Pmel restriction enzymes of the Flexi® vector.

2.3.3.3 Plasmid DNA Extraction

The isolation of plasmid DNA from the overnight culture was done using the Wizard® Plus SV mini-preps DNA Purification System (Promega Corporation). This system incorporates the power of binding DNA on silica membrane technology with the
time tested consistency of alkaline-SDS lysis of bacterial cells to deliver high quality DNA. Plasmid DNA binds effectively to the silica membrane of the DNA Column and can easily be purified from contaminants and enzyme inhibitors. Purified plasmid DNA is simply eluted from the column with elution buffer or nuclease free water.

Thereafter, 5 ml of the overnight culture was centrifuged at 14,000 rpm for 5 minutes and a pellet and supernatant was formed. The supernatant was discarded and excess media was blotted off. 250 μl of Cell Resuspension Solution (50 mM Tris-HCl, 10 mM EDTA and 100 μg/ml RNase A) was added to the tube and the pellet resuspended by vortexing. 250 μl of Cell Lysis Solution (0.2 M NaOH and 1% SDS) was added to the solution, mixed by inverting 4 times and incubated at room temperature until the cell suspension cleared. 10 μl Alkaline Protease Solution was added and incubated for 5 minutes at room temperature. 350 μl of Neutralization Solution (4.09M guanidine hydrochloride, 0.759 M potassium acetate and 2.12 M glacial acetic acid) was added and mixed immediately by inverting the tube 4 times. The bacterial lysate was centrifuged at 14,000 rpm for 10 minutes at room temperature. The supernatant was transferred to a spin column assembled with a collection tube (supplied by Promega Corporation), centrifuged for 1 minute and the flow-through discarded. 750 μl of Column Wash Solution (60% ethanol, 60mM potassium acetate, 8.3 mM Tris-HCl and 0.04mM EDTA) was added to the spin column and centrifuged for 1 minute at maximum speed. The wash procedure was repeated using 250 μl of Column Wash Solution and centrifuged for 2 minutes. The spin column was centrifuged at maximum speed for 2 minutes at room temperature with an open lid. This was done so that the excess ethanol may evaporate from the spin column. The spin column was transferred to a sterile 1.5 ml micro-centrifuge tube. The plasmid DNA was eluted in 30 μl of nuclease-free water.
2.3.4 Characterization of recombinant bacterial expression plasmids

2.3.4.1 Restriction endonucleases digestion

Restriction enzymes, also referred to as restriction endonucleases, are enzymes that recognize short, specific (often palindromic) DNA sequences. They cleave double-stranded DNA (dsDNA) at specific sites within or adjacent to their recognition sequences. A single enzyme digest of the purified DNA product (10 µl) was performed using Flexi® Enzyme Blend (SgfI and Pmel) (2µl) restriction enzymes. 4 µl of 5X Flexi® Digest Buffer (50mM Tris-HCl a of pH 7.9 at 37°C, 250mM NaCl, 50mM MgCl₂, 5mM DTT, 0.5mg/ml acetylated BSA) and 4 µl of nuclease free water were added to micro-centrifuge tubes to make a final volume of 20 µl. The reaction tubes were incubated at 37°C in a thermal cycler for 2 hours. Thereafter, the reaction was heated at 65°C for 20 minutes to inactivate the restriction enzymes.

2.3.4.2 Agarose gel electrophoresis

The size fractionation of the completed restriction endonucleases digest was resolved on a 1% agarose gel electrophoresis in TBE (89mM tris (hydroxymethyl) aminomethane, 0.089mM boric acid, 2mM EDTA (pH 8)) buffer stained with 1 µl ethidium bromide. The entire product of each sample was loaded with a Blue/Orange 6X loading dye (0.03% bromophenol blue, 0.03% xylene cyanol FF, 0.4% orange G, 15% Ficoll® 400, 10mM Tris-HCl at pH 7.5 and 50mM EDTA at pH 8.0) to allow the genes of interest to be visible when viewed under UV light. The agarose gel electrophoresis was run at 90 volts for 1 hour for the appropriate size product, which indicated the correct inserts are present in the cloned gene. Thereafter, the gel pictures were analyzed with the Doc-ItLS® Image Acquisition Software, Version 5.5.5a (Doc-It Life Sciences Software,
UVP™, Inc., San Gabriel, USA) and viewed under a UV Transluminator (UVP™, Inc., San Gabriel, CA 91778, USA).

2.3.4.3 Nucleotide sequencing and sequence analysis

Nucleotide sequencing is a process that is utilized in genetics to determine the order of nucleotide bases along a DNA strand. Since the discovery of the structure of DNA, two distinct methods for sequencing DNA were developed namely the chain termination method and the chemical degradation method. The extracted plasmids were previously sequenced with M13 forward and reverse primers to ensure no base-pair mutations occurred. Sequencing reactions were conducted by Inqaba Biotech. Thereafter, sequence analysis was performed in silico using Finch TV (version. 1.4.0) and extracted sequences were aligned using the multiple sequence alignment function in Clustal X (Version 2.0.12) (Larkin, Blackshields et al. 2007). Sequenced files were subsequently extracted to GeneDoc (Version 2.7.000) for viewing and analysis. The nucleotide sequences that were utilized in this study were obtained from the GenBank sequence database under accession numbers AY360146.1 and DQ846901.1 (Zhu, Qian et al. 2006b).

2.3.5 Protein analysis

2.3.5.1 Protein extraction and Preparation of cell lysates

In addition to centrifugation, cell pellets were resuspended by vortexing in 2ml/g lysis buffer containing 1% (v/v) Triton X, 150mM NaCl, 10mM Tris (hydroxymethyl)-amino-methane (Tris), 5mM EDTA and a protease inhibitor cocktail. Table 2.2 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 successive procedures were done on ice or at 4°C. Subsequent to sonication, insoluble cell fraction was separated from
soluble fraction by centrifugation at 14000 rpm 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 proved to be insufficient.

### Table 2.2: Represents the functions of the components in a cell lysis buffer.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triton X</td>
<td>Non-ionic detergent; improves solubility of GST fusion proteins and prevents aggregation of lysed cells.</td>
</tr>
<tr>
<td>NaCl</td>
<td>Provides an osmotic shock to cells</td>
</tr>
<tr>
<td>Tris</td>
<td>Interacts with lipopolysaccharides in the outer membrane of the cell and thereby increases permeability.</td>
</tr>
<tr>
<td>EDTA</td>
<td>Inhibits divalent cation-dependent proteases</td>
</tr>
</tbody>
</table>

#### 2.3.5.2 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

A 3% stacking and a 15% separation SDS-PAGE gel was used to separate proteins in a SUB Series Horizontal Electrophoresis Units (Hoefer® System). Electrophoresis was run at 15mA/gel for 80 min in a Tris-glycine SDS running buffer (25mM Tris, 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 Western Blot SUB Series Horizontal Electrophoresis Units (Hoefer® System), to identify the GST tag with the appended N-proteins. Coomassie stain was removed using a Coomassie destain solution (50% (v/v) methanol and 10% (v/v) acetic acid).
2.3.5.3 Western Blotting

Western blotting (also referred to as immune-blotted because an antibody is used to specifically detect its antigen) was conducted in accordance to Towbin et al. (Towbin, Staehelin et al. 1979). 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 90 min, after which, the membrane was blocked with a 5% (w/v) milk and 0.05% (v/v) tween 20 in PBS solution for 30 min on a rocker. The membrane was then incubated at 4°C overnight on a roller in 3% (w/v) milk and 0.05% (v/v) tween 20 in PBS solution with the primary antibody, rabbit α 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, horseradish peroxidase (HRP)-labelled 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 colorimetric analysis.

2.3.5.4 Purification of protein constructs

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 Na$_2$HPO$_4$, 2mM K$_2$HPO$_4$, 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 30 min. 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 30 min 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.

2.3.5.5 Quantification of protein constructs

The Qubit® Fluorometer system is a bench-top fluorometer for the quantification of DNA, RNA, and protein, using the highly sensitive and accurate fluorescence-based Qubit™ quantitation assays. The use of the state of the art dyes selective for dsDNA, RNA, and protein minimizes the effects of contaminants in a sample that can affect the quantitation. Furthermore, the very latest illumination and detection technologies used in the Qubit® Fluorometer for attaining the highest sensitivity allow the use as little as 1 μl of sample and still achieve high levels of accuracy, even with very diluted samples.

Subsequently, the purified proteins were quantified using the Quant IT™ Protein assay kit with the Qubit™ Fluorometer System. 5μl of the purified proteins was resuspended into two eppindorfs which contained 195 μl of Quant-iT™ working (Quant-iT™ protein buffer and Quant-iT™ reagent) to make a final volume of 200μl. The working solution was further pipetted into three eppindorfs. These eppindorfs contained 190 μl of the working solution and 10 μl of the Quant-iT™ protein standards (1-3). All
five tubes were briefly vortexed for 2-3 seconds and incubated at room temperature for 15 minutes. The Qubit™ Fluorometer was calibrated using the Quant-iT™ protein three standards and the concentrations of the samples were read using the Quant-iT™ protein assay.

2.3.5.6 Cleavage of fusion proteins

For 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 Plus protease. ProTEV Plus 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 performed according to manufacturer’s instruction; briefly reactions were prepared in 200μl with 10μl 20x ProTEV buffer, 2μl 100mM DTT and 4μl 10U ProTEV Plus. Nuclease free water was replaced with purified fusion protein which was increased to a total volume of 184μl. The reaction was incubated for 2 hours at 30°C. Thereafter, an extra 2μl 10U ProTEV Plus protease enzyme was added and performed a final incubation of 2 hours at 30°C. Subsequently after incubation the sample was separated by 15% SDS-PAGE as previously described and visualized by staining with Coomassie Brilliant Blue R250.
2.4 RESULTS AND DISCUSSION

Coronaviruses, a genus of the Coronaviridae family, are enveloped viruses with a large positive strand RNA genome. The viral RNA genome is 27–32 kb in size; capped; polyadenylated and encapsulated in a helical nucleocapsid. The envelope is studded with long, petal shaped spikes, giving the virus particle a characteristic crown-like appearance. There are three distinct groups of coronaviruses that have been adequately described based on serological affinity and genome sequence. Coronaviruses can infect a wide range of species ranging from humans to a variety of domestic animals, and can cause highly prevalent diseases such as respiratory, enteric, cardiovascular and neurologic disorders (Guy, Breslin et al. 2000; Lai MMC and Holmes KV 2001).

In the 1960s, there were only two Human Coronaviruses (HCoVs) that were indentified and thoroughly studied; these were HCoV-OC43 (group 2) and HCoV-229E (group 1). They are responsible for 10–30% of all common colds and infections occur mainly during winter and early spring (Tyrrell and Bynoe 1965; Hamre and Procknow 1966; Almeida JD and Tyrrell DA 1967). A third novel human coronavirus, SARS-CoV, was then identified as the causal agent of severe acute respiratory syndrome (SARS) during a 2002–2003 outbreak (Drosten, Gunther et al. 2003; Ksiazek, Erdman et al. 2003; Peiris, Lai et al. 2003). Phylogenetic analysis showed that the SARS-CoV did not closely resemble any of the three previously known groups of coronaviruses, and therefore a tentative fourth group of coronaviruses was initially proposed (Marra, Jones et al. 2003; Rota, Oberste et al. 2003). However, an early split-off of the SARS-CoV from the coronavirus group 2 lineage was subsequently accepted (Snijder, Bredenbeek et al. 2003; Gorbunova, Snijder et al. 2004).

A novel human coronavirus associated with respiratory illness, HCoV-NL63, was identified by van der Hoek and colleagues in The Netherlands (van der Hoek, Pyrc et al.
Shortly thereafter, two research groups described the characterization of essentially the same virus (van der Hoek, Pyrc et al. 2004; Fouchier, Hartwig et al. 2005; Esper, Weibel et al. 2005b). The virus was isolated in January 2003 from a nasopharyngeal aspirate of a 7-month-old child suffering from bronchiolitis, conjunctivitis and fever. Real-time RT-PCR assays were designed for screening of respiratory tract samples. Screening of specimens from patients with respiratory symptoms identified seven additional HCoV-NL63 infected individuals (both children and adults) between December 2002 and February 2003. HCoV-NL63 can be considered as a new important causative agent of respiratory illnesses and two different subtypes might be currently co-circulating in the human population (van der Hoek, Pyrc et al. 2004).

The characteristic genome organisations of coronaviruses are: the 5' two-third of the genome contains two large ORF; ORF1a and ORF1b. In the 3' part of the genome, genes encoding four structural proteins are found: S, E, M and N. The haemaglutinin-esterase gene, characteristic for group 2 coronaviruses, is not present in HCoV-NL63. Sequence analysis demonstrated that HCoV-NL63 shares 65% sequence identity with HCoV-229E. Phylogenetic analysis confirmed that HCoV-NL63 is a new group 1 coronavirus, most closely related to HCoV-229E and porcine epidemic diarrhea virus (PEDV) (van der Hoek, Pyrc et al. 2004).

The major function of N protein is to assemble the RNA of coronaviruses. The N protein is one of the major virion structural proteins of SARS-CoV and HCoV-NL63. The predicted N protein of SARS-CoV has a predicted molecular weight of 46 kDa and HCoV-NL63 a molecular weight of 43 kDa. In order to verify the presence of the genes of interest in the NL63-N/pFN2A and SARS-N/pFN2A constructs, a single restriction endonuclease digest of the purified recombinant vectors was performed using Flexi® Enzyme Blend (Sgfl and Pmel) restriction enzymes. The digestion was aimed at separating the HCoV-
NL63 N and SARS-CoV N gene fragments from the Flexi® vector. The size fractionation of the restriction endonuclease digest was resolved on a 1% agarose gel electrophoresis. The results confirmed the presence of the HCoV-NL63 N (~1.1 Kb) and SARS-CoV N (~1.2 Kb) genes on the pFN2A (GST) Flexi® vector (~4.1 Kb) (Figure 2.1).

![Figure 2.1: Verification of insert sizes using restriction endonuclease digest. The recombinant pFN2A (GST) Flexi® vector (~4.1Kb) was digested with Flexi Enzyme blend (Sgf 1 and Pme 1) to release the inserts. Lane 1-4: HCoV-NL63 N-gene (~1.1Kb); Lane M: 1 Kb marker; Lane 5-7: SARS-CoV N-gene (~1.2Kb).](image)

Chromatogram sequencing reactions were conducted by Inqaba Biotec. Results were viewed with FinchTV (www.geospiza.com). The nucleotides were fluorescently-labeled, detected and then represented in the form of chromatograms. All inserts were verified by sequencing the genes of interest in both directions. Subsequently, the obtained consensus sequences were compared with the HCoV-NL63 sequences available in GenBank database release 142.0 using BLAST analysis (NCBI BLAST server). The
cloned genes were sequenced with M13 forward and reverse primers. M13 primer binding sites are located up- and down-stream of the multiple cloning sites allowing for easy sequencing. 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. The gaps and mismatches were removed by sequencing the genes of interest in both directions.

Sequencing was done to screen and check for possible mutations or homologies with the original and cloned genes (Figure 2.2 and 2.3). Thereafter, sequence analysis was performed in silico using Finch TV (version 1.4.0) and extracted consensus sequences were aligned using the multiple sequence alignment function in Clustal X (Version 2.0.12). Sequenced files were subsequently extracted to GeneDoc (Version 2.7.000) for viewing and analysis. All the sequences were confirmed to be the SARS-CoV N gene (Figure 2.2) and HCoV-NL63 N gene (Figure 2.3). All the inconsistencies in the alignment were confirmed to be chromatographic errors where more than one peak existed at the points in question on the chromatogram supplied by Inqaba Biotech. This was resolved by sequencing both strands and visual inspection of chromatograph peaks where inconsistencies occurred. Sequence analysis of the highly conserved nucleocapsid genes were similar to the SARS-CoV (E-value: 0:0 and Max Identity of 99%) and HCoV-NL63 (E-value: 0:0 and Max Identity of 98%) complete genome sequences that were obtained from the NCBI GenBank.
Figure 2.2: Sequence verification of the cloned HCoV-NL63 Nucleocapsid gene in pGEM® vector. Shading indicates conserved regions (black: 100% homology; dark grey: 80% homology; light grey: 60% homology) and gaps were introduced to align sequences. NL63_N represents the gene strand sequenced with M13 forward and reverse primers. All sequences were verified by sequencing the N-gene in both directions. NL63_N_FL is the sequence of the HCoV-NL63 N-gene obtained from NCBI with accession number DQ846901.1, direct submission to GenBank (Zhu, Qian et al. 2006b).
Figure 2.3: Sequence verification of the cloned SARS-CoV Nucleocapsid gene in pGEM® vector. Shading indicates conserved regions (black: 100% homology; dark grey: 80% homology; light grey: 60% homology) and gaps were introduced to align sequences. SARS_N represents the gene strand sequenced with M13 forward and reverse primers. All sequences were verified by sequencing the N-gene in both directions. SARS_N_FL is the sequence of the SARS-CoV N-gene obtained from NCBI with accession number AY360146.1, direct submission to GenBank (Zhu, Qian et al. 2006b).
All constructs 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 (Baneyx 1999). KRX strain of competent *E. coli* cells was used for the protein 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 (Baneyx 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 (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 (Baneyx 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 (Baneyx 1999).

Thereafter, total proteins were extracted from *E. coli* expressing N proteins and separated on 15% SDS-PAGE and stained with Coomassie brilliant blue. They were subsequently transferred from an SDS-PAGE gel to a nitrocellulose membrane for Western Blotting. The blots were detected with specific rabbit anti-GST primary antibody and appropriate secondary antibody. The unpurified total proteins have multiple bands which are indicative of *E.coli* and viral proteins. The total viral proteins were confirmed to be the correct estimated sizes SARS N-GST ~71kDa and NL63 N-GST ~70kDa (Figure 2.4).

![Figure 2.4: Expression of unpurified total viral fusion proteins detected using (A) Coomassie stained SDS-PAGE gel and (B) Western Blot.](image)

(A) Total proteins extracted from *E. coli* expressing N proteins were separated on 15% SDS-PAGE and stained with Coomassie brilliant blue. (B) Proteins were transferred from SDS-PAGE to nitrocellulose membrane with Western Blot. Blots were detected with rabbit anti-GST primary antibody and appropriate secondary antibody. Lane M: Pre-stained molecular weight protein marker; Lane 1: Expression Of ~ 71 kDa SARS-CoV N viral fusion proteins with GST tag; Lane 2: Expression of ~70 kDa HCoV-NL63 N viral fusion protein with GST tag.
Subsequently, protein purification was done using the MagneGST™ purification system. Protein purification is a critical aspect for the recovery of protein with little or no contaminants. The non-specifically GST tagged fusion proteins were removed after protein purification. The GST purification tag used for this study is a 26kDa protein from the parasite Schistosoma japonicum. The final protein concentrations were determined with the aid of the Qubit™ Fluorometer system. The results are depicted in Table 2.3.

Thereafter, the fusion protein sizes were resolved with sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). This method uses polyacrylamide gels and buffers containing sodium dodecyl sulfate (SDS) to analyze small amounts of protein. During SDS-PAGE, proteins are denatured and coated with detergent by heating in the presence of SDS and a reducing agent. SDS-PAGE is commonly performed in reducing conditions (e.g., presence of DTT or beta-mercaptoethanol) in order to reduce the disulfide bonds found within some proteins and to facilitate protein denaturation (Promega). Since proteins have a net negative charge that is proportional to their size, the proteins of interest were separated on the basis of their molecular mass a result of the sieving effect of the gel matrix. The molecular mass of the proteins were estimated by comparing the gel mobility of a band with a pre-stained molecular weight protein marker.

Sharp protein bands were achieved by using a discontinuous gel system, having stacking (3%) and separating gel (15%) layers that differ in both salt concentration and pH. The SDS-PAGE gel was stained with Coomassie Brilliant Blue stain [40% (v/v) methanol, 10% (v/v) acetic acid and 0.025% (w/v) Coomassie Brilliant Blue R250] overnight, to visualize purified proteins expressed in the cell and to identify the GST tag with the appended N-proteins. Coomassie stain was removed using a Coomassie destain solution [50% (v/v) methanol and 10% (v/v) acetic acid]. The fusion viral proteins were confirmed
to be the correct sizes i.e. SARS N-GST ~71kDa and NL63 N-GST ~70kDa (Figure 2.4 and Figure 2.5).

Table 2.3: Concentration of purified total proteins in mg/ml.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Concentration (mg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARS-CoV N-GST</td>
<td>0.44</td>
</tr>
<tr>
<td>HCoV NL63 N-GST</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Figure 2.5: Detection of purified total proteins using (A) Coomassie stained SDS-PAGE gel and (B) Western Blot. (A) Total proteins were purified using the MagnetGST™ purification system and were separated on 15% SDS-PAGE and stained with Coomassie brilliant blue. (B) Proteins were transferred from SDS-PAGE to nitrocellulose membrane with Western Blot. Blots were detected with rabbit anti-GST primary antibody and appropriate secondary antibody Lane M: Pre-stained molecular weight protein marker; Lane 1: Purified ~71 kDa SARS-CoV N with GST tag; Lane 2: Purified ~70 kDa HCoV-NL63 N with GST tag.
A protease enzyme was used to cleave GST affinity tags from viral fusion proteins after protein purification. ProTEV Plus protease which is an improved 48kDa version of the N1a protease from Tobacco Etch Virus (TEV) that has been engineered to be more stable than native TEV protease for prolonged enzymatic activity (Dougherty and Parks 1991; Carrington, Haldeman et al. 1993; Kapust, Tozser et al. 2001). TEV protease is a highly site-specific proteolytic enzyme that recognizes the seven-amino-acid sequence EXXYXQ (G/S), most commonly ENLYFQG, with cleavage occurring between glutamine and glycine or serine (Carrington and Dougherty 1988; Dougherty, Parks et al. 1989). The protease cleaves sequences with a variety of amino acids at the G/S (or P1’) position (Kapust, Tozser et al. 2002). This allows the choice from many different amino acids on the newly formed N-terminus after cleavage. Optimum activity is obtained at pH 7.0 and 30°C, however ProTEV Plus is active over a wide range of pH values (5.5–8.5) and temperatures (4–30°C), allowing a choice of conditions amenable to the protein of interest.

ProTEV Plus Protease is easily removed from the cleavage reaction after cleavage using the HQ-tag located at the N-terminus of the protein. The Pro TEV plus reactions was conducted as described previously and cleaved fusion proteins were subsequently separated by SDS-PAGE and analyzed using Coomassie Brilliant Blue R250 stain. Results showed the GST (~26kDa) affinity tags and the viral fusion proteins SARS-CoV N ~45kDa and HCoV-NL63 N ~44kDa correct sizes compared to a protein marker (Figure 2.6 and Figure 2.7).
Figure 2.6: Digest of the GST affinity tag from HCoV-NL63 full length N protein using ProTEV protease enzyme. Total proteins were separated on 15% SDS-PAGE and stained with Coomassie brilliant blue. Lane 1: ProTEV (-) uncleaved HCoV-NL63 N viral fusion protein (~70 kDa); Lane 2-8: ProTEV (+) HCoV-NL63 N viral fusion protein (~44 kDa) with the GST tag (~ 26 kDa) from the total viral fusion protein.

Figure 2.7: Digest of the GST affinity tag from SARS-CoV full length N protein using ProTEV protease. Total proteins were separated on 15% SDS-PAGE and stained with Coomassie brilliant blue. Lane 1: ProTEV (-) uncleaved SARS-CoV N viral fusion protein (~71 kDa); Lane 2-8: ProTEV (+) SARS-CoV N viral fusion protein (~45 kDa) with GST tag (~ 26 kDa) cleaved from the total viral fusion protein.
2.5 CONCLUSION

In this study, the NL63-N/pFN2A and SARS-N/pFN2A plasmid constructs were confirmed by restriction endonuclease digest and sequence verification. Proteins were expressed in a KRX *Escherichia coli* bacterial system and analysed using 15% SDS-PAGE and Western Blotting. Thereafter, GST-tagged proteins were purified with an affinity column purification system. Purified fusion proteins were subsequently cleaved with Pro-TEV Plus protease, separated on 15% SDS-PAGE gel and stained with Coomassie Brilliant Blue R250. The sequenced chromatographs were consistent with sequences from the NCBI GenBank. 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, the expression protocol had to be optimized to be able achieve the results that are presented in this study. There are several methods of optimization that were employed such as (i) obtaining optical density readings for optimum bacterial growth during protein expression; (ii) utilization of lysozyme to solubilise the proteins and (iii) due to the instability of the N protein, temperature was also optimized. All the previously outlined optimization strategies have resulted in successful protein expression of the nucleocapsid protein in the Flexi® vector system.

A study by Chiu and colleagues (2005) reported that human coronaviruses account for 4.4% of all hospital admissions for acute respiratory infections in children under the age of 18 years (Chiu, Chan et al. 2005). Among these coronaviruses, HCoV-NL63 is the most common coronavirus identified. Furthermore, HCoV-NL63 has been shown to have a worldwide distribution and is observed primarily in the winter season in temperate climates. To date, HCoV-NL63 has been associated with acute respiratory illness and croup in young children, the elderly and immunocompromised individuals. Additionally, HCoV-NL63 can also present as asthma exacerbation, febrile seizures and high fever.
Future prospects in coronavirology include interaction studies of N with other structural proteins such as the M, E and ORF3 proteins. Since this virus can impact the state of our economy due to the illnesses it causes, the development of antiviral drugs is also envisaged; not only for HCoV-NL63, but also for other human coronavirus such as SARS-CoV. The cleaved viral fusion proteins were subsequently used as antigens for HCoV-NL63 and SARS-CoV N antibody production (Chapter 3).
2.6 REFERENCES


Hong Kong, China." Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 40(12): 1721-1729.


CHAPTER THREE

Generation and validation of antibodies

Department of Medical Biosciences, Faculty of Natural Sciences, University of the Western Cape, South Africa.
3.1 ABSTRACT

In 2004 a group 1 Coronavirus, designated Human Coronavirus NL63 (HCoV-NL63), was isolated from a 7 month old Dutch child suffering from bronchiolitis. HCoV-NL63 has since been shown to be a common occurrence worldwide and has been associated with many clinical symptoms and diagnoses, including severe lower respiratory tract infection, croup and bronchiolitis. In addition, HCoV-NL63 causes disease in children (detected in approximately 10% of respiratory tract infections), the elderly and the immunocompromised. The SARS-CoV (associated with an atypical pneumonia) epidemic resulted in 8,422 cases with 916 deaths globally (case fatality rate: 10.9%). In comparison to SARS-CoV, HCoV-NL63 is less virulent with just three reported fatalities. Currently, there are no vaccines available for these respiratory viruses, therefore it is necessary to monitor epidemic patterns and investigate the spread of infections to efficiently identify, control and prevent epidemics. The nucleocapsid protein is a key component of Coronaviruses and is essential for virus formation. It binds to the viral RNA to form a ribonucleoprotein core, which can enter the host cell and interact with cellular processes. Moreover, it is abundantly expressed during infection and it is highly immunogenic.

In the present study five, specific pathogen free (SPF), Balb/c female mice were used to produce polyclonal antibodies for HCoV-NL63 and SARS-CoV nucleocapsid protein. Two mice (nl63-1 and nl63-2) were immunized with NL63 purified viral fusion N protein and the other three (sars-1, sars-2 and sars-3) were subsequently immunized with SARS purified viral fusion N protein. Thereafter, polyclonal antibodies were successfully produced and validated using a direct Enzyme-Linked Immunosorbent Assay (ELISA) on antigen coated plates. The results were analyzed using a microtiter plate reader and the readings were recorded at a wavelength of OD$_{450}$nm. The primary antibodies treated with
SDS indicated an increase in antibody titre, meaning that ELISA was able to detect more antibodies in serum because the antigen was in an “unfolded” confirmation. Furthermore, cross-reactivity was also observed which is consistent with several studies that were performed on the same nucleocapsid protein. Previous studies have also confirmed that this is due to antigenic epitopes that are found across all human coronaviruses. Further studies need to be conducted that will address the extent of cross reactivity to the nucleocapsid protein of human coronaviruses.
3.2 INTRODUCTION

Coronaviruses (CoVs) are large enveloped positive-strand RNA viruses that belong to the family Coronaviridae (Lai MMC and Holmes KV 2001). Based on serological sequence analysis and cross-reactivity, coronaviruses are classified into three distinct groups, namely alpha, beta and gamma (Lai MMC and Holmes KV 2001; Drexler, Gloza-Rausch et al. 2010a). The alpha and betacoronavirus groups contain various mammalian CoVs, whereas avian viruses cluster in the gammacoronavirus group. The earliest-described HCoV strains, HCoV-229E and HCoV-OC43, which are group I and group II coronaviruses, respectively, have now been joined by the more recently described group I and II strains HCoV-NL63 and HCoV-HKU1, respectively (Fouchier, Hartwig et al. 2004; van der Hoek, Pyrc et al. 2004; Woo, Lau et al. 2004; Perlman and Netland 2009). The SARS-CoV constitutes a fifth HCoV, which was in circulation for a limited time during 2002 and 2003, when a novel virus appeared in humans and caused an outbreak affecting at least 8,000 people (Drosten, Günther et al. 2003; Peiris, Yuen et al. 2003). Mortality rate was high, at approximately 10% (Peiris, Yuen et al. 2003). Furthermore, in 2012, a sixth novel HCoV was discovered and sequenced at Erasmus Medical Centre (EMC) in the Netherlands (Corman, Eckerle et al. 2012). The new virus, designated HCoV-EMC, has been responsible for two severe cases in England and Saudi Arabia, one of which was fatal (Corman, Eckerle et al. 2012).

HCoV-229E and HCoV-NL63 belong to the alphacoronavirus group and are the only two human CoVs that have a relatively close phylogenetic relationship (Dijkman and van der Hoek 2009). HCoV-229E was discovered in the mid-1960s in a patient with a common cold, and HCoV-NL63 was isolated for the first time in 2004 from a 7-month-old infant with bronchiolitis (Hamre and Procknow 1966; van der Hoek, Pyrc et al. 2004).
Furthermore, both HCoV-229E and HCoV-NL63 infections have a worldwide distribution, having their peak activity during the winter months causing approximately 10% of all upper and lower respiratory tract infections (LRTI) in hospitalized children (Gerna, Percivalle et al. 2007; Vabret, Dina et al. 2008; Principi 2010; van der Hoek, Ihorst et al. 2010). Seemingly, most cases of these infections do not lead to severe clinical symptoms, although acute infections in infants, elderly individuals, and immunocompromised patients can cause more serious respiratory disease, that may require hospitalization (Chiu, Chan et al. 2005; Milano, Campbell et al. 2010).

The clinical manifestations of HCoV-229E in infected patients are headache, nasal discharge, chills, cough, and sore throat, whereas the symptoms observed in HCoV-NL63 infected patients are more severe, including fever, cough, sore throat, bronchiolitis, and pneumonia. Additionally, HCoV-NL63 has recently been associated with the childhood disease croup (van der Hoek, Sure et al. 2005). Croup is a common manifestation of LRTI in children and has been reported to occur mostly in boys. Furthermore, it indicates peak occurrence in early winter and HCoV-NL63 follows a similar trend of prevalence (van der Hoek, Sure et al. 2005). Reinfections with HCoV-229E and HCoV-NL63 are expected to occur throughout our lifetime. Presently, there are no therapeutic treatments available for any of the HCoVs, and diagnosis is based on virus detection by reverse transcription polymerase chain reaction (RT-PCR) technology (van Elden, van Loon et al. 2004; Bastien, Robinson et al. 2005b; Leung, Li et al. 2009).

Several studies have described the development of immunoassays using the nucleocapsid (N) protein for detection of antibodies to human coronaviruses (Woo, Lau et al. 2004b; Dijkman, Jebbink et al. 2008; Lehmann, Wolf et al. 2008; Severance, Bossis et al. 2008). The N protein is abundantly expressed during infection and it is highly
immunogenic (Hiscox, Cavanagh et al. 1995; Zhao, Wang et al. 2007). The N protein has therefore emerged as a strong candidate to be utilized as an antigen for detecting CoV infection.

Therefore, the aim of the study described in this chapter was to establish a direct enzyme-linked immunosorbent assay (ELISA) using coated antigen capable of detecting and distinguishing between HCoV-NL63 N and SARS CoV N antibodies. Five, specific pathogen free (SPF), Balb/c female mice were used to produce polyclonal antibodies for HCoV-NL63 and SARS-CoV nucleocapsid protein. Serum antibodies were detected on all N coated plates which indicated the production of the polyclonal antibodies by the individual mice. In future, these antibodies could be used to detect and possibly neutralise HCoV-NL63. Moreover, this could lead to the development of anti-viral treatments, rapid detection kits and passive immunisation therapies against respiratory infections. In achieving these end points it will be possible to decrease the prevalence of morbidity due to respiratory tract infections and thus increase productivity and quality of life in both first and third world countries.
3.3 METHODOLOGY

3.3.1 Balb/c mice strain

Specific pathogen free, Balb/c female mice strain were purchased from the University of Cape Town’s Animal Unit (Cape Town, South Africa) after obtaining animal ethical clearance from the University of the Western Cape. All experiments were conducted in accordance with the guidelines of the institutional animal ethics committee. The mice were kept in a well-ventilated animal facility (temperature 18-24°C and 12 hour natural light/12 hour dark cycles) in which they had access to drinking water and fed standard mouse feed (Medical Research Council, Cape Town, South Africa).

3.3.2 Antigen preparation

The recombinant viral fusion N proteins of HCoV-NL63 and SARS-CoV were prepared using the *Escherichia coli* system as described previously (Chapter 2). Briefly, the N genes of each virus were verified by restriction endonucleases, sequenced, expressed in KRX *E. coli* strain; purified to homogeneity and quantified using a Qubit® System. The recombinant N proteins were expressed by transformation in KRX-derived strain competent cells (Promega Corporation). Subsequently, overnight cultures of transformed bacteria were inoculated in media supplemented with 0.05% (w/v) glucose and 0.1% (w/v) Rhamnose. Transformation was conducted in order to verify the presence of the constructs in *E. coli* (KRX) competent cells. This was done using the pFN2A (GST) Flexi® vector system as per manufacturer’s instructions (Promega Corporation). The purified viral proteins were cleaved using ProTEV protease enzyme and analyzed with SDS-PAGE and Commassie staining. The purified viral fusion proteins were used as antigens for immunization of mice.
Subsequently, the viral fusion protein bands were incised out of a SDS gel into small slices using a scalpel and dispensed into their respective 2ml eppindorf tubes. Thereafter, the gel slices were homogenized with a tissue homogenizer to make a thick emulsion. The tissue homogenizer was sterilized with household bleach, dishwasher and tap water prior to the subsequent procedures. The protein and gel slices with the adjuvant were homogenized at two minutes intervals until a homogeneous suspension was obtained. The emulsion was transferred into a syringe (1ml). All the air in the syringe was expelled and an appropriate sized needle was used for subsequent immunization.

3.3.3 Immunization of the mice

Complete Freund’s adjuvant (CFA) (Sigma-Aldrich) was used as a water-in-oil emulsion and was prepared from non-metabolizable oils [paraffin oil (0.85 ml per ml) and mannide monooleate (0.15 ml per ml)]. Additionally, it contained attenuated *Mycobacterium tuberculosis* (strain H37Ra, ATCC 25177) in comparison with the Incomplete Freund’s Adjuvant (IFA) (0.85 ml per ml of paraffin oil and 0.15 ml per ml of mannide monooleate) which is without the bacteria.

The female balb/c strain mice were immunised with an initial intraperitoneal injection of 100 µl antigen emulsified in 100 µl CFA. Antibodies were generated by injecting mice with either SARS or NL63 N antigen. Each mouse was immunised with 200 µl of the antigen and CFA emulsion. Thereafter, two booster immunisations were conducted and doses were prepared in IFA at 2-week intervals. Blood for monitoring of antibody titre was collected from the tail vein on the day of the first immunisation and then again 6 weeks later. Of the 50 µl blood that was collected at each bleed, a volume of 10 µl was used for subsequent experiments. The blood was collected in 0.5 ml eppindorfs. Subsequently, blood plasma and red bloods cells were separated by centrifugation for 5
minutes at 5000 rmp and thereafter, all the samples were screened for antibody titre using antigen-capture Direct ELISA. The animals with the highest antibody titres were used for polyclonal antibody production. The animals were asphyxiated with CO₂ prior to the procedure and sacrificed by cervical dislocation. Anaesthetics could not be used as these would affect the spleen cells that were required for polyclonal antibody production. Asphyxiation and cervical dislocation of the mice were done by the principal investigator. Thereafter, blood was collected by the heart puncture procedure.

3.3.4 Direct ELISA on antigen coated plates

The recombinant proteins were quantified as described previously and standardized to 1 µg/ml in coating buffer (1X PBS). Each well of a Nunc MaxiSorp® flat-bottom 96 well microtiter plates (Nunc-Immuno plate, Serving Life Science, Denmark) was coated with 50 µl of recombinant protein (Table 3.1), covered with Saran wrap and incubated overnight at 4°C. Microtiter plates were washed 5 times with 300ul washing buffer (1 X PBS containing 0, 1% (v/v) Tween), followed by blocking with 200 µl of 1% Albumin in 1 X PBS for 1 hour at room temperature. Plates were subsequently washed 5 times as previously mentioned. The primary antibodies (blood samples) were serially diluted (1:100 to 1:6400) in 1% Albumin in 1 X PBS. The final antigen dilution was determined arbitrarily. Thereafter, 50 µl of the diluted antibodies was added in respective wells. Plates were incubated for 2 hours at room temperature, followed by washing 5 times in a washing buffer. Subsequently, 50 µl of horseradish peroxidase (HRP)-labelled anti-mouse antibody (Southern Biotech, USA) in a dilution of 1:4000 in 0.1 % Albumin and 1X PBS or in 1 X PBS containing 0,05 % SDS was added to the respective wells. Thereafter, the plates were incubated for 45 minutes at room temperature on a Microtiter plate shaker. Plates were then washed 7 times after which 50 µl of the substrate solution (Tetramethylbenzidine
solution, DRG diagnostics, Germany) was added per well. The plates were then incubated for 15 minutes at room temperature. The enzyme reaction was stopped by the addition of 50 µl of 2M sulphuric acid (H$_2$SO$_4$) per well. Absorbance readings were then recorded at 450nm using a Microtiter plate reader and Ascent Software version 2.6 (Original Multiskan EX, Type 355, Thermo Electron Corporation, China).

Table 3.1: Represents the viral fusion proteins that were used in this study.

<table>
<thead>
<tr>
<th>Name of constructs</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARS-N</td>
<td>SARS-CoV full length N (aa 1-423)</td>
</tr>
<tr>
<td>NL63-N</td>
<td>HCoV-NL63 full length N (aa 1-378)</td>
</tr>
</tbody>
</table>

3.3.5 Statistical Analysis

Data were initially entered onto a Microsoft Excel 2007 spreadsheet (Microsoft Corporation, Redmond, USA). Thereafter, all statistical analysis was performed using MedCalc®, Version 11.5.1 (MedCalc Software, Mariakerke, Belgium). ELISA results were recorded at a wavelength of OD$_{450nm}$. Corrected values were obtained by subtracting the optical density readings of pre-immunization from the first and second bleed at a dilution titre of 1/100. Prior to testing for significance, Normal distribution was determined with the Kolmogorov-Smirnov test. Thereafter, the Mann-Whitney (Independent samples) rank sum test was used to determine significance of the differences between the five constructs. Differences were considered statistically significant at P less than 0.05.
3.4 RESULTS AND DISCUSSION

The World Health Organization estimates that about 20% of deaths in children younger than 5 years old are due to severe lung infections caused by bacteria and viruses. In 20-30% of the viral causes, the specific causative organisms remain unidentified. Coronaviruses have been suggested as possible agents of these unidentified causative organisms. Infections of humans by coronaviruses are not normally associated with severe diseases. However, the identification of the coronavirus responsible for the outbreak of severe acute respiratory syndrome (SARS-CoV) showed that highly pathogenic coronaviruses can enter the human population (McBride and Fielding 2012). The SARS-CoV epidemic resulted in 8 422 cases with 916 deaths globally (case fatality rate: 10.9%). Shortly thereafter, another novel human coronavirus (HCoV-NL63) was isolated from a seven-month old infant suffering from respiratory symptoms (van der Hoek, Pyrc et al. 2004). This virus has been shown to have a worldwide distribution and was observed primarily in the winter season in temperate climates like the Netherlands and South Africa, to name a few (Smuts, Workman et al. 2008; van der Hoek, Ihorst et al. 2010). Furthermore, HCoV-NL63 has been shown to infect mainly children and the immunocompromised, who present with either mild upper respiratory symptoms (cough, fever and rhinorrhoea) or more serious lower respiratory tract involvement, such as bronchiolitis and croup; the latter is observed mainly in younger children. Additionally, HCoV-NL63 is the aetiological agent for up to 10% of all respiratory diseases (Fielding 2011).

Enzyme-linked immunosorbent Assays (ELISAs) are the most commonly used methods for the detection and quantitation of antibodies, antigens and a large variety of other molecules (Engvall and Perlmann 1972). As this assay allows for the testing of large
sample numbers, without the need for radioactive material, ELISAs are extremely popular in both research and clinical settings. In addition, only soluble molecules are used as ligands in ELISA; although even purified viruses have been shown to attach directly to the surface of the microtiter plates (Kammer 1983a; Hajimorad and Francki 1991a). Protein preparations made from non-soluble cell membranes, fractionated bacterial extracts, non-soluble hapten-conjugated proteins, immune complexes, etc., cannot be used in ELISA unless completely solubilised with an appropriate detergent or solvent. Therefore, it is crucial that when choosing the method and reagents to solubilise the proteins, three major requirements should be fulfilled. Firstly, the antigenicity of the solubilised proteins must be preserved in order to allow the primary antibody to recognize its epitope. Secondly, treatment with detergent should not cause ligand(s) to bind non-specifically to the primary and secondary antibody. Lastly, the reagents used in the process should not hamper the spontaneous adsorbance of the solubilised ligand(s) to the surface of the ELISA microtiter plates (Kammer 1983a; Hajimorad and Francki 1991a). Therefore, in this study SDS was used to disrupt non-covalent interactions (e.g., ionic, hydrophobic) in the proteins, thereby denaturing them, and causing the molecules to lose their native conformation. This phenomenon therefore enables the antibodies to detect the specific antigen that is coated on the plate. In addition, the SDS coating gives the protein a high net negative charge that is proportional to the length of the polypeptide chain, that stabilizes the protein structure.

This study was designed to establish a direct enzyme-linked immunosorbent assay (ELISA) using coated antigen capable of detecting and distinguishing between HCoV-NL63 N and SARS CoV N antibodies. Five, specific pathogen free (SPF), Balb/c female mice were used to produce polyclonal antibodies for HCoV-NL63 and SARS-CoV nucleocapsid protein. The ELISA assay is a quantitative determination of a specific surface attached ligand, using a primary antibody prepared against the antigen. Additionally,
antibodies raised against a protein are either conformation or sequence specific (Lechtzier, Hutoran et al. 2002). Two mice (nl63-1 and 2) were injected with HCoV-NL63 N purified viral fusion protein and three mice (sars-1, sars-2 and sars-3) were injected with SARS-CoV N purified viral fusion protein. The primary injection was in CFA, which was originally developed by Jules Freund in the 1940’s. Freund’s Adjuvant is designed to provide continuous release of antigens necessary for stimulating a strong, persistent immune response (Freund, Stern et al. 1947). The main disadvantage of Freund’s adjuvant is that it can cause granulomas, inflammation at the inoculation site and lesions. Moreover, the mycobacterium in CFA attracts macrophages and other cells to the injection site which enhances an immune response. For this reason, the CFA was used for the initial immunisation. To minimize side-effects, IFA was used for the subsequent booster injections. The absorbance readings were measured using the corrected values from the readings of microtiter plate reader. All the results were recorded at an OD$_{450}$ nm as previously described. Thereafter, statistical analysis for was performed as previously described. Figures 3.1 and 3.2 represent the detection of viral antibodies against HCoV-NL63 and SARS-CoV N coated plates. Serum antibodies were detected on all N coated plates which indicated the production of the polyclonal antibodies by the respective mice. However, cross-reactivity between SARS-CoV N and the serum HCoV-NL63 N specific antibodies was also observed. The antibody titres of bleed 2 were higher in comparison with bleed 1. This was expected since an exposure of mice to an antigen over time will increase the production of antibodies. There was no significance difference between the antibody titers of the mice represented Figures 3.1 and 3.2. However, the assay results of the effect of SDS on antibody detection indicated a significant difference (P<0.05) between the three mice that were immunized with SARS-CoV N antigen (Figure 3.3 and 3.4). An increase in the antibody titers was observed on the 0.05% SDS treated assays.
SDS has the ability to denature protein structures therefore it enables an increase in the detection of the antigens coated on the microtiter plates.

HCoV-NL63 N immunized mice produced antibodies against HCoV-NL63 N coated plates that bind SARS-CoV N antigen; this indicates that cross-reactivity exists (Figures 3.1, 3.2, 3.3 and 3.4). These results are consistent with other findings from several research groups which have conducted similar studies using serum samples which were subsequently screened for antibodies against HCoV-NL63 using enzyme-linked immunosorbent assays (ELISA) (Shao, Guo et al. 2007; Dijkman, Jebbink et al. 2008; Abdul-Rasool and Fielding 2010). 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). An ELISA assay based on the N protein demonstrated high seropositive results for serum samples from children younger than 20 years old (Shao, Guo et al. 2007; Abdul-Rasool and Fielding 2010). Furthermore, the presence of maternally-acquired N-directed antibodies were detected in serum by ELISA, usually decreasing within the first 4-5 months of life (Shao, Guo et al. 2007; Dijkman, Jebbink et al. 2008). Furthermore, potent neutralizing activity directed against HCoV-NL63 S protein was detected in virtually all serum samples from patients eight years of age or older; suggesting that HCoV-NL63 infection in humans is common and usually acquired during childhood (Hofmann, Pyrc et al. 2005; Abdul-Rasool and Fielding 2010).
Figure 3.1: Detection of viral antibodies against an HCoV-NL63 N coated plate. A comparison of the first bleeds post-immunisation (bleed 1) and second bleed post-immunisation (bleed 2) of anti-NL63 and anti-SARS antibody titres. The y axis represents the absorbance readings (OD$_{450\text{nm}}$) and the x axis represents the respective antibodies. The data is presented as corrected values at a dilution titre of 1/100.

Figure 3.2: Detection of viral antibodies against a SARS-CoV N coated plate. A comparison of the first bleeds post-immunisation (bleed 1) and second bleed post-immunisation (bleed 2) of anti-NL63 and anti-SARS antibody titres. The y axis represents the absorbance readings (OD$_{450\text{nm}}$) and the x axis represents the respective antibodies. The data is presented as corrected values at a dilution titre of 1/100.
Additionally, SARS-CoV infections appear to stimulate cross reactive antibody responses to other human coronaviruses, including HCoV-NL63 (Chan, Cheng et al. 2005). Such cross-reactivity between human coronaviruses has been reported previously for immunofluorescence and complement fixation tests (McIntosh, Chao et al. 1974b; Monto and Rhodes 1977a). In fact, false positive results were also reported for SARS-CoV detection using ELISA tests based on the recombinant N antigen (Woo, Lau et al. 2004a). This cross-reactivity is most likely due to the presence of cross-reactive antigenic epitopes of coronaviruses (Chan, Cheng et al. 2005). Cross-reactivity can be resolved by identifying and directing detection against immunogenic epitopes which increase specificity as the use

Figure 3.3: Detection of viral antibodies treated with SDS against an HCoV-NL63 N coated plate. A comparison of post-immunisation (bleed 2) non-treated with SDS and 0.05% SDS treated anti-NL63 and anti-SARS antibodies. The y axis represents the absorbance readings (OD₄₅₀nm) and the x axis represents the respective antibodies. The data is presented as corrected values at a dilution titre of 1/100. An asterisk (*) designates significant difference when P<0.05.
of synthetic peptides, opposed to the full length protein, may show a reduction in cross-reactivity (Woo, Lau et al. 2004a; 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 (Woo, Lau et al. 2004a). An awareness of this cross-reactivity is imperative when developing antigen and/ or antibody-diagnostic assays and selecting the most suitable antigen or antigen target is important (Abdul-Rasool and Fielding 2010). Furthermore, future studies need to be conducted that will address this cross-reactivity by competitive inhibition experiments with purified nucleocapsid proteins.

Figure 3.4: Detection of viral antibodies treated with SDS against an SARS-CoV N coated plate. A comparison of post-immunisation (bleed 2) non-treated with SDS and 0.05% SDS treated anti-NL63 and anti-SARS antibodies. The y axis represents the absorbance readings (OD \textsubscript{450nm}) and the x axis represents the respective antibodies. The data is presented as corrected values at a dilution titre of 1/100. An asterisk (*) designates significant difference when P<0.05.
3.5 CONCLUSION

Polyclonal antibodies were successfully produced and validated using a direct enzyme-linked immunosorbent assay. Furthermore, additional studies need to be conducted to gain better understanding of cross-reactivity among all human coronaviruses. This will improve serological responses to coronaviruses for purposes of laboratory diagnosis as well as for understanding pathogenesis and immunity. In a study by Severance and colleagues, they revealed in their demographic analysis that race, socioeconomic status, and smoking status were all identified as risk factors for coronavirus exposure (Severance, Bossis et al. 2008). Furthermore, higher rates of sero-positivity were observed in African Americans, smokers, and individuals of low socioeconomic status. All three of these risk factors have been previously shown to render individuals more prone to a variety of respiratory illnesses. The latter then intensifies our pursuit to conduct further studies on this virus. These studies will enable us to create methods of detecting and treating this disease which could become a major problem in South Africa where poverty and AIDS are rife.
3.6 REFERENCES


Drexler, J. F., F. Gloza-Rausch, et al. (2010a). "Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of
coronaviruses based on partial RNA-dependent RNA polymerase gene sequences."


Human Coronavirus NL63 (HCoV-NL63), was discovered in 2004 by a Dutch research team. This virus was isolated from a nasopharyngeal aspirate of a 7 month-old child suffering from bronchiolitis, conjunctivitis and fever. HCoV-NL63 has been shown to infect mainly children and the immunocompromised, who presented with either mild upper respiratory symptoms (cough, fever and rhinorrhoea) or more serious lower respiratory tract involvement such as bronchiolitis and croup. In fact, HCoV-NL63 is the aetiological agent for up to 10% of all lower and upper respiratory diseases. The characteristic genome organisation of coronaviruses can be observed as the 5' two-third of the genome contains two large open reading frames (ORF), ORF1a and ORF1b. In the 3' part of the genome, genes encoding four structural proteins are found: spike (S), envelope (E), membrane (M), and nucleocapsid (N). The SARS-CoV constitutes a fifth HCoV, which was in circulation for a limited time during 2002 and 2003, when this novel virus appeared in humans; it caused an outbreak affecting at least 8,000 people. Mortality rate was high, at approximately 10%. The N protein is one of the major virion structural proteins of SARS-CoV and HCoV-NL63. The major function of N protein is to assemble the RNA of coronavirus. Unlike the helical nucleocapsids of non-segmented negative-strand RNA viruses, coronavirus ribonucleoprotein complexes are quite sensitive to the action of ribonucleases. A significant portion of the stability of the nucleocapsid may derive from N–N monomer interactions.

In this study, the NL63-N/pFN2A and SARS-N/pFN2A plasmid constructs were confirmed by restriction endonuclease digest and sequence verification. The sequenced chromatographs obtained from Inqaba Biotec were consistent with sequences from the NCBI GenBank. Proteins were expressed in a KRX Escherichia coli bacterial system and analysed using 15% SDS-PAGE and Western Blotting. The Flexi® vector system has proven to be a suitable method for the heterologous expression of the SARS-CoV and
HCoV-NL63 N proteins, however, the expression protocol had to be adequately optimized to be able to achieve the results that were presented in this study. There were several methods of optimization that were employed such as (i) obtaining optical density readings for optimum bacterial growth during protein expression; (ii) utilization of lysozyme to solubilise the proteins and, (iii) due to the instability of the N protein the temperature was also optimized. All the previously outlined optimization strategies have resulted in successful protein expression of the N protein in the Flexi® vector system.

Thereafter, GST-tagged proteins were purified with an affinity column purification system. Purified fusion proteins were subsequently cleaved with Pro-TEV Plus protease, separated on 15% SDS-PAGE gel and stained with Coomassie Brilliant Blue R250. The viral fusion proteins were subsequently used to immunize Balb/c mice in order to produce polyclonal antibodies. A direct enzyme-linked immunosorbent assay (ELISA) was used to analyze and validate the production of polyclonal antibodies by the individual mice. Serum antibodies were detected on all N coated plates which indicated the production of the polyclonal antibodies by the individual mice. There was no significance difference between the antibody titers that were not treated with SDS. However, the assay results of the effect of SDS on antibody detection indicated a significant difference between the three mice that were immunized with SARS-CoV- N antigen. An increase in the antibody titers was observed on the 0.05% SDS treated assay. SDS has the ability to denature protein structures therefore it enables an increase in the detection of the antigens coated on the microtiter plates.

Additionally, cross-reactivity was observed on all the ELISA microtiter plate readings. These results are consistent with other findings from several research groups which have conducted similar studies using serum samples which were subsequently
screened for antibodies against HCoV-NL63 using an ELISA. This cross-reactivity is most likely due to the presence of cross-reactive antigenic epitopes of coronaviruses. An awareness of this cross-reactivity is imperative when developing antigen and/or antibody-diagnostic assays and selecting the most suitable antigen or antigen target is important. Furthermore, future studies need to be conducted that will address this cross-reactivity by competitive inhibition experiments with purified nucleocapsid proteins.

Future prospects in coronavirology include interaction studies of N with other structural proteins such as the M, E and ORF3 proteins. Currently, there are no vaccines available for HCoV respiratory viruses. Therefore, it is essential to monitor epidemic patterns and investigate the spread of respiratory infections to efficiently identify, control and prevent epidemics. The development of technologies to accurately identify and detect HCoV-NL63 infections will enhance our knowledge on the true incidence of this virus in the human population. The recent development of the first full-length infectious clone of HCoV-NL63 allows for the systematic experimental study – genes can be modified and/or deleted from the genome – of the functions of the various corresponding HCoV-NL63 proteins, which will lead to a better understanding of the role of the viral genes in infectivity and pathogenicity. This manipulation of the virus genome, in turn, provides a reverse genetics that can lead to the development of HCoV-NL63 based vector vaccines. However, this is hampered by the poor growth of the virus in cell culture, as well as the lack of an appropriate animal model. More studies need to be conducted to obtain enhanced understanding of cross-reactivity among all human coronaviruses. This will improve serological responses to coronaviruses for purposes of laboratory diagnosis as well as for understanding pathogenesis and immunity. In achieving these end points it will be possible to decrease the prevalence of morbidity due to respiratory tract infections and thus increase productivity and quality of life in both first and third world countries.
Appendix

Department of Medical Biosciences, Faculty of Natural Sciences, University of the Western Cape, South Africa.
Appendix 1: An illustration of plasmid maps for eukaryotic expression of viral proteins. (A) pGEM®-T Easy vector map and sequence reference points showing the position of the T-overhangs and the restriction sites in the multiple cloning regions. (B) The vector also contains T7 promoter for bacterial or in vitro protein expression of a protein-coding region. The vector appends an N-terminal glutathione-S-transferase (GST) coding region that can be used to purify the expressed protein. The GST tag contains a TEV protease site for removal of the tag after purification. The vector also contains the lethal barnase gene for positive selection of the insert, an ampicillin-resistance gene for selection of the plasmid and unique endonucleases restriction sites *SgfI* and *Pmel* that allows easy insertion or transfer of the sequence of interest. [Technical Manual pGEM®-T and pGEM®-T Easy and pFN2A (GST) Flexi® Vector Systems. 2009. Promega Corporation].
Appendix 2: P-Values of viral antibodies treated with SDS against an HCoV-NL63 N coated plate. The table represents a comparison between non SDS and SDS treated antibodies. The statistical test that was used to determine these P-values is Mann-Whitney (Independent Samples). There is no statistical significance on the nl63 mice; however the sars mice indicated a significant difference (P<0.05).

<table>
<thead>
<tr>
<th></th>
<th>-SDS</th>
<th>+SDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>nl63-1</td>
<td>0.175</td>
<td>0.995</td>
</tr>
<tr>
<td>nl63-2</td>
<td>0.172</td>
<td>0.573</td>
</tr>
<tr>
<td>P</td>
<td>0.3333</td>
<td></td>
</tr>
<tr>
<td>sars-1</td>
<td>0.282</td>
<td>1.075</td>
</tr>
<tr>
<td>sars-2</td>
<td>0.043</td>
<td>0.649</td>
</tr>
<tr>
<td>sars-3</td>
<td>0.341</td>
<td>0.483</td>
</tr>
<tr>
<td>P</td>
<td>0.0495</td>
<td></td>
</tr>
</tbody>
</table>

Appendix 3: P-Values of viral antibodies treated with SDS against an SARS-CoV N coated plate. The table represents a comparison between non SDS and SDS treated antibodies. The statistical test that was used to determine these P-values is Mann-Whitney (Independent Samples). There is no statistical significance on the nl63 mice; however the sars mice indicated a significant difference (P<0.05).

<table>
<thead>
<tr>
<th></th>
<th>-SDS</th>
<th>+SDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>nl63-1</td>
<td>0.301</td>
<td>1.429</td>
</tr>
<tr>
<td>nl63-2</td>
<td>0.351</td>
<td>0.813</td>
</tr>
<tr>
<td>P</td>
<td>0.1213</td>
<td></td>
</tr>
<tr>
<td>sars-1</td>
<td>0.553</td>
<td>1.224</td>
</tr>
<tr>
<td>sars-2</td>
<td>0.163</td>
<td>0.594</td>
</tr>
<tr>
<td>sars-3</td>
<td>0.321</td>
<td>0.602</td>
</tr>
<tr>
<td>P</td>
<td>0.0495</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 4: Human coronavirus NL63 nucleocapsid protein (N) gene, complete sequence.

CTCTGGTTTCTGATAAGGCACCACATATAGGGTCATCCATGCCTATTGTAAGGGA
ATAAAGATGACAGCAGATTTGCTATTGGAAGTCTAAGAGCTTGACTTGTTGTTTGGTT
GCTAACGAGGCTTAAATCTGTTGATATTTGCACTACGTCATCTCATGTTGATGTTTGGGTT
CTGGTTGAGTTGAGGTATCGCTCTATATACATCACAGCTCCTGCTATAGCTCGTTCTTC
AACTCGTAAACTACAGGACTCTTCTTCGTAAGCACTTCAAGACAAACAGTCTC
GCACTCGTTCTGATTTCTACCAGTCTCTTCTCCGATCTTTGCTGGCTGTACTTTT
GGCTTTAAAGAACATATATGTTTTGATAACCCAGTGAACTGACACTTCTTCTGG
TACCTCACTCTAAGAAACCTAATAAGGCTTTCTCAACCCAGGCTGATAA
GCTTTTCAGTTGAAAGAAACCTCCTGGGAAGGCTGTTCTCCTACCAGAGAGAAA
ATGTTATTCAGTGCATTTCCTCGTGATTTTTAATCACAATAATGGGGGATTCAG
ATCTTGTTCAGAATGTGTGTAGTACCAAAGTTTTCCACAGCTTGCTGAAT
TGAGTTCCTAATCAGGCTGCTGTTATTCTTTGATAGTGAGGTTAGCACTGATGAAG
TGTTGATAATGTTCAGATTACCTACACCTACAAAAATGTCTGTAGCTAAAGGATA
ATAAAGAACCTCTCAATGTTGACAGATTAGTGGTTTACTAACCAGGTTTCAAA
CTACAAAAGAAATGCGACTCACAATCACTTCATGTGTGTCGAAACAGCTACTT
AATGCTTCTATTCCAGAATCTAACCATTGGCTGATGATGATTCAAGCCATTATA
GAAATTGTCAACGAGGTTTTGCAATTAA
Appendix 5: *SARS coronavirus nucleocapsid protein (NP) gene, complete sequence.*

ATGTCTGATAATGGACCCCAATCAAACCAACGTAGTGCCCCCCCACATTACATT
TGTTGGACCCCACAGATCTGAACATAAACCAGAATGGAGGACGCA

TGGGGCAAGGCCAAAACAGCGCCGCACCCCCAAAGGTAAAACCATAATATTCGCT
CTTGGTTACAGCCTACCTCACTGACATGGCAAGGAGAACTTAGATTCCCTCGAG
GCCAGGGCGTCTCAATCAACCAATAGTGCTTACATGACATGACGACGCA

TACCAAGAGACTACCCGACGAGTTGTGTTGTCGACGGCAAAATGAAAGC
TGCCCCAGATGGTACTTCTATTACCTAGGAACTGGGCAACTGAGGAGGACGCT

TGACGGCGCTAACAAGAAGGCATCGTATGGGTTGCAACTGAGGAGGACGCT
CTTGGTTCACAGCCTCACTCAGCATGGCAAGGAGGAACTTAGATTCCCTCGAG
GCCAGGGCGTTCCAATCAACACCAATATGGCTAC

AATACACCCAAAGACCACACTTGCAACCACGCCAACATCCTAAACAAAATATGGCTAC
CGTGCTACAAACTTCTCCTCAAGGAAACAGCATTGCGCAAAAGGCTTCTACGC

AGAGGGAAGCAGAGGCGGCAGCAAGCGCTCTTCTCGCTTCTCATCAGTAGTC
GCGGTAAATTCAGAAGAAATTCACCTGCTGAGGCAGGATAGGGGAATTTCTCCTGCT
CGAATGGCTAGCGGAGGTGTTGAAACTGCCCTCGCTATTTGCTGCTAGACG

ATTGAACCAGCTTGAGAGCAAAGTTTCTGGTAAAGGCCAACAACAACAGGAC
CAAAGTCACTAAGAAATCTGCTGCTGAGGCATCTAAAAGGCTCTGCACAAA
ACGTACTGCGCAAAACAGTACAAACGTCACTCAGCTATTGGGAGACGTGGTC
CAGACAAAGGCGAATTCCGGGAACCAAGACCTAACCAGCAGAAC

TGATTACAAACATTGGCCGCAAATTTGCTCAAATTGCTGCTGCGC

ATTCTTTGGAATGTGACGCTATTGGGACAGTACACCTTGGGAAACATGGC
TGACTTATCATGGAGCCATTAAATTGGATGACAAAGATTCACAAATGAC
AAGGCTATACGTGAACAAGCACTTGGACGCTACAAAAACATTTCCACCAAC
AGACGCTAAAAGGAACAAAAAGAAGACTGATGAGCTAGCCTTGGCCG
CAGAGAcCAAAGAAGCACTGGCCTTTCTCCTTCGTCGACAGTGA
Appendix 6: HCoV–NL63 N protein sequence obtained from NCBI GenBank with the protein id: ABI20791.1.

MANVNWADDRAARKFPFPSSFYMPPLLVSDDKAPYRVI PRNLVPIGKNKDEQIGYWNVQERWRMRGGQRVRDVLPKVFYYLGTGPHKDLKFRQRSDGVVVVAKEGAKTVNTSLGRNKRNQKPLEPKFSIALPPELSVVEFEDRSNNSSRASSRSSTRNNRDSRSRSTTQLSQQSRTQSSDANSQSLVAAAVLAKNLGFDNQSKPSSSGTSTPKKPKPFLSQRADKPSQLKRPWKRVPTREEENVIQCFGPRDFNHNMGSDDLVQNGVDAKGFQLAEHPQAALFFDSEVSTDEGVNDVQITYTYKMLVAKDNKNLPKFIEQISAITK
PSSIKEMQSSHVNVQTNLNASIPSKLADDSSAIIEIVNEVLH

Appendix 7: SARS-CoV N protein sequence obtained from NCBI GenBank with the protein id: AAQ63890.1

MSDNGPQSNQRSAPRIFGPGPTDSTDNNQNGGRNGARPQKRRPQQLPNNTASWFTALTQHGEELRFPRGQVPINTNSGPDDQIGYRRATRRVRRGDDGKMKELSPRWYFYLGTPGAASLYGANKEGIVVWVATEGALNTPKDHIGTRNPNNAAATVLQLPQGTTLPKGFAESRGGSGAAASSRSSRSRSRNSTPGSSRGNSPARMSAGGETALALLLDRLNQLESKVS ngKGGQQQQGQTVTKSSAAEASKKPRQKRTAKQYNVTQAFGRGPEQTQGNGFDQDLIRGQTYKHWPQIAQFAPSAASFGMSRIGMEVTPSGTWLT YHGAIKLDKDQPQKDNNVLLNKIHADAYKTFFPPTEPKKDKKKTDEAQLPQRTQKQPTVTLLPAADMDDFSRALQNSMSGASADSTQA