# Cloning and expression of human cyclophilin A and its interaction with human coronavirus NL63 nucleocapsid protein

by

Anele Gela

Thesis Submitted in Fulfilment of the Requirements for the Degree

MSc

Medical Biosciences

Faculty of Natural Sciences

University of the Western Cape

Supervisor: Professor BC Fielding

Department of Medical Biosciences
University of the Western Cape

Co-Supervisor: Dr C Willemse

Department of Medical Biosciences
University of the Western Cape

November 2011

DECLARATION

I, Anele Gela, declare that this thesis, "***Cloning and expression of human cyclophilin A and its interaction with human coronavirus NL63 nucleocapsid protein***" hereby submitted to the University of the Western Cape for the degree of *Magister Scientiae* (MSc) has not previously been tendered by me for a degree at this or any other university or institution, that it is my own work in design and in execution, and that all materials contained herein have been duly acknowledged.

Anele Gela　　　**:**　..

Date Signed　　**:**　..

UNIVERSITY *of the*
WESTERN CAPE

DEDICATION

This study is dedicated to Xoliswa Dorothy Gela and Singalakha Mlungwana Gela.

ACKNOWLEDGEMENTS

UNIVERSITY *of the*

WESTERN CAPE

ABSTRACT

Coronaviridae family is composed of a number of ribonucleic acid (RNA)-containing viruses currently classified into two genera, the coronavirus and torovirus. The family is classified together with the *Arteviridae* in the order *Nidovirales*. Coronaviruses are enveloped single stranded positive sense RNA viruses about 80-160 nm in diameter. The coronavirus is, as in the case of all positive sense RNA virus, a messenger, and the naked RNA is infectious. The 5′-two thirds of the genome encodes for a polyprotein that contains all the enzymes necessary for replication, whereas the 3′-one third encodes for all the structural proteins that mediate viral entry into the host cell. The structural proteins include spike (S), envelope (E), membrane (M) and nucleocapsid (N) proteins.

Nucleocapsid protein is one of the most crucial structural components of coronaviruses; hence major attention has been focused on characterization of this protein. Some laboratories have demonstrated that this protein interferes with different cellular pathways, thus implying it to be a key regulatory component of the virus (Zakhartchouk, Viswanathan et al. 2005). Furthermore, it has been shown that severe acute respiratory syndrome (SARS)-N protein interacts with cellular proteins, including cyclophilin A (CypA), heterogenous nuclear ribonucleoprotein (hnRNP) A1, human ubiquitin-conjugating enzyme, cyclin dependent kinase (CDK)-cyclin complex protein, Ikappaßalpha (IkBα), cytochrome (Cyt) P450 etc. For the purpose of this study, the focus is based on CypA interaction with human coronavirus (HCoV) NL63-N protein. These interactions might play a role in the pathology of HCoV-NL63. Using glutathione-S-transferase (GST), the interaction of CypA with the nucleocapsid protein can be clearly demonstrated to be direct and specific. Since the N protein is involved in viral RNA packaging to form a helical core, it is suffice to say that both NL63-N and CypA are possibly within the HCoV-NL63 replication/transcription complex and NL63-N/human CypA interaction might function in the regulation of HCoV-NL63 RNA synthesis.

In addition, the results will demonstrate that HCoV-NL63-N has only a specific domain for interacting with CypA.

KEY WORDS

Human coronavirus NL63, severe acute respiratory syndrome (SARS), nucleocapsid protein, human cyclophilin A (CypA), protein-protein interaction, human protein expression, bacterial expression of proteins, GST-pulldown assay.

TABLE OF CONTENTS

UNIVERSITY *of the*
WESTERN CAPE

LIST OF ABBREVIATIONS

%	Percentage
>	More than
<	Less than
A	Adenine
ACE	Angiotensin Converting Enzyme
Ala	Alanine
AMV	Avian Myeloblastosis
APS	Ammonium persulfate
Arg	Arginine
Asn	Aspartate
BCV	Bovine Coronavirus
bp	Base pairs
BSA	Bovine Serum Albumin
C	Cytosine
Ca^{2+}	Calcium ions
CD	Cluster of Designation
CDK	Cyclin Dependent Kinase
cDNA	Complementary Deoxyribonucleic Acid
cfu	Colony Forming Units
cm	Centimeter
cm^2	Square centimetre
CO_2	Carbon Dioxide
COPD	Chronic Obstructive Pulmonary Diseases
CsA	Cyclosporine A
CTD	C-terminal Binding Domain

Cyp	Cyclophilin
CypA	Cyclophilin A
Cyt	Cytochrome
Da	Dalton
DAA	Direct-Acting Antivirals
DEB025	Alisporivir
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphate
E	Envelope
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EIAV	Equitine Infectious Anemia Virus
FIP	Feline Infectious Peritonitis
FIV	Feline immunodeficiency virus
FBS	Fetal Bovine Serum
HE	Hemaglutin Esterase
g	Gram
G	Guanine
GC	Guanine Cytosine
Gly	Glycine
gp	glycoprotein
GST	Glutathione-S-transferase
HCoV	Human Coronavirus
HCoV HKU1	Human Coronavirus Honk Kong U1
HCoV NH	Human Coronavirus New Haven
HCV	Hepatitis C Virus
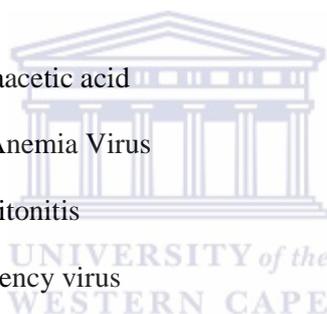
HE	Hemaglutinin esterase
His	Histidine
HIV	Human Immunodeficiency Virus
hnRNP A1	heterogenous nuclear Ribonucleoprotein A1
HQ	Histidine Glutamine
HRP	Horseradish Peroxidase
IBV	Infectious Bronchitis Virus
IF	Immunofluorescence
IgG	Immunoglobulin
Ikßα	Ikappaßalpha
Ile	Isoleucine
IP	Immunoprecipitation
IPTG	Isopropyl-ß-D-thio-galactoside
K	Kilo
Kbp	Kilo bse pair
KDa	Kilo-Dalton
l	litre
LB	Luria-Bertani
μ	Micro
m	meter
M	Membrane
M	Molar
MCF	Michigan Cancer Foundation
MgCl$_2$	Magnesium Chloride
MHV	Mouse Hepatitis Virus
Min	Minutes
mM	milli Molar
mRNA	messenger Ribonucleic Acid

n	nano
N	Nucleocapsid
NaCl	Sodium Chloride
NCBI	National Centre for Biotechnology Information
NF-kß	Nuclear factor kappa light-chain enhance of activated B cells
NLS	Nuclear Localisation Signal
NoLS	Nucleolar Localisation Signals
NRF	National Research Foundation
NSP	Nonstructural protein
NTD	N-terminal Domain
OC	Organ Culture
OD	Optical Density
ORF	Open Reading Frame
p	Pico
PBDG	Porphobilinogene Deaminase
PHEV	Porcine Hemaglutinating Encephalomyelitis Virus
PCR	Polymerase Chain Reaction
PPIA	Peptidylprolyl Isomerase A
Pro	Proline
RBD	RNA Binding Domain
RNA	Ribonucleic Acid
rpm	Revolutions per minute
RT	Reverse Transcription
S	Spike
SARS	Severe Acute Respiratory Syndrome
SDS PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
Sec	Second
Sg	Subgenomic
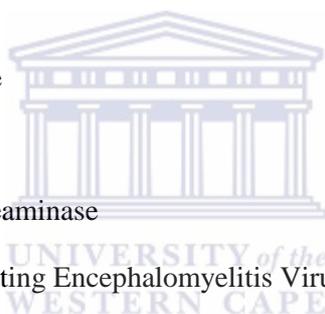
SR	Serine/Arginine
T	Thymine
TGEV	Transmissible Gastroenteritis Virus
Thr	Threonine
TMED	Tetramethylethylenediamine
Trp	Tryptophan
TRS	Transcription Regulatory Sequence
UV	Ultraviolet
Vpr	Viral protein R
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside

UNIVERSITY *of the*

WESTERN CAPE

LIST OF TABLES

LIST OF FIGURES

UNIVERSITY *of the*
WESTERN CAPE

LIST OF APPENDIXES

SARS-N amino acid sequence

NL63-N amino acid sequence

229E-N amino acid sequence

OC43-N amino acid sequence

Feline CoV-N amino acid sequence

TGEV-N amino acid sequence

Bovine CoV-N amino acid sequence

Murine hepatitis virus-N amino acid sequence

HIV-1 capsid protein

HIV-1 p24 protein

HIV-1 Vpr protein

HIV-1 heparan

HCV NS5A protein

HCV NS2 protein
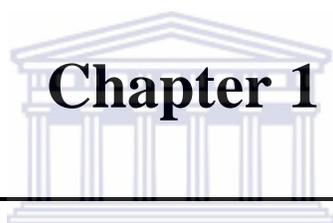
Chapter 1

Literature review

1.1. BACKGROUND

Coronavirus is a genus of animal virus belonging to the family *Coronaviridae*. The name "coronavirus" is derived from the Greek word κορώνα, meaning "crown". The virus envelope is characterized by a ring of small bulbous structures when analyzed by electron microscopy and this gives the virus its characteristic appearance (Liu, Yang et al. 2007; Versteeg, van de Nes et al. 2007; Lin, Feng et al. 2008). This morphology is actually formed by the viral spike (S) peplomers, which are proteins that populate the surface of the virus and determine host tropism (Buchholz, Bukreyev et al. 2004). Coronaviruses are currently classified into two genera, the coronavirus and torovirus. They, together with *Arteriviridea* are grouped in the order *Nidovirales*, named for the Latin *nidus*, meaning "nest", as all viruses in this order produce a 3′ co-terminal nested set of subgenomic mRNA's during infection (Yeung, Yamanaka et al. 2006). In comparison to toroviruses that have a tubular nucleocapsid that appears toroidal in shape, coronaviruses have a helical nucleocapsid 10-20 nm in diameter in their virion. In addition, coronavirus virion is roughly spherical in shape, whereas torovirus virion is disk shaped or rod shaped. Contrary to well-studied, described and published data on the coronavirus, the torovirus, which is comprised of one pathogen of horses, one pathogen of cattle, a presumptive human torovirus, and a possible torovirus of swine, have only been described recently. About three groups of coronaviruses are recognized based on their serologic and genotypic characteristics: group 1 and 2 contain mammalian viruses, whereas group 3 contains only avian viruses (de Haan, Haijema et al. 2008; Woo, Lau et al. 2009; Jackwood, Boynton et al. 2010). Group 2 viruses are also characterized by the presence of a gene encoding the hemaglutin esterase (HE) protein, which is absent in the other groups.

Coronaviruses are enveloped single stranded positive sense RNA viruses about 80-160 nm in diameter. Coronaviruses are somewhat larger in their genome size compared to toroviruses. All coronaviruses employ a common genome organization. The 3′-one third of the genome encodes for structural proteins. These structural proteins include spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins. For the purpose of this study, the structural protein of interest is the nucleocapsid protein. Here we compare the characteristics of severe acute respiratory syndrome coronavirus (SARS)-N with NL63-N, so that we can elucidate whether NL63-N will exhibit the same characteristics as SARS-N.

1.2.   DISCOVERY OF CORONAVIRUSES

The history of coronaviruses began in 1965 when Tyrrel and Bynoe found that they could passage a virus named B814. This virus was accidentally discovered in human embryonic tracheal organ cultures obtained from the respiratory tract of an adult with a common cold. Unfortunately, Tyrrel and Bynoe could not grow the virus in tissue culture at the time due to its unusual properties (Tyrrell and Bynoe 1958; Tyrrell, Bynoe et al. 1959; Bynoe, Hobson et al. 1961). However, at about the same time, Hamre and Procknow were able to grow the virus in tissue culture from volunteers with colds. The literature indicates that both B814 and Hamre′s virus were ether-sensitive and therefore presumably required an envelope for infectivity (Kennedy and Johnson-Lussenburg 1975; Schreiber, Kamahora et al. 1989). Surprisingly, these two viruses were not related to myxo- or paramyxoviruses that have a lipid-containing coat. Using a technique that is similar to Terryl and Beynoe, Hamre and Procknow reported the recovery of multiple strains of ether-sensitive agents from the human respiratory tract (Tyrrell, Bynoe et al. 1960). These viruses were termed OC to designate that they were grown in organ cultures.

Within the same time, Almeida and Tyrrel performed electron microscopy on fluids from organ cultures infected with B814 and found particles that resemble the infectious bronchitis virus (IBV) of chicken (Kapikian 1975). Hamre′s 229E agent and the OC viruses identified by McIntosh had a similar morphology. The coronavirus group was officially accepted as a new genus in the late 1960s after Tyrrel demonstrated human strains and a similar number of animal viruses to be morphologically similar as seen through electron microscopy (Kapikian 1975). However, B814 strain was shown by Bradburne to be serologically non-identical with either OC43 or 229E. Human and animal coronaviruses were segregated into 3 broad groups based on their antigenic and genetic make-up. Group 1 contained coronavirus 229E and other viruses. Group II contained OC43 and group III was made up of avian IBV and other related viruses.

In the three decades after discovery, human strains OC43 and 229E were exclusively studied largely because they were the easiest ones to work with. Given the enormous variety of animal coronaviruses, severe acute respiratory syndrome (SARS) emerged in 2002-2003 as a coronavirus from Southern China and spread throughout the world (Wang, Ji et al. 2003; Lu, Zhao et al. 2004; Hao, Chen et al. 2006). Sequencing of SARS differed sufficiently from any of the known human or animal coronaviruses. It is unclear how the virus entered the human population and palm civets were suspected to be the natural reservoir of the virus. Sequence analysis of the virus isolated from Himalayan palm civets revealed that this virus contained a 29-nucleotide sequence that is not found in most human isolates, in particular those involved in the worldwide spread of the epidemic (Graham and Baric 2010).

Following the outbreak of SARS in 2002-2003, new human coronaviruses (HCoV) have been discovered. After the discovery of SARS it became evident that highly pathogenic strains of coronavirus can still emerge and as a result these viruses have received global attention. Through

increasing efforts in scientific research five human coronaviruses are now known, including OC43, 229E, SARS HCoV, HCoV NL63 and HCoV HKU1. HCoV NL63 was discovered in 2004 by Lia van der Hoek in Amsterdam from a 7 month old girl presenting with respiratory symptoms. Shortly after, Fouchier et al. reported the identification of a coronavirus NL isolated from an 8 month old boy with pneumonia and grown from a clinical sample that was obtained in April 1988. Genome sequence analysis demonstrated the virus belongs to group 1 and is closely related to HCoV NL63 (Lin, Feng et al. 2011). A few months later, Esper et al. found evidence of a human respiratory coronavirus from children younger than 5 years of age, which was designated HCoV New Haven coronavirus (HCoV NH). This was achieved through the use of molecular probes that targeted conserved regions of the coronavirus genome (Shao, Guo et al. 2007). This approach was based on the theory that the viral replicase gene of all coronavirus has conserved genetic sequences that encode indispensable, essential functions and that these sequences could be targeted for virus identification and discovery. It is now accepted that these three viruses are all essentially the same virus.

To date, the latest discovery is a group II coronavirus, known as HCoV HKU1, which is closely related to OC43. It was discovered from a 71 year old man whom had returned from a previously SARS-endemic area in China presenting with fever and productive cough (Chan, Woo et al. 2008; Chan, Tse et al. 2009; Woo, Lau et al. 2009). SARS screening for this patient was negative, and the coronavirus sequence was amplified by reverse transcription-polymerase chain reaction (RT-PCR) from respiratory specimen using primers that targeted conserved regions of the viral replicase.

1.3. HUMAN CORONAVIRUS NL63

1.3.1. Discovery

Human coronavirus NL63 (HCoV NL63) was discovered in 2004 following the outbreak of the deadly SARS outbreak in China. After this outbreak coronaviruses received international attention, and through increasing efforts on research new human coronaviruses have been discovered recently, and HCoV NL63 in the Netherlands is amongst those new discoveries. In 2004, Lia van der Hoek et al. reported the discovery of a group I coronavirus, HCoV NL63, from a 7 month old girl with coryza, conjunctivitis, fever, and bronchiolitis (Shao, Guo et al. 2007; Dijkman, Jebbink et al. 2008; Woo, Lau et al. 2009).

1.3.2. Symptoms and prevalence

Coronavirus infection of human was usually not associated with severe diseases, but following the discovery of SARS it became evident that highly pathogenic strains of human coronavirus can still evolve. Human coronaviruses, except SARS-CoV, generally cause disease much like the common cold, but they have also been associated with more severe lower respiratory tract conditions, especially in frail patients (Woo, Lau et al. 2005; Dijkman, Jebbink et al. 2006; Shao, Guo et al. 2007). This has been implicated more commonly with SARS which causes both upper and lower respiratory infections. The viruses can infect both the enteric and central nervous system, but they are predominantly involved with respiratory illnesses. They are thought to be transmitted by direct contact, droplets, or contaminated human excreta. In the latter, precautionary measures should be instigated immediately and accordingly. The general clinical symptoms associated with the coronavirus include bronchiolitis, rhinitis, pharyngitis, pneumonia, fever, and conjunctivitis

(Hofmann, Simmons et al. 2006). However, the clinical symptoms that are specifically associated with NL63 still need to be determined. Patients suspected of being infected with the coronavirus need to be quarantined and treated with intense care to minimize transmission to others. The same can be said about the very recently described HCoV HKU1, which was identified in a 71 year old patient with chronic obstructive airways disease (COPD) (Chan, Woo et al. 2008).

1.3.3. Seasonal incidence

During the 2002-2003 outbreaks, SARS infection was reported in 29 countries in North America, South America, Europe and Asia. Overall 8 098 infected individuals were identified, with 774 SARS-related fatalities (St-Jean, Jacomy et al. 2004; Patrick, Petric et al. 2006; Bartlam, Xu et al. 2007). Ongoing research using serologic techniques has resulted in a considerable amount of information regarding the epidemiology of the human respiratory coronavirus. Literature indicates that in temperate climates, respiratory coronavirus infections occur more often in the winter and spring than in summer and autumn. There is strong evidence that the spread of this virus is seasonal, and is prevalent during the winter season (St-Jean, Jacomy et al. 2004; Wang, Li et al. 2004). A study that was conducted by Vabret et al. in 2002-2003 indicated that the highest prevalence of coronaviruses was in February (In the northern hemisphere during winter). Some studies have shown that coronavirus infection contributes as much as 35 % of the total respiratory viral activity during epidemics (Stadler, Masignani et al. 2003; Wong and Yuen 2005). Overall, the proportion of adult colds produced by coronaviruses was estimated at 15 %. After the discovery of OC43 and 229E it was revealed that the viruses demonstrated periodicity, with large epidemics occurring at 2-3 year intervals. Strain 229E tended to be epidemic throughout the United States,

whereas the OC43 strain was more predisposed to localized outbreaks (Lau, Woo et al. 2004; Patrick, Petric et al. 2006).

1.3.4. Genome and genomic replication

All coronaviruses employ a common genome organization as indicated in figure 1.1. The genome that is specifically associated with HCoV NL63 has the following gene order: 1a-1b-S-ORF3-E-M-N (Wang, Li et al. 2004). The 5′- two thirds of the genome is comprised of replicase gene (22 kbp) whereas the 3′- one third constitutes the genes that encode for structural proteins.
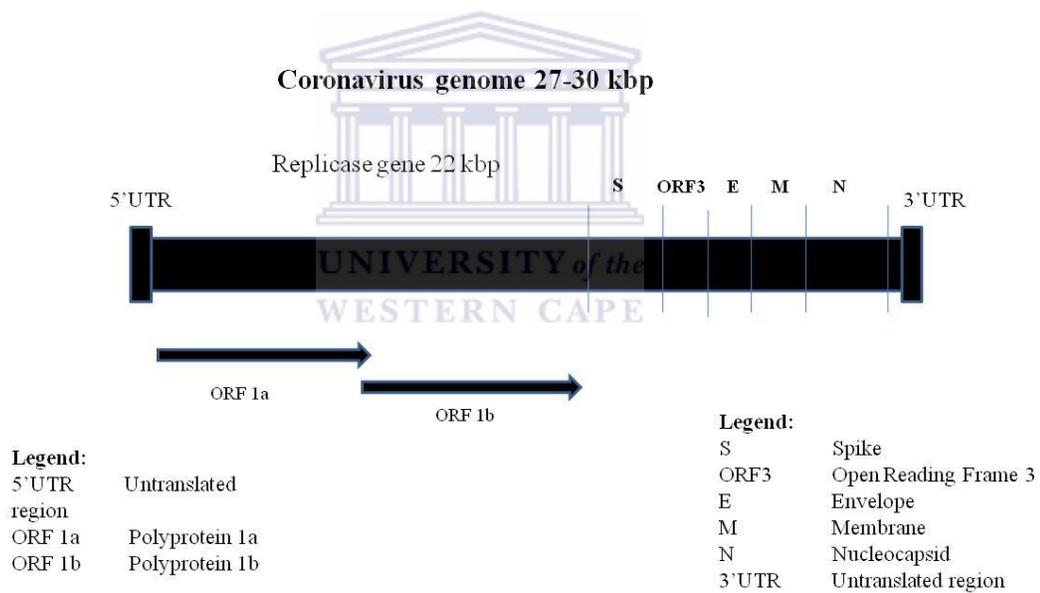
Figure 1.1: Schematic representation of the genomic organization of coronaviruses. The 5′- two thirds of the genome represent two overlapping open reading frames, known as ORF 1a and ORF 1b. These polyproteins constitute approximately 22 kbp of the entire genome. The 3′- one third of the genome, on the other hand, is comprised of structural proteins (adapted from Wang, Li et al. 2004).

The replicase gene encompasses the 5′-two thirds of the genome and is comprised of two overlapping open reading frames (ORFs), ORF 1a and ORF 1b (Woo, Huang et al. 2005; Kim, Lee et al. 2006; Ren, Li et al. 2006). The replicase is the first protein to be made and once the gene encoding the replicase is translated the translation is stopped by a stop codon. This is known as a nested transcript, where the transcript only encodes one gene, it is monocistronic. Expression of the replicase gene is known to be mediated by translation of the genomic RNA that gives rise to the biosynthesis of two large polyproteins, pp 1a (encoded by ORF 1a) and pp 1ab (encoded by ORF 1a and ORF 1b using a ribosomal frameshift at the ORF 1a/1b junction) (Wang, Li et al. 2004). These non-structural proteins are known to have multiple enzymatic functions and might include types of enzymes that are common components of the replication machinery of positive-strand RNA viruses: an RNA-dependent RNA polymerase activity, a 3C-like serine protease activity, a papain-like protease 2 activity, and a superfamily-1 helicase activity (Dijkman, Jebbink et al. 2006). These enzymes are known to be less common in positive-sense RNA viruses, and may therefore be related to the unique properties of coronavirus replication and transcription (Ren, Li et al. 2006).

Expression of the structural gene region is mediated via discontinuous transcription of subgenomic mRNAs, a hallmark of coronavirus gene expression (Zhang, Guy et al. 2007). The capped and polyadenylated RNA is translated into distinct polyproteins that are required for the replication and production of subgenomic mRNAs. The initial polyprotein terminates at a stop codon approximately 12 kb from the 5′ end of the RNA. However, the occurrence of the frequent ribosomal frame shift causes the frame to shift and thus translation continues to the end of the RNA replicase encoding region at 22 kb. The resulting polyproteins are cleaved by virus encoded proteases known as one papain-like and serine-like protease (Chen, Wang et al. 2007). The mechanism by which subgenomic RNAs are produced has been difficult to determine. The first

mechanism proposed was primer-directed synthesis using the negative sense RNA template. In this model, a primer of about 60 nucleotides is transcribed from the 3′ end of the template, which is identical to the 5′ end of the genomic RNA. The primer is proposed to dissociate from the template and to be used by the viral RNA synthetase to reinitiate synthesis at any of the several subgenomic promoters in the (-) RNA template. Evidence to this model includes the fact that each subgenomic RNA has at its 5′ end the same 60 nucleotides that are present at the 5′ end of the genomic RNA, and that there is a short sequence element present at the beginning of each gene that could act as an acceptor of the primer (Shulla, Heald-Sargent et al. 2011).

A more recent model, however, proposes that the bulk of the subgenomic mRNA are produced by independent replication of these RNAs as replicons. Such replication is thought to be possible because the mRNA contains both the 5′ and 3′ sequences present in the genomic RNA, and therefore possess the promoters required for replication. Evidence to this model includes the fact that both positive and negative sense RNAs are present in infected cells (Shulla, Heald-Sargent et al. 2011).

The structural proteins include spike, envelope, membrane and nucleocapsid proteins. The number of subgenomic mRNA produced by a particular coronavirus usually exceeds the number of encoded structural proteins and consequently, coronaviruses are able to express additional, so-called accessory genes (Dijkman, Jebbink et al. 2006; Zhang, Guy et al. 2007; Frieman, Ratia et al. 2009). These genes are interspersed between the structural genes and their number and location varies within the coronavirus genome (Dijkman, Jebbink et al. 2006). The function of these accessory proteins are largely unknown, however, reverse genetic analysis of mouse hepatitis virus (MHV) suggest that they are not required for virus replication (Nedialkova, Ulferts et al. 2009).

1.3.5. Characterized genes

The coronavirus particles consist of four structural proteins that are encoded in the 3′- one third end of the genome: spike (S), membrane (M), nucleocapsid (N) and envelope (E). They also encode for non-structural proteins in the 5′- two thirds of the genome. Some of these proteins are hydrophobic while others are hydrophilic in nature. The hydrophilic proteins are thought to have antigenic properties since they are exposed to the host immune defense mechanism, while hydrophobic proteins are thought to be non-immunogenic. Several studies have shown that most of the nonstructural proteins in SARS are hydrophilic in nature except NSP1, NSP3, X1, X4, which are predicted to have some transmembrane domains, and that most of the structural proteins are hydrophobic except the N protein which is predicted to be highly hydrophilic (Campanacci, Egloff et al. 2003; Nedialkova, Ulferts et al. 2009). These attributes may to some extent reflect the function that is being performed by the viral protein as well as its distribution within the virion. All the nonstructural proteins are synthesized in the cytosol of host cells and then carry out the function of virus replication, including viral RNA replication, viral mRNA transcription and editing, virion assembly, and inhibition of host cell protein synthesis (Neuman, Joseph et al. 2008). Such functional requirements may drive the nonstructural proteins to take soluble form, i.e hydrophilicity.

The structural proteins like E and M are on the surface of a virion, which necessitate them to take hydrophobic form to form a non-polar barrier between the N inside the virion and the outer environment. An exception to this is the S glycoprotein, which is predicted to have a short piece of transmembrane domain near its carboxy (C)-terminus (Kang, Yang et al. 2006; Schaecher, Diamond et al. 2008; Yan, Tan et al. 2009). This transmembrane domain may separate the inner

part from the outer part. For the exception of structural proteins, a hydrophilic form for S- protein is necessary to facilitate an interaction of this nature.

The M protein is a component of the viral envelope that plays a central role in virus morphogenesis and assembly via its interactions with other viral proteins. The E protein, on the other hand, has some similar functions as M protein and there is known interactions between the two proteins during viral replication (Liu, Sun et al. 2010). The E protein seems to play a crucial role in creating the membrane curvature needed to acquire the rounded, stable and infectious particle phenotype. In addition, it induces the formation of hydrophilic pores in cellular membranes (Ren, Li et al. 2006; Yeung, Yamanaka et al. 2006). The S protein mediates viral entry into host cells during infection. The S protein is a type 1 membrane glycoprotein that can be divided into 2 domains viz. S1 (1-717 aa) and S2 (718-1356 aa) based on its similarity to other coronaviruses (Hofmann, Simmons et al. 2006; Lin, Feng et al. 2011). The S1 domain is responsible for binding to cellular receptors, whilst S2 is mainly involved in membrane fusion during viral entry (Kam, Okumura et al. 2009).

1.4. HUMAN CORONAVIRUS NUCLEOCAPSID GENE

Coronavirus N proteins are basic proteins that encapsulate viral genomic RNA to form part of the virus structure. SARS-N protein is known to be highly antigenic and associated with several host cell interactions (Stohlman, Fleming et al. 1983; Hogue and Brian 1986; Zhou, Liu et al. 2008). SARS-N enters the host cell together with the viral RNA and interferes with several cellular processes. These interactions might have a role in the pathology of SARS. Previous studies have shown that protein-protein interactions are critical for viral assembly. Determination of these

interactions and identification of amino acid sequences responsible for such interactions are necessary for the elucidation of the molecular mechanisms of SARS replication and rationalization of anti-SARS therapeutic intervention. SARS-N has been shown to interact with cellular proteins, including cyclophilin A, human cellular heterogenous nuclear ribonucleoprotein A1 (hnRNP A1), and cyclin dependent kinase (CDK)-cyclin complex proteins (Luo, Luo et al. 2004; Luo, Chen et al. 2005; Surjit, Liu et al. 2006; Li, Lai et al. 2011).

The extensively characterized SARS-N protein can be divided into 3 putative domains: N-terminal domain (NTD), RNA binding domain (RBD), and C-terminal domain (CTD). Structural studies of the NTD (residues 45-181) of the SARS-N protein have shown that it acts as a putative RNA-binding domain, whereas the CTD (residues 248-365) acts as a dimerization domain (Tang, Wu et al. 2005). Similar structures to those of the SARS-CoV N protein have also been reported for the NTD and CTD of avian infectious bronchitis virus (IBV) N protein indicating that the aforementioned structural arrangements are common among coronaviruses (Cavanagh 2003). These include NTD residues 19-162 in IBV, analogous to residues 45-181 in SARS-CoV and residues 219-349 in IBV being analogous to residues 248-365 in SARS-CoV (He, Du et al. 2005). This indicates that some degree of conserved regions within the *Coronaviridaea* family, but the significance remains unclear. The NTD and CTD are believed to play an imperative role in interaction with other proteins. Surprisingly, the mid-portion of the protein interacts with hnRNP A1 (Luo, Chen et al. 2005). This association was initially established in mouse hepatitis virus (MHV)-N protein studies. The binding site of hnRNP A1 with MHV-N was determined in two regions of the protein using yeast two hybrid assay. These included amino (N)-terminal region (1-292) and C-terminal end (391-455). The association was mapped to the 161-210 amino acid fragment of SARS-N and to the 203-320 glycine-rich region of hnRNP A1 (Luo, Chen et al. 2005).

The SARS nucleocapsid protein has been shown to have only one binding domain for interacting with hnRNP A1. In contrast, the MHV has two binding domains that are involved in MHV-N/hnRNP A1 interaction, thereby suggesting that the SARS-N might carry out a different binding mode to human hnRNP A1. In addition to hnRNP A1, human cyclophilin A also exhibits high binding affinity to SARS-CoV-N protein. This interaction occurs via the Trp302-Pro310 loop (Luo, Luo et al. 2004). However, the significance of this interaction has not been established.

1.5. CYCLOPHILIN A

1.5.1. Cellular functions
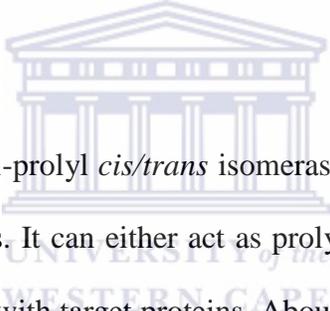
Cyclophilin A (CypA) is a peptidyl-prolyl *cis/trans* isomerase (PPIA) that is involved in several signaling events of eukaryotic cells. It can either act as prolyl bond isomerization catalyst, or it can form stoichometric complexes with target proteins. About 16 different cyclophilin molecules are encoded in the human genome. However, these proteins exhibit differences in tissue distribution, sub-cellular localization and relative abundance, but in terms of their 3-D structure and PPIAse activity they share similarities (Thai, Renesto et al. 2008). Several studies have shown increased expression levels of CypA in tumor cells. Up-regulation of CypA has been found to be a common phenomenon in several tumor types, including hepatocellular carcinoma. However, the role of CypA in tumor cells remains obscure. Some studies have revealed that CypA can up-regulate the expression of many cytokines and drug resistance related genes (Chen, Zhang et al. 2008). Moreover, it has been shown that elevated CypA expressions contribute to drug resistance. This may play a role in clinical resistance to chemotherapy in cancer patients (Chen, Zhang et al. 2008).

CypA has been shown to be secreted in cells in response to inflammatory stimuli and is a potent neutrophil and eosinophil chemo-attractant both *in vitro* and *in vivo* (Yang, Lu et al. 2008). The signaling activity is via CD147 cell surface receptors, but the mechanism underlying CypA-mediated signaling and chemotaxis remain unclear. CypA plays key roles in several different aspects of cellular physiology including the immune response, transcription, mitochondrial function, cell death, and chemotaxis (Watashi and Shimotohno 2007). In addition to these cellular events, a number of reports demonstrate that CypA plays a critical role in the life cycle of viruses, especially human immune virus (HIV)-1, SARS and hepatitis C virus (HCV). These viruses are significant causes of morbidity and mortality worldwide, but current therapies are often insufficient. CypA may provide a novel therapeutic target for the management and possibly cure of these diseases, especially coronaviruses (Luo, Luo et al. 2004).
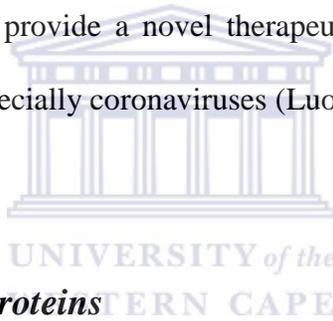
1.5.2. Interaction with viral proteins

Cyclophilin A (CypA) is thought to have numerous cellular functions, including protein folding through *cis/trans* isomerisation, immune response, lipid and protein trafficking or transcription regulation. Despite these attributes, this protein is known to interact with viral proteins and contributes to viral pathogenesis (Saphire, Bobardt et al. 1999). CypA has also been shown to mediate HIV-1 attachment to target cells via heparins. This interaction to heparins was revealed via a domain rich in basic residues similar to known heparin binding motifs. Interaction of CypA and heparins represents the initial step of HIV-1 attachment and is a necessary precursor to gp120 binding to CD4 (Saphire, Bobardt et al. 1999).

Amongst the 16 classes of human cyclophilin, some classes have been demonstrated to bind specifically to HIV-1 capsid proteins. CypA and B have been reported to bind to the HIV-1 gag polyprotein p55gag *in vitro*. The association of HIV-1 virions with CypA was shown along a stretch of amino acid sequences in the capsid domain of p55gag (Thali, Bukovsky et al. 1994). This association was inhibited in a dose dependent manner by cyclosporine A (CsA) as well as SDZ NIM811 ([Melle-4] cyclosporine), a non-immunosuppressive analogue of CsA. However, the latter was shown to be ineffective against SIVMAC, a primate immunodeficiency virus related to HIV-1 (Thali, Bukovsky et al. 1994). The mechanism of action of SDZ NIM811 is clearly different from those of all other anti-HIV agents. In a cell free assay, it has been shown not to be an inhibitor of reverse transcriptase, protease, integrase, and it does not interfere with Rev or Tat function (Steinkasserer, Harrison et al. 1995). SDZ NIM811 inhibits 2-long terminal repeat (LTR) circle formation in a dose-dependent manner, which is indicative of nuclear localization of pre-integration complexes. In addition, it inhibits integration of proviral DNA into cellular DNA (Steinkasserer, Harrison et al. 1995).

In addition to its association with the above mentioned viral protein, CypA has also been reported to interact with HIV-I viral protein R (Vpr) as well as the capsid protein both *in vitro* and *in vivo* (Mascarenhas and Musier-Forsyth 2009). There is a strong correlation between the binding of HIV-1 capsid protein and CypA and the infectivity of certain strains of HIV. The capsid sequence 87His-Ala-Gly-Pro-Ile-Ala92 (87HAGPIA92) encompasses the primary CypA binding site (Gitti, Lee et al. 1996). In contrast to the *cis* proline (Pro) observed in previously reported structures of CypA complexed with model peptides, the Pro in this peptide, Pro-90; bind the CypA active site in a *trans* conformation. In addition to the previously identified Pro-90 site of the capsid protein, two interaction sites between CypA and capsid protein, p24, have also been

reported. Both are located in the C-terminal domain of p24 around Pro-157 and Pro-224 (Saphire, Bobardt et al. 1999). Vpr interaction was demonstrated using surface plasma resonance where a seven-residue motif (RHFPRIW) centered at Pro-35 was revealed for maintaining strong specific binding (Zander, Sherman et al. 2003).

In addition to the above interactions with HIV-1 proteins, CypA has also been shown to interact with hepatitis C virus (HCV) and is a crucial factor for its replication. Their interaction is through domain II of nonstructural protein 5A (NS5A) (Fischer, Gallay et al. 2010). The latter is crucial for HCV replication as a result it constitutes a potential target for antiviral drug development. However, the exact interacting amino acid residues have been determined. Table 1.1 represents the list of known viral proteins that interact with CypA.

Table 1.1: An illustration of viral proteins that are known to interact with cyclophilin A.

Interacting protein	Amino acid residues	Basic/ Acidi	Reference
HIV-1 capsid	pro-90	Neutral	(Zander, Sherman et al. 2003)
HIV-1 p24	pro-90, 157, 224	Neutral	(Fitzon, Leschonsky et al. 2000)
HIV-1 Vpr	pro-35	Neutral	(Zander, Sherman et al. 2003)
HIV-1 heparan	Arg-484, 487, 585 and 588	Basic	(Saphire, Bobardt et al. 1999)
HIV-1 p55	unknown	Unknown	(Thali, Bukovsky et al. 1994)
HCV NS5A	unknown	Unknown	(Chatterji, Lim et al. 2010)
HCV NS2	unknown	Unknown	(Wang and Heitman 2005)
HCV NS5B	unknown	Unknown	(Wang and Heitman 2005)
SARS-N	pro-310	Neutral	(Luo, Luo et al. 2004)

1.6. CORONAVIRUS INFECTION PROCESS

The SARS-CoV nucleocapsid protein is a major antigen in SARS and is known to form a ribonucleoprotein complex with genomic RNA. In addition to its structural role, it is a multifunctional protein involved in viral RNA replication and translation. RNA replication and mRNA transcription occur in virus infected cells. For the infective process to ensue the virus must recognize angiotensin converting enzyme (ACE)-2 receptors on the cell surface (Yeung, Yamanaka et al. 2006). However, group 1 CoV including HCoV-229E utilize aminopeptide N (CD13) as its cellular receptor (Hofmann, Simmons et al. 2006). In addition to host cell receptors, lysosomal cysteine proteases are required for productive infection by some viruses. The SARS-CoV, but not NL63, has been shown to utilize the enzymatic activity of the cysteine protease cathepsin L to infect ACE-2 expressing cells (Huang, Bosch et al. 2006). Therefore, if cathepsin L could be inhibited, SARS infection would not occur. This suggests that 2 coronaviruses that utilize the same receptor nonetheless enter cells through distinct mechanisms. Figure 1.2 represents a schematic diagram of the replication cycle of the coronavirus.
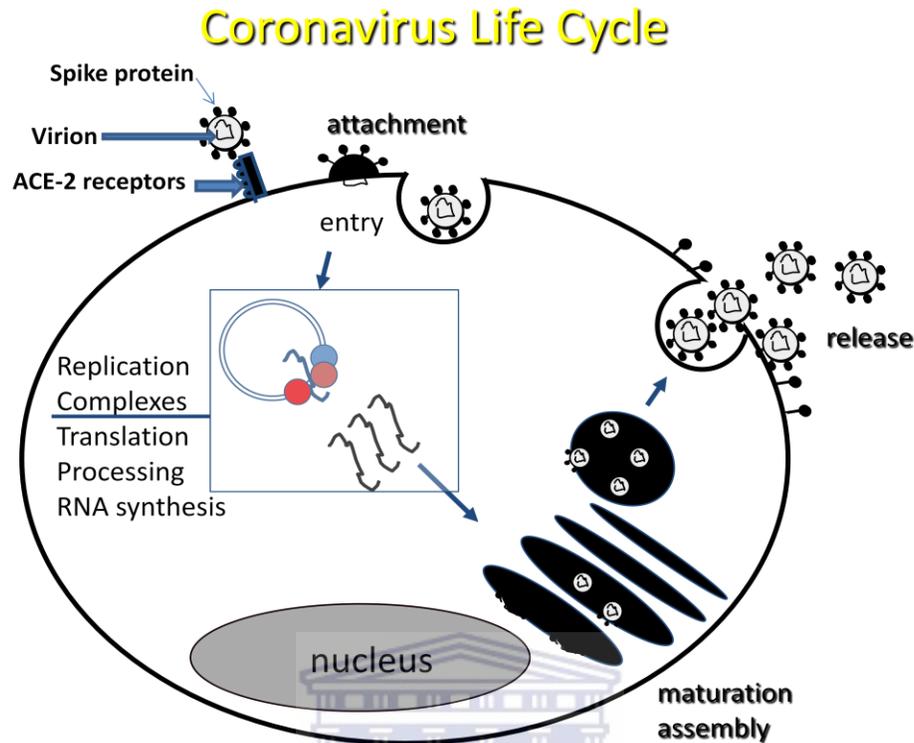
Figure 1.2: Schematic diagram of the replication cycle of coronavirus. Attachment of the virion requires recognition of ACE-2 receptors by spike protein on the cell membrane of a susceptible host. The virus then enters by means of endocytosis with concomitant release of genetic material in the cytoplasm. The viral RNA is then replicated followed by assembly at the budding compartment (adapted from www.iom.edu/~/media/Files/Activity%20Files/.../SARSDenison.ppt, 2005).

Coronavirus attachment onto the host cell receptors is mediated by the S protein, which has the longest nucleotide sequence among other structural proteins. After engagement with the receptors, the virus fuses with host cell membrane by a fusion mechanism. Once the virus enters the cell and fuse with the cellular membrane, the positive stranded coronavirus genome is released and viral

replication is initiated. The viral RNA then takes over the machinery of the host cell and start to replicate. The coronavirus genome has a 5′ methylated cap and a 3′ polyadenylated-A tail to make it look as much like the host RNA as possible (Zhang, Li et al. 2003). This also allows the RNA to attach to ribosomes for translation. During this replication process, copies of the full length genomic RNA and a nested set of subgenomic RNA are generated. These subgenomic mRNA molecules are synthesized by the viral replicase via a discontinuous transcription mechanism, in which a viral leader sequence, derived from the 5′ end of the genome, is added to the 5′ end of each subgenomic mRNA. Adjacent to the leader sequence is a transcription regulatory sequence (TRS) which also precedes each ORF along the genome. The TRS has been shown to act as a high affinity binding site for the N-protein suggesting a role for N in transcription (Zhang, Li et al. 2003). Although replication can take place without N protein it does so at reduced efficiency. These subgenomic mRNA serves as functional templates for the translation of the structural proteins. Once the entire replication process has ensued the viral particles assemble at the budding compartment where envelopment of the nucleocapsid takes place (Zhang, Li et al. 2003).

1.7. AIM OF THE STUDY

This study was designed to clone and express human cyclophilin A (CypA) for interaction with human coronavirus NL63 nucleocapsid proteins. Previous studies have indicated that the SARS-N protein exhibit direct and specific binding with CypA. Comparison of SARS-N and NL63-N protein by means of *in silico* analysis revealed that the two sequences are homologous. Therefore, it is reasonable to infer that HCoV NL63 N will exhibit the same interaction as SARS-N protein.

CHAPTER 2

In silico analysis of human coronavirus NL63

nucleocapsid protein

2.1. ABSTRACT

Cyclophilins (Cyp) are a family of proteins that possess peptidyl-prolyl *cis-trans* isomerase activity and play a key role in the folding of proteins (Lee 2010). About 16 classes of these proteins have been recorded and only 3 have been extensively studied and these include CypA, CypB, and CypC. Cyclophilins were originally discovered for their unique interaction with cyclosporine A, an immunesuppressive drug (Tian, Zhao et al. 2010). CypA has been reported to be associated with several diseases and most commonly with severe types of carcinoma. They have been shown to be up-regulated in hepatocellular carcinoma but their role in pathogenesis remains obscure. Most recently, CypA has been revealed to interact with numerous viral proteins and of outmost importance to this study is the association of CypA with coronaviruses nucleocapsid protein. The coronavirus N proteins are basic proteins that encapsulate viral genomic RNA to form part of the virus structure. SARS-N protein is known to be highly antigenic and associated with several host-cell interactions (Stohlman, Fleming et al. 1983; Hogue and Brian 1986; Zhou, Liu et al. 2008). The SARS-N protein has been recently shown to exhibit high binding affinity to human CypA (Luo, Luo et al. 2004). This study is specifically designed to critically analyze human CypA and its association with human coronavirus (HCoV) NL63 nucleocapsid protein by means of *in silico* examination. *In silico* refers to an analysis that is performed on a computer or via a computer simulation. A variety of computer programs have been used to analyze nucleotide as well as amino acid sequences of CypA, SARS-N, and NL63-N. In the genomic branch of bioinformatics, homology is used to predict the function of a gene. If the sequence of gene A, whose function is known, is homologous to the sequence of gene B, whose function is unknown, then one could infer that gene B may share A′s function. Using bioinformatic tools, the results of this study clearly indicated that there is some degree of homology between the two strains of human coronavirus.

Given this evidence gathered in the results it is suffice to say that NL63-N protein may exhibit the same interaction with CypA.

2.2. INTRODUCTION

Cyclophilins are a group of proteins that have peptidyl-proyl *cis-trans* isomerase activity, collectively known as immunophilins. They are found in both prokaryotes and eukaryotes (Wang and Heitman 2005). Cyclophilins seem to be structurally conserved throughout evolution because they have been found in mammals, plants, insects, fungi, and bacteria. All cyclophilins share a common domain of approximately 109 amino acids, the cyclophilin-like domain (CLD), surrounded by a domain unique to each member of the family that are associated with subcellular compartmentalization and functional specialization (Wang and Heitman 2005). They can be found in most cellular compartments of most tissues and encode unique functions. About 16 classes of cyclophilins have been reported. From the 7 major cyclophilins in human, CypA is the first member of mammalian cyclophilins to be identified (Galat and Bua 2010). CypA plays key roles in several different aspects of cellular physiology including the immune response, transcription, mitochondrial function, cell death, and chemotaxis (Damsker, Bukrinsky et al. 2007). In addition to these key physiological functions of CypA, it has also been reported to interact with viral proteins, especially caspid proteins. Literature indicates that there seems to be a regular pattern of CypA interaction with these viral proteins and several studies have mapped it around the proline (Pro) residue. However, the exact mechanism of action still needs to be determined in order to elucidate strategies for therapeutic intervention. Most recently, CypA has been reported to interact with SARS-coronavirus nucleocapsid protein (Luo, Luo et al. 2004). Therefore, probing the protein-protein interactions of CypA with nucleocapsid protein is of

outmost significance, because the foundation for discovering drugs and/or vaccines relies on these molecular interactions.

Coronaviruses are single stranded positive sense RNA viruses belonging to the *Coronavirideae* family. The latter is known to possess the largest genome of all human RNA viruses with the genome size of approximately 27 to 32 kb and with a guanine and cytosine (GC) contents varying from 32 % to 43 % (Surjit, Liu et al. 2006; Feng and Gao 2007; See, Petric et al. 2008). All coronaviruses have a similar genome organization with the characteristic gene order 5′-replicase ORF1ab, S, E, M, N-3′, although variable numbers of additional ORFs are present in each subgroup of coronavirus. The gene sequences that encode conserved proteins are frequently used for phylogenetic analysis e.g. nucleocapsid gene. The latter is best appreciated and confirmed by constructing phylogenetic trees using different genes in the coronavirus genome. Amongst coronaviruses, the most commonly used genes for phylogenetic studies include chymotrypsin-like protease, Pol, helicase, spike and nucleocapsid, because these genes are present in all coronavirus genome and are of significant length (Wang, Li et al. 2004; Kim, Lee et al. 2006; Woo, Lau et al. 2009). The envelope and membrane genes, although present in all coronavirus genome, are too short for phylogenetic studies.

Due to the complexity of an organism, studying their genome is very challenging and so the bioinformatic tools have made it possible to extract useful information about the genes and make possible predictions about their genome as well as protein structure. In general, genes are complex structures that cause dynamic transformation of one substance into another during the whole life of an individual or even a virus. Even the static information about a particular gene is very difficult to comprehend. When genes are in "action", the dynamics of the processes in which a single gene is involved are thousand times more complex. This is because this gene have to

interact with many other genes, proteins and is influenced by many environmental and developmental factors (van Kampen, van Schaik et al. 2000). Modeling these interactions and extracting meaningful patterns is a key goal of these tools. Therefore, these tools have been designed to broaden our understanding of biological processes.

Genomics, proteomics, and other molecular research technologies and developments in information technology have developed rapidly and have combined to produce a tremendous amount of information within the scientific community. The distinctive part about these tools is that its approach focuses on developing and applying computational intensive techniques in order to expand our understanding of the intricate biological processes (Cai, Han et al. 2005; Hao, Chen et al. 2006; Negi and Braun 2009). Traditionally, viruses were characterized and classified by culture, electron microscopy and serological studies. Using these phenotypic methods, coronaviruses for example were defined as enveloped viruses of 120-160 nm in diameter with a crown-like appearance. The invention of nucleic acid amplification technologies, automated DNA sequencing and bioinformatics tools in the recent decades have revolutionized the characterization and classification of viruses. Bioinformatics apply the principle of information science and technologies to complex life science data (Wang, Ji et al. 2003; Hao, Chen et al. 2006). They have emerged as an important discipline within the biological science that allows scientists to decipher and manage the vast quantities of genomic data that is now available. Previous studies have reported that the ultimate goal of biological researches is to develop knowledge that will allow predictive approaches to all life science. In this regard, the data acquisition, management and analysis technique of bioinformatics are crucial as they provide a lens through which accumulated large-scale data set and knowledge can be viewed.

In this particular study, the *in silico* analysis has been applied in order to predict how human cyclophilin A (CypA) interacts with human coronavirus NL63 nucleocapsid protein. Mapping the interactions of CypA with the viral capsid proteins can provide clues to the initial infection process. Previous studies have showed the involvement of CypA in HIV-1 infection by demonstrating that CypA-deficient viruses could not replicate because they fail to attach to target cells. They demonstrated that CypA is exposed at viral membrane and mediates HIV-1 attachment. Therefore, a comprehensive understanding of the role of CypA in coronavirus can give some indication of viral pathogenesis. To gain insight into the potential role of CypA in human coronavirus NL63 infection, bioinformatics tools have been used to explore potential interactions between NL63 nucleocapsid protein and CypA.

2.3. MATERIALS AND METHODS

2.3.1. Coronavirus DNA sequence alignment

The nucleotide sequences of the genes were extracted from the National Centre for Biotechnology Information (NCBI). Accession numbers for all studied genes are indicated in table 2.3. DNA sequences were displayed in FASTA format and were subsequently used to perform further *in silico* analysis. In order to compare the homologous nucleotide sequences of coronavirus genes, ClustalX version 1.81(Aiyar 2000) was used and the sequences were edited using Gendoc version 2.6.002. These programs provided substantial evidence pertaining to the ancestry of these viruses. Studying evolutionary relationship between all human coronaviruses was achieved by sequence alignment. 100 % homology between these sequences is indicative of conserved regions and these are

described on each as follow: ■ 100 % homology, ■ > 60 %, ▢ < 50 % and white indicates 0 % homology .

2.3.2. Coronavirus amino acid sequence alignment

The primary structures of the proteins were obtained using Expasy (http://web.expasy.org/translate/), which directly transcribed and then translated DNA into amino acid sequence. From the list of amino acid sequences the desirable sequence was chosen by virtue of having the least number of stop codons. The homologous sequences between these amino acid sequences were also compared using ClustalX. Conserved motifs amongst coronavirus nucleocapsid proteins were conducted by aligning sequences. Furthermore, the viral proteins known to interact with CypA were also aligned in order to predict potential motifs for protein-protein interaction.
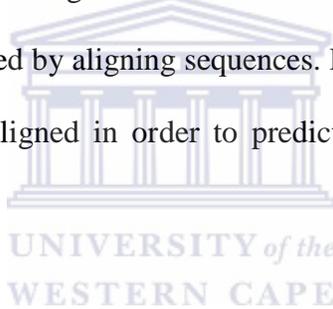
2.3.3. Prediction of nuclear localization signals

The relationship between the different strains of human coronavirus was also studied by comparing their nuclear localization signals, nuclear export signals and their motifs that are involved with interaction of the N protein with host protein and with itself. The nuclear signals were carried out using the PSORT analysis program (http://psort.hgc.jp/form.html) whereas the motif scan was performed using Prosite tools. A motif scan with sufficient information on O-glycosylation and phosphorylation sites was retrieved by means of MOTIF-SCAN (http://hits.isb-sib.ch/cgi-bin/PFSCAN). The program also supplies valuable information about the possible sites that exhibit potential phosphorylation and O-glycosylation sites.

2.3.4. Prediction of cyclophilin A interacting site

All proteins that are known to interact with CypA were retrieved from published literature. These proteins are shown in table 2.1 and their amino acid sequences were obtained from published material. Homologous sequences were compared using ClustalX. From the sequences, the potential motifs were searched manually by comparing homologous sequences. Moreover, the homologous sequences or amino acid residues were compared in terms of their functional groups in order to predict how these proteins interact with CypA.

Table 2.1: An illustration of viral proteins that are known to interact with cyclophilin A.

Interacting protein	Amino acid residues	Basic/ Acidic	Reference
HIV-1 capsid	pro-90	Neutral	(Zander, Sherman et al. 2003)
HIV-1 p24	pro-90, 157, 224	Neutral	(Fitzon, Leschonsky et al. 2000)
HIV-1 Vpr	pro-35	Neutral	(Zander, Sherman et al. 2003)
HIV-1 heparan	Arg-484, 487, 585 and 588	Basic	(Saphire, Bobardt et al. 1999)
HIV-1 p55	unknown	Unknown	(Thali, Bukovsky et al. 1994)
HCV NS5A	unknown	Unknown	(Chatterji, Lim et al. 2010)
HCV NS2	unknown	Unknown	(Wang and Heitman 2005)
HCV NS5B	unknown	Unknown	(Wang and Heitman 2005)
SARS-N	pro-310	Neutral	(Luo, Luo et al. 2004)

2.4. RESULTS

2.4.1. Multiple sequence alignment

The full length genes for all genes were successfully retrieved from NCBI. On the appendices (Appendix 2, page 120-134), the nucleotide and amino acid sequences of these genes are illustrated. Initially, the evolutionary relationship between all human coronaviruses was performed by aligning DNA sequence. In figure 2.1 it is noticeable that these viruses share a common ancestry. The conserved nucleotide sequences are indicated by dark regions and the significance of this relationship is described on each figure. Furthermore, the bovine coronavirus seems to be more homologous with OC43, whereas the rest of the viruses have relatively fair amount of homology.
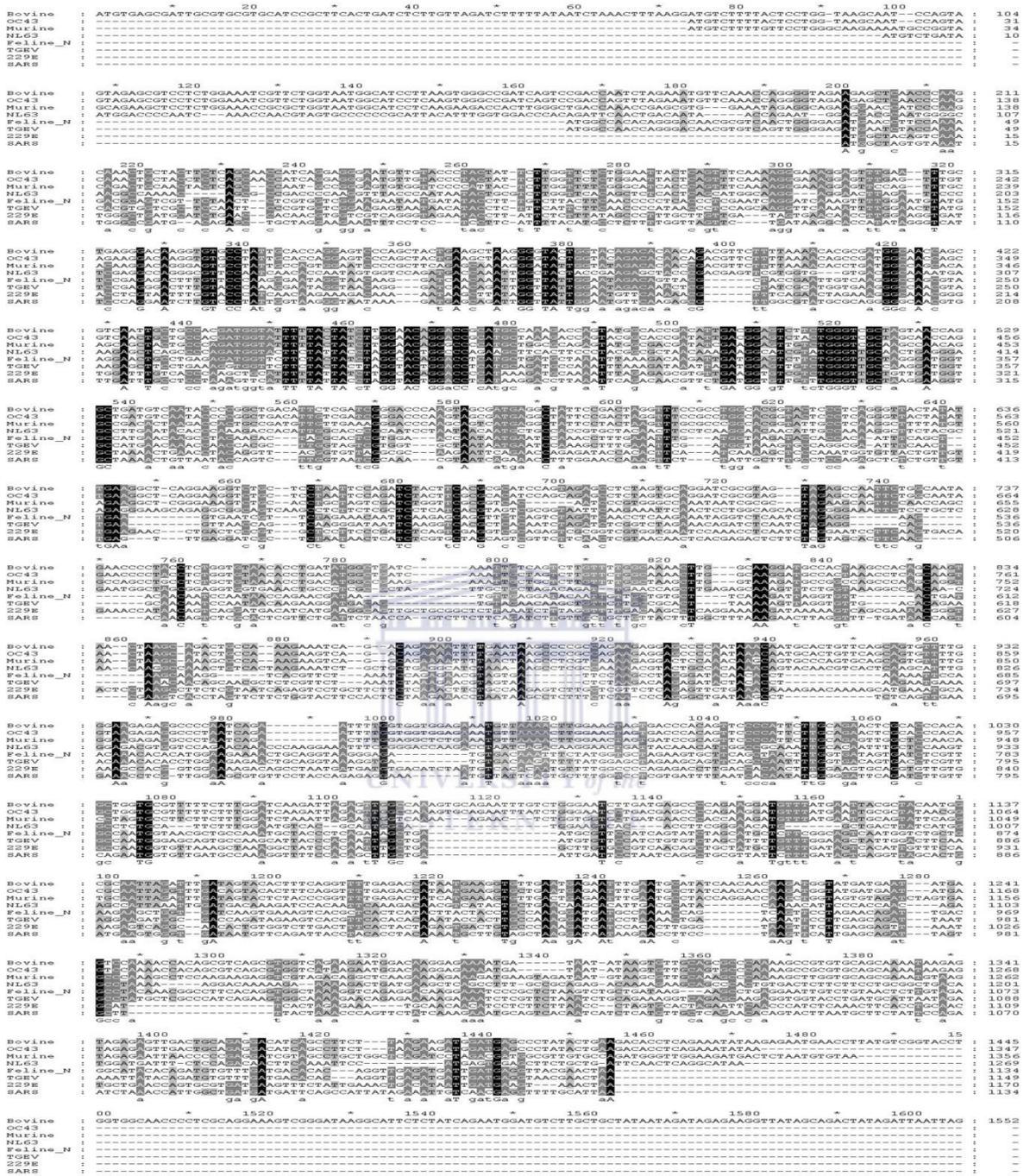
Figure 2.1: Comparison of homologous nucleotide sequences between the different strains of all human coronaviruses. The dark regions indicate some of the conserved regions within the *Coronavirideae* family.

Following the alignment of DNA sequences, this relationship was further compared by aligning amino acid sequences. From figure 2.2, it is clear that different strains expressing N proteins exhibit some degree of variation. It is interesting to note that there are multiple stretches of amino acid residues that share 100 % homology and this is shown in figure 2.2. Moreover, the length of these homologous amino acids is relatively longer in comparison to DNA sequence alignments.
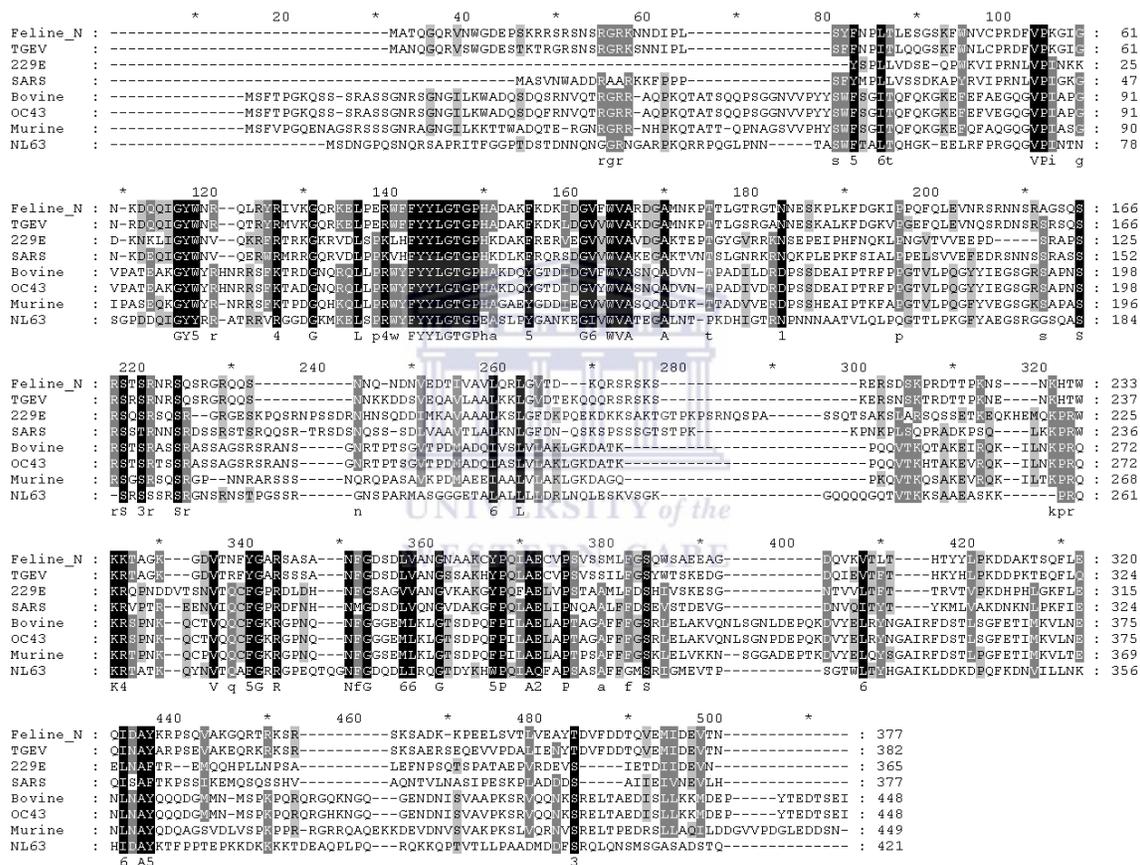
Figure 2.2: Comparison of homologous protein sequences between the different strains of all human coronaviruses. The dark regions indicate some of the conserved regions within the *Coronavirideae* family.

To gain insight into the HCoV-N protein motifs that could potentially interact with CypA, sequence alignment between SARS-N and NL63-N was done. The literature indicates that SARS-N interacts along a specific stretch of amino acid residues (Trp302-Pro310), therefore this region was searched manually and homology determined. In figure 2.3, this region is labeled as "A" and it is clear that there is high degree of homology along these amino acid residues.
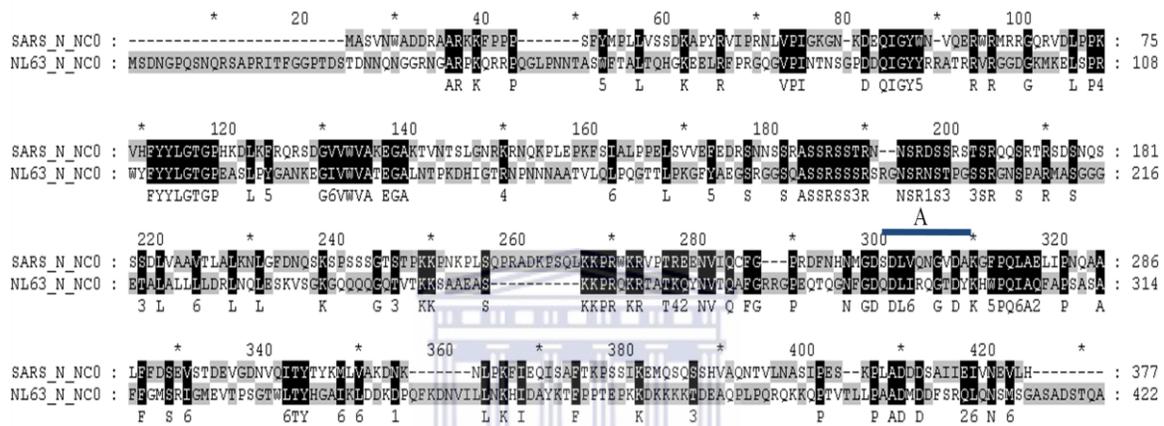
Figure 2.3: Illustration of homologous amino acid sequences between SARS and NL63 nucleocapsid proteins. The dark regions indicate some of the conserved regions within the *Coronavirideae* family. "A" indicates the SARS-N motif that is known to interact with CypA.

Although the two sequences present with conserved residues, there are some notable variations. SARS-N amino acid sequence appears to be longer than NL63-N with amino acid residues of 422 and 377, respectively. In addition, SARS-N protein presents with a larger molecular weight when compared to NL63-N protein. These differences are indicated in table 2.2.

Table 2.2: Comparison of SARS-N protein molecular weight and amino acid sequence to NL63-N protein.

Gene	Molecular weight (kDa)	Amino acid residues
NL63-N	42.26	377
SARS-N	46.04	422

2.4.2. Prediction of nuclear localization signals

The classical nuclear localization signals (NLS) include sequence regions that are rich in basic amino acids and generally conform to one of three types: Pat4, Pat7 and Bipartite motif. "Pat4" NLS consist of a continuous stretch of four amino acids (lysine or arginine) or three basic amino acids associated with histidine or proline. "Pat7", on the other hand, starts with a proline followed within three residues by a segment containing three basic residues. Lastly, bipartite consist of two basic amino acids, a 10 amino acid space, and a five amino acid segment containing at least three basic residues.

It is noticeable that all coronaviruses studied possess the "Pat7" localization signal with the exception of feline coronavirus. It is also appreciable that amongst human coronaviruses, only the recently discovered (SARS and NL63) possess "Pat4" localization signal. In addition, the results reveal that in all human coronaviruses, only SARS has bipartite localization signals. These NLS are crucial in the translocation of nuclear proteins i.e. N proteins that localize in the nucleus, from the cytoplasm, across the nuclear pore complex into the nucleus. From table 2.3 it is clear that amongst all coronaviruses there are distinctions in localization signals.

Table 2.3: Prediction of NLS for N-protein within mammalian coronaviruses using PSORT analysis.

		Group 1 HCoV		
	GenBank	Pat4	Pat7	Bipartite
Feline CoV	NC_007025.1	None	None	326- KRPSEVAKDQRQRRSRS
HCoV 229E	NC_002645.1	None	45-PINKKDK	None
			198-PQEKDKK	
			247-PRWKRQP	
TGEV	AY587884.1	5-RKRK	None	199-KKLGVDTEKQQQRSRSK
		Group 2 HCoV		
	GenBank	Pat4	Pat7	Bipartite
PHEV	NC_007732.1	None	270-PRQKRSP	None
Bovine Coronavirus	EF424619.1	None	269-PRQKRSP	None
Murine hepatits virus	NC_001846.1	None	271-PRQKRTP	None
			392-PQRKGRR	
Human CoV OC43	NC_005147.1	None	270-PRQKRSP	None
SARS-CoV	NC_004718.	257-KKPR	38-PKQRRPQ	373-KKKKTDEAQPLPQRQKK
		373-KKKK	259-PRQKRTA	374-KKKTDEAQPLPQRQKKQ
			369-PKKDKKK	
			384-PQRQKKQ	
HCoV NL63		232-KKPR	234-PRWKRVP	None

2.4.3. Motif scan

Motif scan of the CypA, SARS and NL63 nucleocapsid protein revealed the following. In table 2.5 the NL63-N protein has only one O-glycosylation sites at the threonine residue. Interestingly, the SARS-N protein also exhibits one glycosylation sites at the same amino acid residue. The NL63 and SARS-N protein also exhibit the same kind of properties when analyzing their O-glycosylation sites on the tyrosine and the serine residues. Therefore, SARS-N appears to possess the same O-glycosylation properties as the NL63-N protein. Predictions of the CypA, on the other hand, do not have any O-glycosylation sites. With regards to phosphorylation sites, as indicated in table 2.4, there is also a notable variation in terms of amino acid residues. The two proteins appear to have more phosphorylation sites at the serine residue and then followed by threonine residues. In table 2.4 it is clearly shown that tyrosine residues exhibit the least number of phosphorylation sites. Moreover, the CypA seems to possess the least number of phosphorylation sites.
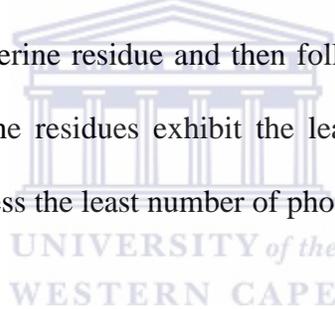
Table 2.4: The prediction of phosphorylation sites using NetPhos.

Protein	Serine	Threonine	Tyrosine
CypA	4	1	0
SARS-N	22	8	3
NL63-N	29	3	2

Table 2.5: The prediction of O-glycosylation sites using NetGlo.

Protein	Serine	Threonine	Tyrosine
CypA	0	0	0
SARS-N	1	1	0
NL63-N	1	1	0

2.4.4. Prediction of interacting sites

The potential sites of interaction for HCoV NL63-N were predicted by comparing sequences of all proteins known to interact with CypA. From the published literature, these sites seem to be centered on the proline residue. However, in SARS-N protein, the mapped site does not involve proline. Whereas the amino acid residues of other proteins seem to be of a consistent functional group, the SARS-N site belongs in multiple functional groups. In figure 2.4, it is noticeable that SARS-N does not share homology with other viral proteins known to interact with CypA.

```
                       *        20         *        40         *        60         *        80         *        100
HIV_1_caps : ------------------------------------------------------------------------------------------------------ :   -
HIV_1_hepa : --MGARASVLIGGKLNPWEKIRLRPGGKKKYRLKLLVWASRELERFALDSGLLETSQGCQQIMEQLQPALRTGTGDLNSLFNTVATLWCVHQRIEVKDTKEALDKI : 104
HIV_1_p24  : ------------------------------------------------------------------------------------------------------ :   -
HCV_NS2    : ------------------------------------------------------------------------------------------------------ :   -
HCV_NS5A   : SGSWLRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFLSCQRGYKGVWRGDGVMHTTCPCGAQITGHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGSCTPSPAPNY : 106
HIV_1_Vpr  : ------------------------------------------------------------------------------------------------------ :   -
SARS_N     : ----------MASVNWADDRAARKKFPPPSFYMPLLVSSDKAPYRVIPRNLVPIGKGNKDEQIGYWNVQERWRMRRGQRVDLPPKVHFYYLG--TGPHKDLKFRQR :  94


                       *        120        *        140        *        160        *        180        *        200        *
HIV_1_caps : -------------------------PIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETIN------ :  74
HIV_1_hepa : EEIPNKKKPQQAAADTGSNSKVSQNYPIVQNAQGQMAYQSISPRTLNAWVKAIEEKAFSPEVIPMFTALSEGATPQDLNMMLNIVGGHQAAMQMLKDTFN------ : 204
HIV_1_p24  : -----------------------PIVQNAQGQMIHQSLSPRTLNAWVKVIEEKAFSPEVIPMFSALSEGATPFDLNMMLNIVGGHQAAMLMLKDTIN------ :  74
HCV_NS2    : -----------------------MDREMAASCGGVVFVGLVFLTMSPHYKVFLARLIW--WLQYFLTIAE-AHLQVWIPPLNIRGGRDAIILLTCVIHP------ :  73
HCV_NS5A   : SRALWRVAAEEYVEVTRVGDFHYVTGMTTDNIKCPCQVPAPEFFTELDGVRLHRYAPACKPLLRDEVTFQVGLNQYVVGSQLPCEPEPDVTVLTSMLTDPSHITAE : 212
HIV_1_Vpr  : -----------------------------MEQAPEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQHIYETYGDTWAGVEAITRILQQLLF------ :  69
SARS_N     : SDGVVWVAKEGAKTVNTSLGNRKRNQKPLEPKFSIALPPELSVEFEDRSNNSSRASSRSSTRNNSRDSSRSTSRQQSRTRSNSNQSSSDLVAAVTLALKN----- : 195
                                                                                                         6


                       220        *        240        *        260        *        280        *        300        *        3
HIV_1_caps : EEAAEWDRVHPVHACPIAPGQMREPRGSDIAGTTS------TLQEQIGWMTNNPPIPVGEIYKR-WIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLR : 173
HIV_1_hepa : EEAAEWDRVLPVQA-CPHPAGQMREPRGSDIAGTTS------TLQEQIGWMTSNPPIPVGEIYKR-WIVLGLNKIVRMYSPVSILDIRQGPKEPFRDYVDRFFKCLR : 303
HIV_1_p24  : EEAAEWDRLHPVHA-PVAPGQMREPRGSD---------------------------------------------------------------------- : 103
HCV_NS2    : ELIFDITKLLLATLCPLLVLQAGIARVPYFVRAHG------LIRACVLLR-----KVAGGHYVQ-MALMKLAALTGTYLYDHLTPLQYWAHAGLRDLAVAVEPVIF : 167
HCV_NS5A   : TAKRRLARGSPPSLASSSASQLSAPSLKATCTTHHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFDPLRAEEDEREVSVPAEILRKTRKFPSALPIWAR : 318
HIV_1_Vpr  : -IHFRIG-CRHSRIEVTRQRRARNGASRS----------------------------------------------------------------------- :  96
SARS_N     : -LGFDNQSKSPSSSCTSTPKKPNKPLSQPRADKPS------QLKKPRWKRVPTREENVIQCFGPRDFNHNMGDSDLVQNGVDAKGFPQLAELIPNQAALFFDSEVS : 294
                                      g


                       20         *        340        *        360        *        380        *        400        *        420
HIV_1_caps : AEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVL------------------------------------------------ : 231
HIV_1_hepa : AEQATQEVKNWMTETLLVQNANPDCKNILKALGPAASLEEMMTACQGVGGPGHSKARVLAEAMSQVHSPNIMMQKGNFRGQKRIKCFNCGKEGHLARNCRAPRKKGC : 409
HIV_1_p24  : ------------------------------------------------------------------------------------------------------ :   -
HCV_NS2    : SDMEIKIIT--------WGADTAACGDIIQGLPVSARRGREILLGPADSLEGQGWRLL------------------------------------------------ : 217
HCV_NS5A   : PDYNPPLLESWRDPDYVPPVVHGCPLPPTKAPPIPPPRRKRTVILTESTVSSALAELATKTFGSSGSSAVDSGTATAPPDGPSDDGDAGSDAESYSSMPPLEGEPG : 424
HIV_1_Vpr  : ------------------------------------------------------------------------------------------------------ :   -
SARS_N     : TDEVGDNVQITYTYKMLVAKDNKNLPKFIEQISAFTKPSSIKEMQSQSSHVAQNTVLNASIPESKPLADDDSAIIEIVNEVLH---------------------- : 377


                       *        440        *        460        *        480        *        500
HIV_1_caps : ------------------------------------------------------------------------------- :   -
HIV_1_hepa : WKCGKEGHQMKDCTERQANFLGRIWPSSKGRPGNFPQSRLEPTAPPAESPGFGEEIAPSPKQETKDKELYPLASLKSLFGNDP : 492
HIV_1_p24  : ------------------------------------------------------------------------------- :   -
HCV_NS2    : ------------------------------------------------------------------------------- :   -
HCV_NS5A   : DPDLSDGSWSTVSEEASEDVVCC-------------------------------------------------------- : 447
HIV_1_Vpr  : ------------------------------------------------------------------------------- :   -
SARS_N     : ------------------------------------------------------------------------------- :   -
```

Figure 2.4: Illustration of homologous amino acid sequences between proteins that are known to interact with CypA.

2.5. DISCUSSION

2.5.1. *Homologous DNA sequences*

The primary purpose of analyzing nucleotide sequences is to increase our understanding of the biology of newly emerging microorganisms. This study was specifically designed to compare SARS-N and NL63-N protein. Research has showed that SARS-N protein interacts with CypA. The results of this study clearly indicate that there is some degree of homology between the two

coronavirus sequences. Therefore, it is suffice to say that NL63-N protein may exhibit the same characteristics as the SARS-N protein. This is attributed to the fact that comparing sequences with known function i.e. the SARS-N protein, with the new sequence i.e. NL63-N protein, enables us to make reasonable inferences about the biology of the new sequence. Thus, sequence analysis helps the scientific community to compare sequences to find both similarities and differences between compared sequences.

2.5.2. DNA guanine-cytosine (GC) content:

The GC content represents the percentage of guanine and cytosine nitrogenous bases on a DNA molecule (Woo, Lau et al. 2005; Mills 2006). This parameter is looked at in order to understand the stability of the DNA molecule because it is believed that a high GC content is associated with more stable DNA whereas a low GC content is associated with less stable DNA (Woo, Lau et al. 2005). However, this popular belief remains controversial. Nevertheless, the difference in GC content between the two species is relatively insignificant. Therefore, for the purpose of this study the difference in GC content is negligible.

The study results show that the SARS-N gene has more GC than NL63. This could be attributed to the fact that the length of the coding sequence is directly proportional to the GC content (Woo, Lau et al. 2005). Also the bias of the stop codon towards the adenine (A) and thymidine (T) nucleotides may contribute to the high GC content (Woo, Lau et al. 2005).

2.5.3. Amino acid sequence

The structural modifications that occur in proteins either during their synthesis or after translation are imperative in understanding the intricate biological processes that occur within the cell. These modifications may either activate proteins or render them inactive. This study proves that there are some post-translational modifications that occur in our proteins of interest. They may undergo glycosylation, phosphorylation or myristilation and other post-translational modification that are implicated in protein structure and function. For the purpose of this study, focus is drawn onto the above mentioned processes.

The process of glycosylation occurs in the cytoplasm and nucleus (Ying, Hao et al. 2004; Han, Lohani et al. 2007). However, O-glycosylation that is implicated in these proteins occurs in the Golgi apparatus and is catalyzed by the enzyme UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosamine transferase (Ko, Kang et al. 2006). The SARS-N and NL63-N protein seem to be glycosylated at the serine and threonine residues. This process occurs by attachment of O-linked glycan to the hydroxyl oxygen of serine and threonine side chains (Shih, Chen et al. 2006). These glycans appear to serve various functions viz. protein folding and stability. Literature indicates that linking of these glycans to target proteins confer some stability to the protein.

Phosphorylation, on the other hand, occurs on serine, threonine and tyrosine residues in eukaryotic cells. Phosphorylation replaces neutral hydroxyl groups on serines, threonines, or tyrosines with negatively charged phosphates (Wang, Ji et al. 2003; Mizutani, Fukushi et al. 2004; Zeng, Ruan et al. 2004). Research shows that serine is the most commonly phosphorylated amino acid, followed by threonine and then tyrosine. This is in agreement with the results of this study. The phosphorylation process is crucial, because, of its ability to effectively turn a hydrophobic

portion of a protein into a polar and extremely hydrophilic portion. This is of particular interest to scientists, because this process may confer antigenic properties to the target protein i.e. SARS-N or/and NL63-N protein.

2.5.4. Analysis of protein interaction using bioinformatics

Protein interactions using bioinformatic tools provide valuable information about the potential interacting site. The probable binding site for HCoV NL63-N protein was determined by comparing the SARS-N interacting domain with NL63-N. However, this site is remarkably different when compared to other viral proteins known to interact with CypA. Therefore, it is suffice to say that the coronavirus N proteins have a unique binding mode to this protein, CypA. Alternatively, CypA might have multiple stereochemically active sites that enable it to interact with these proteins.
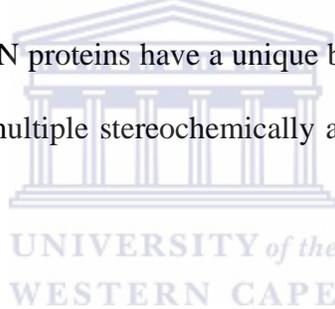
CHAPTER 3

Cloning and expression of human Cyclophilin A

3.1. ABSTRACT

The *Coronaviridae* family is composed of a number of ribonucleic acid (RNA)-containing viruses currently classified into two genera, the coronavirus and torovirus. The family is classified together with the *Arteviridae* in the order *Nidovirales*. Coronaviruses are enveloped single stranded positive sense RNA viruses about 80-160 nm in diameter. The coronavirus is a positive sense RNA virus and the naked RNA is infectious. The 5′-two thirds of the genome encodes for a polyprotein that contains all the enzymes necessary for replication, whereas the 3′-one third encodes for all the structural proteins that mediate viral entry into the host cell. The structural proteins include spike (S), envelope (E), membrane (M) and nucleocapsid (N) proteins.

N is one of the most crucial structural components of coronaviruses; hence major attention has been focused on characterization of this protein. Some laboratories have demonstrated that this protein interferes with different cellular pathways, thus implying it to be a key regulatory component of the virus (Zakhartchouk, Viswanathan et al. 2005). Furthermore, it has been shown that severe acute respiratory syndrome (SARS)-N protein interacts with cellular proteins, including cyclophilin A (CypA), heterogenous nuclear ribonucleoprotein (hnRNP) A1, human ubiquitin-conjugating enzyme, cyclin dependent kinase (CDK)-cyclin complex protein, Ikappaßalpha (IkBα), cytochrome (Cyt) P450 etc. For the purpose of this study, the focus is based on CypA interaction with the human coronavirus (HCoV) NL63 N protein. These interactions might play a role in the pathology of HCoV-NL63. Since the N protein is involved in viral RNA packaging to form a helical core, it is suffice to say that both NL63-N and CypA are possibly within the HCoV-NL63 replication/transcription complex and NL63-N/human CypA interaction might function in the regulation of HCoV-NL63 RNA synthesis. In this chapter, total RNA was extracted from a human

cell line (MCF7) and used to generate first strand cDNA. This was used to amplify human CypA with specific primers for subsequent cloning into a vector and expression of the protein in a bacterial system. Human CypA was expressed, purified and quantified for future studies.

3.2. INTRODUCTION

Cyclophilins are a group of proteins that have peptidyl-proyl *cis-trans* isomerase activity, collectively known as immunophilins. Human CypA has previously been shown to exhibit high binding affinity to the SARS-coronavirus N protein. This interaction occurs via Trp 302-Pro 310 loop (Luo, Luo et al. 2004). However, the significance of this interaction has not been established. CypA is a peptidyl-prolyl *cis/trans* isomerase (PPIase) that is involved in multiple signaling events of eukaryotic cells. CypA catalyses the *cis/trans* isomerisation of proline imide bonds in peptides. They were initially identified as immune modulators through their binding to cyclosporin A (CsA), an immunosuppressant (Daum, Schumann et al. 2009; Chatterji, Lim et al. 2010; Fischer, Gallay et al. 2010; de Wilde, Zevenhoven-Dobbe et al. 2011). The human genome encodes up to 16 different cyclophilins. Despite their differences in tissue distribution, sub-cellular localization and relative abundance, they all share a similar 3-D structure and show *in vitro* PPIAse activity (Zheng, Koblinski et al. 2008; Daum, Schumann et al. 2009; Han, Yoon et al. 2010). *In vivo*, they are thought to have numerous cellular functions, including protein folding through *cis/trans* isomerisation, immune response, lipid and protein trafficking or transcription. Most amino acids have a strong energetic preference for the *trans* peptide conformation, therefore the *cis/trans* activity of CypA is crucial for the stabilization of the *cis* conformation.

Studies have shown that CypA is a potential target for antiretroviral therapy. This is based on the fact that inhibition of CypA was shown to suppress human immune virus (HIV)-1 replication but the mechanism remains unclear. CypA interacts with HIV-1 viral protein R (Vpr) both *in vitro* and *in vivo* (Zander, Sherman et al. 2003; Hatziioannou, Perez-Caballero et al. 2005; Javanbakht, Diaz-Griffero et al. 2007). Moreover, these interactions occur on the N-terminal regions of CypA containing Pro-35, and surface plasmon resonance studies revealed that a seven residue motif centered at Pro-35 consisting of RHFPRIW is sufficient for maintaining strong specific binding (Zander, Sherman et al. 2003; Javanbakht, Diaz-Griffero et al. 2007). Previous studies have also concluded that the ability of CypA to interact with N-terminal Vpr peptides is determined by Pro-35. There is convincing evidence that CypA only binds specifically to peptides containing the seven residue motif RHFPRIW centered at Pro-35 of the N-terminal Vpr (Solbak, Reksten et al. 2010). This region includes the loop region connecting helices 1 and 2 of Vpr, in addition to the Arg-32 and His-33 that terminate the well-defined helix 1 and 2 (Solbak, Reksten et al. 2010). The study revealed that the structural data for N-terminal mutant of Vpr in which Pro-35 is exchanged for Asn indicates a merge of helices 1 and 2 meaning interaction with CypA disrupts the hydrophobic interaction of the two helices. FGPDLPAGP has been shown to be an active site inhibitor for CypA, and nuclear magnetic resonance studies have showed that there is some overlap between the binding epitope and CsA drugs. In human proteins, the linear sequence recognition code was found to be consensus sequence FGPXlp. However, the peptide FGPDLPAG shows inhibition of the CypA isomerase reaction highlighting CypA interaction epitope (Fischer, Gallay et al. 2010).

In addition to Vpr interaction, CypA also interacts with HIV-1 capsid protein. CypA is a highly conserved peptidyl prolyl *cis/trans* isomerase (PPIA) that is incorporated into HIV-1 virions and play a role in the early stages of viral replication (Khan, Garcia-Barrio et al. 2003; Javanbakht,

Diaz-Griffero et al. 2007; Mascarenhas and Musier-Forsyth 2009). The HIV-1 capsid interacting region with CypA encompasses nine residues 85-93 (PVHAGPIAP) belonging to a loop region connecting helices 4 and 5, respectively, of HIV-1 capsid (Li, Tang et al. 2007). Gly-90 and Pro-90 on HIV-1 capsid protein were shown to be the most crucial amino acid targets of CypA binding. Hatziiaoanou et al. (2005) reported that CypA acts as a chaperone with Pro-90 of N-terminal HIV-1 capsid peptide (residues 1-151). The binding of CypA with HIV-1 capsid protein positively affects some early stages of the viral life cycle (Li, Tang et al. 2007). This is because prevention of viral binding either by drugs that occupy that active site of CypA, by mutation in HIV-1 capsid protein, or RNAi that knocks down intracellular CypA levels diminishes viral infectivity. Therefore it′s sufficed to say that viral infectivity is improved by high levels of CypA. CsA is known to exhibit the same binding activity in a loop between the 4th and 5th helices of the capsid protein, and disrupts the CypA-capsid interaction, which leads to attenuation of HIV-1 (Keckesova, Ylinen et al. 2006; Noser, Towers et al. 2006). Literature indicates that there is also a strong correlation between the binding of HIV-1 capsid protein and CypA and the infectivity of certain strains of HIV. In addition to HIV-1, CypA also interacts with other closely related lentivirus such as feline immunodeficiency virus (FIV) and samian immunodeficiency virus (SIV) proteins. The interaction of CypA is more widespread amongst lentivirus (Qi, Yang et al. 2008; Takeuchi 2010).

CsA has been reported to exert anti hepatitis C virus (HCV) activity *in vitro* and *in vivo* and is a crucial factor for HCV replication. Some studies have revealed that even CypB and CypC are indispensible for HCV replication and this interaction is through domain II of nonstructural protein 5A (NS5A) (Hatziioannou, Perez-Caballero et al. 2005; Chatterji, Lim et al. 2010). NS5A is crucial for HCV replication as a result it constitutes a potential target for antiviral drug development. DEB025, a CypA inhibitor, interferes with the interaction between CypA and NS5A, thereby

inhibiting HCV replication. DEB025 sequesters CypA and therefore prevent conformational changes in the replication complex that are crucial for viral replication (Chatterji, Lim et al. 2010). This unique feature makes it by far a more efficient drug than "classical" direct acting antivirals (DAA) inhibitors. Fischer et al. (2010) revealed that interaction of NS5A with CypA is completely abolished in the presence of CsA. Studies have shown that a combination of CsA and DEB025 inhibits the CypA-NS5A interaction in a dose dependent manner. Furthermore, DEB025 was shown to be more potent than CsA meaning that the affinity of DEB025 to CypA is more superior to that of CsA.

3.3. MATERIALS AND METHODS

3.3.1. Reagents
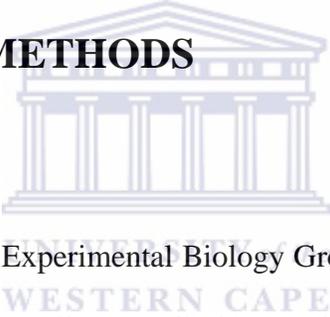
MCF7 cells were obtained from the Experimental Biology Group (University of the Western Cape, Department of Medical Biosciences). HCoV-NL63 RNA (a kind gift from Prof. L van der Hoek, Holland) was extracted from a fifth-passage virus (strain Amsterdam 1) obtained from a clinical sample. SDS gel preparation kit, DMSO, BAP-FLAG and Anti-FLAG M2 antibodies were purchased from Sigma Aldrich. Goat polyclonal Anti-Mouse IgG (H&L) was purchased from Koma Biotech. All other chemicals were purchased from Promega.

3.3.2. Tissue culture

MCF7 cells were grown and maintained as monolayers in 75 cm^2 tissue culture flasks. These cells were cultured in Dulbecco′s modified eagle medium (DMEM) supplemented with 10 % fetal bovine serum (FBS) and 1 % filtered antibiotic mix containing Penicillin (100 U), Streptomycin (0.1

mg/ml), and Amphotericin B (10 U). The flasks were then incubated at 37 ^{o}C in a humidified incubator with a constant supply of 5 % CO_2. The cell monolayers were passaged each time they reached approximately 90 % confluency as confirmed by light microscopy. Once confluent, the MCF7 cells were first washed with sterile 1 x phosphate buffered saline (PBS), followed by treatment with Trypsin-Versene (EDTA) (Gibco) for approximately 2 min with gentle shaking. Once the cells detached from the surface, they were diluted with fresh medium and then seeded into fresh 75 cm^2 flasks. The cells were subcultured for a minimum of three passages and then harvested if there was no contamination after three successive subcultures. The latter involved treating confluent cells with Trypsin-Versene (EDTA) for 2 min followed by addition of fresh medium, vigorous pippeting and then incubation at room temperature for 5 min. Detached cells were then transferred to a 15 ml Greiner tube and centrifuged for 5 min at 3 000 rpm at 4°C. The supernatant was discarded and the pellet kept for RNA extraction.
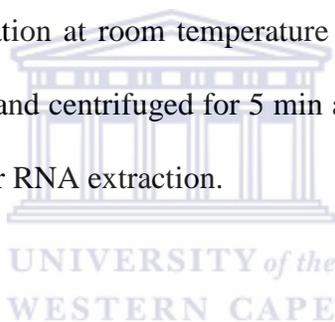
3.3.3. Bacterial strains and plasmids

All bacterial strains used in this study were purchased from Promega and they include KRX, BL21 and JM109 competent cells. KRX is an *E. coli* K12 derivative that has most important features associated with cloning and screening strains. BL21, on the other hand, is a λ (DE3) pLysS strain that can be used with protein expression vectors that are under the control of the T7 promoter. For initial transformation of the recombinant vector, JM109 competent cells were used, whereas KRX and BL21 competent cells were used for optimization and final expression of recombinant protein. The plasmid vectors used were pGEM-T EastVector system II and pFN2A (GST) Flexi vector. 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 vector was utilized because of its convenience for cloning of PCR products.

3.3.4. Generation of recombinant plasmid vectors

3.3.4.1. Design of PCR primers

The human CypA nucleotide sequence was retrieved from NCBI. Using the obtained nucleotide sequence as a template, the forward and reverse primers were designed using Promega primer design tools. For ligation into the Flexi™ protein expression vector a *SgfI* (5´-GCGATCGC-3´) restriction-site was incorporated into the forward primer and a *PmeI* (5´-GTTTAAAC-3´) restriction site was incorporated into the reverse primer. An additional C residue was appended upstream of the ATG in the forward sequence to prevent a frameshift in the reading frame. Porphobilinogene deaminase (PBDG) was used as PCR internal control and primers were donated by the Breast Cancer Research Group (University of the Western Cape, Department of Medical Biosciences). These primers were used as 100 ρ.mole/µl stock solutions and their nucleotide sequences are shown in table 3.1.

Table 3.1: Primers used for the amplification of human CypA and PBDG.

Gene	Primer	Primer sequence
Human CypA	Forward	AGGA<u>GCGATC</u>ATGGCAGTGGCAGAAGATTTTAA
	Reverse	TTGTG*TTTAA*CGTCATCGATACCGTCACCGTCTTCTAAAATTCAAATTTG
PBDG	Forward	TCCTGCACCCTGAGGAAT
	Reverse	CCTCAGGAAGGCCCTTT

Sgf1 restriction site- underlined

Pme1 restriction site- italics

3.3.4.2. Amplification of full length human cyclophilin A gene

Total cellular RNA was extracted from MCF7 cells using the Trizol method. Briefly, the cells were lysed by addition of Trizol reagent to the cell pellet, which was then homogenized by pippeting several times. The tube was then incubated at room temperature for 5 min with occasional pippeting. The sample was then treated with chloroform, mixed gently by inverting for 5 min, and then centrifuged at 3 000 rpm for 10 min. Following spin down, the supernatant was aspirated to a new tube and then washed with isopropanol. This was then incubated at for 10 min at room temperature, followed by centrifugation at 3 000 rpm for 10 min. The supernatant was discarded, while the pellet was washed three times with 70 % ethanol. The latter was air dried and the pellet re-dissolved in nuclease free water. Reverse transcription was done in a mixture of 1 µl RNA template (108-109 copies/ml), 1 µl of Oligo (dT)$_{15}$ primer (100 µM stock), $1\times$ incubation buffer, 2 µl of dNTP mix (10 mM stock), 20 U RNasin$^®$ Ribonuclease inhibitor, 15 U of AMV Reverse Transcriptase and 4 µl MgCl$_2$ (25 mM) in a total volume of 20 µl in nuclease-free water, according to the manufacturer's specifications (Promega). The reaction was

heated at 42 ºC for 60 min, 95 ºC for 5 min and then cooled to 0 ºC for 5 min to deactivate the enzyme. Next the gene was amplified by means of polymerase chain reaction (PCR). CypA was amplified from the cDNA library of MCF7 cells using gene specific primers. In table 3.2 the PCR conditions were designed in such a way that the annealing temperature for the reaction is 5 ºC lower than the melting temperature of the forward primer and the reverse primer. These amplification reaction mixtures are shown in table 3.2.

Table 3.2: Polymerase chain reaction (PCR).

Reagent	Negative control (µl)	Internal control (µl)	Human CypA (µl)
cDNA	0	2	2
dNTP	0.5	0.5	0.5
MgCl$_2$	1.5	1.5	1.5
5X Mg free buffer	10	10	10
Forward primer	1	1	1
Reverse primer	1	1	1
Go Taq	0.5	0.5	0.5

Nuclease free water to final volume of **50**

PCR amplification reactions were run under the following conditions: 95 °C for 3 min followed by 30 cycles of 95 °C for 1 min, 50 °C for 1 min and 72 °C for 1 min. The final elongation was run for 15 min at 72 °C. The PCR product was diluted in a ratio of 1:6 with 6 x Blue/Orange loading dye (0.4 % (v/v) orange G, 0.03 % (v/v) bromophenol blue, 0.03 % (v/v) xylene cyanol

FF, 15 % (v/v) Ficoll® 400, 10 mM Tris-HCl (pH 7.5) and 50 mM EDTA (pH 8.0)) The amplification reactions were run on a 1 % (w/v) agarose gel, containing 0.001 % (v/v) ethidium bromide, by electrophoresis in TBE buffer (89 mM tris (hydroxymethyl) aminomethane, 0.089 mM boric acid, 2 mM EDTA (pH 8)). The gel was run at 90 volts, 2 amperes for a period of approximately 1 hour. The size of the DNA fragment of interest was determined according to the DNA migration pattern in an agarose gel compared to that of a pre-determined DNA molecular weight marker. When the PCR fragment was visualized under the high frequency ultraviolet (UV) light, exposure to UV light was minimized to avoid damage to the DNA. The PCR product to be ligated was gel purified by means of the Wizard® PCR Preps DNA purification systems (Promega) according to the manufacturer′s specifications. Briefly, the bands to be cloned were excised from the agarose gel by means of a clean scalpel and razor blade. A weighed gel slice was completely dissolved by incubation at 50 $^{\circ}$C for 10 min with occasional vortexing in the presence of a membrane binding solution (4.5 M guanidine isothiocyanate and 0.5 M potassium acetate) at a ratio of 1:1. The gel mixture was then transferred into an SV minicolumn, where it was washed with membrane wash solution (10 mM potassium acetate, 80 % (v/v) ethanol and 16.7 μM EDTA). The DNA was finally eluted with nuclease-free water into a sterile micro-centrifuge tube.

3.3.4.3. Cloning and verification of amplification products

3.3.4.3.1. Cloning of amplification product

The product that was purified using Wizard® PCR Preps DNA purification systems was then prepared for ligation into an expression vector. Following PCR product preparation, the amplicon was cloned into pGEM-T easy vector according to the standard specifications of Promega manual

followed by transformation into JM109 competent cells. Table 3.3 represents the ligation reaction that was followed according to Promega standards.

Table 3.3: Ligation into pGEM vector.

Reagent	Positive control (µl)	Background control (µl)	pGEM reaction (µl)
2X rapid ligation buffer	5	5	5
pGEM-T vector (50ng)	1	1	1
PCR product	0	0	2
Control insert DNA	2	0	0
T4 DNA ligase	1	1	1

Following the ligation reaction that is represented in table 3.3, the ligation product was transformed into JM109 competent cells. This involved placing competent cells that were frozen at -80 $^{\circ}$C on ice for 5 min until thawed. The DNA to be transformed was pippeted to the bottom of the tube with 50 µl competent cells and incubated on ice for 10 min. This was followed by a heat shock step whereby the sample was incubated in a 42 $^{\circ}$C water bath for 40 sec and then immediately returned to ice for 2 min. The cells were then transferred to 1 ml of pre-warmed Luria Bertani (LB) broth (10 g pancreatic digest of casein, 5 g yeast extract powder and 5 g NaCl per 1litre water), containing no antibiotics. The mixture was incubated at 37 $^{\circ}$C for 1 hour with shaking at 300 rpm. The cells were then plated on LB-agar selective plates supplemented with ampicillin (100 mg/ml), (Isopropyl-ß-D-thio-galactoside) IPTG (100 mM), and X-gal (3 %). These plates were incubated overnight (12-16 hours) at 37 $^{\circ}$C.

The efficiency of transformation was measured using the following standards: Typically, approximately 100 colonies should be observed (according to Promega's standards) - 10-30 % of which are blue when competent cells that have a transformation efficiency of 1×10^8 cfu/µg DNA are used. Greater than 60 % of the colonies should be white, which is indicative of efficient transformation. If less than 50 % of the white colonies are seen in the positive control reaction, this was regarded as indicative of suboptimal conditions. Cells that were successfully transformed were selected using Blue/White colony selection, where transformants were subjected to colony PCR to verify the presence of the construct before it can be isolated. Briefly, colony PCR involves selecting multiple colonies from a spread plate and mixing with PCR reagents as described in table 3.2. These colonies are inoculated into 15 ml LB broth supplemented with ampicillin and grown overnight. Once the PCR products have been confirmed in agarose gel electrophoresis, the cultures that correspond to the correct genes are selected for plasmid DNA isolation using Wizard® plus SV minipreps DNA purification system.
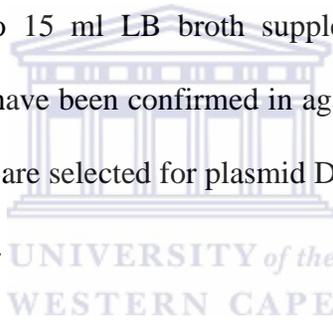
3.3.4.3.2. Plasmid DNA extraction

Plasmid DNA was isolated from the transformed cells using Wizard® plus SV minipreps DNA purification system. Briefly, a single bacterial colony that was confirmed to contain the amplicon by colony PCR was inoculated into 10 ml of LB medium containing antibiotics and incubated overnight at 37 °C. The culture was pelleted for 5 min and then immediately resuspended in 250 µl of cell re-suspension solution (50 mM Tris-HCl, 10 mM EDTA). This was mixed with another 250 µl of cell lysis solution (1 % (w/v) SDS, 0.2 M NaOH) and then mixed by inverting. To this mixture, 10 µl of alkaline protease solution was added, mixed and then incubated for 5 min at room temperature. Following incubation, the sample was mixed with 350 µl of neutralizing solution (80

mM potassium acetate, 8.3 mM Tris-HCl, 40 µM EDTA), inverted and then centrifuged at top speed (13 000 rpm) at room temperature for 10 min. Once the cleared lysate was obtained, the lysate was decanted into a spin column and then centrifuged for 1 min at top speed. Flow through was discarded and the column reinserted into the collection tube. 750 µl of wash solution was added and the sample was centrifuged at top speed for 1 min. The spin column was then washed with 250 µl of wash solution and centrifuged for 2 min. The plasmid DNA was eluted by transferring the spin column into a sterile centrifuge tube followed by addition of 30 µl of nuclease free water to the spin column. The latter was centrifuged at top speed for 1 min and the DNA stored at -20 ^{o}C.

3.3.4.3.3. Restriction endonuclease digestion

To confirm the presence of the exact protein coding region, the purified plasmid was digested with restriction enzyme, *EcoR1*, as shown in table 3.4 to verify the presence of insert.
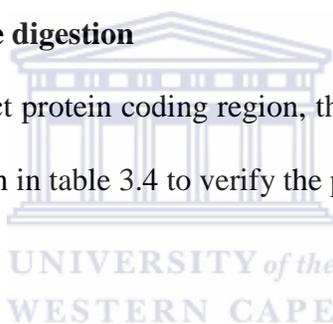
Table 3.4: Recombinant pGEM-CypA construct restriction enzyme digests.

Reagent	pGEM enzyme digest reaction (µl)
DNA	5
Buffer H	2
EcoR1	0.3
BSA	0.4
Nuclease free water	12.6

3.3.4.3.4. Nucleotide sequencing and sequence analysis

Since PCR was used to create the original insert, the sequence of the protein coding region in the initial clone was verified by automated DNA sequencing. Cycle sequencing was performed at Inqaba Biotech. The clone corresponding to the initial nucleotide sequence (NCBI) was selected for further experimentation. Following *in silico* analysis, the construct was digested with pFlexi restriction enzymes (*Sgf1* and *Pme1*) for ligation into pFlexi vector that has been digested with the same restriction enzymes. This reaction is shown in table 3.5. Subsequent to gel purification of the digested product, a ligation reaction was performed whereby the protein coding region was inserted into a bacterial expression vector, pFlexi. The manufacture′s standard conditions for successful ligation of the DNA fragment are represented in table 3.6.

**Table 3.5:** pFlexi restriction enzyme digests.

Reagent	pFlexi enzyme digest reaction (µl)
5X Flexi digest buffer	2
DNA constructs	10
Sgf 1and *PmeI* enzyme blend	2
Nuclease free water	4

Table 3.6: Ligation into pFlexi vector.

Reagent	Background control (µl)	pFlexi reaction (µl)
10X Flexi ligation buffer	2	2
Flexi vector (cut with *Sgf* and *Pme1*)	5	5
T4 ligase	1	1
Cut DNA product	0	12

Successful ligation of the DNA fragment was confirmed by digesting an aliquot of the ligation product and then separated electrophoretically. Once the presence of the amplicon was confirmed in agarose gel, the ligation product was transformed into BL21 and KRX competent cells for protein expression. The conditions for transformation are similar to the ones described in 3.3.4.3.1 above. Upon confirmation of recombinant plasmid by colony PCR, a small-scale protein expression was performed in KRX cells.
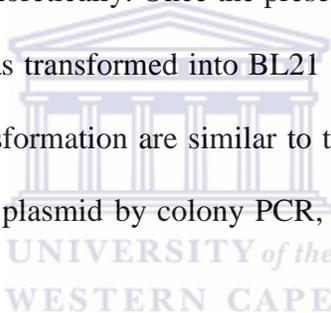
3.3.5. *Expression of full length GST-tagged human cyclophilin A in E.coli cells*

3.3.5.1. Optimization of protein expression conditions

Starter cultures supplemented with 0.4 % Glucose (1:50 dilution of 20 % glucose) were grown overnight in LB medium containing ampicillin at 37 ^{o}C with shaking at approximately 275 rpm. These were diluted 1:100 with LB containing ampicillin, and thereafter they were grown at 37 ^{o}C with shaking until they reached an OD_{600} of 0.8-1.0. This was followed by addition of Rhamnose and IPTG to a concentration of 0.1 % (1:200 dilution of 20 % Rhamnose) to induce protein expression and then incubated overnight at 37 ^{o}C with shaking.

3.3.5.2. Protein analysis

3.3.5.2.1. Preparation of cell lysates

Following the overnight expression reaction, the recovery of cells from LB medium post-expression was done by means of centrifugation, followed by quick freezing and then stored overnight at -80 °C. Frozen cells were thawed, resuspended with 0.5 % TritonX100 and Protease inhibitor, and then lysed by sonication. Cytoplasmic proteins were separated from cell debris (membrane-bound proteins) by means of centrifugation at 6 000 rpm for 30 min. The latter was re-suspended in 30% glycerol at room temperature (20-25 °C) followed by vigorous vortexing and pippeting. This cell suspension was incubated for 20-30 min at room temperature with shaking followed by thorough re-suspension to obtain a uniform suspension. Both cytoplasmic and membrane proteins were diluted with protein loading buffer in a ratio of 1:1 followed by analysis on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and then detected by means of Western Blot.
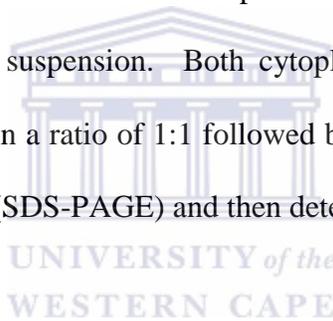
3.3.5.2.2. SDS-PAGE of total proteins

For SDS-PAGE analysis, sterile gel plates were assembled and verified for any leakages. 15 % of sodium dodecyl sulfate (SDS) gel was prepared using these plates according to the Fluka (Sigma) specifications. The 15 % separating gel (Acrylamide/Bisacrylamide-stock solution [30:0.8 %], separating gel buffer [pH 8.8]) and stacking gel (Acrylamide/Bisacrylamide-stock solution [30:0.8 %], stacking gel buffer [pH 6.8]) were each polymerized by addition of 0.8 % (w/v) ammonium persulphate (APS) and 10 % of tetramethylene diamine (TEMED) solution. For quality purpose, APS (10 % w/v) was prepared every week. Prior to sample loading into gel wells, a 1.5:1 mix ratio of protein sample with SDS loading dye was performed. Approximately 25 µl and 10 µl of total protein and protein marker, respectively, were loaded into protein wells. These were separated

using a Hoefer electrophoresis unit in SDS running buffer (25 mM Tris, 192 mM Glycine, 0.1 % (w/v) SDS (pH 8.3)) and proteins were separated according to their electrophoretic ability applying constant current at 20 amperes per gel. Following electrophoresis, the gel was stained overnight with Coomassie brilliant blue ([40 % (v/v) methanol, 10 % (v/v) acetic acid and 0.025 % (w/v) Coomassie brilliant blue R250]) and then de-stained in a solution containing 30 % methanol, 10 % glacial acetic acid and 60 % distilled H_2O, until the protein bands were visible. Alternatively the proteins were subjected to Western Blotting for detection of fusion proteins using specific antibodies.

3.3.5.2.3. Western blot analysis of total proteins

Prior to transfer the nitrocellulose membrane was equilibrated in 20 % (v/v) methanol. Proteins were transferred in transfer buffer (27 mM tris(hydroxymethyl)-amino-methane, 191 mM glycine and 20 % (v/v) methanol) in a submersion system at 100 V for 90 min. Next, the membrane was blocked with 3 % (w/v) milk in 0.05 % (v/v) Tween-20 in PBS solution for 30 min on a rocker at room temperature. The membrane was then incubated in primary antibody rabbit anti-GST in 3 % (w/v) milk in 0.05 % (v/v) Tween-20 in PBS solution (1:5000) overnight at 4 ^{o}C with rolling. This was followed by washing any unbound antibodies with PBS/0.05 % Tween for 45 min at 15 min intervals and then covered with HRP-conjugated goat anti-Rabbit (secondary antibody) for 60 min in 3 % (w/v) milk in 0.05 % (v/v) Tween-20 in PBS solution at a 1:2 000 ratio. The membrane was again washed twice with PBS/0.05 % Tween at 30 min intervals to again remove any unbound secondary antibody and thereafter 1 000 µl of 1 Component TMB membrane peroxidase substrate was added on the membrane to detect the

presence of recombinant protein from the total bacterial proteins. The presence of the protein of interest was indicated by the appearance of a purple color on the nitrocellulose membrane. Once the protein band became distinctly visible, the membrane was rinsed with PBS/0.05 % Tween and air dried.

3.3.5.2.4. Quantification of total protein concentration

For quantification of proteins, a Qubit-iT^{TM} Assay was used. 199 X n µl of Quant-iT^{TM} buffer (where n represents the number of standards plus number of samples) was mixed up with 1 x n µl of Quant-iT^{TM} reagent to a make up a Quant-iT^{TM} working solution. 190 µl of the latter was then mixed with 10 µl of standards, whereas 180-199 µl of working solution was mixed with user samples to make up a final volume of 200 µl. Following equilibration of the Qubit^{TM} fluorometer with known standards, fluorometer readings were taken and the concentrations (mg/ml) were calculated using the formula C = QF x 200/10 (where QF = Qubit fluorometer reading, C = concentration). Protein concentration was also estimated by means of Coomassie stain. Known BSA standards of varying concentration were run with total protein on SDS-PAGE and then stained with Coomassie. The concentration of total protein was estimated based on the degree-of-similarity of band intensity between the known BSA standards and total protein.

3.3.6. Purification of recombinant protein

3.3.6.1. Small scale purification

After induced bacterial expression, 100 ml of bacterial culture was pelleted by means of centrifugation at 14 000 rpm for 30 min. The pellet was resuspended with 2 ml of 1 %Triton and 40

µl of Protease inhibitor per gram of cell pellet. The pre cooled cell suspension was lysed using sonication followed by centrifugation at 14 000 rpm for 30 min. The cleared lysates was subjected to purification according to Promega standards using MagneGST^{TM} protein purification system. 100 µl of the thoroughly re-suspended MagneGST particles was pippeted into a 1.5 ml eppindorf tube. The particles were captured by placing the tube on a magnetic stand and the resulting supernatant was discarded. The particles were re-suspended and washed by 250 µl of Binding/Washing buffer for a total of three washes. After the final wash, the particles were re-suspended with 100 µl of Binding/Washing solution. To bind the cell lysate to the MagneGST particle, 500 µl of the cell lysate was added to the particles and then incubated at room temperature for 30 min with shaking. After incubation, the supernatant was aspirated meanwhile the particles were washed with 250 µl of Binding/Washing buffer for a total of three washes. After the final wash, 100 µl of elution buffer was added followed by incubation at room temperature for 15 min with gentle shaking. The particles were then captured by placing the tube on a magnetic stand while the purified GST-fusion was being aspirated from the supernatant and stored at -20 ^{o}C for further analysis.

3.3.6.2. Large-scale protein purification

After the small-scale purification was performed with successful purification of the protein of interest, a large scale was carried out in order to obtain a higher yield of the protein of interest. For large-scale protein purification the Pierce$^{®}$ GST spin purification kit was used. 500 ml of cell culture was pelleted by centrifugation at 14 000 rpm for 30 min and the supernatant discarded. The bacterial pellet was suspended in 4 ml 1 %Triton and 40 µl of Protease inhibitor by vortexing and/or vigorous pippeting until a homogeneous cell suspension was attained. This was followed by a centrifugation step at 14 000 rpm for 30 min to separate soluble proteins from insoluble proteins.

The protein of interest in the supernatant was then purified according to specification of Pierce® GST spin purification kit.

3.4. RESULTS

3.4.1. Generation of recombinant plasmid

The primers specific for the gene of interest were successfully designed and prepared to the desired concentration of 100 ρ.mole/μl stock solutions. Secondly, mRNA was extracted from MCF7 cells and reverse transcribed into cDNA. The latter was amplified by means of PCR into multiple copies of CypA. In figure 3.1, the sizes of the amplified human gene, CypA, is clearly shown. The lane represented by "M" indicates the 100 bp DNA marker with molecular weights specified along the left margin of the figure. On the same figure the house-keeping gene PBDG served as a positive control. This 396 bp gene is represented in lane 1 and 2 and appears as a faint band. The presence of PBDG on this figure confirms the presence of cDNA in the reaction thereby eliminating any false positive results. Furthermore, the results clearly indicate that the gene of interest was amplified from the cDNA library using these gene specific primers. This 498 bp DNA fragment is clearly indicated in lane 3.
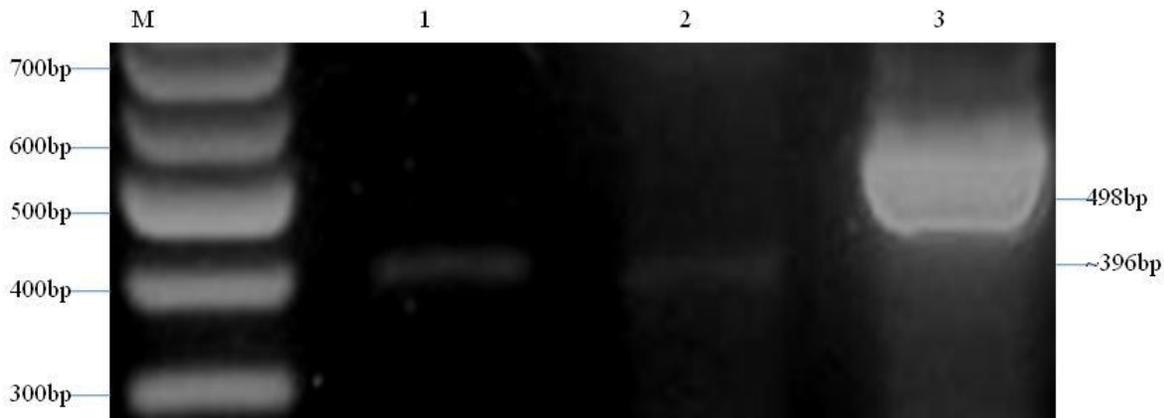
Figure 3.1: Agarose gel electrophoretic analysis of human cyclophilin A that has been amplified by Polymerase Chain Reaction using gene specific primers. "M" represents the 100 bp DNA ladder. The sizes of the DNA molecular weight are indicated on the left of the figure in base pairs. Lane 1 and 2 represent the housekeeping gene, Porphobilinogene Deaminase (PBDG) (396 bp), which serves as a positive control. Lane 3 represents the gene of interest, CypA (498 bp).
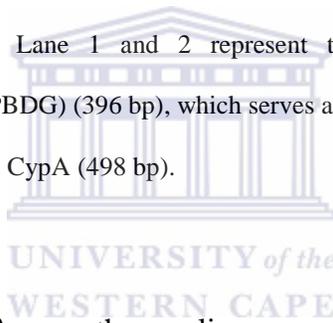
Following amplification of the CypA gene, the amplicon was successfully cloned into pGEM vector and later confirmed by means of blue/white selection following transformation into JM109 competent cells. The presence of white colonies on agar plates represented cells that were transformed, whereas blue/dark colonies represented untransformed cells. Moreover, successfully transformed cells were verified by means of colony PCR as indicated in figure 3.4. Following plasmid isolation from the transformed cells, the presence of the exact protein coding region of CypA was initially verified by restriction digest of pGEM with *EcoR1* and the presence of the DNA fragment is shown in figure 3.2, whereas the correct DNA sequence after sequencing is shown in figure 3.3. In figure 3.2, the 3 kbp DNA band on top of each lane represents the cleaved vector, whereas the bottom bands represent the genes of interest. Lane "M" represents the 100 bp DNA

ladder with molecular weights indicated on the left margin of the figure. Lane 1 and 2, on the other hand, represent cleaved pGEM vector, (3 kbp) that was cloned with hnRNP A1 (900 bp) and Lane 3 and 4 represent cleaved pGEM (3 kbp) that was cloned with CypA (498 bp).
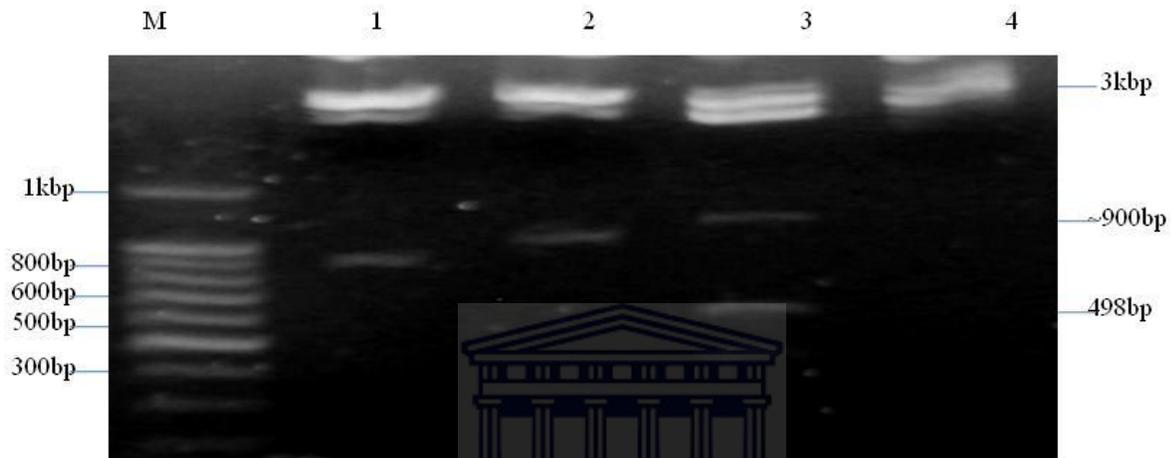
Figure 3.2: pGEM vector cleaved with restriction endonuclease *EcoR1*. The 3 kbp DNA on top of each lane represents the vector, whereas the bottom band represents the gene of interest. Lane "M" represents the 100 bp DNA ladder with molecular weights indicated on the left margin of the figure. Lane 1 and 2 represent cleaved pGEM vector (3 kbp) that was cloned with hnRNP A1 (900 bp), whereas lane 3 and 4 represent cleaved pGEM (3 kbp) that was cloned with CypA (498 bp).

In figure 3.3 below, the dark/black regions in this diagram represent homologous nucleotide sequences between the amplified gene of interest and known NCBI sequences. From this figure it is clearly evident that there is 100 % homology between the two compared sequences, meaning that there were no mutations or mismatching of bases on the gene of interest.

Figure 3.3: Sequence verification by comparing cyclophilin A (CypA) cloned into pGEM vector with known CypA sequence extracted from NCBI. Homologous regions between NCBI sequences and the cloned insert are represented by dark regions that stretch from 25-523 nucleotides. From the figure, it is clear that there is 100% homology between the cloned and known CypA sequence.

After the correct protein coding region was confirmed in pGEM vector it was later shuttled to a bacterial expression vector, pFlexi, which has been cut with the *Pme1* and *Sgf1* restriction enzymes. Successful shuttling of the protein coding into a bacterial expression vector (pFlexi) was verified by means of colony selection and colony PCR subsequent to transformation into KRX competent cells. Figure 3.4 indicates successful transformation of CypA genes onto KRX cells. From figure 3.4 it is clear that pFlexi was successfully cloned and transformed into competent cells.
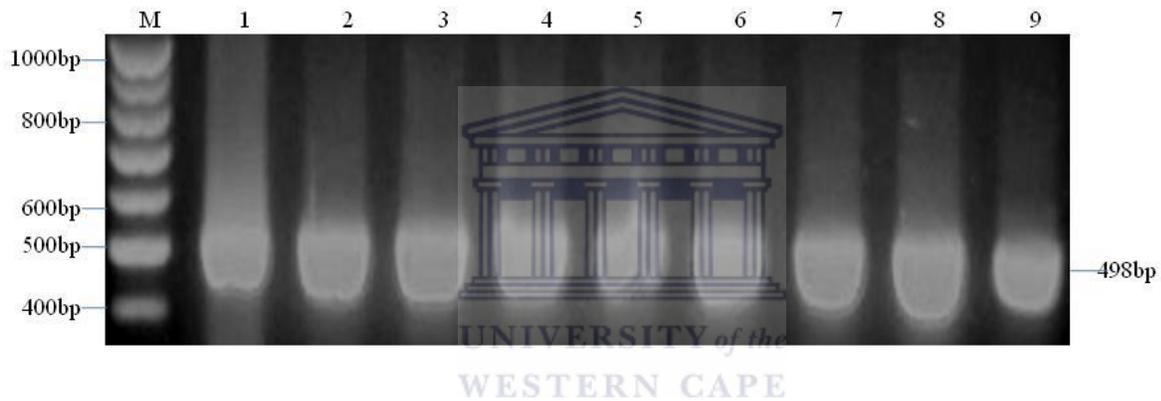
Figure 3.4: Colony PCR confirming the presence of CypA gene that has been transformed into KRX competent cells. On the figure "M" represent the 100 bp DNA ladder with molecular weights indicated along the left margin of the figure. Lane 1-8 represents successful transformation of CypA (498 bp) and lane 9 indicates the positive control.

The presence of the amplicon in pFlexi was again confirmed by means of restriction endonuclease digest and this phenomenon is shown in figure 3.5. This figure shows the presence of the gene and the vector following digestion with restriction enzymes. The approximately 4 kbp vector is presented on top of the figure in each lane of the agarose gel. On the other hand, the gene of interest presents at the bottom of lane 2 as a single band. Lane "M" on the figure represents the 100 bp

DNA ladder. Lane 1 represents the uncut plasmid that has been cloned with CypA. Lastly, lane 2 represents pFlexi/CypA that has been digested with *Pme1* and *Sgf1*.
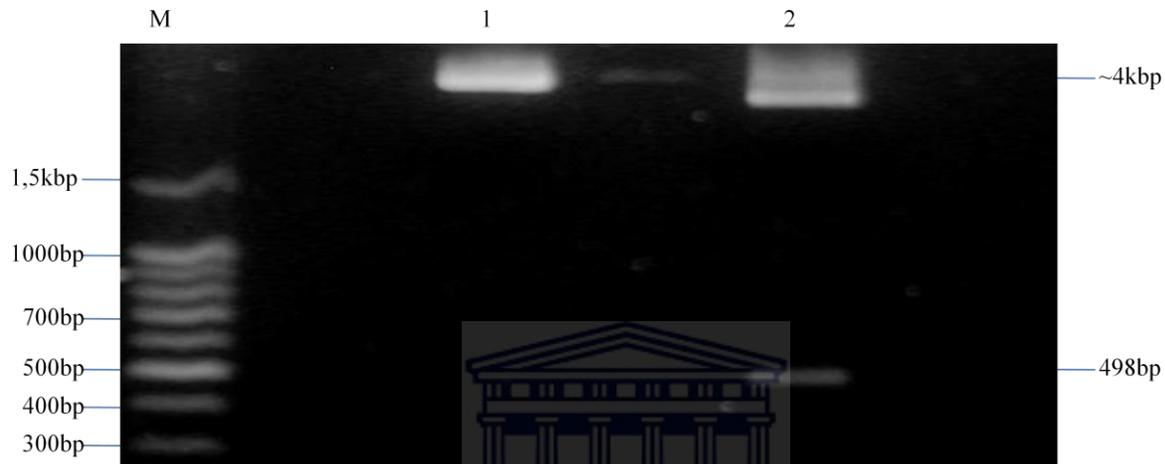
Figure 3.5: pFlexi cleaved with restriction endonucleases, *Sgf*1 and *Pme*1. Lane "M" represents the 100 bp DNA marker with molecular weights indicated along the left margin of the figure. Lane 1 and 2 represent uncut and cut pFlexi vector, respectively. The top bands in lane 1 and 2 with ~4 kbp indicate the cut pFlexi vector, whereas the 498 bp DNA band in lane 2 is CypA.

3.4.2. Expression and protein analysis in E.coli cells

Total proteins were successfully expressed and analyzed using SDS-PAGE, Coomassie blue and Western blot. Figure 3.6 represents total proteins, both cytoplasmic and membrane, expressed within the bacterial cells. From the figure it is evident that there is a multitude of proteins with

different molecular weight. The 42 kDa protein, CypA, is not apparent from total proteins as there are other proteins that are highly expressed in comparison to the protein of interest.
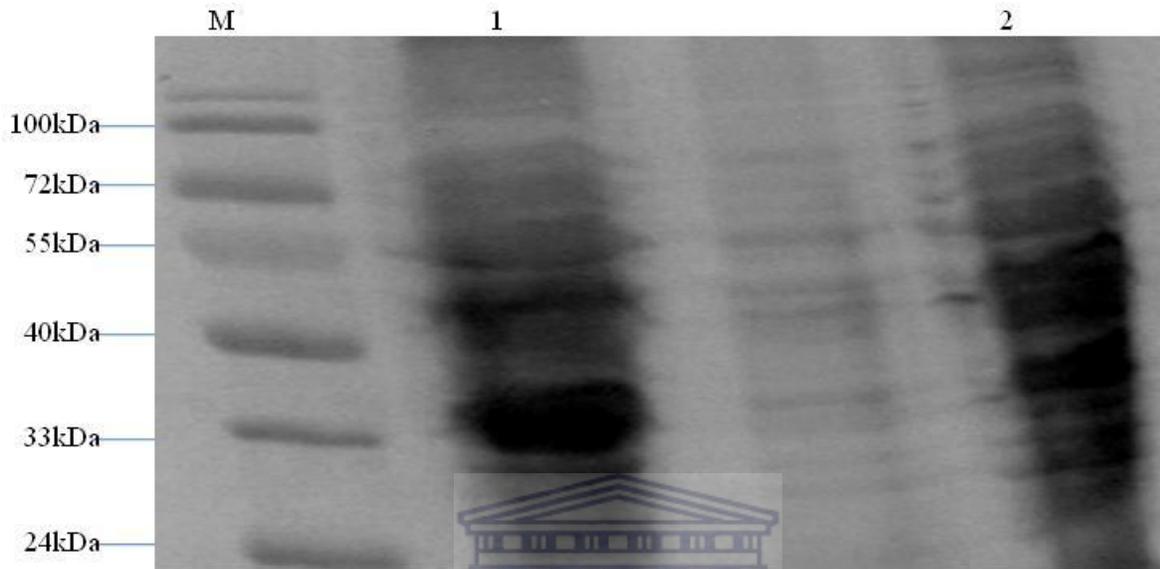
Figure 3.6: Coomassie stain showing the lysate of all expressed proteins. Lane "M" represents the protein marker with molecular weight designated along the left margin of the figure. Lane 1 represents cyctoplasmic proteins of *E. coli*, whereas lane 2 represents membrane proteins of *E. coli*.

Despite the indistinct appearance of CypA gene on Coomassie blue, we have managed to detect its presence from the total proteins by means of Western blotting. The dark band in figure 3.7 below indicates the GST-tagged protein that has been detected in nitrocellulose membrane using specific antibodies. From the results it is evident that the technique is highly specific to the GST tag that was fused to the protein of interest.
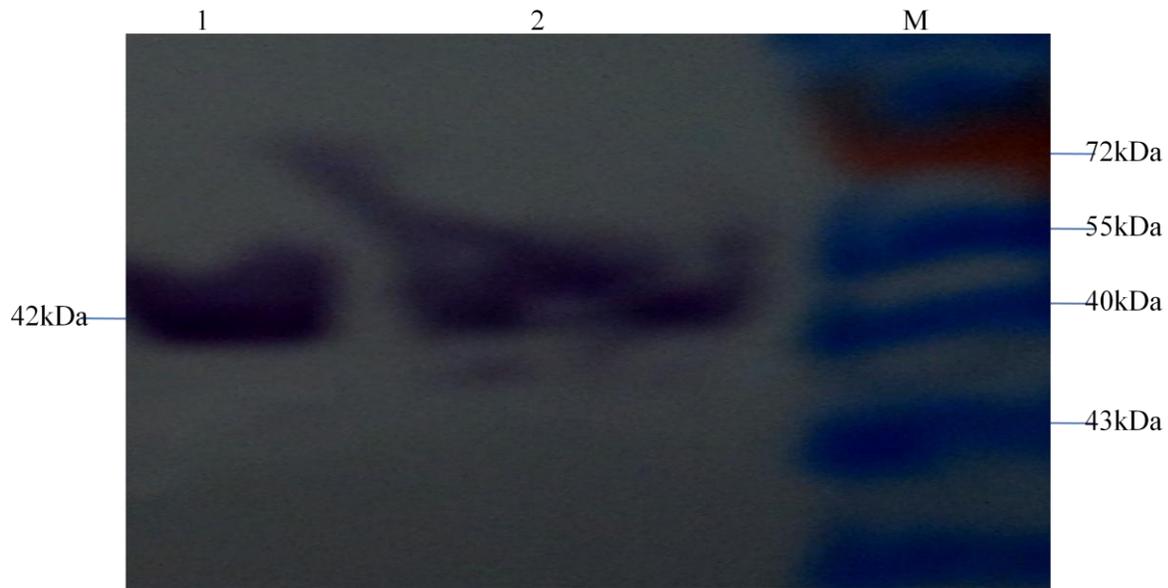
Figure 3.7: Western blot analysis depicting the expression of CypA in *E. coli*. Lane "M" represents the pre-stained protein marker with molecular weights designated along the right hand margin of the figure. The 42 kDa CypA is illustrated in lane 1 and 2 as a distinct purple band.
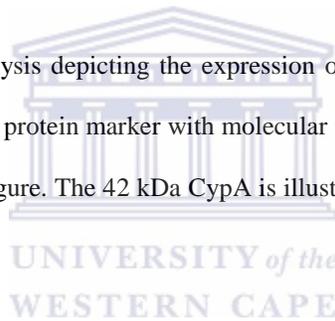
3.4.3. Purification of recombinant protein

The purification of recombinant protein was done using MagneGST protein purification system. Following optimization of protein expression the presence of CypA protein was confirmed by small-scale purification. The fusion proteins were successfully purified in small scale and this is clearly obvious in figure 3.8 below. After comparing the protein concentration of CypA with known BSA standards, it is appreciable that the purification step yielded low concentrations of the protein of interest. In figure 3.8, lanes 2-6 represent the known BSA standards in increasing concentration: 0.125 mg/ml, 0.25 mg/ml, 0.5 mg/ml, 0.75 mg/ml, 1.5 mg/ml, and 2 mg/ml. The measured

concentration of CypA using Qubit fluorometer revealed a concentration of 0.246 mg/ml and this

corresponds to the 0.25 mg/ml BSA standard that is represented in lane 3.
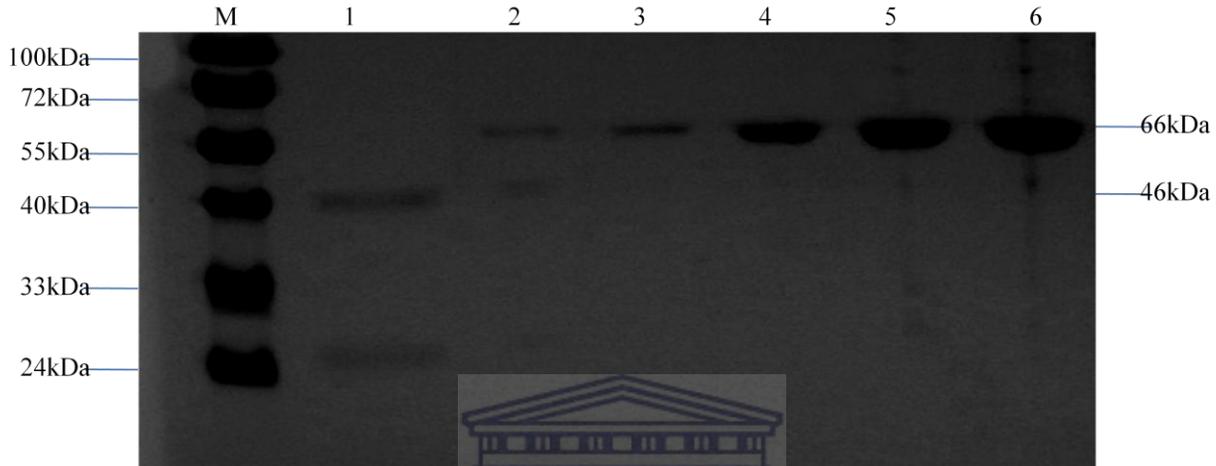
Figure 3.8: Coomassie stain showing small-scale purification of CypA protein. Lane "M"

represents the pre-stained protein molecular marker. Lane 1 indicates purified CypA with

a concentration of 0.246 mg/ml. Lane 2-6 represent bovine serum albumin (BSA)

standards in increasing concentration: 0.125 mg/ml, 0.25 mg/ml, 0.5 mg/ml, 0.75 mg/ml,

1.5 mg/ml, and 2 mg/ml. The intensity of the Cyp A band in lane 1 corresponds more or

less with that of lane 3, which is measuring 0.25 mg/ml.

3.5. DISCUSSION AND CONCLUSION

CypA is a peptidyl-prolyl *cis/trans* isomerase (PPIA) that is involved in several signaling events of

eukaryotic cells. CypA plays key roles in several different aspects of cellular physiology including

the immune response, transcription, mitochondrial function, cell death, and chemotaxis. Because of

their multiple functions elucidation of their mechanism of action received major attention and

therefore, became an active area of research. In comparison to CypB and CypC, which do not seem to exhibit much affinity for viruses, CypA has been shown to interact with capsid proteins of viral origin. In addition, CypA was previously reported to be associated with several tumor types but following its association with HIV-1 capsid protein it became evident that these traits can be exhibited in other viral capsid proteins. The CypA interaction with viral proteins was originally described for HIV-1 capsid protein (Zander, Sherman et al. 2003). Following extensive research on CypA it was later discovered that CypA interacts with other viral capsid proteins and more recently with hepatitis C virus capsid and SARS nucleocapsid protein. Most notably, in some studies these interactions have been mapped around a specific amino acid residue that seems to exhibit considerable affinity for capsid proteins. Most of these studies have revealed that these interactions are centered on the proline residue even though the proline position is not necessarily the same (Zander, Sherman et al. 2003). Therefore, it is suffice to say that proline is the stereochemically active site for interaction with capsid proteins.
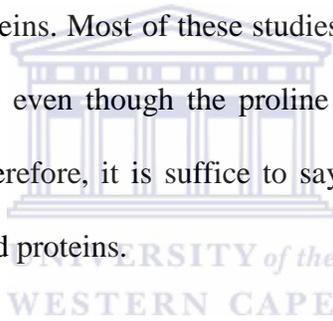
The present study was specifically designed to investigate the interaction of HCoV NL63 nucleocapsid protein and its truncated mutants with human cyclophilin A protein. This study follows the discovery of the SARS-N protein that was reported to exhibit direct and specific binding affinity to CypA. Given the evolutionary relationship between SARS CoV and HCoV NL63, it is reasonable to infer that the same kind of interaction that is observed with SARS-N protein will also be demonstrated by NL63-N protein. Furthermore, the study seeked to identify the exact domain of nucleocapsid protein that exhibits direct and specific interaction with human CypA. The outcomes of these interaction studies will shed light to some of the therapeutic strategies that could be explored to prevent the virus from causing disease.

Initially, the protein of interest was characterized by means of *in silico* analysis. Using an array of molecular techniques, CypA was then characterized both qualitatively and quantitatively. The protein of interest was successfully extracted from MCF7 cells, amplified by means of RT-PCR and then cloned into a bacterial expression vector. The latter was verified by automated genome sequencing, transformed into KRX cells and then expressed in a bacterial system. However, the key challenge for achieving the desirable protein yield was with the expression vector and competent cells. As a result, CypA was expressed in two different tags viz. GST- and HQ. However, it was noticeable that the patterns of expression differ between the two tags. Most notably, the expression vector requires IPTG to bind to the T7 promoter to induce protein expression. Competent cells, on the other hand, seem to be inhibited by addition of the inducer. Therefore, it is reasonable to assume that IPTG has toxic properties towards the competent cells; hence the observed protein yield. Secondly, because of the abovementioned properties of the expression vector the optimum OD_{600} requirements to induce protein expression had to be varying in order to obtain optimum expression. Additionally, the time frames to incubate cultures following protein expression had to fluctuate before attaining the desirable protein yield. The explanation for this variation in expression patterns remains obscure. Lastly, it is noticeable from the results that there is a sharp decline in protein concentration levels between cell lysates and purified protein. This can be attributed to proteolytic cleavage of expressed protein by cell proteases merely because CypA is a human protein, therefore, in a bacterial system it is recognized as foreign.

CHAPTER 4

Interaction of human cyclophilin A with coronavirus

nucleocapsid proteins: A pilot study

4.1. ABSTRACT

Cyclophilins (Cyp) are a family of proteins that possess peptidyl-prolyl cis-trans isomerase activity and play a key role in the folding of proteins (Lee 2010). They have been reported to be associated with several diseases and most commonly with severe types of carcinoma but their role remain unclear. CypA is the first member of mammalian cyclophilins to be identified and it plays key roles in several different aspects of cellular physiology including the immune response, transcription, mitochondrial function, cell death, and chemotaxis (Galat and Bua 2010). In addition to these key physiological functions of CypA, it has also been reported to interact with viral proteins, especially caspid proteins. Most recently, it has been revealed to interact with severe acute respiratory syndrome (SARS) nucleocapsid (N) protein. SARS emerged in 2002-2003 as a coronavirus from Southern China and spread throughout the world (Wang, Ji et al. 2003; Lu, Zhao et al. 2004; Hao, Chen et al. 2006). Coronaviruses are enveloped single stranded positive sense RNA viruses about 80-160 nm in diameter. SARS generally cause disease much like the common cold, but they have also been associated with more severe lower respiratory tract conditions, especially in frail patients (Woo, Lau et al. 2005; Dijkman, Jebbink et al. 2006; Shao, Guo et al. 2007). For the purpose of this study, the focus is based on CypA interaction with HCoV NL63-N protein. These interactions might play a role in the pathology of HCoV-NL63. Using Glutathione-S-transferase (GST), the interaction of CypA with nucleocapsid protein can be demonstrated to be direct and specific. Since the N protein is involved in viral RNA packaging to form a helical core, it is suffice to say that both NL63-N and CypA are possibly within the HCoV-NL63 replication/transcription complex and NL63-N/human CypA interaction might function in the regulation of HCoV-NL63 RNA synthesis. In addition, the study intends to demonstrate that HCoV-NL63-N has a specific domain for interacting with CypA.

4.2. INTRODUCTION

Coronaviruses and toroviruses are members of the *Coronaviridae* family, which along with the *Roniviridae* and *Arteriviridae* families form the order *Nidovirales*. Coronaviruses are responsible for approximately 30% of human upper respiratory infections each year (Yeung, Yamanaka et al. 2006). In November 2002, SARS emerged in China and caused more than 8 000 cases of SARS worldwide (Zhou, Liu et al. 2008). Approximately 10% of these cases were fatal. Coronaviruses are categorized into three groups based on their serologic and genotypic characteristics. Group 1 and 2 contain mammalian viruses, while group 3 consists of avian viruses. Group 2 viruses are also characterized by the presence of a gene encoding the heamaglutinin esterase (HE) protein, which is absent in the other groups (Dijkman, Jebbink et al. 2006; Frieman, Yount et al. 2007; Dijkman, Jebbink et al. 2008).

All coronaviruses are enveloped single stranded positive sense RNA viruses about 80-160 nm in diameter. Their genome has the following gene order: 1a-1b-S-ORF3-E-M-N. The 5′-two thirds of the genome encodes for a polyprotein that contains all the enzymes necessary for replication, whereas the 3′-one third encodes for all the structural proteins that mediate viral entry into the host cell (Hogue and Brian 1986; Hao, Chen et al. 2006; Junwei, Baoxian et al. 2006). The structural proteins include spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins.

Nucleocapsid (N) protein has been shown to be the major antigen in coronaviruses (Lu, Pan et al. 2011). It is encoded by the ninth open reading frame (ORF) of SARS, which is the same ORF that encodes for an accessory protein called ORF9, but they have different ORFs (Hogue and Brian 1986; Almazan, Galan et al. 2004; He, Leeson et al. 2004; Shang, Qi et al. 2006; Muller, van der Hoek et al. 2010). This particular protein has the capability to regulate cellular signal pathways

such as arresting cell cycle progression, inducing apoptosis and actin reorganization and activating the AP-1 signal transduction pathway (Kang, Plescia et al. 2007; Emmott, Rodgers et al. 2010; Li, Lai et al. 2011). Furthermore, N protein plays an imperative role in viral core formation and viral packaging. It has also been shown to interact with cellular proteins including CypA, hnRNP A1, human ubiquitin-conjugating enzyme and CDK-cyclin complex proteins, IkBα, Cyt P450 2A6 etc. Even more interesting, this protein is known to interact with other structural proteins i.e. membrane protein, and also has the capability to self-associate (Klumperman, Locker et al. 1994; Chan, Ma et al. 2007; Bai, Hu et al. 2008; Liu, Sun et al. 2010).

The literature indicates that interaction between SARS-N and M protein *in vivo* takes place along a specific stretch of amino acids found in the N protein. The same region has been discovered to be the region required for multimerization of the N protein suggesting that it may be crucial in maintaining correct conformation of the N protein for self-interaction and interaction with the M protein (Schaecher, Diamond et al. 2008; Liu, Sun et al. 2010). The M-N interactions are known to be crucial for viral assembly and morphogenesis. Luo et al (2004) have mapped M and N protein interacting regions using yeast two hybrid and surface plasmon resonance technique, to residues 197-221 and 351-422, respectively (He, Leeson et al. 2004; Chan, Ma et al. 2007). These fragments were shown to be highly charged at neutral pH, suggesting that their interaction may be of ionic nature. However, this interaction was shown to be weakened by acidification, higher salt concentration (400 mM NaCl) and divalent cations (50 mM Ca^{2+}), which suggests that electrostatic attraction might play a role in this interaction (Liu, Sun et al. 2010). The latter could provide a possible therapeutic strategy aiming at disrupting association between M and N protein. However, in mouse hepatitis virus (MHV) and transmissible gastroenteritis virus (TGEV) the stretch of amino acids where these interactions occur were mapped at different locations. Residues 201-224 of

MHV-M protein and residues 233-257 of TGEV-M protein were mapped to be involved in the M-N protein-protein interaction (Wesseling, Godeke et al. 1993; Versteeg, van de Nes et al. 2007; Ye, Hauns et al. 2007).

The property of self-association of the N protein is of crucial importance, because some laboratories have some interesting discoveries. Dimer formation has been observed both for recombinant proteins as well as during viral infection. Some researchers have shown that the recombinant full length N protein is dimeric with the propensity to form tetramers and higher molecular weight oligomers. However, the homodimer formation of the N-homodimer has been shown to be inhibited by excess truncated mutant containing the dimerization domain (He, Leeson et al. 2004; Tang, Wu et al. 2005). These oligomerization domains of N protein were mapped through studies of the oligomeric states of several truncated mutants. SARS-N carboxy (C)-terminal domain has proven to be the site of dimerization at 285-422 amino acid residues (Schreiber, Kamahora et al. 1989). Therefore, the dimerization domain is mapped to the C-terminal 138 residues, and the homodimer formation is inhibited by excess truncated mutants containing the dimerization domain. The two domains of the N protein viz. amino (N) - and C-terminal domains are believed to play the above mentioned roles. The N-terminal domain, which is known to bind RNA exhibits relatively long sequences of positively, charged amino acids. C-terminal, on the other hand, possesses a lysine rich region (373-390), which is predicted to be nuclear localization signal (NLS) (Timani, Liao et al. 2005; Emmott, Dove et al. 2008). Previous studies have determined the dimerization domain at the C-terminal but at different amino acid residues. Another study determined that the serine/arginine (SR)-rich motif is crucial for N-N oligomer formation and deletion of this motif lead to loss of N-N interaction (He, Leeson et al. 2004; Calvo, Escors et al.

2005; You, Dove et al. 2005). Researchers believe that this self-association may play a role in ribonucleoprotein formation and subsequently in the viral assembly process.

As previously mentioned, the M protein interacts with the mid-portion of the N-protein. Surprisingly, the same region also interacts with hnRNP A1, and structural studies have identified the region between amino acid 45-181 as the putative RNA binding region, which is close to the N-terminus (Luo, Chen et al. 2005). The MHV-N protein was first shown in 1986 to have RNA binding activity by North western assays (Klumperman, Locker et al. 1994; Schickli, Zelus et al. 1997; Luo, Chen et al. 2005). Although there is no sequence homology to previously described RNA-binding motifs, the central domain of N protein has been shown to be the RNA binding region. The N protein is known to interact with viral RNA genome to form a helical nucleocapsid structure. This interaction plays a role in a number of viral processes including viral RNA dependent RNA transcription, genome replication and translation. Nucleic acid binding activity of N-protein has been extensively studied for MHV, bovine coronavirus (BCV) and infectious bronchitis virus (IBV), and this particular interaction was found to be specific to viral RNA and not affected by addition of non-specific competitor RNA. The amino acid residues where N protein interacts with RNA exhibit differences in location, but they have been found to be rich in serine and arginine residues. This motif is known to be conserved in all coronavirus N protein (Cavanagh 2003; Zuniga, Sola et al. 2004; You, Reed et al. 2007).

The main purpose of this study is the interaction of NL63-N protein with CypA. The coronavirus N proteins exhibit high binding affinity to CypA and this association has been established in SARS-N protein studies. In addition to CypA, the nucleocapsid protein is known to exhibit similar binding affinity to hnRNP A1 and this binding was initially established with MHV-N protein. The binding site of hnRNP A1 with MHV-N protein was determined in two regions of

the protein using yeast two hybrid assay (Luo, Chen et al. 2005). These included N-terminal region (1-292) and C-terminal end (391-455). Recently this association was mapped to the 161-210 amino acid fragment of SARS-N protein and to the 203-320 glycine rich region of hnRNP A1. If this interaction is *in vivo*, it may play a significant role in regulation of the viral RNA synthesis. Moreover, hnRNP A1 has been demonstrated to interact direct and specific, *in vitro* and *in vivo*, with IkBα and this interaction is required for maximum activation of nuclear factor kappa light-chain enhance of activated B cells (NF-kB)-dependent transcription. Thus, in addition to regulating splicing, polyadenylation, and mRNA transport, hnRNP A1 also contributes to the fine control of NF-kB-dependent transcription (Luo, Chen et al. 2005). The region in hnRNP A1 which is required for interaction with IkBα is located between residues 95-207. Literature indicates that in cells lacking hnRNP A1, NF-kB activation is defective but re-introduction of hnRNP A1 into these cells restores an efficient NF-kB response to signal transduction (Fang, Gao et al. 2007). NF-kB is a protein complex that acts as a transcription factor, and it plays a key role in regulating immune response to infection. Incorrect regulation of this complex has been linked to cancer, inflammatory and auto-immune diseases, septic shock, viral infection and improper immune development. Therefore, if HCoV-N protein interacts with hnRNP A1, the activation of NF-kB will be defective resulting in the above mentioned consequences. This interaction is mediated by the C-terminal region of IkBα and one of the N-terminal RNA binding domains of hnRNP A1.

In addition to self-association, interaction with viral RNA, hnRNP A1, CypA and M protein, some studies have discovered that SARS-N also interacts with human phosphoprotein B23 both *in vitro* and *in vivo*, and this interaction is dependent on SR-rich domain (175-210) in the N protein (Yuan, Yao et al. 2005; Zeng, Ye et al. 2008). The ultimate result of this interaction is inhibition of phosphorylation of B23 at Thr99. Since B23 plays a role in centrosome duplication

during cell cycle progression, the SARS-N-B23 interaction can lead to cell cycle arrest since the centrosome cannot initiate duplication (Zeng, Ye et al. 2008).

These protein interactions form the basis of this study. Determination of these interactions and identification of amino acid sequences responsible for these interactions are necessary for the elucidation of molecular mechanisms of HCoV-NL63 replication and rationalization of therapeutic intervention. This can be achieved by means of protein tagging, because this technique facilitates protein detection using a peptide. This allows detection and affinity purification of a protein in the absence of a specific antibody against the protein of interest. Previous studies have shown that these interactions amongst various coronaviruses are critical for viral assembly as well as infection process.

4.3. MATERIALS AND METHODS

4.3.1. Plasmid DNA isolation

The glycerol stocks of GST tagged nucleocapsid genes cloned into pFlexi vector were kindly provided by Mr. Berry (University of the Western Cape, Department of Medical Biosciences). The genes included SARS-N and NL63-N gene. In addition to the full length gene of NL63, the gene was also designed in such a way that it was truncated into C-terminal and N-terminal domains. In this study, these genes are designated as SARS-N, N1, N2 and N3 for the full length SARS-N, full length NL63-N, C-terminal NL63-N, and N-terminal NL63-N gene, respectively. Table 4.1 provides a brief description of these genes as well as GST that was used to tag the proteins.

Table 4.1: Plasmid constructs used in this study.

Construct Name	Description of recombinant GST-tagged protein
N/A[#]	GST tagged protein control to confirm plates were adequately coated
GST	GST only to confirm antigenicity detected was not directed against the fusion protein
SARS-N	SARS-CoV full length N (a.a. 1-423)
N1	HCoV-NL63 full length N (a.a. 1-378)
N2	N-terminal half of the HCoV-NL63 N (a.a. 1-189)
N3	C-terminal half of the HCoV-NL63 N (a.a. 191-378)

[#]N/A – Not applicable
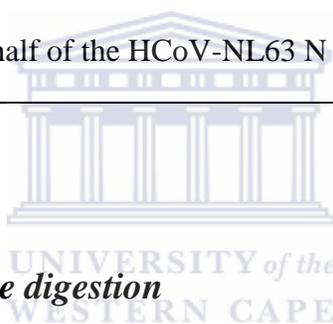
4.3.2. Restriction endonuclease digestion

The glycerol stocks were spread on Luria Bertani (LB)-agar plates Luria Bertani (LB) broth (10 g pancreatic digest of casein, 5 g yeast extract powder, 5 g NaCl, and 15 g of agar bacteriological per 1L water) supplemented with ampicillin (100 mg/ml) and incubated overnight (12-16 hours) at 37^{0}C. Colonies were selected for PCR as well as inoculation onto 15 ml of LB broth. Once the presence of the correct genes was verified in gel electrophoresis, the cultures that correspond to the confirmed colonies were used for plasmid DNA isolation using the Wizard plus SV Minipreps DNA purification system. Furthermore, glycerol stocks were prepared from these cultures. The presence of the exact gene was further confirmed by restriction enzyme digest following the plasmid isolation from the competent cells. Table 4.2 illustrates the standard reaction that was

performed when conducting restriction digests. Aliquots of these constructs were digested with pFlexi restriction enzymes (*Sgf1* and *Pme1*).

Table 4.2: pFlexi restriction enzyme digests.

Reagent	pFlexi enzyme digest reaction (µl)
5 x Flexi digest buffer	2
DNA constructs	10
Sgf 1and *PmeI* enzyme blend	2
Nuclease free water	4

4.3.3. Expression of the nucleocapsid proteins in E. coli cells

4.3.3.1. Optimization of expression conditions

The conditions for protein expression were optimized using the prepared glycerol stocks. Starter cultures supplemented with 0.4 % Glucose (1:50 dilution of 20 % glucose) were grown overnight in LB medium containing ampicillin (100 mg/ml) at 37 ^{o}C with shaking at approximately 275 rpm. These were diluted 1:100 with LB containing ampicillin, and thereafter they were grown at 37 ^{o}C with shaking until they reached an Optical Density (OD$_{600}$) of 0.8-1.0. This was followed by addition of Rhamnose (10 %) and IPTG (100 mM) to a concentration of 0.1 % (1:200 dilution of 20 % Rhamnose) to induce protein expression and then incubated overnight at 37 ^{o}C with shaking.

4.3.3.2. Protein analysis

4.3.3.2.1. Preparation of cell lysate

Following the overnight expression reaction, the recovery of cells from LB medium post-expression was done by means of centrifugation, followed by quick freezing and then stored overnight at -80 ^{o}C. Frozen cells were thawed, resuspended with 0.5 % TritonX100 and Protease inhibitor (1 ml per 4 g of cells), and then lysed by sonication. The cytoplasmic proteins were separated from cell debris (membrane proteins) by means of centrifugation at 6 000 rpm for 30 min. The later was re-suspended by 30 % glycerol at room temperature (20-25 ^{o}C) followed by vigorous vortexing and pippeting. This cell suspension was incubated for 20-30 min at room temperature with shaking followed by thorough re-suspension to obtain uniform suspension. Both cytoplasmic and membrane proteins were diluted with protein loading buffer in a ratio of 1:1 followed by analysis on SDS-PAGE and then detected by means of Western Blot.
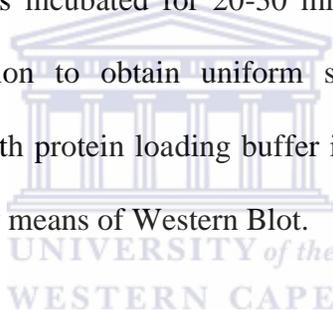
4.3.3.2.2. SDS-PAGE analysis

For SDS-PAGE analysis, sterile gel plates were assembled and verified for any leakages. 15 % of SDS gel was prepared using these plates according to the Fluka specifications. For quality purpose, APS (10 %w/v) was prepared every week. Prior to sample loading into gel wells, a 1.5:1 mix ratio of protein sample with loading dye was performed. Approximately 25 µl and 10 µl of total protein and protein marker, respectively, were loaded into protein wells. These were separated according to their electrophoretic ability applying constant current at 20 amperes per gel. Following this reaction the gel was either submerged in Commassie stain (45 % methanol, 10 % glacial acetic acid, 45 %

water, and 3 g/L Coomasie brilliant blue R250) overnight or subjected to Western blotting for further experiments.

4.3.3.2.3. Western blotting of total proteins

The separated proteins were transferred to a nitrocelullose membrane for Western blot studies. The SDS gel was transferred to a Western blot transfer system covered with nitrocellose membrane that has been sandwiched with Whatmann paper and sponges immersed in pre-cooled transfer buffer. The experiment was carried out at 4 ^{o}C with the transfer system running at 100 V and constant amperes for a period of 120 min.

On completion of the transfer, the nitrocellulose membrane was immersed in Ponceau stain with shaking for 3 min and then blocked with 3 % (w/v) milk for approximately 30 min. The membrane was then exposed to Rabbit anti-GST (primary antibody) in a ratio of 1:5000 overnight at 4 ^{o}C. This was followed by a washing step with PBS/0.05 % Tween for 45 min at 15 min intervals and then covered with Goat anti-Rabbit (secondary antibody) for 60 min in a 1:2 000 ratio. The membrane was again washed twice with PBS/0.05 % Tween at 30 min intervals and thereafter 1 000 µl of 1 Component TMB membrane peroxidase substrate was added on the membrane to detect the presence of recombinant protein from the total bacterial proteins. The presence of the protein of interest was indicated by the appearance of a purple color on the nitrocellulose membrane.

4.3.3.2.4. Quantification of protein concentration

For quantification of proteins, a Qubit-iT^{TM} Assay was used. 199 x n µl of Quant-iT^{TM} buffer (where n represents the number of standards plus number of samples) was mixed up with 1 x n µl of

Quant-iT^{TM} reagent to a make up a Quant-iT^{TM} working solution. 190 µl of the latter was then mixed with 10 µl of standards, whereas 180-199 µl of working solution was mixed with user samples to make up a final volume of 200 µl. Following equilibration of the Qubit^{TM} fluorometer with known standards, fluorometer readings were taken and the concentrations (mg/ml) were calculated using the formula C = QF x 200/10 (where QF = Qubit fluorometer reading, C = concentration).

Protein concentration was also estimated by means of Coomassie stain. Known BSA standards of varying concentration were run with total protein on SDS-PAGE and then stained with Coomassie. The concentration of total protein was estimated based on the degree of band intensity between the known BSA standards and total protein.
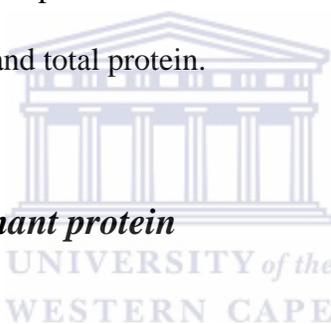
4.3.4. Purification of recombinant protein

4.3.4.1. Small scale purification

For small-scale protein purification the B-PER® GST Fusion Protein Purification Kit was used. 250 ml of cells was pelleted by centrifugation at 14 000 rpm for 30 min and the supernatant removed. The bacterial pellet was suspended in 10 ml of B-PER reagent and 40 µl of Protease inhibitor by vortexing and/or pippeting until a homogeneous cell suspension was attained. This was followed by a centrifugation step at 14 000 rpm for 15 min to separate soluble proteins from insoluble proteins and cell debris. The protein of interest in the supernatant was then purified according to specification of Thermo Scientific using B-PER® GST Fusion Protein Purification Kit.

4.3.4.2. Large scale purification

After induced bacterial expression, 500 ml of bacterial culture was pelleted by means of centrifugation at 14 000 rpm for 30 min. The pellet was resuspended with 2 ml of 1 %Triton and 40 µl of Protease inhibitor per gram of cell pellet. The pre-cooled cell suspension was lysed using sonication followed by centrifugation at 14 000 rpm for 30 min. The cleared lysates was subjected to purification according to Promega standards using MagneGST^{TM} protein purification system.

4.3.4.3. GST-pull down assay

An untagged CypA that was used for GST-pull down assay was purchased from Sigma Aldrich. The molecular weight of this protein was verified comparing it with the GST-fusion CypA. This was done by cleaving GST-tag from CypA with TEV protease, releasing untagged CypA and then ran on SDS PAGE (Results not shown). The cleaved protein molecular weight was compared with the purchased CypA on SDS PAGE.

For particle equilibration, the MagneGST™ Particles were thoroughly re-suspended by inverting the bottle several times to obtain a uniform suspension. 20 µl of MagneGST™ particles were pipette into a 1.5 ml tube. As a pre-cautionary measure, the MagneGST™ particles were not allowed to settle for more than a few minutes during capture of the bait protein as this will reduce binding efficiency. The tubes were placed in the magnetic stand and allow the MagneGST™ particles to be captured by the magnet and capture typically occurs within a few seconds. From the tubes, the supernatant was carefully removed and discarded followed by removal of the tubes from the magnetic stand. 250 µl of MagneGST™ Binding/Wash Buffer was added to the particles and

re-suspend by pipetting or inverting. Again, the supernatant was removed and discarded. This was repeated two more times for total of three washes.

Following the equilibration of the MagneGST particles, 100 µl of MagneGST™ Binding/Wash Buffer was added in order to re-suspend the particles. 30 µg of the GST-fusion proteins was added to the MagneGST particles. This was made up to a final volume of 300 µl with Radio Immuno Precipitation Assay buffer (50 mM Tris HCl pH 8, 150 mM NaCl, 1% NP-40, and 0.5 % sodium deoxycholate). This reaction mixture was incubated (with constant gentle mixing) for 1 hour at room temperature on a roller. After incubation, the tubes were placed on the magnetic stand and allowed the MagneGST™ particles to be captured by the magnet. Once again the supernatant was carefully removed but saved for gel analysis. 250 µl of MagneGST™ Binding/Wash Buffer was added to the particles and mixed gently. This was then incubated at room temperature for 5 min while mixing occasionally by tapping or inverting the tube. The tubes were placed in the magnetic stand and allow MagneGST™ particles to be captured by the magnet. The supernatant was again removed and saved for analysis of wash. 250 µl of MagneGST™ Binding/Wash Buffer was added to the particles and mixed gently by inverting the tube. The tubes were again placed on the magnetic stand and allow the MagneGST™ particles to be captured by the magnet. Supernant was removed and saved for analysis. This step was repeated for a total of three washes. After the last wash, the particles were re-suspended in 20 µl of MagneGST™ Binding/ Wash Buffer. A 3–5 µl aliquot was removed for analysis of the efficiency of immobilization of the GST-fusion protein or GST control onto the particles. To this aliquot, 20 µl 1 x SDS loading buffer was added and proteins eluted by boiling for 5 min and then analyzed by SDS-PAGE.

To each tube of particles carrying GST-fusion protein, GST control and beads 30 µg of CypA was added. 155 µl MagneGST™ Binding/Wash Buffer and 20 µl 10 % BSA was added for each pull-down reaction, and then incubated overnight (12-16 hours) with gentle mixing at 4 ^{0}C on a roller platform. To remove non-specific adherent proteins, the tubes were briefly vortexed at the end of this incubation period. The tubes were placed on a magnetic stand for MagneGST particles to be captured. Once again, the supernatant was removed and saved for analysis. To each tube, 400 µl of MagneGST™ Binding/Wash Buffer was added and mixed gently by inverting the tube. This was incubated at room temperature for 5 min while mixing occasionally by tapping or inverting the tube. All tubes were placed in the magnetic stand and allow the MagneGST™ particles to be captured by the magnet. The supernatant was removed and saved for analysis (it is especially important to keep this fraction during initial optimization for troubleshooting purposes). 400 µl of MagneGST™ Binding/Wash Buffer was added and mixed gently by inverting the tube. The tubes were placed in the magnetic stand and allow the MagneGST™ particles to be captured by the magnet. The supernatant was removed and save for analysis. The washing step was repeated for a total of three washes. Elution of proteins was done by adding 20 µl of 1 x SDS loading buffer; incubate at room temperature for 5 min with occasional mixing. The tubes were placed in the magnetic stand and allow the MagneGST™ particles to be captured by the magnet. The eluate was removed for analysis.

4.4. RESULTS

4.4.1. Generation of recombinant plasmid

The gene specific primers were designed as shown in table 4.1 and 4.2 and then prepared to a concentration of 100 ρ.mole/µl. In figure 4.1 below, it is clearly shown that nucleocapsid genes

were successfully amplified. The genes are arranged as follow: SARS-N (1 320 bp), N1 (1 290 bp), N2 (620 bp), and N3 (620 bp). The results clearly indicate that all the genes of interest were amplified using these gene specific primers and this is evident from figure 4.1.
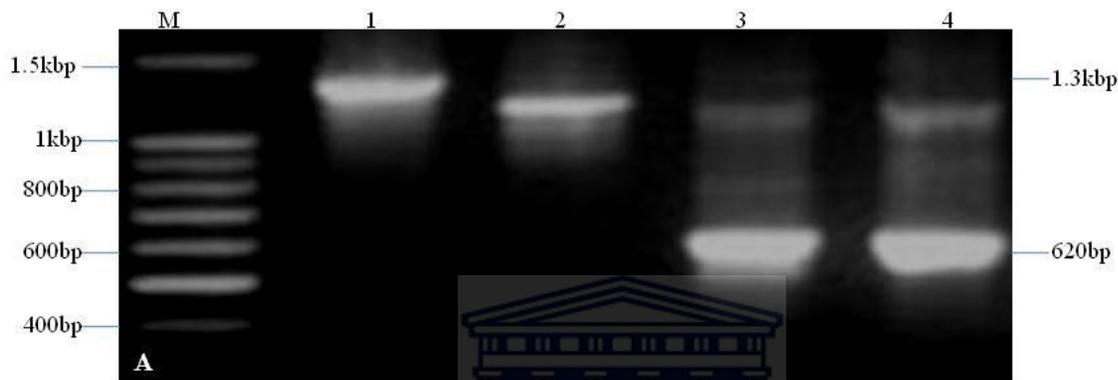
Figure 4.1: Polymerase Chain Reaction of nucleocapsid (N) genes using gene specific primers. "M" represents the 100 bp DNA marker with molecular weights indicated along the left margin of the figure. Lane 1-4 represents N genes in the following order: SARS-N, N1, N2, and N3.

In addition to the polymerase chain reaction (PCR), the presence of the protein coding region was verified by means of restriction enzyme digest. Figure 4.2 illustrates the release of the gene from the plasmid following its treatment with *Sgf1* and *Pme1* restriction enzymes. The 4 kbp DNA on top of each lane represents the vector, whereas the bottom bands represent the genes of interest. From figure 4.2, it is clear that the vector was successfully cleaved at the correct restriction sites with the concomitant release of the amplicon.
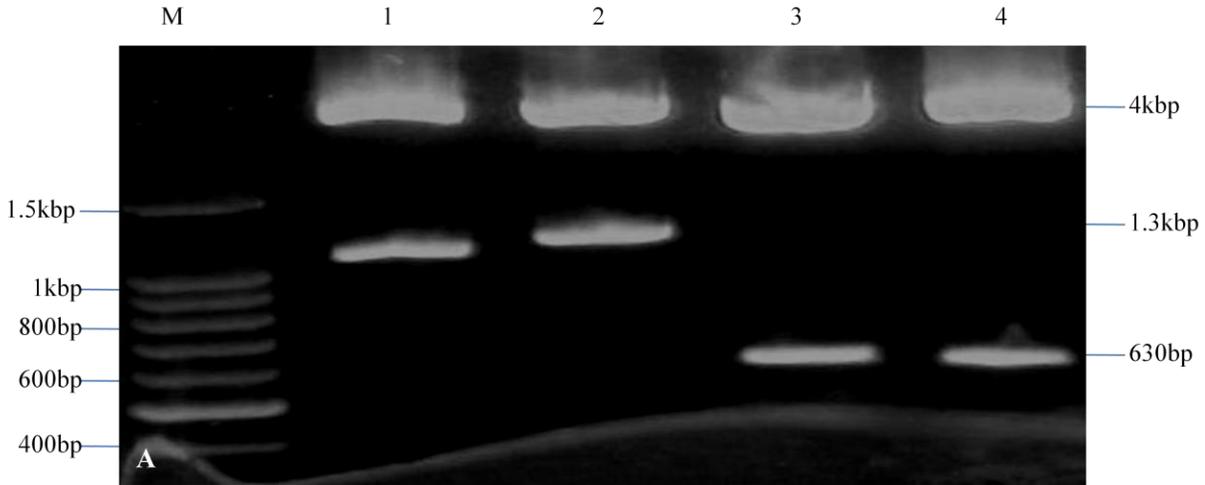
Figure 4.2: pFlexi vector is cleaved with restriction endonuclease *Sgf1 and Pme1.* "M" represents the 100 bp DNA ladder. Lane 1-4 represents cleaved pFlexi in the following order: SARS-N, N1, N2, and N3. The 4 kbp DNA on top of each lane represents the vector, whereas the bottom bands represent the genes of interest.

Following the release of the genes from the plasmid, they were re-ligated into the expression vector. Successful cloning of the exact protein coding region into vector (pFlexi) was verified by means of colony selection and colony PCR subsequent to transformation into KRX competent cells. Figure 4.3 indicates successful transformation of nucleocapsid genes onto KRX cells with the exception of SARS-N, which is shown in figure 4.4. From these figures it is clear that pFlexi was successfully cloned and transformed into competent cells.
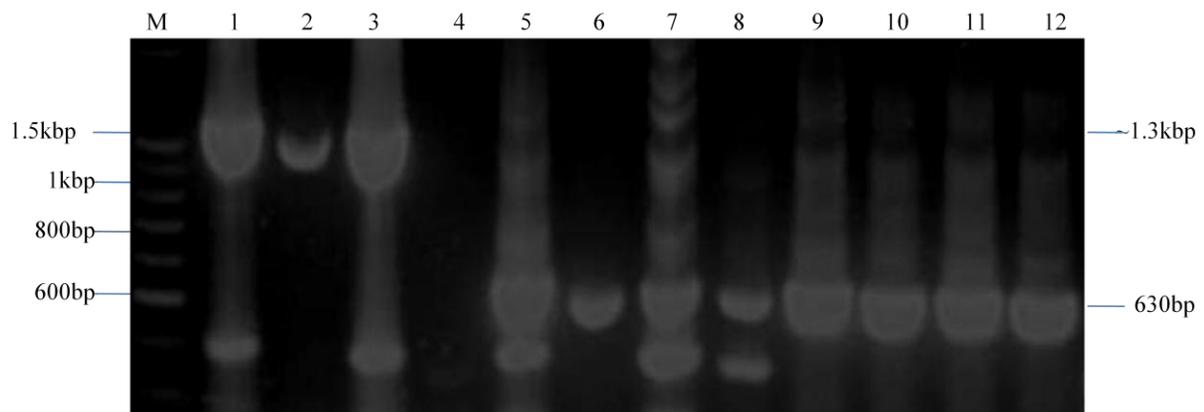
Figure 4.3: Colony PCR confirming the presence of NL63-N genes that have been transformed into KRX cells. "M" represents the 100 bp DNA ladder with molecular weight indicated along the left margin of the figure. Lane 1-3 represents successful transformation of N1/pFlexi. Lane 5-8 represents successful transformation of N2/pFlexi. Lane 9-12 represents successful transformation of N3/pFlexi.
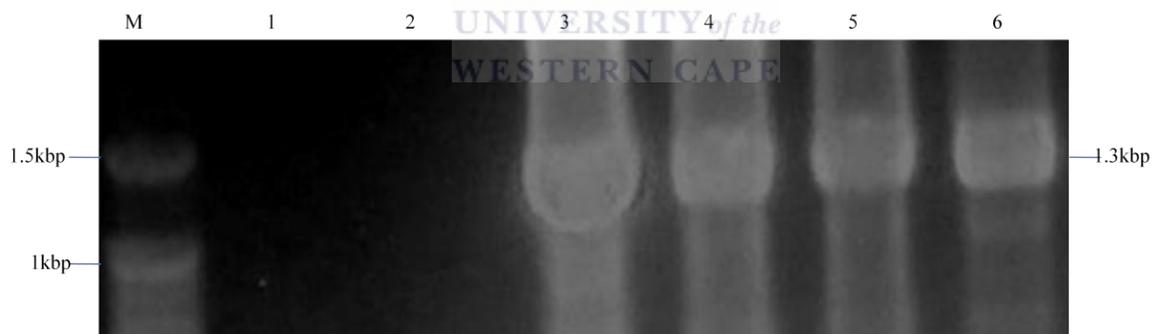
Figure 4.4: Colony PCR confirming the presence of SARS N gene that has been transformed into KRX competent cells. "M" represent the 100 bp DNA ladder. Lane 1 and 2 represent the negative control. Lane 3-6 represents successful transformation of SARS N/pFlexi.

4.4.2. Protein expression and analysis in E. coli cells

Total proteins were successfully expressed and analyzed using SDS-PAGE, Coomassie blue and Western blot. Figure 4.5 represents total proteins that were expressed in a bacterial system. Since bacteria express multiple proteins when they undergo their normal metabolic activities, it is notable from figure 4.5 by the presence of multiple bands. Moreover, the expression profile of the proteins of interests is not distinct in their appearance from the rest of the proteins. All the expressed proteins seem to exhibit the same pattern of expression.

Figure 4.5: Coomassie stain showing the lysate of all expressed proteins. Lane **"M"** represents the protein marker with molecular weight designated along the left margin of the figure. Lane 1-4 represents nucleocapsid proteins in the following order: SARS-N, N1, N2, and N3.

From the Coomassie brilliant blue results, it was noticed that the presence of the nucleocapsid proteins could not be conclusively confirmed. As a result, the cell lysate was subjected to Western blotting. All the nucleocapsid proteins were successfully detected by means of Western blotting and this is evident in figure 4.6 below.
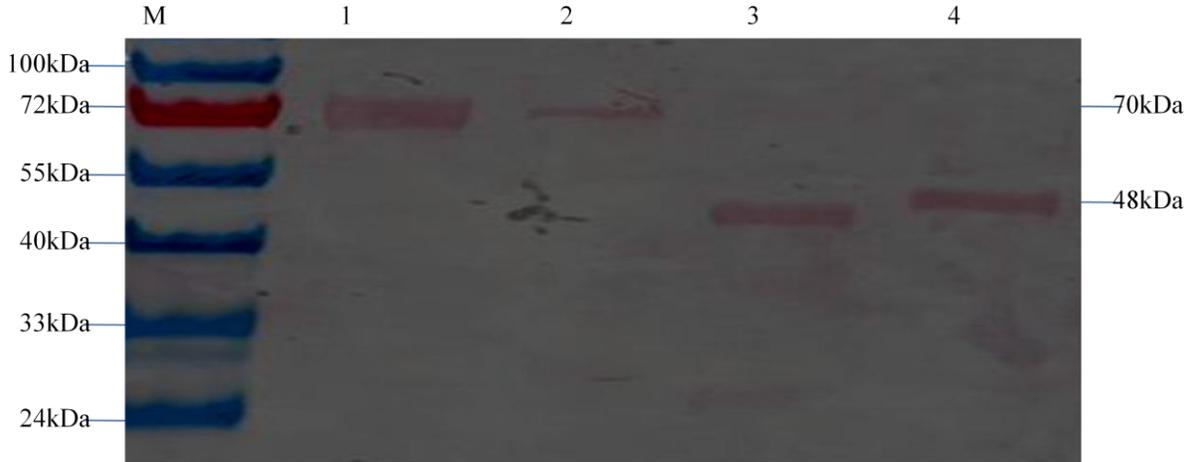
Figure 4.6: Western Blot analysis of nucleocapsid proteins and its trunctaed mutants. On the figure, M represents the prestained protein marker. Lane 1-4 represents detected proteins in the following order: SARS N, N1, N2, and N3.

The dark band in figure 4.6 indicates the GST-tagged proteins that have been detected in nitrocellulose membrane using specific antibodies. From the results it is evident that the technique is highly specific to the GST tag that was fused to the protein of interest.

4.4.3. Purification of the recombinant proteins

The purification of recombinant protein was done using MagneGST protein purification system. Following optimization of protein expression the presence of the nucleocapsid proteins was confirmed by small-scale purification. The fusion proteins were successfully purified in small scale and this is clearly shown in figure 4.7 (A) and (B) below. From the figure below it is clear that the proteins were purified at varying concentrations.
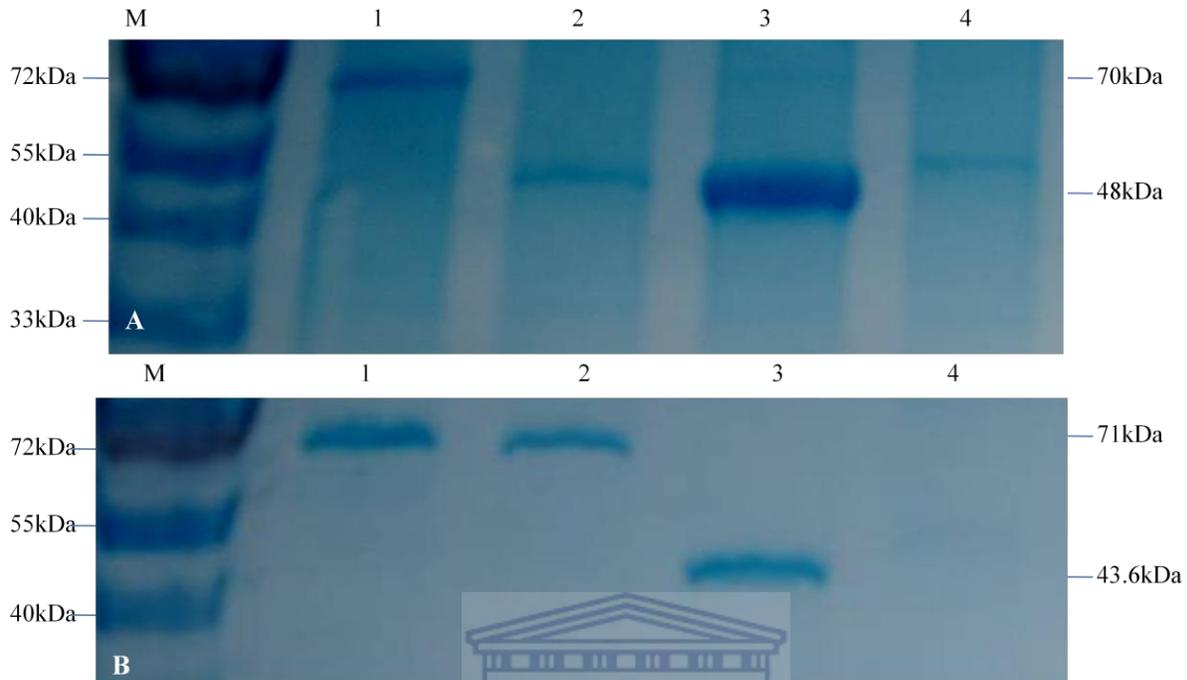
Figure 4.7: Coomassie stain showing small scale purification of nucleocapsid proteins. (A) "M" indicates the protein marker. Lane 1-3 represent purified proteins in the following order: N1, N2, and N3 (B): "M" represents the protein marker. Lane 1-3 represents SARS-N, N1 and N2.

In figure 4.7(A), it is appreciable that N3 is purified at relatively higher concentrations in comparison to N1 and N2, which are indicated in lane 1 and 2, respectively. In figure 4.7(B), on the other hand, there are no significant differences in purification levels as seen from the figure.

Following the successful purification of all proteins of interest, interaction studies between nucleocapsid proteins and CypA were performed by means of GST-pull down assay. From figure 4.8, it is clear that there were noticeable signs of interaction between CypA and nucleocapsid proteins. Instead, an interesting phenomenon was noticed whereby some proteins were either lost or

degraded during the reaction. This is evident in lane 2 and 4 where N1 and N3 disappeared after the reaction. In addition, CypA exhibited the same behavior throughout all reactions.
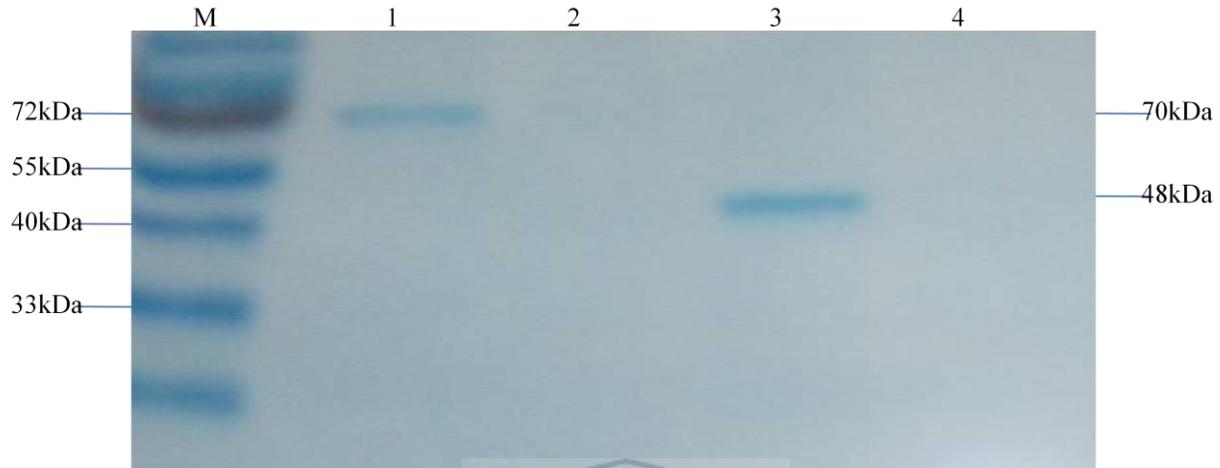
Figure 4.8: Interaction of CypA with nucleocapsid proteins by means of GST-pull down assay. "M" represents the protein marker. Lane 1-4 represents protein interactions in the following order: SARS-N/CypA, N1/CypA, N2/CypA, and N3/CypA.
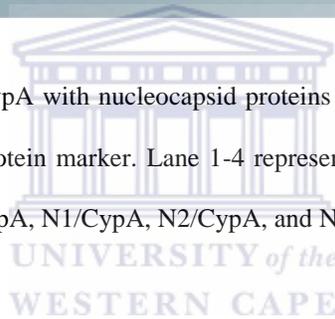
4.5. DISCUSSION AND CONCLUSION

The present study was specifically designed to investigate the interaction of HCoV NL63 nucleocapsid protein and its truncated mutants with human CypA protein. Using an array of molecular techniques, CypA has not been convincingly shown to exhibit a direct and specific binding affinity to the nucleocapsid protein. Furthermore, the study intends to reports on the exact domain that shows direct and specific interaction with human CypA. However, this will only be achieved once the phenomenon that occurred during GST-pull down reaction has been determined.

Despite the observed phenomenon in the GST-pull down reaction, the results of the study clearly indicate that the proteins of interest were successfully extracted from MCF7 cells and viral RNA, amplified by means of RT-PCR and then cloned into a bacterial vector. The latter was verified by genome sequencing, transformed into KRX cells and then expressed in a bacterial system. In addition, CypA protein was expressed in two different tags viz. GST- and HQ. However, it was noticeable that the patterns of expression differ between the two tags hence GST tag was used in this study. The explanation for this variation in expression patterns remains obscure. Secondly, it is apparent from the results that there is a sharp decline in concentration levels between cell lysates and purified protein. This can be attributed to digestion of expressed protein by cell proteases merely because CypA is a human protein. Therefore, in a bacterial system it is recognized as foreign. The nucleocapsid protein, on the other hand, is a viral protein and is known to be antigenic and will thus be destroyed by proteases immediately.

The nucleocapsid protein is known to bind to viral RNA and interfere with replication and even transcription. CypA, on the other hand, has been shown to be secreted in cells in response to inflammatory stimuli and is a potent neutrophil and eosinophil chemo-attractant both *in vitro* and *in vivo* (Damsker, Bukrinsky et al. 2007). The signaling activity is via CD147 cell surface receptors, but the mechanism underlying CypA-mediated signaling and chemotaxis remain obscure. Therefore, it is suffice to say that the cross talk between CypA and NL63 nucleocapsid protein might play a significant role in the pathogenesis of HCoV NL63. In addition to the above mentioned physiological attributes, there is an observed pattern of CypA interaction with an affinity for viral capsid proteins. This has been reported for HIV-1 Vpr protein as well as NS5A protein of HCV (Zander, Sherman et al. 2003; Chatterji, Lim et al. 2010).

Since CypA is engaged in pro-inflammatory reactions, its interaction with nucleocapsid protein may lead to a reduced immunological response, paving the way for the virions to replicate undisturbed and continuously. In addition to its obscure role in immunological reaction, the observed pattern of CypA interaction with capsid proteins may provide clues to its exact role in a myriad of viral infections. It may be possible that CypA has a stereochemical active site for capsid proteins enabling it to interact with such proteins. Therefore, binding of CypA to capsid proteins may result in conformational changes in the structure of CypA, which will then render it inactive from performing its normal physiological pro-inflammatory role. Therefore, determination of this site and its conformation may provide clues to the development of therapeutic strategies.

In conclusion, the interaction of CypA with nucleocapsid protein needs to be determined. Most importantly, the stereochemically active site of CypA with capsid proteins also needs to be determined. Given the normal physiological functions of CypA, HCoV NL63 could be evading the immune response via suppression of inflammatory mediators thus leading to disease progression. Nonetheless, there is still a need to fully comprehend other functional significances of these interactions in the entire pathophysiological process of HCoV NL63. Hopefully, once this phenomenon between the two proteins has been ascertained, this will shed light in order to fully comprehend the infectious process and also provide hints for better understanding of a possible infection pathway.

CHAPTER 5

Summary

Summary

The history of coronaviruses began in 1959 and this group of viruses was officially accepted as a genus in 1960 (Kapikian 1975). The human and animal coronaviruses were segregated into three groups based on their antigenic and genetic make-up. The human coronavirus infection was usually not associated with severe diseases, but following the severe acute respiratory syndrome (SARS) epidemic in 2002 it became evident that highly pathogenic strains can still evolve. Most recently, Lia van der Hoek et al. (2004) reported the discovery of human coronavirus (HCoV) NL63 from a 7 month old girl with coryza, conjunctivitis, fever, and bronchilitis. The pathophysiology of this novel virus still remains to be determined. Some scientists believe that it is through the interaction of viral structural proteins with host cell proteins. Amongst these structural proteins, the nucleocapsid (N) protein has been reported to be highly antigenic and is associated with several host cell interactions (Zhou, Liu et al. 2008). Luo et al (2004) has shown SARS-N to interact with human cyclophilin A (CypA). The latter is a peptidyl-prolyl *cis/trans* isomerase (PPIA) that is involved in cellular signaling events. It has also been shown to be secreted in cells in response to inflammatory stimuli and is a potent neutrophil and eosinophil chemo-attractant (Yang, Lu et al. 2008). Most importantly, CypA has been demonstrated to play a critical role in pathogenesis of viruses such as human immune virus (HIV)-1, hepatitis C virus (HCV), and SARS. Therefore, determination of host and viral protein interactions is crucial for elucidation of strategies for therapeutic intervention. Using an array of molecular and computational techniques, these studies aimed at determing these interactions and most importantly identify the exact nucleocapsid protein domains that are responsible.

Initially, bioinformatic tools were applied in order to predict the function HCoV NL63-N gene and protein. SARS-N is known to interact with CypA, therefore comparing its properties with HCoV NL63-N will enable one to infer the unknown function NL63-N protein. This study has shown that HCoV NL63-N exhibits some degree of homology to SARS-N. However, these proteins exhibit slightly different characteristics in terms of their post-translational modifications. From this evidence of evolutionary relationship between the two strains, it is reasonable to assume that NL63-N protein will interact with CypA.

Following the prediction of NL63-N function, CypA was cloned and expressed in a bacterial system for interaction with HCoV NL63-N protein. Initially, CypA gene was extracted from MCF7 cells and then amplified using reverse transcription polymerase chain reaction (RT-PCR) into multiple genes. The latter was cloned into a sequencing vector and later shuttled into a bacterial expression vector following the confirmation of the correct protein coding region. The cloned vector was transformed into KRX cells and protein expression induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). The protein of interest was detected from the cell lysate by means of Western blotting followed by purification using MagneGST protein purification system.

The nucleocapsid was also cloned and expressed in a bacterial system. Following expression, the proteins were detected and then purified under the conditions described for CypA. Interaction studies between CypA and N proteins were conducted using the GST-pull down assay. However, these interactions could not be convincingly demonstrated to be direct and specific. Therefore, the interaction conditions need to be optimized. Initially, the protein concentrations need to be increased into desirable levels, so as to increase the chances of interaction. Lastly, the stringency of the buffers needs to be optimized so that they do not interfere with protein activity.

REFERENCES

Almazan, F., C. Galan, et al. (2004). "The nucleoprotein is required for efficient coronavirus genome replication." <u>J Virol</u> **78**(22): 12683-12688.

Bai, B., Q. Hu, et al. (2008). "Virus-like particles of SARS-like coronavirus formed by membrane proteins from different origins demonstrate stimulating activity in human dendritic cells." <u>PLoS One</u> **3**(7): e2685.

Bartlam, M., Y. Xu, et al. (2007). "Structural proteomics of the SARS coronavirus: a model response to emerging infectious diseases." <u>J Struct Funct Genomics</u> **8**(2-3): 85-97.

Buchholz, U. J., A. Bukreyev, et al. (2004). "Contributions of the structural proteins of severe acute respiratory syndrome coronavirus to protective immunity." <u>Proc Natl Acad Sci U S A</u> **101**(26): 9804-9809.

Bynoe, M. L., D. Hobson, et al. (1961). "Inoculation of human volunteers with a strain of virus isolated from a common cold." <u>Lancet</u> **1**(7188): 1194-1196.

Cai, C. Z., L. Y. Han, et al. (2005). "Prediction of functional class of the SARS coronavirus proteins by a statistical learning method." <u>J Proteome Res</u> **4**(5): 1855-1862.

Calvo, E., D. Escors, et al. (2005). "Phosphorylation and subcellular localization of transmissible gastroenteritis virus nucleocapsid protein in infected cells." <u>J Gen Virol</u> **86**(Pt 8): 2255-2267.

Campanacci, V., M. P. Egloff, et al. (2003). "Structural genomics of the SARS coronavirus: cloning, expression, crystallization and preliminary crystallographic study of the Nsp9 protein." <u>Acta Crystallogr D Biol Crystallogr</u> **59**(Pt 9): 1628-1631.

Cavanagh, D. (2003). "Severe acute respiratory syndrome vaccine development: experiences of vaccination against avian infectious bronchitis coronavirus." Avian Pathol **32**(6): 567-582.

Chan, C. M., C. W. Ma, et al. (2007). "The SARS-Coronavirus Membrane protein induces apoptosis through modulating the Akt survival pathway." Arch Biochem Biophys **459**(2): 197-207.

Chan, C. M., H. Tse, et al. (2009). "Examination of seroprevalence of coronavirus HKU1 infection with S protein-based ELISA and neutralization assay against viral spike pseudotyped virus." J Clin Virol **45**(1): 54-60.

Chan, C. M., P. C. Woo, et al. (2008). "Spike protein, S, of human coronavirus HKU1: role in viral life cycle and application in antibody detection." Exp Biol Med (Maywood) **233**(12): 1527-1536.

Chatterji, U., P. Lim, et al. (2010). "HCV resistance to cyclosporin A does not correlate with a resistance of the NS5A-cyclophilin A interaction to cyclophilin inhibitors." J Hepatol **53**(1): 50-56.

Chen, S., M. Zhang, et al. (2008). "Oligo-microarray analysis reveals the role of cyclophilin A in drug resistance." Cancer Chemother Pharmacol **61**(3): 459-469.

Chen, Z., Y. Wang, et al. (2007). "Proteolytic processing and deubiquitinating activity of papain-like proteases of human coronavirus NL63." J Virol **81**(11): 6007-6018.

Damsker, J. M., M. I. Bukrinsky, et al. (2007). "Preferential chemotaxis of activated human CD4+ T cells by extracellular cyclophilin A." J Leukoc Biol **82**(3): 613-618.

Daum, S., M. Schumann, et al. (2009). "Isoform-specific inhibition of cyclophilins." Biochemistry **48**(26): 6268-6277.

de Haan, C. A., B. J. Haijema, et al. (2008). "Cleavage of group 1 coronavirus spike proteins: how furin cleavage is traded off against heparan sulfate binding upon cell culture adaptation." J Virol **82**(12): 6078-6083.

de Wilde, A. H., J. C. Zevenhoven-Dobbe, et al. (2011). "Cyclosporin A inhibits the replication of diverse coronaviruses." J Gen Virol.

Dijkman, R., M. F. Jebbink, et al. (2008). "Human coronavirus NL63 and 229E seroconversion in children." J Clin Microbiol **46**(7): 2368-2373.

Dijkman, R., M. F. Jebbink, et al. (2006). "Human coronavirus 229E encodes a single ORF4 protein between the spike and the envelope genes." Virol J **3**: 106.

Emmott, E., B. K. Dove, et al. (2008). "Viral nucleolar localisation signals determine dynamic trafficking within the nucleolus." Virology **380**(2): 191-202.

Emmott, E., M. A. Rodgers, et al. (2010). "Quantitative proteomics using stable isotope labeling with amino acids in cell culture reveals changes in the cytoplasmic, nuclear, and nucleolar proteomes in Vero cells infected with the coronavirus infectious bronchitis virus." Mol Cell Proteomics **9**(9): 1920-1936.

Fang, X., J. Gao, et al. (2007). "The membrane protein of SARS-CoV suppresses NF-kappaB activation." J Med Virol **79**(10): 1431-1439.

Feng, Y. and G. F. Gao (2007). "Towards our understanding of SARS-CoV, an emerging and devastating but quickly conquered virus." Comp Immunol Microbiol Infect Dis **30**(5-6): 309-327.

Fischer, G., P. Gallay, et al. (2010). "Cyclophilin inhibitors for the treatment of HCV infection." Curr Opin Investig Drugs **11**(8): 911-918.

Frieman, M., K. Ratia, et al. (2009). "Severe acute respiratory syndrome coronavirus papain-like protease ubiquitin-like domain and catalytic domain regulate antagonism of IRF3 and NF-kappaB signaling." J Virol **83**(13): 6689-6705.

Frieman, M., B. Yount, et al. (2007). "Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane." J Virol **81**(18): 9812-9824.

Galat, A. and J. Bua (2010). "Molecular aspects of cyclophilins mediating therapeutic actions of their ligands." Cell Mol Life Sci **67**(20): 3467-3488.

Gitti, R. K., B. M. Lee, et al. (1996). "Structure of the amino-terminal core domain of the HIV-1 capsid protein." Science **273**(5272): 231-235.

Graham, R. L. and R. S. Baric (2010). "Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission." J Virol **84**(7): 3134-3146.

Han, D. P., M. Lohani, et al. (2007). "Specific asparagine-linked glycosylation sites are critical for DC-SIGN- and L-SIGN-mediated severe acute respiratory syndrome coronavirus entry." J Virol **81**(21): 12029-12039.

Han, X., S. H. Yoon, et al. (2010). "Cyclosporin A and sanglifehrin A enhance chemotherapeutic effect of cisplatin in C6 glioma cells." Oncol Rep **23**(4): 1053-1062.

Hao, P., M. Chen, et al. (2006). "Bioinformatics research on the SARS coronavirus (SARS_CoV) in China." Curr Pharm Des **12**(35): 4565-4572.

Hatziioannou, T., D. Perez-Caballero, et al. (2005). "Cyclophilin interactions with incoming human immunodeficiency virus type 1 capsids with opposing effects on infectivity in human cells." J Virol **79**(1): 176-183.

He, Q., Q. Du, et al. (2005). "Characterization of monoclonal antibody against SARS coronavirus nucleocapsid antigen and development of an antigen capture ELISA." J Virol Methods **127**(1): 46-53.

He, R., A. Leeson, et al. (2004). "Characterization of protein-protein interactions between the nucleocapsid protein and membrane protein of the SARS coronavirus." Virus Res **105**(2): 121-125.

Hofmann, H., G. Simmons, et al. (2006). "Highly conserved regions within the spike proteins of human coronaviruses 229E and NL63 determine recognition of their respective cellular receptors." J Virol **80**(17): 8639-8652.

Hogue, B. G. and D. A. Brian (1986). "Structural proteins of human respiratory coronavirus OC43." Virus Res **5**(2-3): 131-144.

Huang, I. C., B. J. Bosch, et al. (2006). "SARS-CoV, but not HCoV-NL63, utilizes cathepsins to infect cells: viral entry." Adv Exp Med Biol **581**: 335-338.

Jackwood, M. W., T. O. Boynton, et al. (2010). "Emergence of a group 3 coronavirus through recombination." Virology **398**(1): 98-108.

Javanbakht, H., F. Diaz-Griffero, et al. (2007). "The ability of multimerized cyclophilin A to restrict retrovirus infection." Virology **367**(1): 19-29.

Junwei, G., L. Baoxian, et al. (2006). "Cloning and sequence analysis of the N gene of porcine epidemic diarrhea virus LJB/03." Virus Genes **33**(2): 215-219.

Kam, Y. W., Y. Okumura, et al. (2009). "Cleavage of the SARS coronavirus spike glycoprotein by airway proteases enhances virus entry into human bronchial epithelial cells in vitro." PLoS One **4**(11): e7870.

Kang, B. H., J. Plescia, et al. (2007). "Regulation of tumor cell mitochondrial homeostasis by an organelle-specific Hsp90 chaperone network." Cell **131**(2): 257-270.

Kang, X., B. A. Yang, et al. (2006). "Human neutralizing Fab molecules against severe acute respiratory syndrome coronavirus generated by phage display." Clin Vaccine Immunol **13**(8): 953-957.

Kapikian, A. Z. (1975). "The coronaviruses." Dev Biol Stand **28**: 42-64.

Keckesova, Z., L. M. Ylinen, et al. (2006). "Cyclophilin A renders human immunodeficiency virus type 1 sensitive to Old World monkey but not human TRIM5 alpha antiviral activity." J Virol **80**(10): 4683-4690.

Kennedy, D. A. and C. M. Johnson-Lussenburg (1975). "Isolation and morphology of the internal component of human coronavirus, strain 229E." Intervirology **6**(4-5): 197-206.

Khan, M., M. Garcia-Barrio, et al. (2003). "Treatment of human immunodeficiency virus type 1 virions depleted of cyclophilin A by natural endogenous reverse transcription restores infectivity." J Virol **77**(7): 4431-4434.

Kim, O. J., D. H. Lee, et al. (2006). "Close relationship between SARS-coronavirus and group 2 coronavirus." J Microbiol **44**(1): 83-91.

Klumperman, J., J. K. Locker, et al. (1994). "Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding." J Virol **68**(10): 6523-6534.

Ko, C. K., M. I. Kang, et al. (2006). "Molecular characterization of HE, M, and E genes of winter dysentery bovine coronavirus circulated in Korea during 2002-2003." Virus Genes **32**(2): 129-136.

Lau, S. K., P. C. Woo, et al. (2004). "Detection of severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein in sars patients by enzyme-linked immunosorbent assay." J Clin Microbiol **42**(7): 2884-2889.

Lee, J. (2010). "Role of cyclophilin a during oncogenesis." Arch Pharm Res **33**(2): 181-187.

Li, J., S. Tang, et al. (2007). "HIV-1 capsid protein and cyclophilin a as new targets for anti-AIDS therapeutic agents." Infect Disord Drug Targets **7**(3): 238-244.

Li, S. W., C. C. Lai, et al. (2011). "Severe acute respiratory syndrome coronavirus papain-like protease suppressed alpha interferon-induced responses through downregulation of extracellular signal-regulated kinase 1-mediated signalling pathways." J Gen Virol **92**(Pt 5): 1127-1140.

Lin, H. X., Y. Feng, et al. (2011). "Characterization of the spike protein of human coronavirus NL63 in receptor binding and pseudotype virus entry." Virus Res **160**(1-2): 283-293.

Lin, H. X., Y. Feng, et al. (2008). "Identification of residues in the receptor-binding domain (RBD) of the spike protein of human coronavirus NL63 that are critical for the RBD-ACE2 receptor interaction." J Gen Virol **89**(Pt 4): 1015-1024.

Liu, J., Y. Sun, et al. (2010). "The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes." J Infect Dis **202**(8): 1171-1180.

Liu, M., Y. Yang, et al. (2007). "Spike protein of SARS-CoV stimulates cyclooxygenase-2 expression via both calcium-dependent and calcium-independent protein kinase C pathways." FASEB J **21**(7): 1586-1596.

Lu, H., Y. Zhao, et al. (2004). "Date of origin of the SARS coronavirus strains." BMC Infect Dis **4**: 3.

Lu, X., J. Pan, et al. (2011). "SARS-CoV nucleocapsid protein antagonizes IFN-beta response by targeting initial step of IFN-beta induction pathway, and its C-terminal region is critical for the antagonism." Virus Genes **42**(1): 37-45.

Luo, C., H. Luo, et al. (2004). "Nucleocapsid protein of SARS coronavirus tightly binds to human cyclophilin A." Biochem Biophys Res Commun **321**(3): 557-565.

Luo, H., Q. Chen, et al. (2005). "The nucleocapsid protein of SARS coronavirus has a high binding affinity to the human cellular heterogeneous nuclear ribonucleoprotein A1." FEBS Lett **579**(12): 2623-2628.

Mascarenhas, A. P. and K. Musier-Forsyth (2009). "The capsid protein of human immunodeficiency virus: interactions of HIV-1 capsid with host protein factors." FEBS J **276**(21): 6118-6127.

Mills, J. S. (2006). "Peptides derived from HIV-1, HIV-2, Ebola virus, SARS coronavirus and coronavirus 229E exhibit high affinity binding to the formyl peptide receptor." Biochim Biophys Acta **1762**(7): 693-703.

Mizutani, T., S. Fukushi, et al. (2004). "Phosphorylation of p38 MAPK and its downstream targets in SARS coronavirus-infected cells." Biochem Biophys Res Commun **319**(4): 1228-1234.

Muller, M. A., L. van der Hoek, et al. (2010). "Human coronavirus NL63 open reading frame 3 encodes a virion-incorporated N-glycosylated membrane protein." Virol J **7**: 6.

Nedialkova, D. D., R. Ulferts, et al. (2009). "Biochemical characterization of arterivirus nonstructural protein 11 reveals the nidovirus-wide conservation of a replicative endoribonuclease." <u>J Virol</u> **83**(11): 5671-5682.

Negi, S. S. and W. Braun (2009). "Automated detection of conformational epitopes using phage display Peptide sequences." <u>Bioinform Biol Insights</u> **3**: 71-81.

Neuman, B. W., J. S. Joseph, et al. (2008). "Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3." <u>J Virol</u> **82**(11): 5279-5294.

Noser, J. A., G. J. Towers, et al. (2006). "Cyclosporine increases human immunodeficiency virus type 1 vector transduction of primary mouse cells." <u>J Virol</u> **80**(15): 7769-7774.

Patrick, D. M., M. Petric, et al. (2006). "An Outbreak of Human Coronavirus OC43 Infection and Serological Cross-reactivity with SARS Coronavirus." <u>Can J Infect Dis Med Microbiol</u> **17**(6): 330-336.

Qi, M., R. Yang, et al. (2008). "Cyclophilin A-dependent restriction of human immunodeficiency virus type 1 capsid mutants for infection of nondividing cells." <u>J Virol</u> **82**(24): 12001-12008.

Ren, W., W. Li, et al. (2006). "Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis." <u>J Gen Virol</u> **87**(Pt 11): 3355-3359.

Saphire, A. C., M. D. Bobardt, et al. (1999). "Host cyclophilin A mediates HIV-1 attachment to target cells via heparans." <u>EMBO J</u> **18**(23): 6771-6785.

Schaecher, S. R., M. S. Diamond, et al. (2008). "The transmembrane domain of the severe acute respiratory syndrome coronavirus ORF7b protein is necessary and sufficient for its retention in the Golgi complex." <u>J Virol</u> **82**(19): 9477-9491.

Schickli, J. H., B. D. Zelus, et al. (1997). "The murine coronavirus mouse hepatitis virus strain A59 from persistently infected murine cells exhibits an extended host range." J Virol **71**(12): 9499-9507.

Schreiber, S. S., T. Kamahora, et al. (1989). "Sequence analysis of the nucleocapsid protein gene of human coronavirus 229E." Virology **169**(1): 142-151.

See, R. H., M. Petric, et al. (2008). "Severe acute respiratory syndrome vaccine efficacy in ferrets: whole killed virus and adenovirus-vectored vaccines." J Gen Virol **89**(Pt 9): 2136-2146.

Shang, L., Y. Qi, et al. (2006). "Polymorphism of SARS-CoV genomes." Yi Chuan Xue Bao **33**(4): 354-364.

Shao, X., X. Guo, et al. (2007). "Seroepidemiology of group I human coronaviruses in children." J Clin Virol **40**(3): 207-213.

Shih, Y. P., C. Y. Chen, et al. (2006). "Identifying epitopes responsible for neutralizing antibody and DC-SIGN binding on the spike glycoprotein of the severe acute respiratory syndrome coronavirus." J Virol **80**(21): 10315-10324.

Shulla, A., T. Heald-Sargent, et al. (2011). "A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry." J Virol **85**(2): 873-882.

Solbak, S. M., T. R. Reksten, et al. (2010). "The intriguing cyclophilin A-HIV-1 Vpr interaction: prolyl cis/trans isomerisation catalysis and specific binding." BMC Struct Biol **10**: 31.

St-Jean, J. R., H. Jacomy, et al. (2004). "Human respiratory coronavirus OC43: genetic stability and neuroinvasion." J Virol **78**(16): 8824-8834.

Stadler, K., V. Masignani, et al. (2003). "SARS--beginning to understand a new virus." Nat Rev Microbiol **1**(3): 209-218.

Steinkasserer, A., R. Harrison, et al. (1995). "Mode of action of SDZ NIM 811, a nonimmunosuppressive cyclosporin A analog with activity against human immunodeficiency virus type 1 (HIV-1): interference with early and late events in HIV-1 replication." J Virol **69**(2): 814-824.

Stohlman, S. A., J. O. Fleming, et al. (1983). "Synthesis and subcellular localization of the murine coronavirus nucleocapsid protein." Virology **130**(2): 527-532.

Surjit, M., B. Liu, et al. (2006). "The nucleocapsid protein of severe acute respiratory syndrome-coronavirus inhibits the activity of cyclin-cyclin-dependent kinase complex and blocks S phase progression in mammalian cells." J Biol Chem **281**(16): 10669-10681.

Takeuchi, H. (2010). "Contribution of Cyclophilin A to determination of simian immunodeficiency virus tropism: a progress update." Vaccine **28 Suppl 2**: B51-54.

Tang, T. K., M. P. Wu, et al. (2005). "Biochemical and immunological studies of nucleocapsid proteins of severe acute respiratory syndrome and 229E human coronaviruses." Proteomics **5**(4): 925-937.

Thai, V., P. Renesto, et al. (2008). "Structural, biochemical, and in vivo characterization of the first virally encoded cyclophilin from the Mimivirus." J Mol Biol **378**(1): 71-86.

Thali, M., A. Bukovsky, et al. (1994). "Functional association of cyclophilin A with HIV-1 virions." Nature **372**(6504): 363-365.

Tian, X., C. Zhao, et al. (2010). "Hepatitis B virus (HBV) surface antigen interacts with and promotes cyclophilin a secretion: possible link to pathogenesis of HBV infection." J Virol **84**(7): 3373-3381.

Timani, K. A., Q. Liao, et al. (2005). "Nuclear/nucleolar localization properties of C-terminal nucleocapsid protein of SARS coronavirus." <u>Virus Res</u> **114**(1-2): 23-34.

Tyrrell, D. A. and M. L. Bynoe (1958). "Inoculation of volunteers with J.H. strain of new respiratory virus." <u>Lancet</u> **2**(7053): 931-933.

Tyrrell, D. A., M. L. Bynoe, et al. (1960). "Some virus isolations from common colds. I. Experiments employing human volunteers." <u>Lancet</u> **1**(7118): 235-237.

Tyrrell, D. A., M. L. Bynoe, et al. (1959). "Inoculation of human volunteers with parainfluenza viruses types 1 and 3 (HA 2 and HA 1)." <u>Br Med J</u> **2**(5157): 909-911.

van Kampen, A. H., B. D. van Schaik, et al. (2000). "USAGE: a web-based approach towards the analysis of SAGE data. Serial Analysis of Gene Expression." <u>Bioinformatics</u> **16**(10): 899-905.

Versteeg, G. A., P. S. van de Nes, et al. (2007). "The coronavirus spike protein induces endoplasmic reticulum stress and upregulation of intracellular chemokine mRNA concentrations." <u>J Virol</u> **81**(20): 10981-10990.

Wang, J., J. Ji, et al. (2003). "The structure analysis and antigenicity study of the N protein of SARS-CoV." <u>Genomics Proteomics Bioinformatics</u> **1**(2): 145-154.

Wang, P. and J. Heitman (2005). "The cyclophilins." <u>Genome Biol</u> **6**(7): 226.

Wang, Z. G., L. J. Li, et al. (2004). "Molecular biological analysis of genotyping and phylogeny of severe acute respiratory syndrome associated coronavirus." <u>Chin Med J (Engl)</u> **117**(1): 42-48.

Watashi, K. and K. Shimotohno (2007). "Cyclophilin and viruses: cyclophilin as a cofactor for viral infection and possible anti-viral target." <u>Drug Target Insights</u> **2**: 9-18.

Wesseling, J. G., G. J. Godeke, et al. (1993). "Mouse hepatitis virus spike and nucleocapsid proteins expressed by adenovirus vectors protect mice against a lethal infection." J Gen Virol **74 (Pt 10)**: 2061-2069.

Wong, S. S. and K. Y. Yuen (2005). "The severe acute respiratory syndrome (SARS)." J Neurovirol **11**(5): 455-468.

Woo, P. C., Y. Huang, et al. (2005). "In silico analysis of ORF1ab in coronavirus HKU1 genome reveals a unique putative cleavage site of coronavirus HKU1 3C-like protease." Microbiol Immunol **49**(10): 899-908.

Woo, P. C., S. K. Lau, et al. (2005). "Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia." J Virol **79**(2): 884-895.

Woo, P. C., S. K. Lau, et al. (2009). "Coronavirus diversity, phylogeny and interspecies jumping." Exp Biol Med (Maywood) **234**(10): 1117-1127.

Yan, K., W. Tan, et al. (2009). "SARS-CoV spike proteins expressed by the vaccinia virus Tiantan strain: secreted sq protein induces robust neutralization antibody in mice." Viral Immunol **22**(1): 57-66.

Yang, Y., N. Lu, et al. (2008). "Cyclophilin A up-regulates MMP-9 expression and adhesion of monocytes/macrophages via CD147 signalling pathway in rheumatoid arthritis." Rheumatology (Oxford) **47**(9): 1299-1310.

Ye, Y., K. Hauns, et al. (2007). "Mouse hepatitis coronavirus A59 nucleocapsid protein is a type I interferon antagonist." J Virol **81**(6): 2554-2563.

Yeung, K. S., G. A. Yamanaka, et al. (2006). "Severe acute respiratory syndrome coronavirus entry into host cells: Opportunities for therapeutic intervention." Med Res Rev **26**(4): 414-433.

Ying, W., Y. Hao, et al. (2004). "Proteomic analysis on structural proteins of Severe Acute Respiratory Syndrome coronavirus." Proteomics **4**(2): 492-504.

You, J., B. K. Dove, et al. (2005). "Subcellular localization of the severe acute respiratory syndrome coronavirus nucleocapsid protein." J Gen Virol **86**(Pt 12): 3303-3310.

You, J. H., M. L. Reed, et al. (2007). "Trafficking motifs in the SARS-coronavirus nucleocapsid protein." Biochem Biophys Res Commun **358**(4): 1015-1020.

Yuan, X., Z. Yao, et al. (2005). "Nucleolar localization of non-structural protein 3b, a protein specifically encoded by the severe acute respiratory syndrome coronavirus." Virus Res **114**(1-2): 70-79.

Zakhartchouk, A. N., S. Viswanathan, et al. (2005). "Severe acute respiratory syndrome coronavirus nucleocapsid protein expressed by an adenovirus vector is phosphorylated and immunogenic in mice." J Gen Virol **86**(Pt 1): 211-215.

Zander, K., M. P. Sherman, et al. (2003). "Cyclophilin A interacts with HIV-1 Vpr and is required for its functional expression." J Biol Chem **278**(44): 43202-43213.

Zeng, R., H. Q. Ruan, et al. (2004). "Proteomic analysis of SARS associated coronavirus using two-dimensional liquid chromatography mass spectrometry and one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis followed by mass spectroemtric analysis." J Proteome Res **3**(3): 549-555.

Zeng, Y., L. Ye, et al. (2008). "The nucleocapsid protein of SARS-associated coronavirus inhibits B23 phosphorylation." Biochem Biophys Res Commun **369**(2): 287-291.

Zhang, J., J. S. Guy, et al. (2007). "Genomic characterization of equine coronavirus." Virology **369**(1): 92-104.

Zhang, W. G., J. Q. Li, et al. (2003). "[Genomic characterization of SARS coronavirus: a novel member of coronavirus]." Yi Chuan Xue Bao **30**(6): 501-508.

Zheng, J., J. E. Koblinski, et al. (2008). "Prolyl isomerase cyclophilin A regulation of Janus-activated kinase 2 and the progression of human breast cancer." Cancer Res **68**(19): 7769-7778.

Zhou, B., J. Liu, et al. (2008). "The nucleocapsid protein of severe acute respiratory syndrome coronavirus inhibits cell cytokinesis and proliferation by interacting with translation elongation factor 1alpha." J Virol **82**(14): 6962-6971.

Zuniga, S., I. Sola, et al. (2004). "Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis." J Virol **78**(2): 980-994.

UNIVERSITY of the
WESTERN CAPE
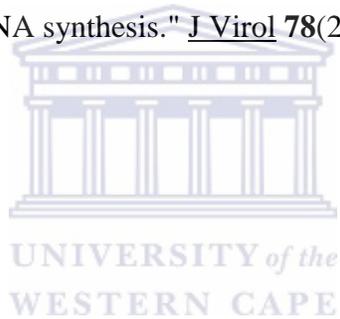
APPENDIX 1

## *Summary of procedures used*

Polymerase Chain Reaction

This reaction was carried out in using the cDNA library obtained using reverse transcription technique. The purpose of this reaction is to make multiple copies of a particular gene. This is essential in order to acquire enough starting template for subsequent experimental procedures. There are 3 major steps in a polymerase chain reaction (PCR): denaturation, annealing, and extension. These are repeated for 30 or 40 cycles in an automated thermocycler.

The initial phase involves melting of the double stranded (ds) DNA into single stranded (ss) DNA followed by annealing of primers and formation of stable ionic bonds between primers and DNA. The reaction is completed with attachment of *Taq* polymerase on the template and the extension of the gene of interest.

Cloning into vectors

DNA cloning is a technique for reproducing DNA fragments. It allows a copy of any specific part of a DNA sequence to be selected among many others and produced in an unlimited amount. The cloning strategy involves 3 steps. The first step is digestion of DNA sample with restriction enzyme followed by digestion of DNA plasmid vector with the same restriction enzyme. Lastly, the DNA sample is ligated into the plasmid vector.
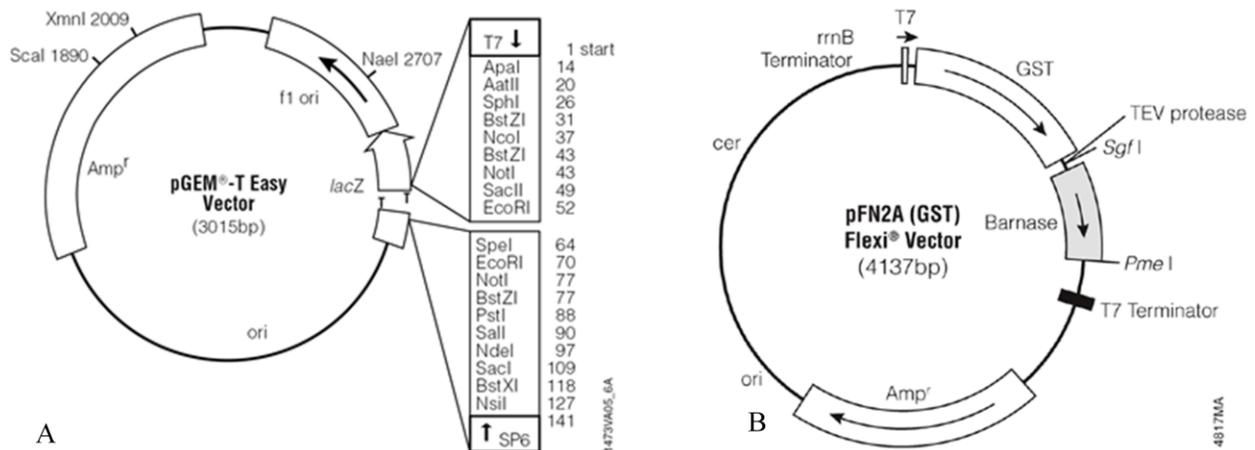
This diagram represents the vectors that were used to clone the genes of interest. The figure marked "A" indicates PGEM-T Easy cloning vector. The latter was chosen because of its convenience for cloning PCR products. The vector provides several options for the removal of the DNA insert with a single restriction enzyme. In this study, *EcoR1* was chosen as the restriction enzyme. The figure marked "B", on the other hand, represents the GST tagged bacterial expression vector that is appended by *Sgf1* and *Pme1* restriction sites. The T7 promoter upstream controls the expression of the DNA insert.

Transformation into competent cells

Transformation is used to transfer exogenous DNA into cells. It is one of the three processes by which exogenous genetic material may be introduced into a bacterial cell. The method involves incubation of competent cells with the DNA on ice and then briefly heat shock ($42^{o}C$ for 45 sec.) thus allowing the DNA to enter the cells. Because the plasmid vector contains an origin of replication, it can replicate independent of the cell. The selection of transformed cells is performed on agar plates supplemented with an antibiotic together with X-gal. This technique is based on the

following principle. In transformed cells, the lacZ gene found on the plasmid vector is disrupted because of the insertion of the gene of interest within the lacZ code. As a result, the plasmid is unable to express lacZ-alpha subunit. Conversely, in untransformed cells the lacZ gene is still intact; hence there is expression of lacZ-alpha subunit and production of a functional ß-galactosidase. Essentially, the white appearance of transformed cells on agar plates is due to the inability to express ß-galactosidase. The blue color of untransformed cells, on the other hand, is due to the action of ß-galactosidase on X-gal that is supplemented on the agar plate.

Protein expression

This technique is used to translate recombinant protein by manipulating the bacterial system. The target DNA that is inserted onto the plasmid vector is under the control of T7 promoter. The transcription of the lac operon is triggered by isopropyl β-D-1-thiogalactopyranoside (IPTG).
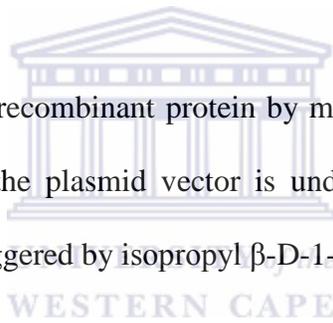
SDS PAGE

The sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) technique is used to separate proteins according to their electrophoretic ability, posttranslational modifications and other factors. The protein complexes are separated according to size by means of SDS PAGE. The procedure involves mixing of the protein complex to be analyzed with SDS. The latter is an anionic detergent which denatures secondary and non-disulfide-linked tertiary proteins, and applies a negative charge to each protein in proportion to its mass. This is important because different proteins with similar molecular weight would migrate differently due to differences in mass charge.

An illustration of an apparatus used for SDS PAGE.
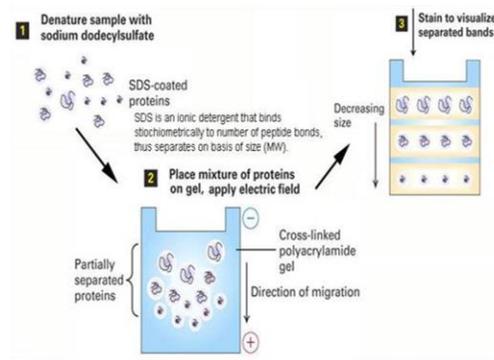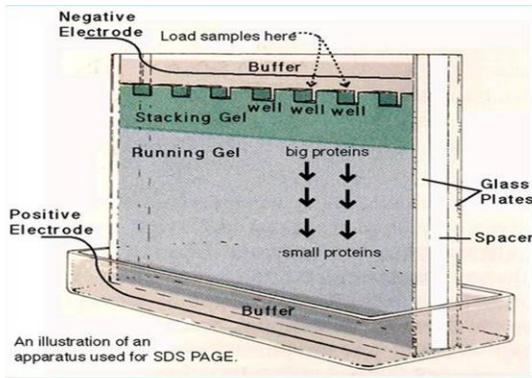
This represents the apparatus used in SDS PAGE analysis as shown on the left hand side. The illustration on the right indicates the principle of the technique. Total proteins are separated according to their electrophoretic ability into different molecular weight.

The following parameters should be kept in check when performing electrophoresis: strength of the electric field, viscosity and temperature of the medium in which the molecules are moving. These are crucial as they influence the rate of migration of charged molecules in a solution in response to electric field.

Western blot

The Western blot or immunoblot is used to detect the specific protein of interest. This analytical technique involves the immobilization of protein on membrane nitrocellulose, and then followed by detection using monoclonal or polyclonal antibody. Western blot detects the specific protein,

whereas SDS-PAGE provides information about the molecular weight and the potential existence of different isoforms of the protein under study.
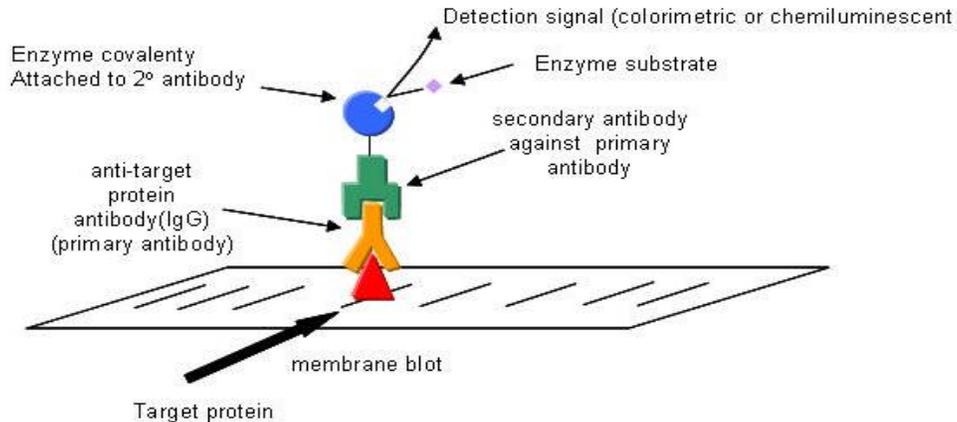
The protein of interest is indicated with the red triangle on nitrocellulose membrane. This protein is initially bound to a primary which in turn binds to a secondary antibody that is covalently attached to an enzyme conjugate. The presence of the protein is detected by addition of a chromogenic substrate which illuminates a purple color when it reacts with secondary antibody.

GST-pull down assay

In the cyctoplamic and membrane compartments of the cell, there is an extensive network of protein- protein interactions that take place. A preliminary step in understanding protein structure and function is to determine which proteins interact with each other, thereby identifying the relevant biological pathways. In this system, the protein of interest is fused with a glutathione-S-transferase (GST) tag that is incorporated upstream into an expression vector. Induction of protein expression from the vector's promoter results in expression of a fusion protein. Since the expression process involves transcription and translation of even housekeeping genes, the system

is used to specifically purify and detect proteins of interest from a myriad of host cytoplasmic and membrane proteins. This GST-fusion protein can then be purified from cells via its high affinity for glutathione.

APPENDIX 2

SARS-N gene

ATGGCTAGTGTAAATTGGGCCGATGACAGAGCTGCTAGGAAGAAATTTCCTCCTCCTTC
ATTTTACATGCCTCTTTTGGTTAGTTCTGATAAGGCACCATATAGGGTCATTCCCAGGA
ATCTTGTCCCTATTGGTAAGGGTAATAAAGATGAGCAGATTGGTTATTGGAATGTTCAA
GAGCGTTGGCGTATGCGCAGGGGGCAACGTGTTGATTTGCCTCCTAAAGTTCATTTTTA
TTACCTAGGTACTGGACCTCATAAGGACCTTAAATTCAGACAACGTTCTGATGGTGTTG
TTTGGGTTGCTAAGGAAGGTGCTAAAACTGTTAATACCAGTCTTGGTAATCGCAAACGT
AATCAGAAACCTTTGGAACCAAAGTTCTCTATTGCTTTGCCTCCAGAGCTCTCTGTTGT
TGAGTTTGAGGATCGCTCTAATAACTCATCTCGTGCTAGCAGTCGTTCTTCAACTCGTA
ACAACTCACGAGACTCTTCTCGTAGCACTTCAAGACAACAGTCTCGCACTCGTTCTGAT
TCTAACCAGTCTTCTTCAGATCTTGTTGCTGCTGTTACTTTGGCTTTAAAGAACTTAGGT
TTTGATAACCAGTCGAAGTCACCTAGTTCTTCTGGTACTTCCACTCCTAAGAAACCTAA
TAAGCCTCTTTCTCAACCCAGGGCTGATAAGCCTTCTCAGTTGAAGAAACCTCGTTGGA
AGCGTGTTCCTACCAGAGAGGAAAATGTTATTCAGTGCTTTGGTCCTCGTGATTTTAAT
CACAATATGGGGGATTCAGATCTTGTTCAGAATGGTGTTGATGCCAAAGGTTTTCCACA
GCTTGCTGAATTGATTCCTAATCAGGCTGCGTTATTCTTTGATAGTGAGGTTAGCACTG
ATGAAGTGGGTGATAATGTTCAGATTACCTACACCTACAAAATGCTTGTAGCTAAGGA

TAATAAGAACCTTCCTAAGTTCATTGAGCAGATTAGTGCTTTTACTAAACCCAGTTCTA

TCAAAGAAATGCAGTCACAATCATCTCATGTTGCTCAGAACACAGTACTTAATGCTTCT

ATTCCAGAATCTAAACCATTGGCTGATGATGATTCAGCCATTATAGAAATTGTCAACGA

GGTTTTGCATTAA

NL63-N gene

ATGTCTGATAATGGACCCCAATCAAACCAACGTAGTGCCCCCGCATTACATTTGGTGG

ACCCACAGATTCAACTGACAATAACCAGAATGGAGGACGCAATGGGGCAAGGCCAAA

ACAGCGCCGACCCCAAGGTTTACCCAATAATACTGCGTCTTGGTTCACAGCTCTCACTC

AGCATGGCAAGGAGGAACTTAGATTCCCTCGAGGCCAGGGCGTTCCAATCAACACCAA

TAGTGGTCCAGATGACCAAATTGGCTACTACCGAAGAGCTACCCGACGAGTTCGTGGT

GGTGACGGCAAAATGAAAGAGCTCAGCCCCAGATGGTACTTCTATTACCTAGGAACTG

GCCCAGAAGCTTCACTTCCCTACGGCGCTAACAAAGAAGGCATCGTATGGGTTGCAAC

TGAGGGAGCCTTGAATACACCCAAAGACCACATTGGCACCCGCAATCCTAATAACAAT

GCTGCCACCGTGCTACAACTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAG

AGGGAAGCAGAGGCGGCAGTCAAGCCTCTTCTCGCTCCTCATCACGTAGTCGCGGTAA

TTCAAGAAATTCAACTCCTGGCAGCAGTAGGGGAAATTCTCCTGCTCGAATGGCTAGC

GGAGGTGGTGAAACTGCCCTCGCGCTATTGCTGCTAGACAGATTGAACCAGCTTGAGA

GCAAAGTTTCTGGTAAAGGCCAACAACAACAAGGCCAAACTGTCACTAAGAAATCTGC

TGCTGAGGCATCTAAAAAGCCTCGCCAAAAACGTACTGCCACAAAACAGTACAACGTC

ACTCAAGCATTTGGGAGACGTGGTCCAGAACAAACCCAAGGAAATTTCGGGGACCAA

GACCTAATCAGACAAGGAACTGATTACAAACATTGGCCGCAAATTGCACAATTTGCTC

CAAGTGCCTCTGCATTCTTTGGAATGTCACGCATTGGCATGGAAGTCACACCTTCGGGA

ACATGGCTGACTTATCATGGAGCCATTAAATTGGATGACAAAGATCCACAATTCAAAG

ACAACGTCATACTGCTGAACAAGCACATTGACGCATACAAAACATTCCCACCAACAGA

GCCTAAAAAGGACAAAAAGAAAAAGACTGATGAAGCTCAGCCTTTGCCGCAGAGACA

AAAGAAGCAGCCCACTGTGACTCTTCTTCCTGCGGCTGACATGGATGATTTCTCCAGAC

AACTTCAAAATTCCATGAGTGGAGCTTCTGCTGATTCAACTCAGGCATAA

229E-N gene

ATGGCTACAGTCAAATGGGCTGATGCATCTGAACCACAACGTGGTCGTCAGGGTAG

AATACCTTATTCTCTTTATAGCCCTTTGCTTGTTGATAGTGAACAACCTTGGAAGGTG

ATACCTCGTAATTTGGTACCCATCAACAAGAAAGACAAAAATAAGCTTATAGGCTAT

TGGAATGTTCAAAAACGTTTCAGAACTAGAAAGGGCAAACGGGTGGATTTGTCACC

CAAGCTGCATTTTTATTATCTTGGCACAGGACCCCATAAAGATGCAAAATTTAGAGA

GCGTGTTGAAGGTGTCGTCTGGGTTGCTGTTGATGGTGCTAAAACTGAACCTACAGG

TTACGGTGTTAGGCGCAAGAATTCAGAACCAGAGATACCACACTTCAATCAAAAGC

TCCCAAATGGTGTTACTGTTGTTGAAGAACCTGACTCCCGTGCTCCTTCCCGGTCTCA

GTCGAGGTCGCAGAGTCGCGGTCGTGGTGAATCCAAACCTCAATCTCGGAATCCTTC

AAGTGACAGAAACCATAACAGTCAGGATGACATCATGAAGGCAGTTGCTGCGGCTC

TTAAATCTTTAGGTTTTGACAAGCCTCAGGAAAAAGATAAAAAGTCAGCGAAAACG

GGTACTCCTAAGCCTTCTCGTAATCAGAGTCCTGCTTCTTCTCAAACTTCTGCCAAGA

GTCTTGCTCGTTCTCAGAGTTCTGAAACAAAAGAACAAAAGCATGAAATGCAAAAG

CCACGGTGGAAAAGACAGCCTAATGATGATGTGACATCTAATGTCACACAATGTTTT

GGCCCCAGAGACCTTGACCACAACTTTGGAAGTGCAGGTGTTGTGGCCAATGGTGTT

AAAGCTAAAGGCTATCCACAATTTGCTGAGCTTGTGCCGTCAACAGCTGCTATGCTG

TTTGATAGTCACATTGTTTCCAAAGAGTCAGGCAACACTGTGGTCTTGACTTTCACTA

CTAGAGTGACTGTGCCCAAAGACCATCCACACTTGGGTAAGTTTCTTGAGGAGTTAA

ATGCATTCACTAGAGAAATGCAACAACATCCTCTTCTTAACCCTAGTGCACTAGAAT

TCAACCCATCTCAAACTTCACCTGCAACTGCTGAACCAGTGCGTGATGAAGTTTCTA

TTGAAACTGACATAATTGATGAAGTAAACTAA

OC43-N gene

ATGTCTTTTACTCCTGGTAAGCAATCCAGTAGTAGAGCGTCCTCTGGAAATCGTTCT

GGTAATGGCATCCTCAAGTGGGCCGATCAGTCCGACCAGTTTAGAAATGTTCAAACC

AGGGGTAGAAGAGCTCAACCCAAGCAAACTGCTACCTCTCAGCAACCATCAGGAGG

GAATGTTGTACCCTACTATTCTTGGTTCTCTGGAATTACTCAGTTTCAAAAGGGAAA

GGAGTTTGAGTTTGTAGAAGGACAAGGTGTGCCTATTGCACCAGGAGTCCCAGCTAC

TGAAGCTAAGGGGTACTGGTACAGACACAACAGACGTTCTTTTAAAACAGCCGATG

GCAACCAGCGTCAACTGCTGCCACGATGGTATTTTTACTATCTGGGAACAGGACCGC

ATGCTAAAGACCAGTACGGCACCGATATTGACGGAGTCTACTGGGTCGCTAGCAAC

CAGGCTGATGTCAATACCCCGGCTGACATTGTCGATCGGGACCCAAGTAGCGATGA

GGCTATTCCGACTAGGTTTCCGCCTGGCACGGTACTCCCTCAGGGTTACTATATTGA

AGGCTCAGGAAGGTCTGCTCCTAATTCCAGATCTACTTCGCGCACATCCAGCAGAGC

CTCTAGTGCAGGATCGCGTAGTAGAGCCAATTCTGGCAATAGAACCCCTACCTCTGG

TGTAACACCTGACATGGCTGATCAAATTGCTAGTCTTGTTCTGGCAAAACTTGGCAA

GGATGCCACTAAACCTCAGCAAGTAACTAAGCATACTGCCAAAGAAGTCAGACAGA

AAATTTTGAATAAGCCCCGCCAGAAGAGGAGCCCCAATAAACAATGCACTGTTCAG

CAGTGTTTTGGTAAGAGAGGCCCTAATCAGAATTTTGGTGGTGGAGAAATGTTAAAA

CTTGGAACTAGTGACCCACAGTTCCCCATTCTTGCAGAACTCGCACCCACAGCTGGT

GCGTTTTTCTTTGGATCAAGATTAGAGTTGGCCAAAGTGCAGAATTTATCTGGGAAT

CCTGACGAGCCCCAGAAGGATGTTTATGAATTGCGCTATAACGGCGCAATTAGGTTT

GACAGTACACTTTCAGGTTTTGAGACCATAATGAAGGTGCTGAATGAGAATTTGAAT

GCCTATCAACAACAAGATGGTATGATGAATATGAGTCCAAAACCACAGCGTCAGCG

TGGTCATAAGAATGGACAAGGAGAAAATGATAATATAAGTGTTGCAGTGCCCAAAA

GCCGCGTGCAGCAAAATAAGAGTAGAGAGTTGACTGCAGAGGACATCAGCCTTCTT

AAGAAGATGGATGAGCCCTATACTGAAGACACCTCAGAAATATAA

Feline CoV-N gene

ATGGCCACACAGGGACAACGCGTCAACTGGGGAGATGAACCTTCCAAAAGACGCAG

TCGTTCTAACTCTCGTGGTCGGAAGAATAATGATATACCTTTGTCTTACTTCAACCCC

CTCACCCTCGAATCAGGATCAAAGTTTTGGAATGTATGTCCGAGAGACTTTGTTCCC

AAGGGAATAGGTAACAAGGATCAACAAATTGGTTATTGGAATAGACAGTTACGCTA

CCGAATTGTCAAGGGCCAGCGTAAGGAACTGCCTGAGAGATGGTTCTTCTACTATCT

AGGCACAGGACCTCATGCTGATGCTAAATTTAAAGACAAGATTGATGGAGTCTTCTG

GGTTGCTAGGGATGGTGCCATGAACAAGCCTACAACACTAGGCACTCGTGGAACCA

ATAATGAATCCAAACCTTTGAAATTTGATGGTAAGATACCACCACAATTTCAGCTTG

AAGTGAATCGTTCTAGAAACAATTCAAGAGCAGGTTCTCAGTCTAGATCAACCTCAA

124

GAAATAGGTCTCAATCTAGAGGAAGACAACAATCCAATAACCAGAACGATAATGTT

GAGGATACAATTGTTGCTGTGCTTCAAAGATTAGGTGTTACTGATAAGCAAAGGTCA

CGTTCTAAATCCAGAGAACGTAGTGATTCCAAGCCTAGGGATACAACACCTAAAAA

TTCCAACAAACACACATGGAAGAAACTGCAGGTAAGGGAGATGTGACAAATTTCT

ATGGTGCTAGAAGTGCTTCAGCTAACTTTGGTGATAGTGATCTCGTTGCCAATGGTA

ACGCTGCCAAATGCTACCCTCAGATAGCTGAATGCGTTCCATCAGTATCTAGCATGC

TCTTCGGCAGTCAATGGTCTGCTGAAGAAGCTGGTGATCAAGTGAAAGTCACGCTCA

CTCATACTTACTACCTGCCAAAGGATGATGCCAAAACCAGTCAGTTCCTAGAACAGA

TTGACGCCTACAAACGGCCTTCACAGGTGGCTAAAGGTCAGAGGACAAGGAAATCC

CGCTCTAAGTCTGCTGATAAGAAGCCTGAGGAATTGTCTGTAACTCTTGTAGAGGCA

TACACAGATGTGTTTGATGACACACAGGTTGAGATGATTGATGAGGTTACGAACTAA

TGEV-N gene

ATGGCCAACCAGGGACAACGTGTCAGTTGGGGAGATGAATCTACCAAAACACGTGG

TCGCTCCAATTCCCGTGGTCGGAAGAGTAATAACATACCTCTTTCATTCTTCAACCCC

ATAACCCTCCAGCAAGGTTCAAAATTTTGGAACTTATGTCCGAGGGACTTTGTACCC

AAAGGAATAGGTAACAGGGATCAACAGATTGGTTATTGGAATAGACAAACTCGCTA

TCGCATGGTGAAGGGCCAACGTAAAGAGCTTCCTGAAGGTGGTTCTTCTACTACTT

AGGTACTGGACCTCATGCAGATGCCAAATTTAAAGATAAATTAGATGGAGTTGTCTG

GGTTGCCAAGGATGGTGCCATGAACAAACCAACCACGCTTGGTAGTCGTGGTGCTA

ATAATGAATCCAAAGCTTTGAAATTCGATGGTAAAGTGCCAGGCGAATTTCAACTTG

AAGTTAACCAGTCAAGGGATAATTCAAGGTCACGCTCACAATCTAGATCTCGGTCTA

GAAACAGATCTCAATCTAGAGGCAGGCAACAATCCAATAACAAGAAGGATGACAGT

GTAGAACAAGCTGTTCTTGCCGCACTTAAAAAGTTAGGTGTTGACACAGAAAAACA

ACAGCAACGCTCTCGTTCTAAATCTAAAGAACGTAGTAACTCTAAGACAAGAGATA

CTACACCTAAGAATGAAAACAAACACACCTGGAAGAGAACTGCAGGTAAAGGTGAT

GTGACAAGATTTTATGGAGCTAGAAGCAGTTCAGCCAATTTTGGTGACAGTGACCTC

GTTGCCAATGGGAGCAGTGCCAAGCATTACCCACAATTGGCTGAATGTGTTCCATCT

GTGTCTAGCATTTTGTTTGGAAGCTATTGGACTTCAAAGGAAGATGGCGACCAGATA

GAAGTCACGTTCACACACAAATACCACTTGCCAAAGGATGATCCTAAAACTGAACA

ATTCCTTCAGCAGATTAATGCCTATGCTCGCCCATCAGAAGTGGCAAAAGAACAGA

GAAAAAGAAAATCTCGTTCTAAATCTGCAGAAAGGTCAGAGCAAGAGGTGGTACCT

GATGCATTAATAGAAAATTATACAGATGTGTTTGATGACACACAGGTTGAGATGATT

GATGAGGTAACGAACTAA

Bovine CoV-N gene

ATGTGAGCGATTGCGTGCGTGCATCCGCTTCACTGATCTCTTGTTAGATCTTTTTATA

ATCTAAACTTTAAGGATGTCTTTTACTCCTGGTAAGCAATCCAGTAGTAGAGCGTCC

TCTGGAAATCGTTCTGGTAATGGCATCCTTAAGTGGGCCGATCAGTCCGACCAATCT

AGAAATGTTCAAACCAGGGGTAGAAGAGCTCAACCCAAGCAAACTGCTACTTCTCA

GCAACCATCAGGAGGGAATGTTGTACCCTACTATTCTTGGTTCTCTGGAATTACTCA

GTTTCAAAAGGGAAAGGAGTTTGAATTTGCTGAGGGACAAGGTGTGCCTATTGCAC

CAGGAGTCCCAGCTACTGAAGCTAAGGGGTACTGGTACAGACACAACAGACGTTCT

TTTAAAACACGCGATGGCAACCAGCGTCAATTGCTGCCACGATGGTATTTTTACTAT

CTTGGAACAGGACCGCATGCCAAAGACCAGTATGGCACCGACATTGACGGAGTCTT

CTGGGTCGCTAGTAACCAGGCTGATGTCAATACCCCGGCTGACATTCTCGATCGGGA

CCCAAGTAGCGATGAGGCTATTCCGACTAGGTTTCCGCCTGGCACGGTACTCCCTCA

GGGTTACTATATTGAAGGCTCAGGAAGGTCTGCTCCTAATTCCAGATCTACTTCACG

CGCATCCAGTAGAGCCTCTAGTGCAGGATCGCGTAGTAGAGCCAATTCTGGCAATA

GAACCCCTACCTCTGGTGTAACACCTGATATGGCTGATCAAATTGTCAGTCTTGTTTT

GGCAAAACTTGGCAAGGATGCCACTAAGCCACAGCAAGTAACTAAGCAGACTGCCA

AAGAAATCAGACAGAAAATTTTGAATAAGCCCCGCCAGAAGAGGAGCCCCAATAAA

CAATGCACTGTTCAGCAGTGTTTTGGGAAGAGAGGCCCCAATCAGAATTTTGGTGGT

GGAGAAATGTTAAAACTTGGAACTAGTGACCCACAGTTCCCCATTCTTGCAGAACTC

GCACCCACAGCTGGTGCGTTTTTCTTTGGATCAAGATTAGAGTTGGCCAAAGTGCAG

AATTTGTCTGGGAATCTTGATGAGCCCCAGAAGGATGTTTATGAATTACGCTACAAT

GGCGCAATTAGATTTGATAGTACACTTTCAGGTTTTGAGACCATAATGAAGGTGTTG

AATGAGAATTTGAATGCATATCAACAACAAGATGGTATGATGAATATGAGTCCAAA

ACCACAGCGTCAGCGTGGTCAGAAGAATGGACAAGGAGAAAATGATAATATAAGTG

TTGCAGCGCCCAAAAGCCGTGTGCAGCAAAATAAGAGTAGAGAGTTGACTGCAGAG

GACATCAGCCTTCTTAAGAAGATGGATGAGCCCTATACTGAAGACACCTCAGAAAT

ATAAGAGAATGAACCTTATGTCGGTACCTGGTGGCAACCCCTCGCAGGAAAGTCGG

GATAAGGCATTCTCTATCAGAATGGATGTCTTGCTGCTATAATAGATAGAGAAGGTT

ATAGCAGACTATAGATTAATTAGTTGAAAGTTTTGTGTGGTAATGTATAGTGTTGGA

GAAAGTGAAAGACTTGCGGAAGTAATTGCCCACAAGTGCCCAAGGGGAAGAGCCA

GCATGTTAAGTTGCCACCCAGTAATTAGTAAATGAATGAAGTTAATTATGGCCAATT

GGAAGAATCACA

Murine hepatitis virus-N gene

ATGTCTTTTGTTCCTGGGCAAGAAAATGCCGGTAGCAGAAGCTCCTCTGGAAACCGC

GCTGGTAATGGCATCCTCAAGAAGACCACTTGGGCTGACCAAACCGAGCGTGGAAA

TAGAGGCAGAAGGAACCATCCCAAGCAGACTGCAACTACTCAGCCCAATGCCGGGA

GTGTGGTTCCCCATTACTCTTGGTTTTCGGGCATCACCCAGTTTCAAAAGGGAAAGG

AGTTCCAGTTTGCACAAGGACAGGGAGTGCCTATTGCCAGTGGAATCCCCGCTTCAG

AGCAAAAGGGATATTGGTATAGACACAACCGACGTTCTTTTAAAACACCTGATGGC

CAGCACAAGCAGCTACTGCCCAGATGGTATTTTTACTATCTTGGAACAGGGCCCCAT

GCTGGCGCAGAGTATGGCGACGATATCGAAGGAGTTGTCTGGGTCGCAAGCCAACA

GGCCGACACTAAGACCACTGCCGATGTTGTTGAAAGGGACCCAAGCAGTCATGAGG

CTATTCCTACTAAGTTTGCGCCCGGCACGGTATTGCCTCAAGGCTTTTATGTAGAAG

GCTCGGGAAAGTCTGCACCTGCTAGTCGATCTGGTTCGCGGTCACAATCCCGTGGGC

CAAATAATCGCGCTAGAAGCAGTTCCAACCAGCGCCAGCCTGCCTCTGCTGTAAAA

CCTGACATGGCCGAAGAAATTGCTGCTCTTGTTTTGGCTAAGCTTGGTAAAGATGCC

GGCCAGCCCAAGCAGGTAACTAAGCAAAGCGCCAAAGAAGTCAGGCAGAAAATTTT

AACTAAGCCTCGTCAAAAGAGGACTCCAAACAAGCAGTGCCCAGTGCAGCAGTGTT

TTGGGAAGAGAGGCCCTAATCAGAACTTTGGAGGCTCTGAAATGTTAAAACTTGGA

ACTAGTGATCCGCAGTTCCCCATTCTTGCAGAGTTGGCTCCAACACCTAGTGCCTTCT

TCTTTGGATCTAAATTAGAATTGGTCAAAAAGAACTCTGGTGGTGCTGATGAACCCA

CCAAAGATGTTTATGAATTGCAGTATTCAGGTGCAATTAGATTTGATAGTACTCTAC

CCGGTTTTGAGACTATCATGAAAGTGTTGACTGAGAATTTGAATGCCTACCAGGACC

AAGCTGGTAGTGTAGATCTAGTGAGCCCAAAGCCTCCAAGAAGAGGTCGTAGACAG

GCTCAAGAAAGAAAGATGAAGTAGATAATGTAAGCGTTGCAAAGCCCAAAAGCTT

GGTGCAGCGAAATGTAAGTAGAGAATTAACCCCCGAGGATCGTAGCCTGCTGGCTC

AGATCCTAGACGATGGCGTTGTGCCAGATGGGTTGGAAGATGACTCTAATGTGTAA

SARS-N amino acid sequence

MASVNWADDRAARKKFPPPSFYMPLLVSSDKAPYRVIPRNLVPIGKGNKDEQIGYWNV

QERWRMRRGQRVDLPPKVHFYYLGTGPHKDLKFRQRSDGVVWVAKEGAKTVNTSLG

NRKRNQKPLEPKFSIALPPELSVVEFEDRSNNSSRASSRSSTRNNSRDSSRSTSRQQSRTR

SDSNQSSSDLVAAVTLALKNLGFDNQSKSPSSSGTSTPKKPNKPLSQPRADKPSQLKKPR

WKRVPTREENVIQCFGPRDFNHNMGDSDLVQNGVDAKGFPQLAELIPNQAALFFDSEV

STDEVGDNVQITYTYKMLVAKDNKNLPKFIEQISAFTKPSSIKEMQSQSSHVAQNTVLN

ASIPESKPLADDDSAIIEIVNEVLH

NL63-N amino acid sequence

MSDNGPQSNQRSAPRITFGGPTDSTDNNQNGGRNGARPKQRRPQGLPNNTASWFTALT

QHGKEELRFPRGQGVPINTNSGPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGT

GPEASLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQGTTLPKGFYAE

GSRGGSQASSRSSSRSRGNSRNSTPGSSRGNSPARMASGGGETALALLLLDRLNQLESK

VSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQD

LIRQGTDYKHWPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDN

VILLNKHIDAYKTFPPTEPKKDKKKKTDEAQPLPQRQKKQPTVTLLPAADMDDFSRQLQ

NSMSGASADSTQ

229E-N amino acid sequence

YSPLLVDSEQPWKVIPRNLVPINKKDKNKLIGYWNVQKRFRTRKGKRVDLSPKLHFYYL

GTGPHKDAKFRERVEGVVWVAVDGAKTEPTGYGVRRKNSEPEIPHFNQKLPNGVTVVE

EPDSRAPSRSQSRSQSRGRGESKPQSRNPSSDRNHNSQDDIMKAVAAALKSLGFDKPQE

KDKKSAKTGTPKPSRNQSPASSQTSAKSLARSQSSETKEQKHEMQKPRWKRQPNDDVT

SNVTQCFGPRDLDHNFGSAGVVANGVKAKGYPQFAELVPSTAAMLFDSHIVSKESGNT

VVLTFTTRVTVPKDHPHLGKFLEELNAFTREMQQHPLLNPSALEFNPSQTSPATAEPVRD

EVSIETDIIDEVN

OC43-n amino acid sequence

MSFTPGKQSSSRASSGNRSGNGILKWADQSDQFRNVQTRGRRAQPKQTATSQQPSGGN

VVPYYSWFSGITQFQKGKEFEFVEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQ

RQLLPRWYFYYLGTGPHAKDQYGTDIDGVYWVASNQADVNTPADIVDRDPSSDEAIPT

RFPPGTVLPQGYYIEGSGRSAPNSRSTSRTSSRASSAGSRSRANSGNRTPTSGVTPDMAD

QIASLVLAKLGKDATKPQQVTKHTAKEVRQKILNKPRQKRSPNKQCTVQQCFGKRGPN

QNFGGGEMLKLGTSDPQFPILAELAPTAGAFFFGSRLELAKVQNLSGNPDEPQKDVYEL

RYNGAIRFDSTLSGFETIMKVLNENLNAYQQQDGMMNMSPKPQRQRGHKNGQGENDN

ISVAVPKSRVQQNKSRELTAEDISLLKKMDEPYTEDTSEI

Feline CoV-N amino acid sequence

MATQGQRVNWGDEPSKRRSRSNSRGRKNNDIPLSYFNPLTLESGSKFWNVCPRDFVPK

GIGNKDQQIGYWNRQLRYRIVKGQRKELPERWFFYYLGTGPHADAKFKDKIDGVFWV

ARDGAMNKPTTLGTRGTNNESKPLKFDGKIPPQFQLEVNRSRNNSRAGSQSRSTSRNRS

QSRGRQQSNNQNDNVEDTIVAVLQRLGVTDKQRSRSKSRERSDSKPRDTTPKNSNKHT

WKKTAGKGDVTNFYGARSASANFGDSDLVANGNAAKCYPQIAECVPSVSSMLFGSQW

SAEEAGDQVKVTLTHTYYLPKDDAKTSQFLEQIDAYKRPSQVAKGQRTRKSRSKSADK

KPEELSVTLVEAYTDVFDDTQVEMIDEVTN

TGEV-N amino acid sequence

MANQGQRVSWGDESTKTRGRSNSRGRKSNNIPLSFFNPITLQQGSKFWNLCPRDFVPKG

IGNRDQQIGYWNRQTRYRMVKGQRKELPERWFFYYLGTGPHADAKFKDKLDGVVWV

AKDGAMNKPTTLGSRGANNESKALKFDGKVPGEFQLEVNQSRDNSRSRSQSRSRSRNR

SQSRGRQQSNNKKDDSVEQAVLAALKKLGVDTEKQQQRSRSKSKERSNSKTRDTTPKN

ENKHTWKRTAGKGDVTRFYGARSSSANFGDSDLVANGSSAKHYPQLAECVPSVSSILF

GSYWTSKEDGDQIEVTFTHKYHLPKDDPKTEQFLQQINAYARPSEVAKEQRKRKSRSKS

AERSEQEVVPDALIENYTDVFDDTQVEMIDEVTN

Bovine CoV-N amino acid sequence

MSFTPGKQSSSRASSGNRSGNGILKWADQSDQSRNVQTRGRRAQPKQTATSQQPSGGN

VVPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTRDGNQ

RQLLPRWYFYYLGTGPHAKDQYGTDIDGVFWVASNQADVNTPADILDRDPSSDEAIPT

RFPPGTVLPQGYYIEGSGRSAPNSRSTSRASSRASSAGSRSRANSGNRTPTSGVTPDMAD

QIVSLVLAKLGKDATKPQQVTKQTAKEIRQKILNKPRQKRSPNKQCTVQQCFGKRGPNQ

NFGGGEMLKLGTSDPQFPILAELAPTAGAFFFGSRLELAKVQNLSGNLDEPQKDVYELR

YNGAIRFDSTLSGFETIMKVLNENLNAYQQQDGMMNMSPKPQRQRGQKNGQGENDNI

SVAAPKSRVQQNKSRELTAEDISLLKKMDEPYTEDTSEI

Murine hepatitis virus-N amino acid sequence

MSFVPGQENAGSRSSSGNRAGNGILKKTTWADQTERGNRGRRNHPKQTATTQPNAGSV

VPHYSWFSGITQFQKGKEFQFAQGQGVPIASGIPASEQKGYWYRNRRSFKTPDGQHKQL

LPRWYFYYLGTGPHAGAEYGDDIEGVVWVASQQADTKTTADVVERDPSSHEAIPTKFA

PGTVLPQGFYVEGSGKSAPASRSGSRSQSRGPNNRARSSSNQRQPASAVKPDMAEEIAA

LVLAKLGKDAGQPKQVTKQSAKEVRQKILTKPRQKRTPNKQCPVQQCFGKRGPNQNFG

GSEMLKLGTSDPQFPILAELAPTPSAFFFGSKLELVKKNSGGADEPTKDVYELQYSGAIR

FDSTLPGFETIMKVLTENLNAYQDQAGSVDLVSPKPPRRGRRQAQEKKDEVDNVSVAK

PKSLVQRNVSRELTPEDRSLLAQILDDGVVPDGLEDDSN

HIV-1 capsid protein

PIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTV

GGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGW

MTNNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRFYKTLRAEQAS

QEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVL

HIV-1 p24

PIVQNAQGQMIHQSLSPRTLNAWVKVIEEKAFSPEVIPMFSALSEGATPFDLNMMLNIVG

GHQAAMLMLKDTINEEAAEWDRLHPVHAGPVAPGQMREPRGSD

HIV-1 Vpr

MEQAPEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWLHGLGQHIYETYGDTWAGVE

AIIRILQQLLFIHFRIGCRHSRIGVTRQRRARNGASRS

HIV-1 heparan

MGARASVLIGGKLNPWEKIRLRPGGKKKYRLKLLVWASRELERFALDSGLLETSQGCQ

QIMEQLQPALRTGTGDLNSLFNTVATLWCVHQRIEVKDTKEALDKIEEIPNKKKPQQAA

ADTGSNSKVSQNYPIVQNAQGQMAYQSISPRTLNAWVKAIEEKAFSPEVIPMFTALSEG

ATPQDLNMMLNIVGGHQAAMQMLKDTFNEEAAEWDRVLPVQAGPHPAGQMREPRGS

DIAGTTSTLQEQIGWMTSNPPIPVGEIYKRWIVLGLNKIVRMYSPVSILDIRQGPKEPFRD

YVDRFFKCLRAEQATQEVKNWMTETLLVQNANPDCKNILKALGPAASLEEMMTACQG

VGGPSHKARVLAEAMSQVHSPNIMMQKGNFRGQKRIKCFNCGKEGHLARNCRAPRKK

GCWKCGKEGHQMKDCTERQANFLGRIWPSSKGRPGNFPQSRLEPTAPPAESFGFGEEIA

PSPKQETKDKELYPLASLKSLFGNDP

HCV NS5A

SGSWLRDVWDWICTVLTDFKTWLQSKLLPRLPGVPFLSCQRGYKGVWRGDGVMHTTC

PCGAQITGHVKNGSMRIVGPKTCSNTWHGTFPINAYTTGSCTPSPAPNYSRALWRVAAE

EYVEVTRVGDFHYVTGMTTDNIKCPCQVPAPEFFTELDGVRLHRYAPACKPLLRDEVTF

QVGLNQYVVGSQLPCEPEPDVTVLTSMLTDPSHITAETAKRRLARGSPPSLASSSASQLS

APSLKATCTTHHDSPDADLIEANLLWRQEMGGNITRVESENKVVILDSFDPLRAEEDERE

VSVPAEILRKTRKFPSALPIWARPDYNPPLLESWRDPDYVPPVVHGCPLPPTKAPPIPPPR

RKRTVILTESTVSSALAELATKTFGSSGSSAVDSGTATAPPDGPSDDGDAGSDAESYSSM

PPLEGEPGDPDLSDGSWSTVSEEASEDVVCC

HCV NS2

MDREMAASCGGVVFVGLVFLTMSPHYKVFLARLIWWLQYFLTIAEAHLQVWIPPLNIR

GGRDAIILLTCVIHPELIFDITKLLLATLGPLLVLQAGIARVPYFVRAHGLIRACVLLRKVA

GGHYVQMALMKLAALTGTYLYDHLTPLQYWAHAGLRDLAVAVEPVIFSDMEIKIITWG

ADTAACGDIIQGLPVSARRGREILLGPADSLEGQGWRLL