UNIVERSITY OF THE
FREE STATE
UNIVERSITEIT VAN DIE
VRYSTAAT
YUNIVESITHI YA
FREISTATA

UFS·UV
HEALTH SCIENCES
GESONDHEIDSWETENSKAPPE

IDENTIFICATION OF ANTIGEN-SPECIFIC SEROLOGICAL CROSS-REACTIVITY AMONG SURVIVORS OF CRIMEAN-CONGO HAEMORRHAGIC FEVER

Azeeza Rangunwala

February 2013

IDENTIFICATION OF ANTIGEN-SPECIFIC SEROLOGICAL CROSS-REACTIVITY AMONG SURVIVORS OF CRIMEAN-CONGO HAEMORRHAGIC FEVER

Azeeza Rangunwala

Dissertation submitted in fulfillment of the requirements for the degree Magister Medical Scientiae (Virology).

In the Faculty of Health Sciences

Department of Medical Microbiology and Virology

University of the Free State.

Promoter: Prof. F. J. Burt

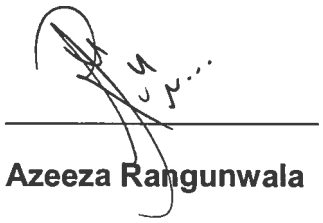
February 2013

Bloemfontein

South Africa

DECLARATION

I declare that the dissertation hereby submitted by me for the **M.Med.Sc Virology** degree at the University of the Free State, Bloemfontein, is my independent effort and had not previously been submitted at another university/faculty. I further more waive copyright of the dissertation in favour of the University of the Free State.



Azeeza Rangunwala

ABSTRACT

Crimean-Congo haemorrhagic fever virus (CCHFV) is a member of the *Nairovirus* genus belonging to the family *Bunyaviridae*, which consists of diverse RNA viruses. The tick-borne virus is widely distributed in Africa, Asia, the Middle East and Europe. CCHFV has the propensity to cause nosocomial infections with a high fatality rate and is endemic in South Africa. Handling of the virus requires biosafety level 4 (BSL-4) conditions, which has limited availability. Advances in molecular techniques have allowed preparation of safe recombinant antigens that have been used in diagnosis and serosurveillance of CCHFV. There is currently no vaccine available; elucidating genetic and antigenic diversity would provide useful information for the development of successful vaccines. The single stranded, negative sense, segmented RNA genome of CCHFV consists of a small (S) medium (M) and large (L) segment. The L segment encodes the viral polymerase, the M segment encodes precursors for glycoproteins and the S segment encodes the viral nucleocapsid protein (NC). Several phylogenetic studies have been performed to determine the genetic diversity of the S segment, but antigenic diversity between geographically distinct isolates has not been well investigated.

The purpose of this study was to examine the global nucleic acid and amino acid diversity between isolates worldwide; clone and express a recombinant CCHFV NC from a southern African CCHFV and distantly related Greek CCHFV strain and determine the antigenic cross-reactivity between the two isolates.

Initially, the aim was to prepare a mammalian expressed recombinant CCHFV NC that has post-translational modifications such as correct folding, disulfide bridge formation, modifications that facilitate binding functions and addition of functional groups that regulate protein activity. South African strain SPU 92/01 RNA was transcribed to cDNA. The native gene encoding the NC of CCHFV was amplified using primers that have been modified for mammalian expression (addition of the Kozak sequence to the forward primer) and the amplicon was cloned into pcDNA3.1 Directional TOPO expression vector that possesses a Histidine- and V5 tag. Nucleic acid sequencing was performed, to determine whether the gene was cloned in frame and also because isolate SPU92/01 had not been previously sequenced.

This sequence was included in a phylogenetic study based on the coding sequences of the S segment of 45 isolates. Nucleotide sequences for the open-reading frame (ORF) of the S segment of 44 isolates were retrieved from GenBank. Phylogenetic analyses using nucleotide and amino acid sequences of the NC revealed six different groups, linked to geographic location and correlating with results from previous studies based on the full-length S segment. Nucleotide sequence diversity and amino acid diversity between groups, within groups and pairwise distances were calculated. Isolate SPU 92/01 grouped with other South African isolates and shares 96.6% nucleic acid homology and 99.4% amino acid homology with SPU 415/85. The most diverse strain, AP92, displayed the greatest amino acid difference with SPU415/85 (8.7%). Lower amino acid diversity suggested synonymous changes in basepairs, resulting in fewer differences at protein level. Protein assays, in turn, could provide more information regarding antigenic diversity. Initially, expression of the NC of isolate SPU92/01 in a TOPO vector construct, using a mammalian cell line was attempted.

A control expression vector pcDNA3.1/V5-His/*LacZ*-TOPO DNA was prepared for mammalian cell transfection. During optimization of mammalian cell transfection, three cell lines were tested: Vero 76, HEK293 and HeLa cells. Two transfection reagents were used: FuGene 6 and GeneJuice. The control expression vector transfection efficiency was determined using a β-galactosidase assay, while immunoflourescence assays (IFA) and Western blot were used to determine successful transfection of cells with the construct containing the NC gene of CCHFV. Proteins were not detected by Western blot analysis and transfection efficiency determined by the β-galactosidase assay did not exceed 20%. The yield of recombinant protein was insufficient for downstream applications.

To obtain a higher protein yield during recombinant protein expression, a bacterial expression system was adopted. A previously expressed codon optimized NC from another closely related South African isolate, SPU 415/85 was subcloned into pColdTF vector and was expressed in OverExpress BL21 (DE3) cells. The protein was purified from the soluble phase. Similarly, to account for codon bias, the gene encoding the NC of AP92 was analyzed using the Rare Codon Analysis Tool and codon optimized for expression in *Escherichia coli* host cells. The codon optimized gene was synthesized by GenScript and cloned into pUC57 vector using *BamH1* restriction sites. After the codon optimized gene was

rescued from pUC57, it was cloned into pColdTF expression vector. The vector consisted of a *cspA* promoter which codes for a cold shock protein. A 106 kDa protein (His tagged TF chaperone protein fused to AP92 NC) was expressed from the construct in the soluble and insoluble phase. Due to the probability of misfolding during the renaturing process, that is compulsory when purifying using denaturing conditions, proteins were purified from the native/soluble phase. Both proteins derived from SPU 415/85 and AP92, respectively, were tested in an ELISA for ability to detect IgG against CCHFV in South African patients who have survived CCHF infection. ELISA is a useful immunoassay to determine whether two different recombinant proteins cross-react against the same panel of sera. The recombinant antigen designated SA NC, detected IgG in all the sera from South African survivors of CCHF, while one sample tested negative for IgG against the AP92 NC. Two of the most distantly related isolates, according to the phylogenetic analyses based on nucleotide and amino acid sequences, cross-reacted when tested against the same panel of sera. The Greek isolate NC detected antibody in South African survivors, suggesting the presence of conserved epitopes.

In addition, the stability of the bacterially expressed recombinant proteins was tested using different storage methods and lyophilization was required to maintain stability for one month.

Development of safe and reliable laboratory reagents is necessary to increase diagnostic capacity worldwide for this emerging virus that has significant public health concerns. The serological cross-reactivity of the two NC antigens suggests that recombinant antigens prepared from geographically specific CCHFV strains will have diagnostic and epidemiological applications worldwide.

ACKNOWLEDGEMENTS

The completion of this study would not have been possible without the assistance of the following institutions and individuals. Your contribution is highly appreciated.

- My supervisor, Professor Felicity Burt, whose work ethic and extensive knowledge in Virology promoted an interactive learning environment. Thank you for having faith in my abilities, helping with troubleshooting and adding value to this project.
- The financial assistance from the following institutions, during my period of study, is hereby acknowledged: The National Research Foundation (NRF), Polio Research Foundation (PRF) and Medical Research Council (MRC). Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to these institutions.
- Fellow students, who have contributed to my knowledge and provided perspective when necessary.
- My brother-in-law, Werner Morris, for design and layout assistance during writing of this dissertation. Tarené Simons, Grant Richardson and Jo-anna Lu for proof-reading assistance.
- My grandmother, Shirley Ayob, my guardian in every sense of the word. My sister, Faheeza Morris and all my friends for their support, understanding and encouragement throughout this study.
- God, for blessings, guidance and solace.

CONTENTS

DECLARATION iii

ABSTRACT iv

ACKNOWLEDGEMENTS vii

CHAPTER 1: LITERATURE REVIEW 5

1.1. Introduction and history 5

1.2. Structural characteristics and physicochemical properties 6

1.3. Molecular biology of CCHFV 8

1.4. Epidemiology 10

1.5. Life cycle of the virus in nature 11

1.6. Signs and symptoms 12

1.7. Clinical pathology 12

1.8. Pathogenesis of the virus 14

1.9. Diagnosis 15

1.9.1. Virus isolation 15

1.9.2. Molecular assays 15

1.9.3. Serological assays 16

1.10. Treatment, prevention and control 17

1.11. Phylogenetic diversity 17

1.12. Neutralizing antibodies and antigenic relationships 19

1.13. Problem identification 20

1.14. Aim and objectives 20

CHAPTER 2: GLOBAL DIVERSITY OF CRIMEAN CONGO HAEMORRHAGIC FEVER NUCLEOCAPSID PROTEINS 21

2.1. Introduction 21

2.2. Materials and Methods 23

2.2.1. Retrieval of sequence data and multiple sequence alignment 23

2.2.2. Determination of sequence of southern African CCHFV isolate 26

2.2.2.1. Viral RNA and cDNA synthesis 27

2.2.2.2. Primer design and Phusion High-Fidelity PCR 28

2.2.2.3. Agarose gel electrophoresis 29

2.2.2.4. DNA purification 29

2.2.2.5. Concentration of DNA 30

2.2.2.6. Cloning of gene encoding NC into pcDNA ™3.1D/V5-His/TOPO® 30

2.2.2.7. Transformation of chemically competent TOP10 cells 31

2.2.2.8. Plasmid purification 32

2.2.2.9. Restriction enzyme digestion reactions to identify positive transformants 33

2.2.2.10. DNA sequencing of CCHF NC gene in TOPO vector 33

2.2.3. Distance analysis using MEGA 35

2.3. Results and discussion 36

2.3.1. Cloning and sequencing 36

2.3.2. Determination of nucleotide sequence data for isolate 92/01 39

2.3.3. Phylogenetic analysis using nucleotide sequence data 42

Summary 52

CHAPTER 3: EXPRESSION OF RECOMBINANT CCHFV NUCLEOPROTEIN 54

3.1. Introduction 54

3.2. Methods and Materials 57

3.2.1. Mammalian expression of CCHFV NC 57

3.2.1.1. Preparation of plasmid constructs for use in transfection 57

3.2.1.2. Mammalian cell lines 59

3.2.1.3. Transfection experiments 60

3.2.1.4. Monitoring transfection 63

3.2.2. Bacterial Expression 67

3.2.2.1. Codon optimization of CCHFV NC genes 67
3.2.2.2. Preparation of chemically competent cells 70
3.2.2.3. Transformation of XL10 cells with AP92 in pUC57 plasmid 70
3.2.2.4. Recovery of CCHFV AP92 NC gene from pUC57 vector and subcloning into pColdTF vector 71
3.2.2.5. Identification of positive transformants 73
3.2.2.6. Induction study: time-course protein expression 75
3.2.2.7. Protein solubility 76
3.2.2.8. Purification of soluble recombinant His-tagged protein under native conditions 77

3.3. Results 79
3.3.1. Preparation of pcDNA3.1/V5-His-TOPO/LacZ DNA plasmid DNA for transfection experiments 79
3.3.2. Optimization of transfection experiments 81
3.3.3. Mammalian cell lysis and characterization of expressed His tagged NC by Western blot analysis 85
3.3.4. Codon optimization of CCHFV AP92 NC gene 86
3.3.5. Cloning of gene encoding codon optimized CCHF AP92 NC from pUC57 into pCold TF bacterial expression vector 89
3.3.6. Bacterial expression of codon-optimized NC protein of isolate SPU415/85 and solubility study 91
3.3.7. Bacterial expression of pCold-TF-optAP92NC using IPTG induction and solubility study 95

Summary 97

CHAPTER 4: ANTIGENIC CROSS-REACTIVITY OF CCHFV NUCLEOCAPSIDS 99

4.1. Introduction 99

4.2. Methods and Materials 101
4.2.1. Serum Samples 101
4.2.2. Enzyme-linked immunoassays 102
4.2.3. IgG ELISA using recombinant antigen 103
4.2.3.1. Selection of cut-off values 105
4.2.4. Determination of antibody titers of CCHF convalescent sera using CCHFV NC ... 105

4.2.5. Stability of bacterially expressed CCHFV NC 106
4.2.6. Statistical analysis/normalization of data. 108

4.3. Results 108
4.3.1. Indirect ELISA for detection of IgG antibody using South African recombinant NC 108
4.3.2. Indirect ELISA for detection of IgG antibody directed against AP92 NC 110
4.3.3. Determination of antibody titers for serum samples from CCHF survivors 112
4.3.4. Stability of bacterially expressed CCHFV NC 113

Summary 116

CHAPTER 5: DISCUSSION 118

REFERENCES 123

APPENDIX 1: PAIRWISE DISTANCES (NUCLEIC ACID AND AMINO ACID) 141

APPENDIX 2: ELISA RAW DATA 144

APPENDIX 3: LIST OF FIGURES, LIST OF TABLES, LIST OF ABBREVIATIONS AND ACRONYMS 147

List of Figures 147

List of Tables 151

Abbreviations and acronyms 153

APPENDIX 4: ABSTRACT OF PRESENTATION AT FACULTY OF HEALTH SCIENCES, UNIVERSITY OF THE FREE STATE FACULTY FORUM AUGUST 2012 156

APPENDIX 5: ETHICS APPROVAL AND PUBLICATIONS 157

SUMMARY 158

OPSOMMING 160

Chapter 1: Literature review

1.1. Introduction and history

Crimean haemorrhagic fever was first described on the Crimean Peninsula in the mid-1940s where agricultural workers were exposed to ticks (Hoogstraal, 1979). In 1956, Congo virus was isolated from a patient in the former Belgian Congo (currently the Democratic Republic of Congo). The causative agent was isolated in suckling mice in 1967 and studies on physicochemical characteristics, morphology and morphogenesis showed that Congo virus and the virus causing Crimean haemorrhagic fever were indistinguishable. The names were subsequently combined and the virus is now known as Crimean-Congo haemorrhagic fever virus (CCHFV) (Casals, 1969; Korolev *et al.*, 1976; Donets *et al.*, 1977; Hoogstraal, 1979).

The first case of CCHF in South Africa was identified in 1981 (Gear *et al.,* 1982); however, in animal surveillance studies, antibody against CCHFV was detected in stored cattle sera that had been collected prior to 1981 (Shepherd *et al.,* 1987), confirming that CCHFV was present in South Africa long before it was identified (Shepherd *et al.,* 1987; Burt *et al.,* 2007).

CCHFV is prevalent in Africa, Asia and Eastern Europe. It is a tick-borne virus transmitted by ticks belonging to the *Hyalomma* genus (Hoogstraal, 1979). The virus is a member of *Nairovirus* genus and the family *Bunyaviridae* (Bishop *et al.,* 1980; Tignor *et al.,* 1980; Clerx *et al.,* 1981). Other genera in the family *Bunyaviridae* include, *Orthobunyavirus, Hantavirus, Phlebovirus* and *Tospovirus* (Bishop *et al.,* 1980; Eley *et al.,* 1989). The *Nairovirus* genus consists of 34 viruses (Morikawa *et al.,* 2007; Weber & Mirazimi, 2008) of which three are known to be human pathogens, CCHFV, Nairobi sheep disease and Dugbe viruses. Nairobi sheep disease virus primarily infects sheep and goats while Dugbe virus causes a mild febrile illness and thrombocytopenia in humans. The viruses are divided into seven different serogroups based on antigenic relationships: The CCHF serogroup consists of CCHFV, Hazara virus from Pakistan and Khasan virus from the former Union of Soviet Socialist Republics (USSR) (Bishop *et al.,* 1980). Phylogenetic

studies, based on the small (S), medium (M) and large (L) segments, have further grouped CCHFV into seven groups/genotypes based on geographic location (Deyde *et al.*, 2006; Mild *et al.*, 2010).

1.2. Structural characteristics and physicochemical properties

Virion morphology of CCHFV was first studied using preparations from newborn white mouse brains (Korolev *et al.*, 1976). CCHFV is the prototype virus of the genus *Nairovirus*. *Nairoviruses* are enveloped, sphere-shaped viruses that are approximately 90-120 nm in diameter (Donets *et al.*, 1977; Bishop *et al.*, 1980; Clerx *et al.*, 1981) and are structurally and antigenically distinct from other genera of the *Bunyaviridae* (Donets *et al.*, 1977; Bishop *et al.*, 1980; Clerx *et al.*, 1981). The virus consists of surface glycoproteins embedded in a lipid bilayer that is approximately 5-7 nm thick (Bishop *et al.*, 1980). Other viral proteins include the nucleocapsid and the RNA-dependent RNA polymerase as depicted in Figure 1.

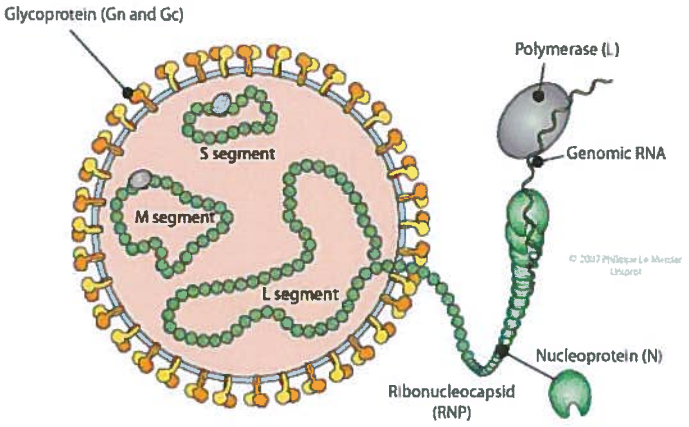


Figure 1: Crimean-Congo haemorrhagic fever virus structure and proteins: nucleocapsid, glycoproteins (embedded in a lipid bilayer), RNA dependent RNA polymerase. The tripartite genome is also illustrated. (Swiss Institute of Bioinformatics. ViralZone www.expasy.ch/viralzone)

The major structural protein of the virus, the 54 kDa nucleocapsid (NC) is known to be the most antigenic protein. Recent data shows that the NC is targeted to the

perinuclear region of infected cells in the absence of native RNA segments and that the targeting is actin filament dependent (Andersson *et al.,* 2004). The structure of the CCHFV NC has been reported: a racket-shaped overall structure that features two major parts: a "head" domain (between amino acid position 1-180 and 300-482) and a "stalk" domain (between amino acid position 181-299) depicted in Figure 2. Additionally, the protein exhibits endonuclease activity (Guo *et al.,* 2012).

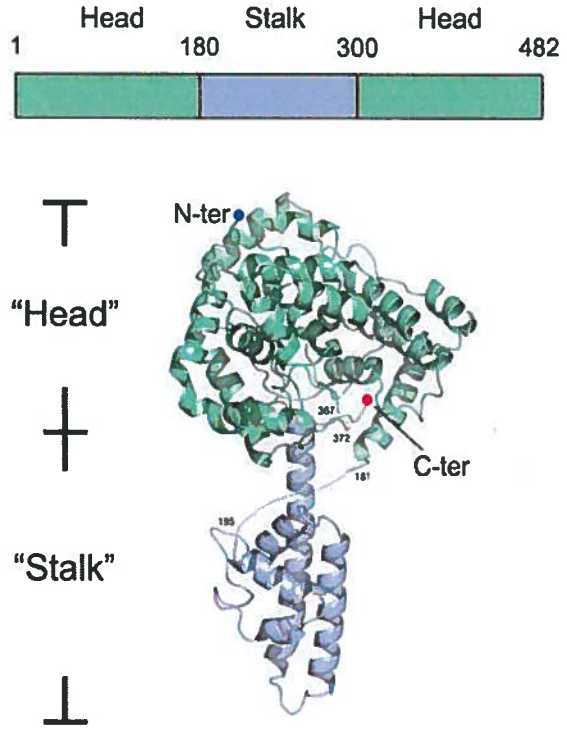


Figure 2: Crystal structure of Crimean-Congo haemorrhagic fever virus nucleocapsid revealing a "head" and "stalk" domain. The head domain is between amino acid position 1-180 and 300-482, while the stalk domain is between 180 and 300 (Guo *et al.,* (2012)).

M segment derived mRNA translates a large precursor protein, which is thought to be co-translationally cleaved into pre-glycoproteins, 140 kDa PreGn and 85 kDa PreGc at the N-terminal and the fifth hydrophobic stretch of the precursor by signalase in the endoplasmic reticulum (ER). The precursor proteins result in the glycoprotein located at the N-terminus of the M segment (Gn), previously known as G2, with a molecular weight of 37 kDa and is smaller than the 75 kDa glycoprotein at the carboxyl-terminus (Gc) of the M segment, which was previously named G1. The

glycoproteins likely influence the vertebrate and tick host usage, cell tropism of the virus and pathogenicity of the virus (Sanchez *et al.,* 2002; Haferkamp *et al.,* 2005) and are also targets for neutralizing antibodies (Ahmed *et al.,* 2005). *Bunyaviruses* bud from Golgi membranes and the budding site seems to be defined by the retention of the glycoproteins Gn and Gc at that particular site. Gn is localized to the Golgi compartment, whereas Gc is found in the ER (Haferkamp *et al.,* 2005). The L segment encodes for a large polypeptide, approximately 200 kDa in size, known as the RNA-dependent RNA polymerase (Marriot & Nuttall, 1996).

CCHFV morphogenesis includes the following main stages: assemblage of virions on the membranes of the Golgi complex, transportation of the virus-containing vacuoles to the cell surface and virus release by exocytosis (Korolev *et al.,* 1976; Andersson *et al.,* 2004). Infectivity of the virus is destroyed by a reduction in pH, which occurs in tissue after death, detergents and specific proteoloytic enzymes. CCHFV is also sensitive to lipid solvents (Donets *et al.,* 1977; Bishop *et al.,* 1980). Infectivity of the virus in cell culture is destroyed by boiling or autoclaving (Hoogstraal 1979; Clerx *et al.,* 1981).

1.3. Molecular biology of CCHFV

The single stranded, negative sense, segmented RNA genome of CCHFV consists of a small (S) medium (M) and large (L) segment of approximately 1672, 5360 and 12 200 basepairs (bp) respectively (Elliott, 1990). The L segment encodes the viral polymerase, the M segment encodes precursors for glycoproteins, Gn and Gc and the S segment encodes the viral NC (Clerx *et al.,* 1981; Marriot & Nuttall, 1996). No additional non-structural proteins have been identified in L or S segments (Hewson, 2007), however, there is evidence of a cell associated non-structural protein that was generated from the full-length M polyprotein, both by transient plasmid expression and by virus (Altamura *et al.,* 2007). The S, M and L segments (depicted in Figure 1) of the tripartite genome each contain an open reading frame that is flanked by non-coding regions. The first eight to thirteen nucleotide bases at the 3' ends of all three RNA segments have a sequence that is conserved in the viruses of the genus Nairovirus, with a complementary consensus sequence at the 5' end. The L

segment encodes the RNA-dependent RNA polymerase, the M segment encodes the mature glycoproteins Gn and Gc and the S segment encodes the NC (Flick & Whitehouse, 2005).

The S segment consists of approximately 1672 bp, and the open reading frame (ORF) encodes a 482 amino acid NC. During evolution, fragmental exchange has likely occurred in the S segment 3' non coding region as a result of homologous recombination (Deyde *et al.,* 2006).

The M segment is on average 5300 bp in length. The viral glycoproteins, named according to their relative proximity to the amino and carboxy terminus of the M segment, are synthesized as a polyprotein precursor that undergoes proteolytic cleavage resulting in the mature glycoproteins, Gn and Gc. The open reading frame of the glycoprotein is illustrated in Figure 3, including Gn and Gc; tetrapeptides RRLL and RKPL are the 5' proteolytic cleavage sites for Gn and Gc respectively. Sanchez *et al* also found that the RKLL tetrapeptides occurred in protein sequences from all the CCHFV isolates analysed and identified RKLL and RSKR as potential 3' cleavage sites. The position of the cleavage sites are shown according to base number (using isolate 10200 as reference) in Figure 3.

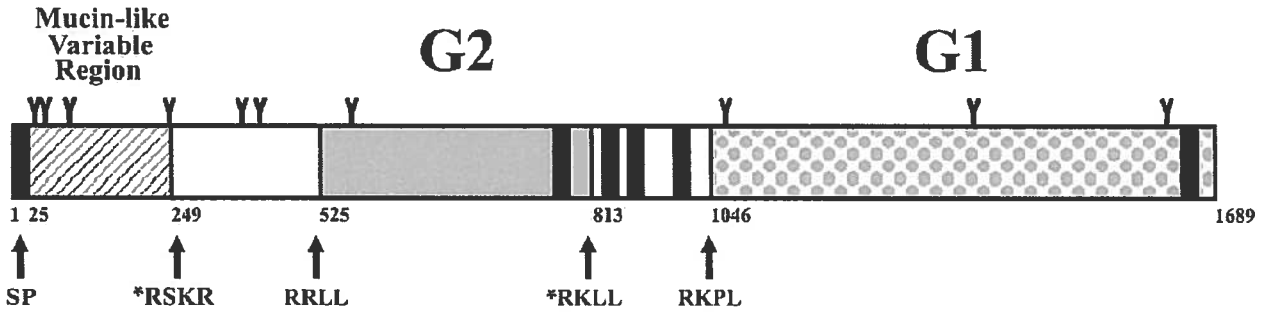


Figure 3: Open reading frame of polyprotein which is cleaved into Gn and Gc. SP - Signal peptide. RKPL and RRLL are proteolytic cleavage sites. RSKR and RKLL are also proposed cleavage sites for Gn and Gc, respectively. (Sanchez *et al.,* (2002))

Sequencing of full-length L segment showed that the segment is nearly twice the size (12164 nucleotides) of those of other genera in the family *Bunyaviridae*. The L

segment of the CCHF virus encodes a large protein with 3944 amino acids in which an ovarian tumor (OTU)-like protease motif is found at N terminus followed by zinc finger motif and helicase domain, and RNA polymerase catalytic domain located in the central region. This suggests that both helicase and polymerase activity stem from one polyprotein (Kinsella *et al.,* 2004).

1.4. Epidemiology

Humans acquire CCHFV by contact with ticks, infected tissue or blood from viremic livestock or other humans with CCHFV infection (Hoogstraal, 1979). The disease caused by the virus can exhibit a high mortality rate (Blackburn *et al.,* 1987) of up to 50% (Swanepoel *et al.,* 1989; Ergonul *et al.,* 2006). Cattle and scrub hares can act as amplifying hosts of CCHFV by transmitting the virus to ticks during viremia (Swanepoel *et al.,* 1983; Shepherd *et al.,* 1987).

The geographic distribution of the principal vector, the *Hyalomma* tick, determines the presence of CCHFV and its potential as a threat. CCHFV has caused outbreaks in South Africa , Pakistan , Kosovo , Turkey , Mauritania , United Arab Emirates (UAE) , Iran, Saudi Arabia , Sudan , India and Afghanistan (Joubert *et al.,* 1985; Jamil *et al.,* 2005; Thomas *et al.,* 2012; Karti *et al.,* 2004; Nabeth *et al.,* 2004; Khan *et al.,* 1997; Sharifi-Mood *et al.,* 2009; El-Azazy & Scrimgeour, 1997; Aradaib *et al.,* 2010; Patel *et al.,* 2011; Mourya *et al.,* 2012; Mustafa *et al.,* 2011). Incidences of naturally acquired human infection have been recorded in many different countries in Africa, Asia, Russia and the Balkans. These countries include: the former Soviet Union, Bulgaria, South Africa, Congo, Mauritania, Burkina Faso, Tanzania, Senegal, Iraq, the United Arab Emirates, Saudi Arabia, Oman, Pakistan, Albania, Yugoslavia, Kazakhstan, Turkey, Greece and Kenya. Serological surveys have confirmed the presence of CCHFV in Greece, India, Egypt, Portugal, Hungary, France and Benin (Al-Tikriti *et al.,* 1981; Morrill *et al.,* 1990; Williams *et al.,* 2000; Papa *et al.,* 2002a; Dunster *et al.,* 2002; Papa *et al.,* 2002b; Gunes *et al.,* 2009; Tall *et al.,* 2009; Knust *et al.,* 2012).

CCHFV infection is mainly an occupational risk, as agricultural workers, abattoir workers and veterinarians are exposed to ticks or potentially infected animals (Swanepoel *et al.,* 1987; Mourya *et al.,* 2012). Healthcare workers are also at risk if diagnosis is delayed.

1.5. Life cycle of the virus in nature

CCHFV circulates in an enzootic tick-vertebrate-tick cycle and humans are the only species where CCHFV causes a disease (Shepherd *et al.,* 1989a). CCHFV persists in the ixodid or argasid tick; mainly from the *Hyalomma* genus, the natural reservoir of the virus, through its life stages from larvae to nymph to adult, referred to as transtadial transmission (Shepherd *et al.,* 1989a; Gonzalez *et al.,* 1992; Williams *et al.,* 2000). Virus strains have also been isolated from field-collected eggs and unfed immature stages which suggests transovarial transmission. Ixodid ticks have three stages or instars in their life cycle, larvae, nymphs and adults. The immatures attach to a host such as ground frequenting birds or small mammals like scrub hares and feed before molting to the next stage and feeding again, either on the same host or on separate hosts. Many *Hyalommas* behave as two-host ticks: the larvae remain attached and molt *in situ* before feeding as nymphs on the same host, and then detach to molt into adults which feed on a second host (Shepherd *et al.,* 1989a).

The most common vector is *Hyalomma marginatum marginatum (H. m. marginatum).* Migrating bird species are known to be infested by *H. m. marginatum* in Europe and Asia and *H. marginatum rufipes, H. truncatum* and other *Hyalomma* spp in Africa (Hoogstraal, 1979; Turell, 2007). Three species of *Hyalomma* occur in South Africa, *H. glabrum*, *H. m. rufipes*, and *H. truncatum* (Burt *et al.,* 2007). Changes in climatic conditions are a factor for tick reproduction. CCHF outbreaks have occurred during favorable environmental conditions that benefit *Hyalomma spp* ticks (Hoogstraal, 1979).

Cattle, sheep and goats generally develop an inapparent or subclinical disease, but viremia persists for approximately one week, supporting transmission of the virus to uninfected ticks or humans that are in contact with infected tissues. During this week, the virus may be transmitted to humans who have close contact with virus-

contaminated blood or tissue (Wilson & Digoutte, 1989). Smaller wildlife species act as hosts for the immature stages of tick vectors (Hoogstraal, 1979). Seroepidemiological studies have shown antibody against CCHFV in cattle and sheep (Umoh *et al.,* 1983; Lotfollahzadeh *et al.,* 2011). Ostriches are the only birds that are susceptible to virus infection and outbreaks have occurred in humans after slaughtering ostriches (Swanepoel *et al.,* 1998). Aerosol transmission has been suspected but not documented and transmission from mother to child has been reported (Ergonul *et al.,* 2010).

1.6. Signs and symptoms

A study by Swanepoel and colleagues (1989), using data from 50 patients was the most comprehensive study on the clinical aspects of CCHF. The disease has a sudden onset, with violent headaches, dizziness, photophobia, myalgia, fever, rigors and chills. In some cases nausea, sore throat, vomiting, abdominal pain and diarrhea would occur. Confusion and aggression were common in the first few days of illness. Fever was intermittent and patients were afebrile and depressed. Hepatomegaly occurred in half of the patients. Petechial rash on the trunk and limbs appear on day three to six of illness and bruises on the upper arms were also common. Viremia was present from day one to day 12 of illness (Swanepoel *et al.,* 1989).

1.7. Clinical pathology

A CCHFV infection usually results in severe haemorrhagic fever. The disease progression is rapid and can be divided into four stages as shown in Figure 4: incubation, pre-haemorrhagic, haemorrhagic and convalescent phase (Ergonul, 2006). The incubation period differs depending on viral dose and route of exposure and ranges between three to seven days.

The pre-haemorrhagic phase is characterized by a sudden onset of fever, headache, myalgia and dizziness and vomiting which lasts for approximately three days. The

haemorrhagic fever phase is also short, develops rapidly, characterized by epistaxis and bleeding from the gastrointestinal, urinary and respiratory tracts. Haemorrhagic manifestations range from petechia and large haematomas on the skin and mucous membranes to fulminant haemorrhage (Gear, 1989). Disseminated intravascular coagulopathy (DIC) is frequently an early feature of infection (Swanepoel *et al.*, 1989; Jamil *et al.*, 2005). The convalescent phase starts 10-20 days after onset of illness. Weakness, nausea, loss of hair, difficulty breathing, loss of hearing and memory has been reported. Thrombocytopenia is also a consistent feature in a CCHFV infection, as well as raised levels of aspartate aminotransferase (AST) and alkaline aminotransferase (ALT) (Ergonul, 2006).

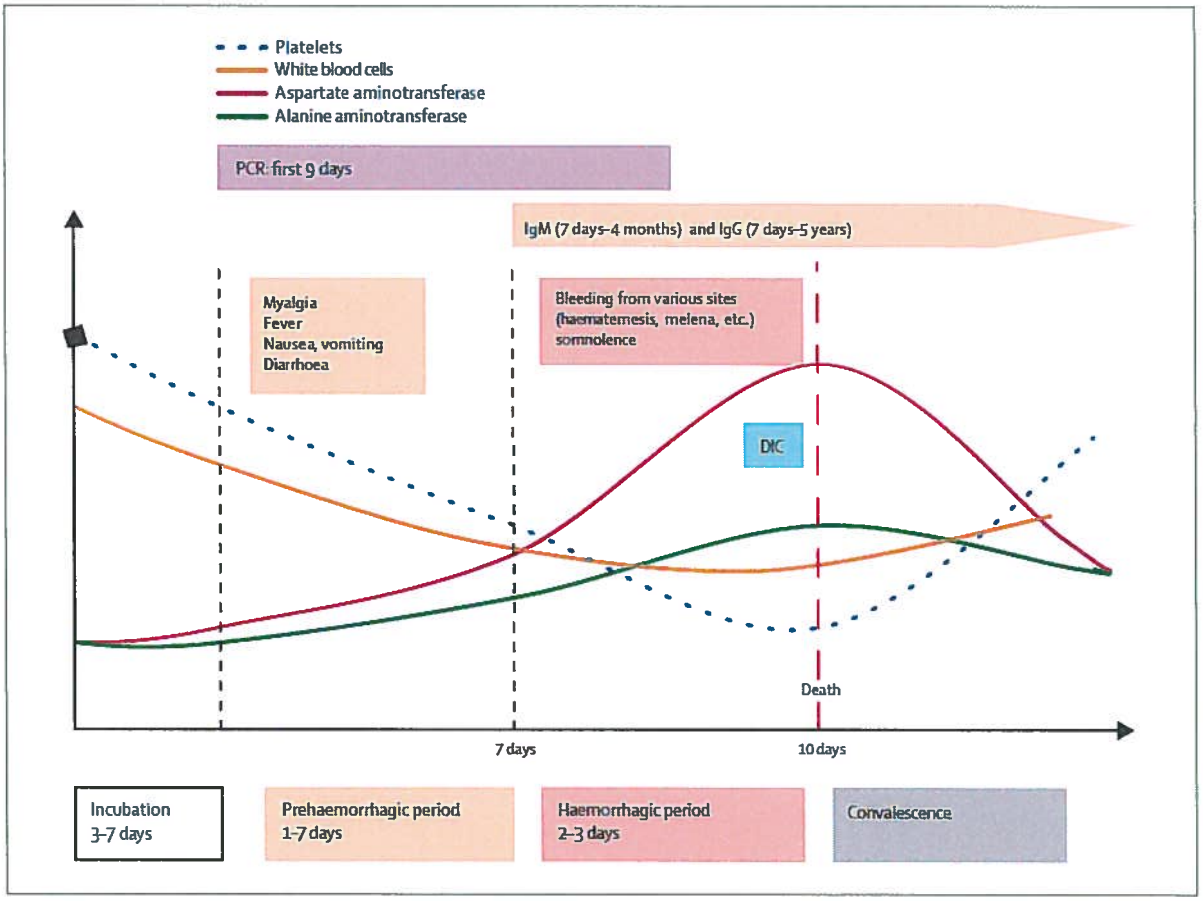


Figure 4: Clinical course of CCHF: The disease progression can be subdivided into four different stages, with different symptoms; incubation, pre-haemorrhagic, haemorrhagic and then convalescence phase. Laboratory analyses demonstrate the elevated liver enzymes aspartate and alanine aminotransferase (AST and ALT) as well as low platelets count (PLTs) in the fatal cases. (Ergonul, (2006)).

Thrombocytopenia is a consistent feature and platelet counts are extremely low at the early stages of illness. In fatal infections; the prothrombin ratio, activated partial thromboplastin time, thrombin time, fibrinogen and fibrin degradation products are highly abnormal (Swanepoel *et al.,* 1989).

The mortality of CCHF is ~30% (Gear, 1989) and has been documented to rise up to 75% (Jamil *et al.,* 2005). Viral loads are linked to the severity of the disease: viral loads in patients with fatal outcomes were a mean of 3 $\log_{10}$ higher than viral loads in survivors of CCHF infection (Cevik *et al.,* 2007). Death usually follows on days 5-14 of illness (Burt *et al.,* 2007).

1.8. Pathogenesis of the virus

The pathogenesis of CCHF infections is not well documented, due to factors such as: sporadically occurring infections in areas where facilities are limited, the virus requires BSL-4 containment for handling and also lack of an animal model for the disease (Weber & Mirazimi, 2008). The glycoproteins likely influence the vertebrate host range and cell tropism of the virus and are the sites where neutralizing antibodies bind (Ahmed *et al.,* 2005). CCHFV attaches to host cell receptors via the glycoproteins, enters the host by endocytosis and replicates in the cytoplasm (Schmaljohn & Hooper, 2001; Chinikar *et al.,* 2012). Infection of the endothelium can occur indirectly by viral factors or virus-mediated host-derived soluble factors that cause endothelial activations and dysfunction, and/or directly by virus infection and replication in endothelial cells. Endothelial damage contributes to haemostatic failure by stimulating platelet aggregation and degranulation, with consequent activation of the intrinsic coagulation cascade. Vascular damage can be induced by immunological mechanisms and/or by direct infection of the vascular tissue. Impairment of endothelial cell function can cause a wide range of vascular effects that lead to changes in vascular permeability or haemorrhage (Haller *et al.,* 2006; Connolly-Andersen *et al.,* 2011; Ergonul, 2012).

1.9. Diagnosis

Laboratory diagnosis is commonly performed on whole blood, plasma, serum or biopsy material (Swanepoel *et al.,* 1989). Due to the potential of CCHFV to cause nosocomial infections, an early diagnosis is important. CCHF can be diagnosed using viral isolation, detection of the viral genome using RT-PCR, antigen detection using enzyme-linked immunosorbent assay (ELISA) and/or detection of specific antibodies (Zeller, 2007; Vanhomwegen *et al.,* 2012). Due to the biohazardous nature of the virus and the associated fatality rate, handling and culture of the virus requires BSL-4 facilities.

1.9.1. Virus isolation

The current gold standard diagnostic test for CCHF is virus isolation through *in vitro* cell culture or mouse inoculation (Vanhomwegen *et al.,* 2012). CCHFV has been isolated by intracranial inoculation of newborn suckling mice (Hoogstraal, 1979). Virus can be isolated using cell lines such as LLC-MCK2 (rhesus monkey, *Macaca mulatta*), Vero (African green monkey kidney, *Cercopithecus aethiops*), BHK-21 (baby hamster kidney) and SW-13 (human small cell carcinoma of adrenal cortex). Vero cell lines are more routinely used however isolation in cells can take two to five days and cell cultures lack sensitivity and require relatively high viremia which is present only during the first 5 days of illness (Ergonul, 2006). CCHFV causes little or no cytopathic effects (CPE) and infection of cells must be determined by immunofluorescence assay. Virus isolation is more frequently performed by intracerebral inoculation of one day-old mice; however mice take 7-9 days to succumb to infection (Hoogstraal, 1979).

1.9.2. Molecular assays

The development of molecular assays for detection of viral nucleic acid has provided a rapid and sensitive method for detection of CCHF infections during the acute stage of illness (Burt *et al.,* 1998). Extraction of viral RNA using a detergent-based protocol can render the sample non-infectious and safe to use at BSL-2 level (Zeller,

2007). RT-PCR is highly specific, sensitive and rapid and can be used to quantify viral load (Papa *et al.,* 2007). Real-time assays have been developed (Duh *et al.,* 2006; Wölfel *et al.,* 2007; Bodur *et al.,* 2010; Ibrahim *et al.,* 2011; Atkinson *et al.,* 2012), with a higher sensitivity and specificity than conventional PCR. However, these assays may be affected by the high genetic diversity among CCHFV isolates and thus detect location-specific strains of CCHFV (Duh *et al.,* 2006).

1.9.3. Serological assays

Two serological tests, ELISA and immunoflourescence assays, are currently the most frequently assays for detection of IgM and IgG antibodies during the convalescent stage of illness (Shepherd *et al.,* 1989b). IgM and IgG antibodies are demonstrable on days seven to nine of illness in patients with non-fatal infections and in some cases earlier (Swanepoel *et al.,* 1989b). Specific IgM becomes undetectable in most patients approximately four to six months post-infection, but has been detected up to one to two years later, and IgG has been detected more than five years after onset and is likely lifelong. Recent or current infection is confirmed by demonstrating seroconversion or a fourfold increase in antibody titer in paired serum samples or detection of IgM antibody. IgM antibody capture ELISA can detect IgM in a single sample. Detection of viral antigen has been described, using reversed passive haemagglutination tests and ELISA (Shepherd *et al.,* 1988). Saijo *et al.,* (2005a), developed a capture ELISA using a monoclonal antibody directed against the NC. Immunogenic proteins engineered using bacterial or eukaryotic systems can also be used in serological assays to detect antibody directed against CCHFV. Recombinant CCHFV NC have been successfully used to develop ELISA to diagnose CCHFV infections (Marriott *et al.,* 1994; Saijo *et al.,* 2005b; Garcia *et al.,* 2006; Samudzi *et al.,* 2012; Dowall *et al.,* 2012), demonstrating that NC based assays can be as specific as using antigen derived from whole virus.

1.10. Treatment, prevention and control

There is currently no treatment for CCHF infection. Ribavirin, a synthetic purine nucleoside analogue with a modified base and D-ribose sugar, has been tested for inhibiting CCHF as post-exposure prophylaxis (Cevik *et al.,* 2008; Koksal *et al.,* 2010; Ascioglu *et al.,* 2011). Ribavirin inhibits replication of a wide range of DNA and RNA viruses, including CCHFV (Watts *et al.,* 1989). Supportive and replacement therapy are also important for treatment (Swanepoel *et al.,* 1989). Supportive therapy includes the administration of thrombocyte suspensions and fresh frozen plasma (containing clotting factors and coagulants) (Ergonul, 2008).

Attempts to develop vaccines against CCHF date back to the 1960s in the former USSR where CCHFV was endemic. Researchers of the Soviet Institute of Poliomyelitis and Viral Encephalitides developed an experimental CCHF vaccine using inactivated brain tissue from infected newborn laboratory mice (Keshtkar-Jahromi *et al.,* 2011). The CCHFV vaccine was licensed for use in military personnel but there is no published data on the efficacy of this vaccine. Recently, Mousavi-jazi and colleagues (2012), studied the cellular and humoral immune response in healthy individuals after administration of an inactivated Bulgarian vaccine. Vaccinees demonstrated anti-CCHFV-specific T-cell activity. Repeated doses were necessary to show neutralizing activity of antibodies against CCHFV (Mousavi-Jazi *et al.,* 2012). The Bulgarian vaccine strain, first described by Papa *et al.,* (2011), is an inactivated suckling mouse brain-derived vaccine.

1.11. Phylogenetic diversity

In 1969, when the virus was first named as CCHFV, the antigenic structures of the viruses from various geographic regions were thought to be indistinguishable (Casals *et al* 1969). However, the development of nucleic acid sequence analysis techniques, including phylogenetics, illustrated high genetic diversity as reported by studies of the S, M and L genes (Hewson *et al.,* 2004a; Burt & Swanepoel, 2005; Deyde *et al.,* 2006).

There have been a limited number of phylogenetic studies based on the full-length S, M and L genes of geographically distinct isolates (Hewson *et al.*, 2004b; Deyde *et al.,* 2006). Genetic analysis using partial sequence data has been determined more extensively. Tree topologies based on complete sequence data and partial sequence data show close relationships within segments however group switching has been shown to occur when comparing tree topologies from different segments particularly the M segment suggesting the occurrence of reassortment of the genes (Yashina *et al.,* 2003a; Yashina *et al.,* 2003b; Seregin *et al.,* 2004; Chinikar *et al.,* 2004; Papa *et al.,* 2005; Burt & Swanepoel, 2005; Sun *et al.,* 2009; Ozkaya *et al.,* 2010; Chamberlain *et al.,* 2005; Anagnostou & Papa, 2009). Data from these studies concur that the CCHFV isolates can be subdivided into seven genetically distinct groups), also referred to as genotypes (Mild *et al.,* 2010).

Based on S segment sequence data for 32 isolates, isolates were divided into the following groups: group I West Africa, group II DRC, group III South Africa/West Africa, group IV Asia/Middle East, the European isolates clustered in group V that excluded the Greek strain AP92, which branched off singularly (group VI) and was the most diverse strain compared to the rest of CCHFV isolates globally (Deyde *et al.,* 2006; Papa *et al.,* 2010; Anagnostou & Papa, 2009).

The groupings were recently expanded by Mild and co-workers (2010) to include 168 partial CCHF S segment sequences and the isolates further subdivided into genotypes as follows: genotype I included isolates from Asia and the Middle East; genotype 2 included isolates from Asia; genotypes 3, 5 and 7 included isolates from Africa; genotypes 4 and 6 comprised of isolates from Europe including AP92 from Greece which formed a distinct lineage from the Turkish isolates.

Different genotypes were shown to circulate in similar geographic regions and in other areas similar genotypes were shown to circulate in geographically distinct regions. Overall CCHFV circulates within and between continents and it is likely that genetic diversity within regions is the result of multiple introductions occurring over time as a result of animal (and therefore tick) movement due to livestock trade and bird migration (Hewson *et al.*, 2004b; Deyde *et al.,* 2006; Anagnostou & Papa, 2009).

A high degree of genetic diversity has been identified within and between the various groups or genotypes of CCHFV. The S segment is the most diverse with nucleotide differences of approximately 30 to 43% reported between groups although amino acid sequences were more homologous with only 3 to 5% differences (Burt *et al.*, 2009). RNA viruses with segmented genomes are capable of genomic reassortment of the segments, resulting in genetically distinct viruses. Evidence of reassortment events have been reported and indicate the existence of additional mechanisms that facilitate genetic diversity (Hewson *et al.,* 2004b; Burt *et al.,* 2009). Studies have also suggested that the S and L segment display co-evolutionary patterns (Chamberlain *et al.,* 2005). Phylogenetic analyses based on amino acid sequence data, could potentially provide insight into antigenic relationships between isolates. The movement of CCHFV isolates within and between continents could impact on the development of targeted recombinant antigens and sub-unit vaccines.

1.12. Neutralizing antibodies and antigenic relationships

An earlier study using virus neutralization assays demonstrated minor antigenic differences among some CCHFV strains that originated in widely separated areas globally (Tignor *et al.,* 1980). Information on whether significant antigenic differences exist between isolates or whether conserved neutralizing epitopes can be found is important for vaccine development (Ahmed *et al.,* 2005).

Epitope mapping provides useful information on antigenic relationships between CCHFV strains. A panel of monoclonal antibodies (mAbs) that react to Gn and Gc of isolate IbAr10200 demonstrated broadly reactive and group-specific neutralizing epitopes on Gn and Gc (Ahmed *et al.,* 2005). The study showed that there were antigenic variations between the M segment groups; however, several neutralizing epitopes in both Gn and Gc proteins were conserved among the strains that had been examined. The viral glycoproteins show a higher antigenic variability than the NC as a result of immune selection and the adaptation needed to efficiently bind to and enter diverse cell types. Results of the study infer that significant antigenic differences do exist between CCHFV strains that often do not correlate with genotypic characteristics or geographical location (Ahmed *et al.,* 2005).

1.13. Problem identification

There is genetic diversity between isolates of CCHFV worldwide based on nucleotide sequence data and predicted amino acid; however it is not clear if these differences influence epitopes that induce detectable antibody responses and subsequently cross-reactivity of antibodies. The identification of diversity and serological cross reactivity between strains from geographically distinct regions will influence development of assays for detection of CCHFV antibody and will influence use of monoclonal antibodies in development of novel treatments and diagnostic assays.

1.14. Aim and objectives

The aim of the study was to determine genetic and antigenic diversity among CCHFV strains based on phylogenetic relationships and antibody cross reactivity against the NC of a representative South African isolate and against the NC of a Greek isolate (AP92), the most genetically diverse CCHFV strain.

The objectives of the study included:

1. Investigation of the nucleotide diversity of the gene encoding the NC of CCHFV and the diversity of the predicted NC amino acid sequence of South African and geographically distinct isolates. The major structural protein of CCHFV, the NC, known to be the most antigenic protein, is used to study antigenic diversity.

2. Identification of isolates of CCHFV that are representative of southern African isolates based on the predicted amino acid sequence of the NC.

3. Preparation and characterization of a recombinant NC protein from a southern African isolate.

4. Preparation and characterization of a recombinant NC protein from a Greek isolate of CCHFV which is the most genetically distinct strain of CCHFV.

5. Development of an in house ELISA for detection of IgG antibody responses using recombinant antigens. Application of ELISA to screen serum samples from survivors of CCHFV to determine if there are serologically distinguishable antigenic differences.

Chapter 2: Global diversity of Crimean Congo haemorrhagic fever nucleocapsid proteins

2.1. Introduction

CCHFV is a member of the Nairovirus genus belonging to the family Bunyaviridae, which consists of diverse RNA viruses (Bishop *et al.,* 1980). The virus causes a medically significant disease with a propensity to cause nosocomial infections with a case fatality rate of up to 50% (Swanepoel *et al.,* 1989). From 1981 to 2012, over 200 cases have been reported from different parts of southern Africa (Burt *et al.*, 2007; Jansen van Vuren *et al.*, 2012). CCHFV had been present in South Africa long before the first human case was identified in 1981, with increased awareness and the availability of specific diagnostic tests likely the reasons for increased recognition of infections (Burt *et al.,* 2007).

Based on classical serological assays such as complement fixation, haemagglutination assays and neutralization assays, isolates from various geographic regions were considered to be antigenically indistinguishable (Casals, 1969). Recent development of molecular techniques for determining the nucleic acid sequence of the genome and analysis of sequence data have shown high genetic diversity (Anagnostou & Papa, 2009; Deyde *et al.,* 2006; Burt *et al.,* 2009; Burt & Swanepoel, 2005; Mild *et al.,* 2010).

Complete genome analysis performed by Deyde and colleagues, (2006), show high genetic diversity among CCHFV isolates, frequent reassortment occurring in the M segment and occasional recombination in the S segment (Hewson *et al.,* 2004b; Deyde *et al.,* 2006). Phylogenetic analyses based on partial and complete sequences of the S segment from different regions of the world have identified seven distinct genotypes (Karti *et al.,* 2004; Hewson *et al.,* 2004a; Chamberlain *et al.,* 2005; Tonbak *et al.,* 2006; Deyde *et al.,* 2006; Sun *et al.,* 2009; Burt *et al.,* 2009; Anagnostou & Papa 2009; Papa *et al.,* 2010; Mild *et al.,* 2010).

Burt & Swanepoel (2005) studied genetic diversity between 70 CCHFV isolates using a 450 bp region of the S segment and genetic diversity up to 18% was identified. In a subsequent study using nucleotide sequence data for a 436 bp

region of the S segment, 413 bp 3' region of the M segment and 754 bp region of the L segment, sequence divergence was determined using Molecular Evolutionary Genetics Analysis (MEGA) to calculate average *P* distances within groups and between groups. The average genetic *P* distances were calculated after phylogeny was constructed and base changes were observed in the portion of the S segment analysed. Nucleotide distances between groups ranged from 29.6% to 43.1% and amino acid distances ranged from 2.8% to 4.9% which indicates that the majority of the changes were synonymous (Burt *et al.,* 2009).

The segment of interest in this study is the complete ORF of the S segment, which codes for the 482 amino acid NC. The NC, which is the major structural protein of the virus, is approximately 53 966 Da in size (Marriot & Nuttall 1992). Characteristically, the protein is known for its high antigenicity and because it is the most conserved, has thus been used to prepare recombinant laboratory diagnostic reagents (Marriott *et al.,* 1994; Samudzi *et al.,* 2012; Dowall *et al.,* 2012). Antigenic variation of arboviruses is important as it provides information on the possible direction of epidemics or endemic spread. Data on antigenic relationships among CCHFV strains is limited. As the NC protein is frequently used for developing diagnostic tools it is important to determine if there are serologically detectable antigenic differences between CCHFV isolates within geographic regions and within geographically distinct regions.

The aim of this study was to investigate the global diversity of CCHFV nucleoproteins using NC sequence data for 44 isolates from Asia, the Middle East, Africa and Eastern Europe. Sequence data for the ORF of the S gene of 44 CCHFV isolates was retrieved from GenBank. In addition, a South African isolate was selected for preparation of recombinant antigen and the complete nucleotide sequence data of the ORF of the S segment was determined. Nucleotide sequence data was aligned using MAFFT alignment software. Phylogenetic trees were constructed to study the nucleotide and amino acid diversity amongst worldwide isolates.

2.2. Materials and Methods

2.2.1. Retrieval of sequence data and multiple sequence alignment

Nucleotide sequences of the S segment of 44 CCHFV isolates representing the broad geographic range of the virus were retrieved from GenBank (http://www.ncbi.nlm.nih.gov). GenBank is a database where nucleotide sequences and amino acid sequences of many species are stored and is hosted by the National Center for Biotechnology Information (Benson *et al.,* 2011). For each isolate, the ORF of the S segment, which codes for the NC, was retrieved. The isolates selected, GenBank accession number, country of origin, source and year of isolation are listed in Table 1.

Table 1: S segment sequences of Crimean Congo haemorrhagic fever virus obtained from GenBank. Isolates are listed according to country of origin. Accession numbers, strain names, sources are also listed.

Accession number	Strain	Source	Country	Year of Isolation
AY277676	Bul/HU517	Human	Bulgaria	1978
AF362080	75024	Human	China	1975
AY029157	88166	Human	China	1988
DQ211642	C-68031	Sheep	China	1968
DQ227496	CLT/TI05146	Tick	China	2005
DQ217602	CYL/TI05035	Tick	China	2005
DQ227495	CYT/TI05099	Tick	China	2005
U88413	Hy13	*Hyalomma asiaticum*	China	1968

Accession number	Strain	Source	Country	Year of Isolation
FJ562093	YL04051	Tick	China	2004
AF358784	79121M18	*Euchoreutes naso*	China	2004
DQ144418	Congo 3010	Human	Democratic Republic of the Congo	1956
DQ211638	AP92	*Rhipecephalus bursa*	Greece	1975
AF428144	955532001	Human	Kosovo	2001
DQ446212	Iran-52	n.a*	Iran	2006
DQ446213	Iran-53	n.a	Iran	2006
DQ446214	Iran-56	n.a	Iran	2006
DQ211641	ArD39554	*Hyalomma marginatum rufipes*	Mauritania	1984
U88410	10200	*Hyalomma excavatum*	Nigeria	1966
DQ211645	Oman	Human	Oman	1997
U88414	JD206	*Hyalomma anatolicum*	Pakistan	1965
AF527810	Matin	Human	Pakistan	1976
DQ211643	Drosdov	Human	Russia	1967
DQ211644	Kashmanov	Human	Russia	1967
DQ206447	ROS/HUVLV-100	Human	Russia	2002
AY277672	ROS/TI28044	Tick	Russia	2000
AF481802	STV/HU29223	Human	Russia	2000
DQ211640	ArD15786	Goat	Senegal	1972

Accession number	Strain	Source	Country	Year of Isolation
DQ211639	ArD8194	*Hyalomma truncatum*	Senegal	1969
DQ211647	SPU103/87	Human	South Africa	1987
DQ076415	SPU128/81/7	Tick	South Africa	1981
DQ076416	SPU4/81	Human	South Africa	1981
DQ211648	SPU415/85	Human	South Africa	1985
DQ211646	SPU97/85	Human	South Africa	1985
GQ862372	Sudan42008	Human	Sudan	2008
GQ862371	Sudan32008	Human	Sudan	2008
AY049083	TAJ/HU8966	Human	Tajikistan	1990
AY297692	TAJ/HU8975	Human	Tajikistan	1990
AY297691	TAJ/HU8978	Human	Tajikistan	1991
DQ211649	Turkey 200310849	Human	Turkey	2003
GQ337053	Turkeykelkit06	Human	Turkey	2006
DQ076413	Semunya	Human	Uganda	1958
AY223475	Hodzha	Human	Uzbekistan	1967
AF481799	Uzbek/TI10145	*Hyalomma asiaticum*	Uzbekistan	1985
DQ133507	Kosova Hoti	Human	Yugoslavia	2001

*n.a – not available

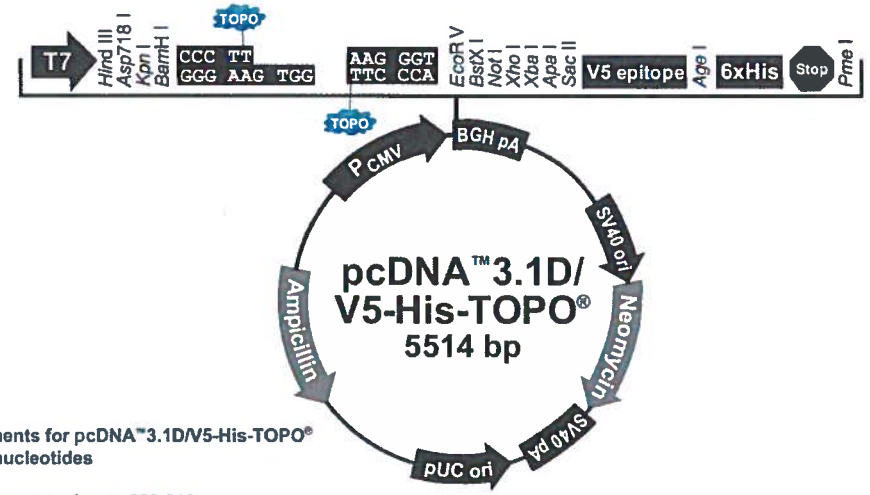
The sequence data were retrieved in FASTA format, a text based format commonly used in Bioinformatics that is compatible with most alignment programs. Multiple sequence alignments were generated using Multiple alignment using Fast Fourier Transform (MAFFT) software (Katoh *et al.,* 2002), version 6 (http://mafft.cbrc.jp). The alignment was performed using default settings for multiple sequence alignment analysis and saved in FASTA format, which is compatible for conversion in MEGA version 4 (Tamura *et al.,* 2007).

2.2.2. Determination of sequence of southern African CCHFV isolate

RNA from a South African CCHFV isolate, SPU 92/01, was available in our laboratory for preparation of recombinant antigen using mammalian expression system pcDNA3.1D V5/His/TOPO directional TOPO expression kit (Invitrogen, Maryland, USA). The vector map, highlighting important features, is shown in Figure 5.

Before expression and characterization of the recombinant protein, the ORF of the gene encoding the NC protein was cloned into the TOPO vector and the nucleotide sequence was determined. The sequence data was included in the phylogenetic analysis to determine if the selected isolate was genetically similar to known South African isolates and therefore, if the NC is representative of South African isolates.

Previously sequenced isolates, included in the phylogenetic analyses, designated SPU415/85 and AP92 respectively, were chosen for expression of bacterially expressed NC. The amino acid differences of CCHFV NC of isolates SPU415/85 and AP92 were determined, thereafter expression and characterization of the proteins was performed.



Comments for pcDNA™3.1D/V5-His-TOPO®
5514 nucleotides

CMV promoter: bases 232-819
T7 promoter/priming site: bases 863-882
TOPO® recognition site 1: bases 930-934
Overhang sequence (complementary strand): bases 935-938
TOPO® recognition site 2: bases 939-943
V5 epitope: bases 1011-1052
Polyhistidine (6xHis) tag: bases 1062-1079
BGH reverse priming site: bases 1102-1119
BGH polyadenylation signal: bases 1108-1332
SV40 early promoter and origin: bases 1833-2142
Neomycin resistance gene: bases 2217-3011
SV40 early polyadenylation signal: bases 3189-3319
pUC origin: bases 3700-4373 (complementary strand)
Ampicillin (*bla*) resistance gene: bases 4518-5378 (complementary strand)
bla promoter: bases 5379-5477 (complementary strand)

Figure 5: Vector map of pcDNA3.1 directional V5/His/TOPO mammalian expression vector showing important features on the vector. The multiple cloning site, restriction enzyme sites and promoter sites are indicated (Invitrogen, Maryland, USA).

2.2.2.1. Viral RNA and cDNA synthesis

CCHF viral RNA was available for a South African isolate (designated SPU 92/01) obtained from a patient. The RNA was extracted from infected Vero cells in a previous unrelated study and kindly supplied by Prof Paweska from the Centre for Emerging and Zoonotic diseases, National Institute for Communicable Diseases (NICD), Johannesburg.

Viral RNA was transcribed to cDNA using SuperScript™ III Reverse Transcriptase (RT) (Invitrogen, Maryland, USA) which is an engineered version of Moloney Murine Leukemia virus (M-MLV) RT with reduced RNase H activity and increased thermal stability. Each reaction consisted of: 1 µl random primers, 1 µl dNTP mix (10 mM each dATP, dGTP, dCTP and dTTP), 2 µl RNA and 9 µl nuclease-free H_2O. The

mixture was heated at 65°C for 5 minutes (min) and then incubated on ice for 1 min. Contents in the microcentrifuge tube were briefly centrifuged and 4 µl First-Strand buffer (5x), 1 µl 0.1 M DTT, 1 µl RNase inhibitor and 1 µl Superscript III RT. The tubes were incubated at 25°C for 5 min, followed by 50°C for 60 min and 70°C for 15 min using Gene Amp PCR system 9700 thermocycler (Applied Biosystems, New York, USA). RNA complementary to the cDNA was removed by adding 1 µl of RNase H (2 units) to the reaction and incubating at 37°C for 20 min.

2.2.2.2. Primer design and Phusion High-Fidelity PCR

A primer pair was identified to amplify the entire ORF of the S segment of the SA CCHFV isolate encoding the NC. The identified primers were modified to include the Kozak sequence at the 5' end (underlined in Table 2) of the ORF and to remove the stop codon at the 3' end. The sequences of the primers are shown in Table 2. The primers were selected to contain approximately 40%- 60% G/C content with similar Tm values, which were calculated using the Biomath calculator on the Promega website (www.promega.com/biomath). The primers were designated MAM S F (forward) and MAM S R (reverse).

Table 2: Nucleotide sequences of primers designed for amplification of the CCHF NC. Positions are indicated relative to reference isolate CCHF SPU 415/85 Accession No. U88415. Kozak sequence included in the forward primer is underlined.

Primer	Nucleotide sequence 5'-3'	Genomic position	Tm
MAM S F	CACCATGGAAAACAAAATTGAGGTGAATAAC	61-82	53
MAM S R	GATAATGTTAACACTGGTGGCATTG	1501-1486	52

PCR amplification was performed using Phusion® HotStart II High Fidelity (HF) DNA polymerase (Thermo Scientific, Illinois, USA). The reaction mixture was set up as follows: 33.5 µl nuclease free H_2O, 10 µl HF buffer (5x), 1 µl dNTPs (10 mM each), 1 µl MAM S F (20 pmol), 1 µl MAM S R (20 pmol), 2 µl template DNA (from cDNA synthesis), 1.5 µl DMSO and 0.5 µl (0.02 U/µl) of HotStart DNA polymerase.

The reactions were cycled in a Gene Amp PCR system 9700 thermocycler (Applied Biosystems, New York, USA). Initial denaturation was done at 98°C for 30 seconds (s), followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 48°C for 30 s, elongation at 72°C for 45 s, and one final elongation cycle at 72°C for 10 min. The samples were held at 4°C.

2.2.2.3. Agarose gel electrophoresis

Electrophoresis of a 10 µl aliquot of the PCR products was done using a 1% agarose gel prepared in Tris-acetate-EDTA (TAE) buffer (pH 8.5) containing ethidium bromide (0.5 µg/ml) (Sambrook & Russel, 2001). O'GeneRuler DNA ladder mix comprising DNA fragments from 100 to 10000 bp fragments (Fermentas, Illinois, USA) was used to determine the size of the amplicons. The samples were loaded in a 6x loading dye solution containing 60 mM Tris, 10 mM EDTA, 0.02% bromophenol blue and 60% glycerol. Gel electrophoresis was performed using a BioRad PowerPac Basic system (BioRad, California, USA) for at 90 V for 1 hour. The DNA stained with ethidium bromide was visualized with a UV transilluminator.

2.2.2.4. DNA purification

The Wizard® SV Gel and PCR Clean-Up System (Promega, Wisconsin, USA) was used to purify the PCR product directly from the amplicon according to the manufacturer's instructions. This system is based on the ability of the DNA to bind to silica membranes in the presence of chaotropic salts and removes excess nucleotides, primers and enzymes. A 40 µl aliquot of the PCR product was added to an equal volume of membrane binding solution (supplied in the kit). One SV minicolumn was placed in a collection tube. The PCR product mix was transferred to the SV minicolumn assembly and incubated for 1 min at room temperature (22-25°C). The column was centrifuged at 16 000 × g for 1 min in a microfuge 16M Spectrafuge (Labnet International, New Jersey, USA). The column was washed by adding 700 µl membrane wash solution (supplied in the kit). The column was centrifuged for 1 min at 16 000 × g. The wash was repeated with 500 µl membrane wash solution and centrifugation of the column was at 16 000 × g for 5 min. An

additional spin at 16 000 × g for 5 min was done to allow evaporation of any residual ethanol. The SV minicolumn was transferred to a 1.5 ml microcentrifuge tube. The DNA was eluted in 40 µl of nuclease free water, centrifugation at 16 000 x g for 1 min, quantified and stored at -20°C.

2.2.2.5. Concentration of DNA

Nucleic acid concentration was determined using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Illinois, USA). The purity was determined based on the ratio of absorbance at 260 nm and 280 nm.

2.2.2.6. Cloning of gene encoding NC into pcDNA ™3.1D/V5-His/TOPO®

CCHFV NC gene was cloned into pcDNA ™3.1D/V5-His/TOPO® expression vector using the Expressionist Kit (Invitrogen, Maryland, USA). The pcDNA™3.1D/V5-His/TOPO® is a 5.5 kb expression vector designed to facilitate rapid directional cloning of blunt-end PCR products for expression in mammalian cells. TOPO works on the principle that Topoisomerase I from Vaccinia virus binds to duplex DNA at specific sites. The following features as indicated in Figure 5 are present on the vector: Human cytomegalovirus (CMV) immediate early enhancer/promoter for high-level constitutive expression of the gene of interest in a wide range of mammalian cells; TOPO® Cloning site for rapid and efficient directional cloning of blunt-end PCR products and the C-terminal peptide containing the V5 epitope and a polyhistidine (6xHis) tag for detection and purification of recombinant protein. Ligation of PCR product into pcDNA ™3.1D/V5-His/TOPO® expression vector reaction was set up as described in Table 3 and incubated at room temperature for 30 min.

Table 3: Ligation reaction for NC amplicon into pcDNA ™3.1D/V5-His/TOPO® expression vector.

Reagents	Volume
Fresh PCR product	2 µl
Salt Solution (1.2 M NaCl; 0.06 M $MgCl_2$)	1 µl
TOPO® vector	1 µl
Total	**4 µl**

2.2.2.7. Transformation of chemically competent TOP10 cells

One Shot® chemically competent TOP10 cells (supplied in the pcDNA3.1D/V5-His expressionist kit) were transformed using a heat-shock method. Chemically competent TOP 10 cells were thawed on ice. The ligation mix was added to the cells and incubated on ice for 20 min. Cells were heat-shocked for 45 s in a water bath at 42°C and returned to ice for 2 min. Pre-warmed super optimal broth with catabolite repression (SOC) media was added and the mix was incubated for 1.5 hours at 37°C with shaking at 200 rpm. Transformation culture was plated on Luria Bertani (LB)/ampicillin (amp) plates and incubated overnight (O/N) at 37°C. Selected colonies were cultured overnight in 5 ml LB broth with amp at 37°C with shaking. The *bla* promoter on the vector allows expression of the amp resistant gene, therefore media containing amp is used as selection media. Control plasmid *pUC19* was also used to transform another vial of TOP10 cells to determine the transformation efficiency. The transformation efficiency was calculated dividing the number of colony forming units (cfu) by the concentration of plasmid DNA (0.001 ng) used in the transformation.

SOC medium (pH 7.0) was prepared as follows: 2 g Bacto-tryptone, 0.5 g Bacto-yeast extract, 1 ml 1 M NaCl and 0.25 ml 1 M KCl were dissolved in 90 ml water. The pH was adjusted to 7 using 10 M NaOH (approximately 10 µl). The mixture was autoclaved at 121°C and cooled to room temperature. A 2 M stock solution of $MgCl_2.6H_20$ and $MgSO_4.7H_20$ and 2 M stock solution of glucose were prepared and added to the tryptone mixture to give a final concentration of 20 mM and a final volume of 100 ml. The final solution was filter sterilized.

LB/amp plates were prepared as follows: 10 g Bacto-tryptone, 5 g-Bacto-yeast, 10 g sodium chloride, 15 g agar and distilled water to a final volume of one liter. The broth was autoclaved at 121°C for approximately 30 min and allowed to cool to 50°C before adding amp to a final concentration of 100 µg/ml. Aliquots of 30-35 ml of agar broth were poured into 85 mm petri dishes and allowed to solidify.

2.2.2.8. Plasmid purification

In order to identify positive transformants, colonies obtained from transformation cultures were selected and each colony was designated a number and was grown O/N in a 5 ml volume of LB/amp at 37°C with shaking at 200 rpm. Overnight cultures were purified using the PureYield™ Plasmid Miniprep System (Promega, Wisconsin, USA) according to manufacturer's instructions. All centrifugation steps were performed at 16000 x g. Briefly, 3 ml of bacterial culture grown O/N in LB/amp media was processed. The pellet was resuspended in 600 µl ultrapure water. A 100 µl aliquot of cell lysis buffer was added and mixed by inverting the tube six times. A 350 µl aliquot of cold (4-8°C) neutralization solution was added the tube, was inverted to mix the solution which was then centrifuged for 3 min. The supernatant was transferred to a PureYield™ Minicolumn which was placed in a PureYield™ collection tube and centrifuged for 1 min. The flow through was discarded. A 200 µl aliquot of endotoxin removal wash was added to the column and the column centrifuged for 1 min. A 400 µl aliquot of column wash solution (supplied in the kit) was added to the minicolumn and the column centrifuged for 1 min. The minicolumn was transferred to a clean 1.5 ml microcentrifuge tube and 30 µl of nuclease free H_2O was added. The column was left to stand for 1 min and then centrifuged to elute the DNA. The purified DNA was stored at -20°C.

Glycerol stocks of each positive transformed culture were prepared (850 µl O/N bacterial culture + 150 µl glycerol) and stored in cryotubes (Nunc, Roskilde, Denmark) at -70°C.

2.2.2.9. Restriction enzyme digestion reactions to identify positive transformants

To verify insertion of the gene of interest in the TOPO vector, each clone was analyzed by restriction enzyme digestion with restriction enzymes (Promega, Wisconsin, USA) selected to excise the CCHFV NC gene. Different restriction enzymes are optimally functional in various buffers with different constituents. Some crucial factors to consider with buffer constituents are pH, type and concentration of monovalent cation (K^+ of Na^+). When double digestions are performed a compromise may have to be made in order to select a single buffer compatible for both enzymes to function optimally. Double digestion reactions were set up according to the recommendations on the Promega website (www.promega.com) as shown in Table 4. Reaction mixtures were incubated for 2 hours at 37°C.

Table 4: Restriction enzyme analysis to identify positive transformants using *BamH1 and Not1sites present on the* pcDNA ™3.1D/V5-His/TOPO® vector.

Reaction component	**Reaction volume**
1 × restriction enzyme buffer E	2µl
BamH1 (10 U/µl)	1µl
Not1 (10 U/µl)	1µl
Bovine serum albumin (BSA)	1µl
70-80ng plasmid DNA	10µl
Nuclease free water	5µl
Total	**20µl**

2.2.2.10. DNA sequencing of CCHF NC gene in TOPO vector

The CCHFV NC gene was sequenced from plasmid DNA using T7 and BGH primer sites available on the vector. Additional primers, designated SR1, SF2, SR2, SF3 were designed using conserved sites on South African isolates for which sequences were retrieved from GenBank: SPU 128/81/7 (DQ076415), SPU 4/81 (DQ0716416), SPU 415/85 (DQ211648), SPU 97/85 (DQ211646) and SPU 103/87 (DQ211647).

Table 5: Primers used for sequencing of CCHFV isolate SPU 92/01 in TOPO pcDNA 3.1D vector. Positions of primers based on TOPO pcDNA 3.1 V5/His/TOPO vector and positions on SPU 415/85 S segment ORF.

Designated primer	Primer sequence From 5' to 3'	GC Content (%)	5' Position of Primer on the vector/gene
T7	TAATACGACTCACTATAGGG	40	48 bases upstream from insert in TOPO vector
SR1	GGTTCCTTCTCCTAATCATGTC	45	Base 512 on S segment ORF
SF2	GGTTTCCGTGTCAATGCAAAC	47	Base 412 on S segment ORF
SR2	CATTGGGGTGCTCAGCAGAG	60	Base 1040 on S segment ORF
SF3	CGACGGTGTCACAGTTCCTC	60	Base 974 on S segment ORF
BGH	TAGAAGGCACAGTCGAGG	55.5	163 bases downstream from insert in TOPO vector

Sequencing was performed using the Big® Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, New York, USA). The reaction mixtures were set up according to the recommendations of the kit as described in Table 6. The reactions were cycled using the Perkin Elmer Thermocycler GeneAmp 9700 as described in Table 7.

Table 6: Sequencing reactions using BigDye Terminator Ready Reaction.

Reaction component	Reaction volume
Terminator Ready Reaction	2µl
Sequencing primers: forward/reverse (0.8 picomolar/ primer)	4µl
1 × Sequencing buffer	2µl
DNA template	10-40ng
Total	**10µl**

Table 7: Thermal cycling conditions for sequencing reaction

Temperature °C	Time	
96	1 min	
96	10 sec	25 Cycles
50	5 sec	
60	4 min	
4	Hold	

EDTA/ethanol precipitation for cleanup of the sequencing reaction was performed. Briefly, the sequencing reaction volume was adjusted to 20 µl and transferred to a 1.5 ml tube that contained 5 µl 125 mM EDTA and 60 µl absolute ethanol. The mixture was vortexed for 5 sec and precipitated at room temperature for 15 min. The tube was centrifuged at 4°C for 20 min at 14 000 x *g*. The supernatant was aspirated completely without disturbing the pellet. A volume 500 µl of 70% ethanol was added to tubes and centrifuged at 4°C for 10 min at 14000 x *g*. The supernatant was aspirated completely and tubes were placed in the incubator at 37°C, with caps open O/N for evaporation of residual ethanol. Purified sequencing reaction products were submitted to the Department of Microbial, Biochemical and Food Biotechnology, University of the Free State (UFS) for electrophoresis. Chromatograms were visualised and sequences were edited using ChromasPro version 1.49 (Technelysium, Brisbane, Australia). The nucleotide sequence of the NC isolate SPU 92/01 was analysed to confirm whether the gene was cloned in frame.

2.2.3. Distance analysis using MEGA

The alignment file of the ORF of the CCHFV gene encoding the NC was analysed using MEGA version 4 (Tamura *et al.,* 2007). MEGA software functions include the creation and exploration of sequence alignments, the estimation of sequence divergence, the reconstruction and visualization of phylogenetic trees, and the testing of molecular evolutionary hypotheses. Nucleotide sequence data and predicted amino acid sequence data were analysed by constructing phylogenetic trees using the neighbour joining algorithm. In addition *P*-distances were calculated.

The *P*-distance is the proportion (*p*) of nucleotide sites at which two sequences being compared are different. It is obtained by dividing the number of nucleotide differences by the total number of nucleotides compared (Tamura *et al.,* 2007)

A total 1000 bootstrap replicates were performed using a bootstrap test algorithm. The principle of the neighbour joining method is to find pairs of operational taxonomic units (OTUs [=neighbors]) that minimize the total branch length at each stage of clustering of OTUs starting with a starlike tree (Saitou & Nei 1987). Taxa and group setup were done using the nomenclature and groupings used in previous publications (Deyde *et al.,* 2006; Burt *et al.,* 2009). Additionally, sequence divergence was determined by calculating the average *p* distances within groups and between groups and pairwise comparisons of all the sequences were deduced.

The nucleotide sequence data was translated to amino acid sequence data using the standard genetic code. A neighbour joining tree was constructed using the same settings as previously mentioned for nucleotide sequence data. Similarly, sequence divergence was determined by calculating the average *p* distances within groups and between groups and pairwise comparisons of all the sequences was determined. Amino acid residues of isolates SPU 415/85 and AP92 were compared. Amino acid changes were studied, to verify which amino acid group the variable residues belong to.

2.3. Results and discussion

2.3.1. Cloning and sequencing

Phusion® HotStart II High Fidelity System was used to amplify the ORF of the S segment of CCHFV isolate SPU 92/01. RNA was transcribed to cDNA which was used as template for the PCR reaction. Four PCR amplification reactions were prepared. After cycling, the PCR products were separated and visualized using a 1% agarose gel stained with ethidium bromide as shown in Figure 6. The amplicon in lanes 2-5 is approximately 1500 bp according to the molecular weight marker,

which is consistent with the expected size of 1443 bp of the gene encoding the NC protein. There was no visible band in the negative control lane 6 indicating that there was no contamination in the amplification reaction. Lane 4 shows non-specific amplification, therefore, the PCR product from that amplification reaction tube was not used in downstream applications. A possible cause for non-specific amplification present in lane 4 may be an excessive amount of template due to pipetting error.

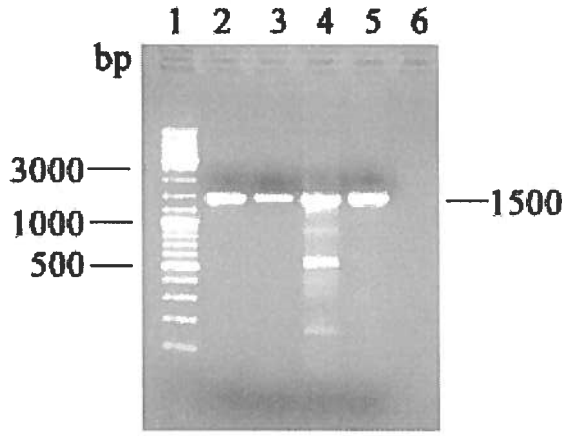


Figure 6: ORF of S segment of isolate SPU 92/01 amplified using Phusion® HotStart II High Fidelity System. Lane 1 – O'GeneRuler DNA ladder mix, Lanes 2-5 – PCR reactions using SPU92/01 cDNA as template, Lane 6 – Negative control PCR reaction

PCR products were purified and PCR reactions in lanes 2, 3 and 5 were pooled post-purification and DNA concentrations were determined using a NanoDrop. The purity of the DNA was determined by calculating the A260/A280 which was 1.94. The ratio should be between 1.8 and 2 for optimal purity of DNA. Nucleic acid concentration of the purified PCR product was 153.9 ng/µl. To confirm the presence of the PCR product after purification, two 5 µl aliquots of the purified PCR product were visualised on a 1% agarose gel stained with ethidium bromide and is shown in Figure 7.

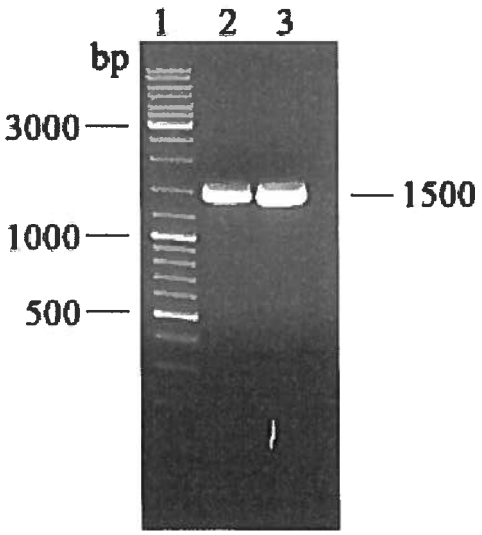


Figure 7: Agarose gel electrophoretic analysis of NC amplicon from the Phusion HotStart PCR reaction after Promega Wizard® SV Gel and PCR Clean-Up. Lane 1 – O'GeneRuler DNA ladder mix; Lanes 2 and 3 – purified PCR product.

For ligation experiments an optimal ratio of PCR product to vector for pcDNA™ 3.1D/V5-His/TOPO®, recommended by the manufacturer, is 2:1. The vector is supplied at approximately 10 ng/µl. The PCR product was diluted to adjust the concentration to approximately 20 ng/µl. The ligation was set up according to manufacturer's instructions. Following transformation of TOP10 cells, six colonies were selected and overnight cultures were prepared. The transformation efficiency of the TOP10 cells was 1 x 10^9 cfu/µg. Overnight cultures were purified and restriction digestion analysis was performed using *BamH1* and *Not1* restriction sites to identify positive transformants. Products from the DNA digestion were separated by electrophoresis on a 1% agarose gel shown in Figure 8. The gene of interest was identified in each of the six colonies selected. On the gel DNA plasmid in lanes 6 and 8 was faintly visible although not depicted clearly in the image.

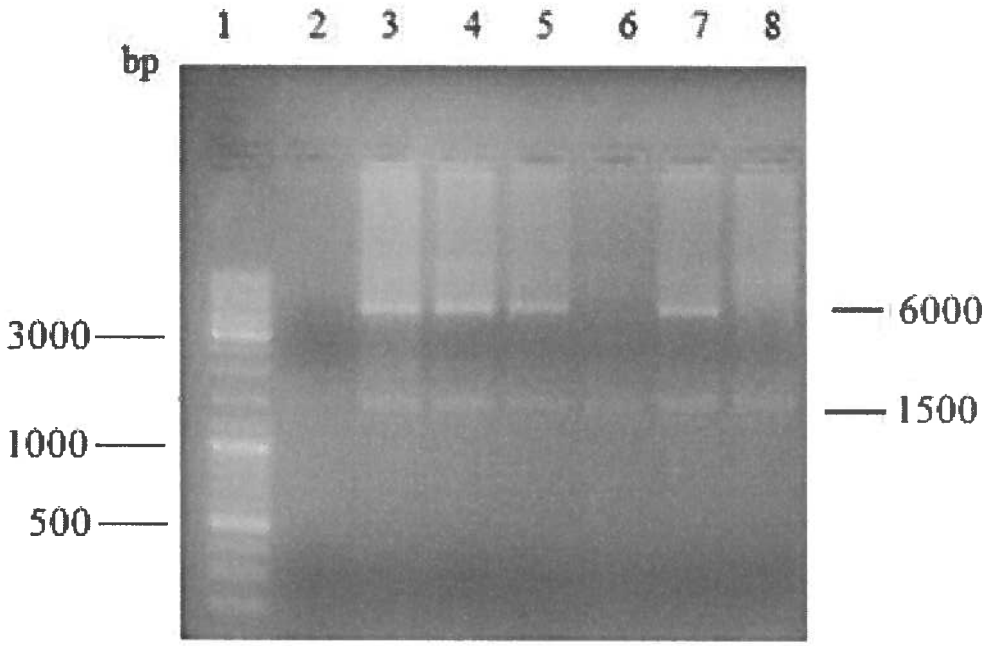


Figure 8: Agarose gel electrophoretic analysis of restriction enzyme analysis of a plasmid obtained from the ligation of the CCHFV NC gene into pcDNA™ 3.1/V5-His TOPO®. Lane 1 – O'GeneRuler DNA ladder mix, Lane 3-8– plasmid DNA from colonies 1-6, double digestions using *BamH1* and *Not1.*

2.3.2. Determination of nucleotide sequence data for isolate 92/01

Nucleotide sequence determination of the ORF of the NC gene of isolate SPU 92/01 was performed using plasmid DNA purified from positive transformants. Raw sequence data was assembled and the data edited using ChromasPro (version 1.49). Duplicate and bidirectional sequencing reactions were performed to obtain data for editing bases. Overlapping regions (depicted in Figure 9) of the NC gene were edited. The primers were designed to cover the entire ORF of the NC gene.

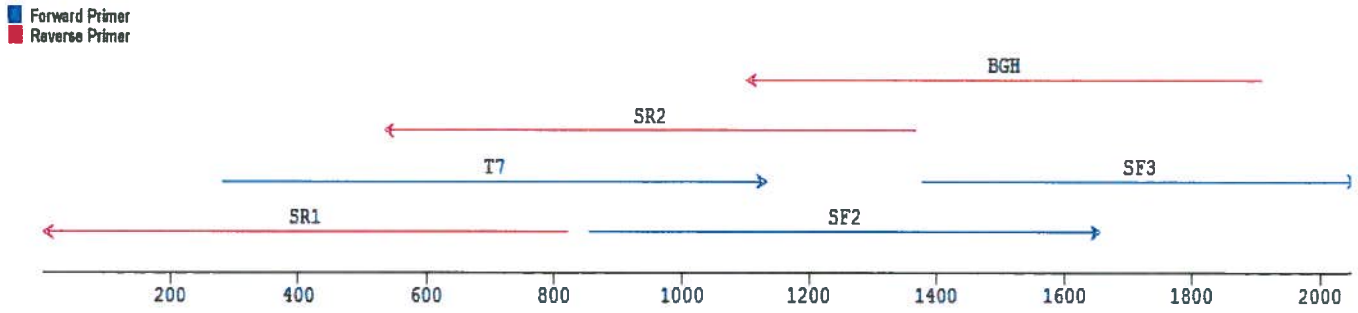


Figure 9: Contiguous sequencing map illustrating the coverage of the primers across the sequence of the NC gene in pcDNA3.1D/V5-His/TOPO vector. Approximately 2000 bases were covered. Primers were designed to ensure overlapping sequences were available for editing. Forward primers are indicated in blue, while reverse primers are indicated in red.

The complete sequence of SPU 92/01 S segment ORF was sequenced in pcDNA3.1D/His-V5/TOPO vector, the sequence of the gene, is depicted in Figure 10. Restriction enzyme sites, BamH1 and *Not1,* used to excise the gene from the plasmid, to identify positive transformants are highlighted.

The sequence data confirmed that the gene had been cloned in frame with the start codon and the His-tag that can be used for purification and detection of the recombinant protein. The Kozak sequence, T7 promoter site, polyhistidine tag and V5 epitope gene is also present as indicated in Figure 10. This construct was designated pcDNA3.1/V5-His/CCHFVNC.

Figure 10: Complete sequence of CCHFV NC of isolate SPU 92/01 cloned in frame in pcDNA3.1D/V5-His/TOPO vector. The T7 promoter site, Kozak sequence and coding region is highlighted, as well as restriction enzyme sites for *BamH1* and *Not1*, used to identify positive transformants. The polyhistidine region and V5 epitope are also highlighted to ensure expression of the tags for protein detection

2.3.3. Phylogenetic analysis using nucleotide sequence data

Phylogenetic trees were constructed using the neighbor joining method. The node values were generated by 1000 bootstrap replications. Node values below 50 were omitted. The isolates clustered into six distinct groups designated: I West Africa, II DRC, III South Africa/West Africa, IV Asia/Middle East, V Europe/Turkey and VI Greece.

Within Group I there were two isolates from Senegal. Within Group II there two isolates from the DRC. Within Group III there were isolates from various regions of Africa including South Africa, Sudan, Nigeria and Mauritania. Within Group IV there were isolates from China, Pakistan, Iran, Tajikistan and Oman. Within Group V there were isolates from Bulgaria, Turkey, Kosovo and Russia. Finally Group VI was comprised solely of an isolate obtained from a tick in Greece. The nomenclature of Deyde *et al* (2006) in which isolates were placed into groups was used as there were too few isolates to clearly show the genotypes indicated by Mild *et al.,* in which 168 partial S gene sequences were analysed.

The DRC group and West Africa group displayed strong node values on the phylogenetic tree. From the phylogenetic tree, it is evident that all the southern African strains of CCHFV included in this analysis were genetically similar and formed a monophyletic branch on the phylogenetic tree with a similarity of 95.5%.

The South African isolate sequenced in this study, SPU 92/01, shared a nucleotide sequence data homology of 96.4% and 95.1% with South African isolates SPU 103/87 and SPU 97/85 respectively. These values correlate with the branch formation on the phylogenetic tree in Figure 11 that was constructed using nucleotide sequence data. Similarly, nucleotide pairwise comparisons showed that SPU92/01 was closely related to SPU128/81/7 and SPU415/85 with homologies of 96.9% and 96.6% respectively.

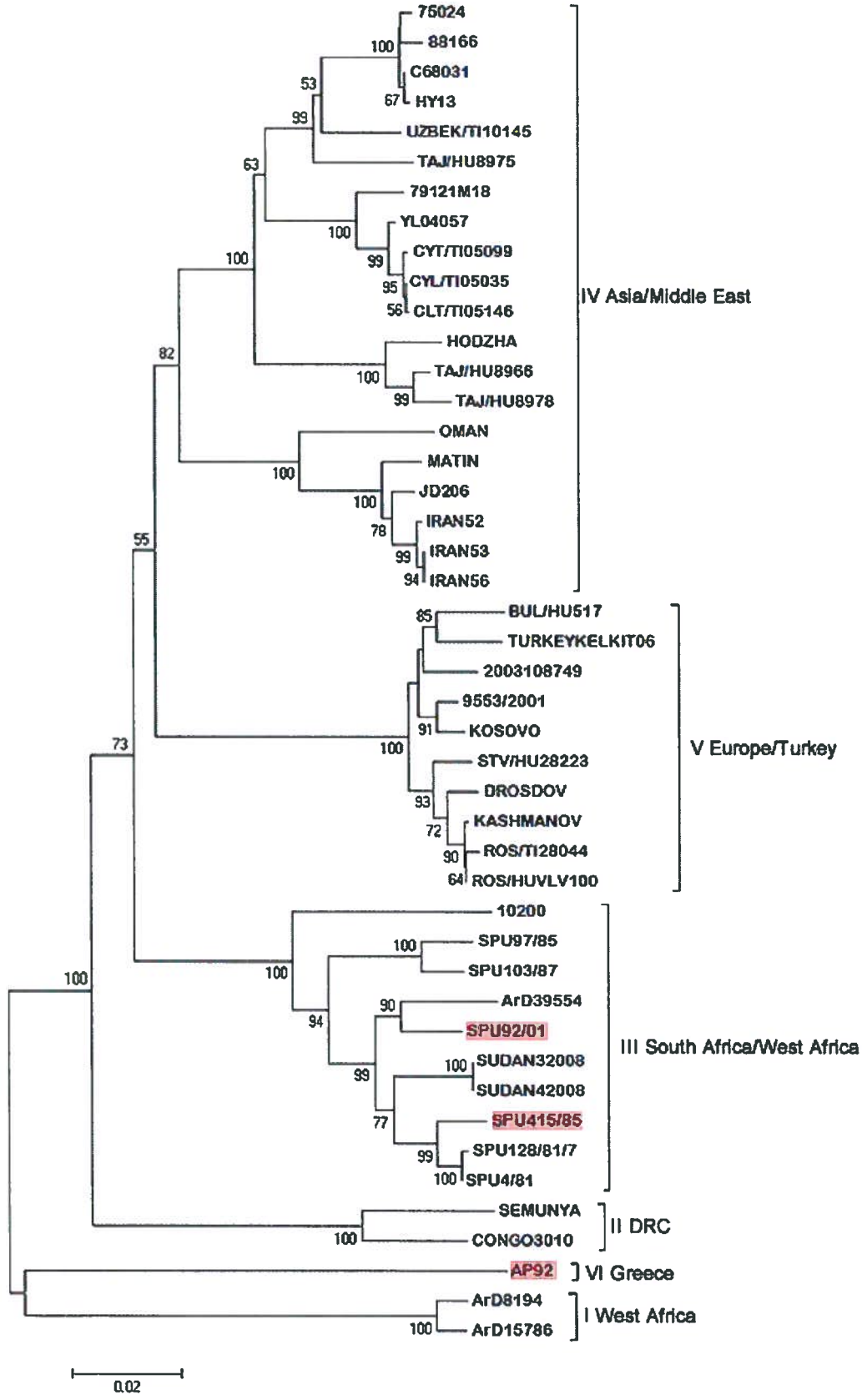


Figure 11: Phylogenetic tree using nucleotide sequence data of ORF of NC genes. There were a total of 1449 positions in the final dataset. The node values indicate the

percentage of 1000 bootstrap replicates, values below 50 are omitted. Horizontal distances are proportional to nucleotide differences. Isolates are labelled according to name of the strain and groups using nomenclature described by Deyde *et al* 2006. Isolates SPU92/01, SPU415/85 and AP92 for which sequence data were used to prepare recombinant antigens were highlighted. Phylogenetic analyses were conducted in MEGA4.

The average genetic *P*-distances, or the proportion of pairwise differences (the proportion of nucleotide or predicted amino acid sites at which the two sequences compared are different), were determined between and within the groups using MEGA. The nucleotide distances are summarized in Tables 8-10. The tables show distances, which may be translated to percentages when multiplied by 100.

Table 8: Mean distance (d) within groups: The number of base differences per site from averaging over all nucleotide sequence pairs within each group and pairwise distance ranges is shown.

Group	d	Pairwise differences (Range)
I West Africa	0.012	0.012
II DRC	0.044	0.044
III South Africa/West Africa	0.045	0.001 - 0.077
IV Asia/Middle East	0.062	0.001 - 0.097
V Europe/Turkey	0.020	0.002 - 0.169
VI Greece	n/c *	n/c *

*n/c – unable to estimate

Sequence diversity within groups ranged from 1.2% to 6.2% (see Table 8). The South Africa/West Africa group had a diversity of 4.5% and Asia/Middle East group, 6.2%. The within group diversity of Group V was 2% compared with the Asia/Middle East group which was much higher, which could be attributed to the larger number of isolates from this group included in the analyses.

Table 9: Mean distance between groups based on nucleotide sequence data: The number of base differences per site from averaging over all sequence pairs between groups is shown.

	I	II	III	IV	V	VI
I West Africa						
II DRC	0.169					
III South Africa/West Africa	0.166	0.145				
IV Asia/Middle East	0.158	0.132	0.114			
V Europe/Turkey	0.165	0.135	0.131	0.108		
VI Greece	0.170	0.179	0.173	0.168	0.170	

The mean net distance between groups was also calculated. This value is given by:

$$d_A = d_{XY} - (d_X - d_Y)/2$$

In this equation, d_{XY} is the average distance between groups X and Y, and d_X and d_Y are the mean within-group distances.

Table 10: Mean Net distance between groups: The number of base differences per site from estimation of net average between groups is shown.

	I	II	III	IV	V	VI
I West Africa						
II DRC	0.141					
III South Africa/West Africa	0.138	0.101				
IV Asia/Middle East	0.121	0.079	0.061			
V Europe/Turkey	0.149	0.103	0.099	0.067		
VI Greece	0.164	0.157	0.151	0.137	0.160	

Table 9 is a summary of the *P*-distance between the groups, with the Asian/Middle East and Europe/Turkey isolates being the most similar with a difference of 10.8% and net difference of 6.7% (Table 10). The net *P*-distance between groups ranged from 6.1% to 16.4%.

A neighbour joining *P*-distance tree was constructed using the predicted amino acid sequences of the 44 isolates listed in Table 1 and isolate SPU 92/01. The tree topology was similar to that obtained using nucleotide sequence data, although monophyletic branches varied. A noteworthy change in grouping was that SPU92/01 and SPU415/85 formed one monophyletic branch.

The overall tree topology showed no obvious correlation between the grouping of the strains and source of infection or year of infection. However, the broad geographic distribution of strains may be related to the distribution and dispersal of the vectors of the virus. Based on the sequence data from the NC gene, isolates circulating in Africa clustered in one of two groups, while isolates circulating in Eastern Europe and Turkey, and the Middle East and Asia and Greece grouped in geographically defined regions.

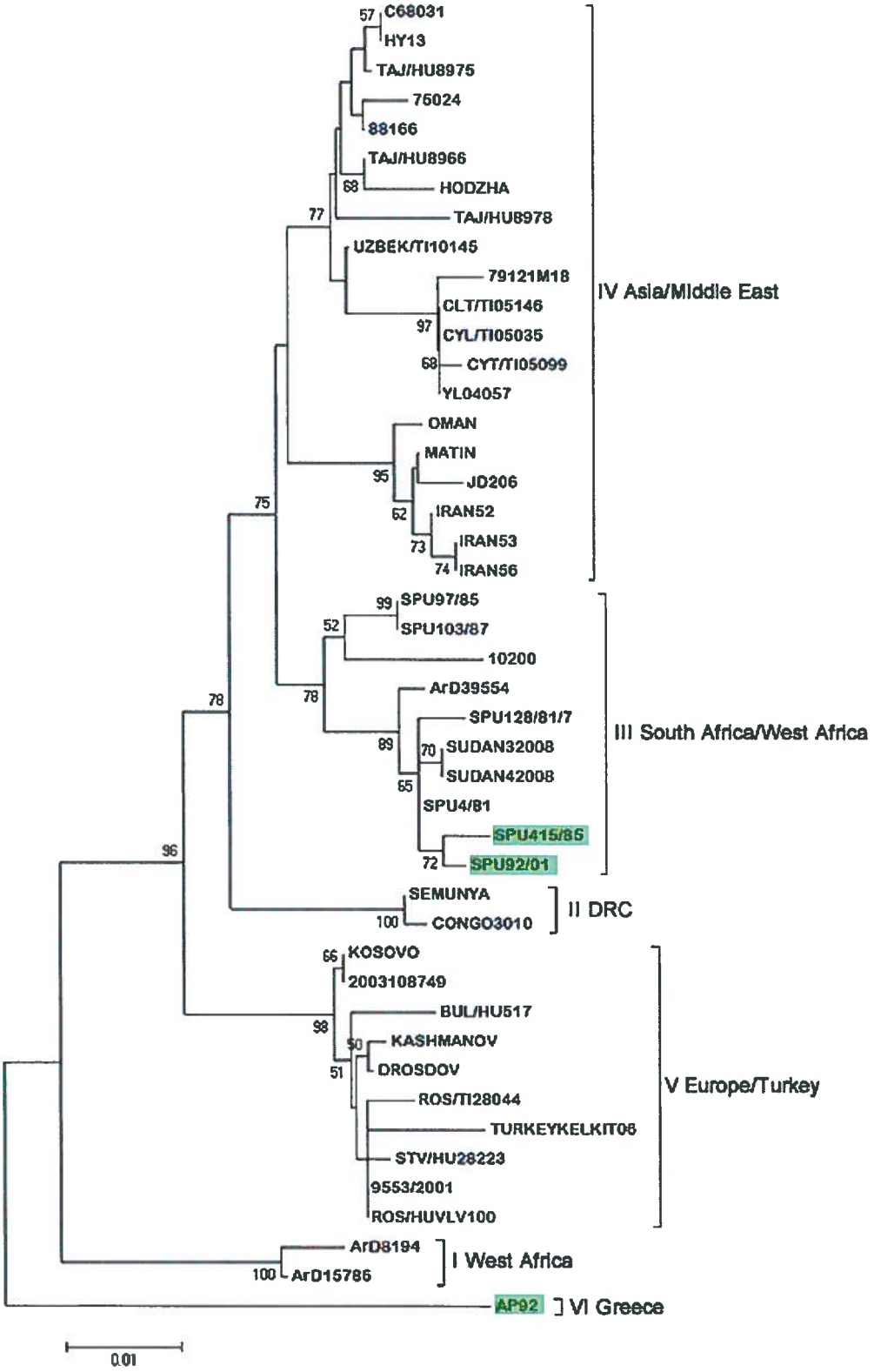


Figure 12: Phylogenetic tree using amino acid sequence data of NC. There were a total of 482 positions in the final dataset. The node values indicate percentage of 1000 bootstrap replicates, values below 50 are omitted. Horizontal distances are

proportional to amino acid differences. Isolates are labelled according to name of the strain and groups using nomenclature described by Deyde *et al* 2006. Isolates SPU92/01, SPU415/85 and AP92 for which sequence data were used to prepare recombinant antigens were highlighted. Phylogenetic analyses were conducted in MEGA4.

Table 11: Mean distance within groups: The number of differences per site from averaging over all amino acid sequence pairs within each group and pairwise distance ranges are shown.

Group	d	Pairwise differences (Range)
I West Africa	0.006	0.006
II DRC	0.002	0.002
III South Africa/West Africa	0.012	0 - 0.029
IV Asia/Middle East	0.017	0 - 0.029
V Europe/Turkey	0.007	0 - 0.062
VI Greece	n/c *	n/c *

*n/c – unable to estimate

Within the groups a 1.2% nucleotide diversity in Group I translated to 0.6% amino acid diversity, 4.4% diversity in Group II translated to 0.02% diversity in amino acids, in Group III 4.5% diversity translated to 1.2%, in Group IV 6.2% diversity translated to 1.7% and in Group V 2% translated to 0.7%. Similarly, isolates in Group II had 0.2% diversity, in Group V 0.7%, Group III 1.2% and Group IV 1.7%. As with the nucleotide differences, the larger diversity shown by isolates from the Middle East and Asia was also influenced by the greater number of isolates available for inclusion in the study. The amino acid diversity between groups was significantly lower than the nucleotide diversity between the different groups.

Table 15: Fugene:DNA complex ratios for optimization of transfection

Transfection Ratio FuGENE 6® (µl) :DNA (µg)	Serum free media (DMEM) (µl)	Total Fugene:DNA complex (µl)	Total volume of media in well (ml)
3:1	96	100	2
3:3	94		
6:1	93		
6:3	91		
9:3	88		
9:6	85		

FuGene 6® transfection reagent was diluted with serum-free, antibiotic-free DMEM medium (Life Technologies, New York, USA) and plasmid DNA (endotoxin free) was added to the tube. The FuGene/DNA complex was incubated at room temperature for 30 min. The complex was then added to the 50-80% confluent cells. For each experiment, two controls were set up: one well containing no transfection reagent or DNA and another well with transfection reagent only. The controls were set up to monitor possible cytotoxicity of the transfection reagent. Cells were incubated at 37°C in a CO_2 incubator and growth was monitored daily. After initial optimization, using ratios described in Table , cells were harvested 48 hours post-transfection.

Transfection experiments using ratios 3:3, 6:3 and 9:6 were repeated using HEK293 cells in a 6-well plate and cells were harvested at intervals of 3, 6 and 10 days post-transfection. Vero cells and HeLa cells were transfected with pcDNA3.1/V5-His/*LacZ*-TOPO DNA at a ratio of 9µl of FuGene per 6µg of DNA. These cells were harvested 3- and 5 days post-transfection for monitoring of transfection efficiency using β-galactosidase assay.

The pcDNA3.1/V5-His/CCHFVNC construct was used to transfect HEK293 and HeLa cells also using a ratio of 9:6. HEK293 cells transfection efficiency was monitored 3 days post-transfection using IFA and HeLa cells were harvested 3 days post-transfection for Western blot analysis.

GeneJuice ® Transfection reagent (Novagen, California, USA) was used to transfect HeLa cells. On the day of transfection, complete growth medium was removed and cells were washed with PBS before media was replaced with serum free and antibiotic free media. For each well, 100 µl serum-free media was added to a sterile tube. A volume of 3 µl of GeneJuice Transfection reagent was added directly to the serum-free medium. The mixture was incubated at 22-25 °C for 5 min. For each 35 mm well, 1 µg, 2 µg, 3 µg of DNA was added to each tube, respectively. The mixtures GeneJuice: DNA complex mixtures were incubated for 30 min and then added dropwise to wells as shown in Figure 16.

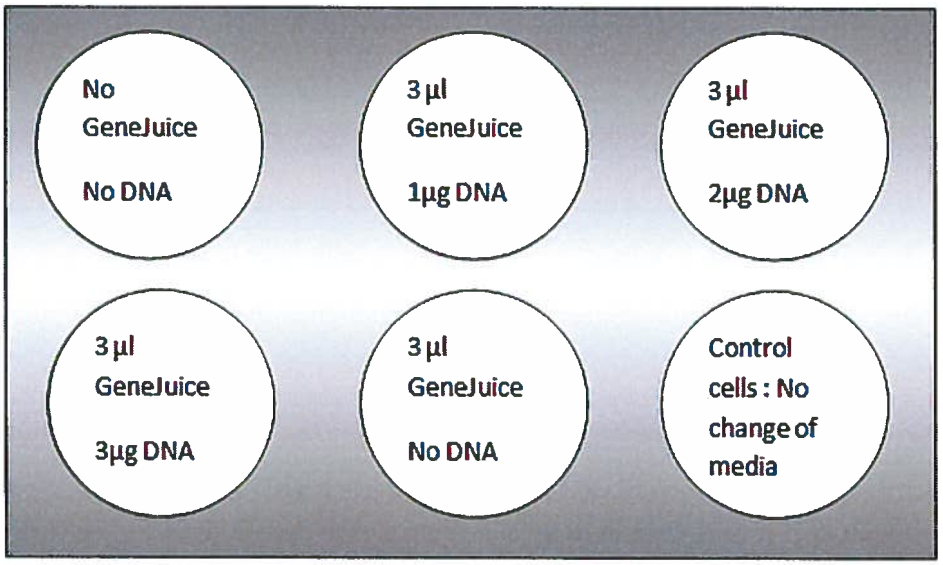


Figure 16: GeneJuice transfection reagent: DNA ratios used in transfection experiment using pcDNA3.1/V5-His-TOPO/*LacZ*

Cells were incubated at 37 °C with 5% CO_2. The media was replaced with complete growth medium 8 hours post-transfection. Cells were harvested for analysis 72 hours post-transfection.

Cells were harvested from 6-well plates using cell scrapers. Harvested cells were centrifuged at 3000 x g for 5 minutes, media was discarded and cells were resuspended in PBS containing 2% FBS. Ten µl aliquots of the cell suspension were

applied to each well of an 8-well multi-test slide (Flow laboratories, California, USA), dried and fixed in cold acetone overnight.

Transfection efficiency of cells transfected with pcDNA3.1/V5-His-TOPO/*LacZ* construct was determined using a β-galactosidase assay, while cells transfected with pcDNA3.1D/V5-His /CCHFVNC were monitored using immunoflourescence.

3.2.1.4. Monitoring transfection

The β-galactosidase assay (Mirus Bio, Wisconsin, USA) was performed according to manufacturer's protocol with modifications during the cell-fixing step. Approximately 20 µl of the freshly prepared cell staining working solution (1 ml staining reagent and 20 µl X-gal reagent) was added to each well of the 8 well slide. The cells were incubated at 37°C protected from light for 4 hours, 5 hours or 24 hours. After staining, the slides were washed in PBS for 3 min and 1 min in distilled water using a bath with a magnetic stirrer. The slides were air dried and mounted with glycerol mounting solution and a cover slip. Blue stained cells were counted using light microscopy (bright-field). Transfection efficiency was determined by calculating the ratio of blue stained cells to total cells which were counted using a cell counter.

To determine the presence of CCHFV NC in HEK293 cells, an immunoflourescence test was performed using serum known to react against CCHFV NC. Cells transfected with the plasmid construct containing the S gene of CCHFV in TOPO vector were harvested, 10 µl aliquots were added to each well of an 8 well slide, dried and fixed on the slide by immersing in cold acetone for 20 min. For the immunofluorescence (IF) tests 10 µl aliquots of patient sera (positive control CCHFV antibody) undiluted and diluted, were applied to each well of an antigen slide. The slides were incubated at 37°C in a moist chamber for 20 min. The slides were washed in PBS for 3 min and 1 min in distilled water and dried. A 10 µl aliquot of flourescein isothiocyanate (FITC) labeled anti-human IgG conjugate (Zymed Laboratories, California, USA) diluted 1:100 in 0.05% Evans blue as a counterstain was applied to each well and slides were incubated at 37°C in a moist chamber for

20 min. The slides were washed in PBS for 3 min and 1 min in distilled water, dried and mounted with glycerol mounting fluid. Glycerol mounting fluid was prepared using 0.715g $NaHCO_3$, 0.16g Na_2CO_3, volume increased to 10 ml using dH_2O and 90 ml glycerol was added. The slides were read with a Nikon ultraviolet light microscope.

FuGene transfection was repeated using a 9:6 ratio for the control expression vector and pcDNA3.1D/V5-His/CCHFVNC construct. Subsequent to transfection, HeLa cells were harvested and lysed. A volume of 375 µl Protease inhibitor cocktail (Complete, Mini, EDTA-free, Roche, Mannheim, Germany) was added to 2.125 ml CelLytic lysis buffer (Invitrogen, Maryland, USA). Media was removed from cells and each well was washed with 2 ml Dulbecco's PBS (without Ca^{2+} or Mg^{2+}) (Life Technologies, New York, USA). After PBS was discarded, 400 µl of the lysis buffer mix was added to each well and the plates were shaken with gentle agitation at room temperature for 15 min. Cells were scraped off with a cell scraper and collected in a sterile autoclaved 1.5 ml microcentrifuge tube. The tubes were centrifuged at 10 000 x g for 15 min at room temperature. The supernatant was collected in pre-chilled (-20°C) 1.5 ml tubes and stored at -80 °C. The pellets were also collected and stored at -20°C.

Due to low ratio of blue cells to unstained cells in the β-galactosidase assay and no positive immunoflourescence from cells transfected with CCHF plasmid, cell lysates were harvested for Western blot analyses. A supernatant and pellet fraction was collected from both expression control vector and pcDNA3.1D/V5-His-TOPO/CCHFVNC construct transfected cells. A 13 kDa His-tagged recombinant yellow fever EDIII protein previously expressed in a bacterial system in an unrelated study was included as a positive control. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of the cell lysates was performed as described by Laemmli (1970). SDS is an anionic detergent that partially denatures proteins. The resolving gel was prepared as shown in Table 16. The resolving gel was poured into the pre-assembled mini-gel electrophoresis apparatus (Bio-Rad, California, USA). A layer of isobutanol was applied to the top before polymerization for approximately 25 min. After the resolving gel solidified the layer of isobutanol was discarded.

Table 16: Preparation of 8% resolving gel

Reaction component	Reaction volume (ml)
30% acrylamide solution/0.8% bisacrylamide stock solution (Merck, USA)	5.35
1M Tris-HCl pH 8.8	7.5
10% SDS	0.2
1.5% ammonium persulphate (freshly prepared)	1
TEMED	0.01
Distilled H_2O	6
Total	**20**

The stacking gel was prepared as shown in Table 17 and was applied on top of the resolving gel. The sample comb was inserted before polymerization of the gel.

Table 17: Preparation of 4% stacking gel.

Reaction component	Reaction volume (ml)
30% acrylamide solution/0.8% bisacrylamide stock solution	1
1M Tris-HCl pH 6.8	0.95
10% SDS	0.075
Glycerol	0.5
1.5% ammonium persulphate	0.35
TEMED	0.02
Distilled H_20	4.6
Total	**7.5**

The electrophoresis tanks were filled with 1× Tris-glycine-SDS electrophoresis buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3) (Fermentas, Illinois, USA). The protein samples were prepared as follows: 20 µl protein sample, 8 µl of 5x protein

loading buffer (0.313 M Tris-HCl, pH 6.8, 10% SDS, 0.05% bromophenol blue, 50% glycerol) and 1 µl 20x reducing agent (Fermentas, Illinois, USA). The addition of reducing agent completely denatures the protein and linearizes the polypeptide. The samples were heated for 5 min at 95°C and 25 µl aliquots of the samples were loaded into each well. The gel was run at 120 V for 1 hour and 30 min. Spectra Broad Range Prestained protein marker (Fermentas, Illinois, USA) comprising proteins ranging in size from 10 to 260 kDa proteins was used to determine the sizes of the expressed proteins. SDS PAGE gel was subsequently used for Western blotting.

The gels were soaked in transfer buffer (25 mM Tris, 192 mM glycine, 20% methanol, pH 7.4) for 15 min. Microporous polyvinylidene diflouride (PVDF) membrane (Roche, Mannheim, Germany) was moistened with methanol for 3 seconds, incubated in distilled water for 2 min and then soaked in transfer buffer for 3 min. The membrane is hydrophobic and electrostatic with a pore size of 0.45 µm. The membrane was transferred to a semi-dry blotter (Bio-Rad, California, USA) and run at 15 V for 15 min for transfer of proteins from the gel to the membrane. The membrane was removed from the transfer apparatus and washed twice with Tris buffered saline (TBS) buffer (50 mM Tris, 150 mM NaCl, pH 7.5). Pico Fast Western Blot Kit (Thermo Scientific, Illinois, USA) was used for Western Blot analysis. The membrane was briefly washed in fast western 1x wash buffer (supplied in the kit) to remove transfer buffer. A mouse anti-His IgG antibody (100µg/ml) (Roche, Mannheim, Germany) diluted 1:200 in the fast western antibody diluent (supplied in the kit) was added to the membrane. The membrane was incubated for 30 min at 22-25°C with shaking. The membrane was removed from the anti-His antibody solution and placed in a clean incubation tray. Anti-mouse IgG horse radish peroxidase (HRP) reagent (supplied in the kit) diluted 1:500 in antibody diluents was added to the membrane and the membrane was incubated for 10 min with shaking at 22-25°C. The membrane was removed from the HRP solution and washed by suspension in approximately 20 ml of 1x wash buffer, agitating it for 5 min. The wash was repeated 3 times. The membrane was then exposed to the supersignal

west pico solution (supplied in the kit) mixed according to the manufacturer's instructions for 1 min at 25°C. The membrane was placed in a clear plastic wrap and exposed to x-ray film (Thermo Scientific, Illinois, USA) for approximately 60 seconds. The film was incubated for 2 min in developer and for 1 min in fixer solution. Lastly, the film was rinsed with water to remove excess fixer.

3.2.2. Bacterial Expression

Bacterial expression is known for producing a higher yield of protein, and was thus adopted due to low protein expression levels in the mammalian expression system. The South African isolate SPU 415/85 NC gene had been codon optimized for bacterial expression and subcloned into bacterial expression vector that encodes for a chaperone, trigger factor (TF), namely pColdTF (Takara Bio, Shiga, Japan) and expressed in an insoluble form in a previous study (Samudzi *et al.,* 2012). The Greek strain of CCHFV, AP92, as described in section 2.3.3 of the study and in previous literature was selected as the most diverse strain of CCHFV (Papa *et al.,* 2002b; Anagnostou & Papa, 2009). The gene encoding the NC of AP92 was codon optimized for bacterial expression, subcloned into pColdTF vector. The two constructs, designated pColdTF-opSPUNC and pColdTF-opAP92NC were expressed in an induction study to determine the optimal time to harvest cells. Following an induction study, a solubility study was performed to determine whether the protein was expressed in a soluble form.

3.2.2.1. Codon optimization of CCHFV NC genes

If nucleotide substitution is random at each site, every site should have one of the four nucleotides AGCT with equal probability. If no selection and no mutation bias exist, it is expected that codons encoding the same amino acid are on average in equal frequencies in protein coding genes. However, some codons are used more frequently leading to codon bias.

CCHFV isolate SPU 415/85 NC gene was analysed in a previous study using the Rare Codon Analysis Tool software available on the GenScript website (http://www.genscript.com). The software optimizes codon usage and GC content and eliminates polyadenylation sites, splicing sites, killer motifs and RNA secondary structure. The gene was codon optimized for expression in a bacterial system and was synthesized by GenScript and supplied cloned in the multiple cloning site (MCS) of pUC57 with *BamH1* and *Pst1* restriction site modifications at the 5' end and 3' end respectively. The codon optimized gene was rescued from the pUC57 plasmid and cloned into pColdTF expression system in a previous study and glycerol stocks of pColdTF-opSPUNC were stored at -80 ºC (Samudzi *et al.,* 2012).

The complete nucleotide sequence for the open reading frame encoding the NC gene for CCHFV isolate AP92 retrieved from GenBank (as described in Chapter 2) was analysed using the Rare Codon Analysis Tool software available on the GenScript website (http://www.genscript.com). As mentioned previously the software optimizes codon usage and GC content and eliminates polyadenylation sites, splicing sites, killer motifs and RNA secondary structure. The sequence data was submitted to GenScript for synthesis of a codon optimized gene.

The two indices that measure codon bias are Codon Adaptation Index (CAI) and Effective number of codons (ENC) and can be used to measure the optimization process. CAI is the most effective of all codon bias measures for predicting gene expression levels and measures the similarity between the codon usage of a gene and the codon usage of the host. Codon usage frequency, percentage distribution of codons and average GC content were determined before and after optimization of the gene. The optimized gene was synthesized by GenScript and supplied cloned in the MCS of pUC57 with *BamH1* restriction site modifications at the 5' end and 3' end. The vector map of pUC57 is shown in Figure 17.

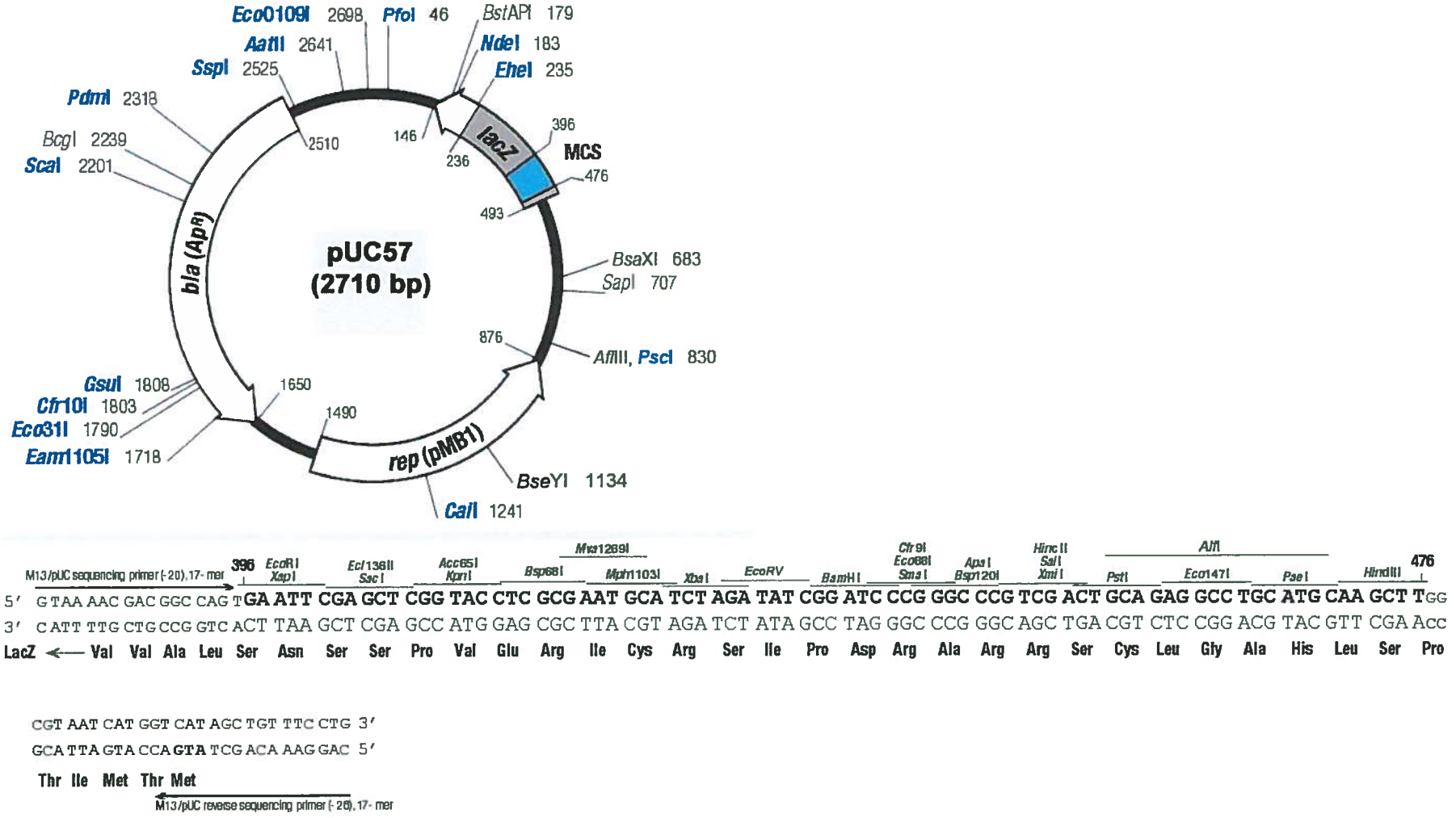


Figure 17: Vector map and multiple cloning sites of cloning vector pUC57 (GenScript Corporation, New Jersey, USA).

3.2.2.2. Preparation of chemically competent cells

To propagate plasmid DNA and for subcloning experiments, chemically competent *E. coli* XL10 and Overexpress BL21 (DE3) competent cells were prepared (Lucigen, Wisconsin, USA). Cells for different strains of *E. coli* were prepared separately to avoid contamination. Chemically competent cells were prepared under sterile conditions utilizing the calcium chloride method. Briefly, a 5 ml O/N culture of cells was prepared in Luria Bertani (LB) broth without ampicillin (amp). A 25 ml aliquot of LB media was inoculated with 250 µl of the overnight culture. The culture was grown at 37 °C with shaking at 200 rpm until the OD_{600} was between 0.2 and 0.4, early log phase of growth. The culture was divided between two 25 ml centrifuge tubes and cells were collected by centrifugation at 3000 x g for 10 min at 4 °C. The cells were kept on ice, and temperature-sensitive steps were performed in the walk-in fridge. Cells were resuspended gently in ½ culture volume (6.25 ml in each tube) of ice-cold freshly prepared 0.1 M $CaCl_2$. The cells were kept on ice for 1 hour at 4°C. Centrifugation step was repeated and the supernatant was discarded. Cells were resuspended on ice in a tenth of the culture volume and combined into one 15 ml centrifuge tube (2.5 ml). Cells were incubated at 4°C on ice. Thereafter, 375 µl of ice-cold sterile glycerol was added to the cells at a final concentration of 15% v/v. After gentle pipetting, the cells were kept on ice at 4°C for 30 min. Cells were divided in microcentrifuge tubes in 100 µl aliquots. The last step was performed in the -20°C walk in freezer and tubes were snap-frozen in liquid nitrogen before being stored at -80°C. The transformation efficiency of the cells was determined post-transformation with pUC19 plasmid DNA as described in section 2.2.2.7.

3.2.2.3. Transformation of XL10 cells with AP92 in pUC57 plasmid

A 100 µl aliquot of XL10 cells were transformed with the plasmid supplied by GenScript using the heat shock method. This method uses the principle that when cells are made competent, their membranes are modified to facilitate the uptake of the DNA plasmid during the heat shock step. Briefly, 10 µl of the ligation reaction was added to a 100 µl aliquot of chemically competent cells, gently mixed and

placed on ice for 20 min. The cells were heat shocked for 45 seconds at 42°C, and immediately transferred to ice for 2 min. A 900 µl aliquot of pre-heated SOC medium was added to the tube and incubated for 1 hour at 37°C with shaking at 200 rpm. SOC medium was prepared as described in Methods and Materials 2.2.2.7. Colonies were selected and overnight cultures were prepared using 5 ml LB/amp broth for identification of positive transformants.

3.2.2.4. Recovery of CCHFV AP92 NC gene from pUC57 vector and subcloning into pColdTF vector

The codon optimized CCHFV AP92 NC gene was rescued from pUC57 cloning vector using the *BamH1* restriction site. A *BamH1* restriction site was added to both the 5' and 3' ends of the AP92 NC gene during synthesis by GenScript. To retrieve the gene form pUC57, a digestion mix was prepared using 2 µl Buffer E (60 mM Tris-HCL pH7.5, 1 M NaCl, 60 mM $MgCl_2$ and 10 mM DTT), 1 µl *BamH1* and 4 µl plasmid DNA and was adjusted to a 20 µl reaction using nuclease free H_2O. The reaction was incubated for 2 hours at 37°C. The digested DNA was separated by electrophoresis on a 1% agarose gel. A DNA band of approximately 1500 bp was excised from the gel and purified using Promega Wizard Cleanup kit as described in 2.2.2.4. The initial step was modified to purify the DNA from agarose gel. Briefly, an equal amount (10 µl per 10 mg) of DNA binding buffer was added to the gel slice which was heated at 65°C for 10 min. The solution was then added to the column and the DNA purified according to manufacturer's instructions. DNA absorbs to the silica membrane in the presence of chaotropic salts while the contaminants flow through the column. DNA concentration of the purified rescued insert was measured using a NanoDrop as described previously.

The pColdTF DNA Vector (Takara Bio, Shiga, Japan) is a fusion cold shock expression vector that expresses Trigger Factor (TF) chaperone as a soluble tag. Trigger Factor is a prokaryotic ribosome-associated chaperone protein (approximately 48 kDa) which facilitates co-translational folding of newly expressed

polypeptides. TF is highly expressed in *E. coli* expression systems due to its *E. coli* origin. The pColdTF DNA Vector comprises of a *cspA* promoter plus additional downstream sequences including a 5' untranslated region (5' UTR), a translation enhancing element, a 5' His-tag sequence, and an MCS. A *lac* operator is inserted downstream of the cspA promoter to ensure strict regulation of protein expression. There are also recognition sites for HRV 3C Protease, Thrombin, and Factor Xa located between TF Tag and the MCS which function to facilitate removal of the tag from the expressed fusion protein. The pColdTF vector is optimal for expression of toxic or transmembrane proteins using the "cold shock" technology. The vector map and MCS of pCold TF is shown in Figure 18.

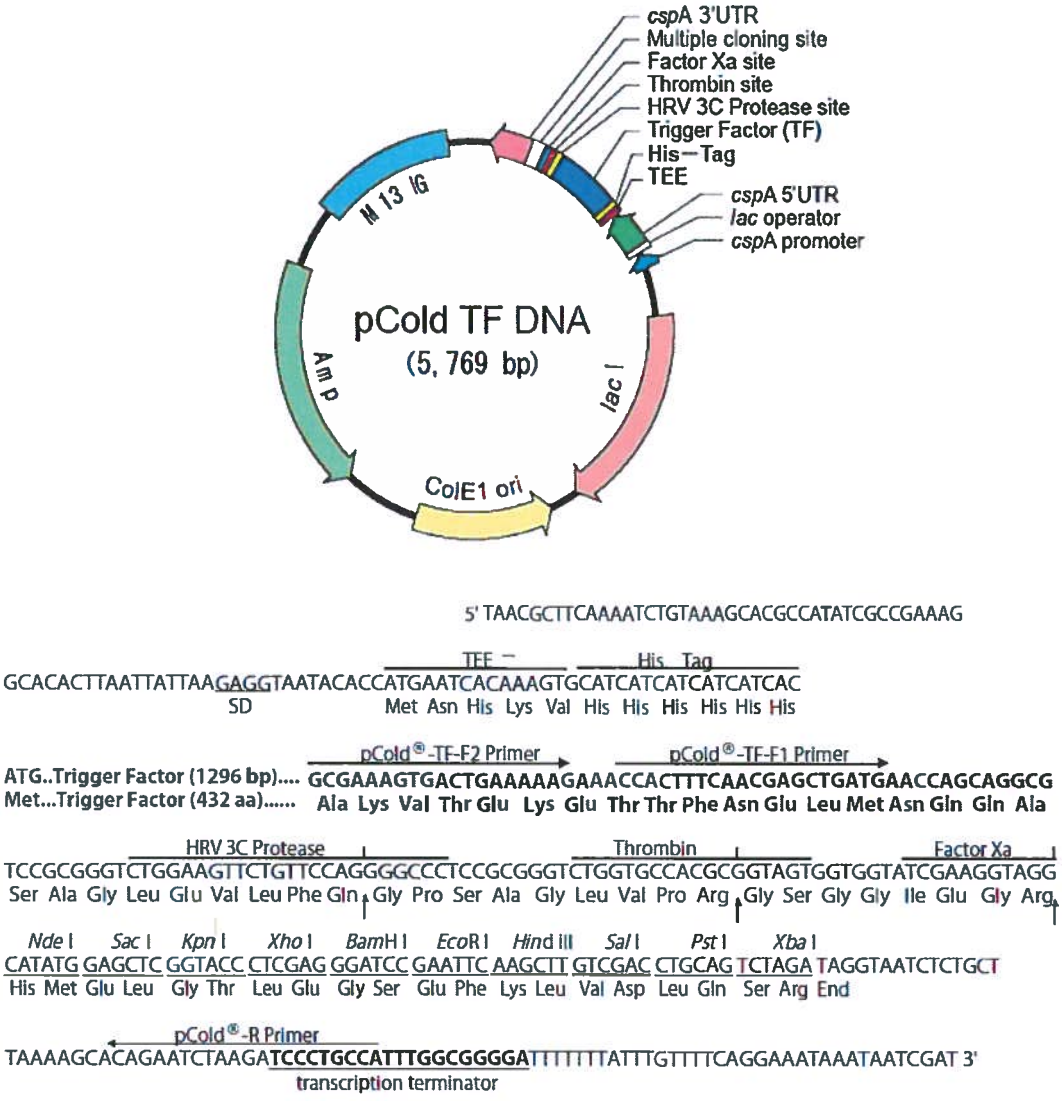


Figure 18: Vector map and MCS of pCold TF bacterial expression system (Takara Bio, Shiga, Japan).

The pColdTF vector was linearized using the *BamH1* restriction site for subsequent ligation reactions with the codon optimized CCHFV AP92 NC gene. The codon optimized CCHF AP92 NC gene was rescued from pUC57 cloning vector using the restriction site (*BamH1*). Following restriction enzyme digestion, the optimized CCHF NC gene was subcloned into pColdTF bacterial expression vector as described in Table. The ligation reaction was incubated for 1 hour at 16°C. OverExpress cells were transformed using the heat-shock method as described in section 3.2.2.3. The transformation culture was spread on LB/amp plates and incubated O/N.

Table 18: Ligation reaction for cloning AP92 NC into pColdTF bacterial expression vector (Fermentas, Illinois, USA).

Reaction component	**Reaction volume**
10 x T4 ligase buffer (400 mM Tris-HCL, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP, pH 7.8)	1µl
T4 ligase enzyme (1U/µl)	1µl
Purified digested AP92 NC DNA	6 µl
Purified double digested pColdTF plasmid	2µl
Total	**10µl**

3.2.2.5. Identification of positive transformants

Colonies were selected from the plates to identify positive transformants using PCR. Briefly, each colony was used to prepare an inoculum using 20 µl nuclease-free H_2O that was heated at 95°C for 10 min and the same colony was used to inoculate 5ml LB/amp broth for propagation of positive transformants. PCR was performed using GoTaq® DNA polymerase (Promega, Wisconsin, USA) according to manufacturer's instructions. The PCR was set up as shown in Table 19, using primers that bind to sites on the on the pCold TF vector depicted in Figure 18: pCold-F1 (5' CCACTTTCAACGAGCT GATG 3') and pCold-R (5' TGGCAGGGATCTTAGATTCTG 3').

Table 19: GoTaq PCR reaction for identification of positive transformants

Component	Final Volume (µl)	Final Concentration
5x Green GoTaq reaction buffer	10	1 x (1.5 mM $MgCl_2$)
dNTP mix	1	0.2 mM each dNTP
PCold-F1	0.8	20 pmol
PCold-R	0.8	20 pmol
GoTaq DNA polymerase	0.25	1.25 U
Template DNA	25	<0.5 µg/ 50 µl
Nuclease free water	12.15	-
Final Volume	**50**	

The PCR reactions were placed in a Perkin Elmer thermocycler and the thermal cycling conditions were set according to manufacturer's instructions as shown in Table 20.

Table 20: Cycling conditions for GoTaq PCR reactions

Step	Temperature (°C)	Time	Number of Cycles
Initial denaturation	95	2 min	1
Denaturation	95	1 min	25
Annealing	46	1 min	25
Extension	72	90 secs	25
Final Extension	72	5 min	1
Hold	4	Indefinite	1

A 10 µl volume of each amplicon was analyzed by gel electrophoresis on a 1% agarose gel stained with ethidium bromide. O'GeneRuler DNA Ladder Mix was used as molecular marker with DNA fragments ranging from 100 to 10000. The predicted size of the amplicon was approximately 1672 bp.

Colonies that produced positive PCR reactions were further confirmed using single restriction enzyme analysis as described in section 2.2.2.9, using *BamH1*. The recombinant plasmid was propagated by preparing an O/N culture and purified using

the Pure Yield™ Plasmid Miniprep System (Promega, Wisconsin, USA) as described in Materials and methods 2.2.2.8. The recombinant plasmid was analyzed using gel electrophoresis and restriction enzyme site digestion (Materials and methods 2.2.2.4) and the DNA concentration measured (Materials and methods 2.2.2.5). Glycerol stocks of the positive transformant and pColdTF were prepared.

3.2.2.6. Induction study: time-course protein expression

An induction study was performed using an O/N culture of the recombinant pColdTF-opSPUNC. The overnight culture was prepared by inoculating a 5 ml aliquot of LB/amp broth with a 20 µl aliquot of a glycerol stock of pColdTF-opSPUNC. An additional 5ml volume of LB/amp was inoculated with 10 µl glycerol stock of pColdTF with no gene insert as the negative control. The following day cultures were diluted 1:20 with pre-warmed LB/amp (2 ml culture in 38 ml LB/amp) and incubated for approximately 2 hours at 37°C with shaking at 200 rpm. The OD readings of the cells were measured with a spectrophotometer. When the OD_{600} reached between 0.4 - 0.5, the cultures were incubated at 16°C for 30 min. The cultures were then induced with IPTG to a final concentration of 1 mM and incubated at 16°C with shaking at 200 rpm. Aliquots of 1 ml of the induced cultures were collected at time intervals, t=0 (before induction), t=2 (2 hours post-induction) t=4 (4 hours post-induction) and t = 24 (24 hours post-induction). The cells were harvested from the samples by centrifugation at 16000 × g for 10 min using the bench top centrifuge. The supernatant was discarded and the cells resuspended in a volume of PBS calculated according to the formula OD/0.5 x 150 µl. Similarly, an induction study following the same protocol was performed for pColdTF-opAP92NC construct that was prepared in section 3.2.2.4. Cells were harvested at t=0, t=2, t=4 and t=24.

The proteins were separated using an 8% resolving and a stacking SDS-PAGE gel as described in section 3.2.1.4 with slight modifications, using a midi-gel assembly. The SDS-PAGE gels were stained using Fairbanks staining method (Wong *et al.*, 1999). The constituents of the different Fairbanks solutions are listed in Table. Briefly, Fairbanks A was added to the gels and incubated overnight at room

temperature with gentle agitation. The following day, the gel was heated with solution A until the staining solution reaches boiling point. Once cooled, Fairbanks A was discarded, the gels were rinsed with distilled water and Fairbanks solution B was added to the gel. The heating was repeated. Fairbanks solution C and D were subsequently added in the same manner and discarded. The gel was soaked in distilled H_2O; pictures were taken using a gel documentation system.

Table 21: Fairbanks's staining solutions

Reagents	Fairbanks Stain			
	A	B	C	D
Coomassie R-250 (g)	0.5	0.05	0.02	0
Isopropanol (ml)	250	100	0	0
Glacial acetic acid (ml)	100	100	100	100
Distilled water	650	800	900	900
Total (ml)	**1000**	**1000**	**1000**	**1000**

3.2.2.7. Protein solubility

A solubility study was performed to determine if the recombinant protein was present in the soluble or insoluble phase. Briefly, a 1ml sample of the pColdTF-opSPUNC culture was collected at t=0, t=4 and t=24 hours. The samples were centrifuged at 16000 × g and the pellets were re-suspended in 500 µl cold PBS. Sarcosyl, a mild detergent that releases membrane bound protein was added to the samples at a final concentration of 7.5%. The cells were sonicated using a Branson 220 ultrasonic cleaner (Branson, Connecticut, USA) using 10 × 15 second bursts with a 15 second cooling period on ice between each burst. An 80 µl aliquot of the cell suspension was collected to represent the total protein (soluble and insoluble fraction) and added to 20 µl of 5 × protein loading buffer. Another 80 µl aliquot of sample was collected from the suspension and centrifuged at 16000 × g for 20 min. The supernatant was collected (soluble fraction) and added to 20 µl of the loading buffer. The pellet (insoluble fraction) was re-suspended in 500 µl cold PBS, 80 µl

was collected and added to 20 µl loading buffer. All the samples were heated for 5 min at 95°C. The proteins were separated using an 8% resolving and a stacking SDS-PAGE gel as described in section 3.2.1.4. The SDS-PAGE gel was stained using a staining solution containing Coomassie Brilliant Blue R-250 that forms a complex with SDS. The staining solution consists of 45% methanol, 10% glacial acetic acid and 0.2% Coomassie Brilliant Blue. The gels were stained O/N while swirling, stain was discarded, and destaining solution (45% methanol, 7% glacial acetic acid, 48% dH_2O) was added for 3 hours. Cells were soaked in a 30% methanol/3% glycerol solution, placed on white filter paper, covered with plastic film and then dried using a gel dryer set at 70°C with constant vacuum applied.

Similarly, a solubility study following the same protocol was performed for pColdTF-opAP92NC construct that was prepared in section 3.2.2.4. Samples were harvested at t=4h and t=24h. However, the SDS-PAGE gel was prepared using a midi-protein gel assembly and was stained using the Fairbanks method described in 3.2.2.6.

3.2.2.8. Purification of soluble recombinant His-tagged protein under native conditions

The recombinant proteins were purified from soluble fractions of the preparation using native conditions for purification of His-tagged proteins on the NI-TED columns. A 5 ml volume of LB/amp was inoculated with 10 µl glycerol stock of pColdTF-opSPUNC. An additional 5 ml volume of LB/amp was inoculated with 10 µl glycerol stock of pColdTF-opAP92NC and as a negative control an aliquot of broth was inoculated with pColdTF with no gene insert. The following day each culture was diluted 1:20 with pre-warmed LB/amp (2 ml culture in 38 ml LB/amp) and incubated for approximately 2 hours at 37°C with shaking at 200 rpm. When the OD_{600} reached between 0.4 - 0.5 (logarithmic growth phase), the cultures were incubated at 16°C for 30 min. The cultures were then induced with IPTG to a final concentration of 1 mM and incubated at 16°C with shaking at 200 rpm. At 24 hours post-induction, culture was centrifuged at 6000 x g at 4°C for 20 min. The pellet was stored at -20°C. The following day, the pellet was thawed on ice and 5 ml

BugBuster protein extraction reagent (final concentration 200mg/ml) (Merck, New Jersey, USA), 1µl *Lysozyme* (final concentration of 1 mg/ml) (Novagen, California, USA) and 5 µl *Benzonase* (final concentration of 50 U/ml) (Novagen, California, USA) was added to the pellet to completely resuspend the pellet. The culture was stirred on ice for 30 min. The suspension was then sonicated using 220 ultrasonic cleaner on ice for 10 15 second bursts with 10 15 second cooling periods in between. Centrifugation at 10 000 x g for 30 min at 4°C was necessary to remove cellular debris. The supernatant was stored in a clean tube on ice.

The protein was purified using Protino®Ni-TED resin according to the manufacturer's instructions (Machery-Nagel, Pennsylvania, USA). Protino®Ni-TED products enable fast and convenient purification of recombinant polyhistidine-tagged proteins by immobilized metal ion affinity chromatography (IMAC). Protino®Ni-TED is a dry silica-based resin precharged with Ni^{2+} ions. Protein binding is based on the interaction between the polyhistidine tag of the recombinant protein and immobilized Ni^{2+} ions. The chelating group of Protino® Ni-TED is based on TED (tris-carboxymethyl ethylene diamine), a strong pentadentate metal chelator. TED occupies five of the six binding sites in the coordination sphere of the Ni^{2+} ion. Compared to TED, other chelating groups such as NTA (nitrilotriacetic acid) have four binding sites available for the Ni^{2+} ion. The additional chelation site of TED with Ni^{2+} minimizes metal leaching during purification and increases specificity for polyhistidine-tagged proteins. Protino nickel columns were equilibrated with 2 ml Lysis-Equilibration-Wash (LEW) buffer (50 mM NaH_2PO_4, 300 mM NaCl), allowed to drain by gravity. The supernatant was added to the column and the column, allowed to drain by gravity. The column was washed with 2 × 2 ml LEW buffer. The protein was eluted with 3.5 ml of elution buffer containing 50 mM NaH_2PO_4, 300 mM NaCl and 250 mM imidazole. The eluates were collected in 500 µl aliquots and labeled using consecutive numbers. Flow through and eluates one to six were separated and visualized on a mini SDS PAGE gel as described in section 3.2.2.6, using Coomassie staining solution overnight. Similarly, protein from pColdTF-opAP92NC cell lysates from a 40 ml 24-hour post-induction culture was purified from the native

state. The eluates were separated on a SDS-PAGE gel and stained using the Fairbanks method described in 3.2.2.6.

The protein concentrations in each eluate were measured using the Quant-iT Protein Assay Kit (Invitrogen, Maryland, USA) according to the manufacturer's instructions. The Quant-iT Protein Assay is supplied with three standards. A Quant-iT working solution was prepared by diluting the Quant-iT Protein reagent 1:200 in Quant-iT Protein buffer. An aliquot of 190 µl of Quant-iT working solution was loaded into each of the tubes used for the standards. A 10µl aliquot of each Quant-iT standard was added to the appropriate tube and mixed by vortexing for 2-3 seconds. Protein samples diluted 1:200 in Quant-iT working solution were briefly mixed by vortexing. All tubes were allowed to incubate at room temperature for 15 min. The Qubit fluorometer was calibrated for each measure using the three standard solutions prepared. The reading was recorded in µg/ml and the dilution was considered when calculating the final concentration.

Preparation of mock antigen:
pCold TF plasmid with no genes inserted was used to prepare mock antigen. Briefly, the plasmid was used to transform *E. coli* OverExpress cells as described using heat-shock method described in Section 3.2.2.3 and recombinant protein was prepared as described in Section 3.2.2.6 and 3.2.2.7. The protein was purified using Ni columns and used as a mock antigen in ELISA.

3.3. Results

3.3.1. Preparation of pcDNA3.1/V5-His-TOPO/LacZ DNA plasmid DNA for transfection experiments

To identify positive transformants following transformation of TOP10 cells with pcDNA3.1/V5-His-TOPO/*LacZ*, selected colonies were cultured overnight and plasmid DNA purified the following morning. DNA was digested with *BamH1* and double digested with *BamH1* and *Not1*. The digested DNA was separated and

visualized on a 1% agarose gel depicted in Figure 19. *BamH1* was used to linearize the DNA. The double digestions were performed to identify positive transformants. Bands of ~3058 bp indicated the presence of the *LacZ* gene. The vector, without insert, is approximately 5500 bp in length when linearized.

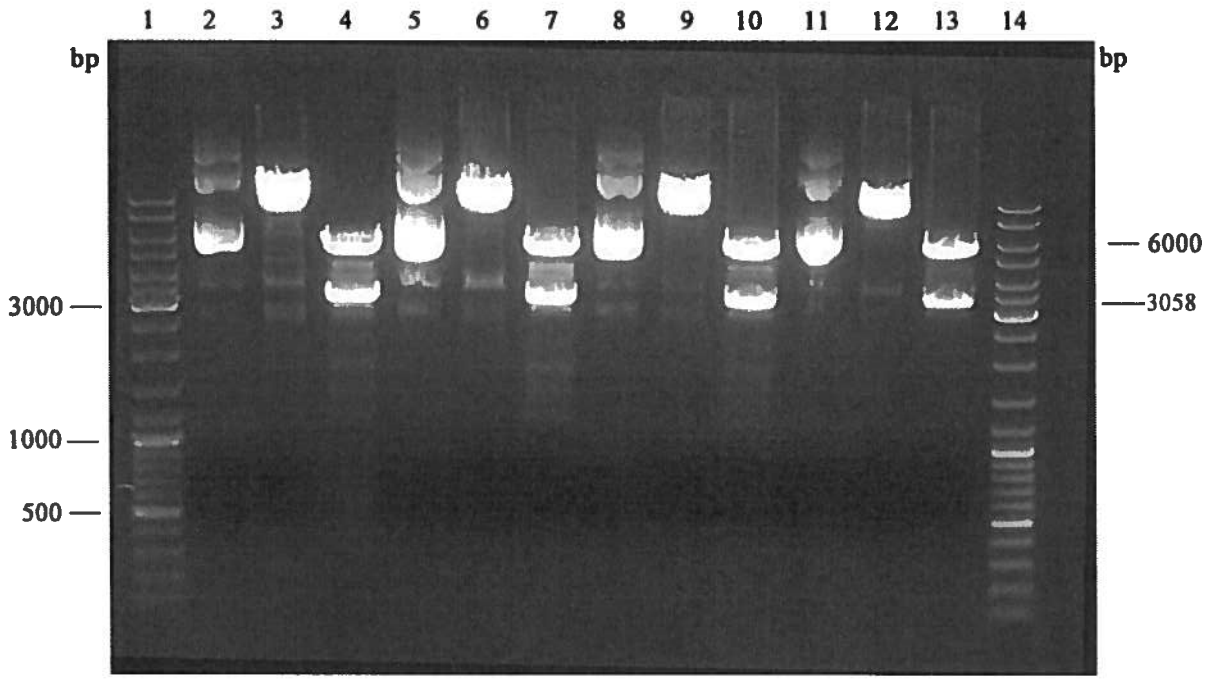


Figure 19: Agarose electrophoretic gel of restriction enzyme analysis of plasmids obtained from purification of an overnight culture inoculated with TOPO *LacZ* V5/His Tag plasmid DNA. Lane 1 – O'GeneRuler DNA ladder mix; Lane 2 – undigested plasmid DNA from Colony 1, Lane 3 – Single digest of Colony 1 using *BamH1*, Lane 4 – Double digest of Colony 1 using *BamH1* and *Not1*, Lane 5 – Colony 2 undigested plasmid DNA, Lane 6 – Colony 2 digested with *BamH1,* Lane 7 – Colony 2 double digested using *BamH1* and *Not1*, Lane 8 – Colony 3 undigested plasmid DNA, Lane 9 – Colony 3 digested with *BamH1*,Lane 10 – Colony 3 double digested using *BamH1* and *Not1*, Lane 11 – Colony 4 undigested plasmid DNA, Lane 12 – Colony 4 digested with *BamH1*, Lane 13 – Colony 4 double digested using *BamH1* and *Not1*, Lane 14 - O'GeneRuler DNA ladder mix.

3.3.2. Optimization of transfection experiments

Mammalian cell lines that were used in the transfection experiments differed in growth and maintenance requirements. HEK293 cells were passaged using trypsin-versene instead of trypsin as these cells proved to be more fragile during handling and passaging cells. Vero cells were constantly passaged and had a constant growth rate. HeLa cells were the most fastidious and did not detach easily from the flask surface as HEK293 cells did. HEK293 cells were used for initial optimization studies. After comparing the different cell lines, it was decided that HeLa cells were suitable for the experiment. Lipofectamines may display cytotoxic effects on certain cell lines and growth of HeLa cells post transfection was similar in the control well containing no transfection reagent or DNA than in wells containing transfection reagent.

During optimization efforts, optimal ratios were determined, different harvesting times post-transfection were tested and different X-gal staining times were also applied, to verify whether cells were not adequately stained. Long staining times can also lead to non-specific staining. Transfection efficiency of cells transfected with pcDNA3.1/V5-His-TOPO/*LacZ* plasmid DNA was determined by obtaining a percentage by dividing the number of blue (stained) cells by the total number of cells in the microscope field. The results, with different ratios were tabulated in Table 22. The low success-rate of transfection with FuGene lipofectamine was equal to that of GeneJuice transfection reagent.

HEK293 cells transfected with pcDNA3.1/V5-His/CCHFVNC were tested by IFA for CCHFV antigen using human serum from a survivor of CCHFV with a known antibody response (lab number VBD 29/10). Antibody-antigen reaction was not demonstrated, suggesting lack of CCHFV antigen expression.

Table 22: FuGene 6 reagent transfection optimization: transfection ratios, type of cell line used and results of β galactosidase assay.

Cell Line	Transfection ratio	Time post-transfection	Staining time	Transfection efficiency
HEK293	Controls	48 hours	4 hours	0%
	3:1	48 hours	4 hours	<10%
	3:3	48 hours	4 hours	<10%
	6:1	48 hours	4 hours	<10%
	6:3	48 hours	4 hours	<10%
	9:3	48 hours	4 hours	<10%
	9:6	48 hours	4 hours	10%
	Controls	72 hours	16 hours	0%
	3:3	72 hours	16 hours	<10%
	6:3	72 hours	16 hours	<10%
	9:6	72 hours	16 hours	10%
	Controls	6 days	16 hours	0%
	3:3	6 days	16 hours	<10%
HEK293	6:3	6 days	16 hours	<10%
	9:6	6 days	16 hours	<10%
	3:3	10 days	5 hours	<10%
	6:3	10 days	5 hours	<10%
	9:6	10 days	5 hours	~10-15%

Cell Line	Transfection ratio	Time post-transfection	Staining time	Transfection efficiency
HeLa cells	Controls	3 days	5 hours	0%
	9:6	3 days	5 hours	10%
	Controls	5 days	5 hours	0%
	9:6	5 days	5 hours	~10-15%
Vero 76 cells	Controls	3 days	5 hours	0%
	9:6	3 days	5 hours	10%
	Controls	5 days	5 hours	0%
	9:6	5 days	5 hours	~10-15%

HeLa cells transfected with GeneJuice transfection reagent at ratios of 3:1, 3:2 and 3:3 µl (transfection reagent: µg of DNA), were harvested 72 hours post-transfection, and stained using β-galactosidase assay for five hours. The control cells and cells transfected at a ratio of 3:1, no blue cells were observed. Blue stained cells were observed in cells transfected with a higher ratio of transfection reagent:DNA (3:3).

Figure 20 shows photographs taken during observation using a light microscope. The photographs depict scattered blue cells, surrounded by translucent cells. Transfection efficiency did not exceed 20% during optimization and after repeated transfections.

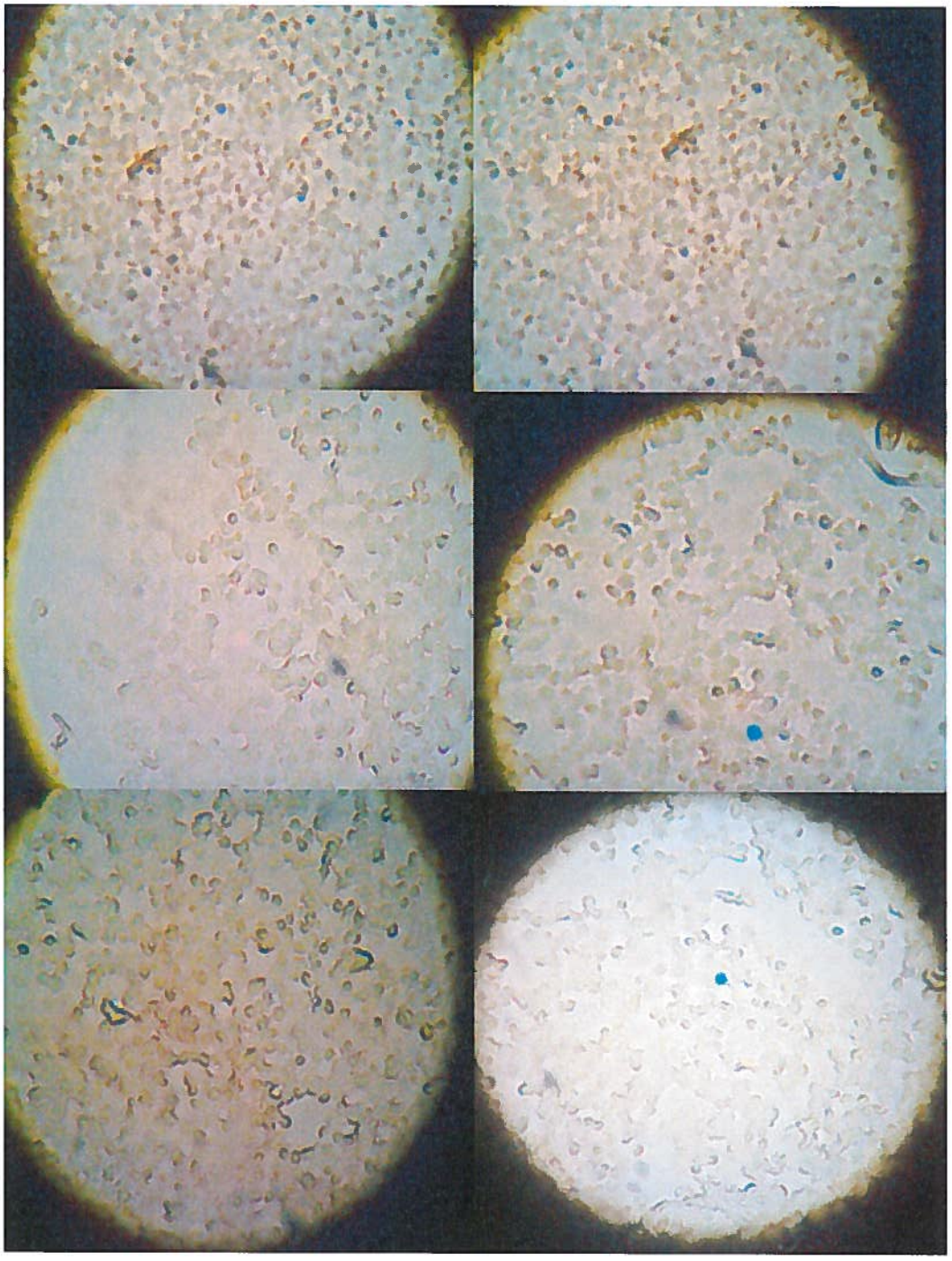


Figure 20: HeLa cells transfected with a vector containing *LacZ* gene coding for β-galactosidase expressed. Cells were stained using X-gal reagent in a B-galactosidase assay resulting in blue cells depicted.

3.3.3. Mammalian cell lysis and characterization of expressed His tagged NC by Western blot analysis

Protein expressed by mammalian cells were characterized by Western blot analysis using anti-Histidine antibody for detection of recombinant antigen. Following cell lysis, the lysates were separated on a 8% SDS-PAGE gel. A bacterially expressed recombinant antigen (yellow fever envelope protein EDIII) expressed using a pQE construct as a fusion protein with a His-tag included in the Western Blot analysis as a positive control. A 13 kDa protein representing the control antigen was detected in lane 2. The control antigen was detected in the Western Blot, however, no His-tagged proteins were detectable in lanes 3-5 of Figure 21.

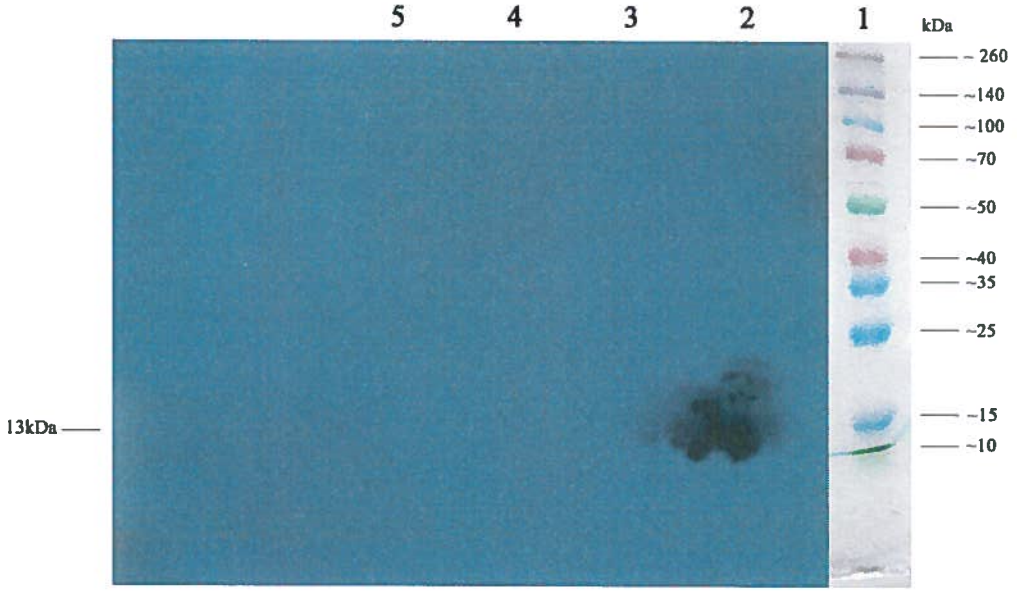


Figure 21: Western Blot analysis of mammalian cell lysates and control antigen EDIII yellow fever containing a His-tag. Lane 1 – Spectra Broad Range protein marker, Lane 2 – Yellow fever EDIII control antigen, Lane 3 – cell lysate from pcDNA3.1D/V5-His/*LacZ* expression control vector, Lane 4 – pcDNA3.1D V5-His/CCHFNC construct pellet, Lane 5-pcDNA3.1D/V5-His/CCHFNC supernatant.

3.3.4. Codon optimization of CCHFV AP92 NC gene

Figure 22 demonstrates the distribution of codon frequency usage along the length of the gene sequence of the S segment of CCHF AP92 NC before optimization. The frequency at which each codon is utilized by the *E. coli* host is expressed as a percentage with 100% being optimal usage. The codon usage for the NC gene of CCHFV ranged from 10% to 70% along the length of the gene. The CAI was calculated as 0.62 indicating that several of the codons used by the native CCHF gene were less frequently recognized by *E. coli.*

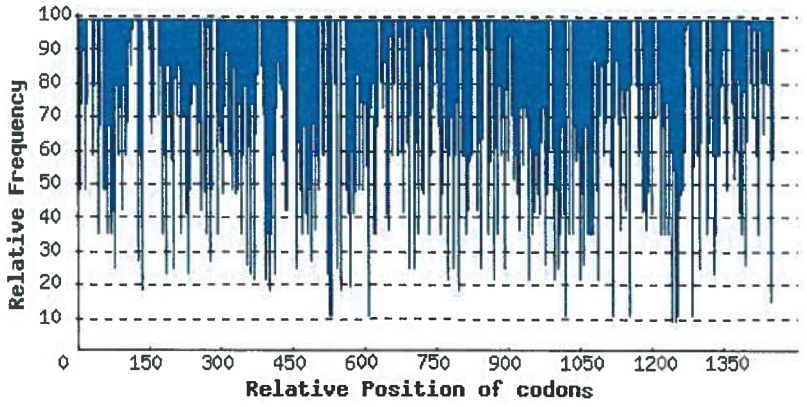


Figure 22: The distribution of codon usage frequency along the length of the gene sequence of the of CCHF AP92 NC before optimization.

After codon optimization the codon usage of *E. coli* along the length of the CCHF NC gene significantly increased with values ranging from approximately 70% to 100%, with the exception of four codons with a 40-50% usage. The CAI of the codon optimized gene was calculated as 0.84 as demonstrated in Figure 23.

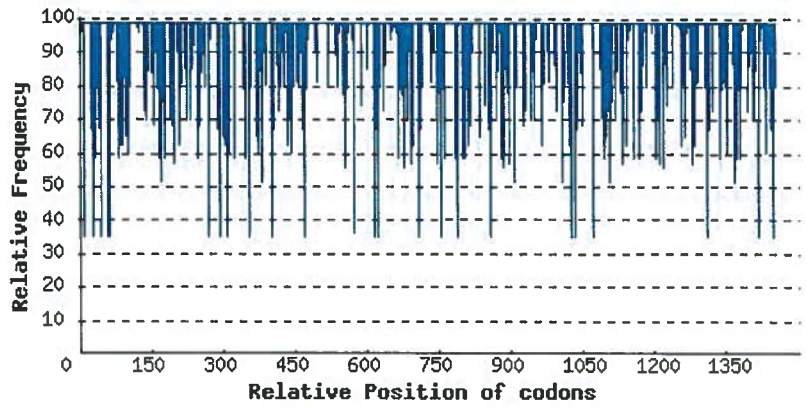


Figure 23: The distribution of codon usage frequency along the length of the gene sequence of the of CCHF AP92 NC after optimization.

To further illustrate the codon usage Figure 24 shows the percentage distribution from the CCHF NC gene recognized by *E. coli*. The codons have been grouped according to usage by *E. coli* where 0-10% is the group less frequently recognized and 91-100% is the group most frequently utilized. Less than half of the codons (38%) are frequently utilized by *E. coli.*

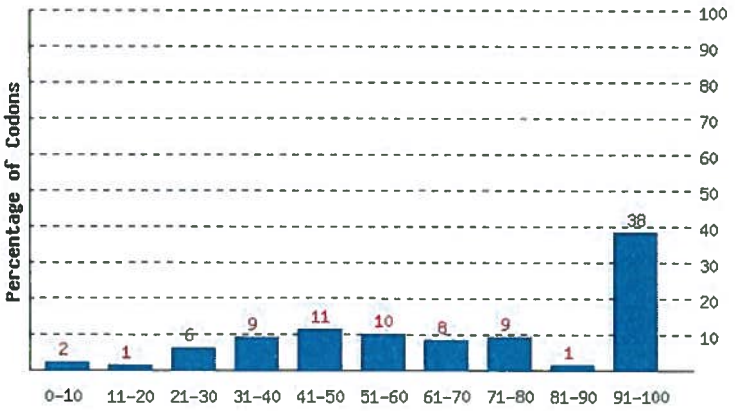


Figure 24: The percentage distribution of codons in computed codon quality groups before optimization.

The proportion of codons frequently utilized by *E. coli* was increased to 66% after codon optimization (Figure 25).

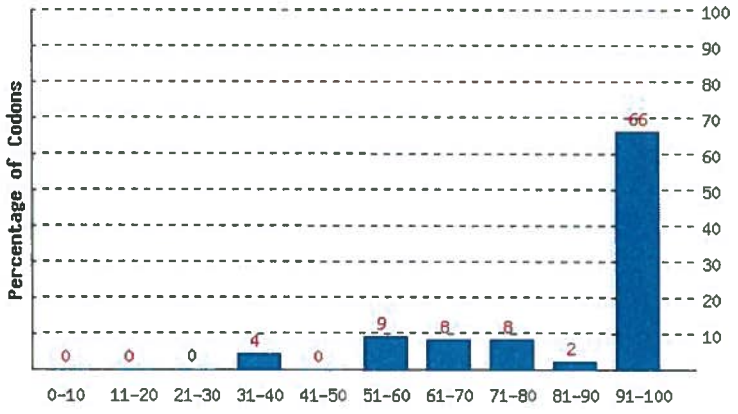


Figure 25: The percentage distribution of codons in computed codon quality groups after optimization

GC content and unfavourable peaks were optimized to prolong the half-life of the mRNA. The average GC content as shown in Figure 26 was 47.78% before optimization and increased to 49.89% after optimization (Figure 27). The stem-loop structures, which impact ribosomal binding and stability of mRNA, were broken.

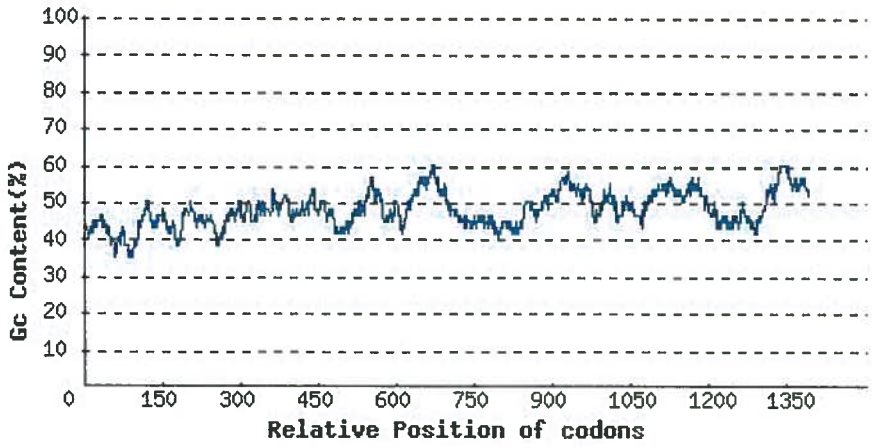


Figure 26: Average GC content before optimization.

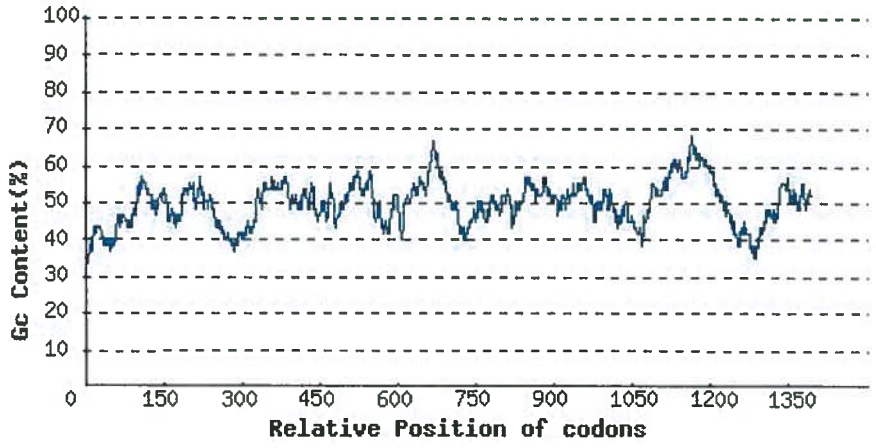


Figure 27: Average GC content after optimization.

The nucleotide sequence of the native gene encoding the NC and the codon optimized gene were aligned and translated to show that the predicted amino acid sequence was identical for both coding regions. The nucleotide sequence of the synthesized gene supplied in pUC57 was confirmed by GenScript.

3.3.5. Cloning of gene encoding codon optimized CCHF AP92 NC from pUC57 into pCold TF bacterial expression vector

Figure 28 depicts restriction enzyme analysis of six colonies from cells transformed with pUC57-opAP92NC plasmid. The expected 1443 bp fragment (lanes 3-8) after digestion with *BamH1* is shown confirming the presence of the opAP92NC gene in pUC57 vector. The bands were excised from the gel and purified from the gel slice using a DNA purification kit.

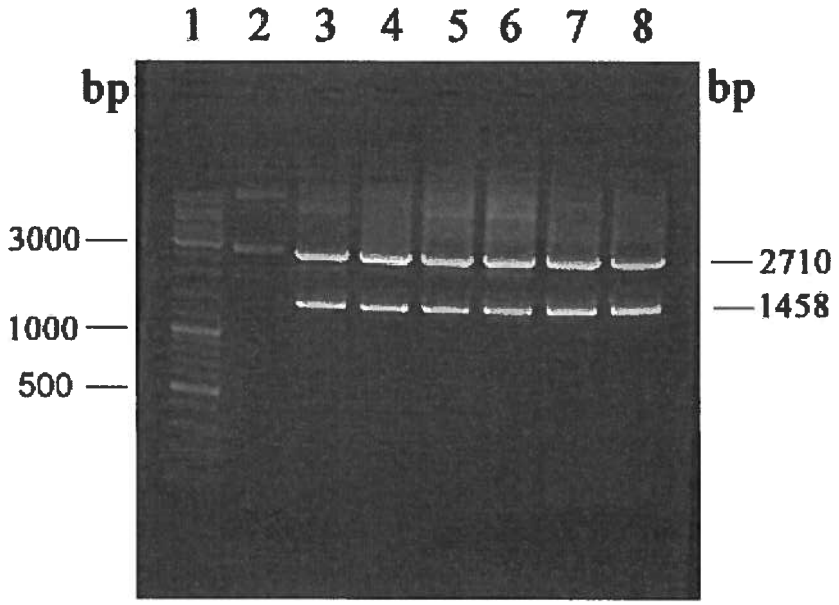


Figure 28: Agarose gel electrophoretic analysis of restriction enzyme analysis of a plasmid obtained from pUC57 vector supplied by GenScript. Lane 1 – O'GeneRuler DNA ladder mix molecular weight marker molecular weight marker; Lane 2 – 1 µl of undigested pUC57 opAP92NC; Lane 3-8 – 20 µl digestion mix from colonies 1-6 pUC57-opAP92NC digested with *BamH1*.

DNA concentration of the purified insert was measured as 72 ng/µl with a A280/A260 ratio of 1.87. The linearized pColdTF vector had a concentration of 98 ng/µl and A280/A260 ratio of 1.85. The purified insert and vector was run on a 1% agarose gel prior to ligation as depicted in Figure 29.

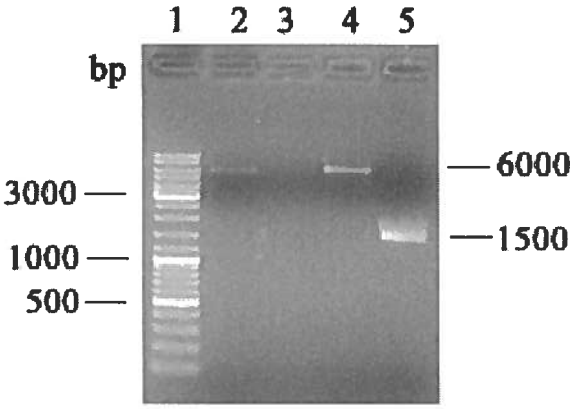


Figure 29: Purified linearized vector and insert rescued from pUC57 vector. Lane 1 – O'GeneRuler DNA Ladder mix, Lane 2 – 1 µl of linearized vector, Lane 3: no sample, Lane 4 – 2 µl of purified linearized vector, Lane 5 – Purified AP92 NC gene insert (2 µl) rescued from pUC57 vector.

After transformation of OverExpress cells using the ligation reaction from vector and insert depicted in Figure 29, colony PCR directly from the colonies on the plate, was performed to establish the presence of a positive transformant prior to DNA purification. GoTaq polymerase was used to amplify DNA from the template after lysis of cells from selected colony in nuclease-free water. One colony was positive as depicted in Figure 30, and the other five were negative (~200bp was amplified from pCold religated without gene of interest).

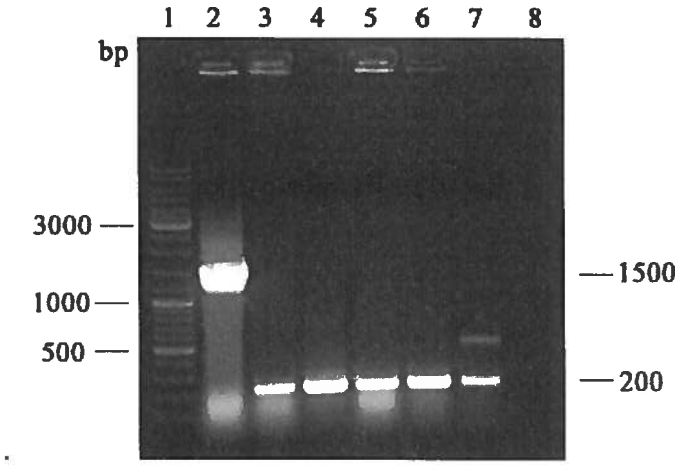


Figure 30: Agarose gel electrophoretic analysis of colony PCR of a plasmid DNA obtained from pCold TF-opAP92NC colonies using GoTaq polymerase. Lane 1 – O'GeneRuler DNA ladder mix molecular weight marker; Lane 2 -7 PCR product, Lane 8 – Negative control.

An overnight culture using the remainder of colony one was prepared, and plasmid DNA was extracted from the bacterial culture. Plasmid DNA concentration was 509 ng/µl and the ratio of A280/A260 was 1.87. The plasmid DNA was digested using *BamH1* to confirm the presence of the AP92 NC gene insert in pColdTF vector. Double digestion resulted in linearized pColdTF vector (~5700 bp) and ~1500 bp gene of interest, depicted in Figure 31. This construct was designated pColdTF-opAP92NC.

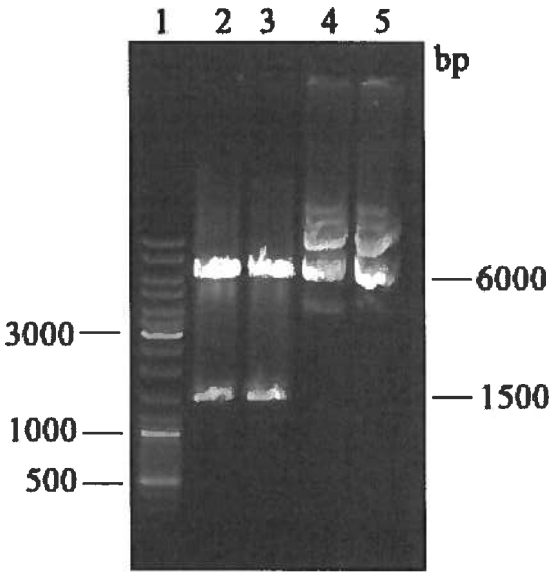


Figure 31: Agarose gel electrophoretic analysis of restriction enzyme analysis of a plasmid that yielded a positive PCR product (1443 bp). Lane 1 – O'GeneRuler DNA ladder mix molecular weight marker, Lane 2-3, op AP92 NC excised from pColdTF vector, Lanes 4-5, undigested plasmid DNA pCold opAP92NC.

3.3.6. Bacterial expression of codon-optimized NC protein of isolate SPU415/85 and solubility study

Protein fractions collected at t = 0, t = 2 h, t = 4 h and t = 24 h during the induction study were separated on an 8% resolving gel and 4% stacking gel to verify the optimal time post- induction for maximum yield of proteins. The expected size of the recombinant pColdTF-opSPUNC was 106 kDa and TF protein is 52 kDa as shown in

Figure 32, Figure 33 and Figure 34. During the solubility study, protein fractions were collected at t = 0, t = 4 and t = 24, to determine in which time frame the protein of interest is expressed in the soluble phase. The time-course of protein expression reveals a high yield of protein 24 hours post-induction. Sarcosyl released protein from inclusion bodies and protein present in the supernatant confirms that the protein can be purified from the soluble phase. The protein was previously expressed and purified from the insoluble phase (Samudzi *et al.*, 2012). In this study the culture was incubated at a lower temperature (22°C compared with 37°C) prior to induction.

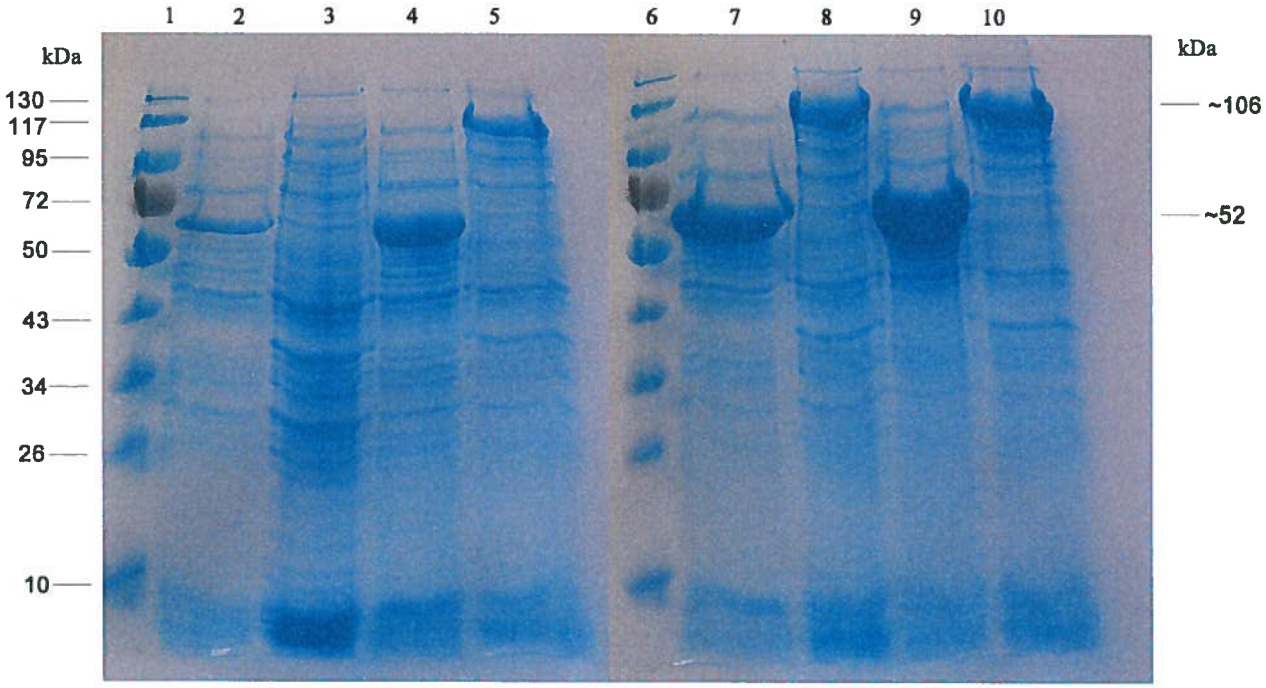


Figure 32: SDS PAGE analysis of proteins expressed using the recombinant expression vector, pColdTF-opSPUNC in OverExpress (DE3) competent cells. Lanes each contain 25µl of sample: Lane 1 – Prestained Protein Ladder, Lane 2 – pColdTF negative control expression before induction; Lane 3 – pColdTF-opSPUNC expression before induction; Lane 4 – pColdTF expression 2 hours after induction; Lane 5 – pColdTF-opSPUNC expression 2 hours after induction; Lane 6 – Prestained Protein Ladder, Lane 7 – pColdTF expression 4 hours after induction; Lane 8 – pColdTF-opSPUNC expression 4 hours after induction. Lane 9 – pColdTF expression 24 hours after induction; Lane 10 – pColdTF-opSPUNC expression 24 hours after induction.

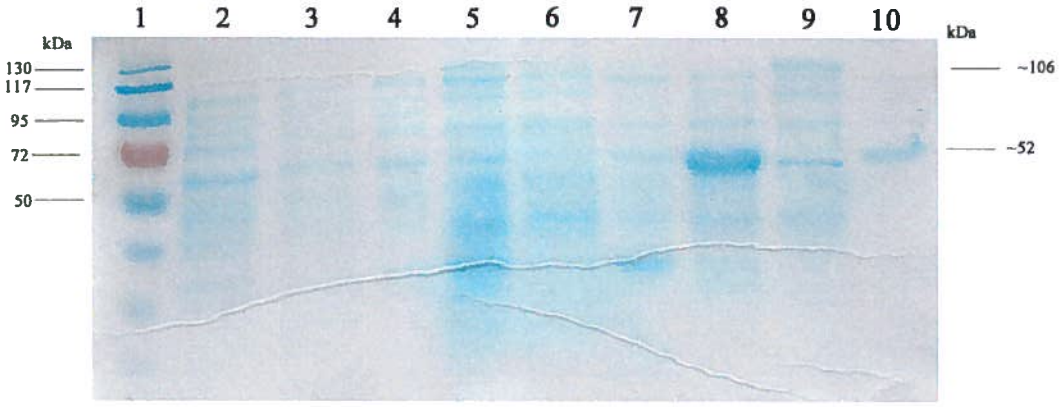


Figure 33: SDS PAGE analysis of recombinant pColdTF-opSPUNC proteins to determine solubility of the expressed protein. Lane 1 – Prestained protein marker; Lanes 2-4, pColdTF expression control before induction: Lane 2 – total cell fraction, Lane 3 – supernatant, Lane 4 – pellet; Lanes 5-7, pColdTF-opSPUNC before induction, Lane 5 – total cell fraction, Lane 6 – supernatant, Lane 7 – pellet; Lanes 8-10, pColdTF expression control 4 hours post induction, Lane 9 – total cell fraction, Lane 9 – supernatant, Lane 10 – pellet.

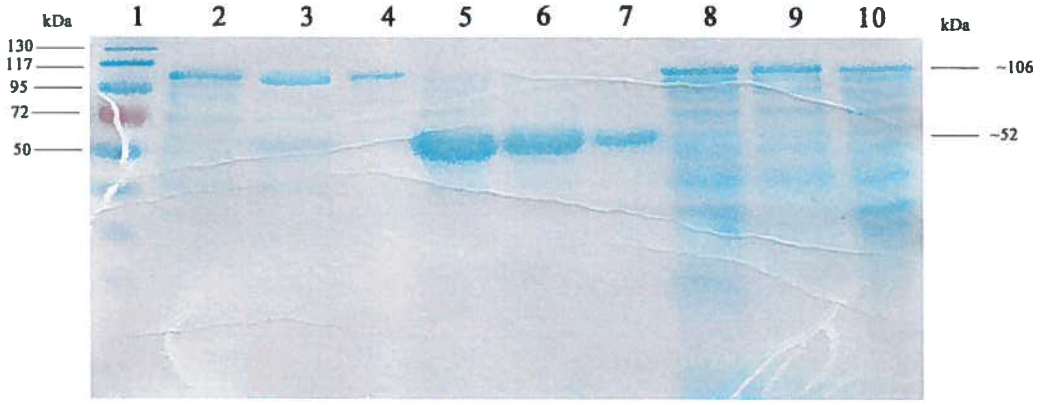


Figure 34: SDS PAGE analysis of recombinant pColdTF-opSPUNC proteins to determine solubility of the expressed protein. Lane 1 – Prestained protein marker; Lanes 2-4, pColdTF-opSPUNC before induction, Lane 2 – total cell fraction, Lane 3 – supernatant, Lane 4 – pellet; Lanes 5-7, pColdTF expression control 24 hours post-induction, Lane 5 – total cell fraction, Lane 6 – supernatant, Lane 7 – pellet; Lanes 8-10, pColdTF-opSPUNC 24 hours post induction, Lane 8 – total cell fraction, Lane 9 – supernatant, Lane 10 – pellet.

After determination of the optimum time required for protein expression, and whether the expressed protein was soluble, protein was purified from a 40 ml culture harvested at 24 hours post-induction. His-tagged proteins were eluted from a Nickel column and the purity of the proteins was analysed by SDS PAGE. The flow-through from the nickel column was included on the gel shown in Figure 35. His tagged protein was eluted with imidazole and 0.5 ml fractions of the eluates were collected and labeled numerically.

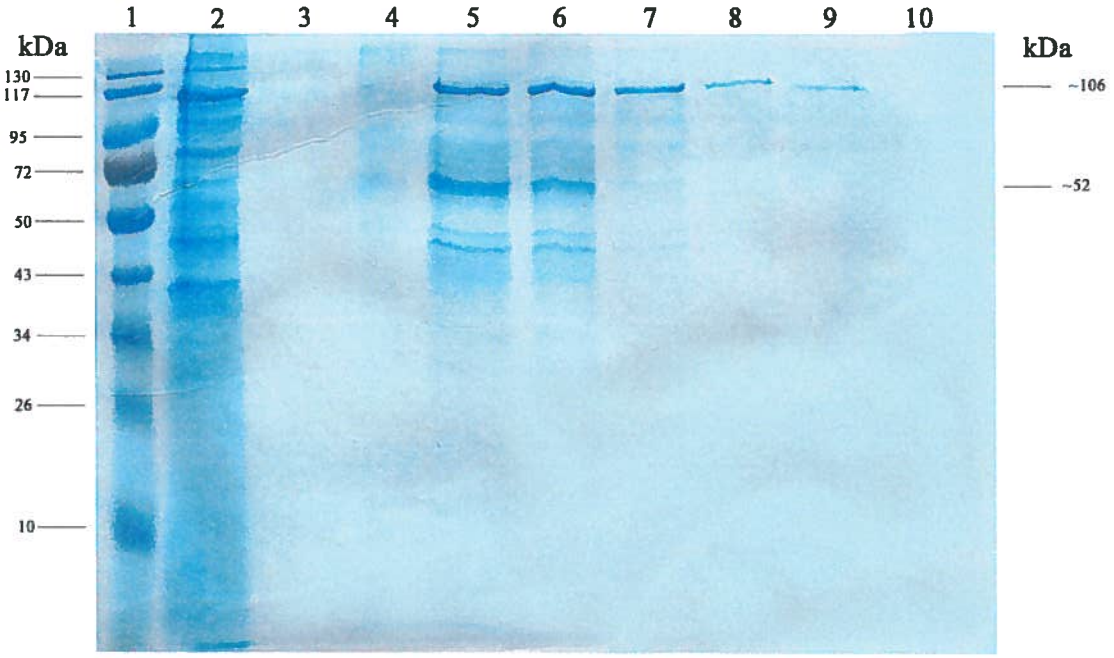


Figure 35: SDS PAGE analysis of His-tag-purified pColdTF-opSPUNC 24 hours after induction. Lane 1 – Prestained protein marker; Lane 2 – flow through, Lane 3-4 – Wash, Lane 5 – eluate 1, Lane 6 – eluate 2, Lane 7 – eluate 3, Lane 8 – eluate 4, Lane 9 – eluate 5, Lane 10 – eluate 6.

Protein concentrations of each fraction were measured as described in Materials and methods 3.2.2.8. Eluate one had a concentration of 590 µg/ml, eluate two 452 µg/ml, eluate three 402 µg/ml and eluate four and five were pooled and had a concentration of 262 µg/ml. Two bands were present in the eluted fractions, a band of approximately 108 kDa the expected size of the expressed CCHFV NC and TF fusion protein and an additional protein at approximately 52 kDa.

3.3.7. Bacterial expression of pCold-TF-optAP92NC using IPTG induction and solubility study

An induction study for the pColdTF-opAP92 construct in OverExpress cells was performed as described for pColdTF-opSPUNC. The protein was optimally expressed at 24 hours post induction as shown in Lane 6, Figure 36. To determine the presence of the protein in the soluble phase, a solubility study was performed. Protein was present in the soluble phase post-induction (see Lane 3, Figure 36). Cells were harvested before induction, 4 hours post-induction and 24 hours post-induction. The results of the induction and solubility study are shown in Figure 36.

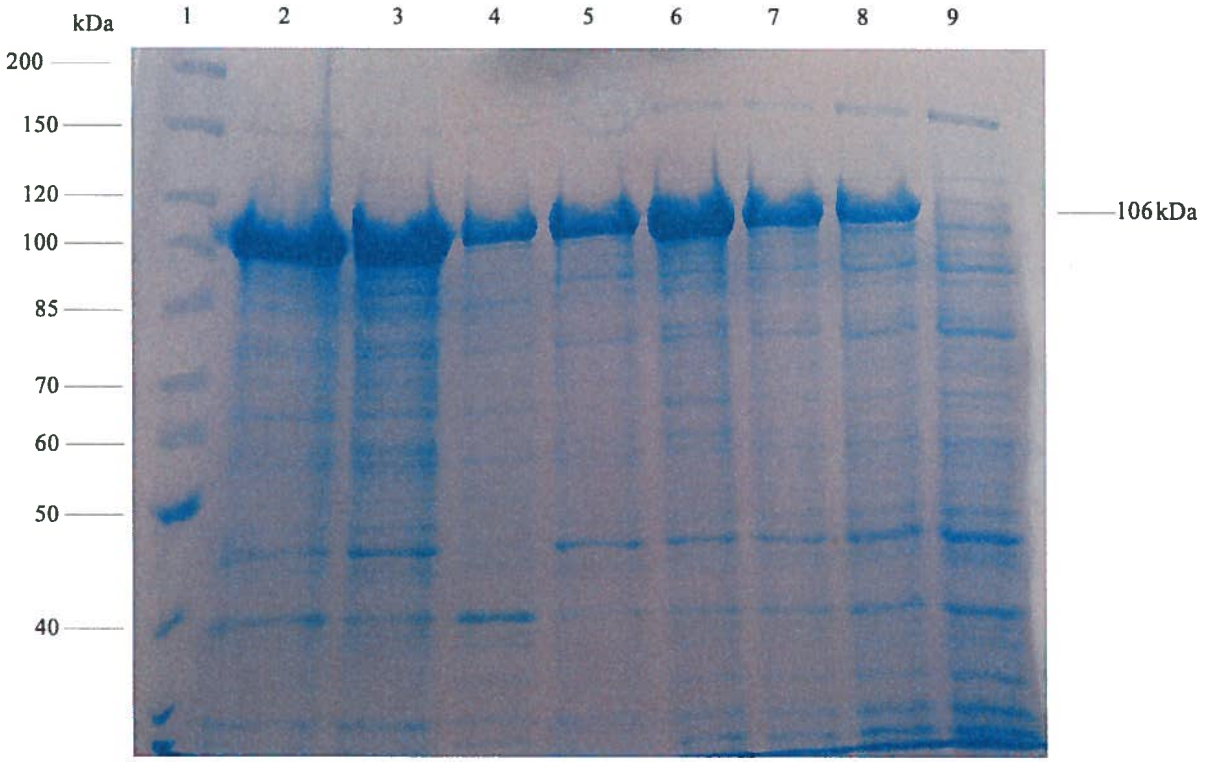


Figure 36: SDS PAGE analysis of proteins expressed in a solubility study using the recombinant expression vector, pColdTF-opAP92NC in OverExpress (DE3) competent cells. Lanes each contain 25µl of sample: Lane 1 – Unstained protein marker; Lanes 2-3 pColdTF-optAP92NC 24 hours post-induction, Lane 2 – pellet, Lane 3 – supernatant; Lanes 4-5, pColdTF-optAP92NC 4 hours post-induction, Lane 4 –pellet, Lane 5 –supernatant; Lanes 6-9, Total cell fraction of pColdTF-optAP92NC, Lane 6 – 24 hours post-induction Lane 7 – 4 hours post-induction; Lane 8 – 2 hours post-induction, Lane 9 – before induction (t=0).

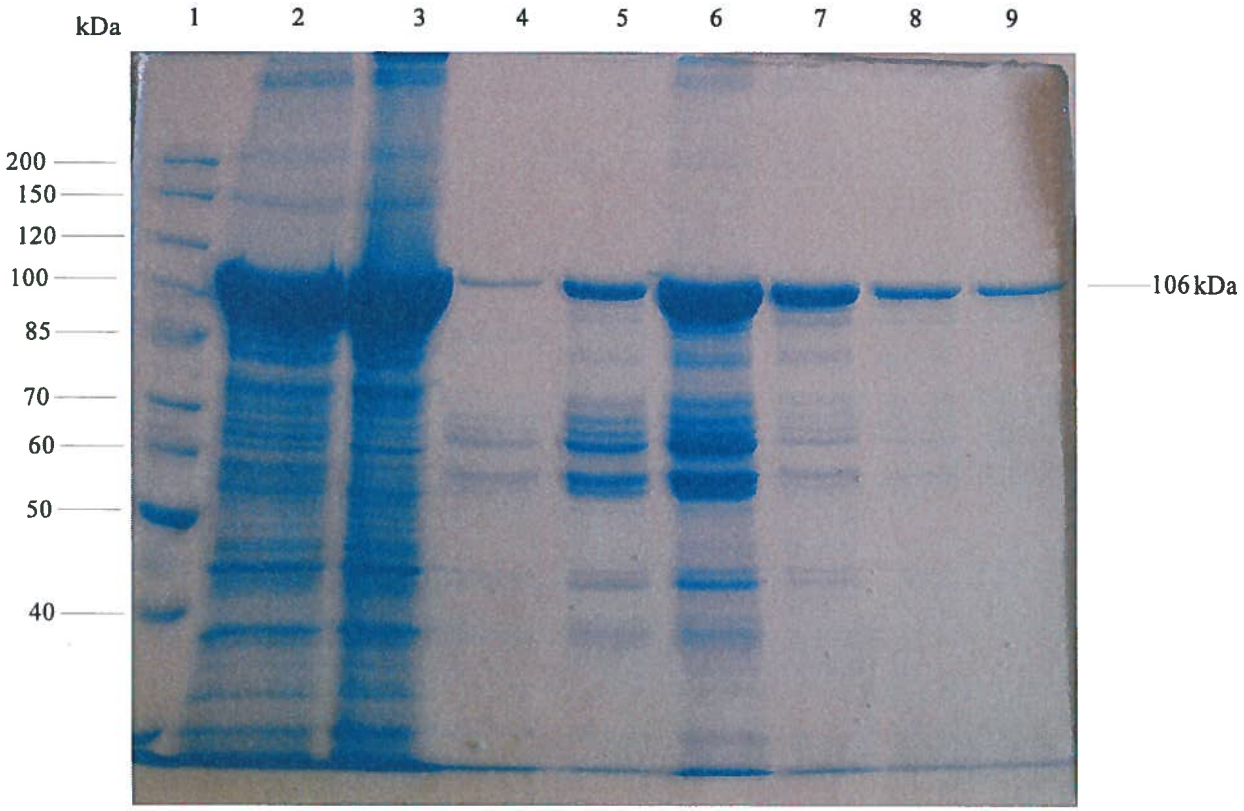


Figure 37: SDS PAGE analysis of pColdTF-opAP92NC protein purified using Ni+ column. Lanes each contain 25 µl of sample: Lane 1 – (Prestained Protein Ladder,); Lane 2 – crude cell lysate, Lane 3 – Flowthrough, Lane 4 – Wash, Lane 5 – Eluate 1, Lane 6 – Eluate 2, Lane 7 – Eluate 3, Lane 8, Eluate 4, Lane 9 – Eluate 5.

Post-purification, protein concentrations of each fraction were measured as described in Materials and methods 3.2.2.8. Eluate one had a concentration of 616 µg/ml, eluate two 1676 µg/ml, eluate three 732 µg/ml and eluate four and five were pooled and had a concentration of 546 µg/ml. Eluates 3-5, revealed higher purity than other eluates, as non-specific bands were present in lanes five and six in Figure 37.

Summary

Mammalian expression systems are known to produce a low yield of recombinant proteins. There are many variables that affect protein expression in mammalian systems. Three different cell types were utilized in this study. Plasmid DNA for transfection into eukaryotic cells must have a high purity and be free from phenol, salts and endotoxins. Hence plasmid was prepared using endotoxin free plasmid purification kit. Contaminants can damage the cells and salt interferes with lipids decreasing transfection efficiency. The CMV promoter on pcDNA3.1D/V5-His/ TOPO vector should allow efficient expression of the gene inserted in frame into the vector (Chen & Okayama 1987). Two different lipofectamine-based reagents were compared. The following factors were considered during mammalian cell transfection: type of cell line, growth media constituents, incubation conditions, possible toxicity of lipofectamine reagents to cells, purity and concentration of plasmid DNA, choice of antibiotics, and ensuring that cells are serum-free during the transfection step. HeLa cells transfected with GeneJuice demonstrated the most promising results, however, the transfection efficiency was too low to produce sufficient yield of recombinant proteins for downstream applications. Initially mammalian expression was selected to optimize the possibility that epitopic sites would be present as in the native protein. In an unrelated study in our laboratory a bacterial expressed NC was shown to adequately detect specific antibody against CCHFV in survivors of infection. Hence mammalian expression was abandoned in favor of bacterial expression system.

Bacterial expression that produces a higher yield of proteins was adopted. pColdTF DNA Vector, which is a fusion cold shock expression vector that expresses Trigger Factor (TF) chaperone as a soluble tag, was selected to express NC genes. It is difficult to predict whether a protein will be expressed from a heterologous host however it is possible to optimize a gene so that the codons are modified to take into account the codon bias of a particular host. The nucleotide sequence for the open reading frame of the Greek AP92 gene encoding the NC was analysed using the Rare Codon Analysis Tool software (download available on GenScript website at http://www.genscript.com). Optimizing the codons for expression in *E. coli* improved

CAI values and the proportion of codons frequently used by *E. coli,* which are generally associated with enhanced protein expression in *E. coli.* Hence it was decided to have a codon optimized gene synthesized. The outcome was a protein that was readily expressed under standard conditions recommended by the suppliers of the pColdTF vector when compared with no protein expression using the native gene. Cloning of the optimized gene was confirmed by restriction site analysis and sequencing. A 106 kDa protein was expressed from the construct likely representing the His-tagged TF chaperone protein fused to the CCHF NC proteins.

During expression of NC of SPU415/85 and AP92, induction was carried out at 25°C to successfully express the protein. At 37°C, optical density of cells was stagnant post-induction. Using a lower temperature likely reduced the rate of growth of the cells resulting in slower synthesis of the protein, which does not overwhelm the export pathway of the bacteria or demonstrate cytotoxicity. The cells grew slower at a lower temperature, and took 4 hours to reach the actively growing phase (OD = 0.4-0.6). The protein purified from a 40 ml culture was sufficient for downstream applications, such as an ELISA. The eluate with higher protein concentration and higher purity would be a suitable candidate, and for each CCHFV strain the recombinant protein was designated, SPU NC and AP92 NC, respectively.

Application of the recombinant antigens to determine antigenic cross reactivity between South Africa and Greek isolates is discussed in the following chapter.

Chapter 4: Antigenic cross-reactivity of CCHFV nucleocapsids

4.1. Introduction

CCHFV isolates are found worldwide and the global diversity of the virus based on S segment sequence data was investigated in Chapter 2. Molecular techniques to clone a target gene and express a recombinant protein are now standard procedures in biotechnology laboratories. The functionality of recombinant antigens can be determined using immunoassays that use known antibody positive serum samples and negative controls. The NC genes of both isolates of CCHFV (SPU415/85 and AP92) were expressed using a bacterial expression system. Bacterial expression systems lack post-translational folding for the recombinant antigen and the biological functionality of the protein was determined using immunoassays (Henry & Sharp 2007).

ELISA are frequently used assays for diagnosis and epidemiological surveillance. ELISA can detect different classes of immunoglobulins, which can differentiate between recent (presence of IgM) and past (IgG antibodies) infection (Shepherd *et al* 1989b). ELISA are rapid, cost effective and readily automated with high sensitivity and specificity and are based on the ability of antibodies or antigens to bind to surfaces of ELISA plates. This is achieved through passive adsorption of the antigen to the plastic surface of the ELISA plate. This process occurs though hydrophobic interactions between the plate and non-polar protein residues (Crowther, 1995).

Although individual proteins may require specific conditions or pretreatment for optimal binding, the most common method for coating plates involves adding a solution of protein dissolved in a buffer such as PBS (pH 7.4). A blocking buffer is used after coating plates and should ideally bind to all potential sites of nonspecific interaction, eliminating background, without altering or obscuring the epitope for antibody binding. The blocking buffer is effective as it improves the sensitivity of an assay by reducing background signal and improving the signal-to-noise ratio. Antibodies interact with proteins forming antigen-antibody complexes. The antibody

is usually diluted in a blocking buffer to prevent non specific attachment of protein in the antiserum on the solid phase. The interactions are detected by a detector antibody labeled with an enzyme (such as horseradish peroxidase). Substrates are critical for the detection and visualization steps of an ELISA. The step involves the addition of suitable substrate solution for the particular enzyme conjugated to the antibodies. The objective is to allow development of color reaction through enzyme catalysis: addition of a substrate which, in the presence of the enzyme, will undergo a colour change that can be detected colorimetrically by measuring optical densities (Crowther, 1995). ELISA is a useful tool for diagnosis and seroepidemiological investigations.

An ELISA using a His-tagged recombinant NC and truncated NC prepared using a baculovirus system has been developed by Saijo *et al* (2002b) for possible use as a diagnostic tool. The NC is a suitable candidate for anti-CCHFV antibody detection and diagnostic ELISA utilizing NC as recombinant antigens have been developed (Saijo *et al.* 2002b; Saijo *et al.* 2005b; Samudzi *et al.* 2012). In addition, another study expressed CCHFV NC in a baculovirus system and developed an ELISA to detect IgG as well as IgM antibodies. The antigen was prepared using the Baghdad-12 strain and serum samples from Turkey, Tajikistan and Kosovo were used to validate the ELISA (Dowall *et al.* 2012).

The genetic- and amino acid diversity between the South African isolate SPU 415/85 and Greek isolate AP92 was determined in Chapter 2. In Chapter 3, recombinant antigens, designated SPU NC and AP92 NC, respectively, were prepared using a bacterial expression system. Collectively, the two proteins were referred to as CCHFV NC. In this chapter the biological functionality of the two recombinant CCHFV NC (isolates SPU 415/85 and AP92) expressed in a bacterial system from codon optimized genes, were assessed to determine antigenic cross reactivity. The proteins were present in both the soluble and insoluble phases, but purified from the soluble phase under native conditions.

In a previous study recombinant NC expressed from SPU415/85 NC gene was purified in our laboratory from the insoluble phase as the yield was too low in the soluble phase. In this study a higher yield was expressed in the soluble phase. Preparation from the insoluble phase requires refolding of the protein and chances of incorrect refolding are high and certain epitopes might be hidden. An in-house indirect ELISA was developed and used in this chapter to establish whether the two recombinant proteins would detect antibodies from the same serum samples from CCHF survivors. The application of a recombinant antigen prepared from a genetically diverse strain of CCHFV for detection of antibodies induced by South African strains of CCHFV would help to determine the usefulness and application of recombinant antigens in different laboratories worldwide.

4.2. Methods and Materials

4.2.1. Serum Samples

A panel of 14 convalescent serum samples from 14 patients that have survived CCHFV infection were used in the study and tested in duplicate. Informed consent was obtained from CCHFV infection survivors, under ethics approval 152/06. A negative panel consisting of six serum samples from six patients with no known history of CCHFV infection was used to determine the cut off value in the ELISA.

Table 23: Positive panel of sera from CCHF survivors in South Africa

Patient VBD number	Year blood collected	Time after onset of CCHF
VBD70/08	2008	10 months
VBD71/08	2008	8 years, 1 month
VBD3/09	2009	5 weeks
VBD6/09	2009	3 months
VBD29/10	2010	19 days
VBD30/10	2010	9 years, 8 months
VBD41/10	2010	8 years, 6 months
VBD42/10	2010	2 years, 6 months
VBD6/11	2011	11 months
VBD12/11	2011	1 year, 3 months
VBD38/11	2011	9 months
VBD39/11	2011	7 years, 3 months
VBD40/11	2011	8 months
VBD51/11	2011	10 years, 10 months

4.2.2. Enzyme-linked immunoassays

ELISA were performed in 96 well immunoassay plates, and optimal working dilutions of the reagents were determined by checkerboard titrations (1:500-1:2000 dilutions) (Crowther, 1995). Eluates one, two and three were compared for both SPU NC and AP92 NC. Throughout the assay, reagent volumes of 100 µl were used, the diluent for reagents was PBS containing 2% skimmed milk powder, incubations were performed for 1 hour at 37°C, wells were blocked after coating with 200 µl PBS containing 10% skimmed milk powder (Pick 'n Pay, Bloemfontein, South Africa) and plates were washed 5 times with PBS containing 0.1% Tween 20 (Merck, New Jersey, USA). Optimal dilutions of selected eluate of the recombinant SPU NC antigen, AP92NC antigen and pColdTF mock antigen were determined using serial fold dilutions from 1:500 to 1:4000.

4.2.3. IgG ELISA using recombinant antigen

Briefly, a 96 well microtiter Polysorb plate (Nunc Immunoplate, Denmark) was coated O/N at 4°C with recombinant SPU NC (Eluate 3, depicted in Figure 35) diluted at 1:1000 and mock pColdTF antigen diluted 1:4000 in PBS. The plates were washed, human serum samples; diluted 1:100 in diluent were added to each well in duplicate. The plates were incubated, washed and anti-human IgG horse radish peroxidase (HRPO) (Zymed laboratories, CA, USA) diluted 1:4000 was added to each well. After further incubation and washing, positive reactors were visualized using the substrate 2,2'-Azino-di-ethyl-benzothiazoline-sulfonic acid peroxidase substrate (ABTS) (Kirkegaard and Perry Laboratories, MD, USA). The plates were incubated at room temperature (22-25°C) in the dark for 30 min and the OD values were read at 405nm. The net OD value (OD on CCHF NC – OD on pColdTF) for each test and negative serum was determined. IgG ELISA using AP92 NC antigen to detect antibody in sera from South African CCHF survivors was performed as described, using HRPO at a dilution of 1:6000. The process of an indirect ELISA is summarized in Figure 38.

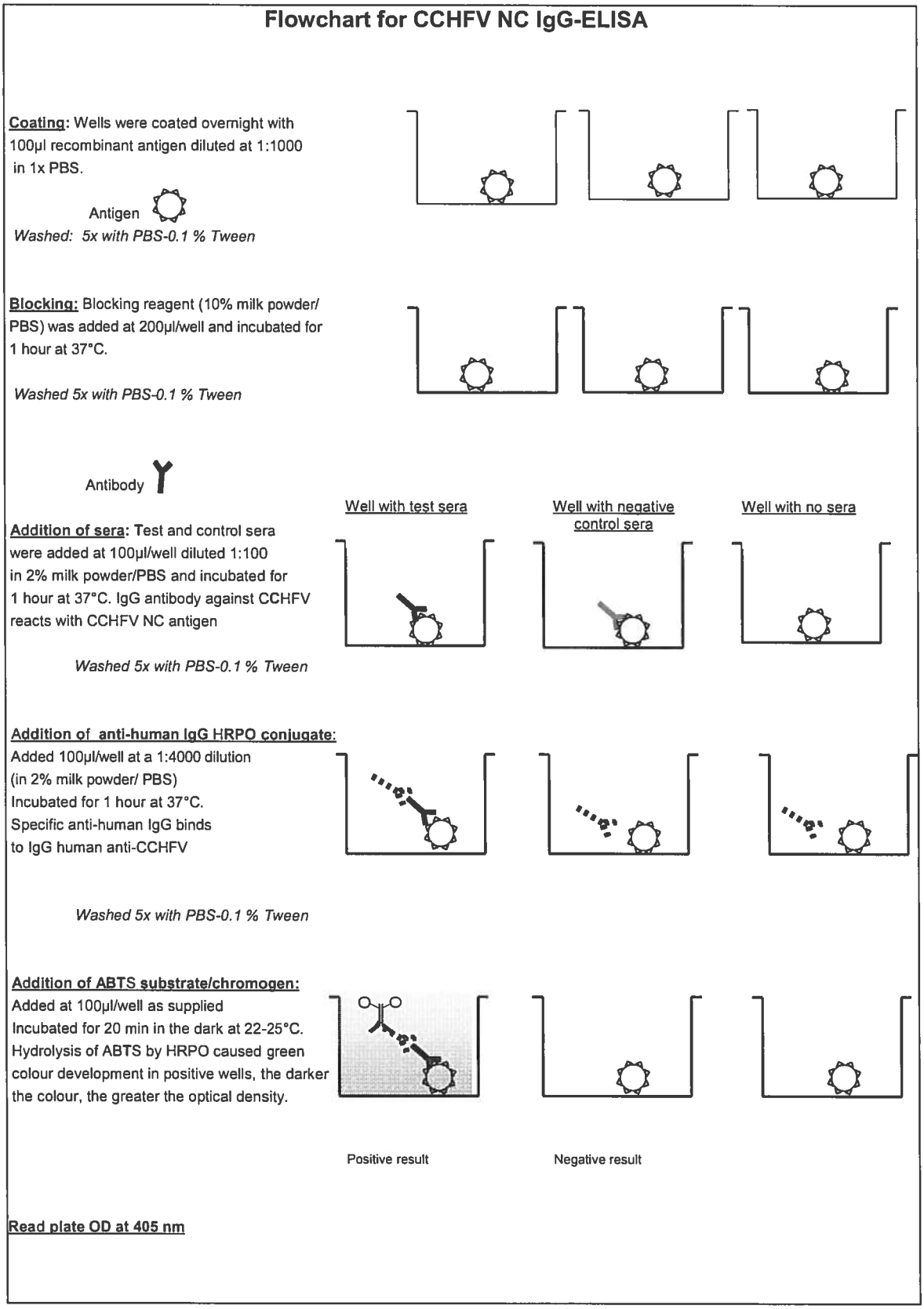


Figure 38: Indirect ELISA procedure: principle and protocol

4.2.3.1. Selection of cut-off values

Cut off values to separate positive results from negative results were determined using a panel of 6 negative control sera. The ELISA was performed on two separate occasions using each serum sample to obtain values for 12 replicates. The mean net OD and standard deviation (SD) obtained from 12 replicates was calculated. A cut off value was determined from the mean net OD of the negative panel plus 2 SD values.

4.2.4. Determination of antibody titers of CCHF convalescent sera using CCHFV NC

Ten-fold dilutions of sera were used to determine antibody titres. The titer of a sample is defined as the reciprocal of the highest dilution that would result in a positive reaction in the assay. Therefore endpoint titers of our patient sera were determined to get a quantitative result. An IgG ELISA was performed as previously described in section 4.2.3 with a few modifications. The patient sera were first diluted 1:10 in PBS in a non binding 96-well plate. Following the blocking and washing step, 180 µl of 2% diluent was added to each well of the ELISA test plate. A 20 µl aliquot of each serum sample diluted 1:10 was transferred from the diluting plate to the ELISA test plate to give a final dilution of 1:100. The sera were further diluted in 2-fold dilutions from 1:100 to 1:800 as illustrated in Figure 39. The plates were incubated and the positive reactors detected as described in section 4.2.3. OD values were read at 405 nm after optimal development of substrate for 20 min at room temperature in the dark and antibody titers were determined from net OD values. Additionally, IgG ELISA to calculate antibody titers of CCHF convalescent sera using AP92 NC antigen was performed as described.

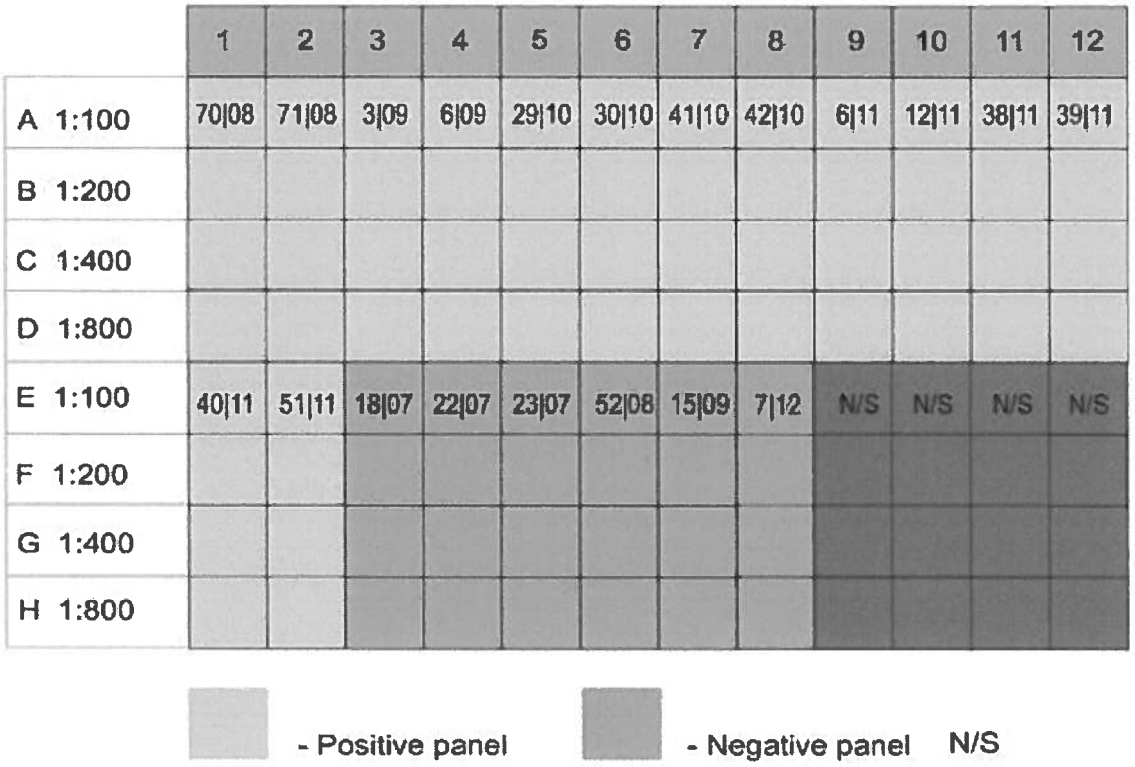


Figure 39: ELISA plate layout for antibody titre determination experiment using positive and negative panel test sera. (N/S no sera)

4.2.5. Stability of bacterially expressed CCHFV NC

A stability study was performed to test whether the recombinant SPU NC protein maintained biological activity when stored using five different methods, storage of precoated plates, freeze drying of recombinant antigens, storage of wet recombinant antigen in high protein solution with and without protease inhibitor. The recombinant SPU NC diluted at 1:1000 was used to coat a PolySorb microtiter plates O/N at 4°C. The following day the plates were washed and blocked in 10% skimmed milk as described in 4.2.3. The plates were washed and were stored in sealed bags at 4 °C before testing. A 200 µl aliquot of the recombinant antigen was lyophilized using Flexi-Dry™ MP micropressor control freeze-dryer (Vacutec, Johannesburg, South Africa). BSA was added to a 200 µl aliquot of recombinant protein (final concentration 0.2%) and stored at -20 °C with protease inhibitor cocktail (Roche, Mannheim, Germany) and a sample with no additives was stored at -20 °C.

A known high positive serum sample was selected and used as the high positive control (C++). Two serum samples were selected from healthy volunteers with no history of CCHFV infection and used as the negative controls (C-). The plates and lyophilized samples were stored at 4°C, the original sample and sample containing BSA were stored at -20 °C and tested at week 4 after preparation. Similar storage methods were tested with AP92 NC, as described for SPU NC.

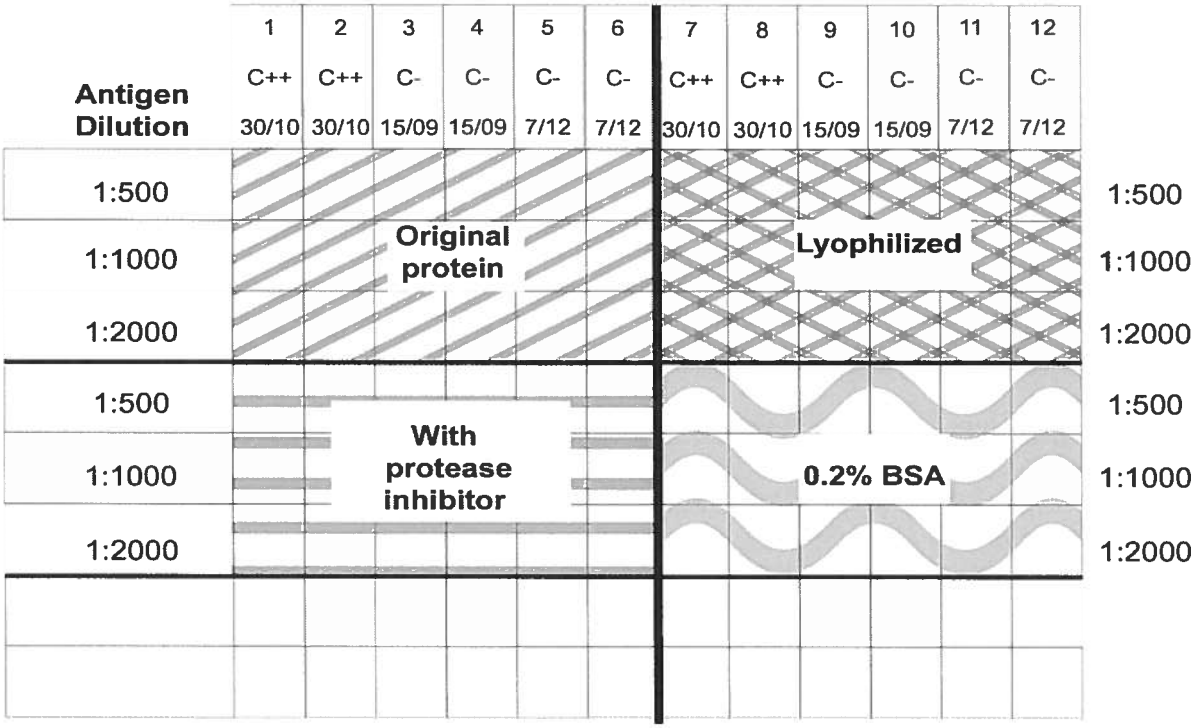


Figure 40: Plate layout for assessment of different storage methods for retention of biological activity of recombinant antigen. Specimen 30/10 was positive for antibody against CCHFV and specimens 15/09 and 7/12 were negative.

The different storage methods were assessed by calculating the ratio between mean OD value for C++ and mean OD value for C- for each serum sample reacted against antigens diluted 1;1000 . Each sample was run in duplicate, therefore, average OD values were used to determine ratios.

4.2.6. Statistical analysis/normalization of data.

OD values are absolute measurements that are influenced by variables such as temperature. To account for variability, results can be expressed as a function of the reactivity of control samples included in each run. Therefore absorbance values for ELISA in which convalescent sera were tested were expressed as percentage positive (PP) relative to a high positive control serum. The following statistical calculations were used in this chapter:

Positive to negative signal ratio:

Pos/Neg ratio = OD_{C++}/OD_{C-}

Net optical density (OD):

Net OD = OD in wells with virus antigen minus OD in wells with mock antigen (TF)

Percent positivity (PP):

PP = (Mean net OD of test sample / mean net OD of C++) x100

C++ = high positive control (VBD30/10)

C - = negative control

4.3. Results

4.3.1. Indirect ELISA for detection of IgG antibody using South African recombinant NC

During optimization of the indirect ELISA, protein fractions prepared as described in 3.3.6 depicted in Figure 35, were compared. Eluate number three, with a concentration of 402 µg/ml, demonstrated no non-specific bands on the SDS-PAGE gel, was thus selected as the protein fraction for use in ELISA for testing of panel of sera from CCHF survivors. Each serum samples was tested in duplicate; the net OD

and PP values of the negative panel used to determine the cut off are shown in Table 24. The cut off was determined using both net OD value and PP value determined from the mean net OD and mean net PP value of the negative panel plus 2 SD. All sera tested reacted positively. Serum samples were collected from patients at various intervals after disease as shown in Table 23 ranging from 5 weeks to 10 years and 10 months. These results demonstrate that IgG antibody against CCHFV can be detected using the South African recombinant NC antigen in sera collected early after onset of illness and up to 11 years.

Table 24: SPU 415/85 recombinant antigen tested against negative panel of sera.
Optical density (OD) ranges and positive percentage (PP) ranges are shown.

Laboratory Number	OD ranges			PP ranges		
	Max	Avg	Min	Max	Avg	Min
VBD*18/07*	0.033	0.0315	0.030	12.09%	12.04%	12.00%
VBD*22/07*	0.095	0.0925	0.090	32.97%	35.48%	38.00%
VBD*23/07*	0.110	0.1050	0.100	40.29%	40.15%	40.00%
VBD*52/08*	0.029	0.0245	0.020	10.62%	9.31%	8.00%
VBD*15/09*	0.034	0.0310	0.028	12.45%	11.83%	11.20%
VBD*7/12*	0.030	0.0200	0.020	10.99%	9.49%	8.00%

Table 25: Calculation of cut off value

Serum	[a]*n*	[b]Mean	[c]SD	Cut off	
				[d]OD	[e]PP
Negative panel	12	0.051583	0.035375	0.122332	46.8%

[a]*n*= number of replicates

[b]Mean = mean NET OD of replicates

[c]SD = standard deviation

[d]OD = optical density

[e]PP = percent positive

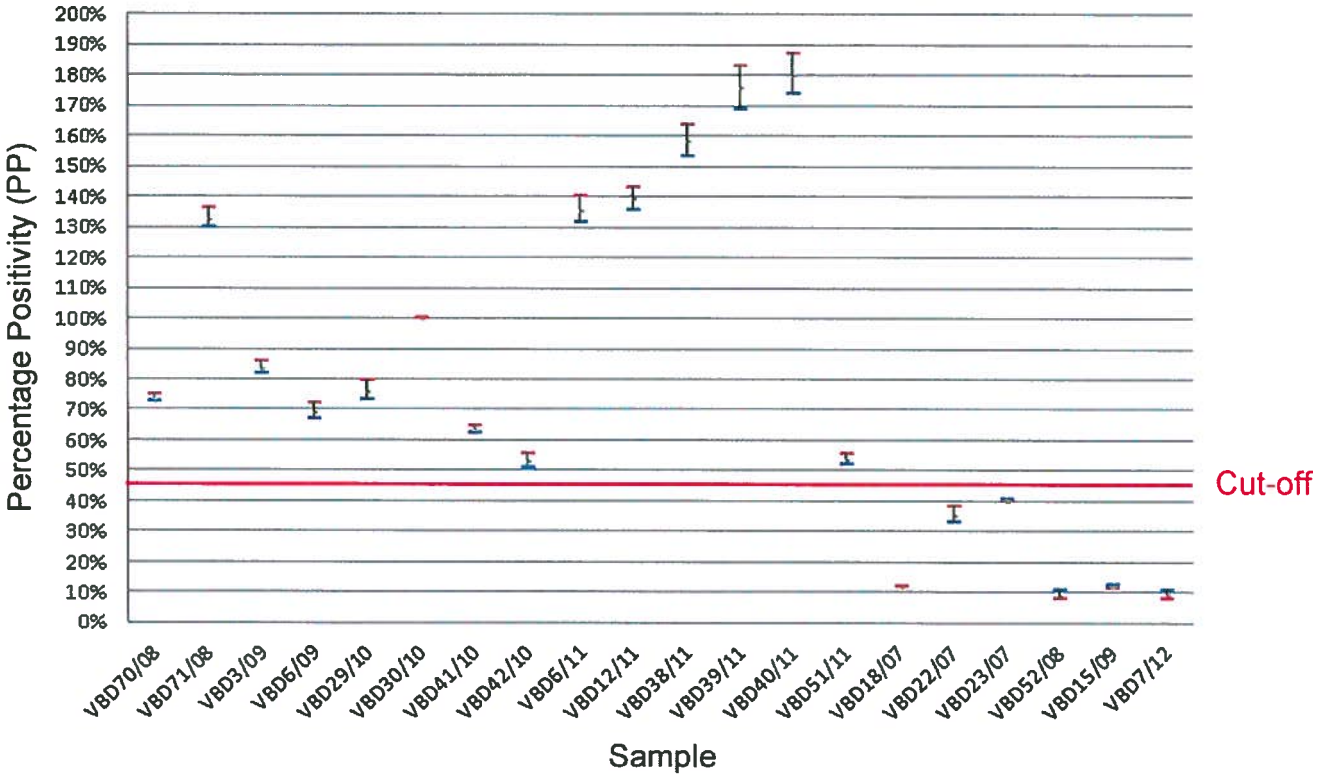


Figure 41: Detection of anti-CCHFV IgG by ELISA using recombinant NC from isolate SPU 415/85. The maximum and minimum percentage positivity (PP) values are shown for each serum sample; the cut off value was calculated as 46.8%. Negative reactors are depicted below the cut-off line.

PP values of positive reactors ranged from 50.92% to 186.8%. All negative reactors are shown below the cut-off, in Figure 41.

4.3.2. Indirect ELISA for detection of IgG antibody directed against AP92 NC

During optimization of the indirect ELISA, protein fractions prepared as described in 3.3.7, depicted in Figure 37 were compared. Eluate number three, had a concentration of 732 µg/ml, to test as an antigen in ELISA. The panel of sera was tested in duplicate; the OD and PP values of the negative panel used to determine the cut off are shown in Table 26. The cut off was determined using both net OD value and PP value determined from the mean net OD and mean net PP value of the negative panel plus 2 SD, as described in Table 27. PP values were obtained to

normalize the data, because OD values are affected by variables such as room temperature and the freeze-thaw cycle of reagents.

Table 26: AP92 recombinant antigen tested against panel of sera from CCHF survivors. Optical density (OD) ranges and positive percentage (PP) ranges are shown.

Laboratory Number	OD ranges			PP ranges		
	Max	Avg	Min	Max	Avg	Min
VBD*18/07*	0.5900	0.0445	0.0300	7.58%	3.85%	5.72%
VBD*22/07*	0.0830	0.0745	0.0660	10.66%	8.48%	9.57%
VBD*23/07*	0.2100	0.1935	0.1770	26.97%	22.74%	24.86%
VBD*52/08*	0.0900	0.0870	0.0840	11.56%	10.79%	11.18%
VBD*15/09*	0.1400	0.1370	0.1340	17.98%	17.21%	17.60%
VBD*7/12*	0.1060	0.1055	0.1050	13.62%	13.49%	13.55%

Table 27: Calculation of cut off value for AP92 antigen ELISA

Serum	[a]*n*	[b]Mean	[c]SD	Cut off	
				[d]OD	[e]PP
Negative panel	12	0.1070	0.051026	0.209052	26.85%

[a]*n*= number of replicates

[b]Mean = mean NET OD of replicates

[c]SD = standard deviation

[d]OD = optical density

[e]PP = percent positive

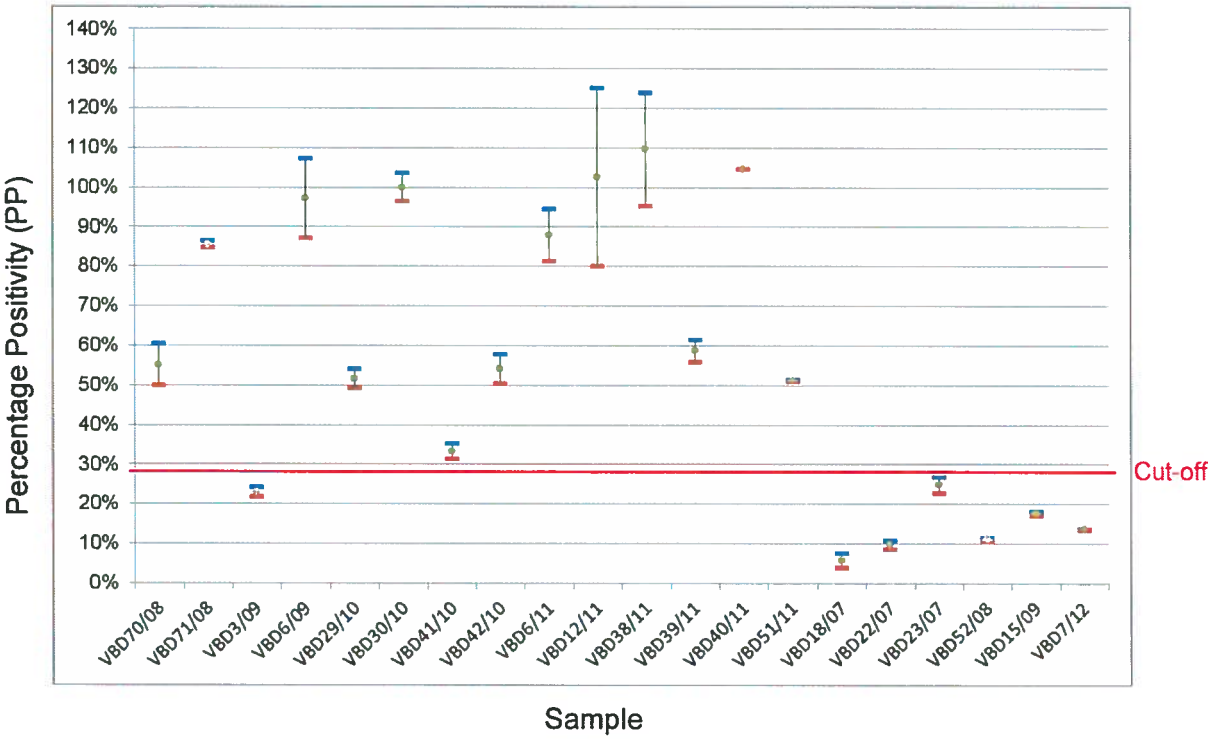


Figure 42: Detection of anti-CCHFV IgG by ELISA using recombinant NC from AP92 isolate. The maximum and minimum PP values are shown for each serum sample. Cut off PP value was calculated at 26.85%. All negative reactors, and one positive reactor against CCHFV NC, fall below the cut-off line.

In total 13/14 serum samples from the positive panel of South African CCHF survivors reacted positively against the AP92 NC antigen. VBD3/09, which was collected 5 weeks after onset of disease, was negative.

4.3.3. Determination of antibody titers for serum samples from CCHF survivors

Antibody titers were determined as described in section 4.2.4. The antibody titers were determined using serum samples diluted two fold from 1:100 and reacted against both antigens. Serum dilutions between 1:100 and 1:800 were performed. There were 5/14 sera that had higher titers using SPU NC antigen and a similar number (5/14) had higher titers using AP92 antigen. One serum was negative against AP92 antigen and three sera had identical titers. When titers varied there was less than a two fold difference as shown in Table 28.

Table 28: Antibody titers of CCHFV NC. Serum samples were diluted 1:100 to 1:800 to determine titer of each serum sample, expressed as reciprocal.

Sample Number	Patient VBD number	Antibody Titer SPU NC	Antibody Titer AP92 NC
1	VBD70/08	100	200
2	VBD71/08	200	400
3	VBD3/09	100	-
4	VBD6/09	100	400
5	VBD29/10	100	100
6	VBD30/10	200	400
7	VBD41/10	100	100
8	VBD42/10	100	100
9	VBD6/11	400	100
10	VBD12/11	400	200
11	VBD38/11	400	200
12	VBD39/11	400	100
13	VBD40/11	>800	>800
14	VBD51/11	100	200

4.3.4. Stability of bacterially expressed CCHFV NC

The protein purified from native conditions was found to be unstable over time with optimal results obtained using freshly purified protein or within 5 days of preparation. The protein, stored at - 20°C and tested 7 days later, had OD values for C++ similar to the OD value for C-, indicating that the protein was very unstable (data not shown).

Therefore, different methods for storing the recombinant protein were employed to determine which storage method would enhance protein stability. As depicted in Figure 40, a known positive control, C+, and two known negative controls, C-, were used to test the antigens. OD values expressed as ratios of mean C+:mean C- from samples tested on the pre-coated plate, protein stored at -20°C and coated 4-weeks post-purification, BSA-fortified aliquot, lyophilized aliquot and aliquot with added

protease inhibitor are shown in Figure 43. Every sample was run in duplicate; ratios were calculated using the average OD between two values. Antigen was diluted two fold from 1:500 to 1:2000 (see Appendix 1) the results for 1:1000 were used in the calculations.

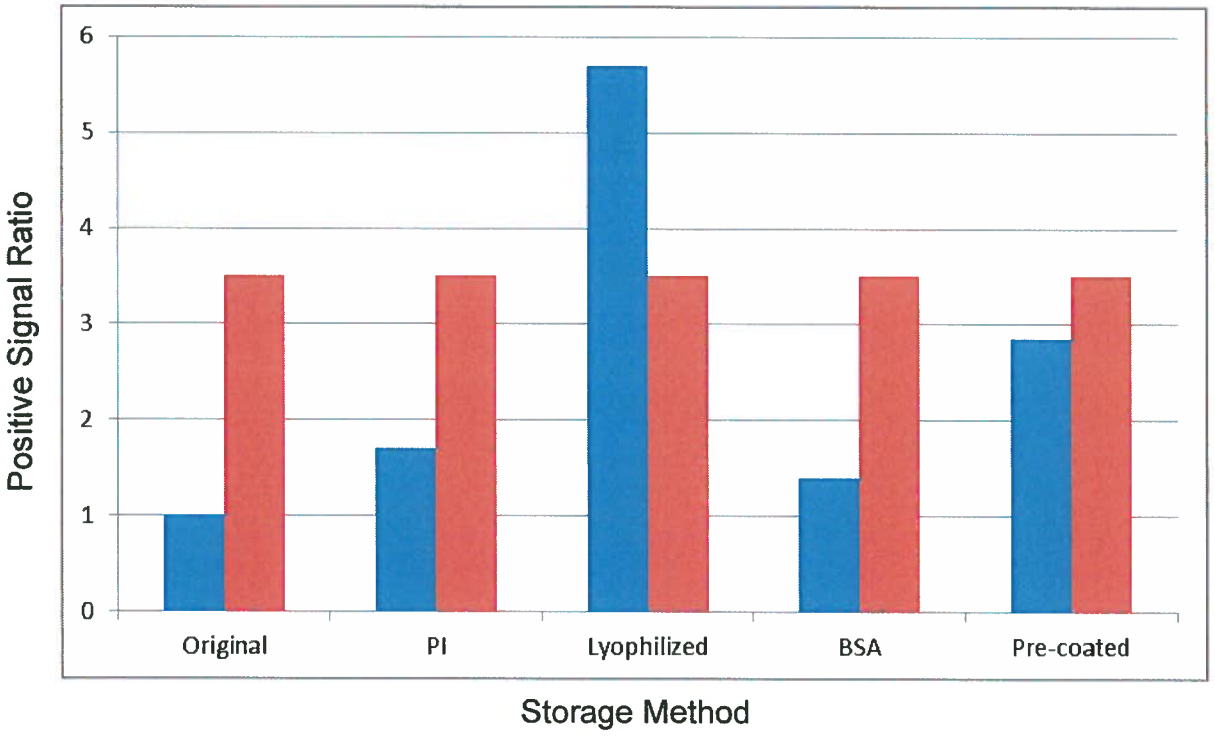


Figure 43: Assessment of different storage methods for retention of biological activity of recombinant antigen SPU NC. Original positive signal ratio a day after purification of SPU NC was 3.5, depicted in red bars, compared to positive signal ratio 4 weeks after storage of the protein using various methods. PI – protease inhibitor, BSA – Bovine serum albumin.

There was a significant decline in biological activity for the protein stored at -20°C, and samples stored with protease inhibitor cocktail and BSA. Precoated plates were more reliable, with positive signal ratios of 2.85. Lyophilized samples were the best storage method for maintaining biological activity for bacterially expressed His-tag purified antigens, however the increase in activity is likely a dilution factor and repeated samples would be useful to confirm the results.

Table 12: Mean distance between groups based on amino sequence data: The number of base differences per site from averaging over all sequence pairs between groups is shown.

	I	II	III	IV	V	VI
I West Africa						
II DRC	0.054					
III South Africa/West Africa	0.053	0.037				
IV Asia/Middle East	0.054	0.032	0.027			
V Europe/Turkey	0.053	0.038	0.039	0.054		
VI Greece	0.070	0.078	0.081	0.078	0.076	

In Table 12, the Asia/Middle East group had a higher amino acid homology with the South Africa/West Africa group, 97.3%, when compared with other groups. The Greek strain, AP92, had the highest amino acid heterogeneity when compared to the other groups. The number of amino acid differences per site from estimation of net average between groups of sequences is shown in Table 13.

Table 13: Mean Net distance between groups: The number of base differences per site from estimation of net average between groups of amino acid sequences is shown.

	I	II	III	IV	V	VI
I West Africa						
II DRC	0.050					
III South Africa/West Africa	0.044	0.030				
IV Asia/Middle East	0.043	0.023	0.013			
V Europe/Turkey	0.046	0.034	0.029	0.026		
VI Greece	0.066	0.077	0.075	0.070	0.073	

Amino acid diversity between groups using mean net distances ranged from 1.3% to 7.7%. The Greek isolate showed the least homology with other isolates ranging from 93.4% to 92.3%.

Additionally, pairwise comparisons using amino acid sequences of the 45 isolates were determined using MEGA4 software. Despite nucleotide differences in all the strains, there were strains that showed identical amino acid homology, namely three European strains YL04045, CYL/TI05099 and CLT/T105146 and two South African strains SPU103/87 and SPU 97/85. The interesting outcome for this study was that as shown previously, despite the relatively high percentage of nucleotide changes there were fewer changes in the predicted protein. The nucleotide changes were synonymous (codon translates into the same amino acid) which could explain the antigenic similarity of the nucleocapsid region observed between different isolates worldwide.

Amino acid sequences for the NC of SPU415/85 and AP92 were aligned for comparison, because the highest amino acid diversity was estimated between these two isolates (8.7%). The differences were highlighted in Figure 13. Majority of changes, such as Valine (V) to Isoleucine (I) and Asparagine (N) to Serine (S) are unlikely to affect the secondary structure of the protein, as the amino acids are from the same group with similar characteristics, nonpolar (V and I) and polar (N and S) amino acid groups, respectively. More significantly there were two amino acid changes from Phenylalanine (F) to Tyrosine (Y) which differ in polarity and are from a non-polar to polar amino acid.

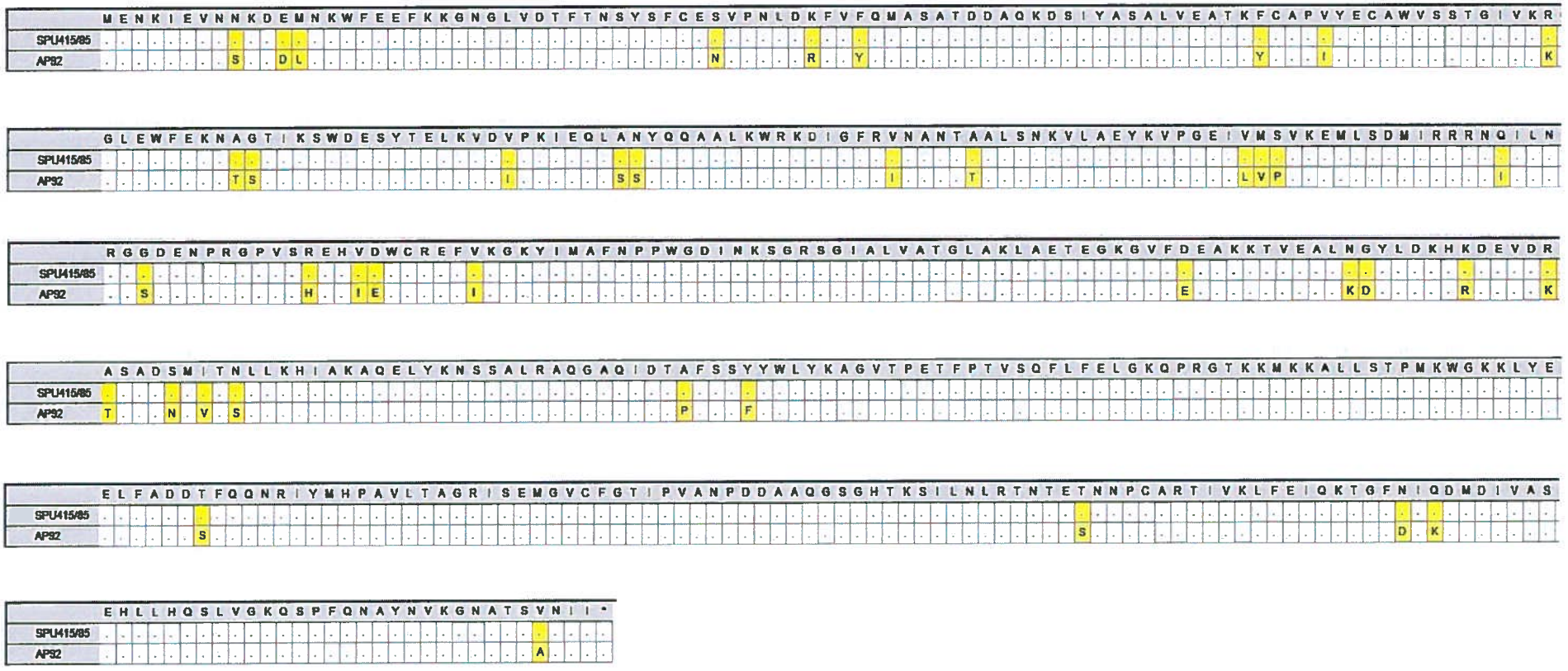


Figure 13: Amino acid residue comparison between NC of isolate SPU 415/85 and AP92, the variable sites are highlighted.

Summary

Sequence data for CCHFV isolates included in this study were selected to represent the broad geographic range where CCHFV is distributed and included all the complete S segment sequence data that was available. Isolates from Africa, Europe, Asia and the Middle East were selected. All available sequence data for South African isolates were included. The grouping of the different isolates correlated with that of Deyde *et al.,* (2006) and Anagnostou and Papa, (2009), and according to geographic distribution. Mild *et al.,* (2010), further subdivided CCHFV into genotypes using partial sequence data for a much larger cohort of 168 isolates. The outcome of this study and previous studies confirm geographical overlap of the different groups or genotypes likely due to migration of CCHFV and geographically distinct lineages. Possible reasons for distinct lineages are the different species of tick vector, genetic isolation or high mountain ranges isolating movement of infected ticks and animals restricting introduction to other regions.

The phylogenetic trees were constructed using a neighbour joining method (Saitou & Nei 1987; Hollingsworth & Ennos, 2004), which is commonly used in studies on genetic diversity. The method has been described by Saitou and Nei, (1987). The model chosen is the p distance model: the *P*-distance is the proportion (*p*) of nucleotide sites at which two sequences being compared are different. It is obtained by dividing the number of nucleotide differences by the total number of nucleotides compared.

AP92 strain from Greece differed greatly from other strains. The strain was originally isolated from *Rhiphicephalus bursa* tick (Papa *et al.,* 2002b; Midilli *et al.,* 2009).

Isolate 92/01 grouped with South African isolates as expected and was most closely related by nucleotide comparison to SPU4/81. According to comparison of amino acid sequence data SPU92/01 was closely related to SPU4/81 and Semunya strain. On the phylogenetic tree using amino acid sequence data this isolate forms a monophyletic branch with SPU415/85 (the two isolates used in this study for preparation of recombinant antigens). Amino acid p distance values were significantly smaller than nucleotide p distance values, that is the nucleotide

difference between SPU92/01 and SPU415/85 was 3.4% which was translated to an amino acid difference of only 0.6%. The similarity of all the South African isolates confirmed that SPU92/01 and SPU 415/85, the strains selected for preparing recombinant antigens, were representatives of the South Africa/West Africa group. The Greek isolate AP92 was the strain with the highest diversity from South Africa isolates. The ability of CCHFV to be dispersed within and between continents emphasizes the need to demonstrate that there is adequate serological cross reactivity between geographically distinct CCHFV isolates. Development of safe recombinant reagents that can be used sensitive, specific, standardized reagents for diagnosis and epidemiological surveys is essential for increasing diagnostic capacity worldwide and monitoring spread of the virus to previously non endemic regions in Europe.

Chapter 3: Expression of recombinant CCHFV nucleoprotein

3.1. Introduction

Recombinant antigens are useful tools with application in research and diagnostics. Nucleic acid sequences can be translated by the genetic code into the amino acid sequence expressing the desired protein. However; the exact sequence chosen can have profound effects on the expression of the encoded protein. Genes can be obtained by cloning from cDNA libraries, PCR amplification from the source organism or synthesized genes. Several prokaryotic and eukaryotic systems are available for heterologous expression, offering flexibility for a variety of protein types and applications. These systems can be further optimized by changing environmental conditions such as temperature or media components, by changing the intracellular environment by altering tRNA levels, and by changing the context and copy number of the gene itself (Baneyx, 1999). Recombinant protein expression is highly dependent on cellular machinery. In prokaryotes (bacterial expression system), the process of transcription and translation occur simultaneously. The translation of mRNA starts even before a mature mRNA transcript is fully synthesized (Makrides, 1996). In eukaryotes (mammalian expression system), the processes are spatially separated and occur sequentially with transcription taking place in the nucleus and translation, or protein synthesis, occurring in the cytoplasm (Baneyx, 1999).

Plasmid DNA delivery to mammalian cells for gene expression can be performed using various methods. Methods to improve efficient gene transfer to a broad range of cell types are important since there are cell types that resist efficient transfection. The main advantage of mammalian expression systems is the presence of post-translational modifications in the recombinant antigen produced and post-translational modifications influence protein structure. However, this method provides a lower yield, higher cost of production and requires more time than a bacterial expression system (Henry & Sharp, 2007).

Bacterial host systems are probably the most frequently used due to simplicity in technology and cost effectiveness however the disadvantages are the lack of post-translational modifications and the formation of inclusion bodies which render the recombinant antigen insoluble (Sahdev *et al.,* 2008). One advantage of this system is the ability to produce protein in large quantities. *E. coli* grow at a very fast rate in comparison to mammalian cells, giving the opportunity to purify, analyse and use the expressed protein in a much shorter time. In addition, transformation of *E. coli* cells with the foreign DNA is easy and requires minimal amounts of DNA. The choice of expression vector, with an inducible promoter is important during production of recombinant proteins. Commonly used promoters include the *lac* promoter, the *trp* promoter and their hybrid, the *tac* promoter that is regulated by the *lac* repressor and is induced by isopropyl-β-galactosidase (IPTG) (Henry & Sharp, 2007; Sahdev *et al.,* 2008).

Another variable to consider is codon bias. It is believed that high-level expression of proteins and highly hydrophobic proteins are more likely to lead to accumulation as inclusion bodies in *E. coli*. Genes in both prokaryotes and eukaryotes show a non random usage of synonymous codons (Henry & Sharp, 2007). Observations are that certain codons are most frequently used and highly expressed genes exhibit a greater degree of codon bias. Codon usage can influence the synthesis and yield of recombinant proteins (Makrides, 1996).

For recombinant antigens part of the functionality of protein to consider is if any epitopes that may induce an immune response are adequately exposed. An epitope can be defined as a region on an antigen that interacts with antibodies thereafter inducing an immune response (Bui *et al.,* 2007). Figure 14 highlights different factors that may influence the expression and ultimately the functionality of a recombinant protein.

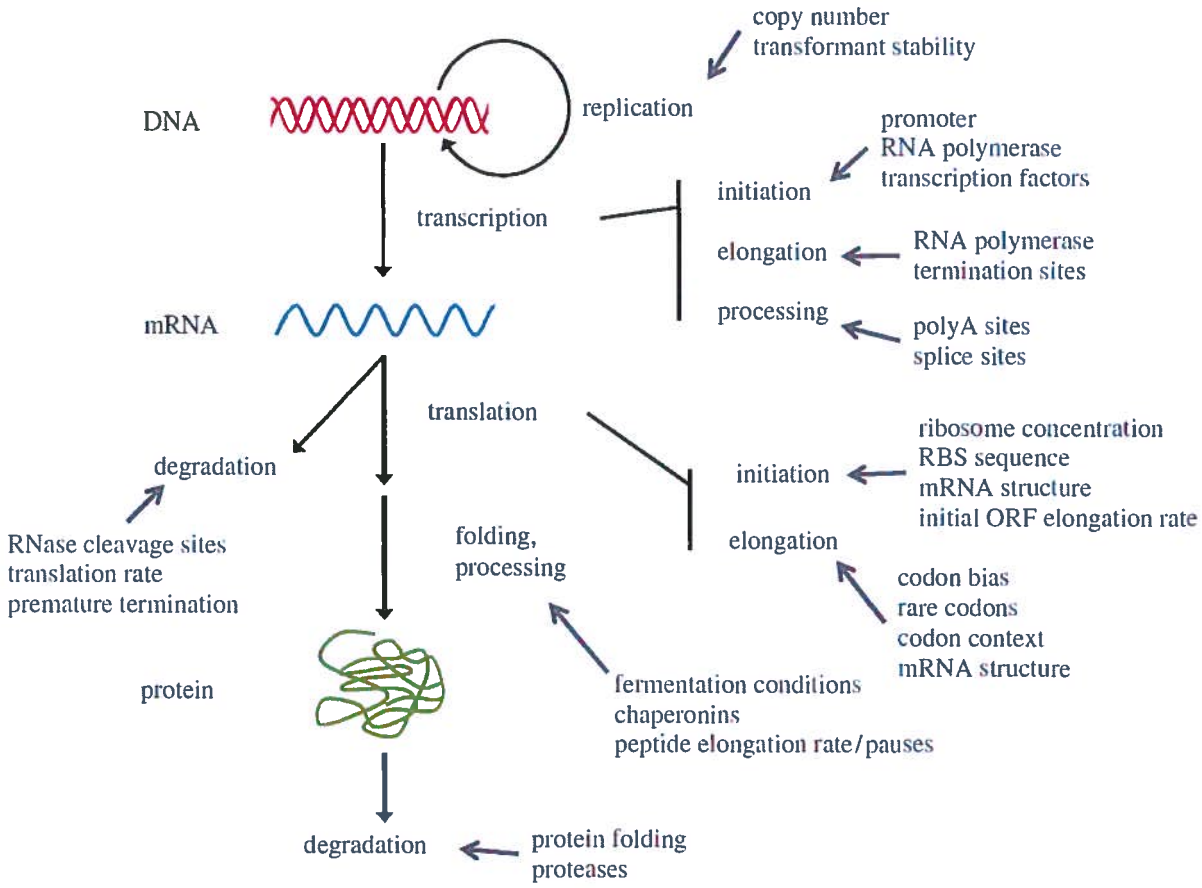


Figure 14: Factors influencing protein expression from mRNA level to final recombinant protein. (Welch *et al.*, (2009))

Chaperones that are co-expressed with the desired protein can enhance folding in bacterial expression systems (Kandror & Goldberg, 1997). To aid solubility of a protein, vectors with genes encoding cold shock proteins are used, which also facilitates slower growth if the recombinant protein is toxic to the cells (Phadtare & Inouye, 2008).

Recombinant antigens could play an important role in development of serological assays for antibody against CCHFV. Currently the reagents for CCHFV require culturing of the virus which is restricted to laboratories with BSL-4 facilities. To increase diagnostic and surveillance capacity worldwide standardized safe reagents are essential. Recombinant antigens could play an important role. However, it is not clear if there are antigenic differences between the NC of African and European strains of CCHFV that need to be considered when preparing recombinant antigens.

The aim of this study was to express biologically active NC of CCHFV from a South African isolate and from a Greek isolate, which is the most genetically diverse strain of CCHFV, and to determine the functionality of these antigens in development of ELISA for detection of IgG antibody against CCHFV. Mammalian and bacterial expression systems were investigated. Mammalian expression system used was described in Chapter 2 as the strain used was sequenced for phylogenetic analyses.

3.2. Methods and Materials

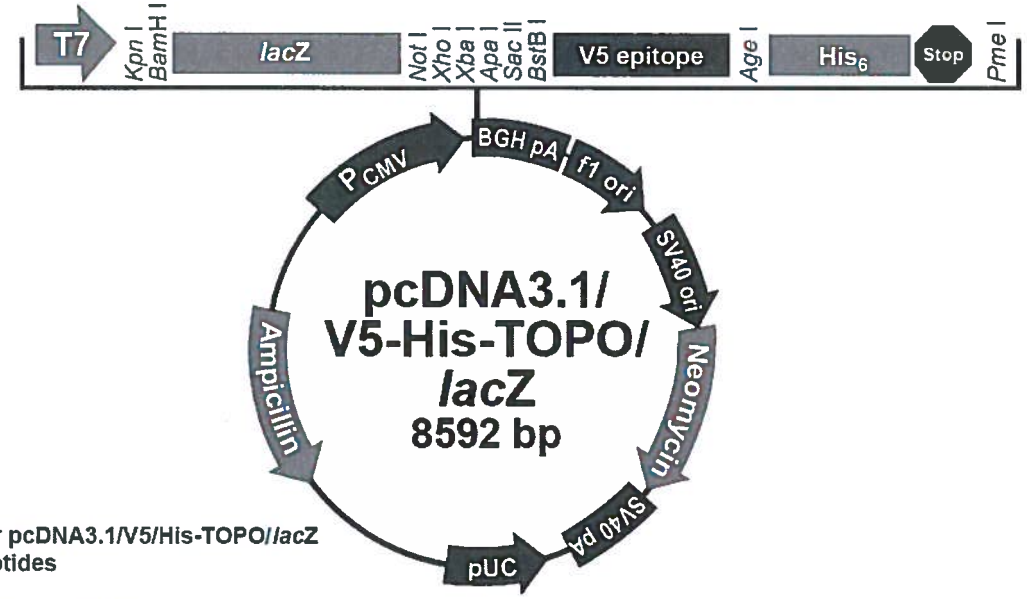
3.2.1. Mammalian expression of CCHFV NC

Correct protein folding is necessary to ensure that all epitopes on a recombinant protein are exposed. Mammalian expression has the advantage of post-translational modifications. The gene encoding the NC of a South African isolate designated SPU 92/01 was cloned into pcDNA3.1D/V5-His-TOPO as described in Materials and Methods 2.2.2. Two lipofectamine-based reagents, FuGene (Roche, Mannheim, Germany) and GeneJuice (Novagen, California, USA) were used to optimize transfection of cells using a control plasmid expressing β-galactosidase and the pcDNA3.1D/V5-His/CCHFNC construct. Transfection efficiency of the control plasmid was determined by assaying for β-galactosidase after transfection with pcDNA3.1/V5-His-TOPO/*LacZ* (Invitrogen, Maryland, USA) and, Western blot analysis. Transfection efficiency of the plasmid expressing the CCHF NC was determined using IFA assays and Western blot.

3.2.1.1. Preparation of plasmid constructs for use in transfection

Expression control plasmid pcDNA3.1/V5-His-TOPO/*LacZ* DNA was supplied in the pcDNA3.1D/V5-His-TOPO expressionist kit. OneShot TOP10 cells were transformed with the plasmid DNA using the heat shock method described in Materials and Methods 2.2.2.7. Colonies were selected and overnight cultures of the transformants were prepared. Expression control plasmid pcDNA3.1/V5-His-

TOPO/*LacZ* DNA was purified using the Pure Yield™ Plasmid Miniprep System with endotoxin free wash included as described in Materials and Methods 2.2.2.8. The expression control vector, pcDNA™3.1/V5-His-TOPO/*LacZ* (vector map shown in Figure 15) provided was used to optimize transfection conditions. The gene encoding β-galactosidase is expressed in mammalian cells under the CMV promoter. Transfection efficiency can be monitored using expression of β-galactosidase that can be easily assayed. DNA concentration of purified plasmid was determined as described in Methods and Materials 2.2.2.5.



Comments for pcDNA3.1/V5/His-TOPO/*lacZ*
8592 nucleotides

CMV promoter: bases 209-863
T7 promoter/priming site: bases 863-882
LacZ portion of fusion: bases 963-4019
V5 epitope: bases 4089-4130
Polyhistidine tag: bases 4140-4157
BGH Reverse priming site: bases 4180-4197
BGH polyadenylation signal: bases 4179-4393
f1 origin of replication: bases 4456-4869
SV40 promoter and origin: bases 4934-5259
Neomycin resistance gene: bases 5295-6089
SV40 polyadenylation signal: bases 6107-6346
pUC origin: bases 6778-7451 (C)
Ampicillin resistance gene: bases 7596-8456 (C)

C = complementary strand

Figure 15: Expression control vector pcDNA3.1/V5-His-TOPO/*LacZ* showing promoter sites, antibiotic resistance genes, restriction enzyme sites and primer sites (Invitrogen, Maryland, USA).

To identify positive transformants of the control plasmid double digestions of purified DNA using restriction sites present on the vector were set up as shown in Table 13. Reaction mixtures were incubated for 2 hours at 37°C. The restriction enzyme digestion reaction excise the *LacZ* gene is shown in Table 14. Digestion products were separated and visualized on a 1% agarose gel.

Table 14: Restriction enzyme analysis of *LacZ* gene by double digestion using *BamH1* and *Not1*.

Reaction component	**Reaction volume**
1 × restriction enzyme buffer D	2µl
BamH1 (10U/µl)	1µl
Not1 (10U/µl)	1µl
Bovine serum albumin (BSA)	1µl
70-80ng plasmid DNA	5µl
Nuclease free water	10µl
Total	**20µl**

3.2.1.2. Mammalian cell lines

Vero 76 cells (monkey kidney cells, American Type Culture Collection, ATCC number CRL-1587), HEK293 cells (human embryonic kidney ATCC number CRL-1573) and HeLa cells (human cervical cancer cells, ATCC number CCL-2) were maintained and used for transfection experiments. These cell lines were routinely maintained, separately, frozen in liquid nitrogen when not in use and passaged regularly before use.

Cells were cultured in T25 flasks and passaged weekly. Briefly, media was discarded from a flask with confluent cells and the cells were washed with Dulbecco's phosphate buffered saline (D-PBS) without calcium and magnesium (Life Technologies, New York, USA). Cells were dissociated using 0.5 ml aliquot of

trypsin or trypsin-versene (Lonza, New Jersey, USA) and incubated at 37°C for five to ten min. To inactivate the trypsin, a volume of 3 ml media was added to the cells and the cell suspension centrifuged at 300 x g for ten min. The supernatant was discarded and cells were resuspended in pre-warmed media. A subcultivation ratio of 1:3 was used. T25 flasks were incubated at 37°C. Vero 76 cells were cultured using Eagle's Minimum Essential Media with L-glutamine (EMEM, Life Technologies, New York, USA) containing 10% foetal bovine serum (FBS), (Delta Bioproducts, Johannesburg, South Africa), 0.5% streptomycin and penicillin (Life Technologies, New York, USA) and 0.5% non-essential amino acids (NEAA), (Invitrogen, Maryland, USA). Dulbecco's Minimum Essential Medium with L-glutamine (DMEM), (Life Technologies, New York, USA) containing 10% FBS, 0.5% streptomycin and penicillin and 0.5% NEAA was used as growth media for HEK293 and HeLa cells. HEK293 cells were trypsinized with trypsin-versene as HEK293 cells are more readily dissociated from the flasks.

Cells were passaged once a week for maintenance. Prior to transfection, cells were passaged every four days to ensure that rapidly growing cells were present. Unless otherwise stated all cell cultures were seeded at 2 x 10^6 cells per well on a 35 mm 6 well plate the day before transfection experiments. A monolayer of cells with 50-80% confluency was ideal for transfection purposes.

3.2.1.3. Transfection experiments

To determine transfection efficiency, mammalian cell line, HEK293 cells that were regularly passaged, were transfected with purified control plasmid pcDNA3.1/V5-His-TOPO/*LacZ* DNA that contained a *LacZ* gene that codes for β-galactosidase. For initial optimization, FuGene 6® Transfection Reagent (Roche, Mannheim, Germany), for the transient and stable transfection of animal cells, was used. The transfection mixes were prepared according to manufacturer's recommendations using the ratios of FuGene 6® and DNA as shown in Table 15.

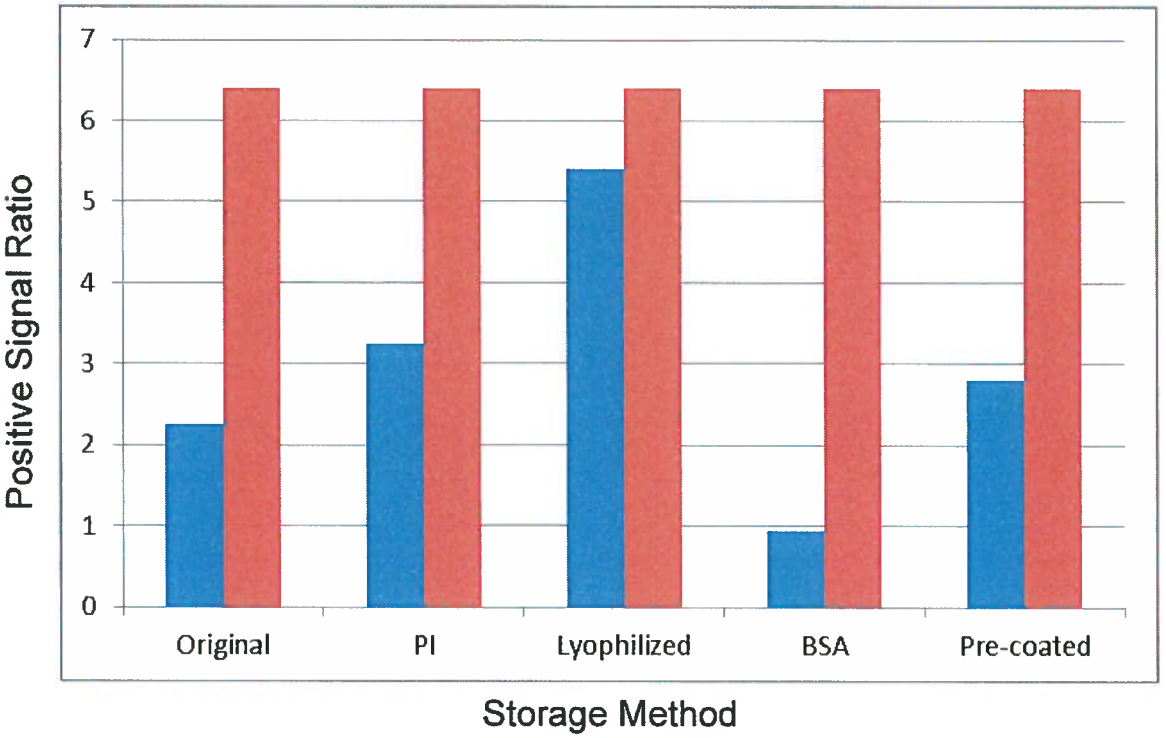


Figure 44: Assessment of different storage methods for retention of biological activity of recombinant antigen AP92 NC. Original positive signal ratio a day after purification of AP92 NC was 6.4, depicted in red bars, compared to positive signal ratio 4 weeks after storing the protein using various methods. At a ratio of 1:1 there is complete loss of biological activity. PI – protease inhibitor, BSA – Bovine serum albumin.

Stability testing of AP92 differed from that of SPU NC. The original sample of AP92 NC did not degrade as significantly as SPU NC. The antigen with added protease inhibitor retained some biological activity after four weeks as did the pre-coated plates. The protein stored in BSA was unstable. These results demonstrate that lyophilizing proteins helped to maintain biological activity.

Summary

ELISA methods are frequently used for seroprevalence- and seroepidemiological studies and also have application in diagnosis of CCHFV. ELISA are the most sensitive assays for demonstrating antibody response. The results suggest that the ELISA method described in this study, using SPU NC, may have application in routine diagnosis. Although for diagnostic purposes a complete validation will be required using significantly larger numbers of negative sera, the results from this study suggest that SPU415/85 NC is a safe reagent for use in ELISA for detection of IgG antibody during acute and convalescent stages of illness.

The SA NC recombinant antigen, detected antibody in all 14 serum samples from South African CCHF survivors. An IgG antibody response against CCHFV was detected in samples collected from day 15 after onset of illness up to 10 years and 10 months. In a previous study in our laboratory, recombinant NC antigen was purified from the insoluble phase by denaturing the antigen and refolding. This protein was very unstable after four weeks (Samudzi *et al.*, 2012). However in this study the antigen was prepared from the soluble phase using a native purification method. Lyophilization of the protein may help to stabilize the antigen over time. Similarly storage of precoated plates may also be useful for stabilizing the protein. Alternatively cell culture pellets can be stored and purified as required. Storage of pellets is not a practical solution for routine use although would be useful for large scale epidemiological studies.

One of the main objectives of the study was to investigate if a recombinant NC antigen expressed from a gene of highly diverse strain of CCHFV would be able to detect antibodies in sera from South African CCHFV convalescents. Almost 100% detection correlation was found with only 1/14 sera negative for IgG antibody using AP92 antigen. The reason why this sample was negative is unclear. It was collected five weeks after onset of illness. The AP92 antigen detected antibody in a sample collected from another patient 15 days after onset so it is unlikely due to early antibody response. Although the assays will require further validation with larger

numbers of samples than is available in our laboratory, the results suggest that there is sufficient serological cross reactivity between genetically and geographically distinct CCHFV isolates to allow recombinant NC to detect antibody in sera from geographically distinct locations. Hence a recombinant antigen expressed from a South African NC gene has possible application in other laboratories worldwide.

Chapter 5: Discussion

CCHFV is an acute, highly contagious, tick-borne viral zoonosis with mortality rates that range from 10% to 50%. Of all tick-borne viruses, CCHFV has the greatest geographic distribution, being endemic in more than 30 countries in Africa, central and southwestern Asia, the Middle East, and Eastern Europe (Sisman, 2012; Ergonul, 2012). Factors such as climate, changes in land and agricultural use, hunting and import and export of livestock, all contribute to the emergence of CCHFV. Emergence may also be caused by an amplification of virus strains by reservoir populations (Midilli *et al.* 2009). Ticks belonging to the genus *Hyalomma* are present in southern and south-eastern Europe which suggests that the virus could spread to currently non-endemic areas in Europe. The establishment and maintenance of an endemic focus requires an environment favoring contact between competent ticks and animal hosts with relatively high prevalence of infection (Maltezou *et al.* 2009). Outbreaks may occur in endemic areas, but new foci may emerge. Imported cases of CCHFV have also increased (Jauréguiberry *et al.* 2007; Atkinson *et al.* 2012). The epidemiology of CCHFV emphasizes the importance of developing standardized reagents for diagnosis and surveillance of the disease.

To develop standardized reagents that can be used worldwide, global diversity of CCHFV isolates, on nucleic acid, protein and antigenic level must be investigated. There is a high genetic variation amongst viruses belonging to the genus *Nairovirus,* which reflects the diversity of their predominant tick hosts (Honig *et al.* 2004). Similar subtypes are often found in distant geographical locations, even though the general pattern of genetic diversity observed in the S segment is related to the geographical distribution of the virus.

The aim of this study was to investigate the global diversity of CCHFV NC, on genetic and amino acid level and antibody cross-reactivity against the NC of a South African isolate and Greek isolate AP92. Phylogenetic analyses based on the S, M and L segment of CCHFV revealed seven defined groups of lineages (Yashina, *et al.* 2003b; Deyde *et al.* 2006; Morikawa *et al.* 2007; Ozkaya *et al.* 2010; Mild *et al.* 2010). Studies have also placed emphasis of the occurrence of recombination and

reassortment events resulting in worldwide genetic diversity of CCHFV (Hewson *et al.* 2004a; Deyde *et al.* 2006; Burt *et al.* 2009). The global diversity of worldwide isolates, based on the ORF of the S segment of 45 isolates in our study, revealed six distinct groups. Neighbour-joining trees displayed groupings that formed according to geographical location. Tree topology of the phylogenetic tree constructed using nucleotide sequence data was similar to the tree topology of phylogenetic tree constructed using amino acid data. Genetic- and amino acid diversity was determined within groups and between groups. The Asia/Middle East group displayed the highest divergence on both nucleotide- and amino acid level, which may be attributed to the large number of isolates from this group included in this study. The Greek isolate (Group VI Greece) showed the highest amino acid divergence with group III South Africa/ West Africa, and highest nucleotide divergence with I West Africa. Divergences between and within groups are significantly lower on amino acid level, due to synonymous changes on gene level. Eight strains that displayed genetic differences with their counterparts in the pairwise comparisons (see Appendix 1), displayed 100% amino acid homology (SPU103/87 and SPU97/85; Iran56 and Iran53; HY13 and C68031; YL04057 and CLT/TI05146; YL04057 and CYL/TI05035; CLT/TI05146 and CYL/TI05035; 9553/2001 and ROS/HUVLV100; KOSOVO and 2003108749).

All South African isolates with available sequence data for the complete ORF of the S segment was included in this study. Nigerian strain 10200, ArD39554 from Mauritania and two isolates from Sudan formed a group with South African strains of CCHFV. The S segment ORF of isolate SPU 92/01 was sequenced and included in the phylogenetic analyses. The isolate grouped with other South African isolates in Group III South Africa/West Africa. In the neighbor-joining phylogenetic tree constructed using amino acid sequence data, SPU92/01 formed a monophyletic branch with SPU415/85 with 99.4% amino acid homology. The most genetically distinct strain, AP92, shares the highest amino acid diversity with SPU415/85, at 8.7%.

The Greek strain, AP92, usually forms a distinct lineage on its own, group VI Greece, and is not associated with fatal disease. Phylogenetic analysis of the S segment

sequence of the CCHFV that caused the first human case with AP92 like CCHFV infection in Turkey has revealed that the strain was closely related to AP92 strain (Midilli *et al.* 2009). CCHFV emerged in Greece when the first non-imported fatal case of CCHF was recorded in June 2008. The causative strain was similar to strains from the Balkan peninsula, Russia and Turkey, but different from the AP92 strain (Papa *et al.* 2008). A seroepidemiological survey in Greece, demonstrated a seroprevalence of 4.2%, (Sidira *et al.* 2012), even though only one confirmed case had been reported. During the last decade CCHFV emerged and re-emerged in several Balkan countries, Turkey, southwestern regions of the Russian Federation, and the Ukraine, with high fatality rates (Maltezou *et al.* 2010). Re-emergence of the virus in Central Africa has also been described (Grard *et al.* 2011). To assess antigenic diversity of AP92 with a South African isolate, recombinant NC of AP92 was prepared for comparison in an ELISA using an antigen derived from a South African strain.

The initial aim was to express the NC protein of SPU92/01 in a mammalian expression system that generates proteins with folding and post-translational modifications identical to native proteins. The complete ORF of the gene encoding the NC was amplified from cDNA (transcribed from RNA) and cloned into pcDNA3.1D/V5-His/TOPO mammalian expression vector. The gene was sequenced to confirm that it had been cloned in frame and in the right orientation. The sequence was included in the phylogenetic analyses, as mentioned previously. Transfection of mammalian cell lines (HEK293, Vero E6 and HeLa cells) was optimized using FuGene 6 transfection reagent (Jacobsen *et al.* 2004). GeneJuice was subsequently tested for optimization of mammalian cell transfection of HeLa cells. A control expression vector, pcDNA3.1/V5-His/*LacZ*-TOPO construct was used for optimization experiments. Transfections using the control expression vector were monitored using a β-galactosidase assay and Western blot, while transfections with the pcDNA3.1D/V5-His/CCHFVNC construct were monitored using IFA and Western blot. Transfection efficiency did not exceed 20%, which resulted in an insufficient yield of protein, undetectable by Western blot technique.

CCHFV NC was previously expressed in HeLa cells using pKS336 vector for detection of IgG using IFA (Saijo *et al.* 2002a) using a baculovirus expression system (Qing *et al.* 2003). Another study utilized the Semliki Forest alphavirus replicon to express the NC in BHK cells (Garcia *et al.* 2006). The baculovirus expression system has a higher yield than mammalian expression systems, but may lack certain post-translational modifications. Bacterial expression of NC has demonstrated a high yield of biologically active protein (Samudzi *et al.* 2012), even though bacterially expressed proteins are known to have no post-translational modifications or proper folding.

A bacterial expression system was adopted, due to low yield of protein produced by the mammalian expression system. The chosen vector, pColdTF utilizes cold-shock technology. TF is a molecular chaperone that is induced at low temperatures, is essential for cell viability in cold conditions and assists with proper protein folding (Kandror & Goldberg, 1997). The vector also has a *cspA* promoter which plays a role in adaptation of cells to low temperature (*E.coli* is mesophilic) and overexpression of proteins in *E. coli* at low temperature improves their solubility and stability (Phadtare & Inouye, 2008). Prior to cloning, in order to express a virus protein in a bacterial system, codon bias had to be accounted for. Optimizing the codons for expression in *E. coli* improved CAI value generally associated with enhanced protein expression in *E.coli*. Codon optimization does not alter amino acid composition: the synthesized gene codes for the same amino acids as the native gene.

As mentioned in Chapter 3, many variables may affect protein expression: copy number and transformant stability, the type of promoter and RNA polymerase transcription factors, RNA polymerase termination sites, splicing sites, mRNA structure, codon bias, elongation rate of peptides and protein folding proteases. Expression of the NC was monitored using a time-course induction study. Bacterial cultures were grown at 25°C until logarithmic growth phase was reached. Using 37°C pre-induction caused growth to stagnate post-induction. Slower growth at 25°C produced bacterial cells expressing the gene of interest with TF, after induction with IPTG. Protein solubility was assessed using sarcosyl, which is a detergent that

releases membrane bound protein. Affinity chromatography was used to purify protein from bacterial cell cultures. Purity of the eluates were analysed using SDS-PAGE, certain eluates contain non-specific bands. Flowthrough also contains the expressed protein, suggesting that the His-tag might be hidden in a percentage of proteins expressed. The eluate with highest purity and sufficient concentration was tested in an ELISA for biological functionality.

A high yield of recombinant protein that was biologically active in ELISA was expressed for both SPU415/85 and AP92. ELISA is a sensitive assay that has been developed for CCHFV serosurveillance and also diagnostic applications (Donets *et al.* 1982; Shepherd *et al.* 1988; Burt *et al.* 1994; Qing *et al.* 2003; Saijo *et al.* 2005a; Dowall *et al.* 2012; Samudzi *et al.* 2012). An in-house ELISA was developed and a panel of positive samples from South African CCHFV survivors was tested against SPU NC. South African isolates are closely related genetic variants that share high amino acid homology. All 14 samples reacted against the SPU NC, while 13 samples reacted against AP92 NC. Cross-reactivity was assessed to determine the usefulness and application of recombinant antigens in different laboratories worldwide.

Recombinant antigens are safe to use and do not require BSL-4 facilities. There is a need for safe and reliable diagnostic reagents for CCHFV to increase diagnostic and surveillance capacity. Although the assays described in this study will require further validation with larger numbers of samples than is available in our laboratory, the results suggest that there is sufficient serological cross reactivity between genetically and geographically distinct CCHFV isolates to allow recombinant NC to detect antibody in sera from geographically distinct locations. Hence a recombinant antigen expressed from a South African NC gene has possible application in other laboratories worldwide.

References

Ahmed, A. A., McFalls, J. M., Hoffmann, C., Filone, C. M., Stewart, S. M., Paragas, J., Khodjaev, S., et al. (2005). Presence of broadly reactive and group-specific neutralizing epitopes on newly described isolates of Crimean-Congo hemorrhagic fever virus. *The Journal of General Virology*, *86*(12), 3327–3336.

Altamura, L. A., Bertolotti-Ciarlet, A., Teigler, J., Paragas, J., Schmaljohn, C. S., & Doms, R. W. (2007). Identification of a novel C-terminal cleavage of Crimean-Congo hemorrhagic fever virus PreGN that leads to generation of an NSM protein. *Journal of Virology*, *81*(12), 6632–6642.

Al-Tikriti, S. K., Al-Ani, F., Jurji, F. J., Tantawi, H., Al-Moslih, M., Al-Janabi, N., Mahmud, M. I., et al. (1981). Crimean-Congo haemorrhagic fever in Iraq. *Bulletin of the World Health Organization*, *59*(1), 85–90.

Anagnostou, V., & Papa, A. (2009). Evolution of Crimean-Congo Hemorrhagic Fever virus. *Infection, Genetics and Evolution*, *9*(5), 948–954.

Andersson, I., Simon, M., Lundkvist, A., Nilsson, M., Holmström, A., Elgh, F., & Mirazimi, A. (2004). Role of actin filaments in targeting of Crimean Congo hemorrhagic fever virus nucleocapsid protein to perinuclear regions of mammalian cells. *Journal of Medical Virology*, *72*(1), 83–93.

Aradaib, I. E., Erickson, B. R., Mustafa, M. E., Khristova, M. L., Saeed, N. S., Elageb, R. M., & Nichol, S. T. (2010). Nosocomial Outbreak of Crimean-Congo Hemorrhagic Fever, Sudan. *Emerging Infectious Diseases*, *16*(5), 837–839.

Ascioglu, S., Leblebicioglu, H., Vahaboglu, H., & Chan, K. A. (2011). Ribavirin for patients with Crimean-Congo haemorrhagic fever: a systematic review and meta-analysis. *The Journal of Antimicrobial Chemotherapy*, *66*(6), 1215–1222.

Atkinson, B., Latham, J., Chamberlain, J., Logue, C., O'Donoghue, L., Osborne, J., Carson, G., et al. (2012). Sequencing and phylogenetic characterisation of a

fatal Crimean - Congo haemorrhagic fever case imported into the United Kingdom, October 2012. *Euro Surveillance*, *17*(48), 6–9.

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*, *10*, 411–421.

Benson, D., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2011). GenBank. *Nucleic Acids Research*, *39*, 32–37.

Bishop, D. H., Calisher, J., Casals, J., Chumakov, M. P., Gaidamovich, S. Y., Hannoun, C., Lvov, D. K., et al. (1980). Bunyaviridae. *Intervirology*, *14*, 125–143.

Blackburn, N. K., Besselaar, T. G., Shepherd, A. J., & Swanepoel, R. (1987). Preparation and use of monoclonal antibodies for identifying Crimean-Congo hemorrhagic fever virus. *American Journal of Tropical Medicine and Hygiene*, *37*, 392–397.

Bodur, H., Akinci, E., Ongürü, P., Carhan, A., Uyar, Y., Tanrici, A., Cataloluk, O., et al. (2010). Detection of Crimean-Congo hemorrhagic fever virus genome in saliva and urine. *International Journal of Infectious Diseases*, *14*(3), e247–e249.

Bui, H. H., Sidney, J., Li, W., Fusseder, N., & Alessandro, S. (2007). Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*, *8*, 361–366.

Burt, F. J., Leman, P. A., Abbott, J. C., & Swanepoel, R. (1994). Serodiagnosis of Crimean-Congo haemorrhagic fever. *Epidemiology and Infection*, *113*(3), 551–562.

Burt, F. J., Leman, P. A., Smith, J. F., & Swanepoel, R. (1998). The use of a reverse transcription-polymerase chain reaction for the detection of viral nucleic acid in the diagnosis of Crimean-Congo haemorrhagic fever. *Journal of Virological Methods*, *70*(2), 129–137.

Burt, F. J., & Swanepoel, R. (2005). Molecular epidemiology of African and Asian Crimean-Congo haemorrhagic fever isolates. *Epidemiology and Infection*, *133*(4), 659–666.

Burt, F. J., Paweska, J., & Swanepoel, R. (2007). Crimean-Congo heamorrhagic fever in South Africa. In C. Whitehouse & O. Ergonul (Eds.), *Crimean-Congo Haemorrhagic Fever*. Dordrecht: Springer Press.

Burt, F. J., Paweska, J. T., Ashkettle, B., & Swanepoel, R. (2009). Genetic relationship in southern African Crimean-Congo haemorrhagic fever virus isolates: evidence for occurrence of reassortment. *Epidemiology and Infection*, *137*(9), 1302–1308.

Casals, J. (1969). Antigenic Similarity between the Virus Causing Crimean Hemorrhagic Fever and Congo Virus (33847). *Proceedings of the Society of Experimental Biology and Medicine (New York)*, *131*, 233–236.

Cevik, M. A., Erbay, A., Bodur, H., Eren, S. S., Akinci, E., Sener, K., Ongürü, P., et al. (2007). Viral load as a predictor of outcome in Crimean-Congo hemorrhagic fever. *Clinical Infectious Diseases*, *45*(7), 96–100.

Cevik, M. A., Elaldi, N., Akinci, E., Ongürü, P., Erbay, A., Buzgan, T., Uzun, R., et al. (2008). A preliminary study to evaluate the effect of intravenous ribavirin treatment on survival rates in Crimean-Congo hemorrhagic fever. *The Journal of Infection*, *57*(4), 350–359.

Chamberlain, J., Cook, N., Lloyd, G., Mioulet, V., Tolley, H., & Hewson, R. (2005). Co-evolutionary patterns of variation in small and large RNA segments of Crimean-Congo hemorrhagic fever virus. *The Journal of General Virology*, *86*(12), 3337–3341.

Chen, C., and Okayama, H. (1987). High-Efficiency Transformation of Mammalian Cells by Plasmid DNA. *Molecular Cell Biology. 7*, 2745-2752.

Chinikar, S., Mirahmadi, R., Moradi, M., Ghiasi, S. M., & Khakifirouz, S. (2012). Crimean-Congo Hemorrhagic Fever (CCHF). In J. Lorenzo-Morales (Ed.), *Zoonosis*. InTech.

Chinikar, S., Persson, S. M., Johansson, M., Bladh, L., Goya, M., Houshmand, B., Mirazimi, A., et al. (2004). Genetic analysis of Crimean-congo hemorrhagic fever virus in Iran. *Journal of Medical Virology*, *73*(3), 404–411.

Clerx, J. P., Casals, J., & Bishop, D. H. (1981). Structural characteristics of nairoviruses (genus Nairovirus, Bunyaviridae). *The Journal of General Virology*, *55*(1), 165–78.

Connolly-Andersen, A. M., Moll, G., Andersson, C., Akerström, S., Karlberg, H., Douagi, I., & Mirazimi, A. (2011). Crimean-Congo hemorrhagic fever virus activates endothelial cells. *Journal of Virology*, *85*(15), 7766–7774.

Crowther, J. R. (1995). ELISA theory and practice. *Methods In Molecular Biology*. Totowa, NJ: Humana Press.

Deyde, V. M., Khristova, M. L., Rollin, P. E., Ksiazek, T. G., & Nichol, S. T. (2006). Crimean-Congo hemorrhagic fever virus genomics and global diversity. *Journal of Virology*, *80*(17), 8834–8842.

Donets, M. A., Chumakov, M., Korolev, M. B., & Rubin, S. G. (1977). Physicochemical Characteristics, Morphology, and Morphogenesis of Virions of the Causative Agent of Crimean Hemorrhagic Fever. *Intervirology*, *8*(5), 294–308.

Donets, M., & Rezapkin, G. (1982). Immunosorbent assays for diagnosis of Crimean-Congo hemorrhagic fever (CCHF). *The American Journal of Tropical Medicine and Hygiene*, *31*(1), 156–162.

Dowall, S. D., Richards, K. S., Graham, V. A., Chamberlain, J., & Hewson, R. (2012). Development of an indirect ELISA method for the parallel measurement of IgG and IgM antibodies against Crimean-Congo haemorrhagic fever (CCHF) virus using recombinant nucleoprotein. *Journal of Virological Methods*, *179*, 335–341.

Duh, D., Saksida, A., Petrovec, M., Dedushaj, I., & Avsic-Zupanc, T. (2006). Novel one-step real-time RT-PCR assay for rapid and specific diagnosis of Crimean-Congo hemorrhagic fever encountered in the Balkans. *Journal of Virological Methods*, *133*(2), 175–179.

Dunster, L., Dunster, M., Ofula, V., Beti, D., Kazooba-Voskamp, F., Burt, F. J., Swanepoel, R., et al. (2002). First Documentation of Human Crimean-Congo Hemorrhagic Fever, Kenya. *Emerging Infectious Diseases*, *8*(9), 1005–1006.

El-Azazy, O. M., & Scrimgeour, E. M. (1997). Crimean-Congo haemorrhagic fever virus infection in the western province of Saudi Arabia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *91*(3), 275–278.

Eley, S. M., Delic, J. I., Henstridge, R. M., Bruce, L. G., Humphreys, C. R., & Moore, N. F. (1989). Bunyaviridae. Serological relationships. *Microbiologica (Pavia)*, *12*(4), 351–367.

Elliott, R. M. (1990). Review article Molecular biology of the Bunyaviridae. *Journal of General Virology*, *71*, 501–522.

Ergonul, O. (2006). Crimean-Congo haemorrhagic fever. *The Lancet Infectious Diseases*, *6*(4), 203–214.

Ergonul, O., Celikbas, A., Baykam, N., Eren, S., & Dokuzoguz, B. (2006). Analysis of risk-factors among patients with Crimean-Congo haemorrhagic fever virus infection: severity criteria revisited. *Clinical Microbiology and Infection*, *12*(6), 551–554.

Ergonul, O. (2008). Treatment of Crimean-Congo hemorrhagic fever. *Antiviral Research*, *78*(1), 125–31.

Ergonul, O., Celikbas, A., Yildirim, U., Zenciroglu, A., Erdogan, D., Ziraman, I., Saracoglu, F., et al. (2010). Pregnancy and Crimean-Congo haemorrhagic fever. *Clinical Microbiology and Infection*, *16*(6), 647–650.

Ergonul, O. (2012). Crimean-Congo hemorrhagic fever virus: new outbreaks, new discoveries. *Current Opinion in Virology*, *2*(2), 215–20.

Flick, R., & Whitehouse, C. A. (2005). Crimean-Congo Hemorrhagic Fever Virus. *Current Molecular Medicine*, *5*(8), 753–760.

Garcia, S., Chinikar, S., Coudrier, D., Billecocq, A., Hooshmand, B., Crance, J. M., Garin, D., et al. (2006). Evaluation of a Crimean-Congo hemorrhagic fever virus recombinant antigen expressed by Semliki Forest suicide virus for IgM and IgG antibody detection in human and animal sera collected in Iran. *Journal of Clinical Virology*, *35*(2), 154–159.

Gear, J. H. S. (1989). Clinical aspects of African Viral Hemorrhagic Fevers. *Reviews of Infectious Diseases*, *11*, S777–S789.

Gear, J. H. S., Thomson, P. D., Hopp, M., Andronikou, S., Cohn, R. J., Ledger, J., & Berkowitz, F. E. (1982). Congo-Crimean haemorrhagic fever in South Africa. Report of a fatal case in the Transvaal. *South African Medical Journal*, *62*(16), 576–580.

Gonzalez, J. P., Camicas, J. L., Cornet, J. P., Faye, O., & Wilson, M. L. (1992). Sexual and transovarian transmission of Crimean-Congo haemorrhagic fever virus in Hyalomma truncatum ticks. *Research in Virology*, *143*(1), 23–28.

Grard, G., Drexler, J. F., Fair, J., Muyembe, J.-J., Wolfe, N. D., Drosten, C., & Leroy, E. M. (2011). Re-emergence of Crimean-Congo hemorrhagic fever virus in Central Africa. (A. D. LaBeaud, Ed.)*PLoS neglected tropical diseases*, *5*(10).

Gunes, T., Engin, A., Poyraz, O., Elaldi, N., Kaya, S., Dokmetas, I., Bakir, M., et al. (2009). Crimean-Congo hemorrhagic fever virus in high-risk population, Turkey. *Emerging Infectious Diseases*, *15*(3), 461–464.

Guo, Y., Wang, W., Ji, W., Deng, M., Sun, Y., Zhou, H., Yang, C., et al. (2012). Crimean – Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(13), 5046–5051.

Haferkamp, S., Fernando, L., Schwarz, T. F., Feldmann, H., & Flick, R. (2005). Intracellular localization of Crimean-Congo Hemorrhagic Fever (CCHF) virus glycoproteins. *Virology Journal*, *2*(1), 42.

Haller, O., Mirazimi, A., & Andersson, I. (2006). Type I Interferon Inhibits Crimean-Congo Hemorrhagic Fever Virus in Human Target Cells. *Journal of Medical Virology*, *78*, 216–222.

Henry, I., & Sharp, P. M. (2007). Predicting Gene Expression Level from Codon Usage Bias. *Molecular Biology and Evolution*, *24*, 10–12.

Hewson, R, Chamberlain, J., Mioulet, V., Lloyd, G., Jamil, B., Hasan, R., Gmyl, A., et al. (2004a). Crimean-Congo haemorrhagic fever virus: sequence analysis of the small RNA segments from a collection of viruses world wide. *Virus Research*, *102*(2), 185–189.

Hewson, R., Gmyl, A., Gmyl, L., Smirnova, S. E., Karganova, G., Jamil, B., Hasan, R., et al. (2004b). Evidence of segment reassortment in Crimean-Congo haemorrhagic fever virus. *The Journal of General Virology*, *85*(10), 3059–3070.

Hewson, R. (2007). Molecular Epidemiology, Genomics and Phylogeny of Crimean-Congo hemorrhagic fever virus. In O. Ergonul & C. A. Whitehouse (Eds.), *Crimean-Congo Hemorrhagic fever* (pp. 45–55). Springer Press.

Hollingsworth, P. M., & Ennos, R. A. (2004). Neighbour joining trees , dominant markers and population genetic structure. *Heredity*, *92*, 490–498.

Honig, J. E., Osborne, J. C., & Nichol, S. T. (2004). The high genetic variation of viruses of the genus Nairovirus reflects the diversity of their predominant tick hosts. *Virology*, *318*(1), 10–16.

Hoogstraal, H. (1979). The epidemiology of tick-borne Crimean-Congo hemorrhagic fever in Asia, Europe and Africa. *Journal of Medical Entomology*, *15*, 307–417.

Ibrahim, S. M., Aitichou, M., Hardick, J., Blow, J., O'Guinn, M. L., & Schmaljohn, C. (2011). Detection of Crimean-Congo hemorrhagic fever, Hanta, and sandfly

fever viruses by real-time RT-PCR. In J. R. Stephenson & A. Warnes (Eds.), *Methods In Molecular Biology*, *665*, (pp. 357–368). Clifton NJ: Humana Press.

Jacobsen, L. B., Calvin, S. A., Colvin, K. E., & Wright, M. (2004). FuGENE 6 Transfection Reagent: the gentle power. *Methods*, *33*, 104–112.

Jamil, B., Hasan, R. S., Sarwari, A. R., Burton, J., Hewson, R., & Clegg, C. (2005). Crimean-Congo hemorrhagic fever: experience at a tertiary care hospital in Karachi, Pakistan. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *99*(8), 577–584.

Jansen van Vuren, P., Weyer, J., & Kemp, A. (2012). Viral Haemorrhagic Fever Outbreaks, South Africa 2011. *Communicable Diseases Surveillance Bulletin*, *10*(1), 1–5.

Jauréguiberry, S., Tattevin, P., Tarantola, A., Legay, F., Tall, A., Nabeth, P., Zeller, H., et al. (2007). Imported Crimean-Congo hemorrhagic fever cases in Istanbul. *BMC Infectious Diseases*, *7*(9), 4905–4907.

Joubert, J. R., King, J. B., Rossouw, D. J., & Cooper, R. (1985). A nosocomial outbreak of Crimean-Congo haemorrhagic fever at Tygerberg Hospital. Part III. Clinical pathology and pathogenesis. *South African Medical Journal*, *68*, 722–728.

Kandror, O., & Goldberg, A. L. (1997). Trigger factor is induced upon cold shock and enhances viability of Escherichia coli at low temperatures. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(10), 4978–81.

Karti, S. S., Odabasi, Z., Korten, V., Yilmaz, M., Sonmez, M., Caylan, R., Akdogan, E., et al. (2004). Crimean-Congo hemorrhagic fever in Turkey. *Emerging Infectious Diseases*, *10*(8), 1379–1384.

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066.

Keshtkar-Jahromi, M., Kuhn, J. H., Christova, I., Bradfute, S. B., Jahrling, P. B., & Bavari, S. (2011). Crimean-Congo hemorrhagic fever: current and future prospects of vaccines and therapies. *Antiviral Research*, *90*(2), 85–92.

Khan, A. S., Maupin, G. O., Rollin, P. E., Noor, A. M., Shurie, H. H., Shalabi, A. G., Wasef, S., et al. (1997). An outbreak of Crimean-Congo hemorrhagic fever in the United Arab Emirates, 1994-1995. *The American Journal of Tropical Medicine and Hygiene*, *57*(5), 519–525.

Kinsella, E., Martin, E., Grolla, S. G., Czub, M., Feldmann, H., & Flick, R. (2004). Sequence determination of the Crimean-Congo hemorrhagic fever virus L segment. *Virology*, *321*, 23–28.

Knust, B., Medetov, Z. B., Kyraubayev, K. B., Bumburidi, Y., Erickson, B. R., Macneil, A., Nichol, S. T., et al. (2012). Crimean-Congo Hemorrhagic Fever, Kazakhstan, 2009–2010. *Emerging Infectious Diseases*, *18*(4), 643–645.

Koksal, I., Yilmaz, G., Aksoy, F., Aydin, H., Yavuz, I., Iskender, S., Akcay, K., et al. (2010). The efficacy of ribavirin in the treatment of Crimean-Congo hemorrhagic fever in Eastern Black Sea region in Turkey. *Journal of Clinical Virology*, *47*(1), 65–68.

Korolev, M. B., Donets, M. A., Rubin, S. G., & Chumakov, M. P. (1976). Morphology and Morphogenesis of Crimean Hemorrhagic Fever Virus. *Archives of Virology*, *50*, 169–172.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*, 680–685.

Lotfollahzadeh, S., Nikbakht Boroujeni, G. R., Mokhber Dezfouli, M. R., & Bokaei, S. (2011). A serosurvey of Crimean-Congo haemorrhagic fever virus in dairy cattle in Iran. *Zoonoses and Public Health*, *58*(1), 54–59.

Makrides, S. C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological Reviews*, *60*(3), 512–538.

Maltezou, H. C., Papa, A., Tsiodras, S., Dalla, V., Maltezos, E., & Antoniadis, A. (2009). Crimean-Congo hemorrhagic fever in Greece: a public health perspective. *International Journal of Infectious Diseases*, *13*(6), 713–716.

Maltezou, H. C., Andonova, L., Andraghetti, R., Bouloy, M., Ergonul, O., Jongejan, F., Kalvatchev, N., et al. (2010). Crimean-Congo hemorrhagic fever in Europe: current situation calls for preparedness. *Euro Surveillance*, *15*(10), e1350.

Marriot, A. C., & Nuttall, P. A. (1992). Comparison of the S RNA segments and nucleoprotein sequences of Crimean-Congo hemorrhagic fever, Hazara and Dugbe viruses. *Virology*, *189*, 795–799.

Marriot, A. C., Polyzoni, T., Antoniadis, A., & Nuttall, P. A. (1994). Detection of human antibodies to Crimean-Congo haemorrhagic fever virus using expressed viral nucleocapsid protein. *Journal of General Virology*, *75*, 2157–2161.

Marriot, A. C., & Nuttall, P. (1996). Molecular biology of nairoviruses. In R. M. Elliot (Ed.), *The Bunyaviridae*. New York: Plenum Press.

Midilli, K., Gargili, A., Ergonul, O., Elevli, M., Ergin, S., Turan, N., Şengöz, G., et al. (2009). The first clinical case due to AP92 like strain of Crimean-Congo Hemorrhagic Fever virus and a field survey. *BMC Infectious Diseases*, *9*(1), 90.

Mild, M., Simon, M., Albert, J., & Mirazimi, A. (2010). Towards an understanding of the migration of Crimean-Congo hemorrhagic fever virus. *The Journal of General Virology*, *91*(1), 199–207.

Morikawa, S., Saijo, M., & Kurane, I. (2007). Recent progress in molecular biology of Crimean-Congo hemorrhagic fever. *Comparative Immunology, Microbiology and Infectious diseases*, *30*, 375–389.

Morrill, J. C., Soliman, A. K., Imam, I. Z., Botros, B. A., Moussa, M. I., & Watts, D. M. (1990). Serological evidence of Crimean-Congo haemorrhagic fever viral infection among camels imported into Egypt. *The Journal of Tropical Medicine and Hygiene*, *93*(3), 201–204.

Mourya, D. T., Yadav, P. D., Shete, A. M., Gurav, Y. K., Raut, C. G., Jadi, R. S., Pawar, S. D., et al. (2012). Detection, isolation and confirmation of Crimean-Congo hemorrhagic fever virus in human, ticks and animals in Ahmadabad, India, 2010-2011. *PLoS Neglected Tropical Diseases*, *6*(5), e1653.

Mousavi-Jazi, M., Karlberg, H., Papa, A., Christova, I., & Mirazimi, A. (2012). Healthy individuals' immune response to the Bulgarian Crimean-Congo hemorrhagic fever virus vaccine. *Vaccine*, *30*(44), 6225–6229.

Mustafa, M. L., Ayazi, E., Mohareb, E., Yingst, S., Zayed, A., Rossi, C. A., Schoepp, R. J., et al. (2011). Crimean-Congo hemorrhagic fever, Afghanistan, 2009., *17*(10), 1940–1941.

Nabeth, P., Cheikh, D. O., Lo, B., Faye, O., Vall, I. O. M., Niang, M., Wague, B., et al. (2004). Crimean-Congo hemorrhagic fever, Mauritania. *Emerging Infectious Diseases*, *10*(12), 2143–2149.

Ozkaya, E., Dincer, E., Carhan, A., Uyar, Y., Ertek, M., Whitehouse, C. A., & Ozkul, A. (2010). Molecular epidemiology of Crimean-Congo hemorrhagic fever virus in Turkey: occurrence of local topotype. *Virus Research*, *149*(1), 64–70.

Papa, A., Bino, S., Llagami, A., Brahimaj, B., Papadimitriou, E., Pavlidou, V., Velo, E., et al. (2002a). Crimean-Congo hemorrhagic fever in Albania, 2001. *European Journal of Clinical Microbiology and Infectious Diseases*, *21*(8), 603–606.

Papa, A., Bożović, B., Pavlidou, V., Papadimitriou, E., Pelemis, M., & Antoniadis, A. (2002b). Genetic Detection and Isolation of Crimean-Congo hemorrhagic fever virus, Kosovo, Yugoslavia. *Emerging Infectious Diseases*, *8*(8), 852–854.

Papa, A., Papadimitriou, E., Bożović, B., & Antoniadis, A. (2005). Genetic characterization of the M RNA segment of a Balkan Crimean-Congo hemorrhagic fever virus strain. *Journal of Medical Virology*, *75*(3), 466–469.

Papa, A., Drosten, C., Bino, S., Papadimitriou, E., Panning, M., Velo, E., Kota, M., et al. (2007). Viral Load and Crimean-Congo Hemorrhagic Fever. *Emerging Infectious Diseases*, *13*(5), 805–806.

Papa, A., Maltezou, H. C., Tsiodras, S., Dalla, V. G., Papadimitriou, T., Pierroutsakos, I., & Kartalis, G. N. (2008). A case of Crimean-Congo Haemorrhagic Fever in Greece, June 2008. *Euro Surveillance*, *13*(7), 13–14.

Papa, A., Dalla, V., Papadimitriou, E., Kartalis, G. N., & Antoniadis, A. (2010). Emergence of Crimean-Congo haemorrhagic fever in Greece. *Clinical Microbiology and Infection*, *16*(7), 843–847.

Papa, A., Papadimitriou, E., & Christova, I. (2011). The Bulgarian vaccine Crimean-Congo haemorrhagic fever virus strain. *Scandinavian Journal of Infectious Diseases*, *43*(3), 225–229.

Patel, A. K., Patel, K. K., Mehta, M., Parikh, T. M., Toshniwal, H., & Patel, K. (2011). First Crimean-Congo hemorrhagic fever outbreak in India. *Journal of the Association of Physicians of India*, *59*, 585–589.

Phadtare, S., & Inouye, M. (2008). Cold-Shock Proteins. In R. Margesin (Ed.), *Psychrophiles: from Biodiversity to Biotechnology* (pp. 191–209). Berlin: Springer-Verlag.

Promega. BioMath Calculator. Retrieved from www.promega.com/biomath

Qing, T., Saijo, M., Lei, H., Niikura, M., Maeda, A., Ikegami, T., Xinjung, W., et al. (2003). Detection of immunoglobulin G to Crimean-Congo hemorrhagic fever virus in sheep sera by recombinant nucleoprotein-based enzyme-linked immunosorbent and immunofluorescence assays. *Journal of Virological Methods*, *108*(1), 111–116.

Sahdev, S., Khattar, S. K., & Saini, K. S. (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry*, *307*, 249–264.

Saijo, M., Qing, T., Niikura, M., Maeda, A., Ikegami, T., Sakai, K., Prehaud, C., et al. (2002). Immunofluorescence Technique Using HeLa Cells Expressing Recombinant Nucleoprotein for Detection of Immunoglobulin G Antibodies to Crimean-Congo Hemorrhagic Fever Virus. *Journal of Clinical Microbiology*, *40*(2), 372–375.

Saijo, M., Qing, T., Niikura, M., Maeda, A., Ikegami, T., Sakai, K., Prehaud, C., et al. (2002a). Immunofluorescence Technique Using HeLa Cells Expressing Recombinant Nucleoprotein for Detection of Immunoglobulin G Antibodies to Crimean-Congo Hemorrhagic Fever Virus. *Journal of Clinical Microbiology*, *40*(2), 372–375.

Saijo, M., Qing, T., Niikura, M., Maeda, A., Ikegami, T., Prehaud, C., Kurane, I., et al. (2002b). Recombinant Nucleoprotein-Based Enzyme-Linked Immunosorbent Assay for Detection of Immunoglobulin G Antibodies to Crimean-Congo Hemorrhagic Fever Virus. *Journal of Clinical Microbiology*, *40*(5), 1587–1591.

Saijo, M., Tang, Q., Shimayi, B., Han, L., Zhang, Y., Asiguma, M., Tianshu, D., et al. (2005a). Antigen-capture enzyme-linked immunosorbent assay for the diagnosis of crimean-congo hemorrhagic fever using a novel monoclonal antibody. *Journal of Medical Virology*, *77*, 83–88.

Saijo, M., Tang, Q., Shimayi, B., Han, L., Zhang, Y., Asiguma, M., Tianshu, D., et al. (2005b). Recombinant Nucleoprotein-Based Serological Diagnosis of Crimean – Congo Hemorrhagic Fever Virus Infections. *Journal of Medical Virology*, *75*, 295–299.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*, *4*(4), 406–425.

Sambrook, J., & Russel, D. W. (2001). *Molecular Cloning: a laboratory manual* (3rd ed.). New York: Cold Spring Harbor.

Samudzi, R. R., Leman, P. A., Paweska, J. T., Swanepoel, R., & Burt, F. J. (2012). Bacterial expression of Crimean-Congo hemorrhagic fever virus nucleoprotein

and its evaluation as a diagnostic reagent in an indirect ELISA. *Journal of Virological Methods*, *179*(1), 70–76.

Sanchez, A. J., Vincent, M. J., & Nichol, S. T. (2002). Characterization of the Glycoproteins of Crimean-Congo Hemorrhagic Fever Virus. *Journal of Virology*, *76*(14), 7263–7275.

Schmaljohn, C. S., & Hooper, J. W. (2001). Bunyaviridae: the viruses and their replication. In D. M. Knipe & P. M. Howley (Eds.), *Fields Virology* (pp. 1581–1602). Philadelphia: Lippincott Williams and Wilkins.

Seregin, S. V, Samokhvalov, E. I., Petrova, I. D., Vyshemirskii, O. I., Samokhvalova, E. G., Lvov, D. K., Gutorov, V. V, et al. (2004). Genetic characterization of the M RNA segment of Crimean-Congo hemorrhagic fever virus strains isolated in Russia and Tajikistan. *Virus Genes*, *28*(2), 187–193.

Sharifi-Mood, B., Metanat, M., Ghorbani-Vaghei, A., Fayyaz-Jahani, F., & Akrami, E. (2009). The outcome of patients with Crimean-Congo hemorrhagic fever in Zahedan, southeast of Iran: a comparative study. *Archives of Iranian Medicine*, *12*(2), 151–153.

Shepherd, A. J., Swanepoel, R., Leman, P. A., & Shepherd, S. P. (1987). Field and laboratory investigation of Crimean-Congo haemorrhagic fever virus (Nairovirus, family Bunyaviridae) infection in birds. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *81*(6), 1004–1007.

Shepherd, A. J., Swanepoel, R., & Gill, D. E. (1988). Evaluation of enzyme-linked immunosorbent assay and reversed passive hemagglutination for detection of Crimean-Congo hemorrhagic fever virus antigen. *Journal of Clinical Microbiology*, *26*(2), 347–353.

Shepherd, A. J., Swanepoel, R., Cornel, A. J., & Mathee, O. (1989a). Experimental studies on the replication and transmission of Crimean-Congo hemorrhagic fever virus in some African tick species. *The American Journal of Tropical Medicine and Hygiene*, *40*(3), 326–331.

Shepherd, A. J., Swanepoel, R., & Leman, P. A. (1989b). Antibody Response in Crimean-Congo Hemorrhagic Fever. *Reviews of Infectious Diseases*, *11*(4), S801–S806.

Sidira, P., Maltezou, H. C., Haidich, A. B., & Papa, A. (2012). Seroepidemiological study of Crimean-Congo haemorrhagic fever in Greece, 2009-2010. *Clinical Microbiology and Infection*, *18*(2), E16–E19.

Sisman, A. (2012). Epidemiologic Features and Risk Factors of Crimean-Congo Hemorrhagic Fever in Samsun Province, Turkey. *Journal of Epidemiology*.

Sun, S., Dai, X., Aishan, M., Wang, X., Meng, W., Feng, C., Zhang, F., et al. (2009). Epidemiology and phylogenetic analysis of crimean-congo hemorrhagic fever viruses in Xinjiang, China. *Journal of Clinical Microbiology*, *47*(8), 2536–2543.

Swanepoel, R., Struthers, J. K., Shepherd, A. J., McGillivray, G. M., Nel, M. J., & Jupp, P. G. (1983). Crimean-congo hemorrhagic fever in South Africa. *The American Journal of Tropical Medicine and Hygiene*, *32*(6), 1407–1415.

Swanepoel, R., Shepherd, A. J., Leman, P. A., Shepherd, S. P., McGillivray, G. M., Erasmus, M. J., Searle, L. A., et al. (1987). Epidemiologic and clinical features of Crimean-Congo hemorrhagic fever in southern Africa. *The American Journal of Tropical Medicine and Hygiene*, *36*(1), 120–132.

Swanepoel, R., Gill, D. E., Shepherd, A. J., Leman, P. A., Mynhardt, J. H., & Harvey, S. (1989). The Clinical Pathology of Crimean-Congo Hemorrhagic Fever. *Reviews of Infectious Diseases*, *11*(4), S794–S800.

Swanepoel, R., Leman, P. A., Burt, F. J., Jardine, J., Verwoerd, D. J., Capua, I., Brückner, G. K., et al. (1998). Experimental infection of ostriches with Crimean-Congo haemorrhagic fever virus. *Epidemiology and Infection*, *121*(2), 427–432.

Swiss Institute of Bioinformatics. Viralzone. Retrieved from www.expasy.ch/viralzone

Tall, A., Sall, A. A., Faye, O., Diatta, B., Sylla, R., Faye, J., Faye, P. C., et al. (2009). Two cases of Crimean-Congo haemorrhagic fever (CCHF) in two tourists in

Senegal in 2004. *Bulletin de la Societe de Pathologie Exotique 1990*, *102*(3), 159–161.

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4 .0. *Molecular Biology and Evolution*, *24*(8), 1596–1599.

Thomas, S., Thomson, G., Dowall, S., Bruce, C., Cook, N., Easterbrook, L., O'Donoghue, L., et al. (2012). Review of Crimean Congo hemorrhagic fever infection in Kosova in 2008 and 2009: prolonged viremias and virus detected in urine by PCR. *Vector borne and Zoonotic Diseases*, *12*(9), 800–804.

Tignor, G. H., Smith, A. L., Casals, J., Ezeokoli, C. D., & Okoli, J. (1980). Close relationship of Crimean hemorrhagic fever-Congo (CHF-C) virus strains by neutralizing antibody assays. *The American Journal of Tropical Medicine and Hygiene*, *29*(4), 676–685.

Tonbak, S., Aktas, M., Altay, K., Azkur, A. K., Kalkan, A., Bolat, Y., Dumanli, N., et al. (2006). Crimean-Congo Hemorrhagic Fever Virus: Genetic Analysis and Tick Survey in Turkey. *Journal of Clinical Microbiology*, *44*(11), 4120–4124.

Turell, M. J. (2007). Role of ticks in the transmission of Crimean-Congo Hemorrhagic Fever virus. In O. Ergonul & C. A. Whitehouse (Eds.), *Crimean-Congo Hemorrhagic fever* (pp. 143–154). US Government.

Umoh, J. U., Ezeokoli, C. D., & Ogwu, D. (1983). Prevalence of antibodies to Crimean-haemorrhagic fever-Congo virus in cattle in northern Nigeria. *International Journal Of Zoonoses*, *10*(2), 151–154.

Vanhomwegen, J., Alves, M. J., Županc, T. A., Bino, S., Chinikar, S., Karlberg, H., Korukluoğlu, G., et al. (2012). Diagnostic Assays for Crimean-Congo Hemorrhagic Fever. *Emerging Infectious Diseases*, *18*(12), 1958–1965.

Watts, D. M., Ussery, M. A., Nash, D., & Peters, C. J. (1989). Inhibition of Crimean-Congo hemorrhagic fever viral infectivity yields in vitro by ribavirin. *The American Journal of Tropical Medicine and Hygiene*, *41*(5), 581–585.

Weber, F., & Mirazimi, A. (2008). Interferon and cytokine responses to Crimean Congo hemorrhagic fever virus; an emerging and neglected viral zonoosis. *Cytokine & Growth Factor Reviews*, *19*, 395–404.

Welch, M., Villalobos, A., Gustafsson, C., & Minshull, J. (2009). You're one in a googol: optimizing genes for protein expression. *Journal of the Royal Society, Interface*, *6*(4), S467–S476.

Williams, R. J., Al-Busaidy, S., Mehta, F. R., Maupin, G. O., Wagoner, K. D., Al-Awaidy, S., Suleiman, A. J., et al. (2000). Crimean-congo haemorrhagic fever: a seroepidemiological and tick survey in the Sultanate of Oman. *Tropical Medicine; International Health*, *5*(2), 99–106.

Wilson, M. L., & Digoutte, J. (1989). Ecology and epidemiology of Crimean-Congo hemorrhagic fever virus transmission in the Republic of Senegal. *Annual Report*, 1–51.

Wölfel, R., Paweska, J. T., Petersen, N., Grobbelaar, A. A., Leman, P. A., Hewson, R., Georges-Courbot, M.-C., et al. (2007). Virus Detection and Monitoring of Viral Load in Crimean-Congo Hemorrhagic Fever Virus Patients. *Emerging Infectious Diseases*, *13*(7), 1097–1100.

Wong, C., Sridhara, S., Bardwell, J. C. A., & Jakob, U. (1999). Heating greatly speeds Coomassie Blue staining and destaining. *Biotechniques*, *28*(3), 426–428.

Yashina, L., Vyshemirskii, O., Seregin, S., Petrova, I., Samokhvalov, E., Lvov, D., Gutorov, V., et al. (2003a). Genetic Analysis of Crimean-Congo Hemorrhagic Fever Virus in Russia. *Journal of Clinical Microbiology*, *41*(2), 860–862.

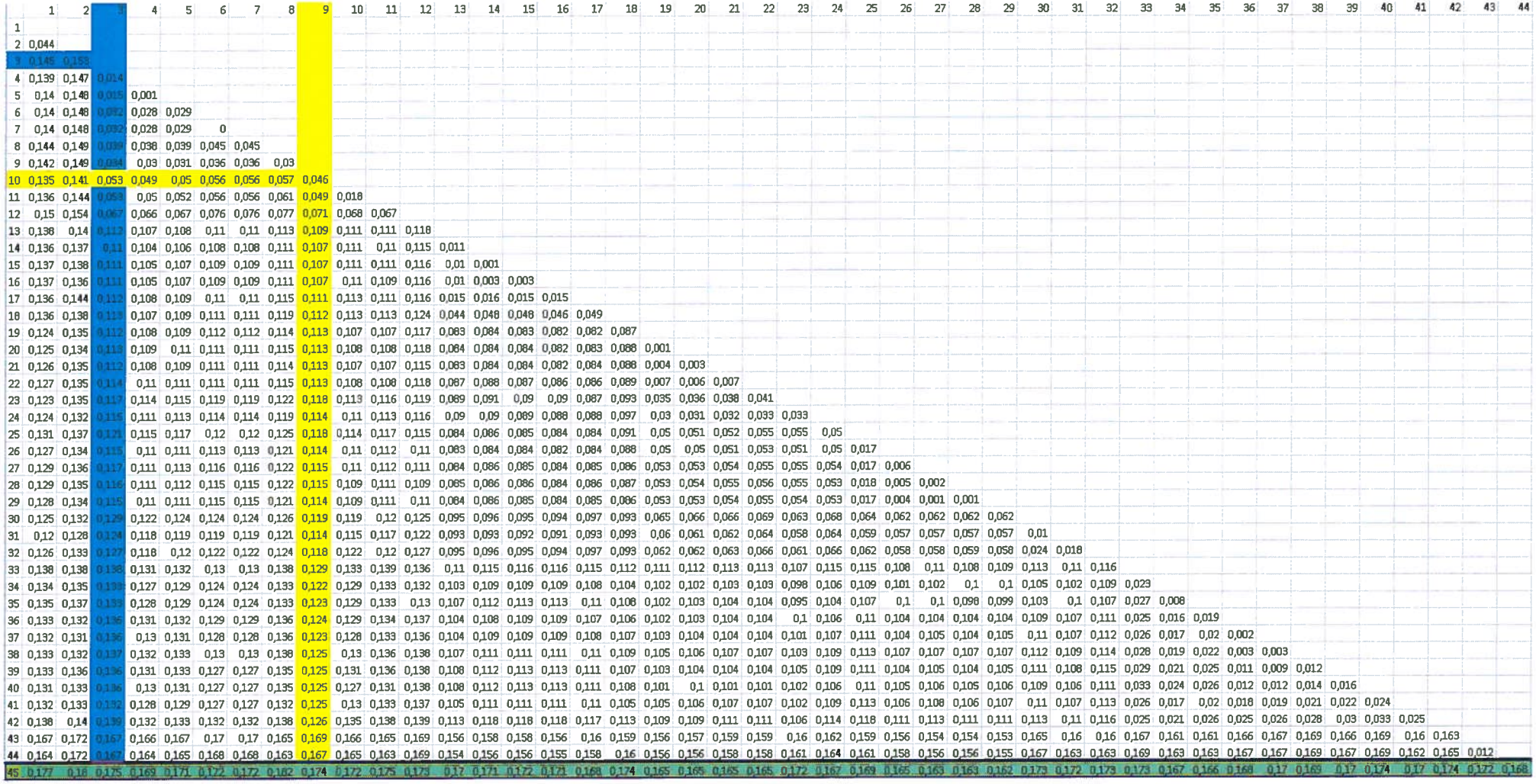
Yashina, L., Petrova, I., Seregin, S., Vyshemirskii, O., Lvov, D., Aristova, V., Kuhn, J., et al. (2003b). Genetic variability of Crimean-Congo haemorrhagic fever virus in Russia and Central Asia. *Journal of General Virology*, *84*, 1199–1206.

Zeller, H. (2007). Laboratory diagnosis of Crimean Congo haemorrhagic fever. In C. Whitehouse & O. Ergonul (Eds.), *Crimean-Congo haemorrhagic fever* (pp. 233–243). Dordrecht: Springer Press.

Appendix 1: Pairwise distances (nucleic acid and amino acid)

Legend for pairwise distance comparisons

1. CONGO3010 {II DRC}
2. SEMUNYA {II DRC}
3. SPU415/85 {III South Africa / West Africa}
4. SPU4/81 {III South Africa / West Africa}
5. SPU128/81/7 {III South Africa / West Africa}
6. SUDAN42008 {III South Africa / West Africa}
7. SUDAN32008 {III South Africa / West Africa}
8. ArD39554 {III South Africa / West Africa}
9. SPU92/01 {III South Africa / West Africa}
10. SPU103/87 {III South Africa / West Africa}
11. SPU97/85 {III South Africa / West Africa}
12. 10200 {III South Africa / West Africa}
13. JD206 {IV Asia / Middle East}
14. IRAN56 {IV Asia / Middle East}
15. IRAN53 {IV Asia / Middle East}
16. IRAN52 {IV Asia / Middle East}
17. MATIN {IV Asia / Middle East}
18. OMAN {IV Asia / Middle East}
19. HY13 {IV Asia / Middle East}
20. C68031 {IV Asia / Middle East}
21. CHINA75024 {IV Asia / Middle East}
22. CHINA88166 {IV Asia / Middle East}
23. TAJ/HU8975 {IV Asia / Middle East}
24. UZBEK/TI10145 {IV Asia / Middle East}
25. 79121M18 {IV Asia / Middle East}
26. YL04051 {IV Asia / Middle East}
27. CYT/TI05099 {IV Asia / Middle East}
28. CYT/TI05146 {IV Asia / Middle East}
29. CYT/TI05035 {IV Asia / Middle East}
30. TAJ/HU8978 {IV Asia / Middle East}
31. TAJ/HU8966 {IV Asia / Middle East}
32. HODZHA {IV Asia / Middle East}
33. Turkeykelkit06 {V Europe / Turkey}
34. 95532001 {V Europe / Turkey}
35. KOSOVO HOTI {V Europe / Turkey}
36. ROS/HULV-100 {V Europe / Turkey}
37. KASHMANOV {V Europe / Turkey}
38. ROS/TI28044 {V Europe / Turkey}
39. DROSDOV {V Europe / Turkey}
40. STV/HU29223 {V Europe / Turkey}
41. TURKEY200310849 {V Europe / Turkey}
42. BUL/HU517 {V Europe / Turkey}
43. ArD8194 {I West Africa}
44. ArD15786 {I West Africa}
45. AP92



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44
1																																												
2	0,044																																												
3	0,145	0,153																																											
4	0,139	0,147	0,014																																										
5	0,14	0,148	0,015	0,001																																									
6	0,14	0,148	0,092	0,028	0,029																																							
7	0,14	0,148	0,092	0,028	0,029	0																																					
8	0,144	0,149	0,039	0,038	0,039	0,045	0,045																																					
9	0,142	0,149	0,084	0,03	0,031	0,036	0,036	0,03																																				
10	0,135	0,141	0,053	0,049	0,05	0,056	0,056	0,057	0,046																																			
11	0,136	0,144	0,053	0,05	0,052	0,056	0,056	0,061	0,049	0,018																																		
12	0,15	0,154	0,067	0,066	0,067	0,076	0,076	0,077	0,071	0,068	0,067																																	
13	0,138	0,14	0,112	0,107	0,108	0,11	0,11	0,113	0,109	0,111	0,111	0,118																																
14	0,136	0,137	0,11	0,104	0,106	0,108	0,108	0,111	0,107	0,111	0,11	0,115	0,011																															
15	0,137	0,138	0,111	0,105	0,107	0,109	0,109	0,111	0,107	0,111	0,111	0,116	0,01	0,001																														
16	0,137	0,136	0,111	0,105	0,107	0,109	0,109	0,111	0,107	0,11	0,109	0,116	0,01	0,003	0,003																														
17	0,136	0,144	0,112	0,108	0,109	0,11	0,11	0,115	0,111	0,113	0,111	0,116	0,015	0,016	0,015	0,015																													
18	0,136	0,138	0,113	0,107	0,109	0,111	0,111	0,119	0,112	0,113	0,113	0,124	0,044	0,048	0,048	0,046	0,049																												
19	0,124	0,135	0,112	0,108	0,109	0,112	0,112	0,114	0,113	0,107	0,107	0,117	0,083	0,084	0,083	0,082	0,082	0,087																											
20	0,125	0,134	0,113	0,109	0,11	0,111	0,111	0,115	0,113	0,108	0,108	0,118	0,084	0,084	0,084	0,082	0,083	0,088	0,001																										
21	0,126	0,135	0,112	0,108	0,109	0,111	0,111	0,114	0,113	0,107	0,107	0,115	0,083	0,084	0,084	0,082	0,084	0,088	0,004	0,003																									
22	0,127	0,135	0,114	0,11	0,111	0,111	0,111	0,115	0,113	0,108	0,108	0,118	0,087	0,088	0,087	0,086	0,086	0,089	0,007	0,006	0,007																								
23	0,123	0,135	0,117	0,114	0,115	0,119	0,119	0,122	0,118	0,113	0,116	0,119	0,089	0,091	0,09	0,09	0,087	0,093	0,035	0,036	0,038	0,041																						
24	0,124	0,132	0,115	0,111	0,113	0,114	0,114	0,119	0,114	0,11	0,113	0,116	0,09	0,09	0,089	0,088	0,088	0,097	0,03	0,031	0,032	0,033	0,033																					
25	0,131	0,137	0,121	0,115	0,117	0,12	0,12	0,125	0,118	0,114	0,117	0,115	0,084	0,086	0,085	0,084	0,084	0,091	0,05	0,051	0,052	0,055	0,055	0,05																				
26	0,127	0,134	0,115	0,11	0,111	0,113	0,113	0,121	0,114	0,11	0,112	0,11	0,083	0,084	0,084	0,082	0,084	0,088	0,05	0,05	0,051	0,053	0,051	0,05	0,017																			
27	0,129	0,136	0,117	0,111	0,113	0,116	0,116	0,122	0,115	0,11	0,112	0,111	0,084	0,086	0,085	0,084	0,085	0,086	0,053	0,053	0,054	0,055	0,055	0,054	0,017	0,006																		
28	0,129	0,135	0,116	0,111	0,112	0,115	0,115	0,122	0,115	0,109	0,111	0,109	0,085	0,086	0,086	0,084	0,086	0,087	0,053	0,054	0,055	0,056	0,055	0,053	0,018	0,005	0,002																	
29	0,128	0,134	0,115	0,11	0,111	0,115	0,115	0,121	0,114	0,109	0,111	0,11	0,084	0,086	0,085	0,084	0,085	0,086	0,053	0,053	0,054	0,055	0,054	0,053	0,017	0,004	0,001	0,001															
30	0,125	0,132	0,129	0,122	0,124	0,124	0,124	0,126	0,119	0,119	0,12	0,125	0,095	0,096	0,095	0,094	0,097	0,093	0,065	0,066	0,066	0,069	0,063	0,068	0,064	0,062	0,062	0,062	0,062														
31	0,12	0,128	0,124	0,118	0,119	0,119	0,119	0,121	0,114	0,115	0,117	0,122	0,093	0,093	0,092	0,091	0,093	0,093	0,06	0,061	0,062	0,064	0,058	0,064	0,059	0,057	0,057	0,057	0,057	0,01													
32	0,126	0,133	0,127	0,118	0,12	0,122	0,122	0,124	0,118	0,122	0,12	0,127	0,095	0,096	0,095	0,094	0,097	0,093	0,062	0,062	0,063	0,066	0,061	0,066	0,062	0,058	0,058	0,059	0,058	0,024	0,018											
33	0,138	0,138	0,136	0,131	0,132	0,13	0,13	0,138	0,129	0,133	0,139	0,136	0,11	0,115	0,116	0,116	0,115	0,112	0,111	0,112	0,113	0,113	0,107	0,115	0,115	0,108	0,11	0,108	0,109	0,113	0,11	0,116										
34	0,134	0,135	0,133	0,127	0,129	0,124	0,124	0,133	0,122	0,129	0,133	0,132	0,103	0,109	0,109	0,109	0,108	0,104	0,102	0,102	0,103	0,103	0,098	0,106	0,109	0,101	0,102	0,1	0,1	0,105	0,102	0,109	0,023									
35	0,135	0,137	0,133	0,128	0,129	0,124	0,124	0,133	0,123	0,129	0,133	0,13	0,107	0,112	0,113	0,113	0,11	0,108	0,102	0,103	0,104	0,104	0,095	0,104	0,107	0,1	0,1	0,098	0,099	0,103	0,1	0,107	0,027	0,008								
36	0,133	0,132	0,136	0,131	0,132	0,129	0,129	0,136	0,124	0,129	0,134	0,137	0,104	0,108	0,109	0,109	0,107	0,106	0,102	0,103	0,104	0,104	0,1	0,106	0,11	0,104	0,104	0,104	0,104	0,109	0,107	0,111	0,025	0,016	0,019							
37	0,132	0,131	0,136	0,13	0,131	0,128	0,128	0,136	0,123	0,128	0,133	0,136	0,104	0,109	0,109	0,109	0,108	0,107	0,103	0,104	0,104	0,104	0,101	0,107	0,111	0,104	0,105	0,104	0,105	0,11	0,107	0,112	0,026	0,017	0,02	0,002						
38	0,133	0,132	0,137	0,132	0,133	0,13	0,13	0,138	0,125	0,13	0,136	0,138	0,107	0,111	0,111	0,111	0,11	0,109	0,105	0,106	0,107	0,107	0,103	0,109	0,113	0,107	0,107	0,107	0,107	0,112	0,109	0,114	0,028	0,019	0,022	0,003	0,003					
39	0,133	0,136	0,136	0,131	0,133	0,127	0,127	0,135	0,125	0,131	0,136	0,138	0,108	0,112	0,113	0,113	0,111	0,107	0,103	0,104	0,104	0,104	0,105	0,109	0,111	0,104	0,105	0,104	0,105	0,111	0,108	0,115	0,029	0,021	0,025	0,011	0,009	0,012				
40	0,131	0,133	0,136	0,13	0,131	0,127	0,127	0,135	0,125	0,127	0,131	0,138	0,108	0,112	0,113	0,113	0,111	0,108	0,101	0,1	0,101	0,101	0,102	0,106	0,11	0,105	0,106	0,105	0,106	0,109	0,106	0,111	0,033	0,024	0,026	0,012	0,012	0,014	0,016			
41	0,132	0,133	0,132	0,128	0,129	0,127	0,127	0,132	0,125	0,13	0,133	0,137	0,105	0,111	0,111	0,111	0,11	0,105	0,105	0,106	0,107	0,107	0,102	0,109	0,113	0,106	0,108	0,106	0,107	0,11	0,107	0,113	0,026	0,017	0,02	0,018	0,019	0,021	0,022	0,024		
42	0,138	0,14	0,139	0,132	0,133	0,132	0,132	0,138	0,126	0,135	0,138	0,139	0,113	0,118	0,118	0,118	0,117	0,113	0,109	0,109	0,111	0,111	0,106	0,114	0,118	0,111	0,113	0,111	0,111	0,113	0,11	0,116	0,025	0,021	0,026	0,025	0,026	0,028	0,03	0,033	0,025	
43	0,167	0,172	0,167	0,166	0,167	0,17	0,17	0,165	0,169	0,166	0,165	0,169	0,156	0,158	0,158	0,156	0,16	0,159	0,156	0,157	0,159	0,159	0,16	0,162	0,159	0,156	0,154	0,154	0,153	0,165	0,16	0,16	0,167	0,161	0,161	0,166	0,167	0,169	0,166	0,169	0,16	0,163
44	0,164	0,172	0,167	0,164	0,165	0,168	0,168	0,163	0,167	0,165	0,163	0,169	0,154	0,156	0,156	0,155	0,158	0,16	0,156	0,156	0,158	0,158	0,161	0,164	0,161	0,158	0,156	0,156	0,155	0,167	0,163	0,163	0,169	0,163	0,163	0,167	0,167	0,169	0,167	0,169	0,162	0,165	0,012
45	0,177	0,18	0,175	0,169	0,171	0,172	0,172	0,162	0,174	0,172	0,175	0,173	0,17	0,171	0,172	0,171	0,168	0,174	0,165	0,165	0,165	0,165	0,172	0,167	0,169	0,165	0,163	0,163	0,162	0,173	0,172	0,173	0,173	0,167	0,166	0,168	0,17	0,169	0,17	0,174	0,17	0,174	0,172	0,168

Figure 45: Pairwise comparisons of 45 isolates of CCHFV based on nucleotide sequence data. Isolates SPU92/01 (9), SPU415/85 (3) and AP92 (45) are highlighted

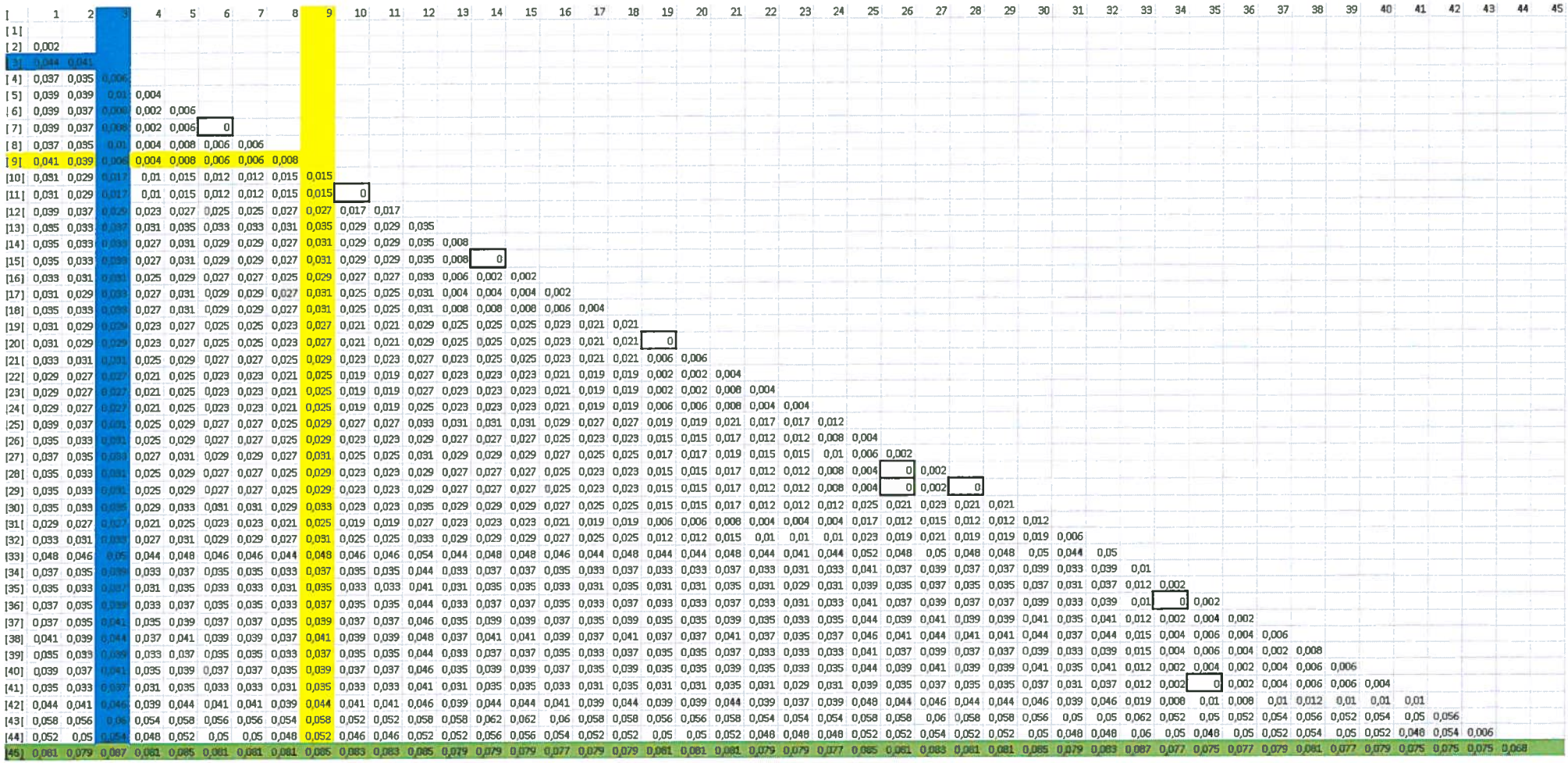


Figure 46: Pairwise comparisons of 45 isolates of CCHFV based on amino acid sequence data. Isolates SPU92/01 (9), SPU415/85 (3) and AP92 (45) are highlighted

Appendix 2: ELISA raw data

SPU NC	PANEL											
A1:100	0,3040	0,4730	0,3400	0,2850	0,3770	0,3560	0,2200	0,2050	0,3700	0,3750	0,4270	0,4760
B1:100	0,5780	0,1950	0,3980	0,1830	0,1180	0,0550	0,1440	0,0860	0,0100	0,0170	0,0180	0,0190

Figure 47: Panel of sera from CCHFV survivors and negative panel tested against SPU NC. Yellow highlights indicate the negative panel, while grey areas depict wells without sera.

AP92NC	PANEL											
A1:100	0,5050	0,7930	0,3140	0,9400	0,5620	0,8570	0,3330	0,4600	0,6530	0,6400	0,7600	0,4960
B1:100	0,9260	0,4570	0,3980	0,1540	0,3170	0,1190	0,2500	0,1610	0,0170	0,0160	0,0150	0,0240

Figure 48: Panel of sera from CCHFV survivors and negative panel tested against AP92 NC. Yellow highlights indicate the negative panel, while grey areas depict wells without sera.

TF	TITRATION											
A1:100	0,1170	0,1330	0,1260	0,1050	0,1790	0,1060	0,0580	0,0670	0,0200	0,0170	0,0180	0,0180
B1:200	0,0930	0,1290	0,0670	0,0590	0,1060	0,0650	0,0240	0,0280	0,0070	0,0090	0,0240	0,0230
C1:400	0,0510	0,0820	0,0640	0,0300	0,0700	0,0460	0,0200	0,0190	0,0040	0,0100	0,0160	0,0200
D1:800	0,0430	0,0670	0,0330	0,0240	0,0540	0,0350	0,0100	0,0170	0,0120	0,0130	0,0250	0,0240
E1:100	0,1110	0,0560	0,3680	0,0880	0,1070	0,0350	0,1100	0,0560	0,0100	0,0170	0,0180	0,0190
F1:200	0,0670	0,0350	0,1780	0,0480	0,0380	0,0110	0,0660	0,0330	0,0140	0,0220	0,0220	0,0220
G1:400	0,0390	0,0280	0,1080	0,0240	0,0370	0,0260	0,0610	0,0370	0,0220	0,0210	0,0220	0,0220
H1:800	0,0360	0,0240	0,0860	0,0330	0,0410	0,0270	0,0460	0,0350	0,0240	0,0170	0,0180	0,0190

Figure 49: Mock antigen tested against positive and negative panel of sera at dilutions 1:100 to 1:800. Yellow regions indicate negative panel and grey areas wells without sera.

SPU NC	TITRATION											
A1:100	0,3150	0,4870	0,3490	0,2870	0,3790	0,3790	0,2280	0,2060	0,3790	0,3870	0,4370	0,4790
B1:200	0,2040	0,2710	0,1760	0,1720	0,2090	0,2430	0,1220	0,1220	0,2280	0,2370	0,2420	0,2650
C1:400	0,1400	0,1930	0,0900	0,0940	0,1190	0,1500	0,0840	0,0800	0,1480	0,1600	0,1580	0,1620
D1:800	0,0870	0,1270	0,0640	0,0720	0,0730	0,0750	0,0500	0,0460	0,0900	0,0940	0,1040	0,1080
E1:100	0,5850	0,1980	0,4010	0,1780	0,2320	0,2070	0,1380	0,0760	0,0160	0,0150	0,0170	0,0170
F1:200	0,4530	0,0990	0,1930	0,0520	0,1810	0,0480	0,0920	0,0700	0,0170	0,0160	0,0170	0,0190
G1:400	0,3190	0,0590	0,0870	0,0380	0,0890	0,0320	0,0650	0,0470	0,0160	0,0160	0,0160	0,0220
H1:800	0,3020	0,0640	0,0750	0,0300	0,0750	0,0290	0,0680	0,0390	0,0190	0,0150	0,0170	0,0170

Figure 50: SPU NC antigen tested against positive and negative panel of sera at dilutions 1:100 to 1:800. Yellow regions indicate negative panel and grey areas wells without sera.

AP92 NC	TITRATION											
A1:100	0,5880	0,8050	0,2960	0,7830	0,6000	0,9120	0,3010	0,5160	0,7560	0,9920	0,9830	0,4530
B1:200	0,3290	0,5160	0,1780	0,4880	0,2860	0,5420	0,0830	0,1240	0,1890	0,2260	0,2580	0,1740
C1:400	0,1870	0,3400	0,0940	0,2530	0,1850	0,3680	0,0510	0,0590	0,1160	0,1290	0,1270	0,1010
D1:800	0,1180	0,2230	0,0580	0,1280	0,1070	0,2220	0,0360	0,0370	0,0690	0,0600	0,0710	0,0590
E1:100	0,9250	0,4530	0,4270	0,1710	0,2840	0,1250	0,2440	0,1620	0,0130	0,0130	0,0160	0,0210
F1:200	0,7830	0,2960	0,4110	0,0990	0,2320	0,0900	0,1570	0,0970	0,0140	0,0140	0,0170	0,0220
G1:400	0,4650	0,2060	0,4030	0,0550	0,1740	0,0580	0,1220	0,0550	0,0120	0,0120	0,0160	0,0210
H1:800	0,3210	0,1480	0,2620	0,0410	0,0980	0,0460	0,1060	0,0610	0,0190	0,0140	0,0250	0,0300

Figure 51: AP92 NC antigen tested against positive and negative panel of sera at dilutions 1:100 to 1:800. Yellow regions indicate negative panel and grey areas wells without sera.

	STABILITY SPU NC											
	C++	C++	C-	C-	C-	C-	C++	C++	C-	C-	C-	C-
A1:500	0,3440	0,3980	0,1780	0,1700	0,2660	0,1550	0,1380	0,0940	0,0200	0,0190	0,0260	0,0210
B1:1000	0,2890	0,2030	0,2410	0,2540	0,2780	0,2110	0,1400	0,0890	0,0180	0,0170	0,0200	0,0260
C1:2000	0,0930	0,1250	0,0480	0,0520	0,0750	0,0600	0,0580	0,0450	0,0150	0,0120	0,0130	0,0170
D:500	0,1220	0,1060	0,1010	0,0990	0,0720	0,0780	0,0470	0,0340	0,0200	0,0130	0,0160	0,0200
E:1000	0,1530	0,1440	0,0890	0,0880	0,0850	0,0760	0,0190	0,0290	0,0180	0,0130	0,0160	0,0200
F:2000	0,1410	0,1480	0,0860	0,0880	0,0880	0,0860	0,0220	0,0290	0,0140	0,0130	0,0150	0,0190

COATED PLATE STABILITY SPU NC					
C++	C++	C-	C-	C-	C-
0,2010	0,2050	0,0650	0,0590	0,0850	0,0740

Figure 52: Assessment of stability of prepared recombinant SPU NC antigen using different storage methods. Plate layout is depicted in Figure 40.

STABILITY AP92 NC												
	C++	C++	C-	C-	C-	C-	C++	C++	C-	C-	C-	C-
A1:500	0,4000	0,4300	0,1370	0,1190	0,0890	0,0940	1,7130	1,6280	0,4950	0,4940	0,2940	0,3020
B1:1000	0,3090	0,2990	0,1750	0,1710	0,1020	0,1160	1,3470	1,2350	0,2890	0,3210	0,2050	0,1850
C1:2000	0,2330	0,2500	0,1600	0,1610	0,1030	0,1080	0,9610	0,8140	0,2290	0,2420	0,1470	0,1500
D:500	0,6000	0,6400	0,2650	0,2670	0,1850	0,1760	0,0860	0,0830	0,1020	0,0970	0,0760	0,0890
E:1000	0,4800	0,5000	0,1850	0,1730	0,1310	0,1240	0,1040	0,1090	0,1170	0,1260	0,0860	0,1000
F:2000	0,3980	0,4800	0,1510	0,1470	0,1070	0,1070	0,1000	0,1230	0,1210	0,1360	0,0960	0,1000

COATED PLATE STABILITY AP92 NC					
C++	C++	C-	C-	C-	C-
1,1480	1,3100	0,6630	0,6850	0,3250	0,3410

Figure 53: Assessment of stability of prepared recombinant AP92 NC antigen using different storage methods. Plate layout is depicted in Figure 40.

Appendix 3: List of Figures, List of Tables, List of Abbreviations and Acronyms

List of Figures

Figure 1: Crimean-Congo haemorrhagic fever virus structure and proteins, nucleocapsid, glycoproteins (embedded in a lipid bilayer), RNA dependent RNA polymerase. 6

Figure 2: Crystal structure of Crimean-Congo haemorrhagic fever virus nucleocapsid revealing a "head" and "stalk" domain. 7

Figure 3: Open reading frame of polyprotein which is cleaved into Gn and Gc. 9

Figure 4: Clinical course of CCHF: The disease progression can be subdivided into four different stages, with different symptoms; incubation, pre-haemorrhagic, haemorrhagic and then convalescence phase. 13

Figure 5: Vector map pcDNA3.1 directional V5/His/TOPO mammalian expression vector showing important features on the vector. . 27

Figure 6: ORF of S segment of isolate SPU 92/01 amplified using Phusion® HotStart II High Fidelity System. 37

Figure 7: Agarose gel electrophoretic analysis of NC amplicon from the Phusion HotStart PCR reaction after Promega Wizard® SV Gel and PCR Clean-Up. 38

Figure 8: Agarose gel electrophoretic analysis of restriction enzyme analysis of a plasmid obtained from the ligation of the CCHFV NC gene into pcDNA™ 3.1/V5-His TOPO®. . 39

Figure 9: Contiguous sequencing map illustrating the coverage of the primers across the sequence of the NC gene in pcDNA3.1D/V5-His/TOPO vector. 40

Figure 10: Complete sequence of CCHFV NC of isolate SPU 92/01 cloned in frame in pcDNA3.1D/V5-His/TOPO vector. 41

Figure 11: Phylogenetic tree using nucleotide sequence data of ORF of NC genes. . 43

Figure 12: Phylogenetic tree using amino acid sequence data of NC. 47

Figure 13: Amino acid residue comparison between NC of isolate SPU 415/85 and AP92, the variable sites are highlighted. 51

Figure 14: Factors influencing protein expression from mRNA level to final recombinant protein. (Source: Welch *et al.*, 2009) 56

Figure 15: Expression control vector pcDNA3.1/V5-His-TOPO/*LacZ* showing promoter sites, antibiotic resistance genes, restriction enzyme sites and primer sites. 58

Figure 16: GeneJuice transfection reagent: DNA ratios used in transfection experiment using pcDNA3.1/V5-His-TOPO/*LacZ* 62

Figure 17: Vector map and multiple cloning sites of cloning vector pUC57 (GenScript Corporation, USA). 69

Figure 18: Vector map and MCS of pCold TF bacterial expression system (Takara Bio, USA). 72

Figure 19: Agarose electrophoretic gel of restriction enzyme analysis of plasmids obtained from purification of an overnight culture inoculated with TOPO *LacZ* V5/His Tag plasmid DNA. 80

Figure 20: HeLa cells transfected with a vector containing *LacZ* gene coding for β-galactosidase expressed. 84

Figure 21: Western Blot analysis of mammalian cell lysates and control antigen EDIII yellow fever containing a His-tag. 85

Figure 22: The distribution of codon usage frequency along the length of the gene sequence of the of CCHF AP92 NC before optimization. 86

Figure 23: The distribution of codon usage frequency along the length of the gene sequence of the of CCHF AP92 NC after optimization. 86

Figure 24: The percentage distribution of codons in computed codon quality groups before optimization. 87

Figure 25: The percentage distribution of codons in computed codon quality groups after optimization 87

Figure 26: Average GC content before optimization. 88

Figure 27: Average GC content after optimization. 88

Figure 28: Agarose gel electrophoretic analysis of restriction enzyme analysis of a plasmid obtained from pUC57 vector supplied by GenScript. 89

Figure 29: Purified linearized vector and insert rescued from pUC57 vector. 90

Figure 30: Agarose gel electrophoretic analysis of colony PCR of a plasmid DNA obtained from pCold TF-opAP92NC colonies using GoTaq polymerase. 90

Figure 31: Agarose gel electrophoretic analysis of restriction enzyme analysis of a plasmid that yielded a positive PCR product (1443 bp). 91

Figure 32: SDS PAGE analysis of proteins expressed using the recombinant expression vector, pColdTF-opSPUNC in OverExpress (DE3) competent cells. 92

Figure 33: SDS PAGE analysis of recombinant pColdTF-opSPUNC proteins to determine solubility of the expressed protein. 93

Figure 34: SDS PAGE analysis of recombinant pColdTF-opSPUNC proteins to determine solubility of the expressed protein. 93

Figure 35: SDS PAGE analysis of His-tag-purified pColdTF-opSPUNC 24 hours after induction. 94

Figure 36: SDS PAGE analysis of proteins expressed in a solubility study using the recombinant expression vector, pColdTF-opAP92NC in OverExpress (DE3) competent cells. 95

Figure 37: SDS PAGE analysis of pColdTF-opAP92NC protein purified using Ni+ column. 96

Figure 38: Indirect ELISA procedure: principle and protocol 104

Figure 39: ELISA plate layout for antibody titre determination experiment using positive and negative panel test sera. 106

Figure 40: Plate layout for assessment of different storage methods for retention of biological activity of recombinant antigen. 107

Figure 41: Detection of anti-CCHFV IgG by ELISA using recombinant NC from isolate SPU 415/85. 110

Figure 42: Detection of anti-CCHFV IgG by ELISA using recombinant NC from AP92 isolate. 112

Figure 43: Assessment of different storage methods for retention of biological activity of recombinant antigen SPU NC 114

Figure 44: Assessment of different storage methods for retention of biological activity of recombinant antigen AP92 NC. 115

Figure 45: Pairwise comparisons of 45 isolates of CCHFV based on nucleotide sequence data. 142

Figure 46: Pairwise comparisons of 45 isolates of CCHFV based on amino acid sequence data. 143

Figure 47: Panel of sera from CCHFV survivors and negative panel tested against SPU NC. 144

Figure 48: Panel of sera from CCHFV survivors and negative panel tested against AP92 NC. 144

Figure 49: Mock antigen tested against positive and negative panel of sera at dilutions 1:100 to 1:800. 144

Figure 50: SPU NC antigen tested against positive and negative panel of sera at dilutions 1:100 to 1:800. Y 145

Figure 51: AP92 NC antigen tested against positive and negative panel of sera at dilutions 1:100 to 1:800. 145

Figure 52: Assessment of stability of prepared recombinant SPU NC antigen using different storage methods. Plate layout is depicted in Figure 40. 145

Figure 53: Assessment of stability of prepared recombinant AP92 NC antigen using different storage methods. Plate layout is depicted in Figure 40. 146

List of Tables

Table 1: S segment sequences of Crimean Congo haemorrhagic fever virus obtained from GenBank. Isolates are listed according to country of origin. Accession numbers, strain names, sources are also listed. 23

Table 2: Nucleotide sequences of primers designed for amplification of the CCHF NC. Positions are indicated relative to reference isolate CCHF SPU 415/85 Accession No. U88415. Kozak sequence included in the forward primer is underlined. 28

Table 3: Ligation reaction for NC amplicon into pcDNA ™3.1D/V5-His/TOPO® expression vector. 31

Table 4: Restriction enzyme analysis to identify positive transformants using *BamH1 and Not1sites present on the* pcDNA ™3.1D/V5-His/TOPO® vector. 33

Table 5: Primers used for sequencing of CCHFV isolate SPU 92/01 in TOPO pcDNA 3.1D vector. Positions of primers based on TOPO pcDNA 3.1 V5/His/TOPO vector and positions on SPU 415/85 S segment ORF. 34

Table 6: Sequencing reactions using BigDye Terminator Ready Reaction. 34

Table 7: Thermal cycling conditions for sequencing reaction 35

Table 8: Mean distance (d) within groups: The number of base differences per site from averaging over all nucleotide sequence pairs within each group and pairwise distance ranges is shown. 44

Table 9: Mean distance between groups based on nucleotide sequence data: The number of base differences per site from averaging over all sequence pairs between groups is shown. 45

Table 10: Mean Net distance between groups: The number of base differences per site from estimation of net average between groups is shown. 45

Table 11: Mean distance within groups: The number of differences per site from averaging over all amino acid sequence pairs within each group is shown. 48

Table 12: Mean distance between groups based on amino sequence data. The number of base differences per site from averaging over all sequence pairs between groups is shown. 49

Table 13: Mean Net distance between groups: The number of base differences per site from estimation of net average between groups of amino acid sequences is shown. 49

Table 14: Restriction enzyme analysis of *LacZ* gene by double digestion using *BamH1* and *Not1*. 59

Table 15: Fugene:DNA complex ratios for optimization of transfection 61

Table 16: Preparation of 8% resolving gel 65

Table 17: Preparation of 4% stacking gel. 65

Table 18: Ligation reaction for cloning AP92 NC into pColdTF bacterial expression vector 73

Table 19: GoTaq PCR reaction for identification of positive transformants 74

Table 20: Cycling conditions for GoTaq PCR reactions 74

Table 21: Fairbanks Staining solutions 76

Table 22: FuGene 6 reagent transfection optimization: transfection ratios, type of cell line used and results of β galactosidase assay 82

Table 23: Positive panel of sera from CCHF survivors in South Africa 102

Table 24: SPU 415/85 recombinant antigen tested against negative panel of sera. Optical density (OD) ranges and positive percentage (PP) ranges are shown. 109

Table 25: Calculation of cut off value 109

Table 27: AP92 recombinant antigen tested against panel of sera from CCHF survivors. Optical density (OD) ranges and positive percentage (PP) ranges are shown. 111

Table 28: Calculation of cut off value for AP92 antigen ELISA 111

Table 26: Antibody titers of CCHFV NC. Serum samples were diluted 1:100 to 1:800 to determine titer of each serum sample, expressed as reciprocal. 113

Abbreviations and acronyms

µg/ml – micrograms per milliliter

µl - microliter

ABTS - 2,2'-Azino-di-ethyl-benzothiazoline-sulfonic acid peroxidase substrate

Amp - ampicillin

ATCC - American Type Culture Collection,

bp – basepair(s)

CAI - codon adaptation index

CCHF - Crimean-Congo haemorrhagic fever

CCHFV - Crimean-Congo haemorrhagic fever virus (CCHFV)

cfu - colony forming units

CMV - human cytomegalovirus

CPE - cytopathic effects

dH_2O – distilled water

DIC - disseminated intravascular coagulopathy (DIC)

DMEM - Dulbecco's Minimum Essential Medium with L-glutamine

DMSO – dimethyl sulfoxide

DNA – deoxyribonucleic acid

ELISA – Enzyme-linked immunosorbent assay

EMEM - Eagle's Minimum Essential Media with L-glutamine (

ENC - effective number of codons

ER - endoplasmic reticulum

FBS - foetal bovine serum

FITC flourescein isothiocyanate

H_2O - water

HF - High Fidelity

HRP - horse radish peroxidase

IFA - immunofluorescence assays

IgG – Immunoglobulin G

IgM – Immunoglobulin M

IMAC - immobilized metal ion affinity chromatography

IPTG - isopropyl-β-galactosidase

kb - kilobases

kDa – kilo daltons

LB - Luria Bertani

LEW - Lysis-Equilibration-Wash

mAbs - monoclonal antibodies

MCS - multiple cloning site

MEGA - Molecular Evolutionary Genetics Analysis (MEGA)

min - minutes

mM – milli molar

M-MLV - Moloney Murine Leukemia virus

mRNA – messenger RNA

NC - nucleocapsid

NEAA - non-essential amino acids

O/N – overnight

ORF - open reading frame

OTU - operational taxonomic units

PCR – polymerase chain reaction

pmol – picomolar

PVDF - microporous polyvinylidene diflouride

RNA – deoxyribonucleic acid

RT – reverse transcriptase

SD - standard deviation

SDS PAGE- sodium dodecyl sulphate polyacrylamide gel electrophoresis

SOC - super optimal broth with catabolite repression

TAE - Tris-acetate-EDTA

TBS - Tris buffered saline

Tm – melting temperature

tRNA – transcriptor RNA

U/µl – units per microliter

Appendix 4: Abstract of presentation at Faculty of Health Sciences, University of the Free State Faculty Forum August 2012

Understanding antigenic relationships based on global diversity of Crimean-Congo Haemorrhagic Fever nucleoproteins

Authors: A Rangunwala, F J Burt
Department of Medical Microbiology and Virology

Introduction and aim:

Crimean Congo haemorrhagic fever virus (CCHFV) is a member of the Nairovirus genus belonging to the family Bunyaviridae, which consists of diverse RNA viruses. The single stranded, negative sense, segmented RNA genome of CCHFV consists of a small (S) medium (M) and large (L) segment. The L segment encodes the viral polymerase, the M segment encodes precursors for glycoproteins and the S segment encodes the viral nucleoprotein. In this study global diversity of CCHF nucleoproteins from Asia, the Middle East, Africa and Greece was investigated; and a South African isolate that had not been sequenced previously was included.

Methodology:

Sequence data for the open reading frame (ORF) the S segment of 45 geographically distinct isolates were retrieved from GenBank. The ORF of a South African isolate was amplified by PCR and the DNA sequence of the amplicon was determined. Neighbour-joining phylogenetic trees were constructed to study the nucleotide and amino acid diversity amongst worldwide isolates. Between groups diversity, within groups diversity and pairwise distances were calculated.

Results:

CCHFV isolates formed six distinct groups. Grouping was according to geographical location. Overall the nucleotide diversity and amino acid diversity ranged from 0% to 18% and 0% to 8.7% respectively. Within groups the maximum diversity was shown in Group IV with Asian and Middle Eastern strains differing by 6.2% for nucleotide data but only 1.7% for amino acid sequences. Most of the nucleotide changes were synonymous.

Conclusion:

The majority of the nucleotide changes were conserved with 91.3% to 100% similarity of predicted amino acid sequences overall. This could explain the antigenic similarity of CCHFV isolates worldwide despite the high nucleotide diversity.

Appendix 5: Ethics approval and publications

Ethics approval was obtained (ETOVS 152/06) under a related study and renewed for inclusion of this study.

An article will be submitted to *Virus Research* describing serological cross reactivity between geographically distinct CCHFV isolates.

Rangunwala A., Burt F. J. Serological cross-reactivity between geographically distinct and genetically diverse strains of CCHFV.

Summary

CCHFV is a tick-borne virus that is widely distributed in Africa, Asia, the Middle East and Europe. This member of the Nairovirus genus has the propensity to cause nosocomial infections with a high fatality rate and is endemic in South Africa. Advances in molecular techniques have allowed preparation of safe recombinant antigens that have been used in diagnosis and serosurveillance of CCHFV, a BSL-4 pathogen. The purpose of this study was to: examine the global nucleic acid- and amino acid diversity between isolates worldwide; clone and express a recombinant CCHFV NC from a southern African CCHFV and distantly related Greek CCHFV strain and determine the antigenic diversity between the two isolates. Initially, the aim was to prepare a mammalian expressed recombinant CCHFV NC that has post-translational modifications. South African strain SPU 92/01 S segment ORF was cloned into pcDNA3.1 Directional TOPO expression vector and sequenced for inclusion in the phylogenetic analyses with S segment sequences of 44 other CCHFV isolates worldwide. Nucleotide sequence diversity and amino acid diversity between groups, within groups and pairwise distances were calculated. Isolate SPU 92/01 grouped with other South African isolates and shared 96.6% nucleic acid homology and 99.4% amino acid homology with SPU 415/85. The most diverse strain, AP92, displayed the greatest amino acid difference with SPU415/85 (8.7%). Lower amino acid diversity suggested synonymous changes in basepairs resulting in fewer differences at protein level. A control expression vector pcDNA3.1/V5-His/*LacZ*-TOPO DNA was also prepared for mammalian cell transfection. Transfection efficiency of the control expression vector did not exceed 20% when monitored using β-galactosidase assay. Western Blot analyses did not detect His-tagged proteins from CCHFV NC construct and attempts to use a mammalian expression system were abandoned in favour of a bacterial expression system. A bacterial expression system was adopted for higher yields of protein. A previously expressed codon optimized NC from another closely related South African isolate, SPU 415/85 that was subcloned into pColdTF vector, was expressed in OverExpress BL21 (DE3) *Escherichia coli* cells and purified from the soluble phase. The gene

encoding the NC of AP92 was analyzed using the Rare Codon Analysis Tool and codon optimized for expression in *E. coli* host cells. Cross-reactivity of proteins were tested in an indirect ELISA, using a panel of 14 serum samples from 14 South African survivors of CCHFV. The recombinant antigen designated SPU NC, detected IgG in all the sera from South African survivors of CCHF, while one sample tested negative for IgG against the AP92 NC. Two of the most distantly related isolates, according to the phylogenetic analyses based on nucleotide and amino acid sequences, cross-reacted when tested against the same panel of sera. The Greek isolate NC detected antibody against South African survivors, suggesting possible conserved epitopes.

In addition, the stability of the bacterial expressed recombinant proteins was tested using different storage methods. Lyophilization was required to preserve stability over a four-week period. Longer stability tests will be required for longer storage periods.

The epidemiology and emergence of CCHFV, particularly in Europe, emphasizes the importance of developing standardized reagents for diagnosis and surveillance of the disease. Recombinant antigens are safe to use and do not require BSL-4 facilities. There is a need for safe and reliable diagnostic reagents for CCHFV to increase diagnostic and surveillance capacity. Although the assays described in this study will require further validation with larger numbers of samples than is available in our laboratory, the results suggest that there is sufficient serological cross-reactivity between genetically and geographically distinct CCHFV isolates to allow recombinant NC to detect antibody in sera from geographically distinct locations. Hence a recombinant antigen expressed from a South African NC gene likely has application in other laboratories worldwide.

Key terms: *Crimean-Congo Haemorrhagic Fever, recombinant antigen, protein expression, serological cross-reactivity, phylogenetic analyses*

Opsomming

Krim Kongo Hemorragiese Koors Virus (KKHKV) word deur bosluise oorgedra en is wyd verspreid in Afrika, Asië, die Midde-Ooste en Europa. Hierdie lid van die Nairovirus genus is geneig om nosokomiale infeksies, wat deur hoë sterftes syfers gekenmerk word, te veroorsaak, en is endemies tot Suid-Afrika. Vordering in molekulêre tegnieke het dit moontlik gemaak om veilige rekombinante antigene teen KKHKV, ('n bioveiligheid-4 patogeen) te ontwikkel. Hierdie antigene kan gebruik word in diagnose so wel as serologiese kontroleringstudies van KKHKV. Die doel van hierdie studie was om die globale nukleïensuur en aminosuur diversiteit tussen verskillende wêreldwye stamme te ondersoek, die kloning en uitdrukking van 'n rekombinante KKHKV nukleokapsied (NK) van 'n Suidelike-Afrika stam en 'n verlangs verwante Griekse KKHKV stam, en die bepaling van die antigeniese diversiteit tussen die twee stamme. Daar is aanvanklik beoog om rekombinante KKHKV NK met post-translasionele modifikasies in 'n soogdier uitdrukkingsmodel te produseer. Die S segment oop lesingsraam (ORF) van die Suid Afrikaanse stam (SPU 92/01) is in die pcDNA3.1 Direksionele TOPO uitdrukkingsvektor gekloneer. Die nukleïensuur opeenvolging was bepaal vir gebruik in 'n filogenetiese analise van die S segment DNS opeenvolgings met 44 ander KKHKV stamme wat wêreldwyd voorkom. Die nukleïensuur opeenvolgings- en aminosuur diversiteit tussen groepe, in groepe en paargewys afstande is bereken. Die SPU 92/01 stam het met ander Suid-Afrikaanse stamme gegroepeer en deel 96.6% nukleïensuur homologie en 99.4% aminosuur homologie met die SPU 415/85 stam. Die mees diverse stam, AP92, vertoon die grootste aminosuur verskil met SPU415/85 (8.7%). Laer aminosuur diversiteit is aanduidend van sinonieme verskille in die DNS basispare, wat lei tot minder verskille op proteïenvlak. 'n Kontrole uitdrukkingsvektor pcDNA3.1/V5-His/*Lac*Z-TOPO DNS is ook voorberei vir soogdiersel transfeksies. Die doeltreffendheid van transfeksie van die kontrole uitdrukkingsvektor was nie meer as 20% wanneer dit met β-galaktosidase toets gemonitor is nie. Western Blot analise het nie His-gekoppelde proteïene van die KKHKV NK konstruk opgetel nie. Gevolglik is hierdie soogdiersel uitdrukkingsisteem vervang met 'n bakteriese sisteem. 'n Bakteriese uitdrukkingsisteem is gebruik op grond van 'n hoër

proteïenproduksie in vergelyking met die soogdier uitdrukkingsmodel. 'n Voorheen uitgedrukte kodon-optimiseerde NK van 'n naby verwante Suid-Afrikaanse stam, SPU 415/85 wat in die pColdTF vektor gesubkloneer is, was in OverExpress B21 (DE3) *Escherichia coli* selle uitgedruk en van die oplosbare fase gesuiwer. Die geen wat die NK van AP92 kodeer is geanaliseer met die *Rare Codon Analysis Tool* en kodon-optimiseer vir uitdrukking in *E. coli* gasheerselle. Die kruisreaktiwiteit van die proteïne is in 'n indirekte ELISA teen 'n stel van 14 serum monsters van 14 Suid-Afrikaanse oorlewendes van KKHKV getoets. Die rekombinante antigeen, waarna as SPU NK verwys word, het IgG in al die monsters van Suid-Afrikaanse oorlewendes van KKHKV opgespoor, terwyl een monster negatief getoets het vir IgG teen die AP92 NK. Twee van die mees verlangs verwante stamme, soos voorgestel deur filogenetiese analise, het kruisreaktiwiteit aangetoon wanneer met dieselfde stel van serum monsters getoets is. Die Griekse stam NK het antiliggame teen Suid-Afrikaanse oorlewendes opgespoor, wat die moontlikheid van gekonserveerde epitope aandui.

Die stabiliteit van die bakteries uitgedrukte rekombinante is deur verskeie bergingsmetodes getoets. Liofilisasie was nodig om stabiliteit vir langer as vier weke te preserveer. Langer stabiliteitstoetse word benodig vir langer stoortye.

Die epidemiologie en verskyning van KKHKV, veral in Europa, beklemtoon die belangrikheid om gestandardiseerde reagense vir diagnose en kontrolering van hierdie siekte te ontwikkel. Rekombinante antigene benodig nie BSL-4 fasiliteite nie en is daarom veilig. Daar is 'n behoefte aan veilige en betroubare diagnostiese reagense vir KKHKV om diagnostiese en kontroleringskapasiteit te verbeter. Alhoewel die toetse wat in hierdie studie uiteengesit is addisionele validering benodig, met 'n groter getal monsters as wat daar in die laboratorium beskikbaar is, toon die resultate aan dat daar voldoende serologiese kruisrekatiwiteit tussen genetiese en geografiese CCHFV stamme bestaan. Dit beteken dat rekombinante NC teenliggame in serum vanaf die onderskeie geografiese areas kan opspoor. 'n Rekombinante antigeen wat van 'n Suid-Afrikaanse NC geen uitgedruk is, het dus moontlike toepassings in ander laboratoriums wêreldwyd.