MOLECULAR DETERMINANTS AND BIOLOGICAL CHARACTERISTICS OF HEPATITIS C VIRUS ISOLATES AT DR GEORGE MUKHARI ACADEMIC HOSPITAL, PRETORIA, SOUTH AFRICA

BY

Maemu Petronella GEDEDZHA

2014
MOLECULAR DETERMINANTS AND BIOLOGICAL CHARACTERISTICS OF HEPATITIS C VIRUS ISOLATES AT DR GEORGE MUKHARI ACADEMIC HOSPITAL, PRETORIA, SOUTH AFRICA

By

Maemu Petronella GEDEDZHA

BSc, BSc (Hons), MSc (Med)

A thesis in fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN MEDICAL VIROLOGY

Submitted

In the Department of Virology
School of Pathology and Pre-clinical Sciences
Faculty of Health Sciences
University of Limpopo (Medunsa Campus)

Supervisor:
Dr. Selokela G. Selabe; PhD

Co-supervisor:
Prof. M. Jeffrey Mphahlele; PhD
DECLARATION

I, Maemu Petronella GEDEDZHA, hereby declare that the work presented in this thesis is original and does not incorporate any material previously submitted for the purpose of diploma or degree to any other University. This thesis is being submitted in fulfillment for the requirements of the degree of Doctor of Philosophy (Medical Virology), in the Department of Virology, School of Pathology and Pre-Clinical Sciences, Faculty of Health Sciences, at the Medunsa Campus of the University of Limpopo. Works of other investigators have been cited and dully acknowledged.

.................................................................

Signature of Candidate

.......... Day of .........................................................2014
DEDICATION

This thesis is dedicated to my late father Mr Munannzhi Phineas Gededzha
ACKNOWLEDGEMENTS

First, I would like to thank my supervisors, Dr SG Selabe and Prof MJ Mphahlele for their intelligent supervision, constructive guidance, inspiration and friendship. Thanks for giving me the opportunity to become the best researcher.

To my co-investigator, Dr T Kyaw, thanks for collecting the specimens, helping with clinical information and all the support and guidance you have provided.

To Dr Jason Blackard, for the endless support with bioinformatics analysis, thank you.

To Lizzy, Andrew, Nare and Tsakani, thanks for the words of support and encouragement, and being there for me when I’m stuck and confused. I have learned a lot from each of you for the duration of my PhD.

To Omphile, thank you for proof reading one of the chapters and Edina thanks a lot for proof reading the thesis.

Thank you to the Department of Virology staff and students for the constant support and assistance, and to the Administrative Officers Queen and Waldelie, thanks for helping with everything, but most importantly for being wonderful friends.

I have to thank my family; my mother, Elisa, for being there whenever I needed support and taking care of my son so that I could have time to focus on my studies, “Vha Mme wa nde me vhutshiloni, Ndi a livhuwa”. To my sister Nkhangweleni and my brother Mulutanyi for your understanding, words of encouragement and for helping me get through the frustrations. To my Dad Munannzhi, although you are no longer with us, thank you for always encouraging me to study since I was 2 years old until my first year of PhD. You always made me laugh whenever I was stressed at school, thank you.
To my wonderful husband Nathaniel without whom I may not have successfully completed this PhD. I am so grateful for your love and support, “Ndi a livhuwa Ishe”. To my son Rudzani thank you for understanding that Mom is busy at school.

Finally, this study would not have been possible without funding from University of Limpopo, Medunsa campus, Medical Research Foundation, National Research Foundation, Stella and Paul Lowenstein Trust, and National Health and Laboratory Service Research Trust.

Thank you to everyone for your help and support. I am eternally grateful to you all.
ABSTRACT

Introduction
Globally, an estimated 150 million people have been infected with hepatitis C virus (HCV). HCV is the major cause of cirrhosis and liver cancer and one of the most common indications for liver transplantation (Simmonds, 2004). In South Africa, the prevalence of HCV is estimated at less than 2% in the general population and data for human immunodeficiency syndrome (HIV) and HCV co-infection is scarce. Additionally, there is paucity of data about HCV diversity. Few studies have investigated the genetic diversity of HCV in South Africa and focused primarily on the 5’ untranslated region (UTR). There is one study that reported genetic diversity on NS5B region. The aim of this study was to explore molecular determinants and biological characteristics of HCV isolates circulating in patients admitted to Dr George Mukhari Academic Hospital (DGMAH) by: (1) investigating HCV co-infection in HIV-positive patients enrolling for HAART (2) investigating the distribution of HCV genotypes in HIV positive and Negative patients (3) characterizing near full-length genome of HCV genotype 5 strains (4) predicting T-cell epitopes in HCV genotype 5a sequences, and (5) characterizing genotype 5a E1 and E2 sequences.

Methods
To investigate HCV co-infection in HIV-positive patients enrolling for HAART, a total of 653 serum samples collected from 2004 to 2006 were screened for anti-HCV using AxSYM and PCR targeting the 5’UTR. To investigate the distribution of HCV genotypes in HIV positive and HIV negative, a total of 78 serum samples from 71 patients (only one was from the HIV positive group) who were anti-HCV positive was employed in a polymerase chain reaction (PCR) assay to amplify HCV RNA in the 5’UTR, C/E1, and NS5B regions. To characterise near full-length genome of HCV genotype 5a strains, near full-length sequences were amplified from six HCV genotype 5a samples. All PCR products were sequenced and those with interesting results were cloned into P-Gem T-easy vector. Sequence analysis was performed with Chromas pro, Clustal X, MAFFT, BioEdit, Simplot, MEGA and BEAST programs. To predict
T-cell epitopes in HCV genotype 5a sequences a suite of online programmes (Propred, Propred 1, VaxiJen and IEDB database) were used to predict the epitopes. Finally to characterise genotype 5a E1 and E2 sequences, the E1 and E2 regions were amplified and sequenced and conserved epitopes and peptides predicted in all two regions using a suite of online programmes (ABCpred, VaxiJen, IEDB database, DataMonkey, NetGlyc, Protparam, GRAVY and NetPhos).

**Results**

HCV serological results in HIV positive patients indicated that only 1.2% (8/653) of samples was positive for anti-HCV and HCV RNA was detected in only one of the 8 patients. The HCV genotype distribution in the studied population was as follows: 59.2% (42/71) were genotype 5, 18.3% (13/71) were genotype 1, 14.1% (10/71) were genotype 4, and 4.2% (3/71) were genotype 3. Three of 71 (4.2%) individuals were infected with mixed genotypes based on the 5'UTR. Apart from major genotypes there was evidence of intragenotypic recombination of 4l/4q. Genotype results were not stratified according to HIV status, since only one strain was from the HIV positive group. Phylogenetic analysis of the six near full-length genome sequences revealed that all genotype 5 sequences formed a close cluster with high bootstrap support. Bayesian analysis of the E1 region was used to estimate the substitution rate and the time to the most recent common ancestor (tMRCA). The tMRCA for HCV genotype 5a was estimated at 114-134 years before the last sampling date. Conserved T-cell epitopes that recognise both major histocompatibility complexes (MHC) I and II were predicted using the near full-length sequences. A total of 24 and 77 antigenic epitopes that recognize HLA I and HLA II respectively were predicted. For conservation analysis 8 and 31 predicted epitopes were conserved in different genotypes for HLA-I and HLA-II alleles respectively. For the E1 and E2 sequences, a total of 26 B cell epitopes that were antigenic were predicted in both the E1 and E2 regions and more than 80% of the predicted epitopes were conserved in genotype 5a sequences. Three and four conserved peptides of 8-16 amino acids (aa) were also derived from the E1 and E2
regions respectively. Only one of the peptides had N-linked glycosylation post-translational modification although at a low probability.

**Conclusion**

This study shows that the majority of HIV/AIDS patients initiating HAART at DGMAH have low exposure to, or active HCV infection and it does not appear that HIV/AIDS is a risk factor for increased detection of HCV co-infection. Genotyping results indicated that genotype 5 is still the most predominant genotype, and genotypes 1 and 4 are encountered in a significant proportion in our population including the evidence of intra-genotypic recombination in South Africa. This study also provided six near full-length nucleotide sequences of HCV genotype 5a strains and also significant insights for understanding the diversity of genotype 5a in the South African population. Finally, the study provided antigenic B- and T-cell epitopes as well conserved peptides that can be useful for designing entry inhibitors and vaccines that will be able to cover the global population.
RESEARCH OUTPUT EMANATING FROM THIS STUDY

Publications


Presentations at Scientific Meetings

Gededzha MP, Selabe SG and Mphahlele MJ. Optimization and amplification of Hepatitis C virus full genome from South African patients. University of Limpopo Faculty of Health Sciences Research Day. University of Limpopo, MEDUNSA, 11-12 August 2010 (Poster).

**Gededzha MP**, Kyaw T, Selabe SG and Mphahlele MJ. Genetic diversity of HCV from infected patients at the Dr George Mukhari Hospital. University of Limpopo Faculty of Health Sciences Research Day. University of Limpopo, MEDUNSA, 11-12 August 2010 (Oral).


**Gededzha MP**, Selabe SG, Blackard JT, Kyaw T and Mphahlele MJ. Near full-length genome analysis of HCV genotype 5 strains from South Africa.
University of Limpopo, Faculty of Health Sciences Research Day. University of Limpopo, MEDUNSA, 20-21 August 2013 (Poster).
# TABLE OF CONTENTS

DECLARATION ...................................................................................................................... ii  
DEDICATION ........................................................................................................................ iii  
ACKNOWLEDGEMENTS ...................................................................................................... iv  
ABSTRACT ............................................................................................................................. vi  
RESEARCH OUTPUT EMANATING FROM THIS STUDY ....................................................... ix  
  Publications ........................................................................................................................... ix  
  Presentations at Scientific Meetings .................................................................................... ix  

| TABLE OF CONTENTS ......................................................................................................... xii |
| LIST OF ABBREVIATIONS .................................................................................................. xix |

1. RESEARCH PROPOSAL AND THESIS ORGANISATION .................................................... 2  
  1.1 Proposal .......................................................................................................................... 2  
    1.1.1 Background information .............................................................................................. 2  
    1.1.2 Study problem ............................................................................................................ 4  
    1.1.3 Overall aim ................................................................................................................ 5  
    1.1.4 Broad objectives ......................................................................................................... 5  
  1.2 Thesis organization ......................................................................................................... 8  

2 LITERATURE REVIEW ON HEPATITIS C VIRUS ............................................................ 11  
  2.1 HCV Classification ......................................................................................................... 11  
  2.2 Genome organization .................................................................................................... 11  
    2.2.1 The 5’ UTR .................................................................................................................. 12  
    2.2.2 Core protein ............................................................................................................... 14  
    2.2.3 The E1 and E2 proteins .............................................................................................. 14  
    2.2.4 The p7 protein .......................................................................................................... 15  
    2.2.5 The NS2 protein ....................................................................................................... 15  
    2.2.6 The NS3 protein ....................................................................................................... 16  
    2.2.7 The NS4A protein .................................................................................................... 16  
    2.2.8 The NS4B protein .................................................................................................... 16  
    2.2.9 The NS5A protein .................................................................................................... 17  

xii
2.2.10 The NS5B protein .................................................................................................. 17
2.2.11 The 3'UTR .......................................................................................................... 18
2.2.12 The ARF protein ................................................................................................ 18
2.3 Epidemiology ........................................................................................................... 19
  2.3.1 Transmission ..................................................................................................... 19
  2.3.2 Prevalence ......................................................................................................... 19
  2.3.3 HCV genotypes and subtypes ......................................................................... 20
2.4 The life cycle of HCV .............................................................................................. 21
  2.4.1 Entry .................................................................................................................. 22
  2.4.2 Translation and polyprocessing ....................................................................... 23
  2.4.3 Replication ........................................................................................................ 24
  2.4.4 Viral assembly and release .............................................................................. 25
2.5 Disease profile ........................................................................................................ 25
  2.5.1 Acute infection ................................................................................................. 26
  2.5.2 Chronic infection ............................................................................................. 27
2.6 Pathogenesis ........................................................................................................... 28
2.7 Laboratory diagnosis ............................................................................................... 29
  2.7.1 Serologic assays ............................................................................................... 29
  2.7.2 Molecular assays .............................................................................................. 30
2.8 Prevention and control ........................................................................................... 31
2.9 Treatment ................................................................................................................ 32
2.10 HIV/HCV co-infection .......................................................................................... 34
  2.10.1 Burden of HIV and HCV co-infections Worldwide ........................................ 34
  2.10.2 Burden of HIV and HCV co-infections in Africa ............................................ 34
  2.10.3 Impact of HCV infection on HIV disease progression .................................... 36
  2.10.4 Impact of HIV infection on HCV disease progression .................................... 36
  2.10.5 Management of HIV/HCV co-infected patients ............................................. 37
2.10.6 Transmission of HCV in HIV positive patients in Africa

3 Materials and Methods

3.1 Ethical Approval

3.2 Study Population

3.3 AxSYM assay

3.3.1 Principle of the assay

3.3.2 Interpretation of results

3.4 HCV RNA Extraction

3.5 HCV cDNA synthesis

3.6 HCV 5'UTR PCR

3.7 HCV C/E1 PCR

3.8 HCV NS5B PCR

3.9 HCV full genome PCR

3.9.1 Fragment A

3.9.2 Fragment B

3.9.3 Fragment C

3.9.4 Fragment D

3.9.5 Fragment E

3.9.6 Fragment F

3.9.7 Fragment G

3.9.8 Fragment H

3.9.9 Fragment I

3.9.10 Fragment J

3.10 HCV E1 and E2 PCR

3.11 Detection of the HCV PCR product

3.12 HCV gene cloning

3.13 Sequencing of the HCV PCR products
3.14 Phylogenetic analysis of HCV sequences
3.14.1 Alignment of sequence data
3.14.2 Model testing
3.14.3 Tree construction
3.15 Bayesian coalescent analysis
3.16 Prediction of T-cell epitopes
3.17 Prediction of B cell epitopes
3.18 Antigenecity of the epitopes
3.19 Epitope conservancy analysis
3.20 Validation of predicted epitopes
3.21 Immune selection pressure
3.22 Analysis of N-linked glycosylation sites
3.23 Peptide design
3.24 GenBank submission

4 INVESTIGATING HCV CO-INFECTION IN HIV-POSITIVE PATIENTS ENROLLING FOR HAART AT DGMAH.
4.1 Introduction
4.2 Methods
4.3 Results
4.4 Discussion

5 INVESTIGATING THE DISTRIBUTION OF HCV GENOTYPES IN HIV POSITIVE AND HIV NEGATIVE PATIENTS AT DGMAH.
5.1 Introduction
5.2 Methods
5.2.1 Study population
5.2.2 Reverse transcriptase PCR and cloning
5.2.3 Sequencing of the PCR products
5.2.4 Phylogenetic tree construction
7.2.4 Epitope conservancy analysis .................................................................94
7.2.5 Validation of predicted epitopes .................................................................94
7.3 Results ..............................................................................................................94
  7.3.1 T-cell epitopes prediction ........................................................................94
  7.3.2 Epitope binding affinity to common South African HLA-I and HLA-II alleles 95
  7.3.3 Validation of epitopes ...............................................................................96
7.4 Discussion .........................................................................................................97
8 CHARACTERISATION OF HCV GENOTYPE 5a ENVELOPE PROTEINS: IMPLICATIONS FOR VACCINE DEVELOPMENT AND THERAPEUTIC ENTRY TARGET ............................................................................................................................101
  8.1 Introduction .....................................................................................................101
  8.2 Methods ..........................................................................................................102
    8.2.1 Amplification and Sequence retrieval .....................................................102
    8.2.2 Prediction of B-cell epitopes .................................................................102
    8.2.3 Antigenecity of the epitopes .................................................................103
    8.2.4 Epitope conservancy analysis ...............................................................103
    8.2.5 Immune selection pressure ....................................................................103
    8.2.6 Analysis of N-linked glycosylation sites ...............................................103
    8.2.7 Peptide design ........................................................................................103
  8.3 Results .............................................................................................................104
    8.3.1 B-cell epitopes prediction .......................................................................104
    8.3.2 Determination of positive selection .......................................................104
    8.3.3 Analysis of E1 and E2 N-linked glycosylation .......................................105
    8.3.4 Peptide design ........................................................................................107
  8.4 Discussion .......................................................................................................109
9 COMBINED DISCUSSION, CONCLUSION AND RECOMMENDATIONS ......112
  9.1 Discussion ......................................................................................................112
  9.2 Conclusion .....................................................................................................115
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.3</td>
<td>Recommendations</td>
<td>115</td>
</tr>
<tr>
<td>9.4</td>
<td>Limitations of the study</td>
<td>116</td>
</tr>
<tr>
<td>10</td>
<td>REFERENCES</td>
<td>118</td>
</tr>
<tr>
<td>11</td>
<td>APPENDIX</td>
<td>176</td>
</tr>
<tr>
<td>11.1</td>
<td>Tables</td>
<td>176</td>
</tr>
<tr>
<td>11.2</td>
<td>Publications</td>
<td>200</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

aa Amino acid
ALT Alanine aminotransferase
Anti-HCV Antibody to hepatitis C virus
apoE apolipoprotein E Binding
ARF Alternative reading frame
AST Aspartate aminotransferase
BEAST Bayesian Evolutionary Analysis by Sampling Trees
BOC Boceprevir
bp Base pair
BSP Bayesian skyline plots
CD81 Cluster of Differentiation 81
CIA Chemiluminescence immunoassays
DAA Direct-acting antiviral
DDB1 Damaged DNA-binding
DGMAH Dr George Mukhari Academic Hospital
dN Non-synonymous
dS Synonymous
E1 Envelope protein 1
E2 Envelope protein 2
EIA Enzyme immunoassays
ER Endoplasmic reticulum
ESS Effective sample size
EVR Early virological response
FDA Food and Drug Administration
FEL Fixed effects likelihood
GRAVY  Grand average of hydropathicity
GTR    Generalized time reversible
HAART  Highly Active Antiretroviral Therapy
HBV    Hepatitis B virus
HCC    Hepatocellular carcinoma
HCV    Hepatitis C virus
HCVcc  Cell culture-derived HCV particles
HCVpp  HCV pseudotyped particles
HDL    high density lipoprotein
HLA    Human leukocyte antigen
HPD    Highest posterior density
HIV    Human immunodeficiency virus
HVR    Hypervariable region
IDU    Intravenous drug use
IFN    Interferon
IU/L   International units per litre
IRES   Internal ribosome entry site
ISDR   Interferon sensitivity-determining region
kDa    Kilodalton
LB     Luria Bertani
LDH    Lactase dehydrogenase
LDL    low-density lipoprotein
MCMC   Markov chain Monte Carlo
MEIA   Microparticle Enzyme Immunoassays
MHC    Major histocompatibility complexes
mRNA   messenger RNA
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>MREC</td>
<td>Medunsa Research and Ethics Committee</td>
</tr>
<tr>
<td>nm</td>
<td>Nanometer</td>
</tr>
<tr>
<td>NHLS</td>
<td>National Health Laboratory Service</td>
</tr>
<tr>
<td>NK</td>
<td>Natural killer</td>
</tr>
<tr>
<td>NS</td>
<td>Non structural</td>
</tr>
<tr>
<td>NT</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>Peg-IFN</td>
<td>Pegylated interferon</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RBV</td>
<td>Ribavirin</td>
</tr>
<tr>
<td>RdRp</td>
<td>RNA-dependent RNA polymerase</td>
</tr>
<tr>
<td>RIBA</td>
<td>Recombinant immunoblot assay</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>RVR</td>
<td>Rapid virological response</td>
</tr>
<tr>
<td>S</td>
<td>Structural</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SEM</td>
<td>standard error of mean</td>
</tr>
<tr>
<td>S/CO</td>
<td>ratio of the sample rate to the cut off rate for each sample and control</td>
</tr>
<tr>
<td>SNARE</td>
<td>soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor</td>
</tr>
<tr>
<td>S.O.C</td>
<td>Super Optimal broth with Catabolite repression</td>
</tr>
<tr>
<td>SRB1</td>
<td>Scavenger receptor class B member 1</td>
</tr>
<tr>
<td>SVR</td>
<td>Sustained virological response</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate-EDTA</td>
</tr>
<tr>
<td>tMRCA</td>
<td>Time to the most recent common ancestor</td>
</tr>
<tr>
<td>TVR</td>
<td>Telaprevir</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>V</td>
<td>Voltages</td>
</tr>
<tr>
<td>VAP</td>
<td>Vesicle-associated membrane protein-associated</td>
</tr>
<tr>
<td>VLDL</td>
<td>Very-low-density lipoprotein</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>Nucleotide base</td>
<td>Standard abbreviation</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Adenine</td>
<td>A</td>
</tr>
<tr>
<td>Cytosine</td>
<td>C</td>
</tr>
<tr>
<td>Thymine</td>
<td>T</td>
</tr>
<tr>
<td>Guanine</td>
<td>G</td>
</tr>
<tr>
<td>Uracil</td>
<td>U</td>
</tr>
<tr>
<td>Amino acid</td>
<td>Standard abbreviation</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>I</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>F</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
</tr>
<tr>
<td>Cysteine</td>
<td>C</td>
</tr>
<tr>
<td>Alanine</td>
<td>A</td>
</tr>
<tr>
<td>Glycine</td>
<td>G</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>W</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>E</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>D</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
</tr>
</tbody>
</table>
CHAPTER ONE

RESEARCH PROPOSAL AND THESIS ORGANISATION

This chapter provides the introductory background to HCV, study problem, aim and objectives and thesis organization.
1. RESEARCH PROPOSAL AND THESIS ORGANISATION

1.1 Proposal

1.1.1 Background information

Globally, an estimated 150 million people have been chronically infected with HCV. HCV is the major cause of cirrhosis and liver cancer and one of the most common indications for liver transplantation (Simmonds, 2004). The World Health Organization (WHO) estimated that Africa has the highest prevalence of 5.3% followed by Eastern Mediterranean with 4.6% while America and Europe have been reported to have low prevalence rates of 1.7% and 1% respectively (WHO, 1999). In South Africa, there is a low prevalence ranging from 0.16% to 1.8% (Ellis et al., 1990; Basetse et al., 1993; Soni et al., 1993) but this is higher in high-risk individuals, e.g. 39.4% in haemophiliacs and 4.8% in chronic dialysis patients (Soni et al., 1993). HCV is common in HIV infected people due to shared routes of transmission. Both HIV and HCV are blood-borne pathogens and can be transmitted by direct contact of infected blood and blood products (Newcomb-Fernandez, 2004). HCV is more infectious than HIV but sexual transmission of HCV is less common than with HIV. Of the 40 million people who are infected with HIV worldwide, an estimated 4 to 5 million are co-infected with HCV (Alter, 2006). In developed countries, the prevalence of HIV/HCV is well understood and one study reported the prevalence of HIV/HCV co-infection to be 33% (Dore and Sasadeusz, 2003).

HCV, a member of the Hepacivirus in the Flaviviridae family, possesses a single stranded, positive-sense RNA genome of approximately 9000 nucleotides (nt) long. HCV RNA has a long open reading frame encoding a polyprotein flanked by relatively short 5' and 3' UTR ribonucleic acid (RNA) segments (Brass et al., 2006). The polyprotein is divided into structural (S) and non-structural (NS) proteins (Tang et al., 2009). The S proteins consist of the core protein and envelope glycoproteins: E1 and E2 (Hijikata et al., 1991), followed by the p7 protein (Lin et al., 1994). The NS protein consists of NS2, NS3, NS4A, NS4B, NS5A and NS5B (Dubuisson, 2007).
HCV displays a high degree of diversity both at the nt and aa level due to its error-prone RNA-dependent RNA polymerase (RdRp), which generates an estimated mutation rate of $10^{-5}$ per replication cycles (Neumann et al., 1998). HCV is classified into at least six major genotypes (genotypes 1 to 6), whose nucleotide sequences differ by 31-33% (Simmonds et al., 2005). A novel genotype 7 has been described as well (Murphy et al., 2007a). HCV genotypes can be divided further into subtypes which differ by 10–30%. Moreover, distinct isolates from the same subtype may differ in nucleotide sequence by 5–15% (Abdo and Lee, 2004). HCV genotypes are frequently assigned by phylogenetic analysis of the 5'UTR, core/E1, NS5B, and/or complete genome sequences (Simmonds et al., 2005). Genotype 1a is common in the United States and Northern Europe. Genotype 1b has a worldwide distribution and is often found to be the most common genotype. Genotypes 2a and 2b are also found worldwide and are relatively common in North America, Europe, and Japan (Simmonds, 2004). Genotype 3 is found in India, United States and Europe. Genotype 4a is most common in North Africa and the Middle East. Genotype 6a occurs in Hong Kong and Southeast Asia, while genotypes 5 and 7 are found in South Africa (Hoofnagel, 2002) and the Democratic Republic of Congo (Murphy et al., 2007a), respectively.

The wide genetic diversity of HCV makes the development of effective vaccines and therapeutic agents difficult (Pawlotsky, 2006). There is no licensed vaccine at the moment, however, several HCV vaccine candidates are in clinical trials and include recombinant proteins, synthetic peptides and DNA-based vaccines (Torresi et al., 2011). Current utilised treatment such as pegylated interferon (Peg-IFN) and ribavirin (RBV) can achieve virologic response rates that range from 41-80% (Manns et al., 2001; Fried et al., 2002). However, virologic response rates are lower for genotypes 1 and 4 compared to genotypes 2 and 3 (Ghany et al., 2009). Since large-scale clinical trials do not include a sufficient number of patients with genotype 5, the optimal treatment schedule for genotype 5 is largely unknown. Nonetheless, a few studies have suggested that treatment response of
genotype 5 is better than for genotype 1 and is very similar to response rates for genotypes 2 and 3 (Legrand-Abravanel et al., 2004). Despite the lack of information, these patients are treated for 48 weeks (Botha et al., 2010). Thus, the determination of genotypes is important for choosing the optimal schedule for treatment. Recently, two NS3 protease inhibitors: telaprevir (TVR) and boceprevir (BOC) have been approved by the Food and Drug Administration (FDA) for the treatment of HCV genotype 1 infections (Bacon and Khalid, 2012). In South Africa, the prevalence of HCV is estimated at less than 2% in the general population and data for HIV/HCV co-infection is scarce. As a result, much remains unknown about HCV prevalence in HIV infected patients. HCV is known to be common in HIV infected people due to shared routes of transmission. However prevalence studies from Africa are very rare. Furthermore, liver disease and hepatotoxicities associated with HCV are the cause of mortality and morbidity in HIV/HCV co-infected patients on highly active antiretroviral therapy (HAART) based on studies from Europe. HCV isolates show four levels of genomic variations: genotypes, subtypes, isolates, and quasispecies. The wide genetic diversity of HCV makes the development of effective vaccines rather difficult and favours the persistence of the virus in humans, and also has therapeutic implications as the continuous generation and selection of fitter or resistant variants within the quasispecies spectra, can allow the virus to escape control by antiviral drugs (Pawlotsky, 2006).

1.1.2 Study problem

Few studies have investigated the genetic diversity of HCV in South Africa and the results indicated that HCV genotype 5 is predominating over others. HCV antiviral therapy is increasingly becoming available in South African tertiary public hospitals through public-private partnerships. It is postulated that South Africa may harbour a variety of unexplored circulating HCV subtypes and quasispecies which could favour persistence of the virus in the host due to differential activity of HCV proteins that could alter the rate of HCV
replication or pathogenicity, or negatively impact on treatment by allowing the virus to escape control by antiviral drugs, or could provide novel genetic materials for inclusion in designing future HCV vaccine candidates. Thus, continuous characterisation of circulating genotypes will aid in monitoring new strains and provide directions on HCV therapy and management. Finally, there is paucity of data about HCV genetic diversity in South Africa as many studies have focused on the 5’ UTR. Recently one study has reported on NS5B region diversity, but less emphasis has been put on other regions e.g. envelope and the full-length genomes. For example, of nearly 200 complete HCV genomes currently available in GenBank, there are only two full genomes of HCV genotype 5 available, one from South Africa.

1.1.3 Overall aim
The overall aim of this study was to explore the molecular determinants and biological characteristics of HCV isolates circulating in patients attending DGMAH, Pretoria, South Africa.

1.1.4 Broad objectives
a. To investigate HCV co-infection in HIV-positive patients enrolling for HAART at DGMAH.

b. To investigate the distribution of HCV genotypes in HIV positive and HIV negative patients at DGMAH.

c. To analyse near full-length genome of HCV genotype 5 strains.

d. To predict immunological T-cell epitopes of HCV genotype 5a sequences.

e. To characterise genotype 5a envelope proteins.
1.1.4.1 Investigating HCV co-infection in HIV-positive patients enrolling for HAART at DGMAH.

Rationale:
Co-infection with HCV and HIV is common due to shared routes of transmission – via blood and blood products and sharing of needles for injecting drugs. Approximately 4 - 5 million people globally are co-infected with HCV and HIV (Alter, 2006). The introduction of HAART dramatically improved the management of HIV patients. However, co-infections with HCV and hepatitis B virus (HBV) remain a major problem. Patients with HIV/HCV co-infection have less immune reconstitution than patients with HIV infection alone (Miller et al., 2005). South Africa has scaled up HAART for treatment of HIV/AIDS in the public health sector since April 2003. However, few data exist on the burden of HCV prevalence in HIV patients. One study demonstrated a low prevalence of 1.9% (Amin et al., 2004), and another, a high prevalence of 13.4% (Parboosing et al., 2008). Therefore, the objective of this chapter was to investigate the burden of HCV co-infections in HIV-positive patients before initiation of HAART.

1.1.4.2 Investigating the distribution of HCV genotypes in HIV positive and HIV negative patients at DGMAH.

Rationale:
HCV genotyping is important for studies regarding HCV epidemiology, evolution and treatment response (Simmonds et al., 2005). HCV antiviral treatment is increasingly available in South African public hospitals through public-private partnerships. Most importantly, the response to HCV treatment is partially dependent on the infecting genotype. HCV treatment such as Peg-IFN and RBV can achieve virologic response rates that range from 41-80% (Manns et al., 2001; Fried et al., 2002). However, virologic response rates are lower for genotypes 1 and 4 compared to genotypes 2 and 3 (Ghany et al., 2009). There have been limited HCV genotyping studies in South Africa. Therefore, the objective of this chapter was to investigate HCV genotypic diversity and recombination based on comparative analysis of the 5’UTR,
C/E1, and NS5B regions in order to understand the change in frequency in circulating genotypes in the country and this will help for policy decision on the introduction of treatment to different populations in South Africa.

1.1.4.3 Near full-length genome analysis of HCV genotype 5 strains.

Rationale:
Genotype 5 is the most predominant genotype in South Africa with increasing emergence in Europe and North America (Antaki et al., 2009). Of the nearly 200 full-length genome sequences available in GenBank, only two are genotype 5, one coming from South Africa and one from the United Kingdom (Chamberlain et al., 1997; Bukh et al., 1998). Very few molecular studies have been performed to sequence and characterise genotype 5 and as such, genotypes 5 variants are not well documented. Pang et al. (2009) determined the evolutionary branching order of 345 full-length genomes from major genotypes in correlation to their respective responses to therapy. The lack of genotype 5 sequences led to the inability to determine a definitive relationship of genotype 5 to other genotypes (Pang et al., 2009). The objective of this chapter was to characterise HCV genotype 5 near full-length genome sequences that can be used as references to design efficient vaccines and for the development of new antiviral agents.

1.1.4.4 Prediction of immunological T-cell epitopes of HCV genotype 5a.

Rationale
Currently, there is no vaccine against hepatitis C and the current treatment does not clear the infection in all patients. In addition treatment is still not affordable in most developing countries. The development of an effective HCV vaccine is a public health priority and requires understanding of the immune response to HCV. A successful HCV vaccine should stimulate the production of neutralizing antibodies and intensify HCV-specific T cell responses (Helle et al., 2011). Because of the characteristic high diversity of HCV, preventative vaccines will have to overcome significant viral antigenic diversity. It is important to have knowledge of the conserved epitope regions
to improve the formulations of vaccine candidates that are to be developed in the future to cover different HCV genotypes. The objective of this chapter was to predict conserved T cell-epitopes in HCV proteins, particularly of the poorly characterized genotype 5 group.

1.1.4.5 Characterization of HCV genotype 5a envelope proteins.

Rationale
The envelope proteins, E1 and E2 are type I transmembrane proteins involved in viral entry. The E2 is the main target for neutralizing antibody responses and variations in this region is thought to be related to the maintenance of persistent infection by emerging escape variants and subsequent development of chronic infection (Farci et al., 2000; Manzin et al., 1998). Many patients do not respond to the current available therapy, therefore there is an urgent need to develop an effective HCV vaccine and specific therapeutic drugs. The specific therapeutic drugs that have been approved recently are mainly focused on the NS proteins such as the NS3 protease and the NS5B RdRp. Recently, considerable progress has been made in understanding HCV entry and the development of entry inhibitors (Baldick et al., 2010). Therefore, the objective of this chapter was to characterise genotype 5a E1 and E2 sequences to determine conserved B-cell epitopes and peptides that could be useful targets in the design of vaccines and entry inhibitors.

1.2 Thesis organization
This thesis consists of Eleven chapters. Chapter One is “Research proposal and thesis organization”, Chapter Two is “General literature review” and Chapter Three is “General materials and methods”. Chapters Four to Eight are organised around the five objectives of the study, with each chapter containing its own independent brief background introduction, methods, results and discussion. In brief, Chapter Four is “Investigating the HCV co-infection in HIV-positive patients enrolling for HAART”, Chapter Five is
“Investigation of the distribution of HCV genotypes in HIV positive and negative patients”, Chapter Six is “Characterization of near full-length genome of HCV genotype 5 strains”, Chapter Seven is “Prediction of immunological T-cell epitopes in HCV genotype 5a sequences” and Chapter Eight is “Characterization of HCV genotype 5a envelope proteins”. Where an overlap in methods occurred, the relevant chapter(s) and section(s) have been referred to too. Finally, Chapter Nine combines the final discussion, conclusions, recommendations and limitations, while Chapters Ten and Eleven contain the references and appendices respectively.
CHAPTER TWO

LITERATURE REVIEW ON HEPATITIS C VIRUS

This chapter provides the general literature review to HCV and epidemiology of HIV/HCV co-infection.
2 LITERATURE REVIEW ON HEPATITIS C VIRUS

2.1 HCV Classification

HCV is the only member of the genus Hepacivirus in the family Flaviviridae (Robertson et al., 1998) and is an enveloped virus with a diameter of approximately 50-60 nanometer (nm) (Figure 2.1). Its genome organization differs from the other members of the Flaviviridae (Simmonds, 2004). Other viruses in the family Flaviviridae include Yellow fever virus, West Nile virus and Dengue fever virus (Baier, 1987). HCV was first discovered by cloning a non-A, non-B hepatitis virus isolated from the plasma of a chimpanzee infected with contaminated factor XIII concentrate (Choo et al., 1989). HCV genomic variations can be categorized as genotypes, subtypes, isolates, and quasispecies (reviewed in Hoofnagle, 2002).

Figure 2.1 Structure of Hepatitis C virion (Adapted from www.accessmedicine.com)

2.2 Genome organization

HCV is a single stranded, positive-sense RNA of approximately 9000 nt in length (Choo et al., 1989). The HCV RNA has a long open reading frame
encoding a polyprotein flanked by relatively short 5' and 3' UTR RNA segments (Brass et al., 2006). The polyprotein is divided into S and NS proteins (Tang et al., 2009). The S proteins consist of the core protein that forms a nucleocapsid and envelope glycoproteins; E1 and E2 (Hijikata et al., 1991), followed by the p7 protein (Lin et al., 1994). The NS protein consists of NS2, NS3, NS4A, NS4B, NS5A and NS5B (Dubuisson et al., 2007) which are released after cleavage by HCV proteases NS2-3 and NS3-4A (Figure 2.2). Additionally, the genome contains an overlapping +1 reading frame that leads to the production of an additional protein termed the alternating reading frame (ARF) (Xu et al., 2001).

Figure 2.2 HCV genome organization (top) and its polyprotein processing (bottom) (Adapted from Lauer and Walker, 2001).
Key: UTR-Untranslated region, HVR-Hypervariable region

2.2.1 The 5' UTR

The 5'UTR is a highly conserved region in HCV with 341 nt (Han et al., 1991), and contains four highly structured domains (I to IV) (Bukh et al., 1992) (Figure 2.3). The 5'UTR contains the secondary structure known as internal ribosome entry site (IRES), which directly binds the 40S ribosomal subunit to the messenger RNA (mRNA), and initiates translation in a cap-independent
approach (Wang et al., 1993). The 40 nts of the 5'UTR upstream of the IRES is not essential for translation but is important for RNA replication (Friebe et al., 2001).

Domains II and III are crucial for IRES activity and sequence conservation downstream of the initiator AUG and are crucial for IRES-directed translation (Honda et al., 1996). It has been shown that domain III is the most appropriate site for blocking HCV and thus can be used for designing therapeutic agents (Kikuchi et al., 2005). Short peptide nucleic acids of 6–10 mer molecules which can be used as anti-HCV bind to IRES conserved IIIId or IV loop regions and inhibit HCV translation (Alotte et al., 2008). The 5’UTR is the preferred region for qualitative and quantitative detection because of its conservative nature and sensitivity (Murphy et al., 2007b).

Figure 2:3 Schematic model depicting predicted RNA secondary and tertiary structure within the 5’UTR and the immediate adjacent capsid-coding (Adapted from Honda et al., 1996).
2.2.2 Core protein

The core protein is a 21 Kilodalton (kDa) RNA-binding protein and its primary function is to form the viral nucleocapsid. During HCV polyprotein translation, core protein yields the immature form of the protein (191 aa’s) (Santolini et al., 1994) and the mature form (179 aa’s) after being removed at the C-terminus by a host signal peptidase cleavage (McLauchlan et al., 2002). The mature form of core is a dimeric alpha-helical protein, which behaves as a membrane protein (Boulant et al., 2005). The HCV core is conserved in all HCV genotypes (Bukh et al., 1994; Tokita et al., 1998). HCV core protein could be involved in cell signaling and could suppress host immune responses (Penin, 2004).

2.2.3 The E1 and E2 proteins

The E1 (31 kDa) and E2 (70 kDa) proteins have N-terminal ectodomains of 160 and 334 aa’s respectively and a C-terminal transmembrane domain of 30 aa’s (Chevaliez and Pawlotsky, 2006). The building blocks for the viral envelope are believed to be formed by non-covalent heterodimers from the E1 and E2 proteins (reviewed in Brass et al., 2006; Deleersnyder et al., 1997). These proteins bind to receptors and induce fusion with a host-cell membrane which is the important part in host-cell entry (Bartosch et al., 2003). For maturation and folding, HCV envelope glycoprotein depends on core protein coexpression and glycosylation (Dubuisson and Rice, 1996; Merola et al., 2001). The E1 and E2 proteins contain 6 and 11 glycosylation sites, respectively (Goffard and Dubuisson, 2003).

The E2 sequences contain three hypervariable regions (HVR); HVR1-HVR3 (Weiner et al., 1991; Kato et al., 2001; Troesch et al., 2006). These HVR’s differ by up to 80% between HCV genotypes and subtypes (Penin et al., 2004). The HVR1 consists of the first 27 aa of the E2 sequences (Weiner et al., 1991). The physicochemical properties and conformation of HVR1 residues are highly conserved in different genotypes (Penin et al., 2001).
Mutations occurring within the HVR1 confer immune escape properties to the virus leading to the development of chronic infection (Farci et al., 2000). HVR2 consists of 7 aa which show sequence diversity of up to 100% (Kato et al., 2001). HVR3 is found at residues 431–466 aa (Troesch et al., 2006). The HVR has been proposed to modulate cell surface receptor binding and viral entry (Roccasecca et al., 2003).

In addition, E2 induces apoptosis through a mitochondrial damage-mediated caspase pathway (Chiou et al., 2006). The E1 derived peptide p35 (aa 315–323) (El-Awady et al., 2006), E2-conserved synthetic peptides p37 (aa 517–531) and p38 (aa 412–419), neutralize HCV particles and as a results, they are important components of a candidate peptide vaccine (El-Awady et al., 2009).

### 2.2.4 The p7 protein

The p7 protein (7 kDa) consists of 63 aa lying between the S and NS regions of the virus polyprotein and is highly hydrophobic in nature (Lin et al., 1994). The p7 protein has two transmembrane domains and both the N- and C-termini oriented near the lumen of the endoplasmic reticulum (ER) (Carrere-Kremer et al., 2002). The intraluminal tails of p7 amino- and carboxyl-terminal contains sequences with genotype-specific function (Sakai et al., 2003). Because of its ion channel activity in cellular membranes and its inhibition by amantadine, p7 can be used as a target for development of anti-viral chemotherapy (Griffin et al., 2004).

### 2.2.5 The NS2 protein

The NS2 protein (21 kDa) is the first NS protein encoded by the HCV genome and has been shown to hinder gene transcription from various cellular promoters (Dumoulin et al., 2003). NS2/3 protease is responsible for the NS2 and NS3 proteins’ intramolecular cleavage (Grakoui et al., 1993; Hijikata et
al., 1993). NS2/3 is required to allow the accumulation of sufficient NS3 for RNA replication to take place and is therefore an important step in the viral life cycle (Welbourn et al., 2005).

2.2.6 The NS3 protein
The NS3 protein (67 kDa) consists of the serine protease domain at the N-terminal which is responsible for cleavage within the NS region and a helicase/NTPase domain at the C-terminal (Grakoui et al., 1993; Hijikata et al., 1993). The NS3 helicase-NTPase domain consists of 442 aa and stimulates RNA NTPase activity and RNA binding (reviewed in Penin et al., 2004). Along with its NS4A, co-factor NS3 activates a catalytic mechanism for the posttranslational processing of the HCV NS proteins (Barbato et al., 1999). The NS3/NS4A proteolytic activity is important for viral RNA replication (Bermudez-Aguirre et al., 2009). NS3 serine protease antivirals are currently licensed for the treatment of HCV genotype 1 (Bacon and Khalid, 2012).

2.2.7 The NS4A protein
The NS4A protein (6 kDa) is a small protein of 54 aa which is required as a co-factor for NS3 proteinase (Kato et al., 2001) and is responsible for the membrane association of NS3/NS4 protease (reviewed in Dubuisson, 2007). NS4A expression leads to mitochondria mediated apoptosis (Nomura-Takigawa et al., 2006) and is also involved in viral pathogenesis by affecting cellular functions (Sharma, 2010).

2.2.8 The NS4B protein
The NS4B protein (27 kDa) is a highly hydrophobic NS protein consisting of transmembrane domains. It is confined to an ER derived membranous compartment with both the N- and C-termini thought to be oriented near the cytosol (Hugle et al., 2001; Welsch et al., 2007). For the formation of the HCV
replication complex, the NS4B stimulates a specific membrane alteration (Egger et al., 2002).

2.2.9 The NS5A protein
The NS5A is a membrane-associated phosphoprotein that exists in two forms: a 56 kDa basally phosphorylated form and a 58 kDa hyperphosphorylated form. It contains structural domains I-II. Domain I is composed of the N-terminal subdomain IA which is basic in nature, and a predominantly acidic C-terminal subdomain IB (reviewed in Penin et al., 2004; Dubuisson, 2007). NS5A has a possible role in altering interferon therapy response via the region called the “interferon sensitivity-determining region” (ISDR) by inhibiting the cellular double-stranded RNA activated protein kinase R in domain II (Gale et al., 1997). The C-terminus of Domain III has been identified as the primary determinant for particle formation in the assembly step (Appel et al., 2008) and was shown to be natively unfolded (Hanoulle et al., 2009). NS5A phosphorylation plays an essential role for RNA replication in the viral life cycle (Appel et al., 2005) and as an RNA binding protein. This suggests that it can be used as a target for the development of drugs for HCV infection (Huang et al., 2005).

2.2.10 The NS5B protein
A membrane-associated protein, NS5B (68 kDa) contains 591 aa with 21 hydrophobic aa at its C-terminus which is responsible for membrane anchorage (Ivashkina et al., 2002) and 530 aa at its N-terminus which exhibits the usual “fingers”, “palm” and “thumb” subdomains of all single-chain polymerases (Bressanelli et al., 1999). NS5B acts as RNA dependent RNA polymerase (RdRp), a core enzyme for HCV replication, which synthesizes RNA using an RNA template (reviewed in Dubuisson, 2007). The RdRp has no proof-reading ability and is error prone in nature. HCV NS5B polymerase has emerged as one of the major targets for development of drugs that
inhibits HCV directly, with several drugs in clinical trials (De Fransesco and Migliaccio, 2005; Chase et al., 2009).

2.2.11 The 3'UTR

The 3'UTR consists of distinct regions; a variable region which follows immediately after the termination codon of the HCV polyprotein with a length that ranges from 27 to 70 nt; a poly(U/UC) stretch that consists of a poly(U) stretch and a C(U)n-repeat region and an X region which is a 98nt highly conserved region which forms three stable stem-loop structures (Han et al., 1991; Tanaka et al., 1995; Kolykhalov et al., 1996; Tanaka et al., 1996) (Figure 2.4). The 3'UTR is believed to play an essential role in initiating viral genomic replication (Tanaka et al., 1996). The poly(U–UC) and the X regions are critical for in vivo infectivity of HCV (Yanagi et al., 1999).

![Figure 2.4 Structure of the 3'UTR. The shaded box indicates RNA elements involved in RNA replication (Adapted from Shi and Lai, 2006).](image)

2.2.12 The ARF protein

The ARF or frameshift frame consists of 160 aa formed as a result of a -2/+1 ribosomal frameshift in the polyprotein N terminal core-encoding region (Xu et al., 2001; Varaklioti et al., 2002). The ARF overlaps the core protein sequences and lacks an in-frame AUG start codon (reviewed in Dubuisson,
Expression of the ARF protein of HCV genotype 1a in vitro or in mammalian cells yields a 17 kDa protein (Xu et al., 2001). It has been shown that the ARF is expressed during natural HCV infection due to the development of antibodies against the ARF protein in chronically infected individuals (Cohen et al., 2007).

2.3 Epidemiology

2.3.1 Transmission

HCV transmission occurs mainly through exposure to infected blood and blood products. The risks for transmission include individuals who receive blood transfusion, especially those that did before 1992 (CDC, 2013), intravenous drug use (IDU) (Johnston et al., 2011; Johnston and Corceal, 2013), high risk sexual behaviour (Van de laar et al., 2010; Bradshaw et al., 2013), nosocomial infection (Bronowicki et al., 1997; Mboto et al., 2010), occupational exposure (Saito et al., 2004), organ transplantation, hemodialysis, household exposure, and from mother to child (Chen and Morgan, 2006). Blood transfusion has been reported to be the main factor associated with HCV infection in developing countries as well as traditional scarifications (Adewole et al., 2009; Diarra et al., 2009; Zeba et al., 2011).

2.3.2 Prevalence

Globally an estimated 150 million people have been infected with HCV. The WHO estimated that Africa has the highest prevalence of 5.3% followed by Eastern Mediterranean with 4.6%. America and Europe have been reported to have low prevalence rates of 1.7% and 1% respectively (WHO, 1999). Within the African region, Egypt has the highest prevalence of up to 20%, while countries with intermediate prevalence include Guinea at 5.5%, Burundi at 11.3% and Cameroon at 13.8%. Southern and Eastern African regions have the lowest prevalence of around 1.6% (WHO, 1999; Madhava et al., 2002). In South Africa, rural Africans have higher prevalence of anti-HCV than urban
Africans (Ellis et al., 1990). In a study conducted in 1990, anti-HCV positivity was found in 1.2%, 0.8% and 0.6% of urban Blacks, Asians and Whites respectively (Ellis et al., 1990). Basetse et al (1993) reported 0.64% of anti-HCV in an elderly urban black population from Ga-Rankuwa, Pretoria. In 1993, 20.3% of HCV RNA was detected in 128 Southern African patients with antibodies to HCV (Bukh et al., 1993).

2.3.3 HCV genotypes and subtypes

HCV is classified into six major genotypes (genotypes 1 to 6), whose nucleotide sequences differ by 31-33% (Simmonds et al., 2005). In 2007, a novel genotype 7 was described as well (Murphy et al., 2007a). HCV genotypes can be divided further into subtypes which differ by 10–30%. Moreover, distinct isolates from the same subtype may differ in nucleotide sequence by 5–15% (Abdo and Lee, 2004). Genotype 1a is common in the United States and Northern Europe. Genotype 1b is the most common genotype found to circulate worldwide, while genotypes 2a and 2b are relatively common in United States, Europe, and Japan and can also be found in other part of the world (Attaullah et al., 2011). Genotype 3 is found in India, United States and Europe. Genotype 4a is the most common genotype in North Africa and the Middle East. Genotype 6a occurs in Hong Kong and Southeast Asia, while genotypes 5a and 7 are found in South Africa and the Democratic Republic of Congo respectively (reviewed in Hoofnagle, 2002; Murphy et al., 2007a).

Apart from major genotypes, recombinant strains of HCV have been reported in different parts of the world. Recombination plays a significant role in the evolution of RNA viruses by creating genetic variation. Several natural intergenotypic recombinants of HCV and only two cases of an HCV intra-subtype recombinant strain have been identified. A natural intergenotypic recombinant (2k/1b) of HCV has been identified in Saint Petersburg, Russia (Kalinina et al., 2002). HCV strains circulating in Peru demonstrated the existence of natural intra-genotypic HCV recombinant strains (1a/1b).
circulating in the Peruvian population (Colina et al., 2004). A recombinant virus with genotype 2 at the 5'end and genotype 5a at the 3'end has also been found (Legrand-Abravanel et al., 2007). In these cases, the recombination events took place in the non-structural region of the HCV genome. Recombination break-points in HCV structural capsid genomic region E1-E2 has also been identified (Cristina and Colina, 2006). In a quasispecies study from patients with sustained response, only one recombinant strain was detected and the recombination break-point was situated on the PKR-binding region of NS5A (Moreno et al., 2006). In another study, evidence of intrapatient recombination was observed in HCV sequences from over 10% of both HCV-mono-infected patients and HCV/HIV co-infected patients, when targeting the E1-E2 and the NS5A region (Sentanderau et al., 2008).

2.4 The life cycle of HCV

Although the mechanism of the HCV life cycle is not fully understood, an overview has been described based on the current knowledge available (Figure 2.5).
2.4.1 Entry

Previously, the study of HCV replication was hampered by the lack of models, but recently several models have been discovered. These include the HCV pseudotyped particles (HCVpp) (Hsu et al., 2003), cell culture-derived HCV particles (HCVcc) (Lindenbach et al., 2005) and HCV clinical isolates (Bartenschlager and Lohmann, 2000). The primary hepatocytes and hepatocarcinoma cells are the main targets for in vitro infection (Regeard et al., 2008; Bartosch et al., 2003). HCV entry requires low pH and depends on temperature. It is thought to occur via a clathrin-dependent pathway (Evans et al., 2007).

The first step of infection in HCV is the attachment to the host cell surface. Several studies have shown that HCV entry may involve the use of several receptors such as the Cluster of Differentiation 81 (CD81) (Pileri et al., 1998), Scavenger receptor class B member 1 (SRB1) (Scarselli et al., 2002), high density lipoprotein (HDL) (Voisset et al., 2005), low-density lipoprotein (LDL)
(Agnello et al., 1999) and Claudin-1 (Evans et al., 2007). Bartosch et al (2003) investigated the role of CD81 and SR-B1, in HCV entry and reported that although CD81 and SR-B1 are important for infection, other hepatocyte-specific co-factors are also important for HCV entry (Bartosch et al., 2003). In the CD81, the large extracellular loop has been shown to allow binding to the ectodomain of soluble E2 (Diedrich, 2006). SRB1 has been shown to bind the E2 HVR1 but not with the E2 lacking HVR1, and also anti-HVR1 competes with SR-B1 for E2 binding (Scarselli et al., 2002). HDL improves SRB1 channelled cell entry and has been shown to protect HCVpp from neutralizing antibodies (Bartosch et al., 2005). Other cellular proteins like apolipoprotein E Binding (apoE) could also play an important role in HCV attachment through interactions with cell surface heparan sulphate (Jiang et al., 2012). After binding of the E1 and E2 to the receptors, endocytosis is induced with a host-cell membrane in a pH-dependent manner (Bartosch et al., 2003).

2.4.2 Translation and polyprocessing

RNA molecules are translated via a cap-independent IRES mediated process in which viral protein expression is regulated by direct recruitment of each ribosome to the start site of translation (Tsukiyama-Kohara et al., 1992; Wang et al., 1993). IRES-dependent translation regulation also involves other viral factors such as the core protein which has been shown to inhibit viral translation through its interaction with the IRES (Zhang et al., 2002).

HCV genome translation leads to a polyprotein precursor that is co- and post-translationally processed by cellular and viral proteases into mature S and NS proteins (Brass et al., 2006; Suzuki et al., 2007). Host cell signal peptidase is responsible for polyprotein processing at the core/E1, E1/E2, E2/p7, and p7/NS2 sites (Brass et al., 2006). NS2/3 protease is responsible for the NS2 and NS3 proteins intramolecular cleavage (Grakoui et al., 1993; Hijikata et al., 1993). Along with its NS4A co-factor, NS3 activates a catalytic mechanism for the posttranslational processing of the HCV NS proteins (Barbato et al., 1999). Processing at the NS3/4A site is intramolecular, whereas cleavage at the other
sites occurs intermolecularly (Bartenschlager et al., 1994; Tanji et al., 1994).

### 2.4.3 Replication

HCV forms a membrane associated replication complex, consisting of viral proteins, replicating RNA, cellular membranes and host cell factors (Wolk et al., 2008). The NS4B protein stimulates a specific membrane alteration within the cytoplasm (Egger et al., 2002; Gosert et al., 2003). Host factors such as vesicle-associated membrane protein-associated (VAP)-A and –B and soluble N-ethylmaleimide-sensitive fusion protein attachment protein receptor (SNARE)-like proteins are known to confine within the ER and Golgi apparatus and are important for HCV replication by binding with both NS5A and NS5B proteins (Tu et al., 1999, Gao et al., 2004; Hamamoto et al., 2005). Although NS proteins are found in both the ER and the Golgi apparatus, replication occurs primarily in the Golgi apparatus (Aizaki et al., 2004; Suzuki et al., 2007). Recently, another cellular substrate, damaged DNA-binding (DDB1) has been found to be required by NS3/4A for HCV replication (Kang et al., 2013).

For HCV replication, NS5B RdRp catalyzes the synthesis of a negative-strand RNA, and the positive-strand RNA which is produced later from a negative-strand RNA template (Dubuisson, 2007; Suzuki et al., 2007). The NS5B RdRp requires the help of the 3'UTR region to initiate RNA synthesis (Tanaka et al., 1996; Choi et al., 2003). Deletion of the 5’proximal stem-loop RNA element, a poly (U/ UC) tract, and the X domains of the 3’UTR resulted in the failure of viral RNA replication (Yanagi et al., 1999, Luo et al., 2003). HCV RNA synthesis occurs on a lipid raft membrane structure (Shi et al., 2003). The NS3/NS4A proteolytic activity has also been shown to be important for viral RNA replication (Bermudez-Aguirre et al., 2009). The NS3 helicase-NTPase domain stimulates RNA NTPase activity and RNA binding (Penin et al., 2004). Additionally, it has been reported that cell culture-adaptive mutations that increase replication efficiency of HCV RNA are found in the central region of the NS5A protein (Lohmann et al., 2001).
2.4.4 Viral assembly and release

As with other Flaviviruses, the mature HCV virion has a nucleocapsid and outer envelope (Suzuki et al., 2007). Various forms of HCV including free mature virions (Kaito et al., 1994), virions bound to LDL and very-low-density lipoprotein (VLDL) (Andre et al., 2002) and non-enveloped nucleocapsids (Maillard et al., 2001) have been reported to circulate in the sera of infected hosts. The core protein is responsible for the formation of the viral nucleocapsid occurring in the presence or absence of ER-derived membrane (Kunkel et al., 2001; Klein et al., 2004; Majeau et al., 2004). The p7 is the main protein essential for assembly and release of infectious virus (Sakai et al., 2003; Steinmann et al., 2007; Gentzsch et al., 2013). Additionally, the NS2 was also shown to be involved in HCV assembly and release (Jones et al., 2007; Phan et al., 2009; Jirasko et al., 2010; Ma et al., 2011; Stapleford et al., 2011). However, only the NS2 domains, but not its NS2 catalytic activity, are essential for HCV assembly (Jones et al., 2007). In addition, several reports have shown specific roles of NS4B and NS3 in HCV assembly (Ma et al., 2008; Jones et al., 2009; Jones et al., 2011). HCV-like particles are packaged into intracellular vesicles as a result of budding from the ER (Baumert et al., 1998). HCV particles are released from the cell through the secretory pathway (Popescu et al., 2009).

2.5 Disease profile

There are two major disease stages of HCV infection; acute infection and chronic infection and these disease stages may lead to cirrhosis and liver cancer (Pawlotsky, 2003a) (Figure 2.6).
2.5.1 Acute infection

Majority of acute infections are asymptomatic and as a result, these are infrequently diagnosed. About 20% of adults with acute HCV infection develop clinical symptoms which occur from 3 to 12 weeks after exposure (Thimme et al., 2001). Symptoms could include malaise, nausea, and jaundice (Farci et al., 1996; Thimme et al., 2001; Jauncey et al., 2004). In some cases serum alanine aminotransferase (ALT) can reach levels of greater than 10 times the upper limits of normal (Maheswari et al., 2008). Among the subjects exposed to HCV, approximately 40% clear the infection within 6 months (Jauncey et al., 2004). A rare case of fulminant hepatitis has been reported in acute infection (Farci et al., 1996; Chu et al., 1999). Almost all patients ultimately develop antibodies to HCV, with low or undetectable titres in immunosuppressed patients (Chen and Morgan, 2006). HCV RNA is detectable from two weeks following exposure (Thimme et al., 2001) (Figure 2.7).
2.5.2 Chronic infection

Approximately 60–85% of patients who have detectable RNA for six months are categorized as chronically infected (Craxi et al., 2008). Normal ALT levels which are persistent are found in 30% of chronically infected patients (Bruce et al., 2006) (Figure 2.8). Approximately 20% of patients with chronic infection ultimately develop liver cirrhosis which may lead to hepatocellular carcinoma (HCC), hepatic failure and even death. The risk of HCC is estimated at 1-4 percent per year (Lauer and Walker, 2001). The risk of liver disease progression is increased by several risk factors including excessive use of alcohol, being infected at an older age, male gender, the extent of liver biopsy inflammation and fibrosis, and also HIV or HBV co-infection and immunosuppression (Chen and Morgan, 2006). In most cases there are no symptoms in chronic infection for a long time (Lauer and Walker, 2001). Several extrahepatic manifestations have been linked to chronic HCV infection. An estimated 40% of patients infected with HCV develop at least one extrahepatic manifestation during the course of the disease (Galossi et al., 2007). These manifestations can involve multiple organ systems, including renal, dermatologic, hematologic and rheumatologic systems (Chen and Morgan,
The most common extrahepatic conditions are mixed cryoglobulinemia (Ali and Zein, 2005; Latt et al., 2012), diabetes (Elhawary et al., 2011) and chronic kidney diseases (Zignego et al., 2007; Latt et al., 2012).

Figure 2:8 Course of acute hepatitis C that evolves into chronic infection (Adapted from Caruntu and Benea, 2006).

2.6 Pathogenesis

After infection of the liver with HCV, the viral particles are continuously released into circulation (Kanto and Hayashi, 2006) and the host factors activate various innate, humoral and cellular immune responses (Irshad et al., 2008; Rosen, 2003). Innate immune responses are generalized as non-specific immune response, involving type I interferon (IFN) secretion and natural killer (NK) cell activation (Heim, 2013). HCV significantly induces type I IFN in the chimpanzee models irrespective of the outcome of infection (Thimme et al., 2002). Binding of E2 to CD81 may hinder NK cell functions and activate their capability to produce IFN-γ (Tseng and Klimpel, 2002; Crotta et al., 2002). The ability of HCV to cause persistent infection indicates that the innate response may be insufficient to prevent the spread of HCV throughout the liver (Liang et al., 2000; Racanelli and Rehermann, 2003).
The role of the neutralizing responses in the management of HCV infection or persistence is not yet understood. HCV specific antibodies are commonly detected seven weeks after infection, while the HCV E2 neutralization antibodies appear 48 weeks after infections and are associated with HCV viral clearance (Raghuraman et al., 2012). A self-limited course of infection may be due to an early strong antibody response to the HVR1 (Zibert et al., 1997). Anti-HVR1 antibodies have been shown to neutralize HCV in vitro and in chimpanzees (Farci et al., 1994). Variation of the HVR1 is thought to be related to the maintenance of persistent infection by emerging escape variants, followed by the development of chronic infection (Manzin et al., 1998; Farci et al., 2000). The role of cellular responses in clearing HCV and/or persistence is well understood. The development and persistence of strong specific responses by cytotoxic T cells (CD8+) and helper T cells (CD4+) is related to HCV clearance (Lechner et al., 2000). HCV specific CD4+ cell response is associated with successful control of HCV in patients with resolved infection (Day et al., 2002). The CD4+ response promotes activity and activation of the CD8+ by production of cytokines such as IFN (Harcourt et al., 2006). The CD8+ T cells have been shown to play a major role in protective immunity in chimpanzees serially infected with HCV (Shoukry et al., 2003). HCV specific CD8+ T cells are responsible for the destruction of infected hepatocytes and also produce inflammatory cytokines (Neumann-Haefelin et al., 2005).

2.7 Laboratory diagnosis

For the diagnosis of HCV infection, two classes of assays are used which are serologic assays (antibody assays and antigen assays) and molecular tests for viral particles. The two assays do not play any role in the assessment of disease severity (Ghany et al., 2009).

2.7.1 Serologic assays

Available immunoassays to detect anti-HCV in plasma or serum include both enzyme immunoassays (EIAs) and chemiluminescence immunoassays (CIAs).
These assays use recombinant antigens derived from HCV core, NS3, NS4 and NS5 regions. However, the limitation is that they do not distinguish between acute, chronic or past infections (Lemon et al., 2007). The detection of anti-HCV is mainly by the use of 3rd generation EIAs, with a specificity of greater than 99% (Colin et al., 2001). In HIV positive, renal failure and HCV-associated essential mixed cryoglobulinemia patients, false negative results have been reported (Lauer and Walker, 2001). Recently, 4th generation assays that detect the presence of the HCV core antigen and anti-HCV have been developed and evaluated (Laperche et al., 2005; Brandao et al., 2013). These new assays can detect HCV infection at an early stage, especially in HIV patients (Nastouli et al., 2009; Yang et al., 2011) and could be a useful alternative to HCV RNA detection for diagnosis or blood screening in a setting where molecular assays are not implemented (Laperche et al., 2005). Although its sensitivity is lower than the molecular assays, a positive result indicates active infection and these assays do not require sample processing as in the case with molecular assays (Kamili et al., 2012). For confirmatory serologic test the use of recombinant immune blot assay (RIBA) is recommended (Lemon et al., 2007), however RIBA is not routinely used mainly because of cost, complexity and labor intensiveness (Chevaliez and Pawlotsky, 2006)

2.7.2 Molecular assays
The gold standard for diagnosis of active HCV infection is the detection of HCV RNA in the serum/plasma. Using this assay, HCV infection can be detected within a week following exposure (Pawlotsky, 2003b). HCV RNA testing is highly specific (up to 99%) for all HCV genotypes (Kamili et al., 2012). HCV RNA testing plays an important role in HIV positive and chronic renal failure patients since it does not suffer false negative results as seen with the serologic assays due to poor antibody production (Pawlotsky, 2002). HCV RNA tests that are currently in routine use include both qualitative and quantitative tests (Scott and Gretch, 2007) with several being approved by different authorities for diagnostic purposes (Kamili et al., 2012).
2.8 Prevention and control

Currently there is no approved vaccine against hepatitis C. Regardless of the difficulties associated with high variability and mutability of HCV, several HCV vaccine types have been developed and tested. However, few of them have entered clinical trials and include, among others, recombinant proteins, synthetic peptides and DNA-based vaccines (Yu and Chiang, 2010; Torressi et al., 2011). A good HCV vaccine should be able to stimulate the production of neutralizing antibodies and intensify HCV-specific T-cell responses (Helle et al., 2011). Both prophylactic and therapeutic vaccine candidates have been developed against HCV infection. Prophylactic vaccines are aimed at inducing a humoral immune response, while therapeutic vaccines are aimed at inducing both humoral and cellular immune responses. There is limited data on the introduction of vaccines that target humoral response in clinical settings (Ip et al., 2012). A phase I clinical trial of HCV E1/E2 with MF59C1 adjuvant prophylactic vaccine was tested in healthy HCV negative adults and found to be safe and induced antibody, cytokine production and CD4+ proliferation responses (Frey et al., 2010). In another phase I clinical trial, healthy individuals who were vaccinated with recombinant E1 and E2 protein and MF59 adjuvant, were found to have mounted E1- and E2-specific antibodies with in vitro virus-neutralizing activity (Ray et al., 2010). The assessment of prophylactic vaccines is a challenge since it should be tested in individuals at high risk of exposure in areas with high HCV prevalence (Strickland et al., 2008).

Several therapeutic vaccine candidates have entered clinical trials. The NS3 protein has been shown to be a good therapeutic vaccine candidate (Torresi et al., 2011; Lang Kuhns et al., 2012; Naderi et al., 2013) since the majority of the HCV epitopes recognized by CD8+ and CD4+ T cells are situated in this protein (Day et al., 2002; Eckels et al., 2002; Wertheimer et al., 2003; Hakamada et al., 2004). Additionally, other proteins such as the E1 have also been found to be good candidates for therapeutic vaccines. In 2004, the HCV E1 protein was tested as a candidate vaccine in a phase I clinical trial, with 20 healthy male volunteers and was shown to induce anti-E1 humoral response and elicit a strong cellular immune response towards E1 (Leroux-Roels et al., 2004). 

phase II clinical trial with a peptide vaccine, IC41, containing five HCV synthetic peptides derived from the core, NS3 and NS4 with an adjuvant poly-L-arginine indicated that IC41 can induce HCV-specific T cell responses in a group of HCV IFN non-responders (Klade et al., 2008).

2.9 Treatment

The management of chronic HCV infection has depended largely on the use of both Peg-IFN and RBV (Ghany et al., 2011). Elimination of detectable circulating HCV is the primary goal of HCV therapy after halting the treatment (EASL, 2011), resulting in a reduction in progression of the disease, development of cirrhosis and the risk of developing HCC (Yee et al., 2012). The end goal for treatment response is a sustained virological response (SVR) which is defined as HCV RNA negative for six months after end of treatment (Craxi et al., 2008). More importantly, the response to HCV treatment is partially dependent on the infecting genotype. Peg-IFN and RBV can achieve virologic response rates that range from 41-80% (Manns et al., 2001; Fried et al., 2002). However, virologic response rates are lower for genotypes 1 and 4 compared to genotypes 2 and 3 (Ghany et al., 2009). Since large-scale clinical trials do not include a sufficient number of patients with genotype 5, the optimal treatment schedule remains unknown. Despite the lack of information, these patients are treated for 48 weeks (Botha et al., 2010). Few studies suggest that treatment response of HCV genotype 5 is better compared to HCV genotype 1 response, and is very similar to HCV genotypes 2 and 3 responses (Legrand-Abravanel et al., 2004).

In 2011, two NS3 protease inhibitors, BOC and TVR were approved by the FDA for the treatment of HCV genotype 1 infection in combination with Peg-IFN and RBV (Bacon and Khalid, 2012). BOC and TVR were the first direct-acting antiviral (DAA) agents, which inhibit HCV replication directly (Yee et al., 2012). There is a significantly improved rate of SVR in both treatment naïve and
experienced patients in clinical trials of HCV genotype 1 patients on combination therapy of Peg-IFN and RBV together with BOC or TVR. Studies showed that SVR was achieved in 72-75% of treatment naïve patients (Sherman et al., 2011, Jacobson et al., 2011), 63-88% of Peg-IFN and RBV relapsers, and in 33% of Peg-IFN and RBV non-responders (McHutchison et al., 2009; Bacon et al., 2011; Manns et al., 2012; Zeuzem et al., 2011). However, treatment with BOC or TVR is associated with several adverse effects including anaemia and rash (McHutchison et al., 2009; McHutchison et al., 2010; Manns et al., 2012; Sherman et al., 2011; Poordad et al., 2011). The choice of which protease inhibitor to use depends on the cost and duration of treatment as well as the probability of achieving a rapid virological response (RVR). The RVR depends on sensitivity to IFN and the IL28B genotype (Shiffman and Esteban, 2012). Several studies have demonstrated a strong association between IL28B genotype and response to treatment with Peg-IFN and RBV in both HCV mono-infected and HIV/HCV co-infected patients (Dayyeh et al., 2011; Moghaddam et al., 2011; Stattermayer et al., 2011).

There are several DAA that are in clinical trials and these have shown promising preliminary results. Studies conducted in patients who received dual DAA treatment for 24 weeks with the NS5A inhibitor, daclatasvir, and NS3 inhibitor, asunaprevir, with and without Peg-IFN and RBV, have shown that high SVR rates can be achieved (Chayama et al., 2012; Lok et al., 2012). However, HCV mutations which can reduce susceptibility to DAA in treatment naïve patients have been reported to occur naturally. These mutations have been found in patients who did not responds to DAA treatment or who had viral breakthrough (Rong et al., 2010; Hafon and Locarcini, 2011; Lok et al., 2012).
2.10 HIV/HCV co-infection

2.10.1 Burden of HIV and HCV co-infections Worldwide

About 40 million people are estimated to be infected with HIV worldwide of which approximately 5.26 million reside in South Africa (SADOH, 2013). Of the 40 million people infected with HIV, approximately 4-5 million are co-infected with HCV (Alter, 2006). Co-infection with HIV and HCV are common due to shared routes of transmission, however, there are differences in the ways that these viruses are transmitted in the developed and developing countries (Ocama and Seremba, 2011). The USA has a prevalence of HIV/HCV co-infection of around 30%, but this rate varies considerably in studies from different states (Alberti et al., 2005).

2.10.2 Burden of HIV and HCV co-infections in Africa

Several studies in developing countries have shown that HCV prevalence is higher among HIV-positive individuals as compared to HIV-negative individuals, mainly as a result of the shared route of transmission (Barth et al., 2010). However, the prevalence differs by countries. Most countries reported HIV/HCV co-infection rates as low as 1% in South Africa and as high as 99% in Mauritius. There are also differences in the distribution of HIV and HCV in different regions of Africa. The North African region for example is the most affected by HCV with Egypt having prevalence rates as high as 20% in the general population (WHO, 2013). There was no evidence of HIV in HCV infected children with haemophilia in Egypt (Abdelwahab et al., 2012). A high prevalence of co-infection was reported in 362 HIV patients in Tunisia with an anti-HCV prevalence of 39.7% (Kilani et al., 2007). In another study, 83% of Libyan people who inject drugs tested positive for HCV (Mirzoyan et al., 2013). Studies conducted in Central Africa reported an average HCV prevalence of 6% in the general population (Madhava et al., 2002). A low prevalence of 0.6% has been reported for HIV/HCV co-infection in this region, with the highest rates of 4.9% in HIV positive women in Uganda (Pirillo et al., 2007). Another study of HIV outpatients in Uganda reported an HCV
prevalence of 3.3%, while a study conducted in the Democratic Republic of Congo reported no HIV/HCV co-infections (Batina Agasa et al., 2010) and another from Rwanda reported a prevalence of 0.6% HIV/HCV co-infections (Pirillo et al., 2007). Data from West Africa has shown an average prevalence of 2.2% in the general population (Madhava et al., 2002). Most of the HIV/HCV co-infection studies in this region have been conducted in Nigeria and the rate of co-infection ranges from 0-33%. In an epidemiological survey of 200 pregnant women in Nigeria, 33% were HIV/HCV co-infected (Duru et al., 2009) while another study of HIV positive patients reported 14.7% of HIV/HCV co-infection (Balogun et al., 2012). In Cameroon, the rate of co-infection was 4.8% in blood donors (Noubiap et al., 2013) and as high as 16.6% in HIV patients initiating antiretroviral therapy (Laurent et al., 2010). In contrast, there was no HIV/HCV co-infection in HIV positive children from Ivory Coast (Rouet et al., 2008) and in blood donors and pregnant women in Burkina Faso (Collenberg et al., 2006). In Eastern Africa, prevalence rates of HIV/HCV co-infections <10% were reported in most studies (Ayele et al., 2002; Tessema et al., 2010; Ramos et al., 2012; Matee et al., 2006; Croce et al., 2007). However, in Tanzania a study of HIV infected paediatrics showed a high seroprevalence of 13.8% of HIV/HCV co-infection in this group (Telaleta et al., 2007). A survey in Ethiopia of individuals presenting for HIV testing reported an HCV prevalence rate of 10.5% in HIV infected individuals as compared to 6% in HIV negative individuals (Alemayehu et al., 2011). In Southern Africa, prevalence rates of HIV/HCV co-infection described in studies conducted among injection drug users and female sex workers show prevalence rates of up to 80% in Mauritius (Johnston et al., 2011; Johnston and Corceal, 2013) and lower rates of 0.8% in Botswana where the prevalence of HIV is very high (Patel et al., 2011). In South Africa, the prevalence of HIV/HCV co-infection varies. Lodenyo et al (2000) reported 1% anti–HCV positivity in AIDS patients at Chris Hani Baragwanath hospital (Lodenyo et al., 2000). In a CAESAR study, participants demonstrated an HIV/HCV co-infection prevalence of 1.9% in South Africa (Amin et al., 2004). Recently a high prevalence of 13.4% was reported in HIV treatment naïve patients from Kwazulu-Natal (Parboosing et al., 2008).
2.10.3 Impact of HCV infection on HIV disease progression

Conflicting results have been reported on the impact of HCV on HIV disease progression. In developed countries, studies have shown that patients with HIV/HCV co-infection have less immune reconstitution than patients with HIV infection alone (Carlos Martin et al., 2004; Miller et al., 2005), and HCV infection with multiple genotypes might influence HIV progression (Van Asten and Prins, 2004). However, different studies have shown no major differences in HIV-related mortality between HIV/HCV co-infected and HIV-mono infected patients if they are on HAART (Rockstroh et al., 2005). HCV has little impact on the response to ART or HIV disease progression (Law et al., 2004). The impact of HCV in HIV disease progression in Africa is still unclear due to the limited data available. Some studies found that co-infected patients have a higher risk of liver progression to death, an increase in mortality and renal morbidity (Parboosing et al., 2008), lower CD4+ count (Forbi et al., 2007; Adewole et al., 2009; Tremeau-Bravard et al., 2012; Hawkins et al., 2013) and increased serum liver enzyme (Telatela et al., 2007; Laurent et al., 2010). However, other studies did not find a difference between patients with and without co-infection with regards to CD4+ cell count (Pirillo et al., 2007; Sagoe et al., 2012) and HIV WHO clinical disease staging of patients (Sagoe et al., 2012). The risk of hepatotoxicity was significantly higher among HIV/HCV co-infected patients compared to those with HIV alone (Agbaji et al., 2013; Hawkins et al., 2013) and experienced slower immunological recovery after starting antiretroviral therapy (Hawkins et al., 2013). Another study reported that HCV did not significantly impact ART response in a Nigerian cohort (Agbaji et al., 2013).

2.10.4 Impact of HIV infection on HCV disease progression

Several studies have demonstrated that HIV/HCV co-infected patients have more rapid fibrosis progression than HCV mono-infected patients (Benhamou et al., 1999; Poynard et al., 2003). In HIV/HCV co-infected patients; the
chances of spontaneous clearance of HCV are lower (Thomas et al., 2000a; Thio et al., 2000; Forns et al., 2006) and those who develop chronic infection have higher HCV viral loads as compared to HCV mono-infected patients (Thomas et al., 2000b). HAART may worsen the outcome of HCV disease through enhancement of drug induced-hepatotoxicity (Sulkowski et al., 2000).

2.10.5 Management of HIV/HCV co-infected patients

For HIV/HCV co-infected patients, the standard treatment is the same as for HCV mono-infected patients which is Peg-IFN and RBV with a recommended fixed course of 48 weeks (Soriano et al., 2007). In HIV/HCV co-infected patients extended treatment with Peg-IFN and RBV have been found to be beneficial in treatment naive patients who achieved early virological response (EVR) (Chung et al., 2012). The response rate is, however, generally poor in HIV/HCV co-infected patients compared to HCV mono-infected patients infected with the same genotype (Hadziyannis et al., 2004; Torriani et al., 2004).

2.10.6 Transmission of HCV in HIV positive patients in Africa

The transmission of HCV is primarily through exposure to infected blood. The main sources of HCV transmission in Africa are not yet identified (Feld et al., 2005). Although sexual transmission rates is well documented in developed countries, few studies in Africa have been conducted reporting a minor role of heterosexual route in HCV transmission (Pirillo et al., 2007; Ndong-Atome et al., 2009; Ayele et al., 2002). However, a study by Simpore et al (2006) showed that the elevated co-infection rate in HIV-positive women demonstrated that they are exposed to Toxoplasma gondii, HCV, and HBV infections mainly by sexual contact (Simpore et al., 2006). A critical factor for increasing the spread of viruses among injection drug users (IDUs) is sharing of needles, syringes and other injecting drug equipment worldwide. In Africa, studies have supported the role of HCV transmission by IDU (Mirzoyan et al., 2013; Johnston and Corceal, 2013). A study conducted in Mauritius (n = 511), observed a high rates of both HCV and HIV infection among IDUs with 99.7%
of those infected with HIV also infected with HCV (Johnston et al., 2011). In contrast a study, conducted in Uganda showed no association between HCV transmission and IDU (O’Reilly et al., 2011). Another mode of transmission which is common in Africa is traditional scarification. The majority of the women in some African countries still believe in traditional rituals that expose them to dangerous cultural practices such as genital excision and ethnic scarifications where unsterile instruments are used. (Adewole et al., 2009; Zeba et al., 2011). In a study conducted in Nigeria, evidence for the nosocomial transmission of HCV through the use of a contaminated ball of cotton wool has been reported (Mboto et al., 2010). Another study reported a higher prevalence of HBV and HCV in medical waste handlers in relation to non-medical waste handlers (Anagaw et al., 2012). Blood and blood safety remains a major issue in most African countries (Noubiap et al., 2013). Many studies have detected HCV infection in blood donors (Matee et al., 2006; Tessema et al., 2010; Nkrumah et al., 2011) and blood transfusion has been shown to be the main factor associated with HCV infection (Adewole et al., 2009; Diarra et al., 2009; Pepin et al., 2010; Zeba et al., 2011). Some countries have already introduced safety measures to reduce transmission of blood borne pathogens (Outarra et al., 2006; Fessehaye et al., 2011).
CHAPTER THREE

MATERIALS AND METHODS

This chapter provides the general materials and methodology used in chapters four to eight.
3 Materials and Methods

3.1 Ethical Approval
The study has been approved by Medunsa Research and Ethics Committee (MREC) as project no MREC/p/142/2009:PG. No information by which samples can be traced back to patients were released or published. Furthermore, only the relevant people involved in the study had access to the results. To ensure confidentiality, patients’ samples and records were delinked from identifiers.

3.2 Study Population
A total of 78 serum samples from 71 patients were collected during a 4 year period from 2007 to 2011; and 653 patients serum samples stored at -70°C collected from 2004 to 2006 at Tshepang Clinic, DGMAH were used. DGMAH is an academic hospital serving a population of around 4 million from both rural and urban areas. It is a referral hospital for patients from the North West, Mpumalanga, Limpopo and the northwest part of Pretoria, Gauteng. Clinical and biochemical information of the study patients were obtained from hospital records.

3.3 AxSYM assay
Anti-HCV test was performed by using Axsym 3.0 assay (Abbott, Chicago, Illinois, USA).

3.3.1 Principle of the assay
The assay is based on the Microparticle Enzyme Immunoassay (MEIA) principle, which is a variation of EIA principle. The solid phase EIA uses antigens and/or antibodies coated on a surface to bind complementary analytes. The bound analyte is detected by a series of antigen-antibody reactions. In the AxSYM final reaction, an antibody coupled to an enzyme acts upon a substrate to produce a fluorescent end product. The fluorescence
produced by the enzyme reaction is measured and is proportional to the amount of bound antibody. AxSYM HCV version 3.0 has been designed for the qualitative detection of antibodies to putative structural and non-structural proteins of the HCV genome in human serum or plasma.

3.3.2 Interpretation of results
Specimens with an S/CO (ratio of the sample rate (S) to the cut off rate (CO) for each sample and control) value of less than 1.00 were considered non-reactive. Specimens with an S/CO value equal to or greater than 1.00 are considered initially reactive. The reactive specimens on the initial testing were retested in duplicate using the AxSYM HCV version 3.0. If neither of the retested specimens was reactive, the specimens were considered non-reactive for HCV. If the specimen were reactive in either of the replicates the specimen was considered consistently reactive. The test requires 150µl of serum. Due to insufficient amount of serum, none of the specimens were retested with AxSYM HCV 3.0.

3.4 HCV RNA Extraction
RNA was extracted from serum with the QIAmp viral mini RNA kit (Qiagen GmbH, Hilden, Germany) using the spin protocol. The samples were first lysed under high denaturation conditions to inactivate RNases and to ensure isolation of intact viral RNA by pipetting 560 µl of buffer AVL containing carrier RNA and 140 µl of serum into a 1.5ml microcentrifuge tube. After mixing by pulse-vortexing the mixture was incubated at room temperature for 10 minutes and centrifuged to remove drops from the inside of the lid. Approximately 560 µl of 96% ethanol was added to the mixture and vortexed for 15 seconds. The mixture was then loaded onto the QIAmp mini spin column and centrifuged at 8000 rpm for one minute during which the RNA binds to the membrane and contaminants then efficiently washed away using two different wash buffers; 500 µl each of buffer AW1 and AW2. Finally, RNA was eluted in an RNase-free buffer AVE, loaded onto each QIAmp mini spin column and centrifuged at 8000 rpm for a minute. The RNA extracts were then
stored at -70°C until needed for further laboratory tests.

3.5 HCV cDNA synthesis

RNA molecules were converted into single stranded cDNA using the enzyme Revertaid™ RT-PCR (Fermentas, Glen burnie, Maryland, USA). Briefly, 17µl cDNA was prepared on ice as follows: 1µl of 10mM anti sense primer M114 (Table 3.1) or oligo dt, 6 µl nuclease free water and 3µl of RNA template followed by incubation at 65°C for 5 minutes. After chilling on ice, 4µl of 5X cDNA buffer and 1µl of each of the following: Revertaid, Ribolock enzyme and 25mM dNTP mix, was added into each tube, mixed and incubated for 1 hour at 42°C. The reaction was terminated by incubating at 72°C for 10 minutes. The cDNA was either used directly in PCR or stored at –70°C for later use.

3.6 HCV 5'UTR PCR

A nested RT-PCR was performed using primers that targeted the 5'UTR (Table 3.1 in the appendix). The final concentration of the 25µl PCR reaction mix for the first round consisted of 1X PCR buffer (Bioline, Luckenwalde, Germany), 1.5mM MgCl₂, 0.2mM dNTP mix, 0.2mM of primers P209 and P939, 0.5U of Taq polymerase (Bioline, Luckenwalde, Germany), 2µl of cDNA template and 18.05µl of nuclease free water. The cycling conditions were as follows: initial denaturation at 94°C for 4 minutes, followed by 35 cycles of 94°C for 20 seconds, 52°C for 30 seconds and 72°C for 30 seconds, with final extension at 72°C for 5 minutes. For second round the final concentration of the 25µl mixture consisted of 1X PCR buffer (Bioline, Luckenwalde, Germany), 1.5mM MgCl₂, 0.2mM dNTP mix, 0.2mM of primers P211 and P940, 0.5U of Taq polymerase (Bioline, Luckenwalde, Germany), 0.5µl of first round product and 19.55µl of nuclease free water. For second round the same PCR conditions were used with the exception of the annealing temperature which was adjusted to 50°C. The expected sizes of the 5'UTR amplicons was 300bp and 250bp for the first and second rounds respectively.
3.7 HCV C/E1 PCR

A nested PCR was performed using primers that targeted C/E1 (Table 3.1 in the appendix). The final concentration of the 25µl PCR reaction mix for the first round consisted of 1X PCR buffer (Bioline, Luckenwalde, Germany), 1.5mM MgCl₂, 0.2mM dNTP mix, 0.2mM of primers DM110 and DM111, 0.5U of Taq polymerase (Bioline, Luckenwalde, Germany), 3µl of cDNA template and 17.05µl of nuclease free water. The cycling conditions were as follows: initial denaturation at 94ºC for 4 minutes, followed by 35 cycles of 94ºC for 20 seconds, 44ºC for 30 seconds and 72ºC for 30 seconds, with final extension at 72ºC for 5 minutes. For second round, the final concentration of the 25µl mixture consisted of 1X PCR buffer (Bioline, Luckenwalde, Germany), 1.5mM MgCl₂, 0.2mM dNTP mix, 0.2mM of primers DM109 and DM108, 0.5U of Taq polymerase (Bioline, Luckenwalde, Germany), 3µl of first round product and 17.05µl of nuclease free water. For second round the same PCR conditions were used with the exception of the annealing temperature which was adjusted to 40ºC. The expected sizes of the C/E1 amplicons were 494bp and 473bp for the first and second rounds respectively.

3.8 HCV NS5B PCR

A nested PCR was performed using primers that targeted the NS5B region (Table 3.1 in the appendix). The final concentration of the 25µl PCR reaction mix for the first round consisted of 1X PCR buffer (Bioline, Luckenwalde, Germany), 1.5mM MgCl₂, 0.2mM dNTP mix, 0.2mM of primers NS5B1s and NS5B3as, 0.5U of Taq polymerase (Bioline, Luckenwalde, Germany), 3µl of cDNA template and 17.05µl of water. The cycling conditions were as follows: initial denaturation at 94ºC for 2 minutes, followed by 35 cycles of 94ºC for 10 seconds, 43ºC for 30 seconds and 72ºC for 1 minute, with final extension at 72ºC for 7 minutes. For second round PCR, the final concentration of the 25µl mixture consisted of 1X PCR buffer (Bioline, Luckenwalde, Germany), 1.5mM MgCl₂, 0.2mM dNTP mix, 0.2mM of primers NS5B2s and NS5B4as, 0.5U of Taq polymerase (Bioline, Luckenwalde, Germany), 3µl of first round product and 17.05µl of nuclease free water. For second round the same PCR
conditions were used as in the first round. The expected sizes of the NS5B amplicons were 450bp and 400bp for the first and second rounds respectively.

3.9 HCV full genome PCR

The HCV genome was amplified in 10 overlapping fragments A-J covering the 5’UTR to 3’UTR (Table 3.1 in the appendix). Primer sequences from Koletzki et al., 2009 and Li et al., 2009 were modified to suit genotype 5 sequences.

3.9.1 Fragment A

A nested PCR was performed using primers that targeted fragment A (5’UTR-E1). The final concentration of the 25µl PCR reaction mix consisted of 1X PCR buffer with 1.75mM MgCl₂ (Roche, Indianapolis, USA), 0.4mM dNTP mix, 1mM of primers [P939 and M101 (first round) and P940 and M102 (second round)], 1U of Expand long template polymerase (Roche, Indianapolis, USA), 3µl of cDNA template and 17.05µl of water. The cycling conditions were as follows: denaturation at 94°C for 2 minutes, followed by 35 cycles of 94°C for 45 seconds, 45°C for 45 seconds, and 72°C for 2 minutes (during the last 25 cycles, the elongation time was increased by 20 seconds per cycle) with final extension at 72°C for 10 minutes. The expected size of fragment A amplicons was 1258bp and 1238bp for the first and second rounds respectively.

3.9.2 Fragment B

A nested RT-PCR was performed using primers that targeted fragment B (E2). The final concentration of the 25µl PCR reaction mix consisted of 1X PCR buffer (Bioline, Luckenwalde, Germany), 1.5mM MgCl₂, 0.2mM dNTP mix, 0.2mM of primers [F1A and R1A (first round) and F2A and R2A (second round)], 0.5U of Taq polymerase (Bioline, Luckenwalde, Germany), 2µl of cDNA template and 18.05µl of water. The cycling conditions were as follows: initial denaturation at 94°C for 2 minutes, followed by 35 cycles of 94°C for 10 seconds, 55°C for 30 seconds and 72°C for 30 seconds, with final extension
at 72°C for 5 minutes. For second round the same PCR conditions were used with the exception of the annealing temperature which was adjusted to 52°C. The expected size of the fragment B amplicons was 337bp and 300bp for the first and second rounds respectively.

### 3.9.3 Fragment C
A nested PCR was performed using primers that targeted fragment C (E2). The final concentration of the 25µl PCR reaction mix for the first round consisted of 1X PCR buffer with 1.75mM MgCl₂ (Roche, Indianapolis, USA), 0.4mM dNTP mix, 1mM of primers [M105 and R1B (first round) and R2B and M106 (second round)], 1U of Expand long template polymerase (Roche, Indianapolis, USA), 3µl of cDNA template and 17.05µl of water. Thermal-cycling conditions for the inner and outer reactions were as follows: denaturation at 94°C for 2 minutes, followed by 35 cycles of 94°C for 45 seconds, 48°C for 45 seconds, and 72°C for 1 minute (during the last 25 cycles, the elongation time was increased by 20 seconds per cycle) with final extension at 72°C for 10 minutes. For second round the same PCR conditions were used with the exception of the annealing temperature which was adjusted to 45°C. The expected size of fragment C amplicons was 769bp and 724bp for the first and second rounds respectively.

### 3.9.4 Fragment D
A nested PCR was performed using primers that targeted fragment D (E2). The final concentration of the 25µl PCR reaction mix consisted of 1X PCR buffer with 1.75mM MgCl₂ (Roche, Indianapolis, USA), 0.4mM dNTP mix, 1mM of primers [F1C and R1C (first round) and F2C and R2C (second round)], 1U of Expand long template polymerase (Roche, Indianapolis, USA), 3µl of cDNA template and 17.05µl of water. The cycling conditions were as follows: denaturation at 94°C for 2 minutes, followed by 35 cycles of 94°C for 45 seconds, 45°C for 45 seconds, and 72°C for 1 minute (during the last 25 cycles, the elongation time was increased by 20 seconds per cycle) with final extension at 72°C for 10 minutes. The expected size of fragment D amplicons
was 893bp and 839bp for the first and second rounds respectively.

### 3.9.5 Fragment E

A nested PCR was performed using primers that targeted fragment E (E2-NS3). The final concentration of the 25µl PCR reaction mix for the first round consisted of 1X PCR buffer with 1.75mM MgCl₂ (Roche, Indianapolis, USA), 0.4mM dNTP mix, 1mM of primers [P101F and P106R (first round) and P102F and P105R (second round)], 1U of Expand long template polymerase (Roche, Indianapolis, USA), 3µl of cDNA template and 17.05µl of water. Cycling conditions were as follows: denaturation at 94°C for 2 minutes, followed by 35 cycles of 94°C for 45 seconds, 48°C for 45 seconds, and 72°C for 2 minutes (during the last 25 cycles, the elongation time was increased by 20s per cycle) with final extension at 72°C for 10 minutes. The expected size of fragment E was 1646bp and 1547bp for the first and second rounds respectively.

### 3.9.6 Fragment F

A nested PCR was performed using primers that targeted fragment F (NS3). The final concentration of the 25µl PCR reaction mix for the first round consisted of 1X PCR buffer with 1.75mM MgCl₂ (Roche, Indianapolis, USA), 0.4mM dNTP mix, 1mM of primers [P103F and P107R (first round) and P104F and P108R (second round)], 1U of Expand long template polymerase (Roche, Indianapolis, USA), 3µl of cDNA template and 17.05µl of water. The cycling conditions were as follows: denaturation at 94°C for 2 minutes, followed by 35 cycles of 94°C for 45 seconds, 48°C for 45 seconds, and 72°C for 4 minutes (during the last 25 cycles, the elongation time was increased by 20 seconds per cycle) with final extension at 72°C for 10 minutes. For second round the same PCR conditions were used with the exception of the annealing temperature which was adjusted to 43°C. The expected size of fragment F was 1162bp and 1133bp for the first and second rounds respectively.
3.9.7 Fragment G
A nested RT-PCR was performed using primers that targeted fragment G (NS3-NS5A). The final concentration of the 25µl PCR reaction mix for the first round consisted of 1X PCR buffer with 1.75mM MgCl₂ (Roche, Indianapolis, USA), 0.4mM dNTP mix, 1mM of primers [M107 and F2AS (first round) and M108 and F7AS (second round)], 1U of Expand long template polymerase (Roche, Indianapolis, USA), 3µl of cDNA template and 17.05µl of water. The cycling conditions were as follows: denaturation at 94°C for 2 minutes, followed by 35 cycles of 94°C for 45 seconds, 45°C for 45 seconds and 72°C for 4 minutes (during the last 25 cycles, the elongation time was increased by 20 seconds per cycle) with final extension at 72°C for 10 minutes. The expected size of fragment G amplicons was 3520bp and 3091bp for the first and second rounds respectively.

3.9.8 Fragment H
A nested RT-PCR was performed using primers that targeted fragment H (NS5A-NS5B). The final concentration of the 25µl PCR reaction mix for the first round consisted of 1X PCR buffer with 1.75mM MgCl₂ (Roche, Indianapolis, USA), 0.4mM dNTP mix, 1mM of primers [595S and M110 (first round) and F1S and M109 (second round)], 1U of Expand long template polymerase (Roche, Indianapolis, USA), 3µl of cDNA template and 17.05µl of water. Thermal-cycling conditions for the inner and outer reactions were as follows: denaturation at 94°C for 2 minutes, followed by 35 cycles of 94°C for 45 seconds, 53°C for 45 seconds, and 72°C for 3 minutes (during the last 25 cycles, the elongation time was increased by 20 seconds per cycle) with final extension at 72°C for 10 minutes. The expected size of fragment H amplicons was 2546bp and 2469bp for the first and second rounds respectively.

3.9.9 Fragment I
A nested RT-PCR was performed using primers that targeted fragment I (NS5B). The final concentration of the 25µl PCR reaction mix for the first
round consisted of 1X PCR buffer with 1.75mM MgCl₂ (Roche, Indianapolis, USA), 0.4mM dNTP mix, 1mM of primers [M111 and M114 (first round) and M112 and M113 (second round), 1U of Expand long template polymerase (Roche, Indianapolis, USA), 3µl of cDNA template and 17.05µl of water. The cycling conditions were as follows: 94°C for 2 minutes, followed by 35 cycles of 94°C for 10 seconds, 43°C for 30 seconds and 72°C for 1 minute, with final extension at 72°C for 7 minutes. The expected size of fragment I amplicons was 828bp and 815bp for the first and second rounds respectively.

3.9.10 Fragment J
A nested RT-PCR was performed using primers that targeted fragment J (NS5B-3'UTR) using RACE PCR kit (Roche, Indianapolis, USA), as per manufacture’s intructions. The first-round amplification used primer P109F and a ‘poly (A)’ primer. The second-round amplification used primer P110F and the ‘NUP’ primer.

3.10 HCV E1 and E2 PCR
The E1 and E2 regions were amplified according to the methods as described in sections 3.9.2 to 3.9.4 (Fragment B-D).

3.11 Detection of the HCV PCR product
PCR products were loaded on a 1% agarose gel stained with ethidium bromide and electrophoresed at 100 volts (V) for 30 minutes in 1X Tris-acetate-EDTA (TAE) buffer (Qiagen Gmbh, Hilden, Germany). Expected bands were viewed against 100bp and 1000bp DNA molecular markers (Fermentas, Glen burnie, Maryland, USA) using ultraviolet (UV) transillumination.
3.12 HCV gene cloning
The PCR products were ligated into a p-GEM T-Easy vector system II (Promega, Madison, Wisconsin, USA) and then transformed into chemically competent Escherichia coli JM109 cells according to the manufacturer’s instruction. To determine the transformation efficiency, 100µl of cells were used. The tubes were gently flicked to mix and placed on ice for 20 minutes. The cells were heat-shocked for 45-50 seconds in a water bath at exactly 42°C. The tubes were immediately returned to ice for 2 minutes. A total amount of 950µl of room temperature super Optimal broth with Catabolite repression (S.O.C) medium was added to each tube containing cells transformed with the ligation reaction and 900µl to the tube containing cells transformed with uncut plasmid. The tubes were incubated for a total of 1 hour and 30 minutes at 37°C while shaking at approximately 150rpm. Following incubation, 100µl of the transformed cells were then cultured by the spread plate method onto Luria Bertani (LB) [NaCl 1% w/v, tryptone 1% w/v, yeast extract 0.5% w/v, agar 1.5% w/v] agar plates containing ampicillin (50µg/ml) and X-Gal (5-bromo-4-chloro-3-indoyl]-D-galactopyranoside) (40mg/ml) and incubated overnight at 37°C. The presence of X-Gal allowed for blue/white differentiation of colonies based on the detection of β-galactosidase activity. The ampicillin selected for colonies that had taken up the vector. White colonies containing the plasmid were chosen for screening via colony PCR.

3.13 Sequencing of the HCV PCR products
All positive PCR products were sequenced with BigDye®Terminator v3.1 Cycle Sequencing on the ABI 3500XL (Inqaba Biotechnological Industry, PTY, Ltd, Pretoria, South Africa) following the Sanger-Coulson method (Sanger et al., 1977). The primers for the nested PCR were used as initial sequencing primers (Table 3.1). For near full-length amplification the remaining nucleotide sequences of the amplicons were obtained by primer walking.
3.14 Phylogenetic analysis of HCV sequences

3.14.1 Alignment of sequence data
The nucleotide sequences were assembled using Chromas pro version 7 (School of Health Science, Griffith University, Australia). Nucleotide sequence alignments were done using CLUSTAL X (Thompson et al., 1997) and MAFFT (Katoh et al., 2002). Nucleotide and amino acid sequences were compared and translated into proteins using BioEdit (Hall, 1999).

3.14.2 Model testing
The substitution models for investigating the sequences were selected with a hierarchical likelihood ratio testing implemented in jModelTest 3.7 (Posada, 2008). The best-fitting model was the general time reversible model assuming four rate categories among sites with a gamma distribution and a proportion of invariable sites.

3.14.3 Tree construction
The evolutionary relationship between the isolated sequences was investigated by constructing a phylogenetic tree. Representative sequences from different HCV genotypes were retrieved from GenBank databases (http://www.ncbi.nlm.nih.gov/genbank/). The genetic distances and evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analysed and a bootstrap value of more than 70% of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. Phylogenetic analyses were conducted in MEGA 5.1 (Tamura et al., 2011).

Additional phylogenetic inference was performed using a Bayesian MCMC approach as implemented in the BEAST v1.5.0 program (Drummond and Rambaut, 2007) under an uncorrelated log-normal relaxed molecular clock and the GTR model with nucleotide site heterogeneity estimated using a
gamma distribution. The BEAST MCMC analysis was run for a chain length of 50,000,000. Three independent runs were performed for each dataset. Combined results were visualized in Tracer v1.4 to confirm adequate chain convergence, and the ESS was calculated for each parameter. All ESS values >200 indicated sufficient sampling. The maximum clade credibility tree was selected from the posterior tree distribution after a 10% burn-in using TreeAnnotator v1.5.0.

3.15 Bayesian coalescent analysis

For coalescent analysis, the number of near full-length sequences was limited with a very narrow range of sample collection dates. Thus, coalescent analysis was conducted with E1 sequences only. The 18 E1 sequences were sampled from 2007 to 2011, and an external substitution rate of $5 \times 10^{-4} \pm 1.5 \times 10^{-4}$ per site per year was used as previously described (Verbeeck et al., 2006). Bayesian skyline plots combined with each of the uncorrelated exponential and uncorrelated lognormal relaxed molecular clocks were run twice for a chain length of 50 million (Drummond et al., 2012).

3.16 Prediction of T-cell epitopes

For MHC class I, prediction for binding alleles were predicted using ProPred I (http://www.imtech.res.in/raghava/propred1/) at a 4% default threshold by keeping the proteosome and immunoproteosome filters on at 5% threshold. ProPred 1 predicts antigenic epitopes for 47 MHC I alleles (Singh and Raghava, 2003). For MHC class II, prediction was performed using ProPred (http://www.imtech.res.in/raghava/propred/) at a 3% default threshold. ProPred predicts antigenic epitopes for 51 MHC II alleles (Singh and Raghava, 2001). All alleles in the Propred I and propred software were included for analysis.
3.17 Prediction of B cell epitopes

For identification of 16-mers B-cell epitopes, the e1 and E2 region of sequence ZADGM2088 was predicted using the program ABCpred (http://www.imtech.res.in/raghava/abcpred/) at a 0.51 default threshold. ABCpred server predicts B cell epitopes using artificial neural network using fixed length patterns (Saha and Raghava, 2006).

3.18 Antigenecity of the epitopes

The Antigenecity of all the predicted epitopes were analysed using VaxiJen v2.0 online antigen prediction (www.ddg-pharmfac.net/vaxijen/). Proteins having antigenic score >0.4 were selected as antigenic. VaxiJen v2.0 allows antigen classification based on the physicochemical properties of proteins without recourse to sequence alignment.

3.19 Epitope conservancy analysis

All predicted epitopes were analyzed for conservation using the IEDB database (http://tools.immuneepitope.org/tools/conservancy/iedb_input) at a threshold of >100% conservation in comparison with randomly selected sequences from each of the HCV genotypes 1, 2, 3, 4 and 6.

3.20 Validation of predicted epitopes

All the predicted epitopes were submitted to IEDB database (http://www.immuneepitope.org/) to confirm if they had been tested previously by other studies. The immuneepitope database contains experimentally confirmed data about antibody, T-cell epitopes, MHC binding, MHC restriction and MHC class.

3.21 Immune selection pressure

The Individual codons under positive selection were detected via fixed effects
likelihood (FEL) methods as implemented in the DataMonkey program for 18 E1 and E2 sequences (Pond and Frost, 2005). FEL is a likelihood-based and statistically rigorous method used to fit an independent non-synonymous (dN) and synonymous (dS) to every site in the context of codon substitution models and it tests whether or not dN is equal to dS. The program directly estimates non-synonymous and synonymous substitution rates at each site.

3.22 Analysis of N-linked glycosylation sites

The N glycosylation sites were predicted for 18 E1 and E2 sequences using the on-line prediction server NetNGlyc version 1.0 (http://www.cbs.dtu.dk/services/NetNGlyc/), which predicts N glycosylation sites in proteins by artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequins. The networks can identify 86% of the glycosylated and 61% of the non-glycosylated sequins, with an overall accuracy of 76%.

3.23 Peptide design

Structure analysis of the sequence was performed using the Protparam online tool (Gasteiger et al., 2005). Protparam computed different parameters including the molecular weight, theoretical pl, AA composition, atomic composition, extinction coefficient, instability index, aliphatic index and grand average of hydropathicity (GRAVY). To check for post-translational modifications, predicted peptides were predicted for N-linked glycosylation as described in section 3.22 and for N-linked phosphorylation using the NetPhos 2.0 (Blom et al., 1999) program. The NetPhos 2.0 produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in sequences. Only those motifs that showed NetPhos score of 0.7 or greater were considered.

3.24 GenBank submission

Sequences were submitted to GenBank under the accession numbers
HQ396466-396518 for the 5'UTR, HQ385855-385885 and JN116558-116566 for the NS5B region, and HQ385886-385914 and JN116567-116577 for the C/E1 region. Near full-length sequences were submitted under the accession numbers KC767829 - KC767834 while E1-E2 sequences were submitted under accession numbers KC767835 - KC767846.
CHAPTER FOUR

INVESTIGATING HCV CO-INFECTION IN HIV-POSITIVE PATIENTS ENROLLING FOR HAART AT DGMAH.

This chapter is based on the publication: Should routine screening of HCV be mandatory in HIV patients enrolling for ARV therapy in South Africa?


Published in South African Medical Journal; 100: 814-815.

Supplementary data is annexed in appendix chapter 11: Publication.
4 INVESTIGATING HCV CO-INFECTION IN HIV-POSITIVE PATIENTS
ENROLLING FOR HAART AT DGMAH.

4.1 Introduction

The South African government has been scaling-up HAART for treatment of HIV/AIDS in the public health sector since April 2003. Despite this development, research programmes to monitor the efficacy of HAART in HIV/AIDS patients co-infected with HBV or HCV do not exist. We have recently shown that 63% of HIV infected South African patients initiating HAART have past or present HBV infection (Lukhwareni et al., 2009). However, there is paucity of data on the burden of HCV prevalence in HIV patients. One study demonstrated a low prevalence of 1.9% (Amin et al., 2004) and another, a high prevalence of 13.4% in Kwazulu-Natal (Parboosing et al., 2008). For this study HCV was investigated in HIV positive patients who were candidates for HAART at DGMAH.

4.2 Methods

This study undertook to investigate the burden of HCV co-infection in HIV-positive patients enrolling for HAART at a tertiary hospital in Pretoria. The study population consisted of 653 serum samples stored at -70°C collected from 2004 to 2006. All sera were screened for anti-HCV using the AxSYM assay version 3.0 (Abbott Laboratories, North Chicago) following manufacturer’s instructions. Due to limited serum volumes of less than 150µl to confirm the initial screening results with a second serological assay, all anti-HCV positives (samples with S/CO [i.e ratio of the sample rate [S] to the cut-off rate [CO] for each sample and control] of above 5.23) and preliminary positives (samples with S/CO between 1.00 and 5.00) were subjected to in-house qualitative RT-PCR assay. Viral RNA was extracted from serum with the QIAmp viral mini RNA kits (Qiagen, Gmbh, Germany), followed by PCR targeting the highly conserved 5’ UTR as previously described with slight modifications (Sithebe et al., 1996). PCR positive sample were sequenced to

4.3 Results

Serological results indicated that only 1.2% (8/653) of samples were positive for anti-HCV, with S/CO values ranging from 5.69 to 37.8. Of these 8 samples, HCV RNA was detected in only one sample, which had the highest anti-HCV titre of 37.08 (Table 4.1). Sequencing confirmed that the RT-PCR product was HCV specific, the genotyping results will be discussed in chapter 5. In contrast, 20.7% (135/653) of samples were preliminary positive, with S/CO values ranging from 1.02 to 4.48. None of these samples tested positive for HCV RNA by RT-PCR. Laboratory diagnosis of HCV still has major limitations as the majority of currently available serological assays do not distinguish between acute, chronic and past infections. HCV RNA positive is regarded to be a marker of replication, whereas HCV RNA negative in anti-HCV positive patients is indicative of either resolved infection or very low viraemia with HCV RNA levels below the detection limit of the RT-PCR assay.

Table 4.1: Summary of patients with reactive anti-HCV

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age</th>
<th>Gender</th>
<th>Anti-HCV (S/CO)</th>
<th>RT-PCR assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>37</td>
<td>F</td>
<td>37.08</td>
<td>Positive</td>
</tr>
<tr>
<td>2</td>
<td>51</td>
<td>M</td>
<td>6.29</td>
<td>Negative</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>M</td>
<td>7.77</td>
<td>Negative</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>F</td>
<td>5.69</td>
<td>Negative</td>
</tr>
<tr>
<td>5</td>
<td>43</td>
<td>M</td>
<td>11.62</td>
<td>Negative</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
<td>F</td>
<td>6.42</td>
<td>Negative</td>
</tr>
<tr>
<td>7</td>
<td>33</td>
<td>F</td>
<td>11.48</td>
<td>Negative</td>
</tr>
<tr>
<td>8</td>
<td>37</td>
<td>F</td>
<td>6.05</td>
<td>Negative</td>
</tr>
</tbody>
</table>
4.4 Discussion

It has been estimated that South Africa is one of the countries with the highest number of people living with HIV (SADOH, 2013). This study has found an HCV sero-prevalence of only 1.2% in HIV/AIDS patients enrolling for HAART in Pretoria. All but one appeared to have past or resolved HCV infection as indicated by a negative RT-PCR assay. Comparable findings were reported in pregnant women from Gabon (Ndong-Atome et al., 2008). It has been reported previously that during the natural history of HCV infection, 20 to 30% of infected individuals eliminate the virus spontaneously (Uto et al., 2006). It is possible that the low prevalence of HCV antibodies in our study could be due to missed HCV antibodies as many of HIV/AIDS patients fail to generate antibodies due to immunosuppression (Sorbi et al., 1996), and RT-PCR assay was only used to confirm anti-HCV positives or preliminary positives. It is also possible that the high number of HCV RNA negative results is the results of false anti-HCV positive with AxSYM 3.0 assay as the initial screening results were not confirmed with another serological assay due to insufficient sera. Other limitations include lack of fresh samples, insufficient volume and that the study was conducted on patients attending a tertiary HIV referral clinic in a hospital setting.

In conclusion this study shows that the majority of HIV/AIDS patients initiating HAART at DGMAH hospital have low exposure to, or active, HCV infection and it does not appear that HIV/AIDS is a risk factor for increased detection of HCV co-infection. While our findings do not support mandatory HCV screening in HIV/AIDS patients initiating HAART in South Africa, consideration should be given to patients who may be at an increased risk such as HIV positive haemophiliacs, injection drug users and diabetics.
CHAPTER FIVE

INVESTIGATING THE DISTRIBUTION OF HCV GENOTYPES IN HIV POSITIVE AND HIV NEGATIVE PATIENTS AT DGMAH.

Part of this chapter is based on the publication: Introduction of new subtypes and variants of hepatitis C virus genotype 4 in South Africa.

Maemu P Gededzha, Selokela G Selabe, Thanda Kyaw, Nare J Rakgole, Jason T Blackard and M Jeffrey Mphahlele.

Published in Journal of Medical Virology 2012; 84:601–607.

Supplementary data is annexed in appendix chapter 11: Publication.
5 INVESTIGATING THE DISTRIBUTION OF HCV GENOTYPES IN HIV POSITIVE AND HIV NEGATIVE PATIENTS AT DGMAH.

5.1 Introduction

HCV genotyping is important for studies of its origin, transmission and evolution (Simmonds et al., 2004). An association has been made between risk of HCC and certain HCV subtypes (Roffi et al., 2001). HCV genotyping is also a marker frequently required for the study and follow-up of patients chronically infected with HCV. HCV isolates are classified into at least six major genotypes (genotypes 1 to 6), whose nucleotide sequences differ by 31-33% (Simmonds et al., 2005). In 2007, novel genotype 7 was described as well (Murphy et al., 2007a). HCV genotypes can be divided further into subtypes which differ by 10–30%. Moreover, distinct isolates from the same subtype may differ in nucleotide sequence by 5–15% (Abdo and Lee, 2004). HCV genotype is frequently assigned by phylogenetic analysis of the 5'UTR, core/E1, NS5B, and/or complete genome sequences (Simmonds et al., 2005). Genotype 1a is common in North America and Northern Europe. Genotype 1b has a worldwide distribution and is often found to be the most common genotype in different parts of the world. Genotypes 2a and 2b are also found worldwide and are relatively common in North America, Europe, and Japan (Simmonds, 2004). Genotype 3 is found in India, United States and Europe. Genotype 4a is the most common genotype in North Africa and the Middle East. Genotype 6a is found mainly in Hong Kong and Southeast Asia, while genotypes 5a and 7 are found in South Africa and the Democratic Republic of Congo respectively (Hoofnagle, 2002; Murphy et al., 2007a).

HCV antiviral treatment is increasingly available in South African public hospitals through public-private partnerships. Importantly, the response to HCV treatment is partially dependent on the infecting genotype. Currently, utilized treatment such as Peg-IFN and RBV can achieve virologic response rates that range from 41-80% (Manns et al., 2001; Fried et al., 2002).
However, virologic response rates are lower for genotypes 1 and 4 compared to genotypes 2 and 3 (Ghany et al., 2009). Since large-scale clinical trials do not include a sufficient number of patients with genotype 5a, the optimal treatment schedule for genotype 5a is largely unknown. Despite the lack of information, these patients are treated for 48 weeks (Botha et al., 2010). Nonetheless, a few studies have suggested that treatment response of genotype 5 is better than for genotype 1 and is very similar to response rates for genotypes 2 and 3 (Legrand-Abravanel et al., 2004).

Genotype 5a has been reported to be the most predominant genotype in South Africa, although genotypes 1-4 have also been reported (Smuts and Kannemeyer, 1995; Prabdial-Sing et al., 2008). Previously, HCV genotype 5a was believed to be found only in South Africa; however, studies demonstrate that HCV genotype 5a infections can be found worldwide (Murphy et al., 1994; Davidson et al., 1995, Levi et al., 2002, Henquell et al., 2004; Verbeeck et al., 2006). HCV genotype 5a strains have been spreading independently in Belgium and South Africa for more than 100 years, with a rate of spread characteristic of an epidemic genotype (Verbeeck et al., 2006). In a recent study in Johannesburg, South Africa, genotype 5a was found to be the dominant genotype in the Liver disease groups and genotype 1 dominated in the Haemophiliac group. Other genotypes such as genotypes 2, 3 and 4 were also found, with subtypes 4c and 4g being identified for the first time in South Africa (Prabdial-Sing et al., 2008). In a study by Sithebe et al., 1996, the most predominant genotype in patients with liver disease was genotype 2 (9/26; 35%), followed by genotype 5a (8/26; 31%) and genotype 1 (2/26; 8%), and an estimated 26.9% (7/26) of this cohort was reported untypable by RFLP (Sithebe et al., 1996).

For identification of new genotypes, classification is based on extensive phylogenetic analysis on the complete sequence of the coding region. However, in order to assign a given sequence to a particular HCV subtype, three or more examples of independently infected individuals with a new proposed subtype is required for a definitive, subtype-specific designation.
based on phylogenetic analysis of at least partial genomes, including both the core/E1 region and the NS5B region (Simmonds et al., 2005). The 5’UTR is the region of choice for qualitative and quantitative HCV RNA detection due to its high level of conservation and sensitivity and it has most often been used by clinical laboratories for routine genotyping of HCV (Murphy et al., 2007b).

Apart from major genotypes, recombinant strains of HCV have been reported in different parts of the world (Kalinina et al., 2002; Colina et al., 2004; Moreau et al., 2006; Legrand-Abravanel et al., 2007). Recombination may present a significant challenge to the treatment of HCV infection. In addition, recombination may affect diagnosis as many of the current genotyping methods focus on the 5’ UTR (Bhattacharya et al., 2011), and most genotyping studies have included only one region such as C/E1 or NS5B. Thus, detection of potential recombination events is unlikely due to methodological constrains (Simmonds et al., 2005). Several population-based studies have analysed the 5’UTR and/or core/E1 genotype versus that of the NS5B region to identify samples with discordant genotypes that may indicate recombination (Xia et al., 2008; Zhou et al., 2010). However, HCV recombination has not been investigated to date in South Africa. Thus, the current study evaluated HCV genotypic diversity and recombination in South Africa based on comparative analysis of the 5’UTR, C/E1, and NS5B regions.

5.2 Methods

5.2.1 Study population

A total of 78 serum samples from 71 patients were analysed. Anti-HCV test was performed by Axsym 3.0 assay (Abbott, Chicago, Illinois, USA) as described in section 3.3 by the National Health Laboratory Service (NHLS) during a 4 year period from 2007 to 2011 at DGMAH. DGMAH is an academic hospital serving a population of around 4 million from both rural and urban areas. It is a referral hospital for patients from the North West, Mpumalanga, Limpopo and the northwest part of Pretoria, Gauteng. Clinical and biochemical information of the study patients were obtained from hospital
records. Anti-HCV positive serum samples were collected prospectively, aliquoted and stored at −70°C for molecular analysis.

5.2.2 Reverse transcriptase PCR and cloning
Viral RNA was extracted from 140μl of serum using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions as described in section 3.4. HCV RNA was detected by reverse transcriptase polymerase chain reaction (RT-PCR) followed by nested PCR using primers corresponding to the 5'UTR, C/E1 and NS5B (Table 3.1 in the appendix) as previously described (Sithebe et al., 1996; Murphy et al., 2007b). Study sequences that did not cluster closely with other HCV reference sequences were subsequently cloned into p-GEM T-Easy vector system (Promega, Madison, WI, USA) following the manufacturer’s instructions as described in section 3.12.

5.2.3 Sequencing of the PCR products
All positive PCR products were sequenced with BigDye® Terminator v3.1 Cycle Sequencing on the ABI 3500XL (Inqaba Biotechnological Industry, PTY, Ltd, Pretoria, South Africa) following the Sanger-Coulson method (Sanger et al., 1977). The primers for the second round PCR were used as sequencing primers.

5.2.4 Phylogenetic tree construction
The evolutionary relationship between the isolated sequences was investigated by constructing a phylogenetic tree. Representative sequences from different HCV genotypes were retrieved from GenBank databases (www.nlm.nih.com). The genetic distances and evolutionary history were inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analysed and a bootstrap value of more than 70% of replicate trees in which the associated taxa clustered together in the bootstrap test is
shown next to the branches. Phylogenetic analyses were conducted in MEGA 5.1 (Tamura et al., 2011).

Additional phylogenetic inference was performed using a Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in the Bayesian Evolutionary Analysis by Sampling Trees (BEAST) v1.5.0 program (Drummond et al., 2012) under an uncorrelated log-normal relaxed molecular clock and the generalized time reversible (GTR) model with nucleotide site heterogeneity estimated using a gamma distribution. The BEAST MCMC analysis was run for a chain length of 50,000,000. Three independent runs were performed for each dataset. Combined results were visualized in Tracer v1.4 to confirm adequate chain convergence, and the effective sample size (ESS) was calculated for each parameter. All ESS values >200 indicated sufficient sampling. The maximum clade credibility tree was selected from the posterior tree distribution after a 10% burn-in using TreeAnnotator v1.5.0.

5.2.5 Recombination analysis
To identify possible recombination events, bootscanning analysis of sequences was performed as implemented in SimPlot version 3.5.1 using the Kimura 2-parameter with a 100 bp window, a 10 bp step increment, and 1000 bootstrap replicates (Lole et al., 1999). Each sequence was compared to reference sequences for genotypes 1 to 7 using the references from GenBank. If >70% of the permuted trees showed similarity to more than one genotype across the region analysed, these “parental” reference sequences were retained in a second bootscanning analysis along with an outlier and the query sequence.

5.2.6 Statistical analysis
Descriptive statistics (namely frequencies of genotypes and gender, and measures of dispersion and central tendency of biochemical variables and age) were calculated using Microsoft Excel (Microsoft Office, 2010). Correlation statistics was used to determine if there is a positive correlation
between mean amino acid diversity (p-distance) and mean nucleotide diversity. Mean values for each group were calculated.

5.3 Results

5.3.1 Patients’ characteristics

A total of 78 serum samples from 71 patients were analysed. Of the 71 patients, 28 were male and 24 were female. There was no information regarding the gender of the remaining patients. The mean age was 54 years (range: 22 to 89 years). The values of ALT varied from 19 international units per litre (IU/L) to 645 IU/L (mean ± standard deviation [SD]: 86 ± 130), aspartate aminotransferase (AST) varied from 21 IU/L to 1489 IU/L (172 ± 322), and lactase dehydrogenase (LDH) varied from 30 IU/L to 2837 IU/L (497 ± 651). ALT, AST, and LDH values higher than normal values were found in 33%, 70%, and 80% of the patients, respectively.

5.3.2 Overview of HCV genotypes

The HCV genotype distribution in the studied population was as follows: 59.2% (42/71) were genotype 5, 18.3% (13/71) were genotype 1, 14.1% (10/71) were genotype 4, and 4.2% (3/71) were genotype 3. Three of 71 (4.2%) individuals were infected with mixed genotypes based on the 5'UTR (Figure 5.1). Of these three patients, one had chronic renal failure and two were diabetic. Clonal sequences of the diabetic patients ZADGM886 and ZADGM909 identified genotype 4 combined with minority genotype 1, while patient with chronic renal failure was infected with genotypes 5 and 1. Although the sample size was modest, there were no significant differences when comparing genotypes and clinical presentation, ALT, AST, and LDH of the patients. Genotype results were not stratified according to HIV status, since only one strain was from the HIV positive group, and the patient was genotype 1b.
5.3.3 Phylogenetic analysis

Only 41 of the 78 sequences were included in the phylogenetic analysis of the 5’UTR together with representative GenBank sequences. Three of the samples ZADGM3460, ZADGM7684 and ZADGM7114 clustered in a group that had 5 varied subtypes of genotype 4 references. ZADGM5566 clustered between GenBank reference sequences for genotype 1b and 1c (Figure 5.2). The 5’UTR sequences were too conserved to distinguish genotypes into significant subtypes. In the C/E1 region 64 of the 78 samples were amplified, of which 10 failed sequencing. Phylogenetic analysis of the 54 C/E1 sequences was able to classify the samples into different subtypes (Figure 5.3).
Figure 5:2 Neighbor-joining tree based on the 5’UTR sequences for 41 HCV samples (indicated by the symbol ●). GenBank reference sequences are indicated by their accession numbers. The nucleotide sequence divergence between sequences can be estimated using 0.005 divergence bar.
Genotype 4 has the most varied subtypes as compared to other genotypes. Samples ZADGM655 and ZADGM1903 clustered with subtype 4k references, ZADGM8690 clustered with subtype 4r, ZADGM4188 and ZADGM3460 clustered with subtype 4q, ZADGM225 and ZADGM7684 clustered with subtype 4c references. However, three sequences (ZADGM9538, ZADGM6426 and ZADGM3771) were identified as 4q in the C/E1 regions (Figures 5.3) but HCV subtype could not be determined by neighbour-joining analysis of NS5B and were therefore considered “unclassified” by this method (Figure 5.4 and 5.5). There was 100% agreement with regards to genotype classification when comparing 5’UTR, C/E1 and NS5B sequences from the same individual. However, the HCV subtype was discordant in 2 samples (ZADGM2739 and ZADGM9300) that were classified as subtype 1b in the 5’UTR analysis but as 1a in the C/E1 and NS5B regions.
Figure 5.4 Neighbor-joining tree based on the NS5B sequences for 42 HCV samples (indicated by the symbol ●). GenBank reference sequences are indicated by their accession numbers. The nucleotide sequence divergence between sequences can be estimated using 0.05 divergence bar.
Figure 5.5 Neighbor-joining tree based on the 12 NS5B sequences for genotype 4 from this study and reference sequences from the GenBank. HCV samples are indicated by the symbol ●). GenBank reference sequences are indicated by their accession numbers. The nucleotide sequence divergence between sequences can be estimated using 0.05 divergence bar.

When the NS5B sequences, were further analysed using a Bayesian inference approach, ZADGM886 was classified as subtype 4q but with a posterior probability <70%. The results of the neighbour joining were
confirmed with the Bayesian approach in samples ZADGM9538, ZADGM6426, and ZADGM3771 and are not related to any other reference sequences highly supported by a posterior probability of 0.99 (Figure 5.6). Samples ZADGM9538, ZADGM6426, and ZADGM3771 were collected from the same patient at different time point.

Figure 5:6 Phylogenetic inference based on a Bayesian MCMC approach as implemented in the BEAST program. Relevant posterior probabilities are shown in italics. The scale bar indicates 0.3 nucleotide substitutions per site.
5.3.4 Recombinant analysis
All NS5B sequences were further analysed for possible inter or intragenotypic recombination using SimPlot (Figure 5.7). NS5B sequences were initially compared to reference sequences for HCV genotypes 1-7. If >70% of the permuted trees showed similarity to more than one genotype across the NS5B region analysed, these “parental” reference sequences were retained in a second bootscanning analysis along with an outlier and the query sequence. The NS5B sequence for subject ZADGM9538 was classified as subtype 4I at the 5’ end but 4q at the 3’ end (i.e., a 4I↔4q recombinant). No other intragenotypic recombinant viruses were detected.

![Bootscanning analysis of recombination within the NS5B region using a 100 bp window, a 10 bp step increment and 1000 bootstrap replicates.](image)

Figure 5.7 Bootscanning analysis of recombination within the NS5B region using a 100 bp window, a 10 bp step increment and 1000 bootstrap replicates.
5.4 Discussion

It has previously been reported that genotype 5a is the most predominant HCV genotype in South Africa (Smuts and Kannemeyer, 1995; Prabdial-Sing et al., 2008). This was confirmed in the current study, in which genotype 5a was observed in 59% of the study population. Initially, genotype 5a was thought to circulate only in South Africa; however, subsequent studies have demonstrated genotype 5a infections worldwide (Levi et al., 2002; Henquell et al., 2004; Verbeeck et al., 2006). Although genotype 5a is the most predominant genotype, our study indicated that genotypes 1 and 4 are encountered in a significant proportion (32.4%) of this population. Genotype 4 was present in 14.1% of the study population, although a very low prevalence of genotype 4 distribution was reported in other studies (Smuts and Kannemeyer, 1995; Prabdial-Sing et al., 2008). However, this is the first report of HCV genotype 4 at the DGMAH in Pretoria.

The prevalence of mixed genotype varies widely depending on the method used and population studied (Giannini et al., 1999). In our study, we were able to detect mixed genotypes by direct sequencing and clonal analysis of the 5’UTR. Not only are major genotypes important for treatment decisions; rather, minor strains in patients with mixed infections might also influence the outcome of therapy and this should be taken into account if patients do not respond as expected to the applied therapy regimen (Schroter et al., 2003).

Phylogenetic analysis of the 5’UTR was accurate in determining genotypes, while the C/E1 and NS5B coding regions were able to differentiate both genotypes and subtypes, including an outlier group. In this study, we have revealed the existence of distinct variants of HCV which were divergent from other genotype 4 subtypes. Furthermore, this analysis has shown, for the first time in South Africa, the presence of HCV subtypes 4k, 4q and 4r. These subtypes were characterised previously among patients in other African countries (Democratic Republic of Congo, Rwanda and Burundi) (Li et al., 2009). Recombinant HCV strains have been identified in other parts of the world (Kalinina et al., 2002; Colina et al., 2004; Moreau et al., 2006; Legrand-
Abravanel et al., 2007; Sentandreu et al., 2008) but none that analysed HCV sequences from the African continent. In this study, evidence for intragenotypic recombinant was found in one of the 42 NS5B sequences analysed. Recombination plays a significant role in the evolution of RNA viruses by creating genetic variation, by reducing mutational load, and by creating viruses with new properties (Worobey et al., 1999). The viability and pathogenicity of intergenotypic and intragenotypic recombinants are more difficult to assess (Simmonds, 2004). Currently, no method exists for classifying recombinant forms of HCV (Simmonds et al., 2005). Our data is in agreement with other studies that reported that analysis of more than one subgenomic region is necessary to avoid missing recombinant strains (Kalinina et al., 2002; Legrand-Abravanel et al., 2007).

In conclusion, genotype 5a is still the predominant genotype in South Africa. Furthermore, the study indicates the introduction of new subtypes and existence of variants of genotype 4 circulating in South Africa.
CHAPTER SIX

CHARACTERIZATION OF NEAR FULL-LENGTH GENOME OF HCV GENOTYPE 5 STRAINS.

This chapter is based on the publication: Near full-length genome analysis of HCV genotype 5 strains from South Africa.

Maemu P Gededzha, Selokela G Selabe, Jason T Blackard, Thanda Kyaw and M Jeffrey Mphahlele.

Published in Infections, Genetics and Evolution 2014; 21:118-123.

Supplementary data is annexed in appendix chapter 11: Publication.
6 NEAR FULL-LENGTH GENOME ANALYSIS OF HCV GENOTYPE 5 STRAINS FROM SOUTH AFRICA

6.1 Introduction

Globally, HCV is a public health concern with an estimated 150 million people infected. Acute HCV infection leads to chronic hepatitis in almost 80% of cases that may then progress to cirrhosis and HCC over several decades (Lauer and Walker, 2001). HCV is an enveloped virus that belongs to the genus Hepacivirus in the family Flaviviridae. The genome consists of 9.5 kilobases of single-stranded, positive-sense RNA that encodes for at least 3 structural (Core, E1 and E2) and 7 non-structural (p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B) proteins which are flanked by 5’ and 3’ untranslated regions (UTR) (Lemon et al., 2007).

HCV exhibits high genetic diversity and is classified into at least 6 genotypes and multiple subtypes (Simmonds et al., 2005), with a putative genotype 7 described in 2007 (Murphy et al., 2007a). Globally, there is regional predominance of certain genotypes. For instance, genotype 1b is found worldwide, whereas genotypes 1a and 3a are found in Europe and North America. Genotype 2 is mainly observed in the Far East, and West Africa, genotype 4 in the Middle East, genotype 5 in South Africa, genotype 6 in South East Asia (Simmonds et al., 2005), and genotype 7 in the Democratic Republic of Congo (Murphy et al., 2007a).

Genotype assignment typically relies on phylogenetic analysis of the HCV core/E1, NS5B, and/or complete genome sequences (Simmonds et al., 2005). Viral genomes of different genotypes may vary by 31-33%, while subtypes may differ by 20–25% in nucleotide sequences (Simmonds et al., 2005). Complete genome sequences are important for accurate HCV classification as well as molecular epidemiology and can also be compared and analysed with any set of sub genomic sequences (Lu et al., 2007). Phylogenetic analysis of one or more complete genome sequences is required to confirm new genotypes, as well as the presence or absence of recombination.
(Simmonds et al., 2005). However, only very limited sequence data are available for genotype 5. Genotype 5 is the most predominant genotype in South Africa, although genotypes 1-4 have also been reported (Prabdial-Sing et al., 2008; Gededzha et al., 2012). It was previously believed that HCV genotype was found only in South Africa, however, recent studies demonstrated that genotype 5 infections can occur elsewhere (Levi et al., 2002; Henquell et al., 2004; Verbeeck et al., 2006). Although highly prevalent in South Africa with increasing reports in Europe, HCV genotype 5 is classified into only one subtype (5a) based on complete genome analysis, and there are only two complete genomes of HCV genotype 5a available in GenBank (Chamberlain et al., 1997; Bukh et al., 1998). Thus, few studies have been performed to characterise genotype 5a. Pang et al. (2009) determined the evolutionary branching order of 345 full-length sequences from major genotypes in correlation to their respective responses to therapy. The lack of genotype 5 sequences led to the inability to determine its relationship to other genotypes (Pang et al., 2009). Therefore, the aim of this study was to characterise HCV genotype 5 sequences from South Africa.

6.2 Methods

6.2.1 Study population
For the current study, near full-length genomes were amplified from six HCV genotype 5 samples, as well as the E1 region from an additional 12 genotype 5 samples that were previously identified based on 5'UTR sequencing (Gededzha et al., 2012).

6.2.2 PCR amplifications
Viral RNA was extracted with QIAmp viral mini RNA kits (Qiagen, Hilden, Germany) as described in section 3.4. HCV RNA was converted into cDNA with the enzyme Revertaid™ RT-PCR (Fermentas, Lithuiana, USA) using 10pmol of an antisense primer M114. For near full-length, nested PCR was performed in nine overlapping PCR fragments (A-I) as described in section
3.9. The E1 region (nucleotide positions 915-1488, numbering according to Accession no. M62321) was amplified by fragment B. Primer sequences from Koletzki et al (2009) were modified to suit genotype 5 sequences (Table 3.1 in the appendix).

6.2.3 Sequencing and phylogenetic analysis

Direct sequencing was performed with ABI 3500XL (Inqaba Biotechnological Industry, PTY, Ltd, Pretoria, South Africa) using second round PCR and “sequence walking” primers for longer fragments. Sequence fragment assembly was performed using Chromas Pro1.5 (www.technelysium.com.au/chromas.html). Representative sequences from different HCV genotypes were retrieved from the GenBank database. Nucleotide sequence alignments were performed with CLUSTAL X. The substitution models for investigating the sequences used were the general time reversible model with a gamma distribution and invariant sites selected using the jModelTest 3.7 (Posada, 2008). The evolutionary relationship was investigated by constructing a phylogenetic tree and pairwise genetic distances with 1000 bootstrap replicates using MEGA 5.1 (Tamura et al., 2011).

Additional phylogenetic inference was performed using a Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in the Bayesian Evolutionary Analysis by Sampling Trees (BEAST) v1.7.5 program (Drummond et al., 2012) under an uncorrelated lognormal relaxed molecular clock. The BEAST MCMC analysis was run for a chain length of 50,000,000 and results were visualized to confirm adequate chain convergence. The effective sample size (ESS) was calculated for each parameter, and all ESS values were >200 indicating sufficient sampling. The maximum clade credibility tree was selected from the posterior tree distribution after a 10% burn-in using Tree Annotator v1.5.0. The near full-length sequences (nucleotide positions 109-9038) and the E1 region (nucleotide positions 915-
numbering according to Accession no. M62321 – were used to construct phylogenetic trees.

6.2.4 Bayesian coalescent analysis

For coalescent analysis, the number of near full-length sequences was limited with a very narrow range of sample collection dates; thus, coalescent analysis was conducted with E1 sequences only. The E1 region sequences were sampled from 2007 to 2011, and an external substitution rate of $5 \times 10^{-4} \pm 1.5 \times 10^{-4}$ per site per year was used as previously described (Verbeeck et al., 2006). Bayesian skyline plots combined with each of the uncorrelated exponential and uncorrelated lognormal relaxed molecular clocks were run twice for a chain length of 50 million (Drummond et al., 2012).

6.2.5 Recombination analysis

Recombination events were assessed by bootscanning analysis with SimPlot version 3.5.1 using the Kimura 2-parameter with a 200 bp window, a 20 bp step increment, and 1000 bootstrap replicates (Lole et al., 1999). Each sequence was compared to individual GenBank references representing HCV genotypes 1 to 7, including genotype 1a (M62321), genotype 1b (AF356827), genotype 2a (AB047639), 3a (D17763), genotype 3b (D49374), genotype 4a (Y11604), genotype 5a (Y13184 and AF064490), genotype 6a (DQ480513), and genotype 7a (EF108306).

6.3 Results

6.3.1 Sequencing and phylogenetic analysis

Six near full-length genomes were sequenced. Excluding ZADGM1104, for which amplification of the 5'UTR-core region failed, while other samples had sequences starting from the 5'UTR to the NS5B. Analysis of the five near full-length sequences showed that the 5'UTRs were 279 nt in length, followed by open reading frames of 8754-8757 nt (2918-2919 amino acids [aa]). For
sequence ZADGM1104, only 8192 nt of the opening reading frame was analysed. Among all genomic regions, only E2 and NS5A showed differences in nucleotide lengths. All sequences had a length of 1350 nt (450 aa) in the NS5A with the exclusion of ZADGM0518 with 1353 nt (451 aa). In the E2, five sequences had 1092 nt (364 aa), while ZADGM1104 had 1089 nt (363 aa). Seven sequences of genotype 5 – five from this study and two from GenBank – showed one additional aa in the E2 and NS5A compared to genotype 1 with the exception of ZADGM1104 that had a deletion of 1 aa in the E2 HVR1 at position 387. ZADGM0518 has an insertion in domain III of NS5A before the sequence SMPPLEGEPEGDPDL that is conserved in all HCV genotypes. ZADGM3013 had a 2 aa changes in the SMPPLEGEPEGDPDL sequence.

A phylogenetic tree was estimated using the near full-length genome sequences. All genotype 5 sequences formed a monophyletic cluster supported by a high (100%) bootstrap value (Figure 6.1).
Figure 6:1 Neighbor-joining tree based on near full-length sequences for 6 HCV samples (indicated by the symbol ●). GenBank reference sequences are indicated by their accession numbers. The nucleotide sequence divergence between sequences was estimated using 0.05 divergence bar.

A second phylogeny was reconstructed using the Bayesian inference method. Genotype 5 sequences again clustered together with a posterior probability of 1.0 (Figure 6.2).
Figure 6.2 Phylogenetic inference of the 6 near full-length HCV genome based on a Bayesian MCMC approach as implemented in the BEAST program. The scale bar indicates 0.07 nucleotide substitutions per site.
In addition to the E1 regions from the 6 near full-length genotype 5a sequences, E1 was sequenced for an additional 12 genotype 5 samples. Phylogenetic analysis of the E1 region indicated that the sequences from this study were closely related to genotype 5a sequences from other regions (Figure 6.3). No evidence of intergenotypic recombination was detected for the sequences evaluated (data not shown).
Figure 6.3: Neighbor-joining tree based on E1 sequences for 18 genotype 5 HCV samples (indicated by the symbol ●). GenBank reference sequences are indicated by their accession numbers. The nucleotide sequence divergence between sequences is estimated using the 0.01 divergence bar.
Nucleotide genetic distances between sequences from this study and previously reported genotype 5 references were also analysed. Intragroup genetic distances for genotype 5 ranged from 8% to 11%, with an average of 9.5%. Comparison to other genotypes showed genetic distances ranging from 31% (ZADGM2088 vs. M62321-1a) to 36% (ZADGM3013 vs. EF108306-7a) at the nucleotide level.

### 6.3.2 Coalescent analysis

The estimation of tMRCA was carried out by coalescent analysis of the E1 sequences. The tMRCA (in years before last sampling date) for the analyses were between 116 years (relaxed lognormal clock) and 134 years (relaxed exponential clock) (Table 6.1).

<table>
<thead>
<tr>
<th>Model</th>
<th>Mean tMRCA (SEM)</th>
<th>Median tMRCA</th>
<th>95% HPD lower</th>
<th>95% HPD lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lognormal</td>
<td>123 (0.3)</td>
<td>116</td>
<td>68</td>
<td>197</td>
</tr>
<tr>
<td>Exponential</td>
<td>147 (0.9)</td>
<td>134</td>
<td>62</td>
<td>258</td>
</tr>
</tbody>
</table>

Key: SEM – standard error of mean; HPD – highest posterior density

The Bayesian skyline plot (BSP) summarizes the spread and epidemic growth of the South African HCV genotype 5. Under different models, similar patterns of growth were shown. BSP plot from both the exponential and lognormal clock showed a relatively constant population size was maintained until around 1950s with a slight increase and again a relatively constant population size was maintained till present (Figure 6.4).
Figure 6.4: Bayesian skyline plot representing the estimates of the effective number of the 18 HCV genotype 5a from South Africa. The left graph is Uncorrected Exponential and the right graph is for Uncorrected Lognormal. The vertical ruler on the left scales the effective population size while the horizontal ruler on the bottom measures time, the solid line represents the median estimate and the filled area represent credibility interval based on 95% highest posterior density (HPD) interval.

6.4 Discussion

Currently only two full-length genomes of genotype 5a have been reported, likely due to its restricted circulation in South Africa, where the prevalence of HCV is low (Gededzha et al., 2010) and its circulation as a minor genotype in some European countries (Nguyen and Keefe, 2005). Thus, there is very limited information about the origin, epidemic history, disease severity, and treatment outcome of HCV genotype 5 infections. Although preliminary studies suggest that genotype 5 responds better to interferon and ribavirin treatment than genotype 1 (Legrand-Abravanel et al., 2004; Bonny et al., 2006), large clinical studies must still be performed in countries such as South Africa with a high proportion of genotype 5 infections to confirm these findings.

This study aimed to characterise genotype 5a in South Africa. Although the sequences were missing part of the NS5B (8754-8757 nt long), two genotype 5 have been described previously with 9042 nt long in the coding region (Chamberlain et al., 1997; Bukh et al., 1998). In this study, an amino acid deletion in the E2 and insertion in the NS5A was observed in each of the two sequences, although these could be natural mutants. The E2 protein is the
main target for neutralizing antibody responses, and mutations in this region are associated with the development of chronic infection (Farci et al., 2000). Moreover, NS5A heterogeneity has been shown to play a role in the response to interferon therapy (Puig-Basagoiti et al., 2005). Including the sequences analysed in the current study, there are only 8 complete or near complete genotype 5 sequences available – seven from South Africa and one from the United Kingdom. This study shows significant relationships between the study sequences and genotype 5 references sequences by phylogenetic analysis and pairwise distance calculations. The divergence of HCV near full-length genome with other references from GenBank was calculated by using the neighbor-joining and Bayesian inference approaches and found to have similar tree topologies for the regions sequenced. Phylogenetic analysis of the E1 sequences with reference sequences from all over the world further support that genotype 5 sequences clustered in one group as a single subtype.

The demographic history was estimated by coalescent analysis of E1 sequences. Since the samples from this study were contemporaneous (i.e. the sample were collected within a 10 year period), an external rate was used to calibrate. The use of priors on substitution rates is subject to bias, although this can be avoided to some extent by using the relaxed molecular clock. Under these priors, the results analysed show a tMRCA ranging from 116 years for the lognormal clock to 134 years for the exponential clock. It has been reported that the use of the non-contemporaneous sequences results in the estimation of the recent tMRCA as compared to the use of priors that yield older tMRCA (Magiorkinis et al., 2009; Re et al., 2011). The tMRCA for genotype 5a infections in this study is similar to that reported by Verbeeck et al. in Belgian and South African patients. Initially, genotype 5 was thought to have originated in South Africa; however, Verbeeck et al., 2006, speculated that it may have originated in Central Africa and spread through travelling between South Africa and Central Africa (Verbeeck et al., 2006). A recent analysis of the E1 region suggests that genotype 5a was introduced in France later than in Belgium (Henquell et al., 2011). Study of the evolutionary
analysis in complete genome of the 6 major genotypes (1-6) showed HCV genotype 5 occurred early after the origin of the common ancestor, and there are three major lineages of HCV genotypes that arose from the common ancestor. Additionally, genotype 5 has evolved much slower than other major genotypes and this could be due to unequal evolutionary rates among different genotypes (Salemi and Vandamme, 2002).

In conclusion, this study provides six near full-length nucleotide sequences of HCV genotype 5a strains that can be used as references to design efficient vaccines and for the development of new antiviral agents. The results of this study concur with the previous reports that demonstrated that genotype 5 is classified into a single subtype currently, and it offers significant insights for understanding the diversity of genotype 5 in South African population.
CHAPTER SEVEN

PREDICTION OF IMMUNOLOGICAL T-CELL EPITOPES OF HCV GENOTYPE 5a.

Maemu P Gededzha, Selokela G Selabe, Thanda Kyaw, M Jeffrey Mphahlele.

Manuscript in preparation

Supplementary data is annexed in appendix chapter 11: Tables.
7 PREDICTION OF IMMUNOLOGICAL T-CELL EPITOPES OF HCV GENOTYPE 5a

7.1 Introduction
Currently, there is no vaccine available for preventing HCV infection. Due to lack of proofreading ability during replication, HCV induces a very high mutation rate of $1.44 \times 10^{-3}$ nucleotide changes in the viral genome/year and this is advantageous for the virus to increase its rate of evolution (Simmonds, 2004). Within an infected individual, the virus exists as a quasispecies of closely related yet distinct variants, resulting in immune escape variants (Shimizu et al., 1994), treatment failure (Ghany et al., 2009) and hindrance of vaccine development. The standard treatment for HCV infection is combination therapy with Pegylated-interferon and Ribavirin (Fried et al., 2002). Recently, two NS3 protease inhibitors (Bocepevir and Telaprevir) and NS5b inhibitor (Sofosbuvir) have been approved by the FDA for the treatment of HCV genotype 1 (Bacon and Khalid, 2012; Keating and Vaidya, 2014). However, this therapy is still not affordable in most developing countries and it remains to be seen if it will be effective against other HCV genotypes. As a result, the development of an HCV vaccine is undoubtedly the best solution for the ultimate control of HCV infections.

The development of an effective HCV vaccine is a public health priority and requires understanding of the immune response. Immune response to HCV is associated with both cell mediated (T-cells) and humoral immune (B cells) responses (Ashfaq et al., 2011a). CD4+ and CD8+ present the viral antigen via the Major Histocompatibility Complex (MHC) or human leukocyte antigen (HLA) class I and class II, respectively (Singh et al., 2007). Different HLA class allele is associated with HCV infection. HLA-A*11, HLA-Cw*04 and HLA-B*53 are associated with HCV persistence (Thio et al., 2002; Fanning et al., 2004). HLA-B*27, HLA-A*1101, HLA-B*57, HLA-Cw*0102 and HLA-A*03 have been associated with spontaneous HCV clearance (McKiernan et al., 2004; Neumann-Haefelin et al., 2006). HLA-DRB1*11 alleles and HLA-DQB1*0301 are always associated with decreased disease severity of HCV.
globally and may present the HCV epitopes more proficiently to CD4+ cells than others and thus capable of viral clearance (Hong et al., 2005). During acute HCV infection, development and persistence of strong specific responses by CD8+ and CD4+ T-cells (Lechner et al., 2000; Spada et al., 2004) and neutralizing antibodies (Lavillette et al., 2005) are associated with viral clearance. However, CD8+ and CD4+ T-cells are usually lost in patients who develop persistence to viral infection. The rate of chronic liver disease progression has been showed to be determined by the magnitude of HCV-specific CD4+ T-cell responses (Kamal et al., 2001).

HCV infection evades the host’s immune system by generating immune escape variants and by altering the virus HLA-restricted epitopes so that they are no-longer recognized by T-cells and neutralizing antibodies (Petrovic et al., 2012). CD8+ T-cells are essential for longterm protection against chronic HCV (Shoukry et al., 2003), while CD4+ T-cells play a role in viral clearance (Kleenermann and Thimme, 2012). Prophylactic vaccines against viral infections are aimed at inducing a humoral immune response, while therapeutic vaccines preferably activate both humoral and cellular immune responses (Ip et al., 2012). Several HCV vaccine types are in clinical trials and they include, among others, recombinant proteins, synthetic peptides and DNA-based vaccines (Torresi et al., 2011). A successful HCV vaccine will need to stimulate the production of neutralizing antibodies and intensify HCV-specific T-cell responses (Helle et al., 2011). Protective epitopes that display minimal cross-genotype amino acid variability should also be the focus of the immune response target as they will provide the highest chance of developing a vaccine with broad potency (Brown et al., 2007). Peptides are the desirable vaccine candidates because they are easy to produce and to construct, and they also do not contain infectious materials (Patronov and Doytchinova, 2013). The first step in the process of epitope-based vaccine design and development is the prediction of in-silico peptide binding affinities to MHC proteins (Dimitrov et al., 2011). Genotype 5 is the most predominant genotype in South Africa accounting for more than 50% of the infections (Gededzha et al., 2012) with increasing emergence in Europe and North America (Antaki et
al., 2010). Genotype 5 is the most conserved genotype as compared to other genotypes as it is classified into only one subtype (5a) (Chamberlain et al., 1997; Bukh et al., 1998). With the growing prevalence of HCV genotype 5a in different parts of the world, it is important to characterize HCV genotype 5a sequences to improve the formulations of vaccine candidates that are in development. Thus, the objective of this chapter was to assess immunological determinants by predicting conserved epitopes in near-full length HCV genotype 5a sequences using a suite of online programmes to help in the designing of new vaccine candidates.

7.2 Methods

7.2.1 Prediction of T-cell epitopes

For T-cell epitopes prediction 2 available Genotype 5a full-length sequences available in the GenBank and 6 of the near full length sequences from the chapter 6 were selected aligned and create consensus using BioEdit (Hall, 1999) for the prediction of T-cell epitopes. Prediction of T cell epitopes was performed as described in section 3.16.

7.2.2 Common South African HLA alleles

For epitope prediction binding to common HLA alleles found in South Africa, the IEDB epitope analysis tool (http://tools.immuneepitope.org/main/jsp/menu) was used for HLA class I using the artificial neural network (ANN) algorithm of (Nielsen et al., 2003) on the IEDB server, while for Class II Prediction of T cell epitopes was performed as described in section 3.16. The most common South African alleles were found in literature search published by (Paximadis et al., 2012).

7.2.3 Antigenicity of the epitopes

The Antigenecity of all the predicted epitopes were analysed as described in section 3.18.
7.2.4 Epitope conservancy analysis

All predicted epitopes were analyzed for conservation using the IEDB database (http://tools.immuneepitope.org/tools/conservancy/iedb_input) as described in section 3.19 in comparison with 406, 221, 98, 33, 45, 45 randomly selected sequences from each of the HCV genotypes 1a, 1b, 2, 3, 4 and 6 respectively.

7.2.5 Validation of predicted epitopes

All the predicted epitopes were checked if they have been tested previously from other studies as described in section 3.20.

7.3 Results

7.3.1 T-cell epitopes prediction

For HLA class I, a total of 24 antigenic epitopes were predicted in the consensus near full-length of genotype 5a (Table 7.1 in the Appendix). Epitope NS3\textsuperscript{1325-1333} covered 30 of the 47 HLA class I alleles that were analysed assuring high binding affinity to different alleles. For conservation analysis with other genotypes, 8 of 24 epitopes were conserved for specific genotypes. Epitopes NS3\textsuperscript{1332-1340} and NS5B\textsuperscript{2557-2565} were highly conserved in all genotypes analysed, and epitope E2\textsuperscript{684-692} was conserved in all genotypes except for genotype 2, while other 5 epitopes were conserved in 2 or 3 genotypes analysed (Table 7.1 in the Appendix). For HLA class II, 77 epitopes were predicted (Table 7.2 in the Appendix). Epitope NS4B\textsuperscript{1879-1887} and NS4B\textsuperscript{1880-1888} covered 51 of 51 HLA class II alleles analysed. For conservation analysis with other genotypes, 31 of 71 epitopes were conserved for specific genotypes. Some epitopes were highly conserved in all genotypes (E2\textsuperscript{507-515}, E2\textsuperscript{509-517}, NS3\textsuperscript{1253-1261}, NS3\textsuperscript{1254-1262}, NS3\textsuperscript{1327-1335}, NS3\textsuperscript{1392-1400}, NS4B\textsuperscript{1916-1924} and NS4B\textsuperscript{1919-1927}), while other epitopes were conserved in at least 1 to 4 of the genotypes analysed (Table 7.2 in the Appendix). The highest number of
HLA class I binding epitopes were predicted within the NS3 (63%), followed by NS5B (21%), and for the HLA class II, the highest number of epitopes were predicted in the NS3 (30%) followed by the NS4B (23%) proteins (Table 7.3 in the Appendix).

7.3.2 Epitope binding affinity to common South African HLA-I and HLA-II alleles

All conserved epitopes and their genotypic variants were further analysed for binding affinity to HLA class I and HLA class II allele most common in South Africa were genotype 5a is predominating. For HLA class I, 12 of the most common HLA-A, HLA-B, HLA-C in South Africa (4 each) were analysed for binding affinity with the conserved epitopes using IEDB, ANN prediction server and for HLA class II 5 most common HLA-DRB alleles were predicted using Propred. Epitope E2 \(^{684-692}\) ALSTGLIHL, conserved in genotypes 1a, 1b, 3, 4, 5a and 6 and its genotype 2 variant ALSTGL\(L\)HL predicted to bind to with high affinity (28 IC\(_{50}\)nm and 37 IC\(_{50}\)nm respectively) to HLA-A*02:01 allele. The NS3 \(^{1359-1367}\) HPNIEEVAL bind with intermediate affinity (53 IC\(_{50}\)nm) to HLA-B*07:02, while its genotype 2, 3 and 4 variant HS\(N\)IEEVAL bind with poor affinity (18334 IC\(_{50}\)nm) and genotype 6 variant HPN\(I\)TE\(T\)AL bind with high affinity (16 IC\(_{50}\)nm). For HLA-B*35:01, the NS3 \(^{1359-1367}\) HPNIEEVAL and genotype 6 variant HPN\(I\)TE\(T\)AL bind with high affinity (12 IC\(_{50}\)nM and 15 IC\(_{50}\)nM respectively) while 2, 3 and 4 variant HS\(N\)IEEVAL bind with poor affinity (3248 IC\(_{50}\)nm) (Table 7.4 in the Appendix).

The limitation of Propred 1 is that it does not cover most of the main HLA-I alleles HLA-A (HLA-A*01:01, HLA-A*30:01 and HLA-A*30:02), HLA-B (HLA-B*08:01) and (HLA-C) HLA-C*04:01, HLA-C*06:01, HLA-C*07:02 and HLA-C*07:02) that are observed in South Africa. As a results the IEDB epitope analysis tool was used to predict epitopes all 12 of the most common HLA-A, HLA-B, HLA-C covering HLA-I allele found in South Africa. Twelve 13 antigenic epitopes covering 3 HLA-A alleles (HLA-A*01:01, HLA-A*02:01 and
HLA-A*30:01), 2 HLA-B alleles (HLA-B*35:01 and HLA-B*07:02) and 1 HLA-C allele (HLA-C*07:01) were predicted with high binding affinity of <50 IC_{50nM} were predicted. Epitope NS5B^{2889-2297} LSAFSLHSY to 2 HLA-A alleles (HLA-A*01:01 and HLA-A*30:01). HLA-A allele HLA-A*30:02, HLA-B allele (HLA-B*08:01) and HLA-C alleles (HLA-C*04:01, HLA-C*06:01 and HLA-C*07:02) bind with intermediate and/or poor affinity to the common South African HLA-I alleles (data not shown) and there were not included for further analysis. For conservation analysis with other genotypes, 9 of 12 epitopes were conserved in one or more genotypes with epitope NS5B^{2763-2771} MTRYSAPP highly conserved in all genotypes (Table 7.5 in the Appendix).

For HLA-II alleles, the HLA-DRB1*15:01 has the highest number of binding epitopes (22) followed by DRB1*0301 and DRB1*1301 (17 each) (Table 7.6 in the Appendix). The highest percentage of optimal binding (61%) was predicted between peptide within NS3^{1415-1423} VAYYRGLDV and the HLA-DRB1*15:01 allele. This epitope is conserved in all genotype except genotype 6. Epitopes NS2^{1327-1335} LGIGTVLDQ, N55B^{1892-1900} LVVGVVCAA, NS5B^{2451-2459} LRHHLTVYS, NS5B^{2696-2704} YRRCRASGV, NS5B^{2883-2891} IQRLHGLSA and NS5B^{2883-2891} variants IQRLHGMAA were predicted to be the highest promiscuous binder covering all 5 HLA-DRB alleles common in South Africa by average percentage score of 26%, 23%, 35%, 20%, 32% and 26% respectively. Epitope NS2^{1327-1335} LGIGTVLDQ is one of the highly conserved epitopes in all genotype. Interestingly, another highly conserved epitopes E1^{507-515} and NS2^{1253-1261} did not bind to any of the HLA-DRB1 allele predicted (Table 6). Epitopes NS3^{1585-1593}, NS5A^{2285-2293} and NS5B^{2889-2297} were predicted to cover both HLA-I and HLA-II alleles.

7.3.3 Validation of epitopes

Seven epitopes predicted epitopes were previously confirmed experimentally by other studies as true positives with 3, 2 and 2 predicted for HLA-I, HLA-II and (both HLA-I and HLA-II) respectively in comparison with the epitopes analysed in the IEDB resource database. Majority of the epitopes predicted in
this study have not been previously tested experimentally. The ‘true epitopes’ are highlighted by (*) in Tables 7.1; 7.2 and 7.5 in the Appendix.

7.4 Discussion

Several studies that have published HCV epitopes focused mainly on genotype 1 (Zhang et al., 2009; Matsueda et al., 2007), but most of these studies do not take into account the diversity in other genotypes that are common in developing countries like most of African countries. In the present study, predicted antigenic epitopes of HCV genotype 5a proteins from South Africa were analysed followed by conservancy with randomly selected genotypes 1-6 references from GenBank. Several studies have confirmed the importance of using immunoinformatics as good predictors for selecting MHC ligands, T-cell epitopes and immunogenicity (De Groot et al., 2002). As a result, several immunoinformatics methods have been developed to assist in the identification of MHC binding peptides (Singh and Raghava, 2001; Singh and Raghava, 2003).

For this analysis, near full-length sequences covering all HCV proteins with the exclusion of the 3’end of the NS5B were included to maximize number of epitopes predicted. The use of the whole viral genome for developing epitope vaccines has a potential control over the immune response and eliminating the side effects (De Groot et al., 2002), and it also increases the chance of detecting a virus at any developmental stage (Toussaint et al., 2008). It has been shown that multiple epitopes from different parts of the HCV genome are important to produce a vaccine that can elicit strong humoral immune responses and multiple specific cellular immune responses (Zeng et al., 2009). A polyepitope-based strategy with multiple components combining core, E1, and E2 proteins; and conserved T-cell epitopes in the NS proteins has been suggested to be a good vaccine candidate for HCV (Hu et al., 2005).
A high number of epitopes was predicted for MHC class II as compared to class I. The findings of this study are consistent with a study by Shehzadi et al (2011) that predicted epitopes in genotype 3 from Pakistan. The study showed that majority of predicted epitopes were found in the NS3 protein for both MHC class I and MHC class II alleles and most of the epitopes were conserved amongst different genotypes (Shehzadi et al., 2011). The NS3 is considered to be a good cellular target candidate for a therapeutic vaccine (Torresi et al., 2011) since majority of the HCV viral epitopes recognized by CD8+ and CD4+ T-cells are located in the NS3 region (Eckels et al., 2002; Day et al., 2002; Hakamada et al., 2004; Wertheimer et al., 2003). The NS3 specific CD4+ and CD8+ T-cell responses were reported in patient responders to interferon therapy (Vertuani et al., 2002) and in spontaneous clearance of HCV (Smyk-Pearson et al., 2006).

Most of the predicted epitopes in the study sequence were found to be conserved across different HCV genotypes with a higher number of conserved epitopes in HCV genotype 5a. Highly conserved epitopes might influence the immunogenic potential since mutations within the epitopes can increase the chance of immune escape (Yusim et al., 2002). For a vaccine to be effective globally the selected epitopes must cover HLAs of different populations and it must also be conserved among different genotypes. The high mutation rates of viral epitopes and HLA polymorphism are some of the challenges that are associated with the development of peptide vaccines (Vider-Shalit et al., 2008). This study attempted to ensure maximal coverage of HLA polymorphism and different genotypes by analyzing conserved epitopes considering different HLA alleles.

Majority of the epitopes predicted from HCV proteins isolated from South African genotype 5 were good binders against MHC alleles that are found worldwide. MHC is both polygenic and polyphormic, and the pool of MHC
molecules differs for every individual. Different MHC alleles bind peptide with a particular sequence pattern (Doytchinova et al., 2006). For an MHC allele to be covered by a set of epitopes, at least one of the epitopes should be capable of inducing an immune response when bound to the corresponding MHC molecule (Toussaint et al., 2008). The epitopes predicted in this study bind to many HLA alleles and can be used for designing good vaccine candidates that will eventually work in genetically diverse population. Very few epitopes were found to be experimentally true positive, however this can be due to the fact that most of the previous studies focused on genotype 1. Limitation of the study was in vivo and in vitro studies to confirm predicted immunogenic epitopes and it will be done in the future. However in silico studies still provide the basis for designing good vaccine candidates.

In conclusion, the results of this study demonstrated antigenic T-cell epitopes that are conserved among genotypes and good MHC binders derived from genotype 5a sequences that can be good candidates for vaccine development. Predicted epitopes analysed in this study will contribute to future design of an efficient vaccine with the use of conserved epitopes to avoid variation in genotypes and as such, it will be able to induce broad HCV specific immune responses. Conserved epitopes among different genotypes will be experimentally tested in the future to show if they are involved in immune response.
CHAPTER EIGHT

CHARACTERISATION OF HCV GENOTYPE 5a ENVELOPE PROTEINS: IMPLICATIONS FOR VACCINE DEVELOPMENT AND THERAPEUTIC ENTRY TARGET.

Maemu P Gedezha, Selokela G Selabe, Thanda Kyaw, M Jeffrey Mphahlele.

Manuscript in preparation.
8 CHARACTERISATION OF HCV GENOTYPE 5a ENVELOPE PROTEINS: IMPLICATIONS FOR VACCINE DEVELOPMENT AND THERAPEUTIC ENTRY TARGET

8.1 Introduction

The E1 and E2 proteins are type I transmembrane proteins with both the N-terminal ectodomain and a C-terminal domain (Chevaliez and Pawlotsky, 2006), and contains 6 and 11 glycosylation sites respectively (Drummer et al., 2003; Goffard and Dubuisson, 2003). These proteins are involved in viral entry by interacting with CD81, SRB1 and HDL (Higginbottom et al., 2000; Scarselli et al., 2002; Mazzocca et al., 2005; Ashfaq et al., 2011b). HCV glycans play an essential role in envelope proteins to ensure correct conformation for virus entry (Goffard and Dubuisson, 2003; Brown et al., 2007) and antigenic variation (Slater-Handshy et al., 2004). HCV E2 glycosylation sites interact with cell surface receptors directly allowing the virus to enter the cell (Pileri et al., 1998; Helle and Dubuisson, 2008). The E2 protein is thought to be primarily responsible for receptor binding (Von Hahn et al., 2007). Glycosylation sites may mask important epitopes from host antibody responses (Schønning et al., 1996; Wei et al., 2003). However, the number of glycosylation sites varies according to genotype (Brown et al., 2007).

The HCV genome harbours a number of epitopes and one of them are B-Cell epitopes (Scarselli et al., 1995; Lee et al., 1999; Giang et al., 2012). B-cell epitopes are essential in increasing the preferred immune responses (Bugli et al., 2001; Giang et al., 2012) and number of epitopes and modulation of immune recognition of antigens can be influenced by deglycosylation of E1 proteins (Fournillier et al., 2001). The E1 derived peptide p35 (aa 315–323) (El-Awady et al., 2006), E2-conserved synthetic peptides p37 (a.a 517–531) and p38 (aa 412–419) have been reported to neutralise HCV particles, as a results, they are important components of a candidate peptide vaccine (El-Awady et al., 2009). The molecular targets for current HCV DAA in
development are mainly focused on the NS proteins such as the NS3 protease, NS5A and the NS5B RdRp (Baldick et al., 2010). Recently, considerable progress has been made in understanding HCV entry (Dorner et al., 2011; Sainz et al., 2012) and the development of entry inhibitors (Baldick et al., 2010; Ciesek et al., 2011; Keck et al., 2011; Dorner et al., 2011, Mittapalli et al., 2012, Zhu et al., 2012). Many patients do not respond to the current available therapy as discussed in section 2.9, therefore there is an urgent need to develop effective HCV vaccine and specific therapeutic drugs. While both E1 and E2 are hypervariable in nature, it is difficult to design vaccines or therapeutic drugs against them. Therefore, the objective of this chapter was to characterise genotype 5a E1 and E2 sequences to determine conserved B-cell epitopes and peptides in HCV that could be useful targets in the design of vaccine and entry inhibitors.

8.2 Methods

8.2.1 Amplification and Sequence retrieval

For amplification, 12 genotype 5a E1 and E2 regions were amplified by fragment B-D and sequenced as described in Appendix A together with 6 sequences from the near full-length described in Chapter 6. Additionally, a total of 224 HCV E1 and ten E2 sequences belonging to genotype 5a were retrieved from the NCBI database. All sequences were aligned by Mafft (mafft.cbrc.jp/alignment/server/) and translated into amino acids using BioEdit (Hall, 1999).

8.2.2 Prediction of B-cell epitopes

For identification of B-cell epitopes, the E1 and E2 region of sequence ZADGM2088 was randomly selected for prediction of 16-mers B cell epitopes using the program ABCpred as described in section 3.17.
8.2.3 Antigenecity of the epitopes
The Antigenecity of all the predicted epitopes were analysed as described in section 3.18.

8.2.4 Epitope conservancy analysis
All predicted epitopes were analysed for conservation as described in section 3.19.

8.2.5 Immune selection pressure
For positive selection, 18 E1 and E2 sequences were analysed. The individual codons under positive selection were detected via the fixed effects likelihood (FEL) method as as described in section 3.21.

8.2.6 Analysis of N-linked glycosylation sites
The N glycosylation sites were predicted for the 18 E1 and e2 sequences using the on-line prediction server NetNGlyc version 1.0 as described in section 3.22.

8.2.7 Peptide design
The sequences of genotype 5a E1 and E2 regions from this study were aligned with reference sequences from different parts of the world to identify the global consensus sequence. Consensus sequences of E1 and E2 sequences were created using BioEdit. Peptide design was performed as described in section 3.23.
8.3 Results

8.3.1 B-cell epitopes prediction

A total of 26 B cell epitopes that were antigenic were predicted in both the E1 and E2 regions (Table 8.1 in the appendix). In the E1, there were 9 B cell epitopes predicted while in the E2, 17 B cell epitopes were predicted. Epitope E2-GERCDLEDRAELSP had the highest antigenic score of 1.754 while E2-GPVYCFTPSPVVGTT and LSTGLIHLHQIVDTQ were 70% conserved in all genotypes used for the analysis. More than 80% of the predicted epitopes were conserved in genotype 5a sequences.

8.3.2 Determination of positive selection

A selection analysis was performed to ascertain which of the HCV E1 and E2 amino acids were subjected to positive selection. In the E1, there were two positively selected amino acids and in the E2 protein, there were five positively selected amino acids with a significance value of p<0.05. Four of the positively selected sites resided within the B cell epitopes (Table 8.2).

Table 8.2: Positively selected amino acid sites with a significance value of p<0.05 in the E1 and E2 regions and their positions in the B cell epitopes

<table>
<thead>
<tr>
<th>Positive selected sites $^a$</th>
<th>B cell epitope $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E1</strong></td>
<td></td>
</tr>
<tr>
<td>331 MMMNWSPPTALLMAQL</td>
<td></td>
</tr>
<tr>
<td>372 AYYASAANWAKVALVL</td>
<td></td>
</tr>
<tr>
<td><strong>E2</strong></td>
<td></td>
</tr>
<tr>
<td>384 -</td>
<td></td>
</tr>
<tr>
<td>395 QTTGGSAARNVGYGTS</td>
<td></td>
</tr>
<tr>
<td>444 -</td>
<td></td>
</tr>
<tr>
<td>532 TYNWGSNETDFLNN</td>
<td></td>
</tr>
<tr>
<td>597 -</td>
<td></td>
</tr>
</tbody>
</table>

Key: $^a$Numbering is based on the M62321 full-length sequence, $^b$ Indicates the epitope where the positive selection sites where found
8.3.3 Analysis of E1 and E2 N-linked glycosylation

The E1 and E2 proteins of the 18 sequences were analysed for possible glycosylation sites. Differences in the probability of glycosylation in the E1 and E2 were observed in most sequences. Whereas other studies have reported five N-linked glycosylation sites in the E1 region, all strains in the current study showed three or four glycosylation sites, with the exception of ZADGM1908 and ZADGM2088 which showed 2 glycosylation sites, with the N325 site absent from all sequences. In the E2 region three sequences (ZADGM1104, ZADGM1707 and ZADGM3013) showed 9 glycosylation sites, while the remaining had variations in the number of glycosylation sites. In ZADGM308, position N430 was replaced by H, while in ZADGM6544, N448 was replaced by D. Site N476 was found in only 6 of the 18 sequences analysed. The E2 sites N423 and N576 were absent in all genotype 5 sequences from this study (Table 8.3).
Table 8.3: Probability of glycosylation in E1 and E2 sequences

<table>
<thead>
<tr>
<th>Isolate no</th>
<th>Probability at glycosylation site **</th>
<th>E1</th>
<th>No of sites</th>
<th>E2</th>
<th>No of sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>196 209 234 305</td>
<td>417 430 448 476 533 541 557 623 645</td>
<td></td>
</tr>
<tr>
<td>ZADGM308</td>
<td></td>
<td>+</td>
<td>++ ++ ++ _ _</td>
<td>++ _ _ ++ _ _ ++ ++ + + _ _</td>
<td>6</td>
</tr>
<tr>
<td>ZADGM0518</td>
<td></td>
<td>+</td>
<td>++ ++ + _ _</td>
<td>++ ++ ++ _ _ ++ ++ + + _ _</td>
<td>6</td>
</tr>
<tr>
<td>ZADGM525gp</td>
<td></td>
<td>_</td>
<td>++ ++ ++ + _</td>
<td>++ + _ _ _ _ ++ + + + _</td>
<td>6</td>
</tr>
<tr>
<td>ZADGM651</td>
<td></td>
<td>+</td>
<td>++ ++ ++ _ +</td>
<td>++ + _ _ _ _ ++ + + + _ _</td>
<td>6</td>
</tr>
<tr>
<td>ZADGM869</td>
<td></td>
<td>+</td>
<td>++ ++ + _ _</td>
<td>++ ++ _ _ _ _ ++ + + + _ +</td>
<td>7</td>
</tr>
<tr>
<td>ZADGM2352</td>
<td></td>
<td>+</td>
<td>++ ++ + _ _</td>
<td>++ ++ _ _ _ _ ++ + + + _ _</td>
<td>6</td>
</tr>
<tr>
<td>ZADGM1104</td>
<td></td>
<td>+</td>
<td>++ ++ ++ _ _</td>
<td>++ ++ + + + + + + + + _ _</td>
<td>9</td>
</tr>
<tr>
<td>ZADGM1707</td>
<td></td>
<td>+</td>
<td>++ ++ + _ _</td>
<td>++ + _ _ _ _ ++ + + + _ _</td>
<td>9</td>
</tr>
<tr>
<td>ZADGM2439</td>
<td></td>
<td>+</td>
<td>++ + _ _ _</td>
<td>++ + _ _ _ _ ++ + + + _ _</td>
<td>8</td>
</tr>
<tr>
<td>ZADGM2582</td>
<td></td>
<td>+</td>
<td>++ ++ + _ _</td>
<td>++ ++ _ _ _ _ ++ + + + _ _</td>
<td>9</td>
</tr>
<tr>
<td>ZADGM3013</td>
<td></td>
<td>+</td>
<td>++ ++ + _ _</td>
<td>++ ++ ++ ++ _ _ ++ + + + _ _</td>
<td>9</td>
</tr>
<tr>
<td>ZADGM4124</td>
<td></td>
<td>+</td>
<td>++ ++ + _ _</td>
<td>++ _ _ _ _ _ _ + _ _ _ _</td>
<td>7</td>
</tr>
<tr>
<td>ZADGM4227</td>
<td></td>
<td>+</td>
<td>++ ++ + _ _</td>
<td>++ ++ _ _ _ _ _ _ + _ _ _ _</td>
<td>6</td>
</tr>
<tr>
<td>ZADGM6485</td>
<td></td>
<td>+</td>
<td>++ ++ + _ _</td>
<td>++ ++ _ _ _ _ _ _ ++ + + + _ _</td>
<td>7</td>
</tr>
<tr>
<td>ZADGM6544</td>
<td></td>
<td>+</td>
<td>++ + _ _ _ _</td>
<td>++ ++ _ _ _ _ ++ + + + _ _</td>
<td>6</td>
</tr>
<tr>
<td>ZADGM7890</td>
<td></td>
<td>+</td>
<td>++ ++ + _ _</td>
<td>++ ++ _ _ _ _ _ _ ++ + + + _ _</td>
<td>7</td>
</tr>
<tr>
<td>ZADGM9150</td>
<td></td>
<td>+</td>
<td>++ ++ + _ _</td>
<td>++ ++ _ _ _ _ _ _ ++ + + + _ _</td>
<td>7</td>
</tr>
<tr>
<td>ZADGM2088</td>
<td></td>
<td>-</td>
<td>++ ++ _ _ _</td>
<td>++ + _ _ _ _ _ _ + _ _ _ _</td>
<td>6</td>
</tr>
</tbody>
</table>

Key, * Numbering is based on the M62321 full-length sequence, ^ Glycosylation probability is shown by +++ (probability > 70%), ++ (probability between 60 and 70%), + (probability between 50 and 60%), and - (not present).
8.3.4 Peptide design

Conserved peptides from genotype 5a were derived from both regions of the E1 and E2. For the E1, three peptides of 8-11 aa’s in length were derived and in the E2, four peptides of 8-16 aa’s were derived. Instability index indicated that majority of the peptides predicted (5 out of 8) were not stable. Only one of the peptides-CNCSIYSGH had post-translation modification which is the N-linked glycosylation, although at a low probability. Peptides YHTNDCPNSSI and VDLAGGA were found to be the best predicted peptides that could be useful for designing entry inhibitors (Table 8.4).
Table 8.4: Predicted peptides for HCV E1 and E2 conserved in genotype 5a

<table>
<thead>
<tr>
<th>Position</th>
<th>Peptides</th>
<th>Length</th>
<th>Molecular weight</th>
<th>Theoretical PI</th>
<th>Extinction coefficient (cm⁻¹ M⁻¹)</th>
<th>Instability Index</th>
<th>Alphatic index</th>
<th>GRAVY</th>
<th>Composition of hydrophobic AA’s</th>
<th>N-linked glycosylation</th>
<th>N-linked Phosphorylation</th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>YHTNDCPNSSI</td>
<td>11</td>
<td>1250.3</td>
<td>5.08</td>
<td>1490</td>
<td>34.81</td>
<td>35.45</td>
<td>-1.082</td>
<td>9%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>304</td>
<td>CNCSIYSGH</td>
<td>9</td>
<td>983</td>
<td>6.72</td>
<td>1615</td>
<td>5.69</td>
<td>43.33</td>
<td>-0.056</td>
<td>0%</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>318</td>
<td>MAWDMMMNWSP</td>
<td>11</td>
<td>1399.6</td>
<td>3.8</td>
<td>11000</td>
<td>45.59ₐ</td>
<td>9.09</td>
<td>-0.164</td>
<td>60%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>262</td>
<td>VDYLAGGA</td>
<td>8</td>
<td>764.8</td>
<td>3.8</td>
<td>1490</td>
<td>-5.22</td>
<td>110</td>
<td>0.75</td>
<td>50%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>502</td>
<td>VCGPYYYCTPSVVVG</td>
<td>16</td>
<td>1623.9</td>
<td>5.48</td>
<td>1615</td>
<td>62.93ₐ</td>
<td>90.62</td>
<td>1.275</td>
<td>37.5%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>482</td>
<td>DKPYCWHY</td>
<td>8</td>
<td>1111.2</td>
<td>6.73</td>
<td>8480</td>
<td>102.84ₐ</td>
<td>0.00</td>
<td>-1.65</td>
<td>12.5%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>491</td>
<td>PRPGGVVPA</td>
<td>9</td>
<td>849</td>
<td>10.18</td>
<td>-</td>
<td>78.4</td>
<td>75.5</td>
<td>0.011</td>
<td>33.3%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>418</td>
<td>GSWHINRTALNC</td>
<td>12</td>
<td>1371</td>
<td>8.26</td>
<td>5500</td>
<td>63.12</td>
<td>73.3</td>
<td>-0.408</td>
<td>36.4%</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Key: ₐ Numbering is based on the M62321 full-length sequence, ᵇ indicates unstable peptides, çı list of hydrophobic aa’s (Leu, Val, Ile, Met, Phe, and Trp), ᵈ Glycosylation probability is shown by +++ (probability > 70%), ++ (probability between 60 and 70%), + (probability between 50 and 60%), and - (not present).
8.4 Discussion

HCV E1 and E2 are highly variable proteins which are the main target for neutralizing antibody and are also required for viral entry. Therefore, this study was designed to identify the conserved sequences of these proteins in order to predict the antigenic epitopes and peptides that could serve as the best targets for vaccine design and potential entry inhibitors. The use of different structural and sequence analyses tools helped with the in silico analysis for the E1 and E2 regions.

The E2 is the main target for neutralizing antibody responses and the variation of this region is thought to be related to the maintenance of persistent infection by emerging escape variants and subsequent development of chronic infection (Manzin et al., 1998; Farci et al., 2000). Recently, a linear region of E2 encompassing amino acids 434 to 446 has been reported to elicit non-neutralizing antibodies that can inhibit the neutralizing activity of antibodies targeting aa’s 412 to 423 (Zhang et al., 2009). However, a study by Tarr et al (2012) reported conflicting results that showed that human antibodies that target the region encompassing aa’s 434 to 446, are not inhibitory but are instead capable of neutralizing HCVpp and HCVcc entry (Tarr et al., 2012). All the B-cell epitopes included in this study were found to be antigenically effective by analyzing them using online softwares, and it can be implied that these epitopes may be important for inducing the desired immune response.

Variations in regions of high variability may be induced by selective forces or may arise due to the fact that some viral genomic regions are simply functionally unconstrained (Humphreys et al., 2009). An excess of non-synonymous substitutions is attributed to positive selection, whereas an excess of synonymous substitutions is considered to result from negative selection (Ray et al., 2005). For this study, 4 of the positively selected amino acid sites were located in B-cell epitopes. The positively selected sites may be vaccine targets since it has also been reported that the sites within the
epitopes should be highly immunogenic and it has been shown to be involved in the clearance of HCV 1b *in vivo* (Suzuki et al., 2001).

The glycosylation sites are highly conserved among different genotypes (Brown et al., 2007). In this study, analysis of the N-linked glycosylation sites revealed that genotype 5a sequences were not conserved in that in most cases there were variations in glycosylation sites for the study sequences. Site N476 which showed a level of conservation of 75% among different genotypes but was absent from the sequences of genotype 5a (Goffard and Dubuisson, 2003), was found in six of the 18 sequences analysed. As reported previously, E2 sites N423 and N576 were absent in all genotype 5a sequences including the 18 sequences from this study, which is notable because these two sites have been reported to be 99-100% conserved across all genotypes (Goffard and Dubuisson, 2003).

Short polypeptides derived from viral envelope sequences of other viruses have been used to investigate protein interactions involved in viral entry and some antiviral agents have been successfully developed (Greenberg and Cammack, 2004). Envelope protein peptide inhibitors for other viruses in the same family with HCV like Dengue and West Nile have been shown to inhibit viral entry (Hrobowski et al., 2005; Bai et al., 2007; Alhoot et al., 2013). In HCV, the post-binding entry step was shown to be prevented by the use of peptides derived from the C terminal region of E2, which plays an important role in the HCV entry process (Liu et al., 2010). For this study, conserved peptides have been derived that can possibly be used as targets for therapeutic purposes. The use of HCV glycoproteins in therapeutic strategies may offer protection against HCV infection (Beyene et al., 2002).

In conclusion, this study provides conserved B cell epitopes and peptides that can be useful for designing entry inhibitors and vaccines that will be able to cover a global population, especially in places where genotype 5a is common. Future investigations will look to further analyse these peptides in order to better understand their involvement in blocking HCV entry.
CHAPTER NINE

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

This chapter summarizes the discussion, conclusion and recommendations
9 COMBINED DISCUSSION, CONCLUSION AND RECOMMENDATIONS

9.1 Discussion

In South Africa, the prevalence of HCV is estimated at less than 2% in the general population and data for HIV/HCV co-infection is scarce. As a result, much is still not known about HCV prevalence in HIV infected patients. HCV antiviral therapy is increasingly becoming available in South African tertiary public hospitals through public-private partnerships, yet there is paucity of data about HCV diversity in South Africa as many studies have focused on the 5’ UTR and recently one on NS5B region but less emphasis has been put on other regions e.g. the envelope and full-length genomes. Hence, this study undertook: (1) To investigate HCV co-infection in HIV-positive patients enrolling for HAART (2) To investigate the distribution of HCV genotypes (3) To characterise near full-length genome of HCV genotype 5 strains (4) To predict T cell-epitopes in HCV genotype 5a sequences (5) To characterise genotype 5a E1 and E2 sequences.

This study investigated HCV co-infection in HIV-positive patients enrolling for HAART. A total of 653 serum samples collected from 2004 to 2006 were screened for anti-HCV. Serological results indicated that only 1.2% (8/653) of samples were positive for anti-HCV, with S/CO values ranging from 5.69 to 37.8. Of these 8 samples, HCV RNA was detected in only one, which had the highest anti-HCV titre of 37.8. Sequencing confirmed that the RT-PCR product was HCV specific. In contrast, 20.7% (135/653) of samples were preliminary positive, with S/CO values ranging from 1.02 to 4.48. None of these samples tested positive for HCV RNA by RT-PCR. It is possible that the low prevalence of HCV antibodies in our study could be due to missed HCV antibodies as many of HIV/AIDS patients fail to generate antibodies due to immunosuppression (Sorbi et al., 1996).

To investigate the distribution of HCV genotypes in HIV positive and HIV negative patients, a total of 78 serum samples from 71 patients who were anti-HCV positive were genotyped by the use of the 5’UTR, C/E1, and NS5B
regions. The HCV genotype distribution in the studied population was as follows: 59.2% (42/71) were genotype 5, 18.3% (13/71) were genotype 1, 14.1% (10/71) were genotype 4 and 4.2% (3/71) were genotype 3. Three of 71 (4.2%) individuals were infected with mixed genotypes based on the 5'UTR. The HIV positive patient was infected by genotype 1b. Phylogenetic analysis of the 5'UTR was accurate in determining genotypes, while the C/E1 and NS5B coding region allowed for differentiation of both genotypes and subtypes, including the recombinant strains. Furthermore, we observed the existence of distinct variants of HCV which were divergent from confirmed genotype 4 subtypes. The response to HCV treatment is partially dependent on the infecting genotype. Recombination plays a significant role in the evolution of RNA viruses by creating genetic variation, reducing mutational load, and creating viruses with new properties (Worobey et al., 1999), and this might lead to the virus escaping immune response and also not responding well to treatment.

To characterise near full-length genome analysis of HCV genotype 5 strains, near full-length genomes of approximately 8000 nt were amplified and sequenced from six HCV genotype 5 samples, as well as the E1 region from an additional 12 genotype 5 samples. In the results phylogenetic analysis indicated that this study concurred with previous reports that demonstrated that genotype 5 is classified into a single subtype (5a). Bayesian coalescent analysis of the E1 showed a tMRCA ranging from 116 years for the lognormal clock to 134 years for the exponential clock. The tMRCA for genotype 5a infections in this study is similar to that reported by Verbeeck et al. in Belgian and South African patients. Initially, genotype 5 was thought to have originated in South Africa; however, Verbeeck et al., 2006, speculated that it may have originated in Central Africa and spread through travelling between South Africa and Central Africa (Verbeeck et al., 2006). Very little information is known about the origin, epidemic history, disease severity, and treatment outcome of HCV genotype 5a. This study provided information on the epidemic history of this rare genotype.
To predict T-cell epitopes in HCV genotype 5a sequences at DGMAH, conserved epitopes that recognise both MHC I and II were predicted in near-full length HCV genotype using a suite of online programmes. A total of 24 and 77 antigenic epitopes that recognize HLA I and HLA II respectively were predicted. For conservation analysis 8 and 31 predicted epitopes were conserved in different genotypes for HLA-I and HLA-II alleles respectively. Several epitopes bind with high affinity for HLA-I alleles and high percentage for HLA-II common in South Africa. Advantages of using the whole viral genome for developing epitope vaccines include the potential for control over the immune response and eliminating side effects (De Groot et al., 2002). It also increases the chance of detecting a virus at any developmental stage (Toussaint et al., 2008). Most epitope vaccines are often designed based on genotype 1 sequence and due to the high variability of HCV genotypes such vaccines may not protect patient infected by other genotypes.

To characterise genotype 5a E1 and E2 sequences at DGMAH, conserved B-cell epitopes and peptides in HCV were determined that could be useful targets in the design of vaccines and entry inhibitors. For the E1 and E2 sequences, a total of 26 B cell epitopes that were antigenic were predicted in both the E1 and E2 regions and more than 80% of the predicted epitopes were conserved in genotype 5a sequences. For amino acids undergoing positive selection, 4 of the positively selected sites resided within the B cell epitopes. The E1 and E2 proteins of the 18 sequences were also analysed for possible glycosylation sites and differences in the probability of glycosylation in the E1 and E2 were observed in most sequences. Three and four conserved peptides of 8-16 aa were also derived from the E1 and E2 respectively. The E2 protein is the main target for neutralizing antibody responses, and mutations in this region are associated with development of chronic infection (Farci et al., 2000). This region is also involved in viral entry during the infection process (Scarselli et al., 2002; Mazzocca et al., 2005).
9.2 Conclusion

This study shows that the majority of HIV/AIDS patients initiating HAART at DGMAH have low exposure to, or active, HCV infection and it does not appear that HIV/AIDS is a risk factor for increased detection of HCV co-infection. Genetic diversity results indicate that genotype 5 is still the most predominant genotype in South Africa, although, genotypes 1 and 4 are also encountered in a significant proportion of the population including the existence of distinct variants of HCV which were divergent from confirmed genotype 4 subtypes. Our findings clearly demonstrate that for the first time in South Africa, we identified the presence of HCV subtypes 4k, 4q and 4r, as well as evidence of intra-genotypic recombination (4q/4l within NS5B). This study also provided six near full-length nucleotide sequences of HCV genotype 5a strains and it concurred with previous reports that demonstrated that genotype 5 is classified into a single subtype (5a). Accounting for these sequences, genotype 5 to date has 8 complete or nearly complete coding sequences and it offers significant insights for understanding the diversity of genotype 5 in the South African population and can be used as references to design efficient vaccines and for the development of new antiviral drugs. Finally, this study demonstrated antigenic B- and T-cell epitopes that were conserved among genotypes and these can be good candidates for vaccine development. Conserved peptides were also identified which can be useful for designing entry inhibitors. This study has described HIV/HCV co-infections, added information regarding circulating HCV genotypes and the diversity of genotype 5a and their epidemic history and finally immunological determinants of HCV genotype 5a.

9.3 Recommendations

While our findings do not support mandatory HCV screening in HIV/AIDS patients initiating HAART in South Africa, consideration should be given to patients who may be at an increased risk such as HIV positive haemophiliacs, injection drug users and diabetics. Further investigations to monitor circulating HCV genotypes in the South African population are required. Thus,
continuous characterization of circulating genotypes will aid in monitoring new strains, therapeutic settings and further investigate the possibilities that lead to mixed genotypes and recombinants. The production of neutralising antibodies that are able to interfere with receptor binding could be beneficial in blocking entry. In future, we aim to conduct laboratory-based studies of these conserved epitopes and peptides so as to better understand their involvement in immune response and blocking HCV entry.

9.4 Limitations of the study

Due to the variability of HCV genotypes we failed to amplify all samples with the 3 regions recommended for genotyping and more full-length genomes for genotype 5. Due to the time constraints, full-length genome analysis of the genotype 4 outlier group was not achieved to explore whether the finding truly represents an HCV intrasubtypic recombinant or a new HCV subtype.
CHAPTER TEN

REFERENCES

This chapter includes all the references cited in the study
10 REFERENCES


Ashfaq UA, Qasim M, Yousaf MZ, Awan MT, Jahan S. Inhibition of HCV 3a genotype entry through host CD81 and HCV E2 antibodies. J Transl Med. 2011b; 9:194.


Fournillier A, Wychowski C, Boucreux D, Baumert TF, Meunier JC, Jacobs D, Muguet S, Depla E, Inchauspé G. Induction of hepatitis C virus E1 envelope


Li C, Lu L, Wu X, Wang C, Bennett P, Lu T, Murphy D. Complete genomic sequences for hepatitis C virus subtypes 4b, 4c, 4d, 4g, 4k, 4l, 4m, 4n, 4o, 4p, 4q, 4r and 4t. J Gen Virol. 2009; 90(Pt 8):1820-6.


Magiorkinis G, Magiorkinis E, Paraskevis D, Ho SY, Shapiro B, Pybus OG, Allain JP, Hatzakis A. The global spread of hepatitis C virus 1a and 1b: a


possessing the capacity to induce cytotoxic T-lymphocytes in HCV1b-infected patients with HLA-A11, -A31, and -A33. Cancer Immunol Immunother. 2007; 56(9):1359-66.


SADOH. South African department of health. 2008 National Antenatal Sentinel


Welsch C, Albrecht M, Maydt J, Herrmann E, Welker MW, Sarrazin C, Scheidig A, Lengauer T, Zeuzem S. Structural and functional comparison of


CHAPTER ELEVEN

APPENDIX

This chapter includes tables and publications
## 11 APPENDIX

### 11.1 Tables

Table 3.1: Sequences of the HCV primers used in the study

<table>
<thead>
<tr>
<th>Fragment</th>
<th>Name</th>
<th>Position</th>
<th>Sequence 5'-3'</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'UTR</td>
<td>P939</td>
<td>349</td>
<td>TA C TCG AGG TCC ACG GTC TA C GAG ACC T</td>
<td>Sithebe et al., 1998</td>
</tr>
<tr>
<td></td>
<td>P209</td>
<td>45</td>
<td>CTG TGA GGA ACT ACT GTC TT</td>
<td>Sithebe et al., 1998</td>
</tr>
<tr>
<td></td>
<td>P940</td>
<td>60</td>
<td>TTC ACG CAG AAA GCG TCT AG</td>
<td>Sithebe et al., 1998</td>
</tr>
<tr>
<td></td>
<td>P211</td>
<td>313</td>
<td>CAC TCT CGA GCA CCC TAT CAG GCA GT</td>
<td>Sithebe et al., 1998</td>
</tr>
<tr>
<td>C/E1</td>
<td>DM111</td>
<td>834</td>
<td>GCA ACA GGG AAY YTD CCY GGT TGC TC</td>
<td>Murphy et al., 2007b</td>
</tr>
<tr>
<td></td>
<td>DM110</td>
<td>1327</td>
<td>GTR GGN GAC CAR TTC ATC ATCA</td>
<td>Murphy et al., 2007b</td>
</tr>
<tr>
<td></td>
<td>DM109</td>
<td>843</td>
<td>AAY YTD CCC GGT TGC TCT TTY TCT AT</td>
<td>Murphy et al., 2007b</td>
</tr>
<tr>
<td></td>
<td>DM108</td>
<td>1315</td>
<td>TTC ATC ATC ATR TTC CAN GCC AT</td>
<td>Murphy et al., 2007b</td>
</tr>
<tr>
<td>NS5B</td>
<td>NS5B1s</td>
<td>8265</td>
<td>TAT GAY ACC CGC TGY TTT GAC</td>
<td>Djebi et al., 2004</td>
</tr>
<tr>
<td></td>
<td>NS5B2as</td>
<td>8650</td>
<td>GARTACCTRGTACTAGCCTC</td>
<td>Djebi et al., 2004</td>
</tr>
<tr>
<td></td>
<td>NS5B3s</td>
<td>8276</td>
<td>CTGYTTGACTCMACRGTCAC</td>
<td>Djebi et al., 2004</td>
</tr>
<tr>
<td></td>
<td>NS5B4as</td>
<td>8639</td>
<td>CAT AGC CTC CGT GAA GRC TC</td>
<td>Djebi et al., 2004</td>
</tr>
<tr>
<td>A</td>
<td>P939</td>
<td>45</td>
<td>TA CTC GAG GTC CAC GGT CTA CGA GAC CT</td>
<td>Sithebe et al., 1998</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>M101</td>
<td>1303</td>
<td>AGG GGA CCA ATT CAT CAT CAT G</td>
<td>Koletzki et al., 2009</td>
<td></td>
</tr>
<tr>
<td>P940</td>
<td>60</td>
<td>TTC ACG CAG AAA GCG TCT AG</td>
<td>Sithebe et al., 1998</td>
<td></td>
</tr>
<tr>
<td>M102</td>
<td>1298</td>
<td>GAC CAT TTC ATC ATC ATG TCC CA</td>
<td>Koletzki et al., 2009</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>F1A</td>
<td>1088</td>
<td>GAC CAT TTC ATC ATG TCC CA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>R1A</td>
<td>1425</td>
<td>TGT ATG CGG CGG CGA ACA AGA CC</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>F2A</td>
<td>1113</td>
<td>CTT CGG AGG GCC GTT GAC TAC TTA GCG</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>R2A</td>
<td>1413</td>
<td>CGA ACA AGA CCC CCC AGT GGG</td>
<td>This study</td>
</tr>
<tr>
<td>C</td>
<td>M105</td>
<td>1292</td>
<td>ATG GCA TGG GAC ATG ATG ATG</td>
<td>Koletzki et al., 2009</td>
</tr>
<tr>
<td></td>
<td>R1B</td>
<td>2061</td>
<td>TAG GCC CTA AGT TGC AGG GTG GA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>M106</td>
<td>1298</td>
<td>TGG GAC ATG ATG AAT TGG</td>
<td>Koletzki et al., 2009</td>
</tr>
<tr>
<td></td>
<td>R2B</td>
<td>2022</td>
<td>CAA ACC CTG TGG AAT TCA TCC AG</td>
<td>This study</td>
</tr>
<tr>
<td>D</td>
<td>F1C</td>
<td>1743</td>
<td>GGC TGG GGA ACT ATC AGC TAT</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>R1C</td>
<td>2636</td>
<td>AAA CCC ATG AGT CCC CGC AGC C</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>F2C</td>
<td>1773</td>
<td>TCG GGC CCC AGT GAT GAC AAG</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>R2C</td>
<td>2612</td>
<td>AGC CGC GTT TAG GAC AAT GAC GTT CT</td>
<td>This study</td>
</tr>
<tr>
<td>E</td>
<td>P101F</td>
<td>2390</td>
<td>GCT CTT AGC ACT GGT CTT ATA CA</td>
<td>Li et al., 2009</td>
</tr>
<tr>
<td></td>
<td>P106R</td>
<td>4036</td>
<td>GGA GCA TGG AGG TGC CCC AC</td>
<td>Li et al., 2009</td>
</tr>
<tr>
<td></td>
<td>P102F</td>
<td>2420</td>
<td>ATA GTA GAC ACC CAA TAC CTC TA</td>
<td>Li et al., 2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>P105R</td>
<td>3960</td>
<td>ACA GGT CAG CGC ATC GTG GTC TCC AGG</td>
<td>Li et al., 2009</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>P103F</td>
<td>3624</td>
<td>GGA CCD GTC AAY CAR ATG TAC AC</td>
<td>Li et al., 2009</td>
</tr>
<tr>
<td></td>
<td>P107R</td>
<td>4786</td>
<td>GCR TCG TGG GGS AOW GTR GTR GTC TC</td>
<td>Li et al., 2009</td>
</tr>
<tr>
<td></td>
<td>P104F</td>
<td>3648</td>
<td>AAT GTA GAC CAA GAC CTR GTG GG</td>
<td>Li et al., 2009</td>
</tr>
<tr>
<td></td>
<td>P108R</td>
<td>4781</td>
<td>YTG GGG RAY AGT RGT RGT CTC AAT</td>
<td>Li et al., 2009</td>
</tr>
<tr>
<td>G</td>
<td>M107</td>
<td>4515</td>
<td>ATC TTC TGC CAC TCT AAG AAA AAG</td>
<td>Koletzki et al., 2009</td>
</tr>
<tr>
<td></td>
<td>F2AS</td>
<td>8035</td>
<td>AAV ACC TCG TTT TTD GCC ATG ATG</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>M108</td>
<td>4521</td>
<td>TGC CAT CCT AAG AAA AAG TGT GA</td>
<td>Koletzki et al., 2009</td>
</tr>
<tr>
<td></td>
<td>F7AS</td>
<td>7615</td>
<td>TAW GAC ATK GAG CAG CGS ACG TTC TCC</td>
<td>This study</td>
</tr>
<tr>
<td>H</td>
<td>595S</td>
<td>6069</td>
<td>GAG GGG GCN GTN CAG TGG ATG AA</td>
<td>de Mello Malta et al., 2010</td>
</tr>
<tr>
<td></td>
<td>M110</td>
<td>8615</td>
<td>AGC GGA GTA CCT GGT CAT AGC CTC ATG GAA</td>
<td>Koletzki et al., 2009</td>
</tr>
<tr>
<td></td>
<td>F1S</td>
<td>6141</td>
<td>TAC GTG CCC GAG AGC GAT GC</td>
<td>de Mello Malta et al., 2010</td>
</tr>
<tr>
<td></td>
<td>M109</td>
<td>8610</td>
<td>AGT ACC TGG TCA TAG CCT CAG TGA AGG CTC</td>
<td>Koletzki et al., 2009</td>
</tr>
<tr>
<td>I</td>
<td>M111</td>
<td>8245</td>
<td>TGG CTT TCT CCT ATG ATA CCC GCT GCT TTG A</td>
<td>Koletzki et al., 2009</td>
</tr>
<tr>
<td></td>
<td>M114</td>
<td>9072</td>
<td>AAA GGG GGT ACC CCA AGC TTC CTG</td>
<td>Koletzki et al., 2009</td>
</tr>
<tr>
<td></td>
<td>M112</td>
<td>8253</td>
<td>TCC TAT GAT ACC CGC TGC TTT GAC TCA AC</td>
<td>Koletzki et al., 2009</td>
</tr>
<tr>
<td></td>
<td>M113</td>
<td>9067</td>
<td>CTC TCA AAG GGG GTA CCC CAA GCT</td>
<td>Koletzki et al., 2009</td>
</tr>
</tbody>
</table>

178
<table>
<thead>
<tr>
<th></th>
<th>Poly(A)</th>
<th></th>
<th>AAG CAG TGG TAT CAA CGA GTA CAA-AAA AAA AAA A AAA AAA AAA AAA AVN</th>
<th>Li et al., 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>P109F</td>
<td>8964</td>
<td>ACT CCG CTG CAT CTT CCA GCT A</td>
<td>This study</td>
</tr>
<tr>
<td>NUP</td>
<td>P110F</td>
<td>9003</td>
<td>GGA CTT AGC GCG TTT TCA CT</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9236</td>
<td>AAG CAG TGG TAT CAA CGC AGA GT</td>
<td>Li et al., 2009</td>
</tr>
</tbody>
</table>
Table 7.1: HLA class I predicted epitopes of HCV genotype 5a and their Antigenicity prediction score, number of allele and conservation (in percentage) in different genotypes.

<table>
<thead>
<tr>
<th>Position</th>
<th>Epitope sequence</th>
<th>No of allele</th>
<th>Antigenecity score</th>
<th>Genotype 1a</th>
<th>Genotype 1b</th>
<th>Genotype 2</th>
<th>Genotype 3</th>
<th>Genotype 4</th>
<th>Genotype 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>651</td>
<td>RCDLEDRDR</td>
<td>9</td>
<td>2.7428</td>
<td>46</td>
<td>55</td>
<td>10</td>
<td>9</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>677</td>
<td>CSFTPTPAL</td>
<td>27</td>
<td>0.8772</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>684</td>
<td>ALSTGLIHL*</td>
<td>29</td>
<td>0.9005</td>
<td>70</td>
<td>99</td>
<td>0</td>
<td>97</td>
<td>84</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>(Himoudi et al., 2002; Engler et al., 2004; Ohno et al., 2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>861</td>
<td>VPPLQVRGG</td>
<td>6</td>
<td>1.4312</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>NS3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1032</td>
<td>YAQQTRGVL</td>
<td>26</td>
<td>0.7638</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1033</td>
<td>AQQTRGVLG</td>
<td>5</td>
<td>0.7751</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1128</td>
<td>ADLYLVTRH</td>
<td>5</td>
<td>1.0277</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>1325</td>
<td>TILGIGTVL</td>
<td>30</td>
<td>0.7555</td>
<td>0</td>
<td>53</td>
<td>88</td>
<td>6</td>
<td>82</td>
<td>16</td>
</tr>
<tr>
<td>1332</td>
<td>VLDQAETAG</td>
<td>5</td>
<td>0.5004</td>
<td>96</td>
<td>98</td>
<td>96</td>
<td>100</td>
<td>93</td>
<td>71</td>
</tr>
<tr>
<td>1357</td>
<td>TPHPNIEEV</td>
<td>24</td>
<td>0.8423</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>0</td>
<td>29</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>1359</td>
<td>HPNIEEVAL*</td>
<td>24</td>
<td>1.1654</td>
<td>88</td>
<td>76</td>
<td>21</td>
<td>3</td>
<td>24</td>
<td>0</td>
</tr>
</tbody>
</table>

(Ibe et al., 1998; Giugliano et al., 2009; Park et al., 2012; Boucherma et al., 2013)

| 1370 | EGEIPFYGR | 9 | 1.0560 | 0 | 0 | 14 | 36 | 0 | 0 |    |    |    |    |    |    |
| 1374 | PFYGRAIPL | 7 | 0.9479 | 0 | 1 | 13 | 15 | 49 | 7 |    |    |    |    |    |    |
| 1642 | TKYIMACMS | 6 | 0.7008 | 0 | 70 | 0 | 70 | 93 | 0 |    |    |    |    |    |    |

**NS4B**

<p>| | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1712</td>
<td>SASLPYMDE</td>
<td>5</td>
<td>0.5488</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1819</td>
<td>QIAPPTAAT</td>
<td>5</td>
<td>0.5293</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1832</td>
<td>SGMAGAAVG</td>
<td>5</td>
<td>0.8750</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1848</td>
<td>LIDILAGYG</td>
<td>5</td>
<td>0.8431</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1854</td>
<td>GYGAGVAGA</td>
<td>9</td>
<td>0.5562</td>
<td>99</td>
<td>99</td>
<td>0</td>
<td>6</td>
<td>91</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NS5B**

<p>| | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2522</td>
<td>GYGAKEVRS</td>
<td>7</td>
<td>1.3540</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2557</td>
<td>TIMAKNEVF*</td>
<td>12</td>
<td>0.6303</td>
<td>99</td>
<td>80</td>
<td>83</td>
<td>100</td>
<td>84</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Doi et al, 2009)

<p>| 2568 | EPSKGKKKP | 6 | 1.4218 | 0 | 0 | 0 | 0 | 0 | 0 |    |    |    |    |    |    |
| 2720 | LASCRAAKL | 28 | 0.5264 | 0 | 0 | 0 | 0 | 0 | 0 |    |    |    |    |    |    |</p>
<table>
<thead>
<tr>
<th></th>
<th>LHGLSAFSL</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2886</td>
<td></td>
<td>11</td>
<td>0.6732</td>
<td>99</td>
<td>98</td>
<td>0</td>
<td>12</td>
<td>7</td>
</tr>
</tbody>
</table>

*-Indicates that the epitope has been experimentally proven to be a true positive

Bold- indicates percentage of epitope that is 100% conserved in more than 70% of the sequences analysed in each genotype
Table 7.2: HLA class II predicted epitopes of HCV genotype 5a and their Antigenicity prediction score, number of allele and conservation (in percentage) in different genotypes.

<table>
<thead>
<tr>
<th>Position</th>
<th>Epitope sequence</th>
<th>No of allele</th>
<th>Antigenecity score</th>
<th>Genotype 1a</th>
<th>Genotype 1b</th>
<th>Genotype 2</th>
<th>Genotype 3</th>
<th>Genotype 4</th>
<th>Genotype 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>320</td>
<td>WDMMMNWSP</td>
<td>10</td>
<td>1.3610</td>
<td>99</td>
<td>97</td>
<td>28</td>
<td>97</td>
<td>93</td>
<td>93</td>
</tr>
<tr>
<td>E2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>507</td>
<td>YCFTSPVVV</td>
<td>6</td>
<td>1.2112</td>
<td>93</td>
<td>98</td>
<td>88</td>
<td>100</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>509</td>
<td>FTPSPVVG</td>
<td>17</td>
<td>1.3553</td>
<td>94</td>
<td>98</td>
<td>88</td>
<td>100</td>
<td>89</td>
<td>73</td>
</tr>
<tr>
<td>665</td>
<td>LLHTTTQWA</td>
<td>5</td>
<td>0.6616</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>666</td>
<td>LHHTTQWAI</td>
<td>25</td>
<td>0.5863</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>692</td>
<td>LHQNIVDTQ</td>
<td>6</td>
<td>0.8564</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>739</td>
<td>LLVCQAEAA</td>
<td>22</td>
<td>0.5506</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>P7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>803</td>
<td>LPHRALALD</td>
<td>7</td>
<td>0.6966</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NS2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>860</td>
<td>WVPPLQVRG</td>
<td>7</td>
<td>1.1998</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>861</td>
<td>VPPLQVRGG</td>
<td>25</td>
<td>1.4312</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>864</td>
<td>LQVRGGRDA</td>
<td>20</td>
<td>0.6368</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>880</td>
<td>FHPALGFEI</td>
<td>19</td>
<td>1.4374</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>892</td>
<td>LLGILGPLY</td>
<td>14</td>
<td>0.5940</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Sequence</td>
<td>p-value</td>
<td>NAI</td>
<td>0%</td>
<td>1%</td>
<td>5%</td>
<td>9%</td>
<td>19%</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-----------</td>
<td>---------</td>
<td>-----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>895</td>
<td>ILGPLYLLQ</td>
<td>5</td>
<td>0.8264</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>938</td>
<td>LLHLGRLTG</td>
<td>25</td>
<td>0.5554</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>964</td>
<td>LRDLAVATE</td>
<td>9</td>
<td>1.0497</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>997</td>
<td>LAGLPVSAR</td>
<td>11</td>
<td>0.9120</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>1025</td>
<td>LLAPITAYA</td>
<td>22</td>
<td>0.5253</td>
<td>99</td>
<td>30</td>
<td>33</td>
<td>79</td>
<td>89</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td><strong>NS3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1047</td>
<td>LTGRDKNEA</td>
<td>22</td>
<td>1.5633</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1131</td>
<td>YLVTRHADV*</td>
<td>15</td>
<td>0.7569</td>
<td>99</td>
<td>99</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>1152</td>
<td>LLSPRPISY</td>
<td>5</td>
<td>2.2999</td>
<td>99</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1153</td>
<td>LSPRPISYL</td>
<td>5</td>
<td>1.3109</td>
<td>99</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1253</td>
<td>VLNPSVAAT</td>
<td>14</td>
<td>1.2489</td>
<td>100</td>
<td>100</td>
<td>97</td>
<td>100</td>
<td>96</td>
<td>89</td>
</tr>
<tr>
<td>1254</td>
<td>LNPSVAATL</td>
<td>6</td>
<td>0.9169</td>
<td>100</td>
<td>100</td>
<td>98</td>
<td>100</td>
<td>95</td>
<td>89</td>
</tr>
<tr>
<td>1258</td>
<td>VAATLGFGA</td>
<td>11</td>
<td>1.2617</td>
<td>90</td>
<td>85</td>
<td>98</td>
<td>9</td>
<td>67</td>
<td>29</td>
</tr>
<tr>
<td>1262</td>
<td>LGFGAYMSK*</td>
<td>14</td>
<td>0.8784</td>
<td>87</td>
<td>82</td>
<td>78</td>
<td>0</td>
<td>67</td>
<td>13</td>
</tr>
<tr>
<td>1264</td>
<td>FGAYMSKAY</td>
<td>14</td>
<td>0.6561</td>
<td>0</td>
<td>22</td>
<td>3</td>
<td>0</td>
<td>33</td>
<td>2</td>
</tr>
</tbody>
</table>

(Scognamiglio et al., 1999; Chang et al., 1999; Sreenarasimhaiah et al., 2003)
<p>| | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1327</td>
<td>LGIGTVLDQ</td>
<td>42</td>
<td>0.5607</td>
<td>97</td>
<td>88</td>
<td>98</td>
<td>100</td>
<td>93</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1373</td>
<td>IPFYGRAIP</td>
<td>16</td>
<td>0.5679</td>
<td>0</td>
<td>2</td>
<td>13</td>
<td>36</td>
<td>44</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1375</td>
<td>FYGRAIPLA</td>
<td>22</td>
<td>1.0280</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>26</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1392</td>
<td>IFCHSKKKC</td>
<td>22</td>
<td>1.5437</td>
<td>93</td>
<td>94</td>
<td>95</td>
<td>88</td>
<td>80</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1415</td>
<td>VAYYRGLDV</td>
<td>21</td>
<td>0.8250</td>
<td>98</td>
<td>96</td>
<td>92</td>
<td>79</td>
<td>91</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1418</td>
<td>YRQLDVAVI</td>
<td>51</td>
<td>1.0522</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1482</td>
<td>VSRSQRRGR</td>
<td>12</td>
<td>2.0649</td>
<td>0</td>
<td>93</td>
<td>96</td>
<td>97</td>
<td>98</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1561</td>
<td>VFTGLTNID</td>
<td>9</td>
<td>1.1689</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1583</td>
<td>FPYLVAYQA</td>
<td>24</td>
<td>0.6584</td>
<td>47</td>
<td>85</td>
<td>5</td>
<td>0</td>
<td>40</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1585</td>
<td>YLVAYQATV*</td>
<td>17</td>
<td>0.5020</td>
<td>99</td>
<td>85</td>
<td>29</td>
<td>24</td>
<td>87</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Zhang et al., 2002; Matsui et al., 2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1586</td>
<td>LVAYQATVC</td>
<td>28</td>
<td>0.6052</td>
<td>99</td>
<td>85</td>
<td>28</td>
<td>24</td>
<td>82</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1631</td>
<td>VQNEITLTH</td>
<td>35</td>
<td>1.1979</td>
<td>16</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1641</td>
<td>ITKYIMACM</td>
<td>8</td>
<td>0.6964</td>
<td>0</td>
<td>58</td>
<td>0</td>
<td>18</td>
<td>75</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1645</td>
<td>IMACMSADL</td>
<td>22</td>
<td>0.8406</td>
<td>1</td>
<td>82</td>
<td>0</td>
<td>70</td>
<td>93</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NS4B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1733</td>
<td>LGLIGTAGQ</td>
<td>43</td>
<td>1.2450</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1765</td>
<td>MWNFVSGIQ</td>
<td>18</td>
<td>1.2300</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>11</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1766</td>
<td>WNPVSGIQY</td>
<td>22</td>
<td>0.6246</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>11</td>
<td>91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1791</td>
<td>MSFTAAVTS</td>
<td>8</td>
<td>0.7414</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>93</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>1812</td>
<td>LGGWVASQI</td>
<td>16</td>
<td>0.608</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>96</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1849</td>
<td>IDILAGYGA</td>
<td>9</td>
<td>0.7525</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1863</td>
<td>LVAPKIMCG</td>
<td>47</td>
<td>1.0160</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1879</td>
<td>LVNLLPSIL</td>
<td>51</td>
<td>0.7592</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1880</td>
<td>VNLLPSILC</td>
<td>51</td>
<td>0.9874</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1883</td>
<td>LPSILCPGA</td>
<td>6</td>
<td>0.7104</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1886</td>
<td>ILCPGALVV</td>
<td>9</td>
<td>0.9365</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1892</td>
<td>LVVGVICAA</td>
<td>30</td>
<td>0.7097</td>
<td>0</td>
<td>0</td>
<td>97</td>
<td>94</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1893</td>
<td>VVVVICAAC</td>
<td>50</td>
<td>0.9530</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1896</td>
<td>VCIAVLRRR</td>
<td>35</td>
<td>0.9857</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1897</td>
<td>ICIAVLRRH</td>
<td>12</td>
<td>1.2316</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1916</td>
<td>MNRLIAFAS</td>
<td>48</td>
<td>0.5291</td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>100</td>
<td>93</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1919</td>
<td>LIAFASRGN</td>
<td>24</td>
<td>2.0235</td>
<td>100</td>
<td>99</td>
<td>98</td>
<td>100</td>
<td>93</td>
<td>96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1922</td>
<td>FASRGNSVH</td>
<td>18</td>
<td>1.2938</td>
<td>100</td>
<td>98</td>
<td>0</td>
<td>94</td>
<td>82</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**NS5A**

<p>| | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>WQQAKLLPQ</td>
<td>33</td>
<td>1.1603</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2101</td>
<td>YITGVTQDN</td>
<td>6</td>
<td>1.1358</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2105</td>
<td>VTQDNLKKCP</td>
<td>6</td>
<td>1.0178</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2237</td>
<td>MGGNITRVE</td>
<td>20</td>
<td>1.2656</td>
<td>92</td>
<td>96</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2285</td>
<td>LPVVARPGY</td>
<td>9</td>
<td>0.6143</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

186
**NS5B**

<table>
<thead>
<tr>
<th></th>
<th>Amino Acid</th>
<th>Length</th>
<th>Conserved</th>
<th>Total</th>
<th>85</th>
<th>91</th>
<th>97</th>
<th>85</th>
<th>91</th>
</tr>
</thead>
<tbody>
<tr>
<td>2422</td>
<td>MSYSWTGAL</td>
<td>9</td>
<td>0.5749</td>
<td>76</td>
<td>3</td>
<td>97</td>
<td>85</td>
<td>91</td>
<td>42</td>
</tr>
<tr>
<td>2451</td>
<td>LRHHNLVYS</td>
<td>40</td>
<td>1.2321</td>
<td>89</td>
<td>0.9</td>
<td>0</td>
<td>97</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>2523</td>
<td>YGAKEVRSL</td>
<td>24</td>
<td>1.1739</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>2559</td>
<td>MAKNEFAV</td>
<td>18</td>
<td>0.8466</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>2564</td>
<td>VFAVEPSKG</td>
<td>10</td>
<td>0.7925</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2637</td>
<td>FSYDTRCFD</td>
<td>11</td>
<td>1.5808</td>
<td>99</td>
<td>46</td>
<td>95</td>
<td>100</td>
<td>93</td>
<td>87</td>
</tr>
<tr>
<td>2660</td>
<td>YQSCDLQPE</td>
<td>11</td>
<td>1.3802</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2696</td>
<td>YRRCRASGV</td>
<td>41</td>
<td>1.2642</td>
<td>99</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>2778</td>
<td>YDLELVSC</td>
<td>6</td>
<td>1.0923</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2850</td>
<td>FSVLQSSEQ</td>
<td>21</td>
<td>0.8451</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2871</td>
<td>VYVSPTPLDL</td>
<td>29</td>
<td>2.0113</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2881</td>
<td>IQRLHGLSA</td>
<td>50</td>
<td>0.4322</td>
<td>98</td>
<td>72</td>
<td>0</td>
<td>0</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>2889</td>
<td>LSAAFLHHSY*</td>
<td>6</td>
<td>0.6007</td>
<td>99</td>
<td>97</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Mizukoshi et al., 2002*

*Indicates that the epitope has been experimentally proven to be a true positive

**Bold** indicates percentage of epitope that is 100% conserved in more than 70% of the sequences analysed in each genotype
Table 7.3: Distribution of genotype 5a HLA class I and II predicted epitopes in each of the HCV gene

<table>
<thead>
<tr>
<th>HLA class</th>
<th>No of predicted epitopes</th>
<th>Distribution of the predicted epitopes in each of the HCV gene *</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>I</td>
<td>24</td>
<td>0 (0)</td>
</tr>
<tr>
<td>II</td>
<td>77</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

* Number of epitopes and their percentage in brackets

b HCV gene with highest number of binding epitopes-bold
Table 7.4: Binding affinity scores of predicted epitopes and their variants to common HLA I alleles types prevalent in South Africa

<table>
<thead>
<tr>
<th>Gene</th>
<th>Epitope sequence</th>
<th>Gene type of epitope</th>
<th>HLA I allele types</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLA-A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A*01:01</td>
</tr>
<tr>
<td>E2</td>
<td>ALSTGLIHL</td>
<td>1a, 1b, 3, 4, 5a, 6</td>
<td>20245</td>
</tr>
<tr>
<td></td>
<td>ALSTGLIHL</td>
<td>2</td>
<td>18629</td>
</tr>
<tr>
<td>NS3</td>
<td>TILGIGTVL</td>
<td>1b, 2, 4, 5a</td>
<td>21706</td>
</tr>
<tr>
<td></td>
<td>SILGIGTVL</td>
<td>1a, 3</td>
<td>21632</td>
</tr>
<tr>
<td></td>
<td>TVLGIGTVL</td>
<td>6</td>
<td>21566</td>
</tr>
<tr>
<td></td>
<td>VLDQAETAG</td>
<td>1a, 1b, 2, 3, 4, 5a, 6</td>
<td>19514</td>
</tr>
<tr>
<td></td>
<td>HPNIEEVAL</td>
<td>1a, 1b, 5a</td>
<td>20801</td>
</tr>
<tr>
<td></td>
<td>HSNIEEVAL</td>
<td>2, 3, 4</td>
<td>17424</td>
</tr>
<tr>
<td></td>
<td>HPNITEFAL</td>
<td>6</td>
<td>20712</td>
</tr>
<tr>
<td></td>
<td>TKYIMACMS</td>
<td>1b, 3, 4, 5a</td>
<td>21634</td>
</tr>
<tr>
<td></td>
<td>TKYIMTCMS</td>
<td>1a, 6</td>
<td>21505</td>
</tr>
</tbody>
</table>

189
<table>
<thead>
<tr>
<th></th>
<th>GYGAGVAGA</th>
<th>1a, 1b, 4, 5a</th>
<th>22476</th>
<th>18978</th>
<th>8122</th>
<th>11107</th>
<th>21895</th>
<th>21157</th>
<th>37908</th>
<th>25486</th>
<th>40525</th>
<th>34892</th>
<th>27309</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GYGAGVSGA</td>
<td>3, 6</td>
<td>22088</td>
<td>20417</td>
<td>8866</td>
<td>12347</td>
<td>20861</td>
<td>20857</td>
<td>37791</td>
<td>27653</td>
<td>38649</td>
<td>31751</td>
<td>25129</td>
</tr>
<tr>
<td></td>
<td>GYGAGISGA</td>
<td>2</td>
<td>22118</td>
<td>19605</td>
<td>8977</td>
<td>11370</td>
<td>20252</td>
<td>20644</td>
<td>38005</td>
<td>26563</td>
<td>39368</td>
<td>30757</td>
<td>26422</td>
</tr>
<tr>
<td><strong>NS5B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TIMAKNEVF</td>
<td>1a, 1b, 2, 3,</td>
<td>20008</td>
<td>18433</td>
<td>18014</td>
<td>13671</td>
<td>17941</td>
<td>1603</td>
<td>1126</td>
<td>6660</td>
<td>3999</td>
<td>2297</td>
<td>2199</td>
</tr>
<tr>
<td></td>
<td>LHGLSAFSL</td>
<td>1a, 1b, 5a</td>
<td>21421</td>
<td>22806</td>
<td>18388</td>
<td>24928</td>
<td>15382</td>
<td>14861</td>
<td>28473</td>
<td>24356</td>
<td>4496</td>
<td>1842</td>
<td>7382</td>
</tr>
<tr>
<td></td>
<td>LHGLSAFL</td>
<td>3, 4</td>
<td>21330</td>
<td>22496</td>
<td>19372</td>
<td>26433</td>
<td>17894</td>
<td>17297</td>
<td>26323</td>
<td>26432</td>
<td>2706</td>
<td>1081</td>
<td>4801</td>
</tr>
<tr>
<td></td>
<td>LHGMAAFSL</td>
<td>6</td>
<td>20901</td>
<td>22156</td>
<td>16279</td>
<td>15813</td>
<td>13509</td>
<td>13310</td>
<td>23476</td>
<td>16026</td>
<td>3185</td>
<td>1253</td>
<td>11830</td>
</tr>
</tbody>
</table>

*a* Italic and bold- Indicates variation with one or two amino acids in comparison to the predicted epitope.

*b* IC$_{50}$ <50- <50 IC$_{50}$nm, bold, high binding affinity, >50 IC$_{50}$nm, <500 IC$_{50}$nm, italic, intermediate binding affinity, >500 IC$_{50}$nm, poor binding affinity, No value indicates program produced no binding score
Table 7.5: Predicted epitopes of HCV genotype 5a binding to common HLA-I prevalent in South African and their Antigenicity prediction score, HLA allele types and conservation (in percentage) in different genotypes.

<table>
<thead>
<tr>
<th>Position</th>
<th>Epitope sequence</th>
<th>Antigenicity score</th>
<th>HLA-I allele</th>
<th>Binding affinity score</th>
<th>Genotype 1a</th>
<th>Genotype 1b</th>
<th>Genotype 2</th>
<th>Genotype 3</th>
<th>Genotype 4</th>
<th>Genotype 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2</td>
<td>ALSTGLHL*</td>
<td>0.9005</td>
<td>HLA-A*02:01</td>
<td>28</td>
<td>70</td>
<td>99</td>
<td>0</td>
<td>97</td>
<td>84</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>(Himoudi et al., 2002; Engler et al., 2004; Ohno et al., 2006)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS2</td>
<td>LLAPITAYA</td>
<td>0.5253</td>
<td>HLA-A*02:01</td>
<td>13</td>
<td>99</td>
<td>30</td>
<td>33</td>
<td>79</td>
<td>89</td>
<td>69</td>
</tr>
<tr>
<td>NS3</td>
<td>YAQQTRGVL</td>
<td>0.5125</td>
<td>HLA-B*07:02</td>
<td>30</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HLA-A*30:01</td>
<td>19</td>
<td>98</td>
<td>97</td>
<td>9</td>
<td>0</td>
<td>90</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>AAYAAQGYK</td>
<td>0.6361</td>
<td>HLA-A*35:01</td>
<td>19</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>33</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>FGAYMSKAY</td>
<td>0.6561</td>
<td>HLA-B*35:01</td>
<td>19</td>
<td>9</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>33</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>HPNIEEVAL*</td>
<td>1.1654</td>
<td>HLA-B*35:01</td>
<td>12</td>
<td>88</td>
<td>77</td>
<td>21</td>
<td>3</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>(Ibe et al., 1998; Giugliano et al., 2008; Park et al., 2012; Boucherma et al., 2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LPSEGEIPF</td>
<td>0.6922</td>
<td>HLA-B*35:01</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1585</td>
<td>YLVAYQATV* (Zhang et al., 2002; Matsui et al., 2004)</td>
<td>0.5020</td>
<td>HLA-A*02:01</td>
<td>5</td>
<td>99</td>
<td>85</td>
<td>29</td>
<td>24</td>
<td>87</td>
<td>91</td>
</tr>
<tr>
<td>------</td>
<td>---------------------------------------------------</td>
<td>--------</td>
<td>-------------</td>
<td>---</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>1927</td>
<td>NHVSPTHYV</td>
<td>1.5172</td>
<td>HLA-C*07:01</td>
<td>50</td>
<td>99</td>
<td>98</td>
<td>0</td>
<td>94</td>
<td>82</td>
<td>96</td>
</tr>
<tr>
<td>2285</td>
<td>LPVWARPGY</td>
<td>0.6143</td>
<td>HLA-B*35:01</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2696</td>
<td>YRRCRASGV</td>
<td>1.2642</td>
<td>HLA-C*07:01</td>
<td>34</td>
<td>99</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>2763</td>
<td>MTRYSAPPG</td>
<td>1.0997</td>
<td>HLA-A*30:01</td>
<td>5</td>
<td>99</td>
<td>100</td>
<td>99</td>
<td>100</td>
<td>87</td>
<td>77</td>
</tr>
<tr>
<td>2889</td>
<td>LSAFSLHSY* (Mizukoshi et al., 2002)</td>
<td>0.7552</td>
<td>HLA-A<em>30:01, HLA-A</em>01:01</td>
<td>11, 38</td>
<td>99</td>
<td>98</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Indicates that the epitope has been experimentally proven to be a true positive

Bold- indicates percentage of epitope that is 100% conserved in more than 70% of the sequences analysed in each genotype
Table 7.6: Percentage binding affinity scores of HLA II epitopes to common alleles prevalent in South Africa.

<table>
<thead>
<tr>
<th>Position</th>
<th>Predicted Epitopes</th>
<th>Genotypes</th>
<th>HLA II- DRB1 allele types</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>DRB1*-0301</td>
</tr>
<tr>
<td><strong>E1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>320</td>
<td>WDMMMNNWSP</td>
<td>1a, 1b, 3, 4, 5a, 6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>WDMMMLNWSP</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td><strong>E2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>507</td>
<td>YCFTPSPVV</td>
<td>1a, 1b, 2, 3, 4, 5a</td>
<td>0</td>
</tr>
<tr>
<td>509</td>
<td>FTPSPVVVG</td>
<td>1a, 1b, 2, 3, 4, 5a</td>
<td>19</td>
</tr>
<tr>
<td><strong>NS2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1025</td>
<td>LLAPITAYA</td>
<td>1a, 3, 4, 5a</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>LLAPITAYS</td>
<td>1b</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>LLAPITAYT</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td><strong>NS3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1131</td>
<td>YLVTRHADV</td>
<td>1a, 1b, 5a</td>
<td>0</td>
</tr>
<tr>
<td>1152</td>
<td>LLSPRPISY</td>
<td>1a, 5a</td>
<td>15</td>
</tr>
<tr>
<td>1153</td>
<td>LSPRPISYL</td>
<td>1a, 5a</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>----------</td>
<td>------</td>
</tr>
<tr>
<td>1253</td>
<td>VLNPSVAAT</td>
<td>1a, 1b, 2, 3, 4, 5a, 6</td>
<td>0</td>
</tr>
<tr>
<td>1254</td>
<td>LNPSVAATL</td>
<td>1a, 1b, 2, 3, 4, 5a, 6</td>
<td></td>
</tr>
<tr>
<td>1258</td>
<td>VAATLGFGA</td>
<td>1a, 1b, 2, 5a</td>
<td></td>
</tr>
<tr>
<td>1262</td>
<td>LGFGAYMSK</td>
<td>1a, 1b, 2, 5a</td>
<td></td>
</tr>
<tr>
<td>1327</td>
<td>LGIGTVLDO</td>
<td>1a, 1b, 2, 3, 4, 5a, 6</td>
<td></td>
</tr>
<tr>
<td>1392</td>
<td>IFCHSKKKK</td>
<td>1a, 1b, 2, 3, 4, 5a, 6</td>
<td></td>
</tr>
<tr>
<td>1415</td>
<td>VAYYRGLDVE</td>
<td>1a, 1b, 2, 3, 4, 5a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>VAFFYG</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>1482</td>
<td>VSRSQRRGR</td>
<td>1b, 2, 3, 4, 5a, 6</td>
<td></td>
</tr>
<tr>
<td>1585</td>
<td>YLVAYQATV</td>
<td>1a, 1b, 4, 5a, 6</td>
<td></td>
</tr>
<tr>
<td>1586</td>
<td>LVAYQATV</td>
<td>1a, 1b, 4, 5a, 6</td>
<td></td>
</tr>
<tr>
<td>1645</td>
<td>IMACMSADL</td>
<td>1b, 3, 4, 5a</td>
<td></td>
</tr>
</tbody>
</table>

**NS4B**

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1765</td>
<td>MWNFVSGIQ</td>
<td>3, 5a</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>MWNFSGIQ</td>
<td>1a, 1b, 4</td>
<td></td>
<td>0</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>1766</td>
<td>WNFVSGIQY</td>
<td>3, 5a, 6</td>
<td></td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>WNFSGIQY</td>
<td>1a, 1b, 4</td>
<td></td>
<td>0</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>1892</td>
<td>LVVGVICAA</td>
<td>2, 3, 5a</td>
<td></td>
<td>40</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LVVGV</td>
<td>VCAA</td>
<td>1a, 1b, 4, 6</td>
<td>45</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>1916</td>
<td>MNRLIAFAS</td>
<td>1a, 1b, 2, 3, 4, 5a, 6</td>
<td>17</td>
<td>0</td>
<td>23</td>
<td>41</td>
</tr>
<tr>
<td>1919</td>
<td>LIAFASRGN</td>
<td>1a, 1b, 2, 3, 4, 5a, 6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NS5A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1922</td>
<td>FASRGNHVS</td>
<td>1a, 1b, 3, 4, 6</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>2237</td>
<td>MGGNITRVE</td>
<td>1a, 1b, 5a</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NS5B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2422</td>
<td>MSYSWTGAL</td>
<td>1a, 2, 3, 4, 5a</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>MSYWTGAL</td>
<td>1b</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2451</td>
<td>LRHHNLVYS</td>
<td>1a, 3, 5a</td>
<td>42</td>
<td>25</td>
<td>13</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>LRBBHNMVYA</td>
<td>1b, 4</td>
<td>18</td>
<td>17</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>LRHHNLVYS</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2637</td>
<td>FSYDTRCFD</td>
<td>1a, 2, 3, 5a</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2696</td>
<td>YRRCHASGV</td>
<td>1a, 1b, 2, 3, 5a</td>
<td>24</td>
<td>23</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>2883</td>
<td>IQRHLHGLSA</td>
<td>1a, 1b, 4, 5a</td>
<td>31</td>
<td>41</td>
<td>34</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>IQRHLHGMAA</td>
<td>6</td>
<td>24</td>
<td>12</td>
<td>31</td>
<td>18</td>
</tr>
<tr>
<td>2889</td>
<td>LSAFSLHSY</td>
<td>1a, 1b, 5a</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>LSAFTLHGY</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
</tbody>
</table>
Italic and bold- Indicates variation with amino acids in comparison to the predicted epitope.
Table 8.1: Predicted B cell epitopes of HCV genotype 5 and their antigenecity score, number of allele and conservancy in different genotypes.

<table>
<thead>
<tr>
<th>Position</th>
<th>Predicted Epitopes</th>
<th>Antigen score</th>
<th>Genotype 1</th>
<th>Genotype 2</th>
<th>Genotype 3</th>
<th>Genotype 4</th>
<th>Genotype 5</th>
<th>Genotype 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>222</td>
<td>HTPGCVPCVROGNVSS</td>
<td>1.2778</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>229</td>
<td>CVRQGNVSSCWQITP</td>
<td>1.1928</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>CWVQUITPTLSAPSGA</td>
<td>0.8932</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>286</td>
<td>LVAQMFTYRPRQHTTV</td>
<td>0.4089</td>
<td></td>
<td></td>
<td></td>
<td>LVAQMFTYRPRQHTV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>299</td>
<td>TTVQDCNCSIYSGHVT</td>
<td>0.4806</td>
<td></td>
<td></td>
<td></td>
<td>TTVQDCNCSIYSGHVT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>314</td>
<td>TGHRMAWDMMMNWSPT</td>
<td>0.8501</td>
<td>TGHMAWDMMMNWSPT</td>
<td>TGHMAWDMMMNWSPT</td>
<td>TGHMAWDMMMNWSPT</td>
<td>TGHMAWDMMMNWSPT</td>
<td>TGHMAWDMMMNWSPT</td>
<td>TGHMAWDMMMNWSPT</td>
</tr>
<tr>
<td>322</td>
<td>MMMNWSPTTALLMAQL</td>
<td>0.9287</td>
<td>MMMNWSPTTALLMAQL</td>
<td>MMMNWSPTTALLMAQL</td>
<td>MMMNWSPTTALLMAQL</td>
<td>MMMNWSPTTALLMAQL</td>
<td>MMMNWSPTTALLMAQL</td>
<td>MMMNWSPTTALLMAQL</td>
</tr>
<tr>
<td>344</td>
<td>VVDIIAGAHWGVLFAA</td>
<td>0.5973</td>
<td></td>
<td></td>
<td></td>
<td>VVDIIAGAHWGVLFAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>360</td>
<td>AYYASAANWAKVAL</td>
<td>0.7181</td>
<td></td>
<td></td>
<td></td>
<td>AYYASAANWAKVAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>386</td>
<td>QTTGGSAARNVYG VTS</td>
<td>0.7586</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>411</td>
<td>LQLINTNGSWHINRTA</td>
<td>0.5242</td>
<td>LQLINTNGSWHINRTA</td>
<td>LQLINTNGSWHINRTA</td>
<td>LQLINTNGSWHINRTA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>421</td>
<td>HINRTALNCNDSDLHTG</td>
<td>0.7383</td>
<td>HINRTALNCNDSDLHTG</td>
<td>HINRTALNCNDSDLHTG</td>
<td>HINRTALNCNDSDLHTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>469</td>
<td>WGPISYASVSGPSDDK</td>
<td>0.550</td>
<td>WGPISYASVSGPSDDK</td>
<td>WGPISYASVSGPSDDK</td>
<td>WGPISYASVSGPSDDK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>475</td>
<td>VSGPSDDKPYCWHYPP</td>
<td>0.9420</td>
<td>VSGPSDDKPYCWHYPP</td>
<td>VSGPSDDKPYCWHYPP</td>
<td>VSGPSDDKPYCWHYPP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>486</td>
<td>CWHYPPRCGVPAPAQE</td>
<td>0.6602</td>
<td>CWHYPPRCGVPAPAQE</td>
<td>CWHYPPRCGVPAPAQE</td>
<td>CWHYPPRCGVPAPAQE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>504</td>
<td>GPVYCFTPSPVVGGTT</td>
<td>1.1613</td>
<td>GPVYCFTPSPVVGGTT</td>
<td>GPVYCFTPSPVVGGTT</td>
<td>GPVYCFTPSPVVGGTT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>526</td>
<td>TYNWGSNETDIFLLNN</td>
<td>0.4859</td>
<td>TYNWGSNETDIFLLNN</td>
<td>TYNWGSNETDIFLLNN</td>
<td>TYNWGSNETDIFLLNN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>566</td>
<td>APPCNLGPTGNNSLKC</td>
<td>1.4119</td>
<td>APPCNLGPTGNNSLKC</td>
<td>APPCNLGPTGNNSLKC</td>
<td>APPCNLGPTGNNSLKC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>629</td>
<td>VRMYIGGLEHRLDAC</td>
<td>1.1496</td>
<td>VRMYIGGLEHRLDAC</td>
<td>VRMYIGGLEHRLDAC</td>
<td>VRMYIGGLEHRLDAC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>638</td>
<td>HRLDAACNWTRGE</td>
<td>0.8961</td>
<td>HRLDAACNWTRGE</td>
<td>HRLDAACNWTRGE</td>
<td>HRLDAACNWTRGE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RCD</td>
<td>RGERCD</td>
<td>GERCD</td>
<td>ERCD</td>
<td>ERCD</td>
<td>RCD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>--------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>649</td>
<td>GERCDLEDRDAELSP</td>
<td>1.7546</td>
<td>GERCDLEDRDAELSP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>655</td>
<td>EDRDRAELSPLLHTTT</td>
<td>0.8442</td>
<td>EDRDRAELSPLLHTTT</td>
<td></td>
<td>EDRDRAELSPLLHTTT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>668</td>
<td>TTTQWAILPCSFTPTP</td>
<td>0.7815</td>
<td></td>
<td></td>
<td>TTTQWAILPCSFTPTP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>675</td>
<td>LPCSFTPTPALSTGLI</td>
<td>0.5340</td>
<td>LPCSFTPTPALSTGLI</td>
<td>LPCSFTPTPALSTGLI</td>
<td>LPCSFTPTPALSTGLI</td>
<td>LPCSFTPTPALSTGLI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>685</td>
<td>LSTGLIHLHQIVDTQ</td>
<td>0.6639</td>
<td>LSTGLIHLHQIVDTQ</td>
<td>LSTGLIHLHQIVDTQ</td>
<td>LSTGLIHLHQIVDTQ</td>
<td>LSTGLIHLHQIVDTQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>709</td>
<td>LVLWAIKWEYIVLAF</td>
<td>1.3221</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bold indicates 100 percent conserved as compared to the predicted epitopes.
11.2 Publications