

https://dspace.org/

In vitro and in vivo diversity of HIV-1 subtype C envelope proteins and correlation with changes in biological properties of viral isolates.

Coetzer, Maria Elizabeth

2006-10-31T12:37:19Z
http://hdl.handle.net/10539/1540

Downloaded from a DSpace repository. https://dspace.org/

In vitro and *in vivo* diversity of HIV-1 subtype C envelope proteins and correlation with changes in biological properties of viral isolates.

Maria Elizabeth (Mia) Coetzer

A thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, for the degree of Doctor of Philosophy in Medicine.

2005

ABSTRACT

HIV-1 gains entry into host cells by binding to CD4 and a coreceptor, predominantly CCR5 or CXCR4. Viruses that use CCR5 are termed R5, those able to use CXCR4 are termed X4 while viruses able to use both coreceptors are referred to as R5X4. Accelerated CD4 decline and disease progression within an infected HIV-1 subtype B infected individual is often associated with the emergence of viruses able to use CXCR4. However, CXCR4 coreceptor usage appears to occur less frequently among HIV-1 subtype C viruses, the most predominant strain circulating globally, including South Africa. The aim of this study was to investigate the genetic determinants of CXCR4 usage in HIV-1 subtype C isolates.

The V3 region of the envelope glycoprotein is the major determinant of coreceptor usage. In Chapter 2, 32 subtype C isolates with known phenotypes (16 R5, 8 R5X4 and 8 X4 isolates) were assessed using a subtype C specific V3-heteroduplex tracking assay. Results indicated that there were sufficient genetic differences to discriminate between R5 viruses and those able to use CXCR4 (both R5X4 and X4). In general, R5 isolates had a mobility ratio >0.9 whereas CXCR4-using isolates were usually <0.9. Sequence analysis of the V3 region showed that CXCR4-using viruses were often associated with an increased positive amino acid charge, insertions and loss of a glycosylation site, similar to HIV-1 subtype B. In contrast, where subtype B consensus V3 has a GPGR crown motif irrespective of coreceptor usage, all 16 subtype C R5 viruses had a conserved GPGQ sequence at the tip of the loop, while 12 of the 16 (75%) CXCR4-using viruses had substitutions in this motif, most commonly arginine (R). Thus, the rare occurrence of CXCR4-using viruses in subtype C may be due to the highly conserved

nature of the GPGQ crown that may limit the potential for the development of X4 viruses.

The usefulness of available genotype-based methods for predicting viral phenotypes in subtype C was explored in Chapter 3. Results indicated that commonly used prediction methods could detect R5 viruses, but were not very sensitive at identifying X4 viruses. We therefore developed a subtype C specific predictor based on position specific scoring matrices (PSSM). Similar methodology, as used in developing the subtype B PSSM, was applied on a training set of 280 subtype C sequences of known phenotype (229 NSI/CCR5 and 51 SI/CXCR4). The C-PSSM had a specificity of 94% (C.I. [92%-96%]) and sensitivity of 75% (C.I. [68%-82%]), indicating that the C-PSSM had improved sensitivity in predicting CXCR4 usage. This method also highlighted amino acid positions within V3 that could contribute differentially to phenotype prediction in subtypes B and C. A reliable phenotype prediction method, such as the C-PSSM, could provide a rapid and less expensive approach to identifying CXCR4 variants, and thus increase our knowledge of subtype C coreceptor usage.

In Chapter 4 we examined the genetic changes in full-length gp160 envelope genes of 23 sequential isolates from 5 patients followed for two to three years. Three of the patients' isolates used CCR5 at all time points while 2 patients underwent a coreceptor switch with disease progression. The genetic changes observed over time indicated changes in length of variable loops particularly the V1, V4 and V5 and shifting N-glycosylation sites, particularly in the 2 patients that used CXCR4. Changes in the V3 were only noted in the 2 patients' that used CXCR4 which included substitutions of specific amino acids including those in the crown and increased amino acid charge in the V3 region. Both of

these patients were dually infected suggesting that recombination may contribute to the rapid emergence of X4 viruses.

The *in vitro* and *in vivo* development of CXCR4 usage was analysed in a pediatric patient that experienced a coreceptor switch during disease progression (Chapter 5). Biological and molecular clones were generated and the V1-V5 regions sequenced. Analyses of the V3 region indicated that the evolution to CXCR4 usage happens in a step-wise manner that included increased charge and changes in the crown motif. The intermediate variants with predicted dualtropism were also associated with increased V1-V2 lengths, suggesting that other regions may contribute to coreceptor switching. Furthermore, the development of CXCR4 usage within this patient was due to two mutational pathways, in which one resulted in R5X4 viruses and the other X4 variants.

In Chapter 6, the impact and treatment of acute TB on HIV-1 diversity in co-infected patients was investigated, specifically to determine the genetic characteristics of the viral populations present before, during and after TB treatment. Plasma samples from 18 HIV-1 infected patients were analysed using the C2V3 region, six of whom showed a high degree of variation using a V3-HTA and were selected for further analyses. All patients were predicted as R5 with no evidence of coreceptor switching over time. There was no correlation between the degree of genetic diversity and viral load, although both showed fluctuations over time. Phylogenetic and pairwise genetic distance analysis indicated that there was amplification of existing variants in 3 patients while in the other 3 patients there were dramatic shifts in viral populations suggesting selection of viral sub-populations over time. Thus in some co-infected patients, TB can affect HIV-1 genetic

heterogeneity although there was no evidence of a shift towards CXCR4 usage despite the presence of an AIDS defining illness.

Observations in this study have shown that the V3 region is the major determinant of coreceptor usage within HIV-1 subtype C, similar to HIV-1 subtype B. Characteristics such as increased charge length variability of the V3 region and loss of the glycosylation site within this region are associated with CXCR4 usage. The limited number of X4 viruses in subtype C does suggest some restricting mechanisms for CXCR4 usage. In this study we looked at genetic determinants and found that the rare occurrence of CXCR4-using viruses in subtype C, may be due to the highly conserved nature of the GPGQ crown that may limit the potential for the development of subtype C X4 viruses. Furthermore, the development of CXCR4 usage happened in a step-wise manner, with R5X4 viruses intermediates, in which an increased V1-V2 was observed suggesting that other regions within the envelope protein do contribute to coreceptor usage. Thus, regions such as V1-V2 and V4-V5 did contribute to coreceptor usage, but the V3 region remained the most important determinant of coreceptor usage in HIV-1 subtype C isolates. Collectively these findings have provided important data on the genetic determinants of CXCR4 usage in HIV-1 subtype C and an understanding of how they might evolve within a patient.

DECLARATION

I declare that this thesis is my own work unless specified and submitted for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted for any other degree or examination at any other university.

____ day of _____, _____

_____

Maria Elizabeth (Mia) Coetzer

I would like to dedicate this thesis to my parents, Riaan and Alet, as well as my sister Elmien and brother Riaan for your faithful support and boundless love.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DECLARATION	vi
DEDICATION	vii
ACKNOWLEDGMENTS	xv
PUBLICATIONS AND MANUSCRIPTS FROM THIS THESIS	xvi
OTHER PUBLICATIONS	xvii
PRESENTATIONS AT MEETINGS	xviii
LIST OF FIGURES	xix
LIST OF TABLES	xxii
LIST OF ABBREVIATIONS	xxiii
CHAPTER 1 - INTRODUCTION	**1**
1.1 BACKGROUND	2
1.2 ENTRY OF HIV-1 INTO THE HOST CELL	5
1.2.1 Viral genome and structure	5
1.2.2 Viral entry	6
1.2.3 Coreceptors	8
1.2.4 Minor receptors	10
1.2.5 Phenotype and tropism	10
1.2.6 Importance of coreceptor determination	11

1.2.7 Coreceptors and entry inhibitors 12

1.2.8 Coreceptor usage of South African subtype C isolates 13

1.3 HIV-1 ENVELOPE GLYCOPROTEIN 14

1.3.1 Structure of the envelope protein 14

1.3.2 Glycosylation of the envelope protein 15

1.3.3 Genetic diversity of the envelope glycoprotein 16

1.3.4 Envelope protein diversity and disease progression 17

1.3.5 Conformational changes during entry 18

1.3.6 The V3 region as a major determinant of coreceptor usage 19

1.3.7 Other regions involved in coreceptor usage 22

1.4 PHENOTYPE PREDICTION METHODS BASED ON THE V3 REGION 23

1.4.1 The 11/25 rule 23

1.4.2 Briggs method 24

1.4.3 Pillai method 25

1.4.4 Neural networks 26

1.4.5 Geno2Pheno method 26

1.4.6 Position specific scoring matrix method (PSSM) 27

1.4.7 Prediction of subtype C data sets 27

1.5 NATURAL HISTORY OF HIV-1 INFECTION 28

1.5.1 Immune activation 29

1.5.2 Tuberculosis 31

1.5.2.1 Interaction between TB and HIV-1 31

1.6 OBJECTIVES OF THIS STUDY 32

CHAPTER 2 - WHAT GENETIC CHANGES IN V3 ARE ASSOCIATED WITH CXCR4 USAGE IN HIV-1 SUBTYPE C ISOLATES? 33

2.1 INTRODUCTION 34

2.2 MATERIALS AND METHODS 36

2.2.1 Isolation and coreceptor usage of HIV-1 subtype C viruses 36

2.2.2. Viral RNA isolation and RT-PCR 36

2.2.3 V3-HTA 37

2.2.4 Cloning of single populations 38

2.2.5 Sequence analysis 38

2.3 RESULTS 39

2.3.1 Investigation of sample complexity using a V3-HTA 39

2.3.2 Analysis of isolates with multiple variants 43

2.3.3 V3 sequence variability of CCR5- and CXCR4-using variants 45

2.3.4 Determining subtype C V3 characteristics associated with CXCR4 usage 49

2.4 DISCUSSION 52

CHAPTER 3 - A RELIABLE PHENOTYPE PREDICTOR FOR HIV-1 SUBTYPE C BASED ENVELOPE V3 SEQUENCES 58

3.1 INTRODUCTION 59

3.2 MATERIALS AND METHODS 61

3.2.1 Data set compilation 61

3.2.2 Performance of four genotypic algorithms on subtype C sequences 61

3.2.3 How is a PSSM matrix constructed? 62

3.2.4 Determining the PSSM score for a target sequence 63

3.2.5 Development and validation of a C-PSSM 64

3.2.6 Overlap coefficient analysis determining differences between subtypes B and C 66

3.2.7 Comparison between predictions from V3-HTA and C-PSSM 68

3.2.8 C-PSSM for public use 68

3.3 RESULTS 69

3.3.1 HIV-1 subtype C sequences 69

3.3.2 Predicting phenotypes from genetic sequence data 70

3.3.3 Performance of C-PSSM prediction 71

3.3.4 Site-wise differences in phenotype between subtypes B and C 74

3.3.5 Comparison of C-PSSM scores with a genotype-based assay 76

3.4 DISCUSSION 77

CHAPTER 4 - CHARACTERISATION OF THE gp160 REGION FROM HIV-1 SUBTYPE C ISOLATES AND CORRELATION TO CORECEPTOR SWITCHING 81

4.1 INTRODUCTION 82

4.2 MATERIALS AND METHODS 83

4.2.1 Isolation and coreceptor usage of HIV-1 subtype C viruses 83

4.2.2 Viral RNA isolation and gp160 sequencing 83

4.2.3 Sequence analysis 84

4.3 RESULTS 85

4.3.1 Clinical information of patients 85

4.3.2 Phylogenetic analysis 87

4.3.3 Genetic variation within a patient during disease progression 89

4.3.4 Variable loops V1-V5 96

4.3.5 Predicted N-glycosylation sites 97

4.3.6 V3 region and coreceptor usage 98

4.3.7 Monoclonal antibody recognition sites 99

4.4 DISCUSSION 100

CHAPTER 5 - MOLECULAR AND BIOLOGICAL HETEROGENEITY IN SEQUENTIAL HIV-1 ISOLATES FROM A PATIENT THAT ACQUIRED THE ABILITY TO USE CXCR4 104

5.1 INTRODUCTION 105

5.2 MATERIALS AND METHODS 107

5.2.1 Viral isolation and coreceptor usage 107

5.2.2 Generating biological clones 107

5.2.3 Amplification of the V1-V5 region 107

5.2.4 Generating molecular clones 108

5.2.5 Sequencing and analysis 109

5.2.6 Phenotype prediction 109

5.3 RESULTS 110

5.3.1 Coreceptor determination of biological clones 110

5.3.2 True coreceptor usage versus predicted phenotype 112

5.3.3 Phenotype prediction of molecular clones 113

5.3.4 Distribution of phenotypes at different time points 114

5.3.5 Sequence analysis of the V3 region 115

5.3.6 Evolution of CXCR4 usage within the V3 loop 117

5.3.7 Other regions contributing to coreceptor usage 119

5.4 DISCUSSION 120

CHAPTER 6 - THE IMPACT OF ACTIVE TUBERCULOSIS ON HIV-1 SUBTYPE C GENETIC DIVERSITY 124

6.1 INTRODUCTION 125

6.2 MATERIALS AND METHODS 127

6.2.1 Patient information 127

6.2.2 Viral RNA isolation and RT-PCR 127

6.2.3 C2V3-HTA 128

6.2.4 Cloning 129

6.2.5 Subtyping 129

6.2.6 Phenotype prediction 130

6.3 RESULTS 131

6.3.1 Characterization of HIV-1 populations in HIV/TB patients 131

6.3.2 Screening of HIV/TB patients for genetic diversity 134

6.3.3 Molecular clone analysis of viral populations 134

6.3.3.1 Genetic diversity within patient samples 135

6.3.3.2 Phylogenetic analysis within a patient 137

6.3.3.3 Sequence analysis of patients with sub-populations 145

6.4 DISCUSSION 150

CHAPTER 7 - CONCLUSION **154**

APPENDICES 161

Appendix A: Ethical clearance 162

Appendix B: Additional ethical clearance 163

Appendix C: Amino acid abbreviations 164

Appendix D: Reagents and recipes 165

Appendix E: HTA probe sequences 166

Appendix F: Vector map 167

Appendix G: Determining coreceptor usage in transfected cell lines 168

CHAPTER 8 - REFERENCES **169**

AKNOWLEDGEMENTS

I would like to thank Prof Lynn Morris for giving me this incredible opportunity, and allowing me the freedom to discover. Thank you for your support and guidance!

I would also like to thank the following people who contributed to this thesis:

Viral isolation, coreceptor determination and generating of biological clones - Mary Phoswa, Dr Tonie Cilliers and Melene Smith.

Development and script writing of C-PSSM - Mark Jensen.

Prof Salim Abdool Karim and Dr Quarraisha Abdool Karim for providing sponsorship though the Fogarty Training Program to visit Dr Ronald Swanstrom's laboratory at University of North Carolina, Chaplin Hill, USA and Dr Jim Mullins laboratory at University of Washington, Seattle, USA.

I would like to thank, Drs Ronald Swanstrom, Angelique van't Wout, Mark Jensen and Jim Mullins for allowing me to visit their laboratories and sharing their knowledge, they were instrumental in this study. I would also like to thank Sue Herrmann and Sheila Doig for their support and help through these years, as well as Sarah Cohen, and my fellow colleagues at the AIDS Unit. Hazel Saevitzon and Sandy Hutchinson in the library, and Prof Schoub for allowing me to study at the National Institute for Communicable Diseases.

Thank you to Elsie for your support. All glory to Jesus Christ.

This work was funded by: South African AIDS Vaccine Initiative, Poliomyelitis Research Foundation, Fogarty International Center (TWO-0231) and Wellcome Trust.

PUBLICATIONS AND MANUSCRIPTS FROM THIS THESIS

1. **COETZER, M.**, Cilliers, T., Swanstrom, R. and Morris, L. What genetic changes in V3 are associated with CXCR4 usage in HIV-1 subtype C isolates? SUBMITTED.

2. Jensen*, M.A., **COETZER*, M.**, Van' t Wout, A.B., Morris, L. and Mullins, J.I. A reliable phenotype predictor for HIV-1 subtype C based on envelope V3 sequences. Journal of Virology Volume 80, Number 10, pp 4698-4704. *Joint first authors.

3. **COETZER, M.**, Cilliers, T., Papathanasopoulos, M., Ramjee, G., Abdool Karim, S. and Morris, L. Characterisation of the gp160 region from HIV-1 subtype C isolates and correlation to coreceptor switching. (In preparation).

4. **COETZER, M.**, Smith, M., Cilliers, T., Myers, T. and Morris, L. Molecular and biological heterogeneity in sequential HIV-1 isolates from a patient that acquired the ability to use CXCR4. (In preparation).

5. **COETZER, M.**, Grant, A., Day, J., Charalambous, S., Puren, A., Fielding, K., Churchyard, G. and Morris, L. The impact of active tuberculosis on HIV-1 subtype C V3 genetic diversity. (In preparation).

OTHER PUBLICATIONS

Bredell, H., Hunt, G., Casteling, A., Cilliers, T., Raudemeyer, C., **COETZER, M.**, Miller, S., Johnson, D., Tiemessen, C., Martin, D., Williamson, C. and Morris, L. HIV-1 subtype A, D and G sequences identified in South Africa. 2002. AIDS Research and Human Retroviruses Volume 18, Number 9, pp 681-683.

Cilliers, T., Nhlapo, J., **COETZER, M.**, Orlovic, D., Ketas, T., Olson, W.C., Moore, J.P., Trkola, A. and Morris, L. The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C. 2003. Journal of Virology Volume 77, Number 7, pp 4449-4456.

Cilliers, T., Willey, S., Sullivan, M., Patience, T., Pugach, P., **COETZER, M.**, Papathanasopoulos, M.A., Moore, J.P., Olson, W.C., Trkola, A., Clapham, P. and Morris, L. Use of alternate coreceptors on primary cells by two HIV-1 isolates. 2005. Virology 339, pp136-144.

Cilliers, T., Moore, P.L., **COETZER, M.**, and Morris, L. *In vitro* resistance of HIV-1 isolates to Enfuvirtide. 2005. AIDS Research and Human Retroviruses Volume 21, Number 9, pp 776-783.

PRESENTATIONS AT MEETINGS

COETZER, M., Cilliers, T., Papathanasopoulos, M.A., Ramjee, G., Abdool-Karim, S., Williamson, C. and Morris, L. Analysis of co-receptor usage among sequential HIV-1 subtype C isolates from acutely infected sex workers in South Africa. 2001. AIDS VACCINE Conference, Philadelphia: (Poster).

COETZER, M., Cilliers, T., Papathanasopoulos, M.A., Ramjee, G., Abdool-Karim, S., Williamson, C. and Morris, L. Analysis of co-receptor usage among sequential HIV-1 subtype C isolates from acutely infected sex workers in South Africa. 2001. IUBMB / SASBMB Special meeting on the Biochemical & Molecular basis of disease, Cape Town. (Poster).

COETZER, M., Jensen, M.A., Van‘ t Wout, A.B., Mullins, J.I. and Morris, L. Genetic determinants of co-receptor usage of HIV-1 subtype C isolates from South Africa. 2004. HIV Pathogenesis Program, Biennial Conference, Durban, South Africa. (Oral presentation).

COETZER, M., Grant, A., Day, J., Charalambous, S., Puren, A., Fielding, K., Churchyard, G. and Morris, L. Impact of TB on HIV genetic diversity in co-infected patients. 2004. Aurum Scientific Review Meeting, Johannesburg, South Africa. (Oral presentation).

LIST OF FIGURES

Page

1.1: Global distribution of adults and children living with HIV-1. 2
1.2: The evolutionary history of the primate lentiviruses. 3
1.3: Global distribution of the predominant HIV-1 subtypes and recombinants. 4
1.4: HIV-1 viral genome and structure. 5
1.5: A model for HIV-1 entry into the target host cell. 6
1.6: Key aspects of HIV-1 viral life cycle. 7
1.7: Two-dimensional structure of CCR5 and CXCR4. 9
1.8: HIV-1 phenotype determined by coreceptor usage. 11
1.9: Three-dimensional structure of the gp120 glycoprotein. 14
1.10: A schematic diagram of the gp120 envelope glycoprotein and potential glycosylation sites. 15
1.11: A schematic representation of the V3 loop within gp120 envelope protein. 20
1.12: Modeled trimer and coreceptor schematic. 21
1.13: Decision tree constructed of amino acid position 12. 25
1.14: The natural history of HIV-1 infection. 28
1.15: Consequences of immune activation on the biology of HIV-1 infection. 30

2.1: V3-HTA of HIV-1 subtype C isolates with different biological phenotypes. 41
2.2: Comparison of mobility ratio and coreceptor usage of subtype C isolates. 42
2.3: Analysis of molecular clones from HIV-1 subtype C isolates with multiple variants. 44
2.4: Comparison of V3 length and charge of HIV-1 subtype C isolates. 46
2.5: Variation in the V3 region of CCR5- and CXCR4- using isolates. 47
2.6: Distribution of amino acids in the V3 region. 48
2.7: Consensus V3 sequences from CCR5- and CXCR4-using isolates. 49
2.8: Comparison of amino acid variation in the V3 crown of subtype C isolates. 51

3.1: Constructing a PSSM matrix. 62
3.2: Calculating the PSSM score. 63
3.3: Determining sensitivity and specificity of the PSSM. 65
3.4: Determining the overlap coefficients within V3. 67
3.5: Sequence logos of HIV-1 subtype C V3 sequences used in this study. 69
3.6: Comparison of C-PSSM score distribution of 229 NSI and 51 SI subtype C sequences. 72
3.7: Comparison of C-PSSM on a subtype C data set with B-PSSM on a subtype B data set. 73
3.8: Effect of sample size on sensitivity and specificity of the C-PSSM. 74
3.9: V3 overlap coefficient *p*-values for subtypes C and B. 75
3.10: The HTA mobility ratio compared to the C-PSSM score. 76

4.1: Neighbor joining tree of HIV-1 gp120 and gp41. 88
4.2: Recombination analysis using of Du151(JUN99) using SimPlot. 89
4.3 A: Du123 amino acid alignment of the gp160 region. 90
4.3 B: Du151 amino acid alignment of the gp160 region. 91
4.3 C: Du179 amino acid alignment of the gp160 region. 92
4.3 D: Du368 amino acid alignment of the gp160 region. 93
4.3 E: Du422 amino acid alignment of the gp160 region. 94
4.4: Percentage genetic variability of the envelope region within a patient. 95
4.5: Frequency and position of N-glycosylation sites over time. 97
4.6: Changes in the V3 amino acid sequence alignment correlated to isolate biotype and amino acid charge. 98

5.1: Comparison between experimentally determined and predicted phenotypes of the biological clones. 112
5.2: Selection of molecular clones within TM18 A-C. 113
5.3: Distribution of phenotypes within the molecular and biological clones at various time points. 114
5.4: Schematic phylogenetic tree showing the evolution of CXCR4-using viruses. 118

6.1: Phylogenetic analysis of 18 HIV/TB and 5 HIV infected patients. 132
6.2: C2V3-HTA profiles of 6 HIV/TB patients over time. 135
6.3: CD4 count and viral load compared to genetic diversity over time. 136
6.4 A: Phylogenetic tree of variants from PC0122 in the C2V3 region. 139
6.4 B: Phylogenetic tree of variants from PC0261 in the C2V3 region. 140
6.4 C: Phylogenetic tree of variants from PC0703 in the C2V3 region. 141
6.4 D: Phylogenetic tree of variants from PC0137 in the C2V3 region. 142
6.4 E: Phylogenetic tree of variants from PC0213 in the C2V3 region. 143
6.4 F: Phylogenetic tree of variants from PC0614 in the C2V3 region. 144
6.5 A: Consensus sequences and distribution of viral variants within PC0137. 146
6.5 B: Consensus sequences and distribution of viral variants within PC0213. 148
6.5 C: Consensus sequences and distribution of viral variants within PC0614. 148
6.6 Summary of results from the 6 HIV/TB patients. 151

LIST OF TABLES

Page

2.1: Clinical information of 16 R5, 8 R5X4 and 8 X4 HIV-1 subtype C isolates. 40

3.1: Comparison of the performance of available prediction methods. 71

4.1: Clinical information of the five patients followed longitudinally. 86

4.2: The length of the variable loops and number of predicted N-glycosylation sites within the gp160 region of each patient. 96

5.1: Clinical information of patient TM18 followed for 2 years. 110

5.2: Biological clones of TM18B and TM18C and the experimentally determined coreceptor usage. 111

5.3: Predicted phenotype of 18 V3 sequences representing identical molecular and biological clones. 116

5.4: Variations within V1-V2 and V4-V5 that may contribute to phenotype switching. 119

6.1: Phenotype prediction based on the V3 region of 18 HIV/TB patients. 133

6.2: The pairwise genetic distance for 6 HIV/TB patients. 138

LIST OF ABBREVIATIONS

ag	Antigen
AIDS	Acquired immunodeficiency syndrome
ARV	Antiretroviral therapy
bp	Base pair
B-PSSM	Subtype B specific-PSSM
C	Constant region
CRF	Circulating recombinant form
C2	Second constant region
C2V3-HTA	Second constant region-third variable region specific-heterodupplex tracking assay
CD4	Cluster differentiation four
cDNA	Complementary deoxyribonucleic acid
C-PSSM	Subtype C specific-PSSM
CTL	Cytotoxic T-lymphocytes
ECL	extracellular loop
ENF	Enfuvirtide
Env	Envelope protein
env	Envelope gene
gag	Group-specific antigen gene
gp	Glycoprotein
GPR1	G-protein receptor 1
HAART	Highly active antiretroviral therapy
HIV-1	Human immunodeficiency virus type 1

HIV-2	Human immunodeficiency virus type 2
HR-1	First heptad repeat region
HR-2	Second heptad repeat region
HTA	Heteroduplex tracking assay
IL-2	Interleukin 2
k	Mobility ratio from the V3-HTA
kb	Kilo base
Mab	Monoclonal antibody
MEGA	Molecular Evolutionary Genetic Analysis
mg	Milligram
min	minute
MIP-1α	Macrophage inflammatory protein type 1 alpha
MIP-1β	Macrophage inflammatory protein type 1 beta
ml	Millilitre
ng	Nanogram
NH	Amino terminus
NSI	Non-syncytium inducing
nt	Nucleotide
OC	Overlap coefficient
ºC	Degrees Celsius
PBMC	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PHA	Phytohemagglutinin
pol	Polymerase protein

r5	Predicted CCR5 usage
R5	CCR5-using virus
R5-like	Suggestive of R5 virus
R5X4	CCR5- and CXCR4-using virus
r5x4	Predicted R5X4 virus
RANTES	Regulated upon activation, normal T-cell expressed and secreted
RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase polymerase chain reaction
SI	Syncytium inducing
SIV	Simian immunodeficiency virus
SVM	Support vector machines
TB	Tuberculosis
µl	Microlitre
V	Variable region
V3-HTA	Third variable region specific-heteroduplex tracking assay
x4	predicted CXCR4 usage
X4	CXCR4-using virus
X4-like	Suggestive of CXCR4 usage

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

The Acquired Immune Deficiency Syndrome (AIDS), was first described in 1981 among young homosexual men who presented with rare opportunistic infections, such as pneumonia caused by *Pneumocystic carinii* (Gottlieb, Schroff *et al*. 1981). This led to the isolation of the causative agent human immunodeficiency virus (HIV) in 1983 (Barre-Sinoussi, Chermann *et al*. 1983). HIV is a retrovirus that infects and replicates in human $CD4^+$ T cells and macrophages, causing immune deficiency and ultimately death over many years. Presently, about 40 million people are infected with this virus of which 70% live in sub-Saharan Africa (Figure 1.1).

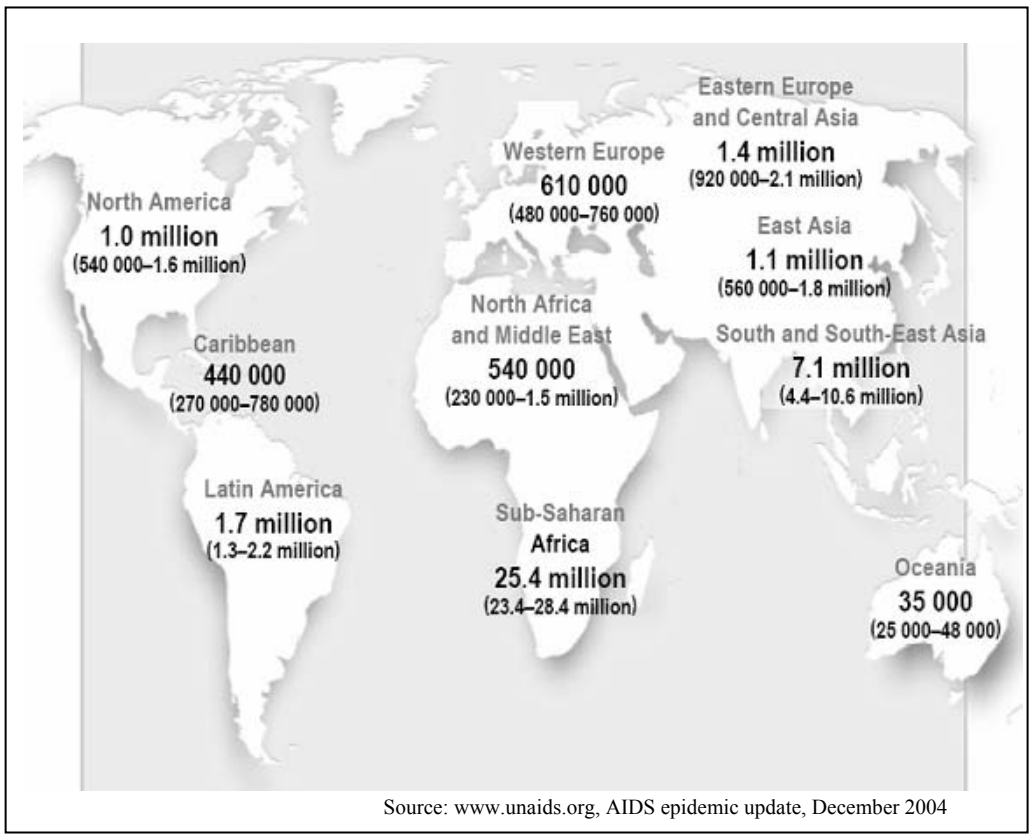


Figure 1.1: Global distribution of adults and children living with HIV-1 at the end of 2004.

The origin of HIV has been linked to the simian immunodeficiency viruses (SIV) from the genus *Lentiviruses*, belonging to the *Retroviridae* family. SIV is found in a variety of African primates, but these viruses do not cause disease in their primate hosts. In the human population there are two types of genetically distinct circulating HIV (HIV-1 and HIV-2). HIV-1 is closely related to SIVcpz isolated from a sub-species of chimpanzee (Gao, Bailes *et al*. 1999; Santiago, Rodenburg *et al*. 2002)**,** and HIV-2 is related to SIVsm found in sooty mangabey monkeys (Gao, Yue *et al*. 1992; Rambaut, Posada *et al*. 2004) (Figure 1.2).

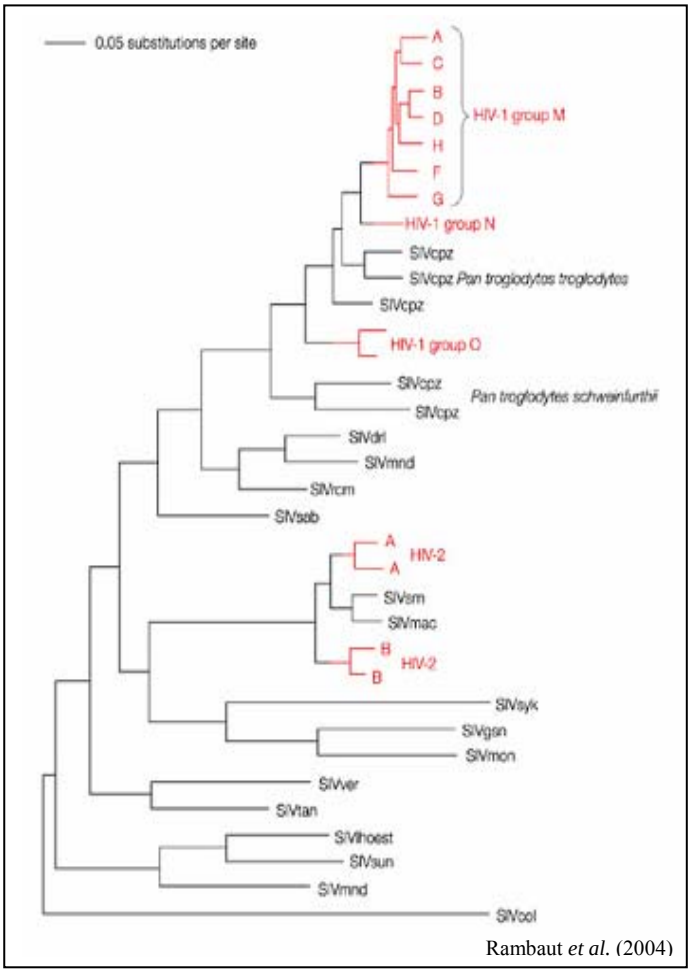


Figure 1.2: The evolutionary history of the primate lentiviruses. HIV-1 and HIV-2 lineages (red branches) group with SIVcpz and SIVsm respectively, representing independent cross-species transmission events. The tree also indicates that HIV-1 groups M, N and O originated from separate transfers of SIVcpz.

HIV-1 is further divided into three groups (M, N and O), of which group N and O are mostly restricted to West Africa. Group M consist of 9 subtypes (A - J) and 18 circulating recombinant forms (CRFs), and this group is more globally distributed (Figure 1.3).

In South Africa, HIV-1 subtype C is the most predominant circulating strain with an estimated 5.3 million infected people (www.unaids.org). The primary route of transmission is through heterosexual intercourse and from mother to child. This subtype has some genotypic and phenotypic characteristics that differ from other subtypes, such as the predominant use of CCR5 as a coreceptor (Bjorndal, Sonnerborg *et al*. 1999; Ping, Nelson *et al*. 1999; Cilliers, Nhlapo *et al*. 2003). The genetic changes within HIV-1 that influence viral phenotype and cell tropism are focused on the envelope glycoprotein of the virus.

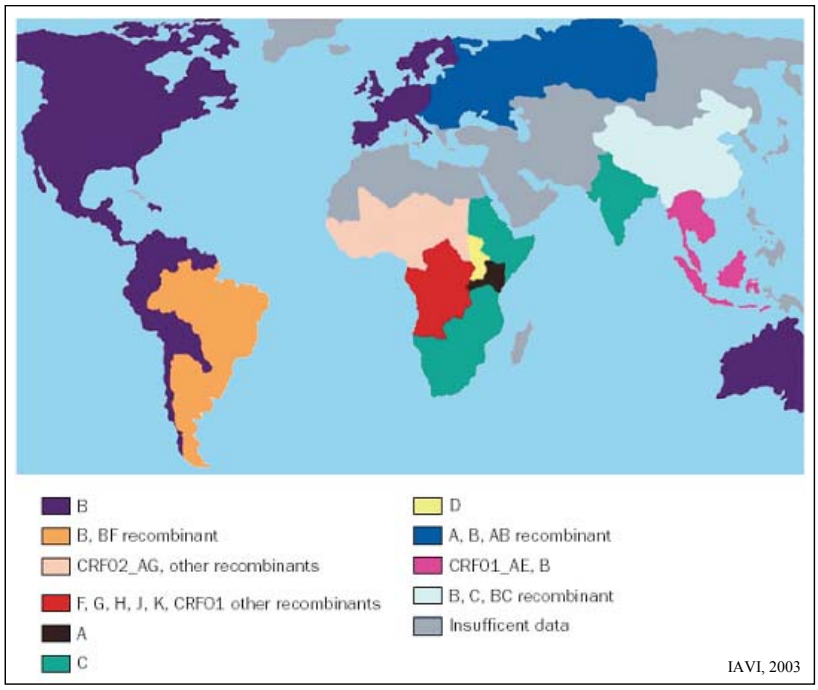


Figure 1.3: Global distribution of the most predominant HIV-1 subtypes and recombinants.

1.2 ENTRY OF HIV-1 INTO THE HOST CELL

1.2.1 Viral genome and structure

HIV-1 has a 9.8kb genome that encodes structural and enzymatic proteins (gag, pol and env), RNA-binding regulatory proteins (tat, rev and nef) and accessory proteins (vif, vpr, vpu) (Figure 1.4 A).

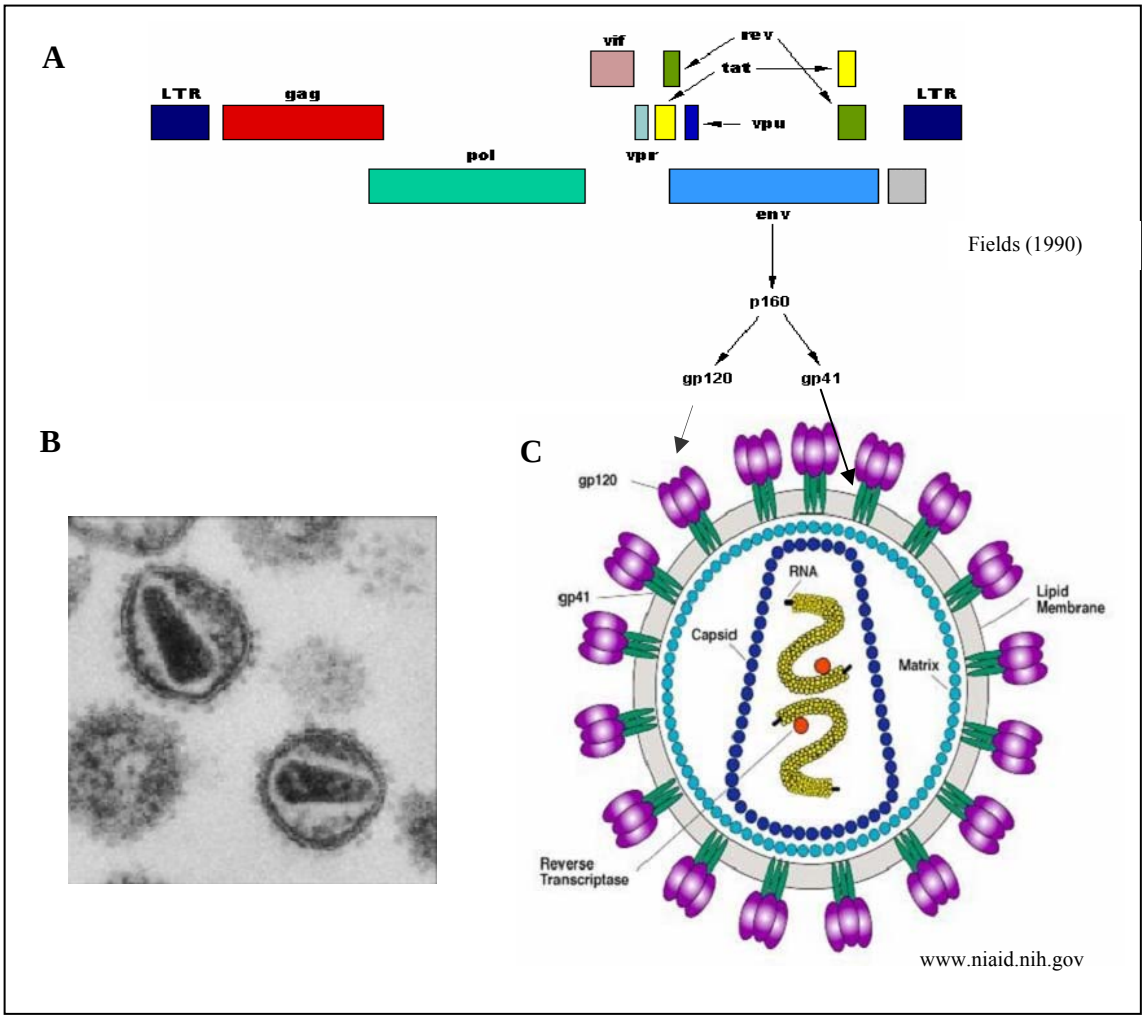


Figure 1.4: HIV-1 viral genome and structure. **A,** schematic representation of the HIV-1 genome depicting the structural and enzymatic proteins. **B,** Electron microscopy of viral particles and **C,** schematic picture of HIV-1 virion indicating the glycoproteins gp120 and gp41 that are non-covalently linked on the virion surface as trimers.

The gag (group-specific antigen) gene encodes the components of the inner capsid protein, the pol (polymerase) gene produces the enzymes such as reverse transcriptase and protease that are used in viral replication, and the env (envelope gene) encodes for the glycoprotein precursor gp160. The glycoprotein precursor gp160 is cleaved intracellularly to form the functional gp120 and gp41 subunits, that are non-covalently linked on the virion surface as trimers (Chan, Fass *et al*. 1997; Wyatt and Sodroski 1998) (Figure 1.4 B-C). These subunits are involved in viral entry and play a pivotal role in coreceptor usage.

1.2.2 Viral entry

The entry of HIV-1 into a host cell is facilitated by the gp120 subunit that binds to the primary receptor, the CD4 glycoprotein (Figure 1.5). CD4 is expressed on the surface of T lymphocytes, monocytes, dendritic cells and brain microglia, all of which are infected by HIV-1. This binding leads to conformational changes in the gp120 that results in the exposure and/or formation of a binding site for a chemokine receptor (Kwong, Wyatt *et al*. 1998).

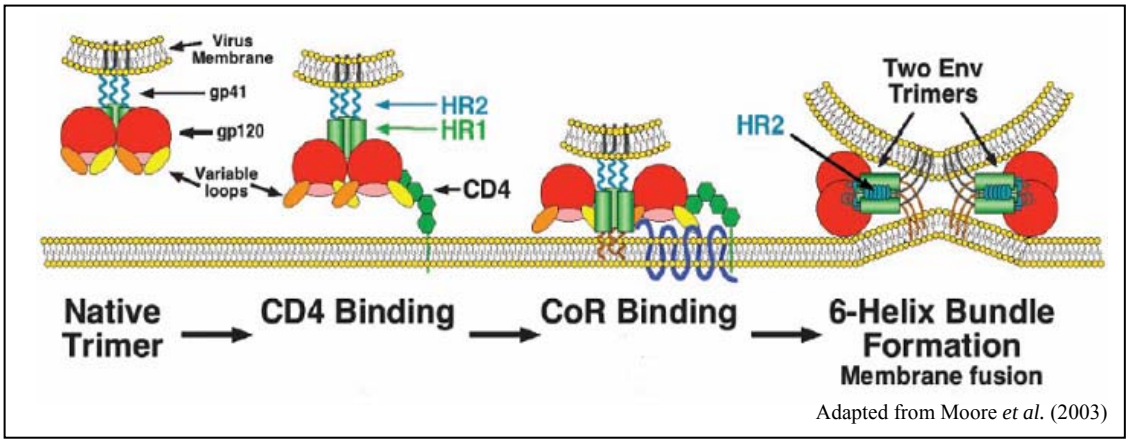


Figure 1.5: A model for HIV-1 entry into the target host cell. The variable loops cover the coreceptor binding sites, which are exposed through conformational changes. The gp120 then binds to CD4, followed by coreceptor binding and the formation of a 6-helix bundle and fusion of the cell and viral membranes.

The two major coreceptors involved in HIV-1 entry are CCR5 and CXCR4 (Deng, Liu *et al*. 1996; Dragic, Litwin *et al*. 1996; Feng, Broder *et al*. 1996). Most HIV-1 isolates that are transmitted and predominate during early infection use CCR5 as coreceptor (Michael, Chang *et al*. 1997). The gp41 subunit undergoes conformational changes that result in the fusion of the viral and host membranes (Sattentau, Zolla-Pazner *et al*. 1995; Chan, Fass *et al*. 1997). RNA and enzymes needed for viral replication are injected from the viral core to the host cell cytoplasm where it undergoes reverse transcription to cDNA (Figure 1.6). The cDNA is incorporated into the host genome followed by a series of events including viral replication and budding from the host cell to infect new cells (Greene 1993).

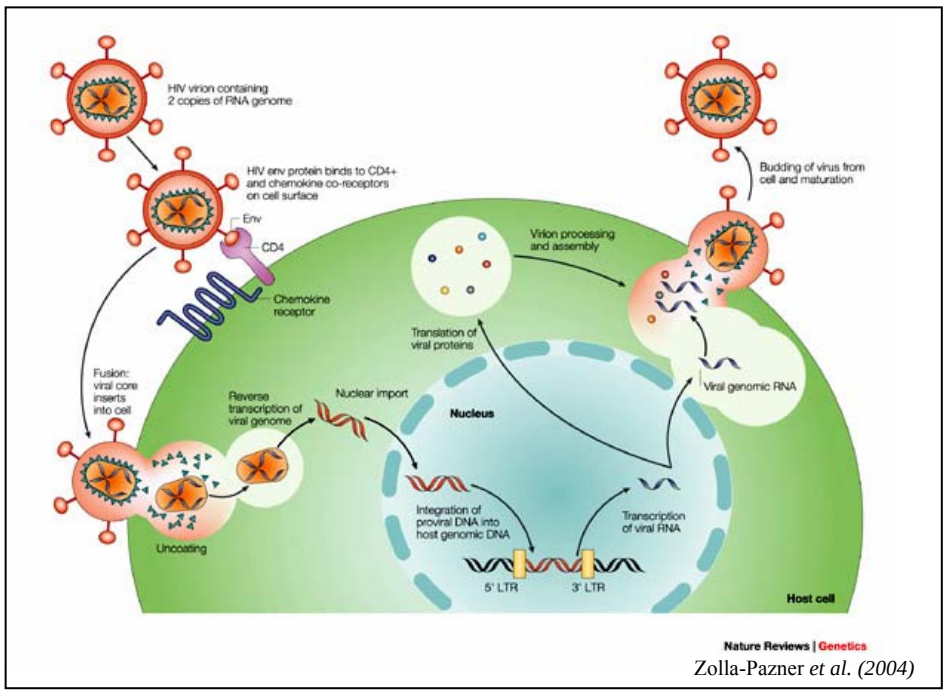


Figure 1.6: Key aspects of HIV-1 viral life cycle. Shown are the binding and entry of HIV-1 into the host cell, uncoding of the viral RNA and integration of proviral DNA into the host genomic DNA. This results in the amplification and transcription of viral RNA followed by viral processing, assembly of new virions and budding from the host membrane.

1.2.3 Coreceptors

The two major coreceptors involved in HIV-1 binding, CCR5 and CXCR4 are seven-transmembrane molecules from the G-protein-coupled receptor family (Berger 1997) (Figure 1.7). Both receptors consist of 352 amino acids with three extracellular loops on the cell surface, but have different functions within the immune system (Berson, Long *et al*. 1996). CXCR4 plays an important role in re-circulation of lymphoid cells from tissues to secondary lymphoid organs, whereas CCR5 is a key component of inflammatory processes during immune defence mechanisms (Loetscher, Moser *et al*. 2000).

Studies have shown different regions of these coreceptors to be involved in gp120 binding and entry. A cluster of residues, in particular tyrosine at the N-terminal of CCR5, participates in entry of R5 and R5X4 viruses (Rucker, Samson *et al*. 1996; Dragic, Trkola *et al*. 1998; Farzan, Choe *et al*. 1998; Rabut, Konner *et al*. 1998). Dispersed residues from the extracellular region of CXCR4 are involved in entry of HIV-1, especially the second extracellular loop (ECL-2) (Lu, Berson *et al*. 1997; Kajumo, Thompson *et al*. 2000). These receptors also have different cell surface charges with CCR5 almost neutral, while CXCR4 is negatively charged, suggesting charge-related interaction between the coreceptor and gp120, with higher positive charge of gp120 associated with CXCR4 usage (Moore and Stevenson 2000). Of note, is that most chemokine receptors described to date have negatively charged regions in their extracellular domains, but do not mediate HIV-1 entry and the specific features that make these receptors permissive for HIV-1 remain unknown (Cormier and Dragic 2002).

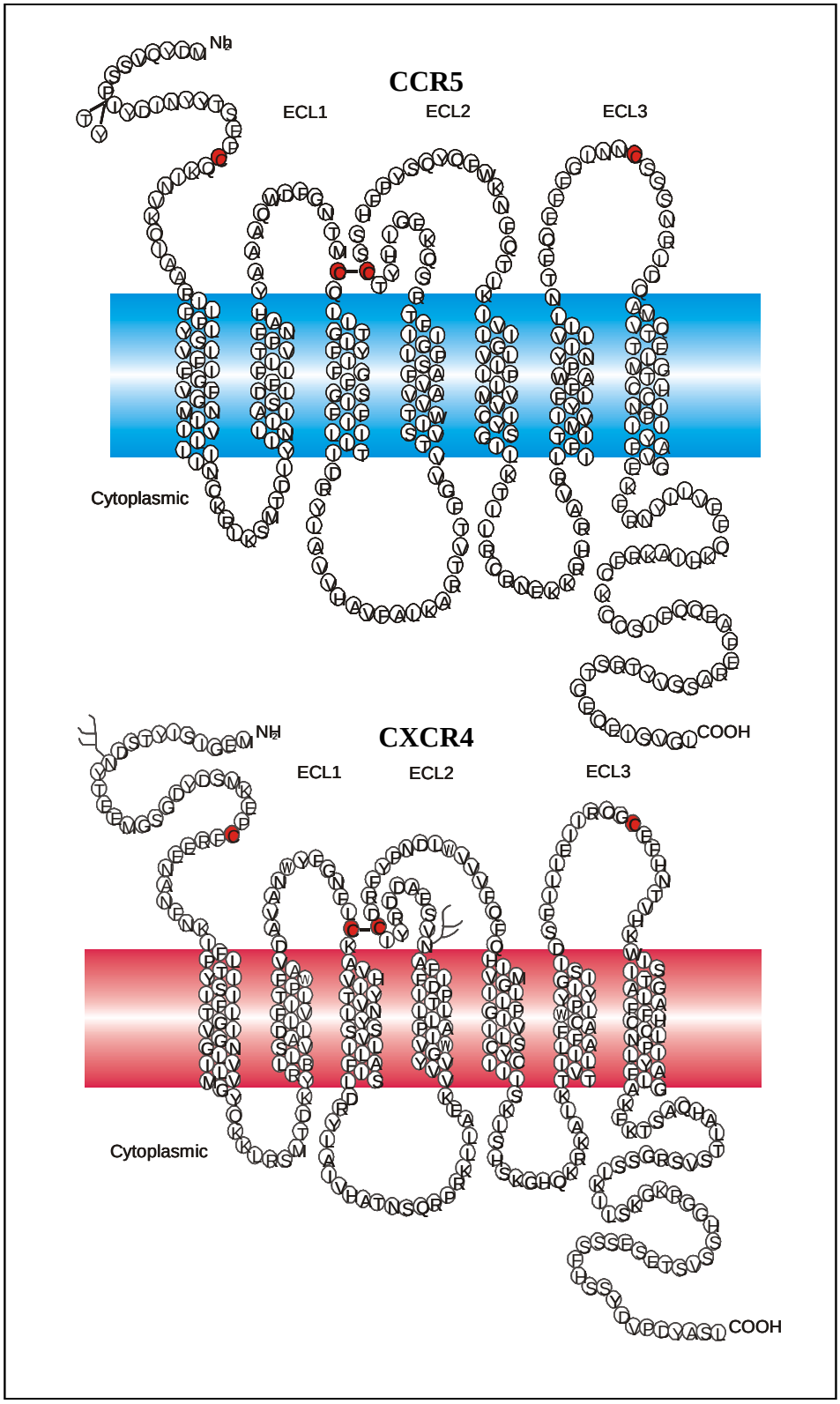


Figure 1.7: Two-dimensional structure of CCR5 (Doranz, Rucker et al. 1996) and CXCR4 (Berson, Long et al. 1996). These coreceptors are seven-transmembrane molecules from the G-protein-coupled receptor family and the major coreceptors involve in viral entry. The extracellular loops (ECL) are indicated, and cysteine residues are shown in red.

1.2.4 Minor receptors

Some rare HIV-1 isolates are able to use other coreceptors for entry such as CCR1, CCR2b, CCR3, CCR8, Bonzo (STRL33), BOB (GPR15) and GPR1 (Edinger, Hoffman *et al*. 1998; Zhang, Dragic *et al*. 1998). The use of minor coreceptors has more commonly been reported among HIV-2 and SIV isolates (Guillon, van der Ende *et al*. 1998; McKnight, Dittmar *et al*. 1998; Willey, Reeves *et al*. 2003). In subtype C, a few HIV-1 isolates have been identified that are able to use these receptors (Choge, Cilliers *et al*. 2005; Cilliers, Willey *et al*. 2005). However, sequence analysis of the gp160 region of these isolates provided no clues as to what genetic changes within the envelope might give the virus the ability to use these minor receptors (Cilliers, Willey *et al*. 2005). The use of minor receptors is infrequent and associated with late stage disease and their role *in vivo* is still unclear (Pohlmann, Krumbiegel *et al*. 1999). Nevertheless, they could complicate the efficacy of entry inhibitors that focus on CCR5 and CXCR4 coreceptors.

1.2.5 Phenotype and tropism

HIV-1 isolates can be differentiated based on their ability to use particular coreceptors R5 viruses (formerly known as non-syncytium-inducing or NSI viruses) utilising CCR5, X4 viruses using CXCR4 and R5X4 viruses able to use both receptors (collectively know as syncytium-inducing or SI isolates) (Berger, Doms *et al*. 1998) (Figure 1.8). CCR5 is most commonly used during transmission and early infection (Michael, Chang *et al*. 1997). Viruses able to use CXCR4 emerge later in the course of infection in about 50% of individuals infected with subtype B viruses (Richman and Bozzette 1994; Connor, Sheridan *et al*. 1997; Scarlatti, Tresoldi *et al*. 1997).

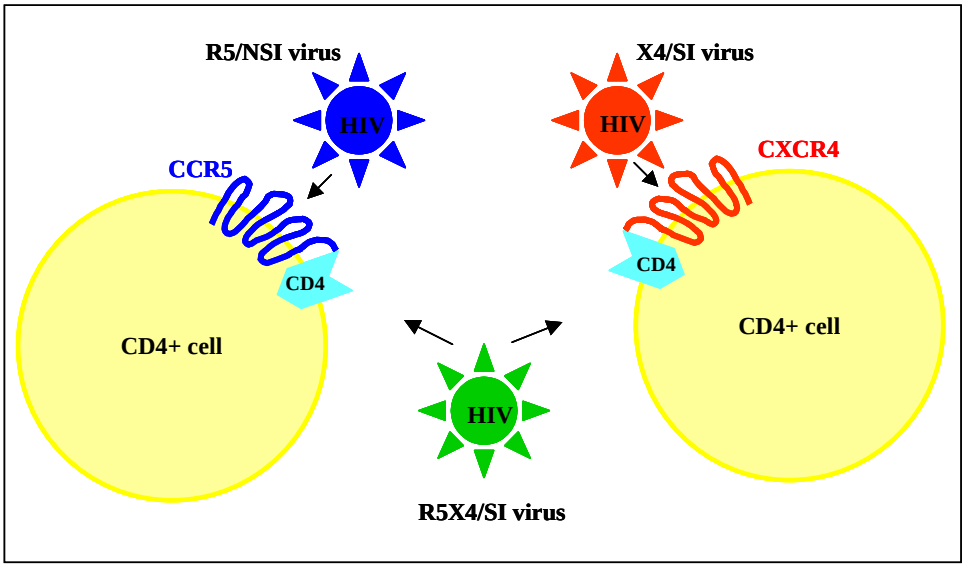


Figure 1.8: HIV-1 phenotype determined by coreceptor usage. R5/NSI viruses (blue) use CCR5 for entry, X4/SI viruses (red) use CXCR4 for entry and R5X4 viruses (green) use CCR5 and/or CXCR4 for entry into the host cell.

The ability of HIV-1 to use CXCR4 as a coreceptor is associated with accelerated disease progression (Koot, Keet *et al*. 1993; Glushakova, Grivel *et al*. 1998; Berkowitz, Alexander *et al*. 2000), therefore the evolution of R5 viruses to X4 viruses may have important implications for pathogenesis. Dualtropic viruses are seen as the intermediate step in the transition from R5 to X4, as these viruses have the ability to use both CCR5 and CXCR4 as receptors. There seems to be a fitness cost involved in being able to use more than one receptor; as dualtropic viruses gain the ability to utilise CXCR4 interaction with CCR5 becomes less efficient and more sensitive to entry inhibitors (Lu, Berson *et al*. 1997). Investigating the probability and dynamics of an R5 virus to X4 virus switch, could assist in understanding what genotypic adaptations within the envelope are required for specific coreceptor usage.

1.2.6 Importance of coreceptor determination

Identification of the coreceptors used in HIV-1 entry is important because it is indicative of viral tropism and pathogenicity of HIV-1 *in vivo* (Koning, van Rij *et al*. 2002). Coreceptor utilisation also needs to be considered in drug treatment strategies.

It has been shown that the presence of X4 viruses in patients on HAART is suggestive of a poor response to treatment (Brumme, Dong *et al*. 2004). In addition, antiretroviral treatment may create an environment for the emergence of CXCR4-using HIV-1 viruses (Johnston, Zijenah *et al*. 2003). This might be due to the increasing survival of patients with low CD4 cells on ARV (Gervaix, Nicolas *et al*. 2002; Pierdominici, Giovannetti *et al*. 2002), the reduced expression of CCR5 on PBMC during ARV treatment (Nicholson, Browning *et al*. 2001) or pre-existing X4 variants in naive $CD^{+}4$ T cells (Delobel, Sandres-Saune *et al*. 2005). In countries, such as South Africa, where ARV is not accessible to most HIV-1 infected persons, there is a low prevalence of X4 viruses. Whether there is a direct correlation between ARV treatment and the emergence of X4 viruses is unclear and requires further study. With the increased availability of ARV, it will be important to investigate the determinants of coreceptor usage as well as other factors that might contribute to the development of X4 viruses.

1.2.7 Coreceptors and entry inhibitors

Natural ligands for CCR5 and CXCR4 exist that can block HIV-1 infection. The CCR5 chemokine ligands MIP-1α, MIP-1β and RANTES block R5 viruses, while SDF-1, the ligand for CXCR4, block X4 viruses from entering T-cells (Cocchi, DeVico *et al*. 1995; Oberlin, Amara *et al*. 1996). This has lead to the design of synthetic ligands that target viral entry. Entry inhibitors that act at all three steps in this process have been developed. The CD4-gp120 interaction is inhibited by PRO542, a CD4-immunoglobin fusion protein (Allaway, Davis-Bruno *et al*. 1995). PRO140 (an anti-CCR5 monoclonal antibody) and TAK779 (a small molecule that binds to ECL2 of CCR5) prevent the binding of HIV-1 to the coreceptor CCR5 (Wu, LaRosa *et al*. 1997; Baba, Nishimura *et al*. 1999) and AMD3100 inhibits the binding to CXCR4 (Bleul, Farzan *et al*. 1996;

Labrosse, Brelot *et al*. 1998). T-20 (based on the HR-2 region of gp41), also known as Enfuvirtide or Fuzeon, is the first entry inhibitor used as an antiretroviral drug in humans (Kilby, Hopkins *et al*. 1998; Lazzarin, Clotet *et al*. 2003). It binds to the HR-1 region of gp41 and prevents fusion of the viral and host membranes (Wild, Shugars *et al*. 1994). Some studies have shown that V3 sequences influence the sensitivity of HIV-1 to T-20, and on average that X4 viruses were more sensitive than R5 isolates (Derdeyn, Decker *et al*. 2000; Reeves, Gallo *et al*. 2002), although Cilliers *et al.* (2004) did not observe this in subtype C isolates tested.

1.2.8 Coreceptor usage of South African subtype C isolates

Subtype C is the most prevalent subtype globally and R5 viruses dominant at all stages of disease, including late stage AIDS. To date only limited numbers of CXCR4-using viruses have been described and characterised (Abebe, Demissie *et al*. 1999; Ping, Nelson *et al*. 1999; Batra, Tien *et al*. 2000; Cilliers, Nhlapo *et al*. 2003; Johnston, Zijenah *et al*. 2003). This suggests there may be factors limiting the development of X4 viruses in subtype C. Whether this is due to host immune or virological constraints remains unknown. Determining what defines a subtype C X4 virus might increase the likelihood of identifying these viruses, improve understanding of the factors involved in their development, as well as the role CXCR4 usage plays in disease progression of subtype C HIV-1.

1.3 HIV-1 ENVELOPE GLYCOPROTEIN

1.3.1 Structure of the envelope protein

The envelope glycoprotein is organised into oligomeric, most probably trimeric spikes on the virion surface anchored into the viral membrane by the gp41 transmembrane envelope protein (Kwong, Wyatt *et al*. 1998). The gp120 protein structure consists of three domains, an inner and an outer domain, connected by a third domain, called the bridging sheet (Figure 1.9).

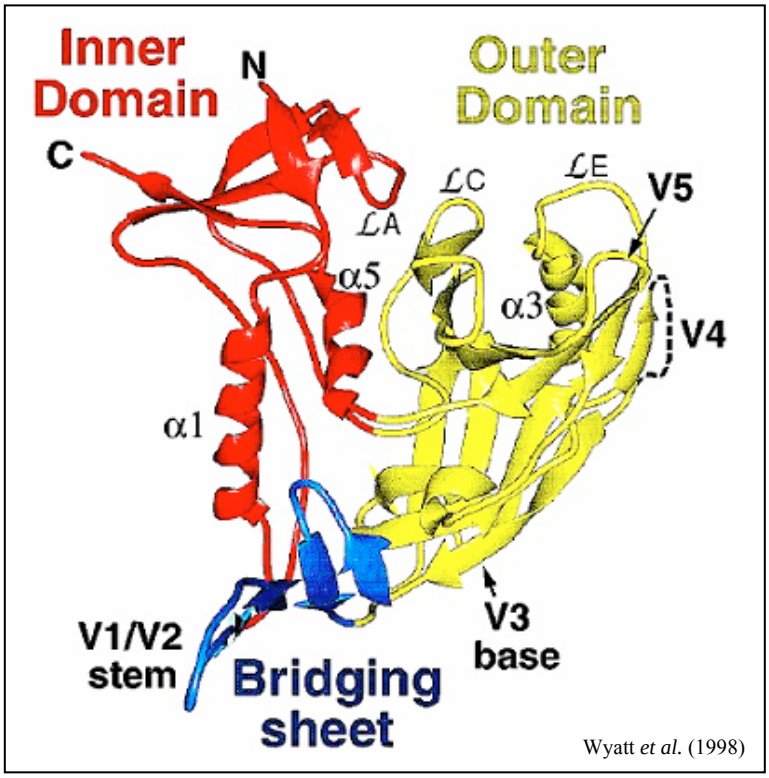


Figure 1.9: Three-dimensional structure of the gp120 glycoprotein. The inner domain (in red) with the variable loops V1-V2, outer domain and V3-V5 loops (yellow) and bridging sheet (blue) that are involved in binding of gp120 to a chemokine receptor.

The inner domain that contains the N- and C-termini, is thought to interact with gp41 and is inaccessible to antibodies (Cao, Bergeron *et al*. 1993). The outer domain consists of the variable regions (V1-V5) interspaced between the conserved regions (C1-C5) and is exposed on the surface of the trimer (Wyatt and Sodroski 1998). The variable regions

form surface exposed loops and are important for virus binding to the host cell receptors. The conserved regions are involved in interaction with gp41 and the receptors on the target cell (Kwong, Wyatt *et al*. 1998). The bridging sheet consists of four anti-parallel β-strands and is involved in the binding of gp120 to the chemokine receptor (Kwong, Wyatt *et al*. 1998).

1.3.2 Glycosylation of the envelope protein

The gp120 protein is highly glycosylated with complex glycans in the variable regions and high-mannose or hybrid glycans present in the conserved regions of gp120 (Leonard, Spellman *et al*. 1990; Zhu, Borchers *et al*. 2000) (Figure 1.10).

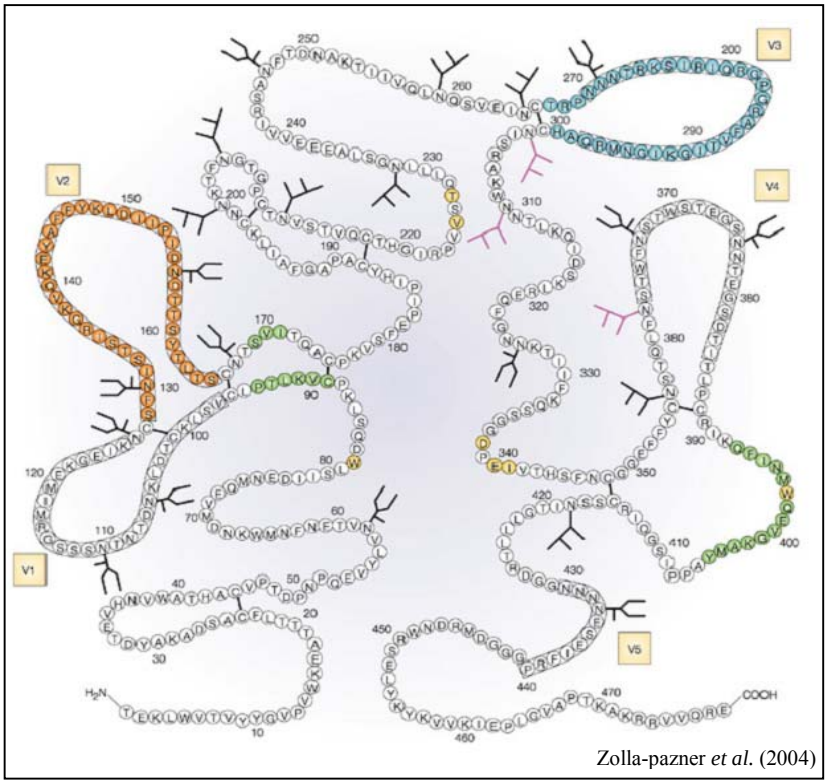


Figure 1.10: A schematic diagram of the gp120 envelope glycoprotein and potential glycosylation sites. High mannose-type and/or hybrid-type oligosaccharide glycosylation indicated by the branched structures, and complex-type oligosaccharide glycosylation sites indicated by the U-shaped branches. The variable loops are indicated (V1–V5). Epitopes able to induce neutralising antibodies are highlighted in color including the V2 loop (orange) and the V3 loop (blue).

Glycosylation is essential for correct folding and processing of the viral envelope protein (Land and Braakman 2001). It also provides protection from recognition by neutralising antibodies (known as a glycan shield) (Wei, Decker *et al*. 2003). The number and position of potential glycosylation sites varies over time and are thought to be an escape mechanism. Some glycans within the envelope regions have been correlated to viral phenotype, in particular the presence of a glycosylation site within the V3 region associated with CCR5 usage (Zhang, Gaschen *et al*. 2004).

1.3.3 Genetic diversity of the envelope glycoprotein

HIV-1 is characterised by a high degree of genetic diversity particularly in the envelope gene. HIV-1 populations within an individual are continually evolving and can differ by as much as 10% in sequence at end-stage disease (Shankarappa, Margolick *et al*. 1999). Viral diversity is influenced by various factors. These include viral characteristics such as error-prone reverse transcriptase and high viral turnover, host factors such as HLA type of patient as well as cellular and humoral immune responses to HIV-1. External influences such as opportunistic infections can also contribute to viral diversity by promoting viral replication. The rapid viral turnover (10^{10} viral particles/day) in an HIV infected individual, as well as the high rate of incorrect nucleotide substitutions during HIV reverse transcription (10^{-4}/nt, resulting in one nucleotide miss-incorporation per replication cycle within a 10kb genome) in the absence of proof-reading mechanisms, results in extensive viral heterogeneity (Preston, Poiesz *et al*. 1988; Keulen, Nijhuis *et al*. 1997; Drosopoulos, Rezende *et al*. 1998). Under the selective pressure of immune responses, neutralising antibody and CTL escape variants with mutations in the gp120 and gp41 regions arise (Wolfs, Zwart *et al*. 1991; Hogervorst, de Jong *et al*. 1995; Yoshida, Nakamura *et al*. 1997). Characterising

these viral genetic adaptations within the envelope might shed some light on the biological consequences of this genetic diversity, as well as its impact on cellular and humoral responses in an infected host.

1.3.4 Envelope protein diversity and disease progression

Various studies have investigated the correlation between disease progression and viral heterogeneity, focusing on the envelope protein and in particular the V3 region as this region contains recognition sites for humoral and cellular immune responses as well as cell tropism (Goudsmit, Debouck *et al*. 1988; Safrit, Lee *et al*. 1994). HIV-1 has various adaptive methods to escape this immune response resulting in viral heterogeneity which is an important determinant of disease progression (Lukashov and Goudsmit 1998; Kalpana, Srikanth *et al.* 2004). Transmission of HIV-1 is associated with the transfer of relatively homogenous populations to the recipient. (Korber, Wolinsky *et al*. 1992; McNearney, Hornickova *et al*. 1992; Zhang, MacKenzie *et al*. 1993). This indicated that there might be selection for specific variants during transmission influenced by cell tropism (as most HIV-1 viruses transmitted are CCR5-using). Primary infection is associated with low sequence diversity, which rapidly increases with disease progression and as the host immune response deteriorates, diversity decreases in late stage disease (Delwart, Pan *et al*. 1997). Genotypic adaptations within the gp160 region, include change in length and charge of variable loops, as well as changes in the number and position of potential glycosylation sites (Wei, Decker *et al.* 2003). Derdeyn *et al*. (2004) showed in a heterosexual cohort (with donor and recipient pairs), that the transmitted variants had significantly shorter V1-V4 regions, suggesting that loop length increases with disease progression. Longer V2 length has also been correlated to slower disease progression (Shioda, Oka *et al*. 1997;

Masciotra, Owen *et al*. 2002). It has been proposed that longer variable loops provide a sheild to the more neutralising sensitive regions in V3 and the bridging sheet (Wolinsky, Korber *et al*. 1996; Delwart, Pan *et al*. 1997).

1.3.5 Conformational changes during entry

During entry of HIV-1 into a host cell, gp120 interacts with CD4. This interaction takes place within a cavity of gp120 formed by the three domains (inner domain, outer domain and bridging sheet) that are very conserved with no carbohydrates. The cavity is also close to the base of the V3 and V1-V2 loops that could mask the gp120 site before CD4-gp120 interaction, possibly explaining their involvement in coreceptor preferences. The actual CD4-gp120 contact is between 22 CD4 amino acid residues (concentrated between residues 25-64) and 26 gp120 amino acid residues distributed within gp120 (Kwong, Wyatt *et al*. 1998). This binding orientates the viral spike and induces conformational changes within gp120. This includes movement of the V2 loop that results in exposure of the V3 loop. Sattentau *et al.* (1993) suggested that the CD4 binding also changes the V3 region, making it more susceptible to proteolytic activity. The uncovered V3 region participates in the coreceptor binding and variation in this region determine which coreceptor is used.

The coreceptor binding causes further conformational changes in the HIV-1 envelope that exposes the gp41 ectodomain (Weissenhorn, Dessen *et al*. 1997; Kwong, Wyatt *et al*. 1998). The HIV-1 gp41 ectodomain has a trimeric coiled coil structure that consists of a fusion peptide and 2-terminal heptad regions (HR-1 and HR-2), which fold into each other to form a six-helical bundle. This results in a hydrophobic gp41 NH_2-

terminus that is inserted into the target cell membrane, leading to target cell and viral membrane fusion (Sattentau, Zolla-Pazner *et al*. 1995; Chan, Fass *et al*. 1997).

1.3.6 The V3 region as a major determinant of coreceptor usage

The third variable loop (V3) of gp120 is generally 35 amino acids long (Figure 1.11) and the major determinant of coreceptor usage (Hartley, Klasse *et al*. 2005). Within the V3 region, variation of specific amino acids, increased positive amino acid charge and glycosylation are some of the characteristics that have been identified as influencing coreceptor usage in subtype B. In particular, the amino acids at positions 11 and 25 are used to distinguish between NSI- (R5) and SI-like (R5X4 and X4) viruses with SI viruses often having a positive amino acid at these positions (De Jong, De Ronde *et al*. 1992; Fouchier, Brouwer *et al*. 1995; Hoffman, Seillier-Moiseiwitsch *et al*. 2002). As a result, the net V3 charge is often a good indicator and determinant of viral tropism, with a high positive charge correlating with CXCR4 usage. Another factor influencing CCR5 usage appears to be the predicted N-linked glycosylation site at positions 6-8 within the V3 region, as loss of this glycan is associated with less efficient usage of CCR5 and in some cases enhanced ability to use CXCR4 for entry (Ogert, Lee *et al*. 2001; Polzer, Dittmar *et al*. 2002).

The role of these V3 genetic characteristics and the influence they may have on subtype C coreceptor usage remain to be evaluated, as to date only a few isolates able to use CXCR4 have been identified (Abebe, Demissie *et al*. 1999; Batra, Tien *et al*. 2000; Cilliers, Nhlapo *et al*. 2003; Johnston, Zijenah *et al*. 2003). Presently, R5 (NSI) subtype C HIV-1 is characterised by a GPGQ crown motif, an overall charge of +3 to +5 and 35 amino acids in length. The X4 viruses have an increased charge from +6 to +9, with

some variation in the crown and in some cases an insertion between positions 13 and 14 of mostly isoleucine and glycine (Coetzer, Cilliers *et al*. 2005; Huang, Tang *et al*. 2005).

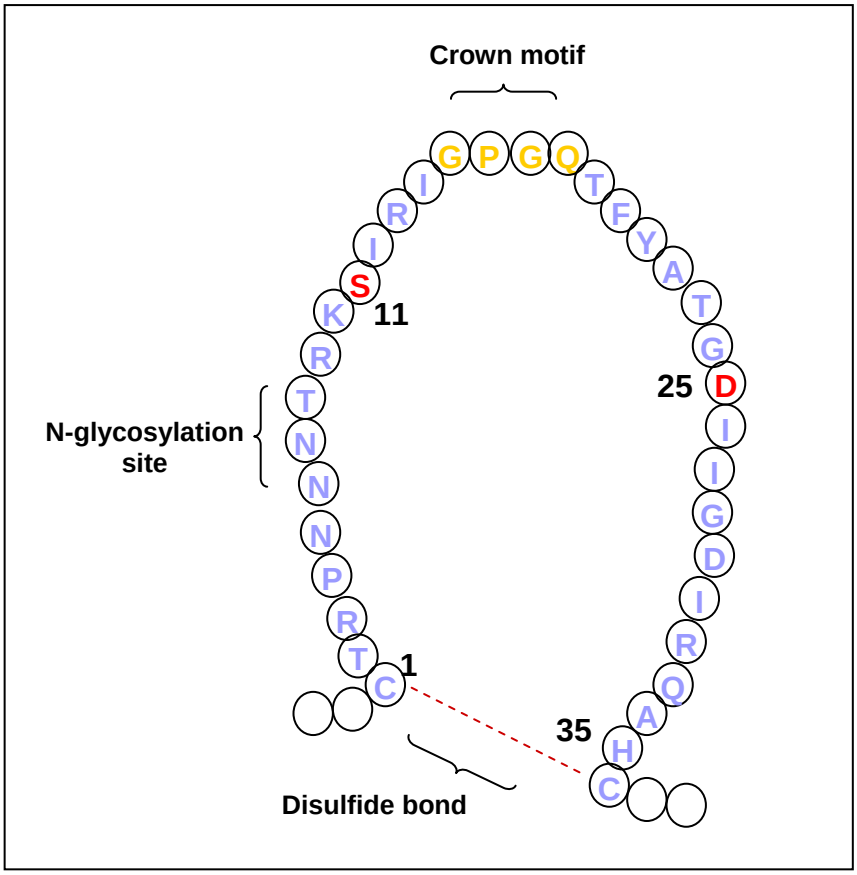
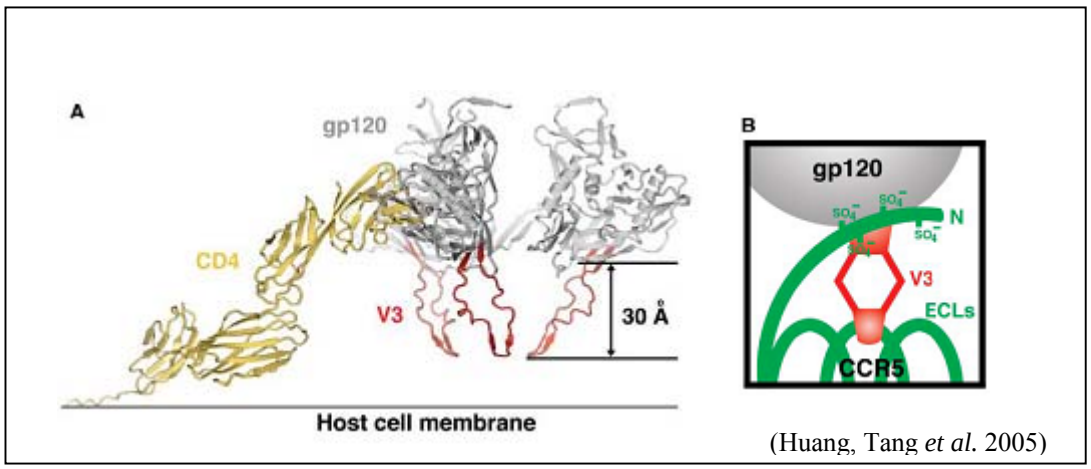


Figure 1.11: A schematic representation of the V3 loop. This region usually consists of 35 amino acids indicated inside the loop and positions 11 and 25 also highlighted as these positions play a role in coreceptor usage.

The conformation of the V3 domain might also be an important determinant of tropism/biological properties of gp120 (Fouchier, Groenink *et al*. 1992; Shioda, Levy *et al*. 1992). Predicted secondary structure based on neural network modelling, suggests that the V3 folds into a structure containing two antiparallel β-strands and a short carboxyl-terminal α-helix (LaRosa, Davide *et al*. 1990). The role of the V3 loop during viral entry into the host cell is not clear, as there is no consistent evidence that the V3

loop interacts directly with the primary gp120-CD4 binding step (Hartley, Klasse *et al*. 2005). The positively charged V3 suggests that it does interact with the anionic phospholipid head-groups on the cell membrane to reduce the repulsion effect between membranes, but this is still unproven (Callahan, Phelan *et al*. 1991). Therefore, it is not necessarily sequence-specific requirements that influence coreceptor usage, but the overall conformation of the V3 that define to which coreceptor the CD4-gp120 complex will interact, either CCR5 of CXCR4 (Hartley, Klasse *et al*. 2005).

A crystal structure of gp120 with the V3 region intact has recently been described (Huang, Tang *et al.* 2005). This model shows that the V3 loop extends from the gp120 structure and the base of the loop interacts with the N-terminus of CCR5 and the tip binds to ECL2 (Figure 1.12). The author propose that V3 acts as a 'molecular hook' to engage CCR5 (Huang, Tang *et al.* 2005). Therefore the amino acids within the base and tip of the V3 are important factors in coreceptor selection.



(Huang, Tang *et al.* 2005)

Figure 1.12: Modeled trimer and coreceptor schematic. (A) A trimer with the orientation of the V3 region (in red) towards the target cell. (B) Schematic of interaction between coreceptor (in green) with V3 loop (in red). The N-terminus of the coreceptor binds to the base of the V3 and V3 tip reaching the ECL2 of CCR5.

1.3.7 Other regions involved in coreceptor usage

Other regions have also been implicated in the efficiency of coreceptor usage, such as the V1-V2 loop and the constant regions. In particular glycosylation sites near the V1-V2 have been identified, although these are less important than the glycosylation site in V3. The V2 regions of X4 variants have been associated with length variation and charge increases (Groenink, Fouchier *et al*. 1993; Cornelissen, Mulder-Kampinga *et al*. 1995; Fouchier, Broersen *et al*. 1995; Jansson, Backstrom *et al*. 2001). Position 440 in the C4 region of gp120 has also been linked to viral tropism in some isolates (Carrillo and Ratner 1996; Milich, Margolin *et al*. 1997; Hoffman, Seillier-Moiseiwitsch *et al*. 2002). In this case R5 genotype is associated with arginine and lysine at position 440 (HXB2 numbering) and X4 genotype over represented with glutamic acid and serine, as well as glycine, threonine and glutamine. The V4-V5 region has been shown to contribute to viral tropism, but usually in addition to changes within other regions of the envelope protein (Smyth, Yi *et al*. 1998; Hu, Barry *et al*. 2000).

1.4 PHENOTYPE PREDICTION METHODS BASED ON THE V3 REGION

Phenotype predictions are increasingly being applied to identify and understand HIV-1 biological phenotypes. As a major determinant of coreceptor utilisation, the V3 region has been used in various studies to predict viral phenotype (Jensen and van 't Wout 2003). These bioinformatic methods are very applicable in various research aspects, such as the identification of X4 viruses, relating their presence to disease status and increased understanding of the R5 to X4 transition and the evolution of X4 viruses (Jensen, Li *et al*. 2003). Prediction methods are also applicable in studies where information of coreceptor utilisation within a patient, for example those receiving ARVs or small molecule inhibitors, will facilitate improved treatment. Pharmaceutical companies are also focusing on faster / easier methods in determining coreceptor usage particular with the increased interest in developing coreceptor inhibitory drug therapy. Coreceptor phenotypic assays are expensive and very laborious and a reliable phenotype prediction method, based on sequence, could provide for rapid and less expensive screening. There are numerous methods to predict HIV-1 phenotype from genotype and each has its own advantages and disadvantages, some of which are described below. It should be noted that predicting phenotype based on only 35 amino acids (from the V3 region) underscores the importance of other determinants for coreceptor usage within the gp120.

1.4.1 The 11/25 rule

This method is based on the presence of basic amino acids (R and K) at positions 11 and or 25, associated with the SI phenotype, and neutral or acid amino acids (D and E) for NSI viruses (highlighted in Figure 1.11) (Fouchier, Groenink *et al*. 1992). This method

was developed using subtype B sequences and has been used extensively in studies to differentiate between the NSI and SI phenotype (De Jong, De Ronde *et al*. 1992; Fouchier, Groenink *et al*. 1992; Hoffman, Seillier-Moiseiwitsch *et al*. 2002). Resch *et al.* (2001) have shown this method to be the most reliable predictor of phenotype (NSI/SI), but loses sensitivity when predicting X4 tropism.

1.4.2 Briggs method

This is a multiple regression method based on positive, negative and net V3 charge to determine phenotype and coreceptor usage (Briggs, Tuttle *et al*. 2000). This method was developed using 43 subtype B sequences with known phenotype. Four genotype variables were identified as predictors of phenotype, (i) the number of positively charged amino acids, (ii) number of negatively charged amino acids, (iii) the net V3 charge, and (iv) an isoleucine residue at position 292 (amino acid 30 in V3 loop). An equation to predict viral phenotype based on these genetic variables was calculated as:

Predicted phenotype = 0.94 + [1.68 X (V3 net charge)]
- [1.37 X (total positive charges)]
+ [1.54 X (total negative charges)]
- [1.19, if aa292 = I]

Values	Phenotype
0.5 - 1.4	R5
1.5 - 2.4	R5X4
2.5 - 3.4	X4

1.4.3 Pillai method

This is a machine-learning method, using phenotype classifiers (Pillai, Good *et al*. 2003). Classifiers are rules generated from a data set using decision tree methodology (Figure 1.13). Various classifiers were tested but the support vector machines (SVM) performed the best in determining phenotype. These classifiers were trained to make the distinction between viruses able to use CXCR4 versus those unable to use CXCR4, and hence dualtropic viruses could not be identified. This method is a web based method (http://genomiac2.ucsd.edu:8080/wetcat/v3.html).

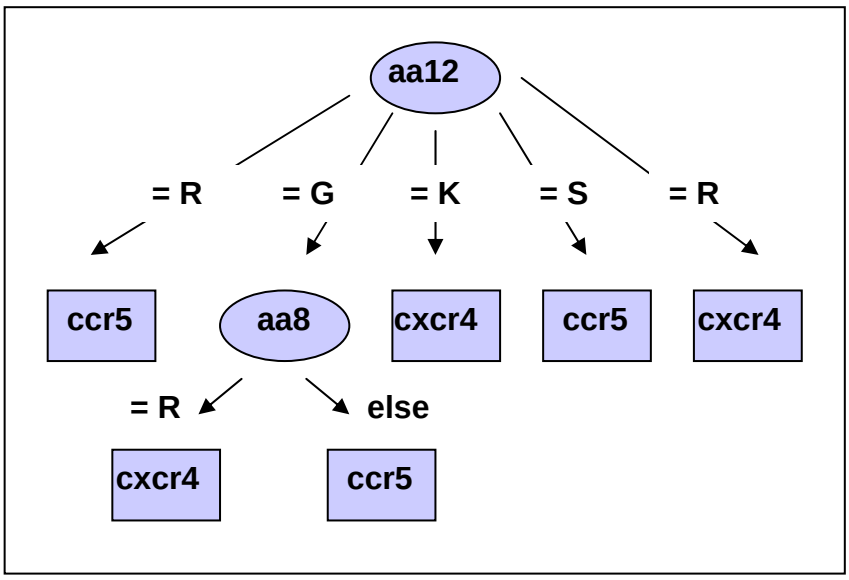


Figure 1.13: Decision tree constructed of amino acid position 12. Depending on the amino acid at position 12 the sequence is classified as 'ccr5' or 'cxcr4', but other positions can also influence the decision at this site, as indicated for amino acid position 8. 'ccr5' indicated that the particular amino acid was associated with CCR5 viruses, similarly 'cxcr4' suggest that the same for the CXCR4-using isolates.

1.4.4 Neural networks

Neural network method is similar to the machine learning method where classifiers are compiled from data with a specific characteristic, in this case coreceptor usage. Neural networks were trained using subtype B data sets with known coreceptor usage (Resch, Hoffman *et al*. 2001). Two networks were developed, an R5/X4 network and a NSI/SI network. The NSI/SI network had similar reliability to the 11/25 rule in predicting phenotype. The R5/X4 network had increased sensitivity in predicting X4 viruses. Resch *et al.* (2001) thus suggested that phenotype may be largely determined by changes at position 11 and 25, but coreceptor usage is more complex and other regions such as V1-V2 and C4 might be involved (Groenink, Fouchier *et al*. 1993; Carrillo and Ratner 1996; Milich, Margolin *et al*. 1997).

1.4.5 Geno2Pheno method

This support vector machine (SVM) learning method was developed for coreceptor prediction, and similar methodology was also applied to phenotypic drug resistance prediction (Beerenwinkel, Daumer et al. 2003; Sing, Beerenwinkel et al. 2004). This is a web-based tool available at (http://www.genafor.org). The majority of the V3 sequences used in developing this method were obtained from the Los Alamos sequences database (http://www.hiv.lanl.gov) and included different subtypes. SVM is a two-class classification method trained to differentiate between CXCR4 versus CCR5 usages. This tool is different from the Pillai method, in that it includes the possibility of choosing the prediction stringency (significance level) and confidence levels (p-value) are given for each prediction. In comparison with other methods the developers of the Geno2Pheno method did observe that this method does perform significantly worse than

the support vector or PSSM methods, but comparable to neural networks (Sing, Beerenwinkel *et al*. 2004).

1.4.6 Position specific scoring matrix method (PSSM)

PSSM detects a non-random distribution of sequences at a specific site within an alignment (Jensen, Li *et al*. 2003) and has been used to identify DNA or protein motifs in various studies (Gribskov, McLachlan *et al*. 1987; Henikoff, Wallace *et al*. 1990). A target sequence is compared to a group of sequences with a known property (in this case CXCR4 or CCR5) and a score is obtained that indicates how closely the target sequence resembles that property. Therefore the higher the score the more X4-like the sequence. This method was demonstrated to have improved predictive power, as well as increased applicability in understanding the transition between R5 to R5X4 to X4 development in subtype B (Jensen and van 't Wout 2003). The method is available at http://ubik.microbiol.washington.edu/computing/pssm.

1.4.7 Predictions on subtype C data sets

Most of these prediction methods have been developed with subtype B sequences, and whether they are applicable to other subtypes such as subtype C is not known. In particular, with the low prevalence of CXCR4 usage in subtype C and not enough sequences available, these methods might not be sensitive enough to detect these X4 viruses in subtype C. Therefore the prediction methods need to be tested on all available subtype C sequences with known coreceptor usage.

1.5 NATURAL HISTORY OF HIV-1 INFECTION

The clinical course of HIV infection can be divided into three stages: primary/acute infection, chronic/asymptomatic and AIDS (Figure 1.14). The primary phase is associated with high levels of viral replication and a decrease in $CD4^+$T cell count (Cooper, Tindall *et al*. 1988; Daar, Moudgil *et al*. 1991; Mellors, Rinaldo *et al*. 1996).

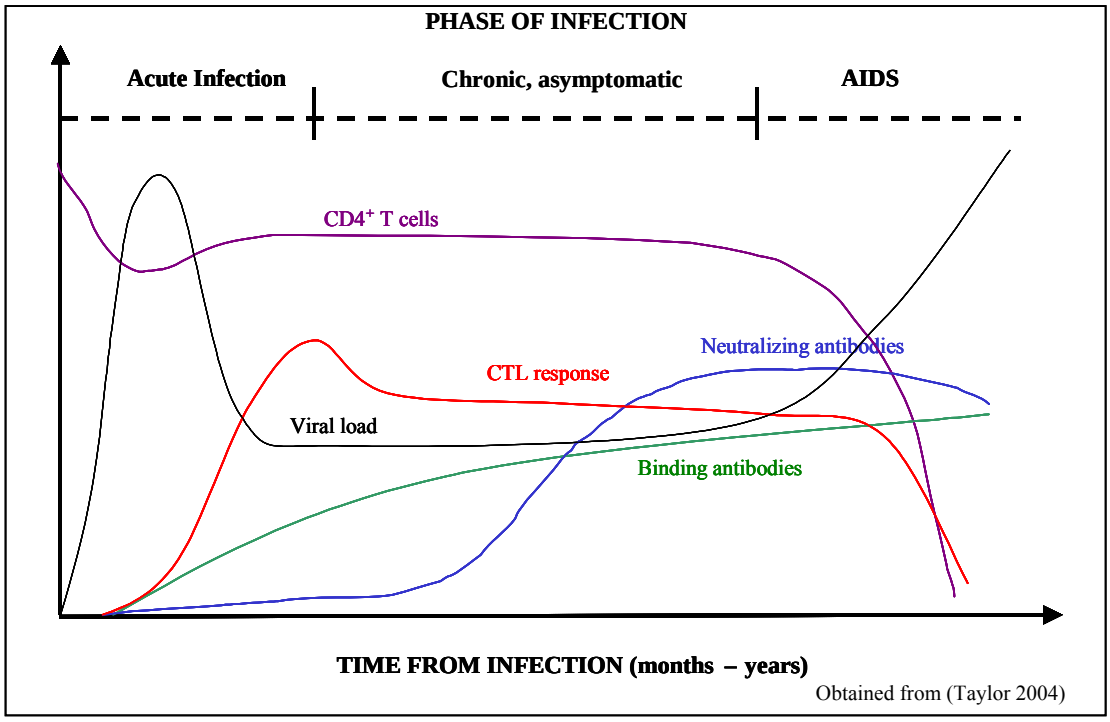


Figure 1.14: The natural history of HIV-1 infection. Viral load increases dramatically in the acute phase and the CD4 T-cells. Within a few weeks, an immune response to HIV develops which curtails viral replication and a return of CD4 T-cell numbers to near normal levels occurs. The control of viremia is attributed to the CTL response and to a lesser extent the neutralising antibody response, which takes longer to develop. As a result of these responses individuals remain clinically well for many years.

Cytotoxic T-lymphocytes are the first immune response to HIV-1, and occur within days of primary infection. It is thought that the decline in viremia after primary infection is due to CTL response (Koup, Safrit *et al*. 1994). Binding antibodies (usually

against gag) appear within weeks after infection and are used in diagnosing HIV-1 within an individual (using an ELISA assay). The CTL defence is followed by neutralising antibodies that block the infection of HIV-1 to host cells (Moore, Sattentau *et al*. 1994; Moore and Sodroski 1996). These antibodies can take up to three months to develop, and although they can neutralise the infecting virus, they often have little to no activity against other strains of virus. It is only during the course of infection, that more broadly neutralising antibodies appear (Wyatt and Sodroski 1998; Richman, Wrin *et al*. 2003; Wei, Decker *et al*. 2003). Due to the HIV-1 specific cellular immune response, viral loads decline and the asymptomatic period follows (Koup, Safrit *et al*. 1994; Moore, Cao *et al*. 1994). During this period there is a gradual deterioration of the immune system and progressive loss of $CD4^+$T cells. Viral load remains relatively constant at a set-point that is indicative of a patient's clinical outcome (Mellors, Rinaldo *et al*. 1996). Most HIV-1 infected individuals develop AIDS after a median of 11 years (Munoz, Wang *et al*. 1989), although this is two to three years shorter in developing countries such as South Africa (Grant and De Cock 2001). This phase is associated with increased viral replication, further loss of $CD4^+$ T cells (below 200) and the onset of opportunistic infections.

1.5.1 Immune activation

There is a complex balance between viral and host factors within a HIV-1 infected individual, as reflected by steady-state level of plasma viremia during the asymptomatic phase. Exogenous immune-activating stimuli such as immunisation, viral, bacterial and parasitic infections in HIV-1 infected patients can disrupt this balance. This disruption is usually associated with transient increases (viral bursts) in HIV-1 viral load (Goletti, Weissman *et al*. 1996; Sulkowski 1998) (Figure 1.15). Immune activation due to

opportunistic infections not only impacts on the dynamics of viral replication, but could promote the expression of variants with specific tropisms (Ostrowski, Krakauer *et al*. 1998). Clerici *et al.* (2000) suggest that the predominance of HIV-1 R5 variants in Africa were immunologically driven due to environmental factors (such as bacterial and parasitic infections) that promote macrophage stimulation. These macrophages contain CCR5 coreceptors increasing potential target cells, and therefore creating a suitable environment that promotes HIV-1 viral replication and disease progression.

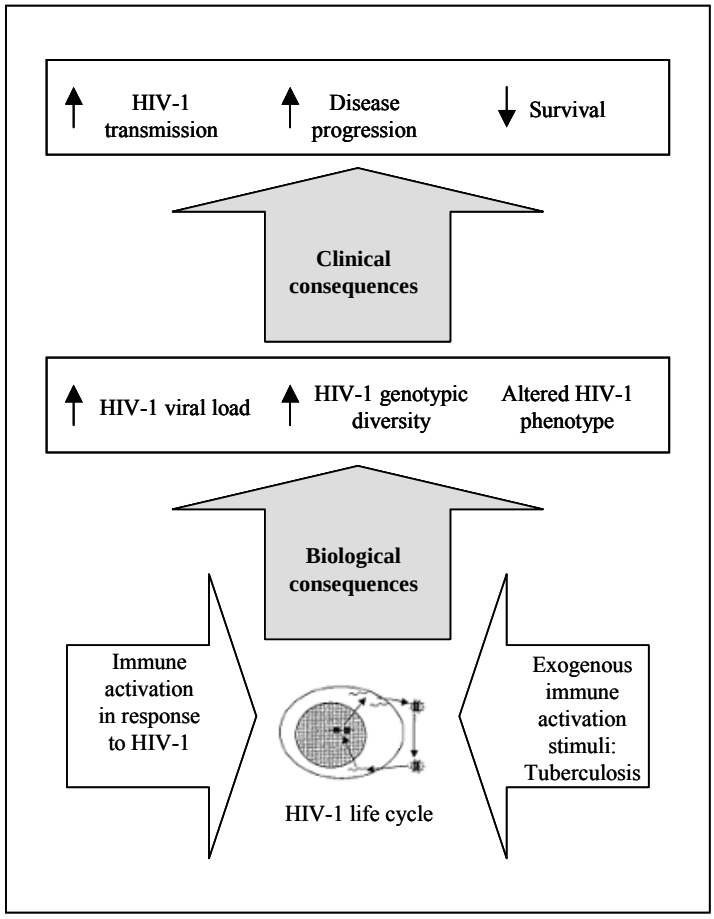


Figure 1.15: Consequences of immune activation on the biology of HIV-1 infection. The subsequent effect of immune activation on HIV-1 transmission, disease progression and survival in HIV-1 infected persons.

1.5.2 Tuberculosis

Tuberculosis (TB) is the most common opportunistic infection in HIV-1 infected patients in South Africa. *Mycobacterium tuberculosis* (MTB) is a slow growing acid-fast bacillus, transmitted primarily by the respiratory route and although it can cause disease in most organs, pulmonary tuberculosis is the most common form (Flynn and Chan 2001). It is estimated that one third of the worlds population will react positively when given a purified protein derived (PPD) skin test, indicating that these individuals have been or still are infected with the latent or chronic phase of TB (Russell 2001). Most people develop an immune response to TB. However HIV-1 infected individuals are unable to mount an effective immune response because of a compromised immune system and are susceptible to TB infection. (Flynn and Chan 2001).

1.5.2.1 Interaction between TB and HIV-1

HIV-1 infection influences the probability, severity and frequency of a tuberculosis infection (Porco, Small *et al*. 2001). The stage of HIV infection and degree of immunodeficiency influences the clinical picture of tuberculosis. In early HIV infection, TB is associated with characteristics of post-primary TB (due to reactivation or re-infection), whereas advanced stages of HIV-1 are associated with an increased frequency of pulmonary disease resembling primary pulmonary tuberculosis and extrapulmonary disease (Raviglione, Harries *et al*. 1997).

In contrast, the effect of TB on HIV-1 infection is less clear, but there is a general consensus that tuberculosis enhances HIV-1 replication (Whalen, Horsburgh *et al*. 1995; Goletti, Weissman *et al*. 1996; Del Amo, Malin *et al*. 1999; Toossi, Johnson *et al*. 2001). As the host immune responses to TB infection, T cells and macrophages are

activated and these may harbour reservoirs of HIV-1 (Wallis, Vjecha *et al*. 1993). HIV-1 entry is dependant on the receptors expressed on host cells and immune activation (due to opportunistic infection) can change the expression of coreceptors, making these cells more susceptible to HIV-1 infection (Lawn, Shattock *et al*. 1999). This might result in the preferential amplification of specific variants with increased fitness, resistance to antiretroviral treatment, neutralising antibodies or expansion of coreceptor usage and these factors could contribute to HIV-1 disease progression.

1.6 OBJECTIVES OF THIS STUDY

The aim of this study was to investigate the *in vivo* and *in vitro* diversity of HIV-1 subtype C isolates from South Africa and correlate genotype with biological phenotype, in particular focusing on the genetic determinants that influence coreceptor usage in subtype C. In this study the envelope gene (and specifically the V3 region) was used for these analyses to determine (1) the criteria for CXCR4 usage and whether it is the same as has been reported in subtype B, (2) the applicability of prediction methods, as well as development of a subtype C specific PSSM, (3) changes in the gp160 region including those regions outside the V3 in patients, some of whom have acquired the use of CXCR4 over time, (4) the evolutionary pathways in coreceptor switching from R5 to X4 within a single patient, and (5) the effect of TB and TB treatment on HIV-1 genetic diversity, particularly the influence of this selection pressure on viral populations and coreceptor usage.

CHAPTER 2

WHAT GENETIC CHANGES ARE ASSOCIATED WITH CXCR4 USAGE IN HIV-1 SUBTYPE C ISOLATES?

2.1 INTRODUCTION

HIV-1 enters the host cell by binding to CD4 and a coreceptor. The two major coreceptors involved are CCR5 and CXCR4 (Deng, Liu *et al*. 1996; Dragic, Litwin *et al*. 1996; Feng, Broder *et al*. 1996), with CCR5 being the most commonly used during transmission and early infection (Michael, Chang *et al*. 1997). HIV-1 isolates can be differentiated based on their ability to use these coreceptors with R5 viruses utilising CCR5, X4 viruses using CXCR4 and R5X4 viruses able to use both receptors (Berger, Doms *et al*. 1998). In subtype C R5 viruses dominate at all stages of disease, including late stage AIDS. Limited numbers of CXCR4-using viruses have been described (Abebe, Demissie *et al*. 1999; Bjorndal, Sonnerborg *et al*. 1999; Ping, Nelson *et al*. 1999; Batra, Tien *et al*. 2000; Cilliers, Nhlapo *et al*. 2003; Johnston, Zijenah *et al*. 2003), suggesting that there may be obstacles that prevent or limit the development of X4 viruses in this subtype. Defining the molecular factors that contribute to CXCR4 usage within this subtype will assist in identifying viruses with these phenotypes, as well as understand the role CXCR4 usage plays in disease progression of subtype C HIV-1.

The third variable loop (V3) of gp120 is generally 35 amino acids long, highly variable and a critical determinant of coreceptor usage (Hartley, Klasse *et al*. 2005). Within the V3 region the variation of specific amino acids (such as positions 11 and 25), increased positive amino acid charge and glycosylation are some of the characteristics that have been identified as influencing coreceptor usage in subtype B. Whether these characteristics are also important in subtype C coreceptor usage remain to be evaluated.

The V3-specific heteroduplex tracking assay (V3-HTA) has been used to rapidly identify V3 variants that are frequently associated with CXCR4 usage in subtype B (Nelson, Fiscus *et al*. 1997). This assay is based on the formation of heteroduplexes between the isolate V3 sequence and a consensus V3 sequence (probe). Divergence from the consensus sequence is then measured by the mobility ratio of the heteroduplexes formed and represents the variation within the V3 region, with CXCR4-using viruses usually the most divergent (Nelson, Fiscus *et al*. 1997; Nelson, Baribaud *et al*. 2000). A subtype C-specific V3-HTA has previously been applied to subtype C samples, but no CXCR4-using viruses were identified due to the low frequency of X4 viruses in this subtype (Ping, Nelson *et al*. 1999). In this study we selected 32 subtype C isolates with experimentally determined coreceptor usage profiles (16 R5 and 16 R5X4 or X4) and used the V3-HTA to screen for sequence diversity and population complexity of these biological variants. Furthermore, the sequences of these isolates were studied to determine if there were any distinct changes in the V3 region that might be used to predict CXCR4 usage in subtype C.

2.2 MATERIALS AND METHODS

2.2.1 Isolation and coreceptor usage of HIV-1 subtype C viruses

Viral isolates were selected from previously described cohorts from our laboratory. This included isolates from adult patients with advanced HIV-1 disease [SW, CM, PCP] (Cilliers, Nhlapo *et al*. 2003), sex workers with acute HIV-1 infection [Du] (Williamson, Morris *et al*. 2003) and one patient failing response to anti-retroviral treatment (Cilliers, Patience *et al*. 2004). Some isolates originated from a paediatric cohort of slowly progressing (TM) and rapidly progressing children (RP) (Choge, Cilliers *et al*. 2005) (for ethical clearance see Appendix A). Levels of virus in plasma were measured using the Versant HIV-1 RNA 3.0 assay (bDNA from Bayer Nucleic Acid Diagnostics) and CD4 counts were determined using a FACS count (Becton Dickinson, San Jose, CA). Viral isolates were tested for their ability to replicate in U87.CD4 cells transfected with either CCR5 or CXCR4, as previously described (Morris, Cilliers *et al*. 2001; Cilliers, Nhlapo *et al*. 2003).

2.2.2. Viral RNA isolation and RT-PCR

Viral RNA of each isolate was extracted from PBMC culture supernatant using a MagnaPure LC Isolation station and the Total Nucleic Acid isolation kit (Roche Applied Science, Penzberg, Germany). RT-PCR was performed with primers C+V3 (5'- ATA GTA CAT CTT AAT CAA TCT GTA GAA ATT -3') and C-V3 (5'- CCA TTT ATC TTT ACT AAT GTT ACA ATG TGC -3'), generating a 159 bp product as described (Nelson, Fiscus *et al*. 1997; Ping, Nelson *et al*. 1999). PCR products were purified using the High Pure PCR Product Purification kit (Roche Diagnostics GmbH, Mannheim, Germany).

2.2.3 V3-HTA

V3-HTA probe construction and labelling was done as previously described by Nelson *et al*. (Nelson, Fiscus *et al*. 1997) and Ping *et al*. (Ping, Nelson *et al*. 1999). The probe from the plasmid (D516-11) originating from a subtype C R5 virus and with only three nucleotide differences from the subtype C V3 consensus was used (Ping, Nelson *et al*. 1999). Single stranded probe labelling was done by digesting plasmid D516-11 with *Bam*H1 (Amersham Pharmacia Biotech, UK), end-labelling at room temperature with a mixture containing 12.5µCi ^{35}S-dATP (Amersham Pharmacia Biotech, UK), unlabeled dGTP and Klenow DNA polymerase I (Amersham Pharmacia Biotech, UK). The probe was removed from the vector by digestion with *Spe*I (Amersham Pharmacia Biotech, UK) and purified using the High Pure PCR purification kit (Roche Diagnostics GmbH, Mannheim, Germany) into a final volume of 50 µl. Heteroduplexes were formed between the probe and PCR product in a 10 µl reaction containing 5 µl PCR product, 3 µl labelled probe, 1 µl annealing buffer (1M NaCl, 100mM Tris-HCL [pH7.5], 20mM EDTA) and 1 µM of the C+V3 primer denatured at 95°C for 2 minutes. The reactions were then cooled at room temperature for 10 minutes. The heteroduplexes were separated in non-denaturing 12% polyacrylamide gels as described by Nelson *et al*. (Nelson, Fiscus *et al*. 1997). The dried gels were exposed to autoradiograms (BioMax MR, Kodak). Heteroduplex mobility ratio was determined by measuring the mobility of the slowest heteroduplex (highest band in a gel) of each sample and dividing it by the mobility of the probe homoduplex. The PCR product from samples with single bands was sequenced using an ABI PRISM 3100 genetic analyser with ABI PRISM BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems).

2.2.4 Cloning of single populations

Isolates with multiple variants were selected for cloning. Purified PCR product was cloned into the pGEMTeasy vector (Promega, USA) and individual molecular clones were screened by HTA and sequenced as described above.

2.2.5 Sequence analysis

All sequences were aligned with ClustalX, predicted protein translations were performed using BioEdit (version 5.0.9), phylogenetic analysis and genetic distances were determined using the MEGA program (version 2.1). The consensus sequences for isolates that used CCR5 and CXCR4 were determined using BioEdit. Additional subtype B and C V3 sequences were downloaded from the Los Alamos database (http://www.hiv.lanl.gov). The subtype B data set contained 129 sequences with known coreceptor usage (99 CCR5 and 30 CXCR4) representing only one sequence per patient. The subtype C data set represented single patient sequences from 91 CCR5 viruses and 17 SI viruses, since limited numbers of subtype C CXCR4 viruses were available.

2.3 RESULTS

2.3.1 Investigation of sample complexity using a V3-HTA

To investigate the variants present within the V3 region of CCR5- and CXCR4-using viruses a V3-HTA was used. Sixteen R5 isolates, eight R5X4 and eight X4 isolates from patients at various stages of disease, including late stage AIDS, were selected for analysis (Table 2.1). Two isolates were obtained from a single patient (Du179) that were one year apart and were included as they showed different coreceptor profiles. The V3 region from all isolates was amplified and hybridised to a subtype C R5 radiolabelled probe and resolved on a polyacrylamide gel. In general, the heteroduplexes that formed between the probe and the R5 viruses migrated faster through the gel, close to the probe homoduplex, while the R5X4 and X4 viruses migrated more slowly through the gel at variable distances and often closer to the single stranded probe (Figure 2.1). At the extreme, one R5X4 (SW30) and one X4 (TM2) had heteroduplexes that were above the single-stranded probe suggestive of a high degree of genetic difference and/or an insertion or deletion relative to the probe. Apart from one R5 isolate (SW4) and one X4 isolate (SW7) all the R5 and X4 isolates had single bands indicative of homogeneous populations within the V3 region. In contrast, 5 of the 8 R5X4 isolates had multiple variants often migrating at different distances between the single strand and homoduplex probe, indicative of heterogeneous mixtures.

Table 2.1: Clinical information of the 16 R5, 8 R5X4 and 8 X4 HIV-1 subtype C isolates used in this study.

R5 isolates					R5X4/X4 isolates				
Isolate	Clinical status	CD4 count (cells/µl)	Viral load (copies/ml)	Biotype	Isolate	Clinical status	CD4 count (cells/µl)	Viral load (copies/ml)	Biotype
Du151(11-98)	Acute	367	>500 000	R5	CM9	AIDS	24	NA	R5X4
PCP1	AIDS	2	NA	R5	Du179(05-99)	Chronic	279	2 640	R5X4
CM1	AIDS	43	146 514	R5	Du36	Acute	25	54 944	R5X4
CM4	AIDS	47	163 755	R5	RP1	Rapid progressor	7	178 830	R5X4
SW2	AIDS	84	157 150	R5	SW20	AIDS	2	43 595	R5X4
SW3	AIDS	53	261 880	R5	SW30	AIDS	2	73 860	R5X4
SW4	AIDS	76	1 496 620	R5	TM1	Slow progressor	NA	190 000	R5X4
SW5	AIDS	40	1 374 235	R5	TM18b	Slow progressor	202	500 000	R5X4
SW8	AIDS	67	1 198 880	R5	DR28	AIDS, on treatment	173	269 000	X4
SW9	AIDS	65	301 605	R5	Du179(05-00)	Chronic	231	2 228	X4
TM3	Slow progressor	329	11 178	R5	Du55	Acute	13	6 589	X4
TM4	Slow progressor	692	25 815	R5	SW12	AIDS	27	68 410	X4
TM5	Slow progressor	1378	22 488	R5	SW7	AIDS	10	NA	X4
TM6	Slow progressor	846	108 716	R5	TM2	Slow progressor	NA	25 156	X4
TM10	Slow progressor	1158	685	R5	TM46b	Slow progressor	4	28 613	X4
TM12	Slow progressor	976	21 976	R5	TM9	Slow progressor	11	296 865	X4

NA: Not available
Slow progressor - child diagnosed with HIV-1 >5 years

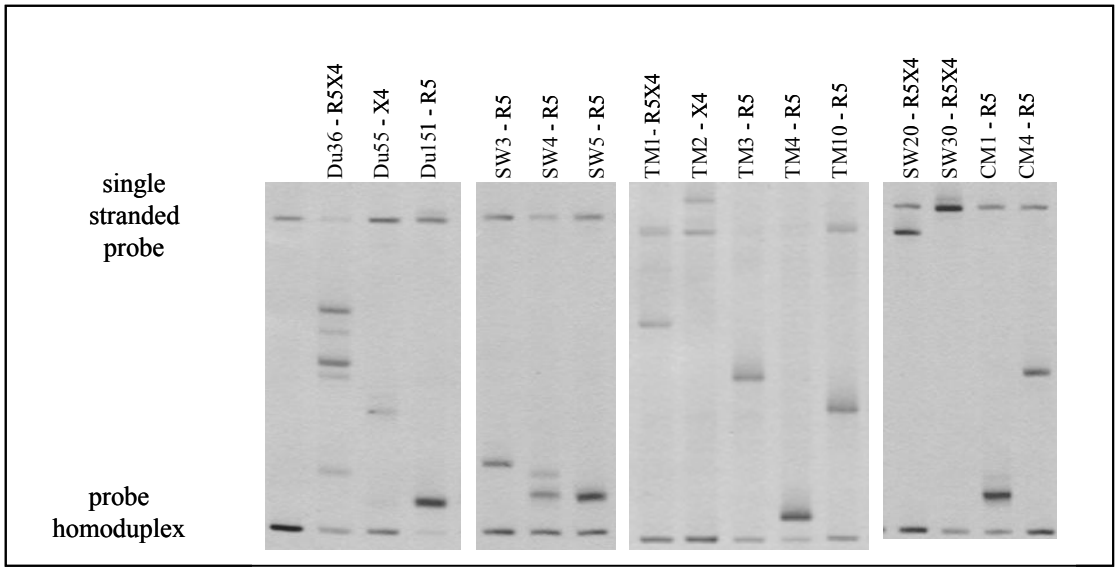


Figure 2.1: V3-HTA of HIV-1 subtype C isolates with different biological phenotypes. PCR products from viral isolates were hybridised to a radiolabelled subtype C R5 probe and separated on a polyacrylamide gel. Heteroduplexes formed between the isolate and probe usually migrated between the single stranded probe and probe homoduplex as shown in four separate gels.

A heteroduplex mobility ratio (k) was calculated for each isolate based on the rate of migration of heteroduplex bands. In isolates with multiple bands, the mobility ratio for each variant was calculated, although for the purposes of analysis the lowest k value was used. The majority of R5 isolates clustered closer together with a median mobility ratio of 0.94 (range 0.76 - 0.96) (Figure 2.2A). The three R5 isolates (TM3, CM4 and TM10) with low mobility ratios ($k<0.90$) had a single deletion within the V3 region at amino acid positions 23, 24 and 25 respectively compared to the probe sequence, thus resulting in retarded migration through the gel. The mobility ratios of R5X4 and X4 isolates showed a broader range (0.52 - 0.91) with a median of 0.64. There was no significant difference in k values between R5X4 and X4 isolates, although almost all fell below 0.90. The one isolate with a mobility ratio above 0.90 (DR28) was from a drug treated patient and had 35 amino acids similar to the probe. A graphical display of

these data relative to coreceptor usage is shown in Figure 2.2B. A dashed line at 0.90 separated most R5 isolates (samples above the line) from the R5X4 and X4 ($p<0.001$).

A

R5 isolates				R5X4/X4 isolates			
Isolate	Mobility ratios	# amino acids	Charge	Isolate	Mobility ratios	# amino acids	Charge
Du151(11-98)	0.96	35	3.5	CM9	0.85/0.9*	35	5.5
PCP1	0.96	35	3.5	Du179(05-99)	0.78	34	4.5
CM1	0.94	35	2.5	Du36	0.68/0.71/0.76/0.77/0.91*	35, 36	6.5, 7.5
CM4	0.8	**34**	2.5	RP1	0.62/0.84*	35, 37	4.5, 7.5
SW2	0.96	35	3.5	SW20	0.58	37	5.5
SW3	0.9	35	3.5	SW30	0.54/0.93*	35,37	7.5, 8
SW4	0.92/0.94	35	4*	TM1	0.61/0.68	37	7.5
SW5	0.95	35	3.5	TM18b	0.88	35	6
SW8	0.93	35	3.5	DR28	0.91	35	7.5
SW9	0.94	35	3.5	Du179(05-00)	0.61	32	5
TM3	0.76	**34**	3	Du55	0.83	34	6
TM4	0.96	35	4.5	SW12	0.61	37	7.5
TM5	0.91	35	4	SW7	0.65/0.78	36	7.5
TM6	0.95	35	3.5	TM2	0.52	37	5.5
TM10	0.81	**34**	4.5	TM46b	0.72	34	6
TM12	0.95	35	2	TM9	0.54	37	7
Median	0.94	35	3.5	Median	0.64	35	6

Number
* see Figure 3 for further analysis of multiple bands

B

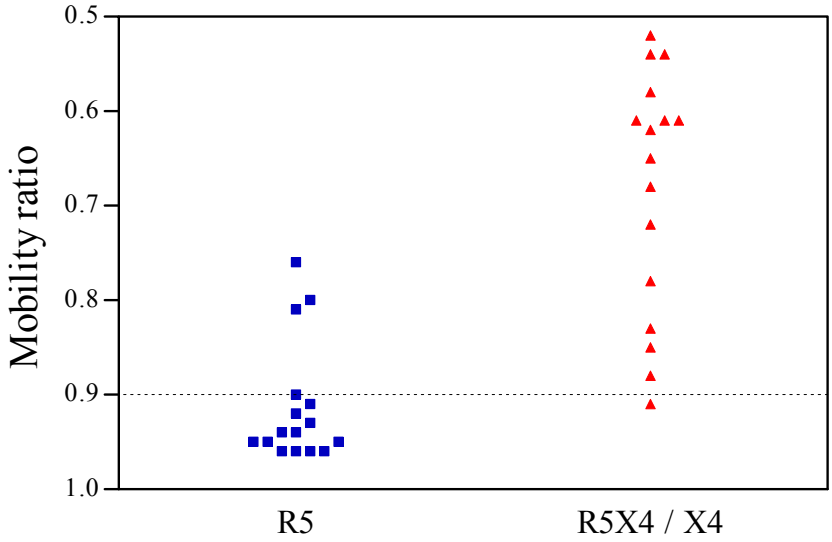


Figure 2.2: Comparison of mobility ratio and coreceptor usage of subtype C isolates. **(A)** V3-HTA mobility ratio, number of amino acids and charge in V3 of the variants present within the R5, R5X4 and X4 isolates. In isolates with multiple bands, all the mobility ratios are listed. **(B)** Association between mobility ratio (k) and biotype. Only the lowest mobility ratio was used for samples with multiple bands. The dashed line at 0.90 separated most R5 from R5X4/X4 viruses.

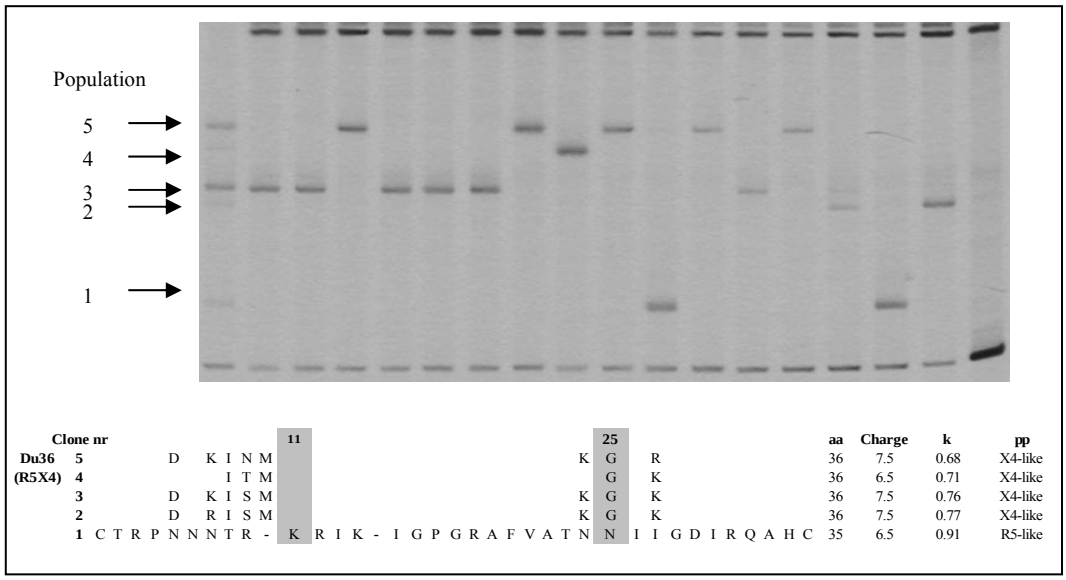
2.3.2 Analysis of isolates with multiple variants

Seven isolates in this study had multiple variants in the V3 region that were clearly distinguishable on a V3-HTA. This included one R5 isolate (SW4), 5 R5X4 isolates (CM9, Du36, RP1, SW30, TM1) and one X4 isolate (SW7). In order to analyse the different variants, molecular clones were selected from five of these isolates (TM1 and SW7 were not cloned). The V3 nucleotide (not shown) and amino acid alignment for each of the variants were then compared. The phenotype of individual clones were predicted based the V3 amino acid charge, sequence analysis and mobility ratio (where $k>0.90$ was considered R5-like and a $k<0.90$ considered to be X4-like).

A total of five variants were present in the dualtropic isolate Du36 (Figure 2.3A). These variants were present in different proportions with variants 3 and 5 being the most common. The amino acid charge for all 5 variants was between 6.5 and 7.5, indicative of CXCR4 usage, however, the mobility ratio of variant 1 was 0.91 suggesting it was more R5-like. A lack of a predicted N-glycosylation site at the N-terminus of the V3 region was noted in 3 of the X4-like variants (2, 3 and 5). Collectively the X4-like variants were genetically more similar to each other than to the R5-like variant (data not shown). This patient was previously shown to be dually infected with two subtype C viruses, which might explain the high degree of genetic divergence between the different populations in this sample (Grobler, Gray *et al*. 2004).

Sequence analysis of clones from 1 R5 and 3 R5X4 isolates with multiple bands were also investigated (Figure 2.3B). The amino acid sequences from both clones of SW4 were identical with high mobility ratios consistent with an R5 phenotype and this was

A



Clone nr		11	25	aa	Charge	k	pp
Du36	5	D K I N M	K G R	36	7.5	0.68	X4-like
(R5X4)	4	I T M	G K	36	6.5	0.71	X4-like
	3	D K I S M	K G K	36	7.5	0.76	X4-like
	2	D R I S M	K G K	36	7.5	0.77	X4-like
	1	C T R P N N N T R - K R I K - I G P G R A F V A T N N I I G D I R Q A H C		35	6.5	0.91	R5-like

B

		11	25	aa	Charge	k	pp
SW4	2	* *		35	4	0.92	R5-like
(R5)	1	C T R P S N N T R K S I R - - V G P G Q S F H A T G E I I G D I R Q A H C		35	4	0.94	R5-like
CM9	2	I R T I		35	5.5	0.85	X4-like
(R5X4)	1	C T R P G N N T R K R I R - - I G P G Y A F Y A K E A I V G D I R Q A H C		35	5.5	0.9	R5-like
SW30	2	N K I G R H K V		37	8	0.54	X4-like
(R5X4)	1	C T R P K N N T R R S V R - - I G P G Q A F Y A T G R S I G N I R Q A H C		35	7.5	0.93	R5-like
RP1	2	R L G	R R	37	7.5	0.62	X4-like
(R5X4)	1	C T R P G N N T R K S V R - - I G P G Q T F Y A T G Q V I G D I R Q A H C		35	4.5	0.84	R5-like

(*) Indicates the synonymous changes responsible for the variants in SW4
(-) Indicates amino acids not present
(aa) Number of amino acids

Figure 2.3: Analysis of molecular clones from HIV-1 subtype C isolates with multiple variants. The variants were labelled from the bottom (highest mobility ratio) to the top of the gel. (**A)** Amino acid alignment of 5 populations present in Du36, as seen in accompanying gel. Amino acid alignments indicating only the differences within these populations, number of amino acids and charge of the V3 region are shown. Predicted phenotype according to charge, sequence analysis and mobility ratio is also indicated. (-) Indicates amino acids not present, (aa) number of amino acids in the V3 region and positions 11 and 25 are highlighted. (**B)** Amino acid alignments of variants present in 1 R5 and 3 R5X4 isolates.

supported by the charge and length, which were typical of R5 viruses. Thus, the differences between these 2 variants of SW4 visible on the V3-HTA were due to synonymous nucleotide changes. Among the 3 R5X4 isolates amino acid differences were seen between the 2 clones of each isolate, suggestive of a mixture of R5-like and X4-like variants. Predicted CXCR4 usage was associated with different genetic characteristics in each case although some commonalities in the positions were noted. In one patient there were positively charged amino acids at positions 11 and 25 in the X4-like variant only. Changes in the crown motif were seen in two and insertions were observed in all X4-like variants from the 3 patients. There were also differences in the k ratios between the 2 clones from the same patient, with the X4-like variant having lower k values compared to the R5-like variant.

2.3.3 V3 sequence variability of CCR5- and CXCR4-using variants

Two data sets were compiled representing CCR5- and CXCR4-using sequences. The CCR5-using data set was obtained from the 16 patients with R5 viruses that had single populations as seen in the V3-HTA (SW4 had 2 populations and the sequence with the highest mobility ratio was used). The CXCR4-using data set contained sequences from the 6 patients with X4 viruses and 3 dualtropic patients (that had homogenous populations), as well as two population-based sequences (SW7 and TM1). To increase the number of sequences in the CXCR4 data set, single sequences from the five dualtropic patients, with multiple variants were also included. The sequences with the highest mobility ratios were selected as representative of the CXCR4-using variant within a patient. These V3 sequences from the 16 CCR5- and 16 CXCR4-using variants were compared for length and charge. The majority of CCR5-using sequences had 34-35 amino acids with no insertions, whereas CXCR4-using sequences consisted

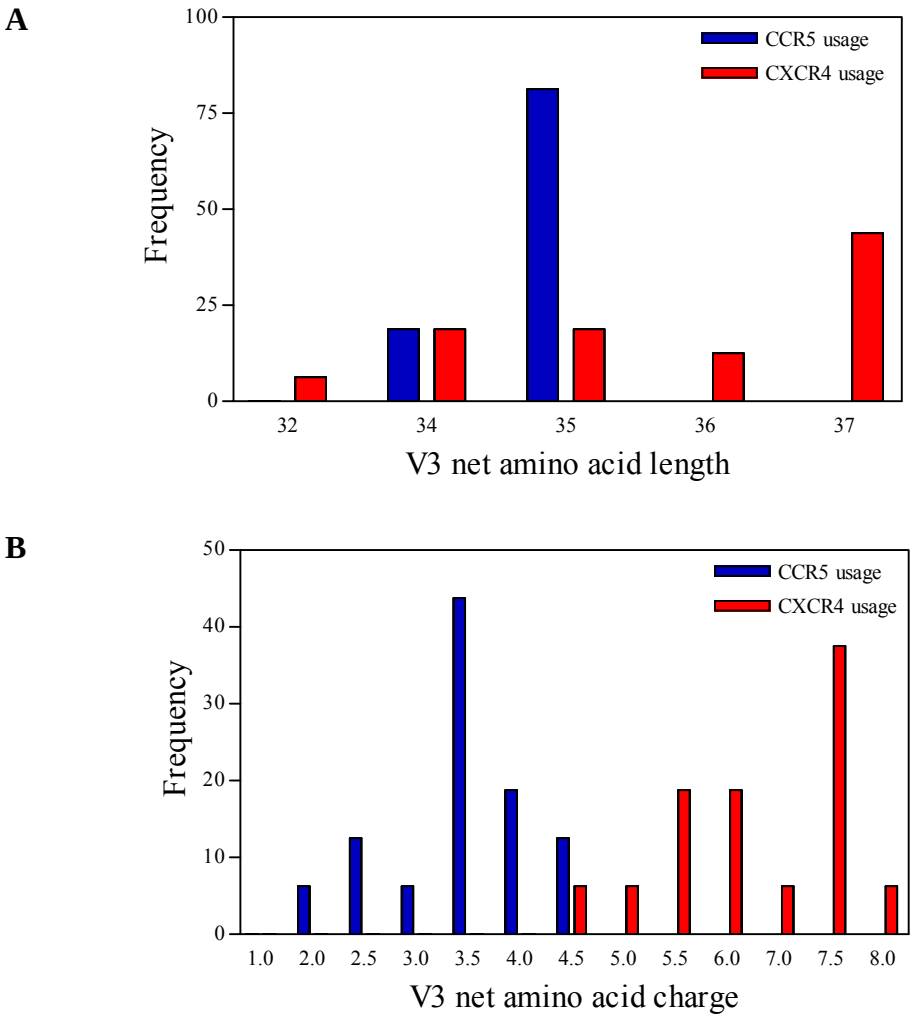


Figure 2.4: Comparison of (**A**) V3 length and (**B**) V3 charge of HIV-1 subtype C isolates able to use CCR5 and CXCR4.

of 32-37 amino acids due to insertions and deletions (Figure 2.4A). The V3 net amino acid charge between CCR5 and CXCR4 usage were distinct with little overlap (Figure 2.4B). The CCR5-using viruses had an amino acid charge between +2 and +4.5, with the majority of samples having +3.5. CXCR4-using variants ranged between +4.5 and +8 with the highest frequency of variants having a charge of +7.5. Thus, CXCR4 usage in subtype C was associated with an increased V3 length and increased number of positively charged amino acids.

Entropy plots were performed to compare the amount of sequence variability between the CCR5- and CXCR4-using variants (Figure 2.5). Entropy plots graphically measure the amount of variability at a specific site of an alignment, with higher entropy indicative of more variation at a specific site. In general, the CXCR4 variants were more variable across the V3 region at each specific site, with high variation at positions 11, 12, 24 and 25. The predicted N-glycosylation site (at positions 6-8) associated with CCR5 usage was conserved in all 16 R5 isolates but highly variable in the CXCR4-using samples. Similarly, the crown motif within the V3 for all R5 isolates was GPGQ, compared to CXCR4-using variants that showed variation in this motif (Fischer exact test, p<0,0001).

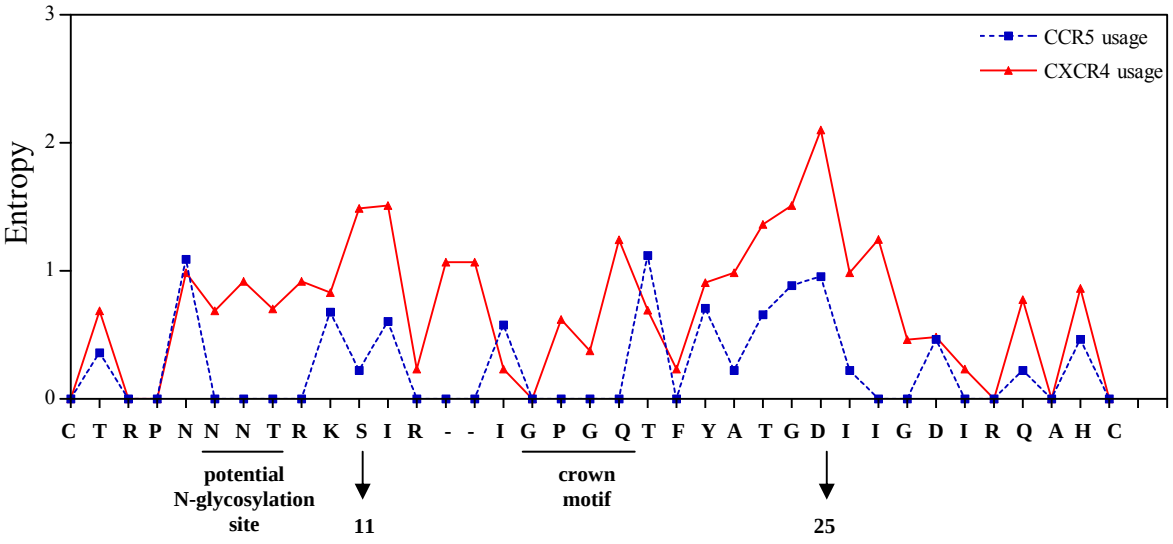


Figure 2.5: Variation in the V3 region of CCR5 and CXCR4-using isolates of HIV-1 subtype C. **(A)** Entropy plot representing the variation at each amino acid site for CCR5 usage and CXCR4 usage. The potential N-glycosylation site N-[X]-T, amino acid positions 11 and 25, and crown motif are indicated.

The frequency and type of amino acid substitutions at each site was determined for the CCR5- and CXCR4-using isolates (Figure 2.6). Although these results represent small numbers (16 CCR5- and 16 CXCR4-using isolates), CCR5 usage was associated with less amino acid substitutions compared to CXCR4 usage. In particular, a variety of amino acids were seen at positions 11 and 25 within CXCR4-using viruses, and these were not necessarily positively charged. Nine of the 16 CXCR4-using viruses had insertions between amino acid positions 13 and 14. These insertions were either T/K/L or more commonly I at the first position and if a second insertion followed, it was usually glycine (G). Twelve of the 16 CXCR4-using viruses (75%) had changes in the crown motif (GPGQ). These changes were mostly at position 18, but amino acid substitutions were also seen at positions 16 and 17.

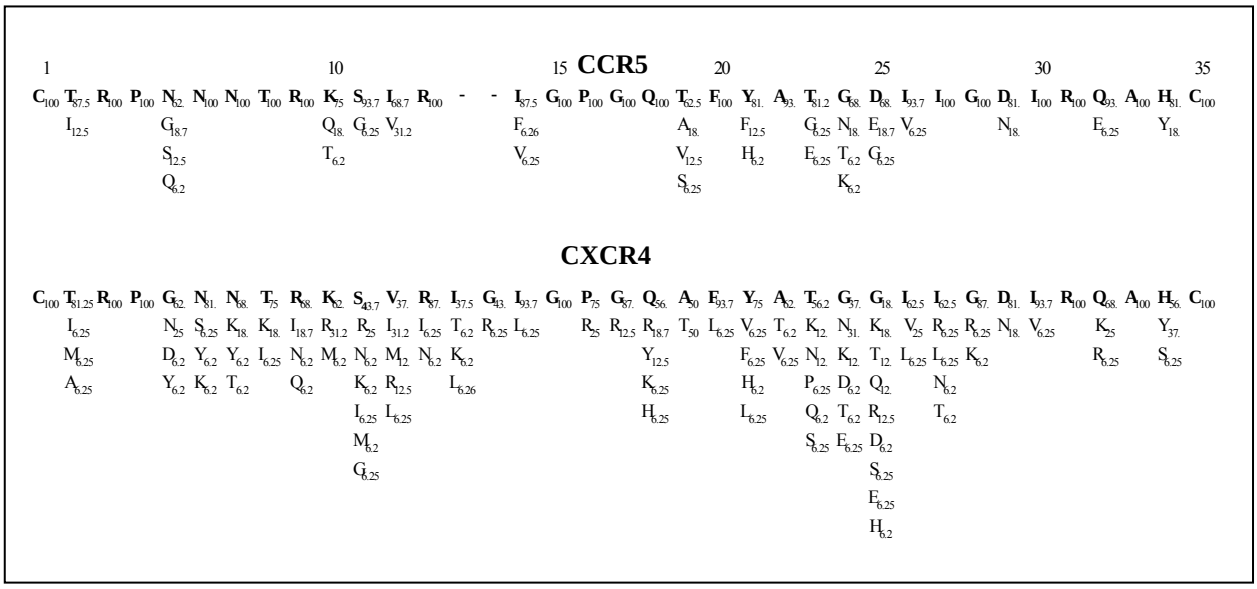


Figure 2.6: Distribution of amino acids in the V3 region of 16 CCR5- and 16 CXCR4-using variants (numbers indicate frequency in total of 16 viruses).

2.3.4 Determining subtype C V3 characteristics associated with CXCR4 usage

Consensus sequences were compiled for these 16 CCR5 and 16 CXCR4 representative sequences investigated in this study. There was sequence homology for 26 of the 35 amino acids between the CCR5- and CXCR4-using isolates (Figure 2.7). Amino acid substitutions associated with the CXCR4-using data set were seen at position 5 with N being replaced by G and position 19 associated with either A or T. The consensus sequence of CXCR4-using viruses had a 2 amino acid insertion between positions 13 and 14, as well as a further seven variable (X) amino acid positions (11, 12, 18, 23, 24, 25 and 34) compared to the CCR5 consensus. The crown motif of the CXCR4 consensus changed from GPGQ to GPGX, where X was usually a charged amino acid such as R, Y, K and H.

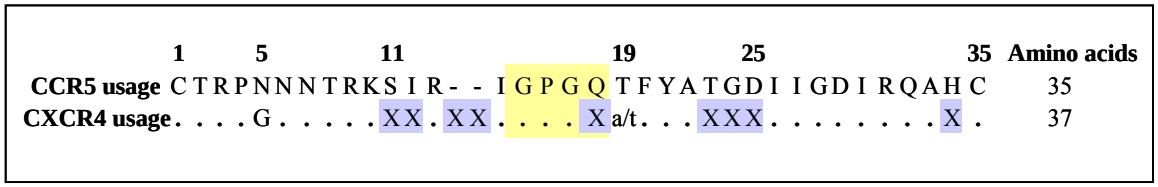


Figure 2.7: Consensus V3 sequences from CCR5- and CXCR4-using HIV-1 subtype C isolates. Amino acids within the CXCR4-using consensus that differ from the CCR5 consensus are highlighted and the crown motif is boxed. Variable (X) and identical (.) amino acids are indicated.

Further analysis of the crown revealed additional genetic differences between CCR5 and CXCR4-using isolates that were of particular interest as such differences have not previously been fully explored (Figure 2.8A). In order to extend this analysis we combined this data with previously published subtype C sequences from the Los Alamos database. These included 91 single patient sequences with determined CCR5 coreceptor usage and 17 sequences of which the majority only had an SI phenotype determined and were assumed to be CXCR4-using. Using this larger data-set (n=107)

the GPGQ consensus remained highly conserved among the CCR5-using viruses (Figure 2.8B). Similarly, we noted a significant correlation between CXCR4 usage and changes in the crown motif in 33 HIV-1 subtype C isolates ($p<0.0001$, Fisher's exact test).

Given that CXCR4-usage is more common in subtype B and that sufficient data is available we performed a comparison with HIV-1 subtype B sequences from isolates with known coreceptor usage (obtained from Los Alamos database). The consensus for both CCR5 and CXCR4-using subtype B isolates was GPGR. Data showed that while there was some variation at positions 16 (P) and 18 (R), this did not differ significantly between isolates that used CCR5 and those that used CXCR4 (Figure 2.8C). This is in contrast to HIV-1 subtype C CXCR4-using isolates which showed marked variation particularly at position 18 from the subtype C GPGQ consensus.

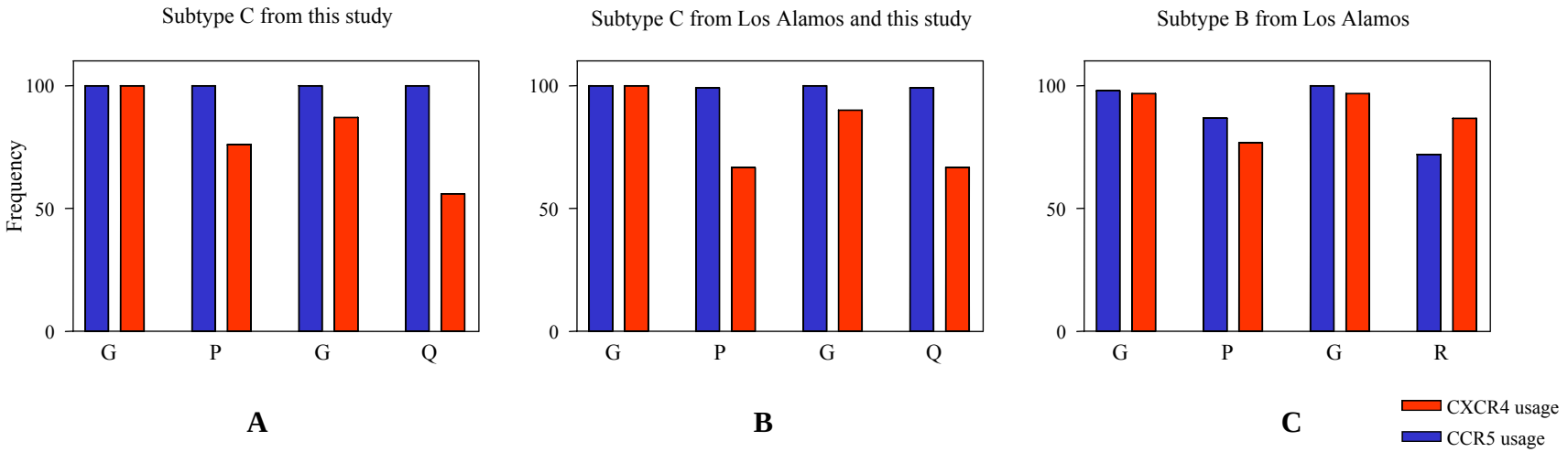


Figure 2.8: Comparison of amino acid variation in the V3 crown of subtype C isolates from this study (16 CCR5 and 16 CXCR4) (p>0.0001) (**A**); subtype C isolates from Los Alamos and this study (106 CCR5 and 33 SI) (p<0,0001) (**B**) and subtype B isolates from Los Alamos (99 CCR5 and 30 CXCR4) (**C**).

2.4 DISCUSSION

A subtype C specific V3-HTA was used to examine 32 subtype C isolates with known biological phenotypes (16 R5 and 16 R5X4 or X4 isolates). Results indicated that there were sufficient genetic differences to discriminate between R5 viruses and those using CXCR4 (both R5X4 and X4). Sequence analysis of the V3 region showed that CXCR4-using viruses were often associated with an increased number of positively charged amino acids and an increased length due to amino acid insertions. Compared to HIV-1 subtype B V3 sequences, where the consensus sequence at the GPGR crown did not differ between CCR5 and CXCR4-using isolates, the GPGQ subtype C consensus was heavily substituted in CXCR4-using viruses.

The V3-HTA has proven to be a rapid genotype-based method to detect V3 evolutionary variants of HIV-1 subtype B and C viruses (Nelson, Fiscus *et al*. 1997; Ping, Nelson *et al*. 1999; Nelson, Baribaud *et al*. 2000). This assay measures distinct genetic features such as insertions, deletions or clustered amino acid changes that influence the mobility of the V3 heteroduplex. These characteristics are frequently associated with the X4-like phenotype, and using this assay it has been possible to screen for X4 variants (Nelson, Fiscus *et al*. 1997). In a previous study of subtype C (Ping, Nelson *et al*. 1999) no X4-like subtype C isolates were identified using a V3-HTA. This is because this study did not focus on subjects with low CD4 cell counts where X4 variants are more likely to be found even in subtype C where such variants are rare. Here we selected subtype C viruses with experimentally determined phenotypes including a large collection of X4 variants. Using a subtype C V3-HTA most R5 isolates had homogeneous populations with mobility ratios above 0.90. This

included R5 isolates from patients with advanced disease where more heterogeneous sequences might be expected (McNearney, Hornickova *et al*. 1992). The highly conserved nature of subtype C V3 R5 isolates was also noted in the study by Ping and others (1999). The R5X4 and X4 isolates had a broader range of mobility ratios that were generally lower than 0.90. The X4 isolates had homogenous populations in the V3-HTA, whereas the R5X4 isolates were associated with multiple variants as evidenced by multiple bands. Further analysis of these dualtropic isolates indicated that they usually comprised of mixtures of CCR5- and CXCR4-like viruses. Thus, we confirmed that the V3-HTA assay could be used to identify genetic variants associated with R5 and X4-like variants in subtype C.

A limiting factor within this study was the biased selection of samples that may have influenced the sensitivity and specificity of this assay. The ratio of R5 and X4 variants selected in this study is not reflective of prevalence of these viruses within the general populations of HIV-1 subtype C. Among 231 HIV-1 subtype C viral isolates in our laboratory, 10% were found to be CXCR4-using and this was significantly correlated with a CD4 count of <200 cells/μl ($p = 0.0021$, Fisher's exact test, unpublished). Although V3-HTA was sensitive in detecting most of the slower migrating X4 variants, this sensitivity would decrease with larger data sets due to the low abundance of these variants, as well as the increase sampling of CCR5 variants with genotypic characteristics that cause slow migration (such as deletions and insertions). Similarly the frequency of false positives (i.e. slow migrating R5 variants) would decrease the specificity. Thus, since the V3-HTA is more reflective of sample complexity the mobility ratio criteria could vary if a larger data set was used. Despite these considerations these data suggest that the V3-HTA was an

applicable screening tool to evaluate sample complexity and could assist in cloning strategies to investigate the V3 differences between CCR5- and CXCR4-using viruses.

Comparisons between the representative CCR5 and CXCR4 data sets revealed differences in charge and length between CCR5- and CXCR4-using subtype C viruses. The V3 loops of CXCR4-using viruses were usually longer and more positively charged, previously shown to be associated with CXCR4 usage (De Jong, De Ronde *et al*. 1992). The increased length was usually due to one or two amino acid insertions between position 13 and 14, with amino acids I and G being the most common. Almost all CXCR4-using isolates had a high V3 charge above +4.5. This was due to the presence of increased numbers of K and R residues that were scattered throughout the V3 of CXCR4-using viruses, including positions 11 and 25 which are indicative of SI viruses in subtype B (Fouchier, Groenink *et al*. 1992). Although the 11/25 rule is used for tropism determination, it is not clear whether basic amino acid substitutions at these sites are sufficient or necessary for CXCR4 usage (Kuiken, de Jong *et al*. 1992; Nelson, Baribaud *et al*. 2000). In this study, these positions were not necessarily associated with positively charged amino acids in CXCR4-using viruses, although there was increased variation at these positions compared to R5 viruses. Nevertheless, similar to subtype B, CXCR4 usage in subtype C was rarely due to a single amino acid change but rather to changes in 3-5 amino acids that increased the length and charge of the V3 loop (Shioda, Levy *et al*. 1992).

All the R5 isolates in this study had a potential N-glycosylation site at positions 6-8 within the V3 region. Most early viruses, able to use CCR5, have this glycan,

suggesting that it is needed for CCR5 interaction (Polzer, Dittmar *et al*. 2002). As immune pressure decreases with disease, viruses lacking this glycan that are able to use CXCR4 have been shown to emerge (Polzer, Dittmar *et al*. 2002; Pollakis, Abebe *et al*. 2004). The loss of this glycan has also been shown to assist in more efficient use of CXCR4, and thus might be an important factor in the switching of R5X4 to X4 viruses (Polzer, Dittmar *et al*. 2002; Nabatov, Pollakis *et al*. 2004). Four of the 16 CXCR4-using viruses in this study lacked this potential glycosylation site suggesting that in subtype C this site may play a similar role. The highly conserved nature of this glycan in R5 subtype C viruses suggests it is crucial to CCR5 interaction possibly by masking surrounding positively charged amino acids at the N-terminus of the V3 region (Hartley, Klasse *et al*. 2005). Pollakis and others (2001) have speculated that the high frequency of V3 glycosylation within subtype C viruses might constrain the envelope structure promoting the use of CCR5 and thereby increasing its transmission efficiency.

Most subtype C isolates had a GPGQ crown motif, whereas the consensus subtype B generally contains a GPGR motif irrespective of coreceptor usage (Milich, Margolin *et al*. 1997). In this study, CCR5-using viruses had the expected GPGQ crown, whereas this changed to GPGX, with X being R, K, H or Y in CXCR4-using isolates. Although changes within the crown of subtype C viruses have been noted previously, the possible importance of this for coreceptor usage has not been highlighted due to the limited numbers of CXCR4-using viruses in this subtype C (Abebe, Demissie *et al*. 1999; Bjorndal, Sonnerborg *et al*. 1999; Ping, Nelson *et al*. 1999; Batra, Tien *et al*. 2000; Cilliers, Nhlapo *et al*. 2003; Johnston, Zijenah *et al*. 2003). The inclusion of additional subtype C sequences within this analysis resulted in no changes within the

consensus CCR5 usage sequence, suggesting that CCR5 sequences are homogenous despite disease status. There were very limited viruses with biologically determined CXCR4 coreceptor usage available with most of the sequences annotated as SI phenotype (which are suggestive of R5X4 and X4 viruses). Nevertheless the changes in the crown motif associated with CXCR4 usage were confirmed in the larger dataset. Secondary structure prediction suggests that the GPGX forms a beta turn and amino acid changes within this motif are critical determinants of coreceptor usage (Shimizu, Haraguchi *et al*. 1999; Hu, Trent *et al*. 2000; Cormier and Dragic 2002; Suphaphiphat, Thitithanyanont *et al*. 2003; Pollakis, Abebe *et al*. 2004; Hartley, Klasse *et al*. 2005). Position 18 in the crown of subtype B (GPGR) was found to be less variable in X4 viruses, suggesting a functional role for R in CXCR4 usage (Resch, Hoffman *et al*. 2001). In subtype C arginine (R) was the most frequent amino acid at position 18 in CXCR4-using isolates. Thus one route to CXCR4-usage may require GPGQ to first undergo a change at position 18 to arginine (R) increasing the charge and/or altering the conformation of V3, as proposed by Hartley (Hartley, Klasse *et al*. 2005). Previous studies have shown that the transition from a R5 to X4 requires few genetic changes (at least within the V3) although these transitional intermediates may be less fit, accounting for the low frequency of X4 viruses in subtype C (Pastore, Ramos *et al*. 2004). Conversely the presence of GPGR may predispose subtype B viruses to CXCR4 usage. Although the numbers in this study are limited, these data are suggestive of the crucial role of GPGQ in restricting HIV-1 subtype C viruses from using CXCR4. The fact that mutations occur in this region *in vitro* as well as *in vivo* suggests that it is not immune-mediated. Whether this is a cause or consequences of the predominant use of CCR5 by HIV-1 subtype C is unclear and can only be addressed with further studies.

In conclusion, changes within the V3 region such as increased amino acid charge, insertions, specific amino acid variation and loss of the potential glycosylation site, are all factors that play a role in the use of CXCR4 by subtype C viruses. The most noteworthy difference between CCR5- and CXCR4-using viruses observed in this study was changes within the crown motif. This suggests that increased virological adaptation within subtype C viruses, in particular within the crown, allows these viruses to acquire the ability to use CXCR4 as a coreceptor. The highly conserved crown motif in subtype C R5 viruses might also have a restrictive characteristic, limiting the development of CXCR4 using viruses.

CHAPTER 3

A RELIABLE PHENOTYPE PREDICTOR FOR HIV-1 SUBTYPE C BASED ON ENVELOPE V3 SEQUENCES

3.1 INTRODUCTION

Viruses establishing HIV-1 infection generally use CCR5 as a coreceptor for entry into host cells (Deng, Liu *et al*. 1996; Dragic, Litwin *et al*. 1996). In some individuals, viruses are transmitted or viral genetic changes arise over time that permit the virus to use other coreceptors, in particular CXCR4. The ability to screen large subtype C infected cohorts for CXCR4-using viruses is vital to better understand why these viruses are far less common, even in individuals with more advanced disease (Bjorndal, Sonnerborg *et al*. 1999; Ping, Nelson *et al*. 1999; Cilliers, Nhlapo *et al*. 2003; Coetzer, Cilliers *et al*. 2005). However, coreceptor phenotypic assays are expensive, labor intensive and require specialized laboratories not always available in the developing countries where subtype C predominates. A reliable phenotype prediction method, based on genetic sequence analysis, could provide for rapid and less expensive screening.

Distinct genetic differences within the V3 region, between CCR5- and CXCR4-using viruses have been described that influence coreceptor usage (Shioda, Levy *et al*. 1991; De Jong, De Ronde *et al*. 1992; Fouchier, Groenink *et al*. 1992). These differences have been used with varying degrees of success to predict tropism using bioinformatic approaches and were recently reviewed by Jensen and van 't Wout (2003) using subtype B data sets. They include the 11/25 rule to distinguish between NSI and SI-like viruses (Fouchier, Groenink *et al*. 1992), a multiple regression method based on positive, negative and net V3 charge (Briggs, Tuttle *et al*. 2000), a neural network strategy (Resch, Hoffman *et al*. 2001), a machine-learning method (Pillai, Good *et al*. 2003), and a subtype B position specific scoring matrix (B-PSSM) (Jensen, Li *et al*.

2003). Another program that has recently been developed is Geno2Pheno (Sing, Beerenwinkel *et al*. 2004), although the PSSM showed improved predictive power over all these methods, and was also useful in the analysis of the transition between R5 to X4 in subtype B.

In this chapter, we tested 4 of these predictors on a subtype C data set of V3 sequences with known phenotypes to determine the applicability of these methods to subtype C sequences. The poor performance of these methods in predicting SI/CXCR4 usage suggested that a predictor based on subtype C sequences was necessary. Given that the B-PSSM was shown to have improved positive predictive value (Jensen and van 't Wout 2003), we therefore developed such a predictor using V3 sequences from subtype C isolates of known phenotype. Predictions based on the C-PSSM, exhibited increased reliability and sensitivity over subtype B-based predictors.

3.2 MATERIALS AND METHODS

3.2.1 Data set compilation

A training set of 280 HIV-1 subtype C V3 sequences was compiled from the Los Alamos HIV database (http://hiv-web.lanl.gov) and from our own laboratory [(Cilliers, Nhlapo *et al*. 2003; Choge, Cilliers *et al*. 2005; Coetzer, Cilliers *et al*. 2005) and unpublished data]. Only sequences for which the corresponding biological phenotypes were determined on coreceptor-transfected cell lines (for CCR5 and/or CXCR4 usage) or MT-2 cells (for NSI/SI phenotypes) were used in this study. Where only coreceptor usage data was available, R5 viruses were assumed to be NSI while R5X4 and X4 viruses were assumed to be SI. In some cases only MT-2 data was available therefore, the training set was divided into NSI (R5) and SI (R5X4 and X4) viruses. There were 229 NSI V3 sequences (from 200 subjects) and 51 SI V3 sequences (from 20 subjects).

3.2.2 Performance of four genotypic algorithms on subtype C sequences

A subset of the data set representing only the unique sequences from 220 subjects (200 with NSI and 20 with SI viruses) was submitted to the following four prediction methods: the 11/25 rule (Fouchier, Groenink *et al*. 1992); a multiple regression method referred to as Briggs' method (Briggs, Tuttle *et al*. 2000); a machine-learning method referred to as the Pillai method (Pillai, Good *et al*. 2003), and the B-PSSM (Jensen, Li *et al*. 2003) as discussed in Chapter 1. The percentage of sequences with correctly predicted phenotype was calculated for each algorithm. To compare the effectiveness of these methods on subtype C sequences with each other and to the C-

PSSM, a correct overall prediction value (percentage specificity X percentage sensitivity) was determined.

3.2.3 How is a PSSM matrix constructed?

PSSM detects non-random distributions of amino acids at a specific site within a group of aligned sequences that has a desired property, such as NSI/R5 or SI/X4/R5X4 phenotype (Gribskov, McLachlan *et al*. 1987; Henikoff, Wallace *et al*. 1990; Jensen, Li *et al*. 2003). A matrix or profile is compiled from a data set of sequences by determining the likelihood ratio of a specific amino acid at a specific site (Figure 3.1). This site-specific score reflects the difference in abundance of a particular amino acid at a specific site in a group of SI sequences, compared to that same site in a group of NSI sequences.

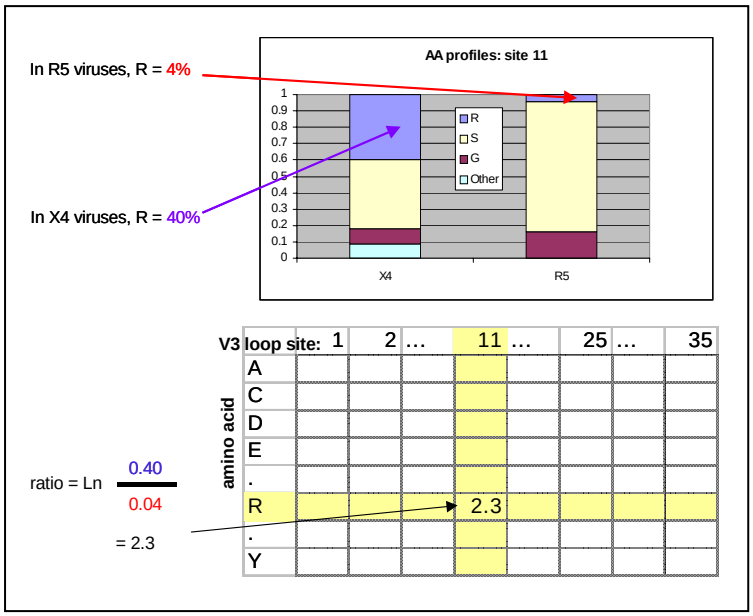


Figure 3.1: Constructing a PSSM matrix, specifically for the likelihood ratio of amino acid arginine (R) at position 11 in R5 and X4 viruses. Arginine was observed in 4% of the R5 viruses and in 40% of X4 viruses at position 11. The log ratio is then determined for R at this position, also indicating that R is more prevalent at position 11 in X4 viruses. This ratio (2.3) is then placed at position 11 for R in the matrix. The ratios for all possible amino acids at this position are calculated. This process is then repeated for each position within the V3 region thus compiling a PSSM matrix.

3.2.4 Determining the PSSM score for a target sequence

A target sequence is compared to this group of aligned sequences (matrix) with known properties. The ratio at each site for the target sequence is added to get a final score. This score indicates the likelihood that the target sequences have the property of interest, in this case, SI phenotype. Therefore, the higher the score (more positive), the more likely the target sequence is a SI virus (Figure 3.2). An optimal cut-off score is calculated from this dataset that differentiate between SI and NSI sequences, in such a manner that the best specificity (number of NSI samples correctly predicted) compared to the best sensitivity (number of SI samples correctly predicted) is obtained. But sensitivity and specificity are not independent of each other and increased sensitivity would result in lower specificity and visa versa.

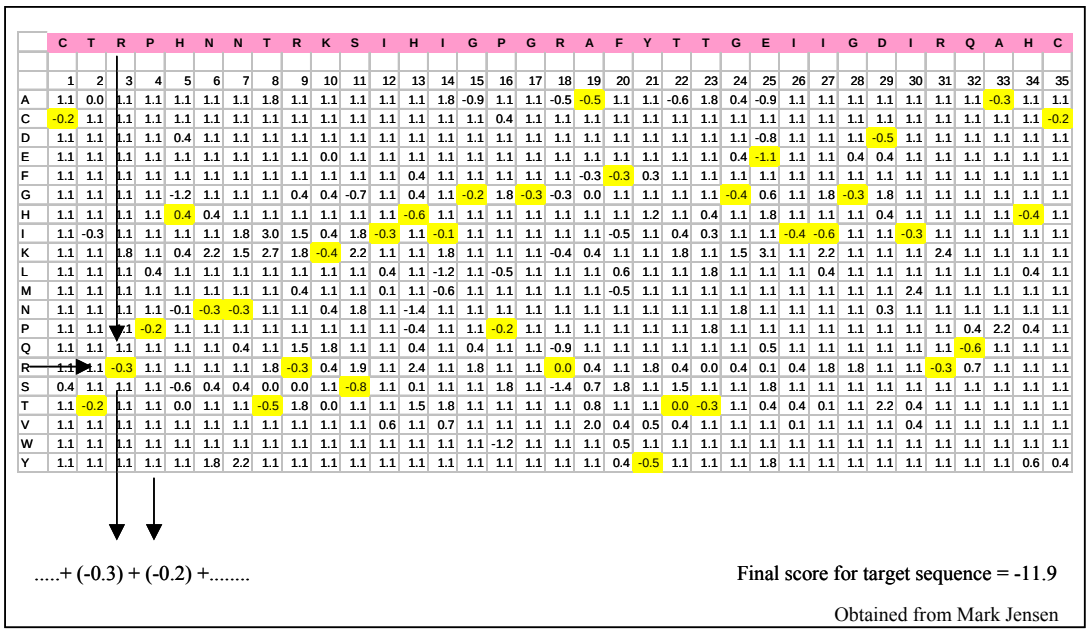


Figure 3.2: Calculating the PSSM score. A target sequence is compared to the PSSM and the likelihood score for each site is determined (highlighted in yellow). These scores are then added together to determine the likelihood of the target sequence to have the specific property, in this case X4 phenotype characteristic. If the score is more than the optimal cut-off, it is predicted as X4 and below the cut-off as R5.

3.2.5 Development and validation of a C-PSSM

We derived a predictor from position specific scoring matrices (PSSM), calculated as described in (Jensen, Li *et al*. 2003), based on the subtype C training set of 280 V3 sequences.

To determine the optimal cut-off score to differentiate between a R5 and X4 sequence, distributions of specificity (fraction of NSI sequences correctly predicted) and sensitivity (fraction of SI sequences correctly predicted) were estimated by combining data set bootstrapping with leave-one-out cross-validation. The PSSM is dependant on which sequences were used in making the matrix, therefore the bootstrapping analysis test the effect of sequence sampling on the performance of the PSSM. Because all the sequences in the data set were used to develop the PSSM, the matrix might be biased towards this set of data. Thus cross validation is done to determine if the matrix can predict the phenotype of a sequence not used in its development. Previously in the B-PSSM, the data set was randomly partitioned into subsets, but in the C-PSSM this analysis was improved by combining of the bootstrap analysis with leave-one-out cross-validation (Jensen, Li et al. 2003).

In this procedure, a target sequence was removed from the data set, and single sequences from the remaining individuals were randomly sampled with replacement (Figure 3.3). The randomly selected samples were used to calculate a PSSM predictor, and the PSSM was used in turn to predict the phenotype of the target sequence (that was removed).

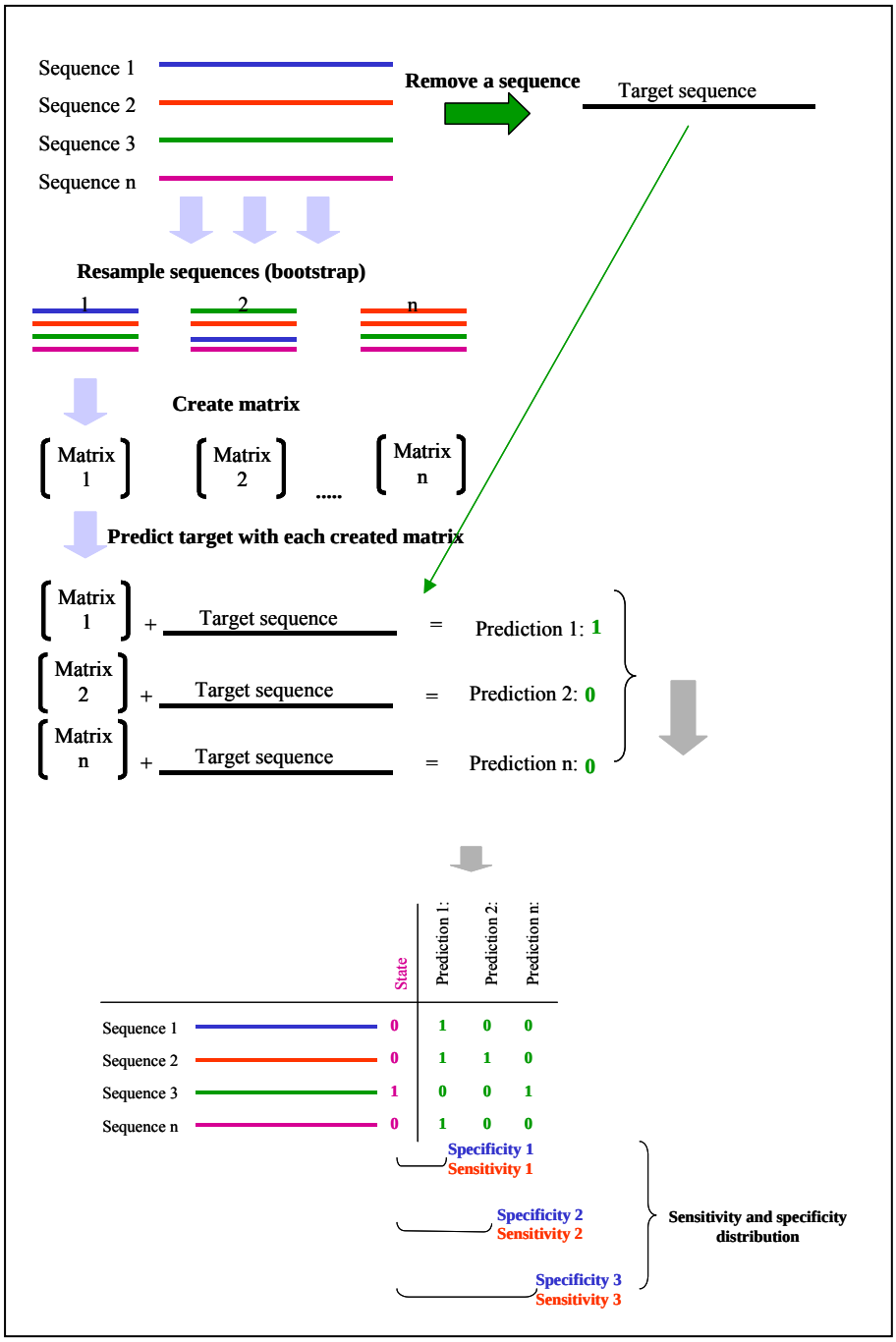


Figure 3.3: Determining sensitivity and specificity of the PSSM. Combining the leave-one out cross validation and data set bootstrapping to determine the variability of scores as well as the optimal cut-off score whereby (specificity X sensitivity) is maximised. A target sequences was removed from the data set, the remaining sequences resampled (100X) and a PSSM matrix calculated for each. The target sequence was then predicted with these matrices, as well as sensitivity and specificity for each prediction.

Resampling was repeated 100 times to obtain an empirical prediction probability for the target sequence. Each sequence in the data set was treated as a target in turn. The phenotype prediction for the target sequence was made by comparing the score of this sequence to a cut-off score: an SI prediction was called if the target score was greater than the cut-off, and an NSI prediction was called if the score was less than the cut-off. Therefore, the optimum cut-off score was calculated as that score which maximised the product of the sensitivity and specificity (sensitivity X specificity) of predictions.

For between-subtype comparison, we performed this analysis on a HIV-1 subtype B data set described in Resch *et al*. (Resch, Hoffman *et al*. 2001), consisting of 187 NSI and 70 SI sequences from 107 infected subjects.

3.2.6 Overlap coefficient analysis determining differences between subtypes B and C

To investigate the potential differences between subtypes B and C we examined the overlap coefficient [OC, (Dybul, Daucher *et al*. 2003)] between SI and NSI V3 amino acid profiles. The training sets described the amino acid frequency distribution for each site. The OC, in this context, is a site-wise measure of the difference between the SI and NSI amino acid distributions for V3 (equations for the OC are given in (Dybul, Daucher *et al*. 2003). If the OC = 0, the distributions are identical, and if OC = 1, there is no amino acid overlap between the distributions (i.e., the SI amino acids at a site are completely distinct from NSI amino acids at that site) (Figure 3.4). Thus, the OC is a measure of the ability of a V3 site to discriminate between the two phenotypes, based on the training set.

To determine whether an OC is significantly high, we compared it to a distribution of OCs generated by randomly assigning training set sequences to SI or NSI categories. OCs were calculated for 250 random permutations, and the *p*-value of the training set OC was reported as 1 minus its percentile within the random OC distribution.

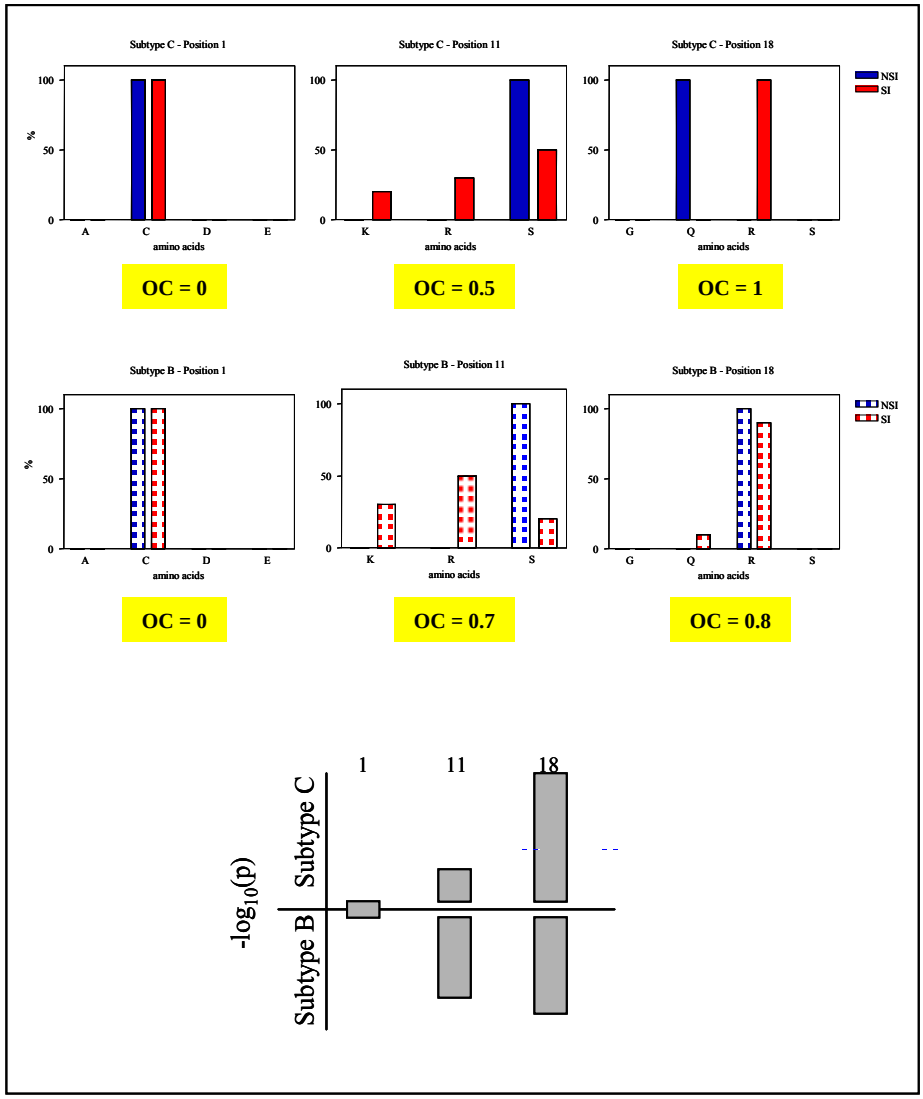


Figure 3.4: Determining the overlap coefficients within V3. A simplified example to show how OC is a measure of the ability of a V3 site to discriminate between the two phenotypes, based on the training set. The OC difference between NSI and SI for each site is highlighted in yellow (positions 1, 11 and 18 shown). OC=0, indicated identical distributions and OC=1 that there are no amino acid distribution overlap. The V3 overlap coefficient *p*-values for subtypes C (above X-axis) and B (below X-axis) were then compared to determine if different amino acids contribute to phenotype in these subtypes.

3.2.7 Comparison between predictions from V3-HTA and C-PSSM

The V3-based heteroduplex tracking assay (V3-HTA) has been used as a rapid genotype-based method to identify genetic variation associated with NSI- and SI-like viruses in subtypes B and C (Nelson, Fiscus *et al*. 1997; Ping, Nelson *et al*. 1999). The mobility of the heteroduplex reflects the differences between the probe and the sample sequence. This is measured by the mobility ratio, which is the distance traveled by the heteroduplex divided by the distance traveled by the homoduplex. The greater the genetic difference between the probe and the sample, the slower the migration of the heteroduplex and the smaller the mobility ratio. HTA mobility ratios were available for 13 NSI and 8 SI subtype C viral isolates from South Africa using a NSI probe (Coetzer, Cilliers *et al*. 2005). The C-PSSM score of each of these isolates was calculated and correlated to the V3-HTA mobility ratio to determine the relationship between the genotypic algorithm and a genotype-based molecular assay.

3.2.8 C-PSSM for public use

We have made a C-PSSM predictor available online based on the computational techniques presented in this chapter at the URL (and available for public use upon publication of the manuscript):

http://ubik.microbiol.washington.edu/computing/pssm/.

3.3 RESULTS

3.3.1 HIV-1 subtype C sequences

The data set compiled for this study consisted of 280 HIV-1 subtype C V3 sequences with known biological phenotypes consisting of 229 NSI V3 sequences (from 200 subjects) and 51 SI V3 sequences (from 20 subjects). A graphical representation of these sequences using sequence logos (Schneider and Stephens 1990) is shown in Figure 3.5.

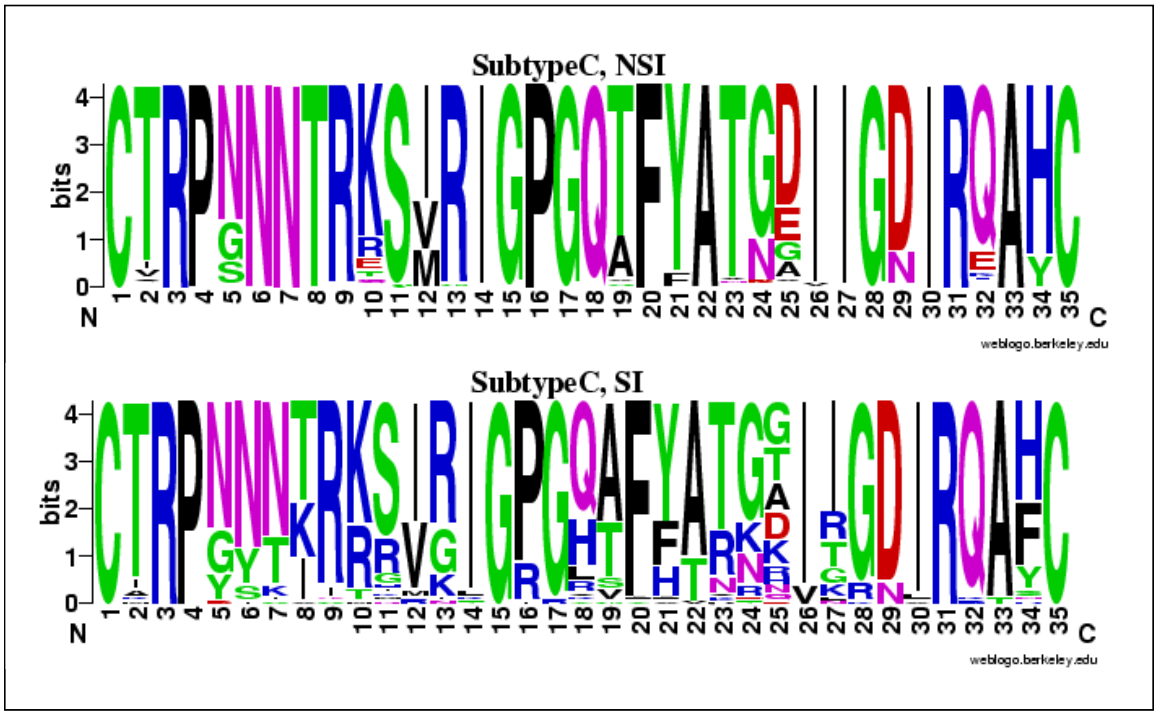


Figure 3.5: Sequence logos of HIV-1 subtype C V3 sequences used in this study. The character and size of each logo represents the proportion of an amino acid at the specific site. The subtype C NSI data set is represented by 229 NSI V3 sequences, and the SI data set corresponds to 51 SI V3 sequences.

There was a high degree of amino acid variation within the V3 region of subtype C SI sequences that distinguished them from the NSI sequences. Overall 29 of the 35 amino acid loci among SI viruses showed variation compared to only 12 in NSI viruses. Given the fact that more than 4 times the number of NSI sequences was available, the limited V3 variability of NSI as compared to SI viruses is unlikely to result from sampling error. The GPGQ crown motif, typical of subtype C, was highly conserved among the NSI sequences but showed variation at positions 16 and 18 among SI sequences; in particular the Q at position 18 was heavily substituted. There was also increased variation at positions 11, 13 and 25 and 27 in the SI sequences. In subtype B, positions 11 and 25 are often positively charged amino acids in SI viruses (Hoffman, Seillier-Moiseiwitsch *et al*. 2002), but in this subtype C data set this was less evident.

3.3.2 Predicting phenotypes from genetic sequence data

The performance of 4 commonly used phenotype predictors was evaluated on a subset of 220 sequences, representing unique sequences (one randomly selected from each individual) from the data set (Table 3.1). All algorithms predicted the NSI phenotype with a high degree of accuracy (99.5%), except for the Briggs method where only 52% of the NSI sequences were correctly identified. For SI viruses all 4 algorithms performed poorly in predicting biological phenotypes. This may be due to the fact that all four algorithms were developed largely with subtype B sequences. Of these algorithms the PSSM had increased predictive power compared to other methods (Jensen and van 't Wout 2003) and we therefore chose this method to develop a subtype C specific predictor.

Table 3.1: Comparison of the performance of available prediction methods to determine viral phenotype of the subtype C unique data set. Pillai method and B-PSSM are web-based tools and the pre-trained classifiers selected for analysis are in brackets.

	% Phenotype correctly predicted		**Correct overall prediction value** (% specificity X % sensitivity)
Method	**NSI** (n=200)	**SI** (n=20)	
11/25	99.5	47.8	0.48
Briggs	52.0	34.8	0.18
Pillai (SVM)*	99.5	52.0	0.52
B-PSSM (sinsi)#	99.5	52.0	0.52

* http://genomiac2.ucsd.edu:8080/wetcat/index.html
# http://ubik.microbiol.washington.edu/computing/pssm/

3.3.3 Performance of C-PSSM prediction

A subtype C PSSM was derived using the data set from 280 V3 sequences containing 35 amino acids. The C-PSSM scores of the NSI viruses ranged from –27.7 to –9.0 (median of –24.8), while for SI viruses the range was –28.1 to 9.3 (median of –9.8) (Figure 3.6). The median C-PSSM score distributions for NSI and SI differed significantly ($p<10^{-15}$) by the Kruskal-Wallis test.

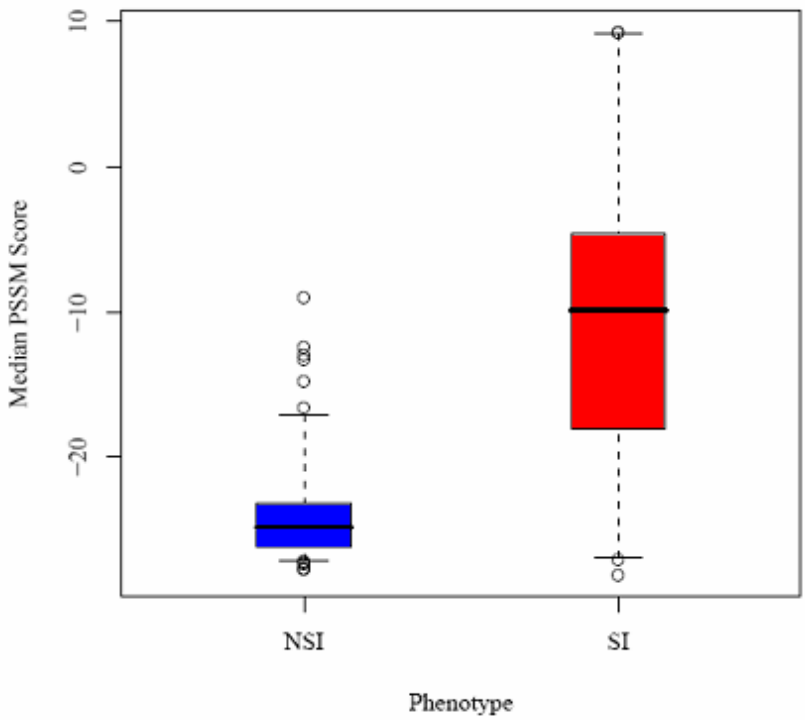


Figure 3.6: Comparison of C-PSSM score distribution of 229 NSI and 51 SI subtype C sequences. Score is the median PSSM score over 100 bootstrapped data sets, as described in Materials and Methods. Box boundaries: interquartile range; central line; median over sequences; error bars extend from the 2.5th and 97.5th percentiles; beyond these, sequences are represented as outlier points. Kruskal-Wallis test: chi-squared=78.9, $p < 10^{-15}$.

The performance of the C-PSSM on the subtype C data set was compared to that of the B-PSSM on a subtype B data set (Jensen, Li *et al*. 2003), (Figure 3.7). The C-PSSM had a specificity of 94% (C.I., [92%-96%]) and sensitivity of 75% (C.I., [68%-82%]). Comparison with the B-PSSM in Table 1 showed the C-PSSM to have a lower specificity but a higher sensitivity. The correct overall prediction value for the C-PSSM was higher (0.71; 94% x 75%) than the B-PSSM predictor on subtype C sequences (0.52; 99.5% x 52%). The B-PSSM, applied to subtype B targets within the Resch *et al*. (2001) data set.

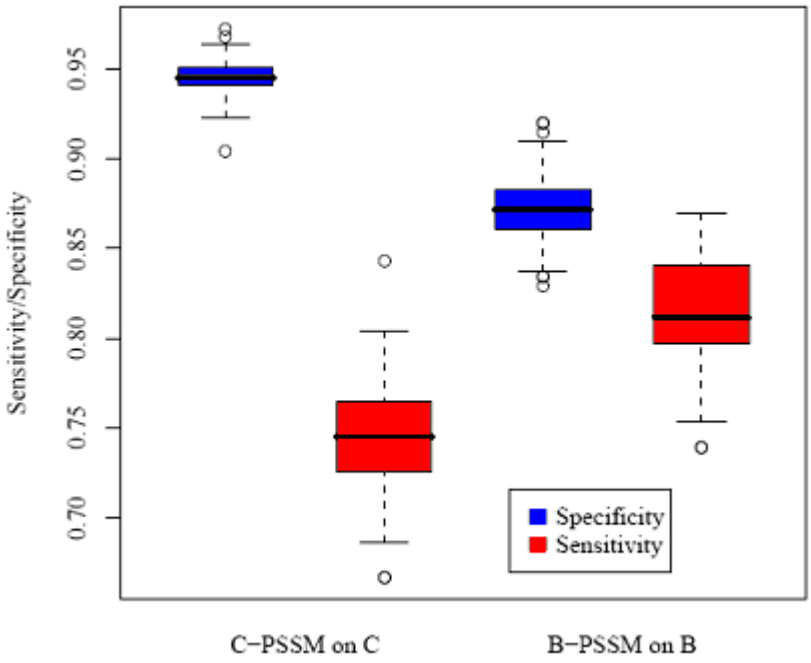


Figure 3.7: Comparison of C-PSSM on a subtype C data set with B-PSSM on a subtype B data set, using leave-one-out/bootstrap predictions.

To determine whether the sensitivity and specificity might be significantly improved by increasing the number of training sequences, we performed leave-one-out/bootstrap analysis on subsets of the data set. Total size of the subsets was increased incrementally, and unique SI and NSI sequences were randomly selected in a ratio of 1:10, comparable to the ratio in the total data set (Figure 3.8). The specificity of the C-PSSM for predicting NSI phenotypes was high at even the smallest total sample size, and it did not improve significantly when the sample number was increased. The sensitivity of the C-PSSM for predicting SI phenotypes at low sample numbers was poor but appeared to approach a limit as the sample size increased to approximately 100.

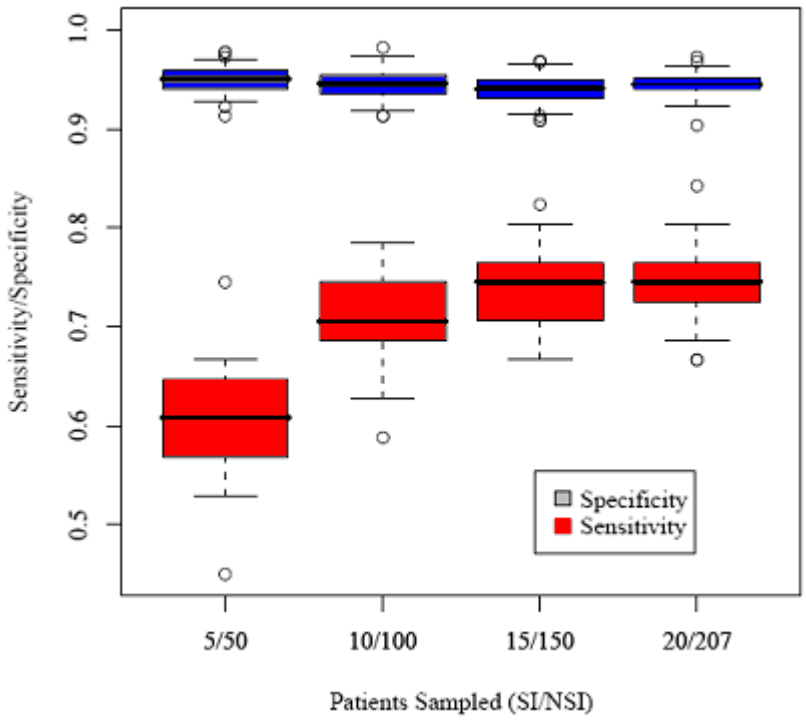


Figure 3.8: Effect of sample size on the sensitivity and specificity of the C-PSSM. Boxes as in Figure 3.6.

3.3.4 Site-wise differences in phenotype between subtypes B and C

Because SI viruses are so much less prevalent in subtype C than subtype B populations, it is possible that different, less evolutionarily labile sites influence the manifestation of phenotype in subtype C. The availability of both subtype B and subtype C training sets afforded us a chance to investigate potential differences using overlap coefficients (OC). Figure 3.9 displays the *p*-values for the OC for each V3 site, comparing subtype B [based on the SI/NSI data set of Resch *et al*. (2001)] and subtype C using our data set. Sites with OCs that exceed the 95th percentile of the random permutation distribution (depicted in Figure 5 as bars that extend beyond the dotted lines) have amino acid distributions that are significantly different between SI and NSI viruses. Sites with non-significant OCs are less informative for purposes of

discriminating between SI and NSI viruses by genotype. Under this interpretation, the OC analysis highlights sites that are potentially different in their influence on phenotype between the two subtypes. In particular, V3 sites 12, 15, 16, 26, 27, 28, 33, and 34 (highlighted in grey in Figure 3.9) have significant OC values in one but not the other subtype.

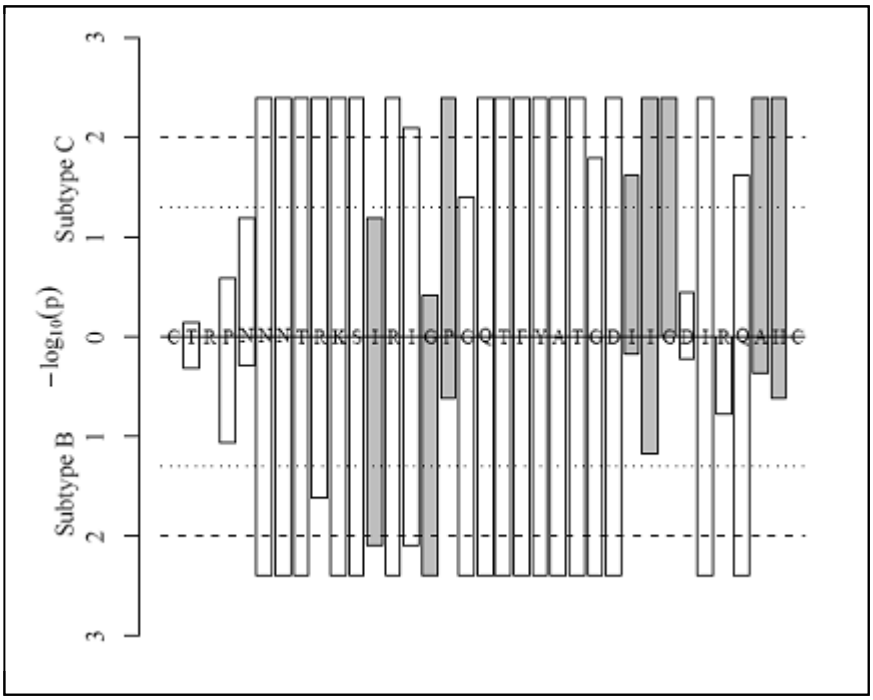


Figure 3.9: V3 overlap coefficient *p*-values for subtypes C (above X-axis) and B (below X-axis). Dotted line: $p = 0.05$; dashed line: $p = 0.01$. X-axis: V3 site, labelled with subtype C NSI consensus residue. Grey pairs: sites in which the measured OC was significant in one subtype and non-significant in the other.

3.3.5 Comparison of C-PSSM scores with a genotype-based assay

V3 heteroduplex tracking assays (V3-HTA) have been used to identify NSI- and SI-like viruses when hybridised with a R5 probe (Nelson, Fiscus *et al*. 1997; Ping, Nelson *et al*. 1999; Coetzer, Cilliers *et al*. 2005). Samples with a similar sequence to the probe have a high mobility ratio (k>0.90) and are usually NSI whereas samples with a low mobility ratio (k<0.90) are different from the probe and are frequently SI (Coetzer, Cilliers *et al*. 2005). The mobility ratio of 8 SI (5 R5X4 and 3 X4) and 13 NSI isolates were compared to their C-PSSM scores (Figure 3.10). There was a highly significant correlation between the mobility ratio and PSSM score of the samples ($p < 0.0001$; r^2: 0.53). NSI viruses had a high mobility ratio and decreased PSSM score and most of these clustered tightly together. The two NSI samples that had a lower mobility ratio (0.76 and 0.80) had amino acid deletions that probably accounted for this, but this did not affect the C-PSSM score. The SI samples had a broader distribution of mobility ratios and C-PSSM scores.

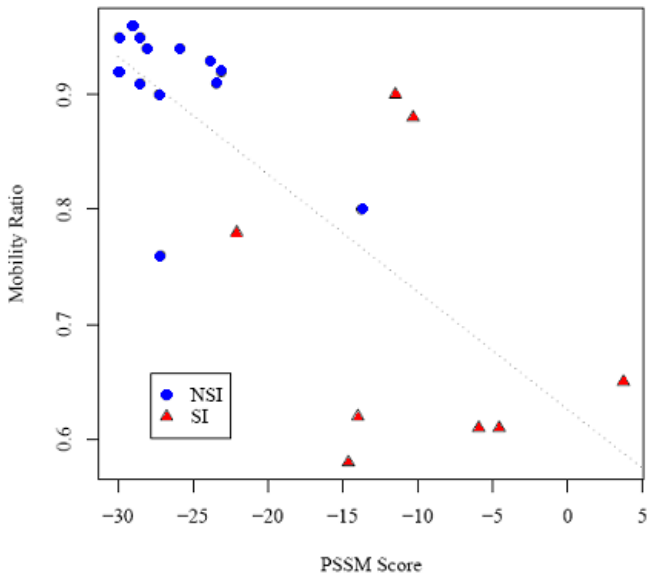


Figure 3.10: The HTA mobility ratio compared to the C-PSSM score for 13 NSI and 8 SI isolates. The NSI samples had high mobility ratio and very low PSSM score, whereas the SI samples had variable mobility ratios (but usually below 0.90) and higher PSSM scores (above optimal cut-off of -21.64). Dotted line: linear regression ($p < 0.0001$; r^2: 0.53).

3.4 DISCUSSION

We tested 4 available phenotype prediction methods developed for HIV-1 subtype B, on HIV-1 subtype C sequences with known phenotypes and found them to be highly accurate in predicting NSI usage, but less so in predicting CXCR4 usage. We therefore derived a subtype C specific phenotype predictor that performs nearly as well on subtype C V3-loops as do existing subtype B-specific methods on subtype B V3-loops. This correlated well with a genotype-based method for detecting NSI and SI viruses (V3-HTA) suggesting that the C-PSSM can be applied to subtype C V3 sequences of unknown coreceptor usage.

Sequence logos highlighted appreciable differences between V3 sequences of NSI and SI subtype C viruses. The NSI data set was very homogenous with little or no variation at many of the amino acid sites, while in the SI data set there was greater variation at most sites. These data suggested that sufficient genetic variation between NSI and SI subtype C sequences exists that can be used to differentiate these phenotypes. However, none of the available prediction methods were adequately able to exploit these differences in differentiating NSI from SI viruses. The 11/25 rule is based on the presence of positively charged amino acids at positions 11 and/or 25 (Fouchier, Groenink *et al*. 1992; Milich, Margolin *et al*. 1993). However, more than 50% of the subtype C SI sequences in this study did not have a positively charged amino acid at these positions. Furthermore, while this method is considered to be a reliable sequence-based phenotype predictor, other studies have shown that more than two amino acid positions need to be considered when assigning phenotype (Resch, Hoffman *et al*. 2001). The Briggs' method performed the least well of all the algorithms evaluated. This method is based on genotype variables in the V3 region

derived from subtype B sequences with NSI viruses having a net charge <+4 and SI viruses >+4. The net charge of the subtype C NSI data set ranged from +0 to +6 and the SI data set had a net charge between +3 to +9 which probably explains the poor performance of this algorithm. The Pillai method is a two-way classification method that differentiates between viruses able or unable to use CXCR4. Limitations of this method include the fact that it misclassifies R5X4 viruses (Pillai, Good *et al*. 2003). These dual tropic viruses represent an intermediate stage of coreceptor evolution in subtype B, (Jensen, Li *et al*. 2003; Yi, Shaheen *et al*. 2005) and are usually grouped into the SI data set, as was done in this study. Our SI data set contained 9 dual tropic viruses of which 6 were incorrectly predicted as NSI, indicative of the low predictive power of the Pillai method on this data set.

While most methods could accurately identify NSI viruses it was clear that a new method was needed to improve the sensitivity of prediction of SI viruses from subtype C. The PSSM method has proven to be a simple yet reliable phenotype method based on V3 amino acids with multiple applications (Jensen, Li *et al*. 2003; Jensen and van 't Wout 2003). The subtype C specific PSSM addressed some of the limitations of other prediction methods including the B-PSSM (Jensen, Li *et al.* 2003). In particular, the C-PSSM identified SI viruses more reliably than other methods resulting in a major increase in sensitivity. However, the sensitivity and specificity in the PSSM are not independent of each other, but are determined by the cutoff value that differentiates between NSI and SI sequences. Thus the specificity of the C-PSSM was slightly less than other prediction methods as show in Table 3.1. We therefore compromised specificity for increased sensitivity in identifying SI viruses. To compare all the various prediction algorithms we multiplied the specificity and

sensitivity to give an overall correct prediction value. For all the methods described in Table 1 the values were less than 0.52. For the C-PSSM this value was 0.71 similar to the subtype B PSSM on subtype B sequences. Overall, the C-PSSM gives better prediction for subtype C sequences than all other methods including the B-PSSM, but is similar to the B-PSSM using subtype B sequences.

It was previously shown by Jensen *et al*. (2003) that the PSSM score represents the "X4 potential" of a sequence and intermediate scores correlate well with the evolution of viruses within an individual. This method has also contributed to a better understanding of the role of intermediates (R5X4) in the transition of R5 to X4 in subtype B (Jensen, Li *et al*. 2003) and can now be applied to subtype C viruses where this transition has not often been reported. To improve the prediction quality of the C-PSSM vigorous sampling of more patients will be required, at least in the present ratio of SI to NSI sequences. Therefore, future sampling should focus on acquisition of more X4 sequences from new individuals.

Potential site differences between subtype B and C within the V3 region were investigated using current training sets by overlap coefficient analysis. This suggested that changes in the crown at site 15 will influence coreceptor usage in subtype B, but changes at site 16 will have more influence in subtype C. Other sites that had significant OC values in one but not the other subtype were 12, 16, 26, 27, 28, 33, and 34. These are unlikely to be simple artifacts of sampling, since the majority of sites are congruent (either both significant or non-significant) between the subtypes, and both possibilities (B significant, C non-significant, and vice versa) are represented at the incongruent sites. However, we are not claiming that incongruent sites constitute

a rejection of any explicit null model. Rather, this simple analysis suggests the possibility of differential phenotypic effects of mutations at certain sites that could be evaluated in future studies.

The C-PSSM represents an improvement over currently available methods for predicting SI viruses in subtype C. This could aid in the better identification and understanding of subtype C coreceptor usage. With the increased availability of small molecule fusion inhibitors and use of antiretroviral therapy in patients infected with subtype C, there is concern that certain therapies may increase the risk of developing X4 viruses during subtype C infections. PSSM scores may be a useful tool for assessing baseline risk (Jensen and van 't Wout 2003). Although the number of HIV-1 subtype C SI viruses available is a limiting factor, this study has shown that currently available data provide a good initial basis for subtype C coreceptor usage predictor.

CHAPTER 4

CHARACTERISATION OF THE gp160 REGION FROM HIV-1 SUBTYPE C ISOLATES AND CORRELATION TO CORECEPTOR SWITCHING

4.1 INTRODUCTION

HIV-1 is characterised by extreme genetic diversity that lends the virus the ability to escape from selective forces such as host immune response or antiretroviral drugs, thus the virus is continually adapting and evolving. This genetic diversity in turn, contributes to the lack of immune control and increased disease progression. Understanding the complex interaction between virus and host immune system are very necessary, in particular characterisation of the viral genetic adaptations, in the development of vaccine and drug strategies.

Although the CD4 and coreceptor binding sites are not very accessible and only exposed for brief periods of time during viral entry, they are very neutralisation sensitive (Derdeyn, Decker *et al*. 2004). This has resulted in high genetic diversity of these envelope regions to escape immune pressure. The diversity within HIV is the result of mutations during replication (due to the high error-prone rate), as well as recombination between viral genomes (Preston, Poiesz *et al*. 1988; Perelson, Neumann *et al*. 1996; Rambaut, Posada *et al*. 2004). The genetic adaptations of the envelope gene include changes within the variable loops, net amino acid charge, tropism, and glycosylation sites.

In this study the genetic diversity over time within the gp160 region in five patients infected with HIV-1 subtype C were investigated. In particular, we aimed to determine if other regions or specific sites within gp160 could be associated with coreceptor changes in HIV-1 subtype C isolates.

4.2 MATERIALS AND METHODS

4.2.1 Isolation and coreceptor usage of HIV-1 subtype C viruses

Five female commercial sex workers, from Kwa-Zulu Natal, South Africa participating in a Phase III clinical trial of vaginal microbicide (N9) were followed for between two to four years. Ethical clearance was obtained from the University of Witwatersrand Committee for Research on Human Subjects (see Appendix A). Twenty-three samples were collected for virus isolation, viral load and CD4 counts, as previously described in Chapter 2.2.1. Viral isolates were tested for their ability to replicate in U87.CD4 cells transfected with either CCR5 or CXCR4, as previously described (Morris, Cilliers *et al*. 2001; Cilliers, Nhlapo *et al*. 2003).

4.2.2 Viral RNA isolation and gp160 sequencing

Viral RNA from the 23 isolates was extracted from PBMC culture supernatant using a MagnaPure LC Isolation station and the Total Nucleic Acid isolation kit (Roche Applied Science, Penzberg, Germany). A reverse transcription step was performed and the cDNA was used to amplify the gp160 region with primers envA and envM as previously described (Gao, Morrison *et al*. 1996). PCR products were purified using the High Pure PCR Product Purification kit (Roche Diagnostics GmbH, Mannheim, Germany). Sequencing reactions were done in a MicroAmp 96 well optical reaction plate using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA) and resolved on an ABI 3100 automated genetic analyser.

4.2.3 Sequence analysis

Sequences were assembled and edited using Sequencher (version 4.0) software (Genecodes, Ann Arbor, MI). The envelope sequences were aligned with CLUSTAL X (version 1.8.1) (Thompson, Gibson *et al*. 1997) and manually edited in BioEdit (version 5.0.9). Phylogenetic trees were constructed from the gp120 and gp41 regions using MEGA (version 2.1) (Kumar, Tamura *et al*. 2001). Subtype reference sequences were downloaded from the Los Alamos database (www.hiv.lanl.gov). Predicted N-glycosylation sites were determined with the web base program N-GLYCOSITE (www.hiv.lanl.gov). Recombination within samples from Du151 was investigated with the SimPlot program (version 2.5).

4.3 RESULTS

4.3.1 Clinical information of patients

Twenty-three samples from five patients were investigated in this study. These patients had acute HIV-1 infection and were followed for two to four years (Table 5.1). During this follow-up period all patients had a decline in CD4 count. The viral load declined over time as these patients were entering the steady-state phase, except in Du151 who maintained a very high viral load. Three patients (Du123, Du368 and Du422) had R5 viruses at all stages and two patients (Du151 and Du179) underwent coreceptor changes. Du151 switched coreceptor usage from R5 to R5X4, while Du179 had an R5X4 virus at the time of first isolation (19.8 months) and then lost the ability to use CCR5. These two patients (Du151 and Du179) were also dually infected as they were found to have two genetically distinct viruses (Gottlieb, Nickle *et al*. 2004; Grobler, Gray *et al*. 2004). Du151 was a rapid progressor and died within 2 years of infection. Isolates from patients Du151 and Du422, have been selected as vaccine strains based on their similarity to a derived South African consensus sequence (Williamson, Morris *et al*. 2003).

Table 4.1: Clinical information of the five patients followed longitudinally for between two to four years.

Isolate	Biotype	Months of infection	CD4 (cells/µl)	Viral load (copies/ml)
Du123				
Nov-98	R5	3.5	841	19 331
Mar-00	R5	20	546	6 558
May-00	R5	21.6	779	2 568
Sep-00	R5	25.6	340	6 018
Nov-00	R5	28.2	616	3 695
Du151				
Oct-98	R5	1.1	NA	NA
Nov-98	R5	2.1	367	>500 000
Jun-99	R5	8.8	239	>500 000
Mar-00	R5X4	18.6	143	>500 000
May-00	R5X4	19.9	66	>500 000
Du179				
Mar-99	R5X4	19.8	435	4 895
May-99	R5X4	21.9	279	2 640
Feb-00	R5X4	31.2	259	3 131
May-00	X4	33.8	231	2 228
Oct-03	X4	74	NA	NA
Du368				
Nov-98	R5	8.3	670	13 933
Feb-00	R5	24	628	NA
Aug-00	R5	30	531	11 247
Du422				
Jan-99	R5	4.4	397	67 982
May-99	R5	8.2	NA	16 098
Mar-00	R5	17.7	355	11 346
May-00	R5	20.4	325	6 717
Sep-00	R5	24.5	220	8 114

NA: not available
Months of infection determined from midpoint between last antibody negative and antibody positive sample

4.3.2 Phylogenetic analysis

Phylogenetic analysis based on gp160 sequences showed that all samples clustered with subtype C references (Figure 4.1). In four of the patients, isolates from a patient grouped together in the gp120 and gp41 trees. Although isolates from Du151 clustered within the subtype C group, some of the time points were separated by unlinked sequences, indicating dual infection with two different subtype C viruses. Comparable results were seen in the gp41 tree. Thus the Du151(JUN99) isolate clustered in the gp120 tree with the earliest isolates of Du151 and in the gp41 tree with the later isolates of Du151, indicating inter-patient recombination of the subtype C strains within Du151. Although Du179 was found to be dually infected (Grobler, Gray *et al.* 2004), this was not observed in these phylogenetic trees probably due to selective growth of viral populations in our culture system.

Further recombination analysis was done to determine the breakpoint within the Du151(JUN99) (Figure 4.2). This analysis showed that there was a single breakpoint at amino acid position 680 situated within the gp41 region. The first part of Du151(JUN98) sequence showed high similarly with the OCT98 and NOV98 sequences. After the break point the JUN98 sequence was more similar to the MAR00 and MAY00 sequences.

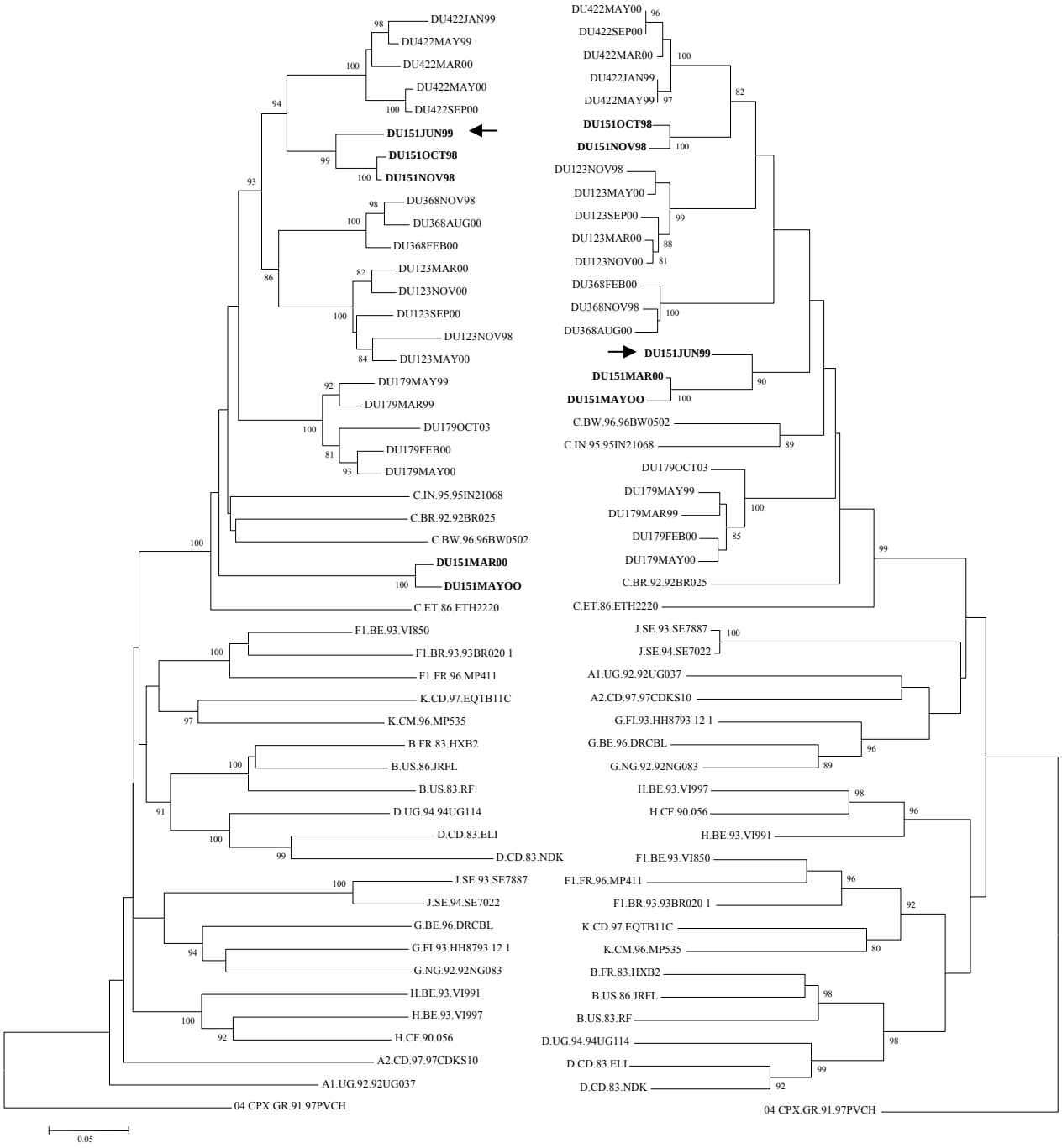


Figure 4.1: Neighbor joining tree of HIV-1 gp120 (left) and gp41 (right). Reference sequences from Los Alamos database (available at http://www.hiv.lanl.gov) were also included. Bootstrap values (1000 replicates) >80% are shown and scale denotes 5% divergence. Dual infected patient Du151 in bold. Arrow denotes Du151(JUN99) with inter-patient recombination.

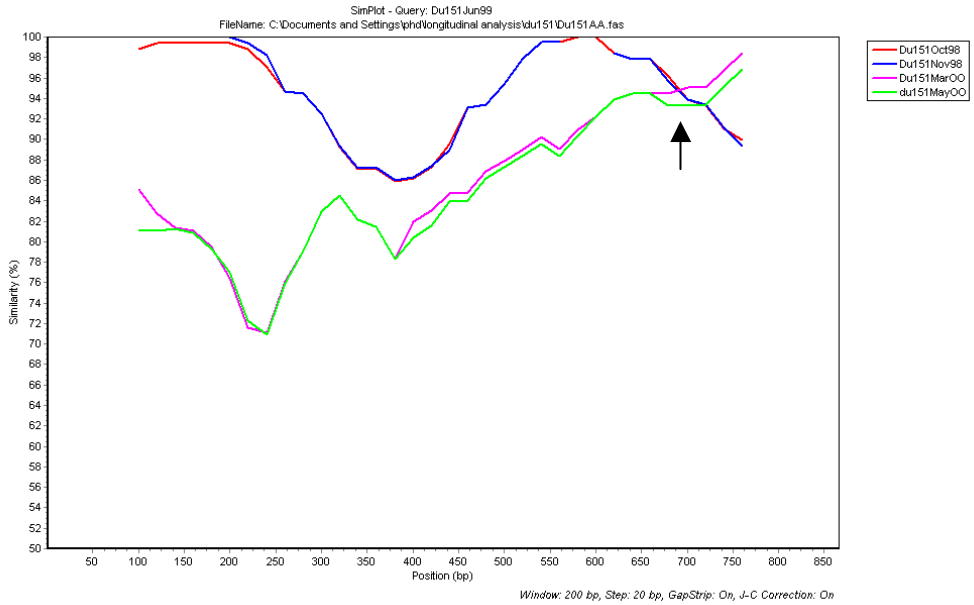


Figure 4.2: Recombination analysis of Du151(JUN99) using SimPlot. Du151(JUN99) sequence is a recombinant with the most similarity to Du151(OCT98) and Du151 (NOV98). After the breakpoint (indicated with an arrow) at amino acid position 680 the Du151(JUN99) sequence has more similarity with Du151(MAR00) and Du151(MAY00).

4.3.3 Genetic variation within a patient during disease progression

The gp160 sequences from various time points within a patient were compared (Figure 4.3 A-E) to investigate genetic changes during disease progression and whether these changes could be associated with coreceptor switching. This investigation focused on the genetic variation of different regions within gp160, variable loop length differences and predicted N-glycosylation sites. Genetic changes that might influence MAb (monoclonal antibody) recognition sites over time were also studied.

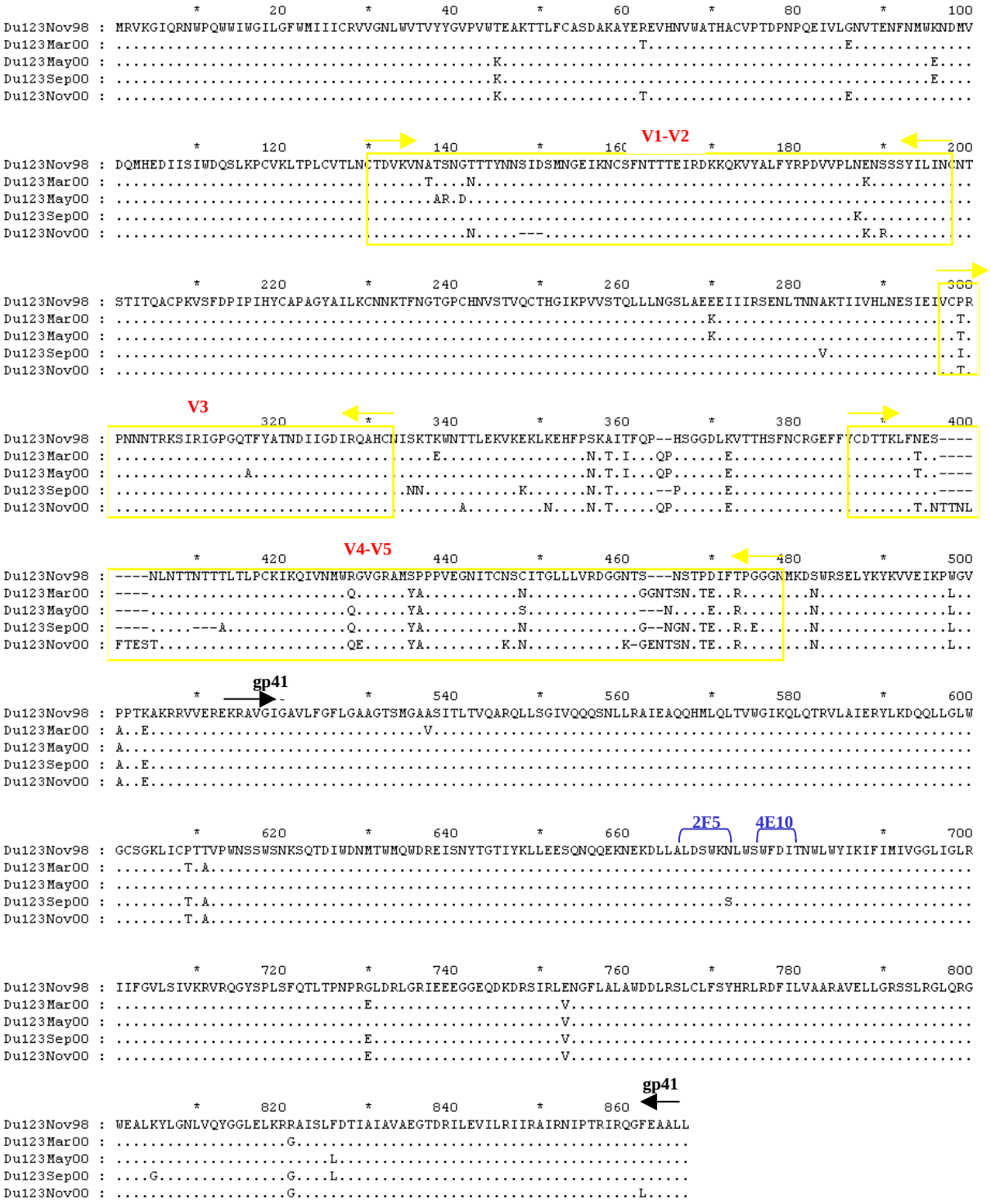


Figure 4.3 A: Du123 amino acid alignment of the gp160 region from five follow-up samples. Variable loops shown with yellow boxes, MAb epitopes 2F5 and 4E10 in blue and gp41 region indicated.

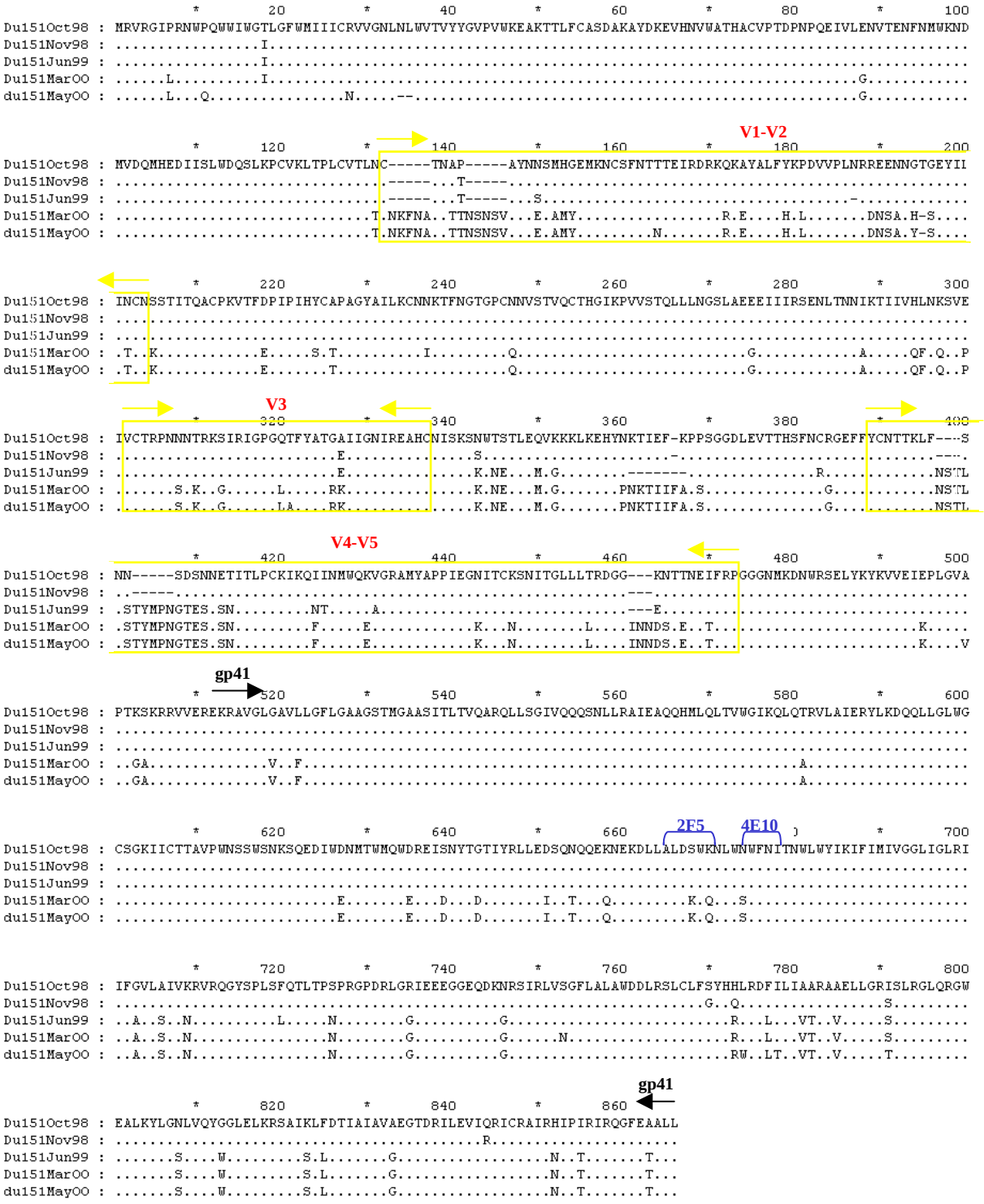


Figure 4.3 B: Du151 amino acid alignment of the gp160 region from five follow-up samples. Variable loops shown with yellow boxes, MAb epitopes 2F5 and 4E10 in blue and gp41 region indicated.

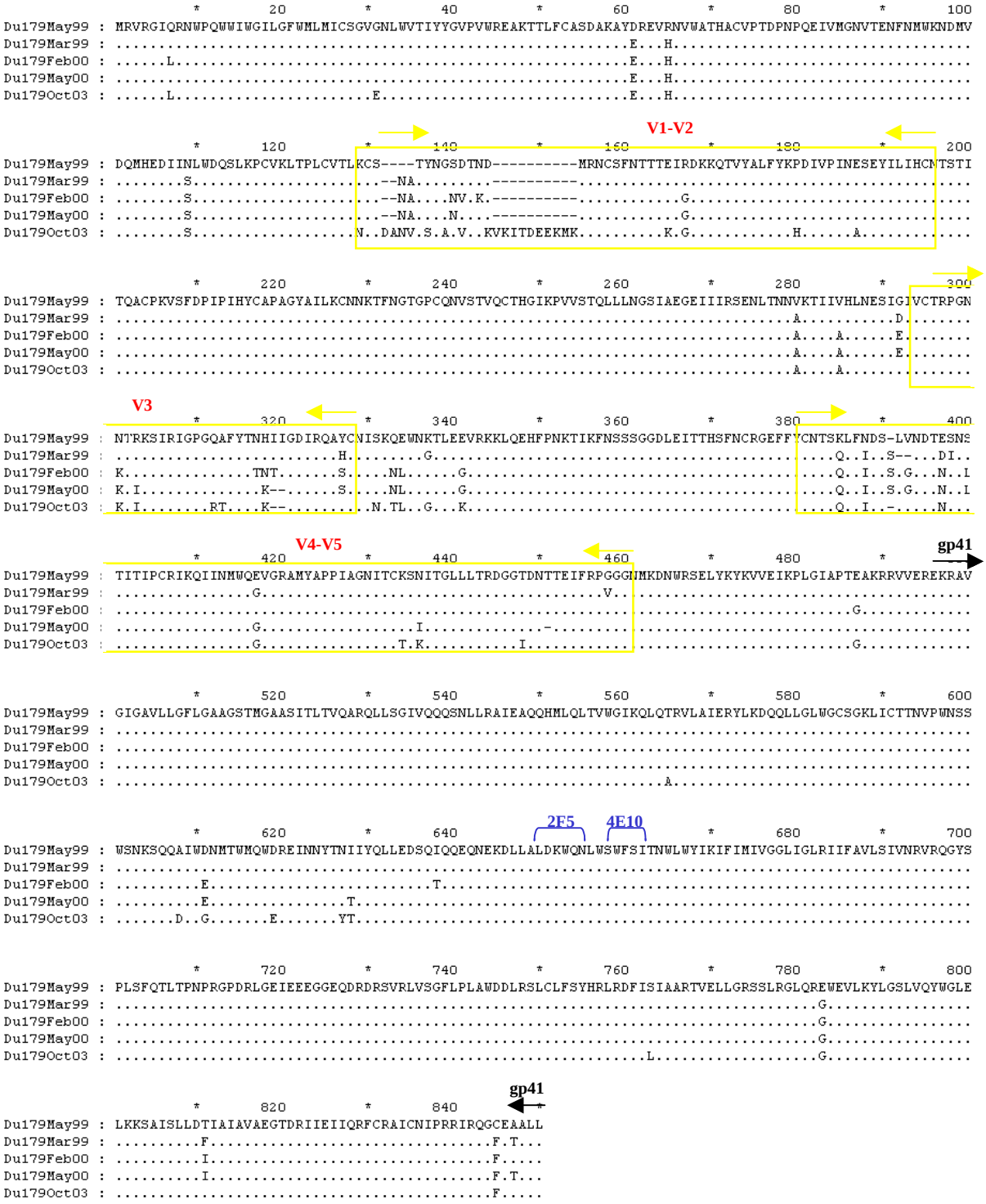


Figure 4.3 C: Du179 amino acid alignment of the gp160 region from five follow-up samples. Variable loops shown with yellow boxes, MAb epitopes 2F5 and 4E10 in blue and gp41 region indicated.

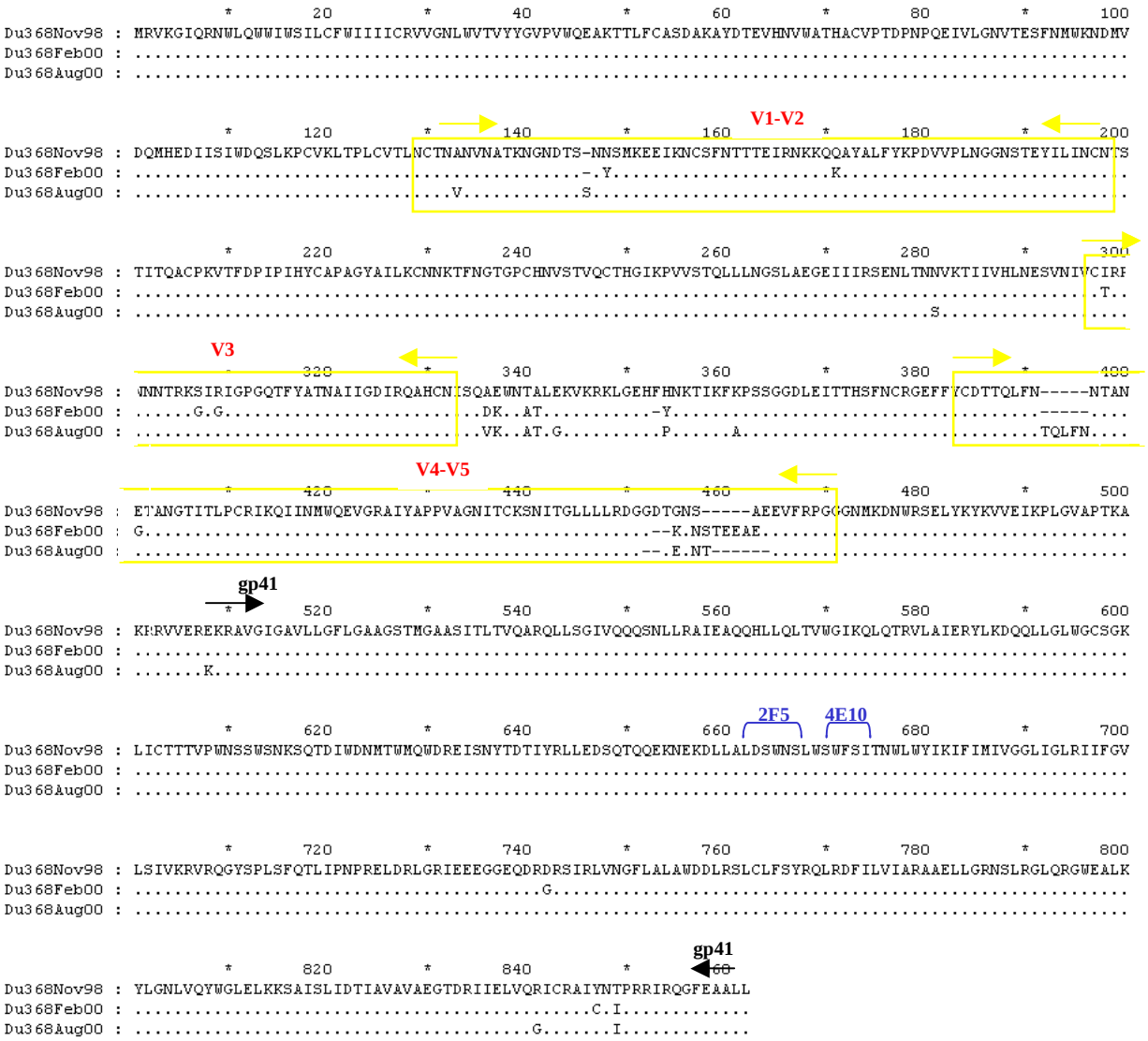


Figure 4.3 D: Du368 amino acid alignment of the gp160 region from three follow-up samples. Variable loops shown with yellow boxes, MAb epitopes 2F5 and 4E10 in blue and gp41 region indicated.

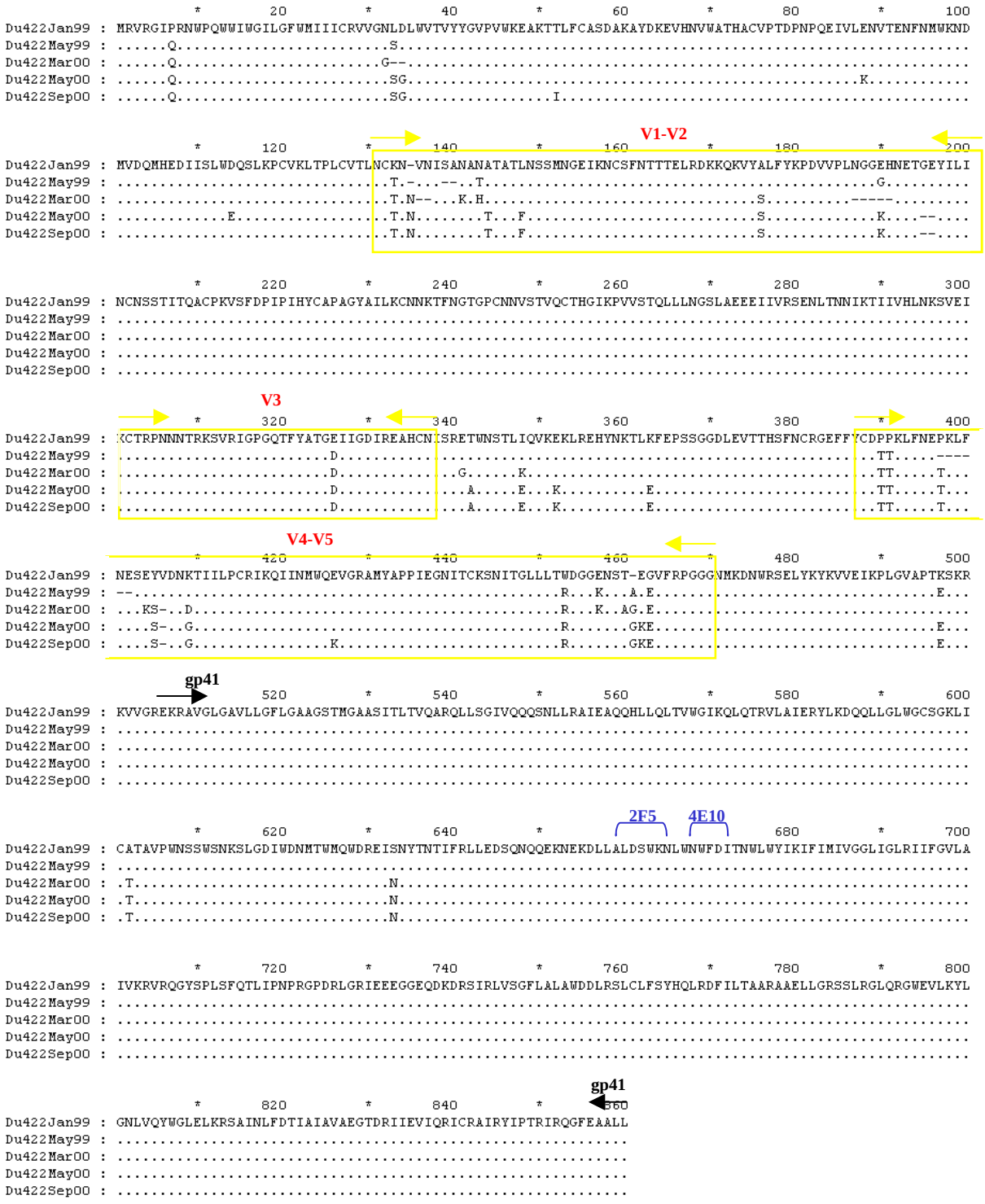


Figure 4.3 E: Du422 amino acid alignment of the gp160 region from five follow-up samples. Variable loops shown with yellow boxes, MAb epitopes 2F5 and 4E10 in blue and gp41 region indicated.

The genetic variability within a patient was calculated for the V1-V2, V3, V4-V5 and gp41 region to determine the degree of genetic variation within the envelope region over time (Figure 4.4). As expected, the gp120 region was more variable than gp41. Comparing the different regions within gp120, V1-V2 and V4-V5 loops had the most genetic variation (both median 9%, range 5-31% and 3-20% respectively,) between patients, compared to the V3 region with the least (median 6%, range 3-26%). Individual analysis showed that Du151 had the most variable sequences in all the different regions, followed by Du179.

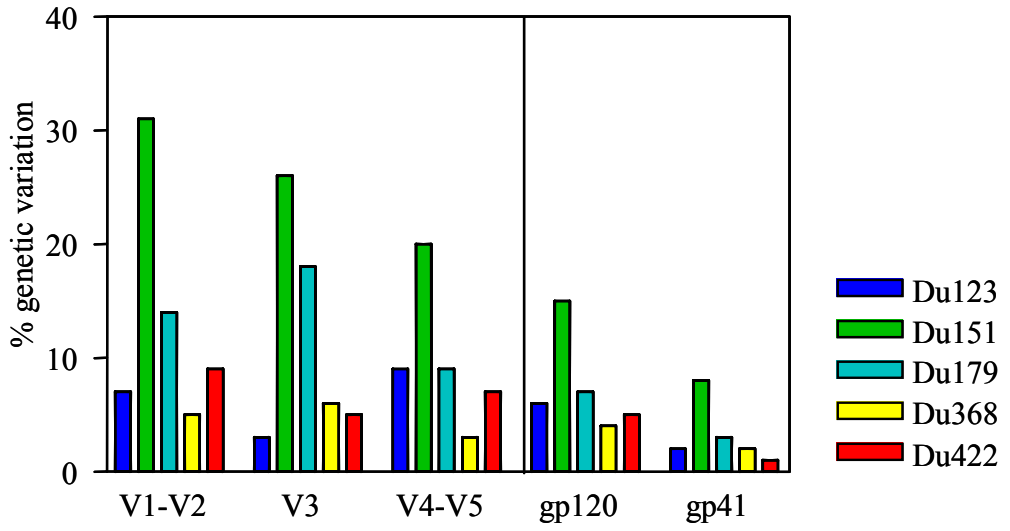


Figure 4.4: Percentage genetic variability of the envelope region within a patient over time. Du151 had overall the highest genetic variability in all regions (green bars). The V1-V2 and V4-V5 had the most genetic variable within gp160.

4.3.4 Variable loops V1-V5

Changes within the length of the variable loops (V1-V5) were determined within each patient over time (Table 4.2). In four patients (Du123, Du151, Du179 and Du368) the total number of amino acids within V1-V5 increased. The V1 and V4 loops had the most length variation between and within patients, ranging from median 17-30 amino acids, whereas the V2 to V5 had relatively constant length within a patient over time. The V3 region remained constant in all patients at all time points. There were no obvious differences noted in sequences from patients able to use CXCR4 as coreceptor.

Table 4.2: The length of the variable loops and number of predicted N-glycosylation sites within the gp160 region of each patient.

Isolate	Length (nr of amino acids)						Predicted N-glycosylation sites
	V1	V2	V3	V4	V5	V1-V5	
Du123							
Nov-98	30	40	35	23	12	140	28
Mar-00	30	40	35	23	15	143	30
May-00	30	40	35	23	12	140	29
Sep-00	30	40	35	20	13	138	29
Nov-00	27	40	35	31	14	147	31
Du151							
Oct-98	18	45	35	22	11	131	29
Nov-98	18	45	35	22	11	131	28
Jun-99	18	44	35	30	11	138	29
Mar-00	28	44	35	30	13	150	29
May-00	28	44	35	30	13	150	28
Du179							
Mar-99	17	38	34	24	10	123	30
May-99	15	38	34	25	10	122	28
Feb-00	17	38	34	26	10	125	29
May-00	17	38	34	26	10	125	28
Oct-03	29	38	34	25	10	136	29
Du368							
Nov-98	27	41	35	23	9	135	27
Feb-00	27	41	35	23	14	140	31
Aug-00	28	41	35	28	7	139	29
Du422							
Jan-99	27	44	35	28	9	143	27
May-99	24	44	35	22	10	135	27
Mar-00	24	39	35	27	10	135	28
May-00	28	42	35	27	10	142	30
Sep-00	28	42	35	27	10	142	30

4.3.5 Predicted N-glycosylation sites

The number and location of the predicted N-glycosylation sites within the envelope glycoprotein were investigated using a web base program N-GLYCOSITE (www.hiv.lanl.gov) (Table 4.2). The number of sites within a patient varied slightly during disease progression, with a median number of 29 predicted N-glycosylation sites within a patient. The position of these glycosylation sites was mostly conserved in the V3 and gp41 region, with most of the shifting glycans observed in V1-V2 and V4-V5. The most noteworthy differences were seen in Du151 and Du179 with glycans in the gp41 region, which were not observed in the other patients (Figure 4.5).

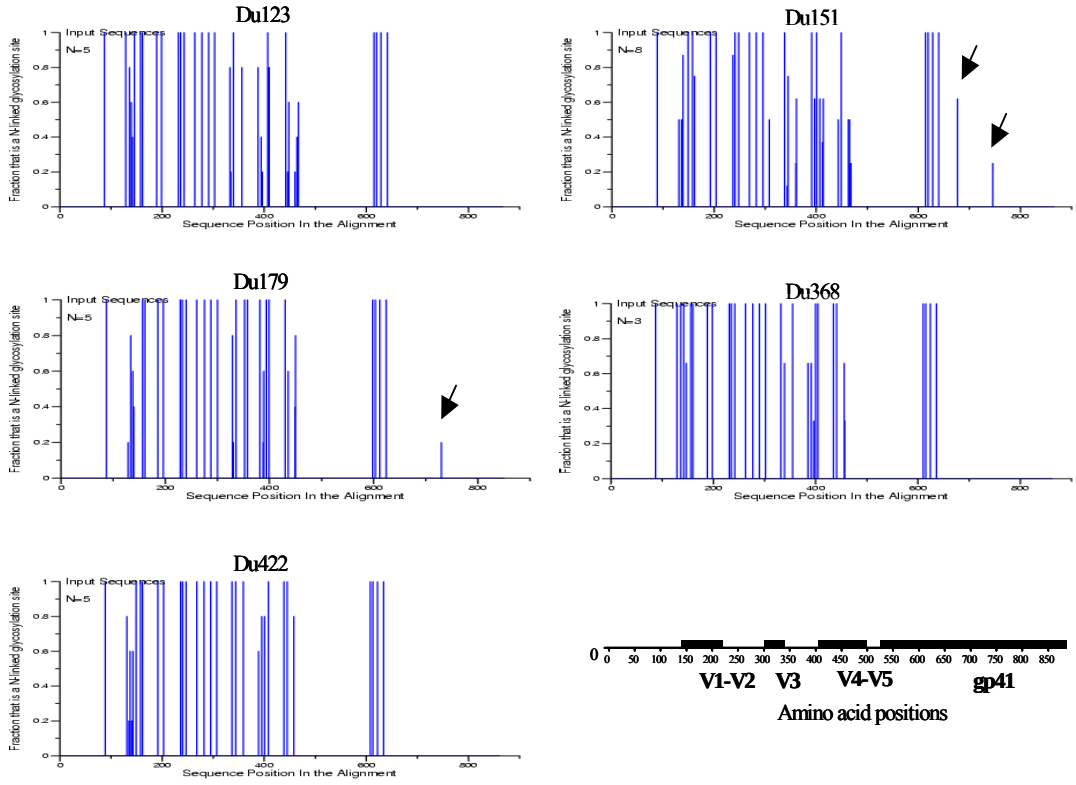


Figure 4.5: Frequency and position of N-glycosylation sites within HIV-1 subtype C patients over time. Arrows in Du151 and Du179 indicated predicted N-glycosylation sites in gp41 not observed in the other patients. Conserved glycosylation sites have a fraction of 1, and shifting glycan positions have variable fractions, as not all samples within a patient were glycosylated at that specific site.

4.3.6 V3 region and coreceptor usage

Amino acid variation and net amino acid charge within the V3 region was correlated with isolate biotype (Figure 4.6). The R5 isolates (Du123, Du368 and Du422) had very few amino acid changes per time point compared to the consensus subtype C sequence, with a low net V3 amino acid charge (between 2.5 and 4.5). The CXCR4-using samples from Du151 and Du179 had various changes to the consensus subtype C V3 sequence. Most of these amino acid variations were due to substitutions that resulted in positively charged amino acids and increased net V3 charge. These two patients also had changes in the crown motif at later time points, as well as loss of potential glycosylation site with change in coreceptor usage.

	Biotype	1										11														25										35	Charge
Consensus C		C	T	R	P	N	N	N	T	R	K	S	I	R	I	G	P	G	Q	T	F	Y	A	T	G	D	I	I	G	D	I	R	Q	A	H	C	
DU123																																					
Nov-98	R5		P																						N											3.5
Mar-00	R5																								N											3.5
May-00	R5																			A					N											3.5
Sep-00	R5		I																						N											3.5
Nov-00	R5																								N											3.5
DU151																																					
Oct-98	R5																								A				N			E				4.5
Nov-98	R5																								E				N			E				3.5
Jun-99	R5												M							S					E											2.5
Mar-00	R5X4						S		K		R	G	V						L	S		F			R	K											8.5
May-00	R5X4						S		K		R	G	V						L	A		F			R	K											8.5
DU179																																					
Mar-99	R5X4					G														A			-		N	H									Y		5
May-99	R5X4					G														A			-		N	H											4.5
Feb-00	R5X4					G		K			R									A			-		T	N	T									S		5
May-00	X4					G		K		I	R				L					A			-		N	K	-	-								S		6
Oct-03	X4					G		K		I	R								R				-		N	K	-	-								Y		7
DU368																																					
Nov-98	R5		I																						N	A											4.5
Feb-00	R5											G		G											N	A											3.5
Aug-00	R5																								N	A											4.5
DU422																																					
Jan-99	R5												V												E							E				2.5
May-99	R5												V																			E				2.5
Mar-00	R5												V																			E				2.5
May-00	R5																												N			E				3.5
Sep-00	R5																												N			E				3.5

(-) Indicates a deletion

Figure 4.6: Changes in the V3 amino acid sequence alignment correlated to isolate biotype and amino acid charge. Sequences were aligned with a subtype C consensus sequence, position 11 and 25, as well as the crown motifs have been highlighted.

4.3.7 Monoclonal antibody recognition sites

Epitope sites recognised by the monoclonal antibodies 2G12, 2F5 and 4E10 were investigated to determine if any sequence changes occured at these sites with disease progression, which might influence the neutralisation sensitivity of these viruses. 2G12 is directed at the outer region of gp120 and depends on glycans at residues in positions N295, N332, N339, N386 and N392 (according to HXB2 numbering) (Moulard, Phogat *et al*. 2002; Sanders, Venturi *et al*. 2002; Scanlan, Pantophlet *et al*. 2002). 2F5 recognise the conserved sequence ELDKWA (Muster, Steindl *et al*. 1993; Zwick, Jensen *et al*. 2005), and 4E10 (close to the 2F5 epitope) recognises the sequence NWF(D/N)IT) (Stiegler, Kunert *et al*. 2001). Previous studies have shown that the initial isolate from these five patients were resistant to 2G12 and 2F5, and sensitive to 4E10 [(Bures, Morris *et al.* 2002), Montefiori, unpublished]. This correlated with the absence (2G12 and 2F5) or presence (4E10) of these epitopes (Gray, unpublished). Analysis of the sequences from later isolates in this study (see Figure 4.3) showed that these epitopes did not change with disease progression in four patients, suggesting that the sensitivity to the MAbs remain unchanged. However, Du151 showed a change in the 2F5 epitope over time that would predict neutralising sensitivity to this Mab.

4.4 DISCUSSION

In this study the genetic changes within the gp160 region of five patients followed longitudinally for two to four years were investigated and correlated to changes in viral phenotype. This analysis showed that although there were variations within the gp160 associated with disease progression, changes within regions other than the V3 could not be linked to coreceptor switching.

Phylogenetic analysis indicated that all the samples were subtype C, which is the most predominant subtype is South Africa. Previous studies have shown that Du151 and Du179 were dually infected and Du151 experienced rapid disease progression (Gottlieb, Nickle *et al*. 2004; Grobler, Gray *et al*. 2004). Interestingly, viruses from these patients also had the ability to use CXCR4 as a coreceptor. The correlation between dual infection and ability to use CXCR4 is unclear, but both factors influence disease progression. Although CXCR4 usage is not necessary for disease progression (de Roda Husman, van Rij *et al*. 1999), the ability of the virus to use this coreceptor is more commonly seen is late disease (Richman and Bozzette 1994; Connor, Sheridan *et al*. 1997; Scarlatti, Tresoldi *et al*. 1997). In this study, dual infection in Du151 was confirmed in both the gp120 and gp41 phylogenetic trees. The OCT98 and NOV98 samples clustered together (both CCR5-using), but separate from the MAR00 and MAY00 sample clustering (CXCR4 usage), indicative of infection with two genetically different viruses. Clustering of the JUN99 sample was suggestive of recombination between the two infecting strains as was seen in the recombination analysis. This sample had the most sequence similarity with the CCR5 samples (OCT98 and NOV98) and the breakpoint was determined at amino acid position 680 within the gp41 region. This similarity was also reflected in the viral phenotype of

JUN99 that was CCR5-using. Although Du179 was reported as dually infected, all isolates in this study grouped together in both trees. This discrepancy might be due to the population based sequencing of cultured isolates used in this study, whereas the previous study was based on molecular clones derived from plasma of a smaller region (C2-C3) (Grobler, Gray *et al*. 2004). When sequences from this smaller region were used in phylogenetic analysis, we did not detect dual infection either. This might suggest that the virus from the second infection was lost during co-culture.

The variable loops play an important role in viral binding and entry, and are constantly changing to escape the immune response. It has been observed that viruses with shorter V1-V4 are preferentially transmitted or grow out in recipients, and are more sensitive to antibody neutralisation (Derdeyn, Decker *et al*. 2004). This might be because shorter loops influence the orientation and packaging of glycans resulting in increased exposure of the CD4 binding domain. The number and sites of glycans also play an important role in viral adaptation. A glycan can mask epitopes or a loss of glycan can result in tighter packaging of the envelope protein, thereby preventing recognition by neutralising antibodies (Ye, Si *et al*. 2000). Thus changes in length of variable loops, as well as the number and sites of glycans are strategies that the virus uses to escape the host immune responses. We found increases in length of the variable loop during disease progression in two patients. Both patients were dually infected and provided the opportunity for recombination between the different strains, as was seen with patient Du151. HIV-1 is known for its high recombination rate (at least 2.8 crossovers during each cycle of replication) (Zhuang, Jetzt *et al*. 2002). Recombination is a method to increase genetic diversity during immune pressure and this process might be used to increase the length of variable loops during a coreceptor

switch, ensuring survival of these variants. The number of glycosylation sites did not vary significantly within or between patients and ranged from 27 - 31 within the gp160 region. The V1-V2 length from transmitted viruses could not be compared between donor and recipient in this study, but patients infected for <6 months (Du123, Du151 and Du422) did not have significantly shorter V1-V2 loops when compared to isolates later in infection. There was also no difference in the number of predicted N-glycosylation sites in samples from patients infected less than 6 months compared to samples from infections of more than 2 years.

Both Du151 and Du179 changed coreceptor usage over time. This was associated with changes in the V3 loop including increase in charge, amino acid substitutions, loss of the glycosylation site and changes within the crown as they acquired the ability to use CXCR4 either together with CCR5 (Du151) or exclusively (Du179). This supports observations in Chapters 2 and 3 showing that such changes occur in R5X4 and X4 envelopes and probably facilitates gp120 interaction with the CXCR4 coreceptor. However in both cases neither showed an increase in V3 length, which is often typical of CXCR4-using isolates (see Chapter 2). Both viruses did however show increases in length in other variable loops. For Du151 there were an additional 19 amino acids mostly in the V1, V4 and V5 loops while Du179 had an additional 13 amino acids almost all in V1. Furthermore only Du151 and Du179 had glycosylation sites in gp41. Whether or not these features are involved in CXCR4 usage remains to be explored by analysis of gp160 genes of other CXCR4-using viruses and by site-directed mutagenesis.

A previous study determined the sensitivity of the early isolates from Du123, Du151, Du179 and Du422 to the monoclonal antibodies (MAbs) 2G12, 2F5 and 4E10 (Bures, Morris *et al*. 2002). 2G12 and 2F5 were found to be ineffective against these isolates. In this study, sequence analysis of later isolates showed similar results with all samples lacking the glycosylation site at amino acid 295 that is important for recognition by 2G12 (Binley, Wrin *et al*. 2004). Similarly 2F5 is also likely to be ineffective against later viruses due to the natural polymorphism associated with subtype C (Binley, Wrin *et al*. 2004). There were no changes in the epitope for 4E10 and all isolates would be expected to be sensitive to this MAb.

In conclusion, there were changes in length of the variable loops from samples over time and these changes were associated with patients that were dually infected and switched coreceptor usage. There were no significant changes in the number of predicted N-glycosylated sites within a patient over time. No genetic changes that could result in loss of sensitivity to these MAbs were observed with disease progression. An increase in the net V3 amino acid charge were associated with the ability to use CXCR4 as a coreceptor by HIV-1 subtype C, but the impact of other regions remains to be explored. Thus, coreceptor switching was largely influenced by changes in the V3 region and therefore used in further analysis to investigate viral tropism of HIV subtype C isolates.

CHAPTER 5

MOLECULAR AND BIOLOGICAL HETEROGENEITY IN SEQUENTIAL HIV-1 ISOLATES FROM A PATIENT THAT ACQUIRED THE ABILITY TO USE CXCR4

5.1 INTRODUCTION

HIV-1 viral diversity within an individual increases with time and can result in the appearance of viruses with different biological phenotypes (Delwart, Pan *et al.* 1997; Shankarappa, Margolick *et al.* 1999). The evolution of viral quasispecies from R5 to X4 has been described previously, with R5X4 variants seen as an intermediate between these phenotypes (Doranz, Rucker *et al.* 1996). Due to the rare occurance of CXCR4-using isolates described within HIV-1 subtype C, the sequential development of CXCR4 usage within a single individual has not previously been described.

Although the V3 region is the major determinant of viral tropism, other regions such as V1-V2 and V4-V5 have also been implicated (Koito, Harrowe *et al.* 1994; Carrillo and Ratner 1996). Changes within the variable loops associated with phenotype switching included increased length, loss of potential N-glycosylation sites and positively charged amino acids. However, it has been shown that an increase in charge of the V3 region is not sufficient for CXCR4 utilisation, but that changes in V1-V2 as well as the loss of an N-glycosylation site in V3 is also necessary (Pollakis, Kang *et al.* 2001; Nabatov, Pollakis *et al.* 2004). In the absence of the loss of this glycosylation site, changes within the V1-V2 region resulted in R5X4 tropism. Similarly deletion of charged amino acids, loss of glycosylation sites and specific amino acid substitutions in the V4-V5 region have also been shown to impact on coreceptor usage (Carrillo and Ratner 1996; Smyth, Yi *et al.* 1998; Hu, Barry *et al.* 2000).

The aim of this study was to investigate the variants present within a patient that had switched coreceptor usage from CCR5 to CXCR4 during disease progression. Biological and molecular clones were generated and clones were screened using a V3-HTA. The V1-V5 region of selected clones were sequenced and analysed to determine which genetic changes were associated with coreceptor switching. We also hypothesised whether the X4 viruses present were evolutionary variants of the R5 viruses with R5X4 as the intermediates.

5.2 MATERIALS AND METHODS

5.2.1 Viral isolation and coreceptor usage

The patient (TM18) was part of a cohort of perinatally infected children who had survived for >4 years and was classified as a slow progressor. This patient received no anti-retroviral therapy, developed advanced AIDS and has subsequently died. Three samples were obtained (TM18 A-C) at one-year intervals during 1999 - 2002. Ethical clearance was obtained from the University of Witwatersrand Committee for Research on Human Subjects (see Appendix A). Levels of virus in plasma were measured using the Versant HIV-1 RNA 3.0 assay (bDNA from Bayer Nucleic Acid Diagnostics) and CD4 counts were determined using a FACS count (Becton Dickinson, San Jose, CA). Viral isolates were made from all three time points using phytohemagglutinin (PHA) and IL-2 stimulated PBMC and tested for their ability to replicate in U87.CD4 cells transfected with either CCR5 or CXCR4, as previously described (Morris, Cilliers *et al.* 2001; Cilliers, Nhlapo *et al.* 2003).

5.2.2 Generating biological clones

Biological clones of HIV-1 isolates were generated by limiting dilution as described (Schuitemaker, Koot *et al.* 1992; Berger 1997). Briefly, 24 wells of a 96 well plate at each dilution of 1:200, 1:400 and 1:800 were established in complete medium and co-cultured with PHA stimulated healthy donor PBMC at 2 X 10^6 cells/plate and incubated at 37ºC for two weeks. Culture supernatants were tested weekly using an in-house p24 antigen assay (Cilliers, Nhlapo *et al*. 2003). Cells were assumed to be clonal if fewer than 37% of wells were positive (i.e. 8/24 wells). These clones were expanded in 12 well plates with PBMC and tested for their ability to replicate in

U87.CD4 cells transfected with either CCR5 or CXCR4, mentioned earlier. Biotype was assigned based on whether or not the clone grew in the CCR5 (R5) or CXCR4 (X4) cell lines. Clones able to use both coreceptors with comparable efficiencies or within 10% of the major coreceptor were considered dualtropic (R5X4) (Berger, Doms *et al.* 1998).

5.2.3 Amplification of the V1-V5 region

Viral RNA was extracted from plasma or cultured supernatant from the biological clones, using a MagnaPure LC Isolation station and the Total Nucleic Acid isolation kit (Roche Applied Science, Penzberg, Germany). The V1-V5 region was amplified using primers ED5 (5'-ATG GGA TCA AAG CCT AAA GCC ATG TG-3') and ES8 (5'-CAC TTC TCC AAT TGT CCC TCA-3'). Briefly, RNA was reverse transcribed with Reverse Transcriptase AMV (Roche Molecular Biochemicals) at 42ºC for 60 minutes. Amplification followed with Super-Therm Polymerase (Southern Cross Biotechnologies) for 30 cycles of 94ºC 1 minute, 55ºC 45 seconds and 72ºC for 1 minute.

5.2.4 Generating molecular clones

PCR products (V1-V5 region) from the plasma samples were purified using the High Pure PCR Product Purification kit (Roche Diagnostics GmbH, Mannheim, Germany) and cloned into the pGEMTeasy vector (Promega, USA). Molecular clones were screened by V3-HTA (as described in Chapter 2).

5.2.5 Sequencing and analysis

The V1-V5 region of the selected molecular and biological clones was sequenced using an ABI PRISM 3100 genetic analyzer with ABI PRISM BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). Sequences were aligned with ClustalX and predicted protein translations were performed using BioEdit. Phylogenetic analysis and genetic distances were determined using MEGA (version 2.1; Molecular evolution Genetic Analysis).

5.2.6 Phenotype prediction

Viral phenotypes of the molecular and biological clones were predicted from the sequenced V3 region using various prediction methods: V3 charge, 11/25 rule (Fouchier, Brouwer *et al.* 1995), observed changes in the crown motif (Coetzer, Cilliers *et al.* 2005) and C-PSSM (Jensen, Coetzer *et al.* 2005). A sample was then predicted as r5 or x4 based on the results from these four methods. However, the C-PSSM score was used as the primary method for phenotype prediction. Those with intermediate C-PSSM scores were considered to have an r5x4 phenotype. To differentiate between predicted and known phenotypes, 'R5' (in capital letters) depicts experimentally determined phenotypes and 'r5' indicates predicted phenotypes.

5.3 RESULTS

An HIV-1 infected child (TM18), followed for two years who underwent a switch in coreceptor usage from CCR5 to CXCR4 was investigated in this study. Three samples were obtained at one-year intervals and during this time both the CD4 and viral load decreased (Table 5.1). The coreceptor usage was determined for the three viral isolates, which showed a coreceptor switch over the two-year period from R5 to R5X4 and to X4.

Table 5.1: Clinical information of patient TM18 followed for 2 years.

Samples	Age (years, months)	CD4 count	CD4%	CD4:CD8 ratio	Viral Load	Biotype
TM18A	5.0	1 239	20	0.31	699 740	R5
TM18B	6.6	202	11	0.13	500 000	R5X4
TM18C	7.4	7	1	0.01	177 797	X4

5.3.1 Coreceptor determination of biological clones

Biological clones were generated from TM18B and TM18C isolates and coreceptor usage was determined (Table 5.2). For time point B all 10 biological clones used both CCR5 and CXCR4, five of them efficiently. Thirty-three biological clones were generated from time point C (14 R5, 10 R5X4 and 9 X4 viruses). The majority of the R5X4 viruses could use both receptors efficiently. Twenty-seven clones (8 R5X4 from TM18B, 8 R5, 4 R5X4 and 7 X4 from TM18C) were randomly selected for further investigation (asterisked in Table 5.2).

Table 5.2: Biological clones of TM18B and TM18C and the experimentally determined coreceptor usage. Clones selected for further analysis are indicated with (*).

Clone	U87-CCR5 (p24 ng/ml)	U87-CXCR4 (p24 ng/ml)	Biotype
TM18B (n=10)			
4*	6.7	57.5	R5X4
5	22.2	16.6	R5X4
7*	1.5	0.56	R5X4
8*	70.1	77.5	R5X4
9	13.8	15.7	R5X4
11*	42.2	42.9	R5X4
12*	16.7	5.6	R5X4
13*	33.9	2.1	R5X4
14*	13.9	4.1	R5X4
15*	13.9	24.8	R5X4
TM18C (n=33)			
7*	>100	<0.3	R5
8*	>100	1.3	R5
14*	>100	<0.3	R5
17*	>100	<0.3	R5
18*	>100	<0.3	R5
23*	>100	<0.3	R5
24*	73.9	<0.3	R5
27*	>100	<0.3	R5
29	>100	<0.3	R5
30	>100	<0.3	R5
31	5.3	<0.3	R5
33	>100	<0.3	R5
35	>100	<0.3	R5
36	>100	<0.3	R5
9	11.1	>100	R5X4
10*	45.8	>100	R5X4
11*	5.6	26.7	R5X4
15	>100	>100	R5X4
20	39.8	>100	R5X4
25	10.8	>100	R5X4
21*	14.9	37.5	R5X4
28*	24.8	>100	R5X4
34	>100	12.2	R5X4
39	26.6	>100	R5X4
1*	1.4	10.4	X4
2	0.5	89.2	X4
4	<0.3	19.6	X4
12*	1.5	60.2	X4
16*	0.3	7.9	X4
19*	3.5	>100	X4
26*	0.5	4.7	X4
37*	0.8	6.8	X4
38*	7.7	>100	X4

5.3.2 True coreceptor usage versus predicted phenotype

The PSSM score for each of the biological clones was calculated and compared to the experimentally determined coreceptor usage (Figure 5.1). All 27 biological clones were predicted to use CXCR4 as a coreceptor. The eight R5 viruses (from TM18C) had intermediate C-PSSM scores between (-27.45 to -15.55). This suggested that these late stage R5 viruses might be similar in the V3 region to CXCR4-using viruses or that regions outside of V3 might be involved in CXCR4 usage. The 12 R5X4 viruses (from time points B and C) were predicted as x4 and in some cases had higher C-PSSM scores than the X4 clones. The C-PSSM failed to predict the r5x4 as the method was not optimised for this. The seven X4 viruses from time point C were correctly predicted by the C-PSSM. Overall, 70% of the biological clones were correctly predicted as CXCR4-using

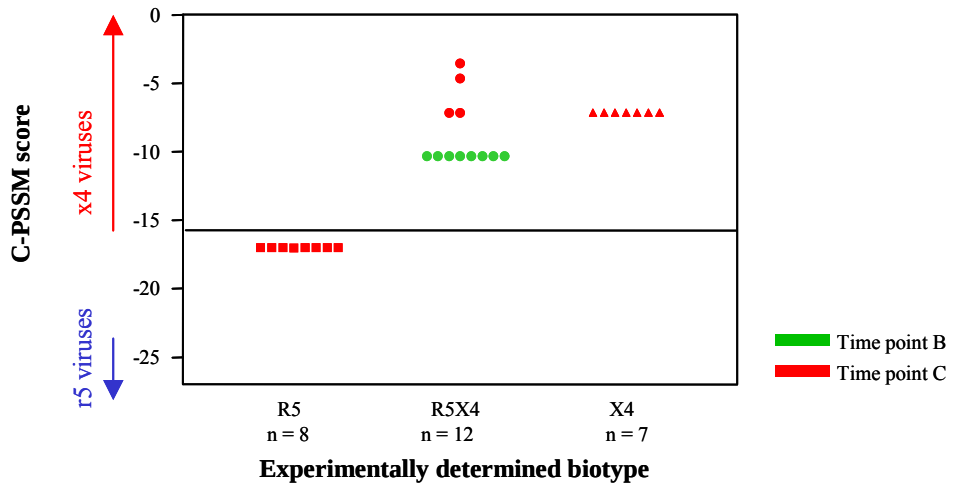


Figure 5.1: Comparison between experimentally determined and predicted phenotypes (PSSM score) of the biological clones. Samples with C-PSSM scores >-15.55 were predicted as x4.

5.3.3 Phenotype prediction of molecular clones

In order to access the *in vivo* genetic diversity a V3-HTA was performed on plasma samples from the three time points. The V3-HTA profiles showed a large degree of genetic diversity over the two-year period. Three variants were present at time point A. Time points B and C shared a variant with A (slowest migrating band indicated by arrow in Figure 5.2A). Time point B also had three variants at different mobilities, whereas time point C had only a single variant. Molecular clones were made from each sample representing the variants present *in vivo* over time. Using a V3-HTA, 44 molecular clones (20 from TM18A, 16 from TM18B and 8 from TM18C) were selected for further analysis. Results from prediction methods based on the V3 region, indicated that the 16 TM18A clones (from lowest two bands) were r5 and the slowest migrating band (shown with arrow in Figure 5.2A) presented r5x4 (3 clones) and x4 (1 clone) variants. Similarly, this top band in the TM18B clones was predicted as either r5x4 (4 clones) or x4 (2 clones). The other 10 clones representing the lower bands were r5. All eight clones from time point C were predicted as CXCR4 using.

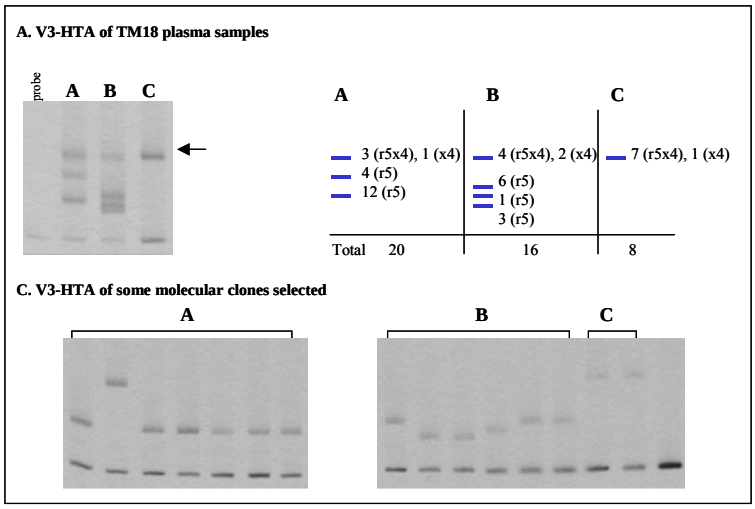


Figure 5.2: Selection of molecular clones within TM18 A-C. (**A**) V3-HTA of variants present in plasma of TM18 A-C (arrow depicts slowest migrating bands). (**B**) Schematic representation of plasma variants (blue bands) and number of clones selected with predicted phenotype. (**C**) An example of V3-HTA of molecular clones selected for further analysis.

5.3.4 Distribution of phenotypes at different time points

The distributions of phenotypes at the three time points were compared between molecular and biological clones (Figure 5.3). Only molecular clones were available for TM18A and were mostly r5 with a few r5x4 and x4 variants. The fact that CXCR4-using variants were not present in the original viral isolate from TM18A (Table 5.1) suggests they may not be highly infectious or were outgrown by the R5 variants in culture. Most of the molecular clones at time point B were predicted as r5 with a few r5x4 and x4 viruses. However, all the biological clones at this time point were dualtropic, concordant with the biotype of the original viral isolate (Table 5.1). Time point C contained CXCR4-using molecular and biological clones, but the R5 variants were only observed *in vitro*. The molecular clones were largely dualtropic, whereas the biological clones within this time point were more diverse despite the fact that only X4 variants were detected in the original viral isolate.

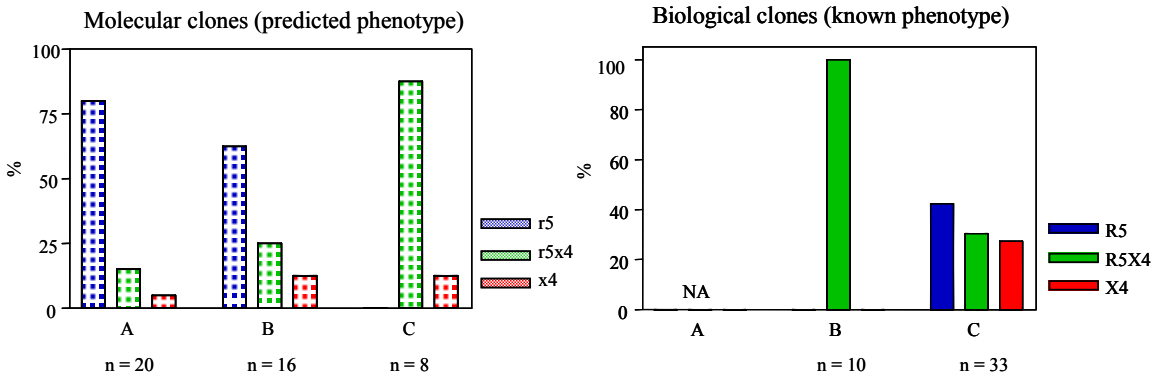


Figure 5.3: Distribution of phenotypes within the molecular and biological clones at various time points. (NA - not available).

5.3.5 Sequence analysis of the V3 region

The V3 sequences from all 44 molecular and 27 biological clones were aligned and identical sequences grouped to give 18 representative V3 sequences from the different time points. These sequences thus represented all the variants present within TM18 over time. The sequences were labelled as a sequence type (1 - 18) and consisted of either molecular or biological clones from a specific time point, except sequence 8 that represented sequences from both time points A and C, as well as 15 and 17 that contained both molecular and biological clones.

Phenotypes were predicted for these 18 sequences based on the V3 region using various methods such as the V3 charge, 11/25 rule, crown motif and C-PSSM (Table 5.3). The 18 sequences were listed based on their C-PSSM score and showed increased amino acid changes compared to the subtype C CCR5 consensus. Sequences 1 - 14 had a low amino acid charge (+3 or +4) and sequences 15 - 18 had a much higher charge (+6 and +7). Sequences 14 - 18 had an R at position 11, but no changes at position 25. Variations of the crown motif (GPGQ) associated with CXCR4 usage were seen in sequences 7 - 18. Sequences 1 - 3 were predicted as r5 and sequences 14 - 18 as x4 using the C-PSSM. Sequences 4 - 13 could not be accurately predicted with the C-PSSM due to the transitional score (highlighted in grey in Table 5.3) and were considered r5x4. Phenotypes were assigned when three or four of the prediction methods concurred. Thus, variants similar to sequences 1 - 6 were CCR5-using (r5), sequences 7 - 11 CCR5- and CXCR4 using (r5x4) and sequences 12 - 18 CXCR4-using (x4).

Table 5.3: Predicted phenotype of 18 V3 sequences representing identical molecular and biological clones. Sequences were aligned with a subtype C CCR5 sequence and listed according to their PSSM score, which was associated with increased ability to use CXCR4 as coreceptor. Predictions shown in red are associated with CXCR4 usage.

| Sequence type | Nr of identical clones | Time point of clones | Origin of clones+ | 1 | | | | | | | | | | 11 | | | | | | | | | | | | | | | | | | | | | | | 25 | | | | | | | | | | | | | | | | | | | | 35 | Prediction methods | | | | Predicted phenotype | |
|---|---|---

5.3.6 Evolution of CXCR4 usage within the V3 loop

The 18 representative sequences were used to draw a schematic phylogenetic tree to determine the genetic relationships between these different V3 sequences (Figure 5.4). Variation in the different time points, predicted and known phenotype, amino acid charge, 11/25 rule and changes within the crown motif are also shown. The tree topology suggested two main lineages: lineage A consisted of sequence types 1 - 5 representing molecular clones from TM18A and TM18B. They were characterised by a GPGQ crown motif, V3 amino acid charge of +3 and S/D at position 11 and 25, typically seen in CCR5 using isolate. In lineage B, consensus sequence 6 was of particular interest as this r5 predicted sequence might have provided the basis from which CXCR4-using strains evolved. Variants from this consensus were present in TM18A and had all the characteristics of CCR5 usage. In this lineage there were two routes to developing CXCR4 usage (indicated as B-1 and B-2 on tree). The first (B-1) was associated with increased V3 charge (to +7) and change in the crown motif to GPGY. These changes resulted in dualtropic viruses but also X4 viruses as shown within sequence 16 (that contained X4 biological clones), thus suggesting that changes within the V3 region were sufficient to confer CXCR4 usage. The second possibility of development towards CXCR4 usage (B-2) contained r5x4 predicted sequences. These consensus sequences were associated with various crown changes, but no increase in charge, and this might have been the main obstacle in preventing exclusive CXCR4 usage. The PSSM score of these predicted dualtropics increased with time (as reflected in the colour of the branches), with the sequences closer to the root more r5-like and moving towards the tips of the tree more x4-like. Although sequence 12 and 13 were predicted r5x4, biologically

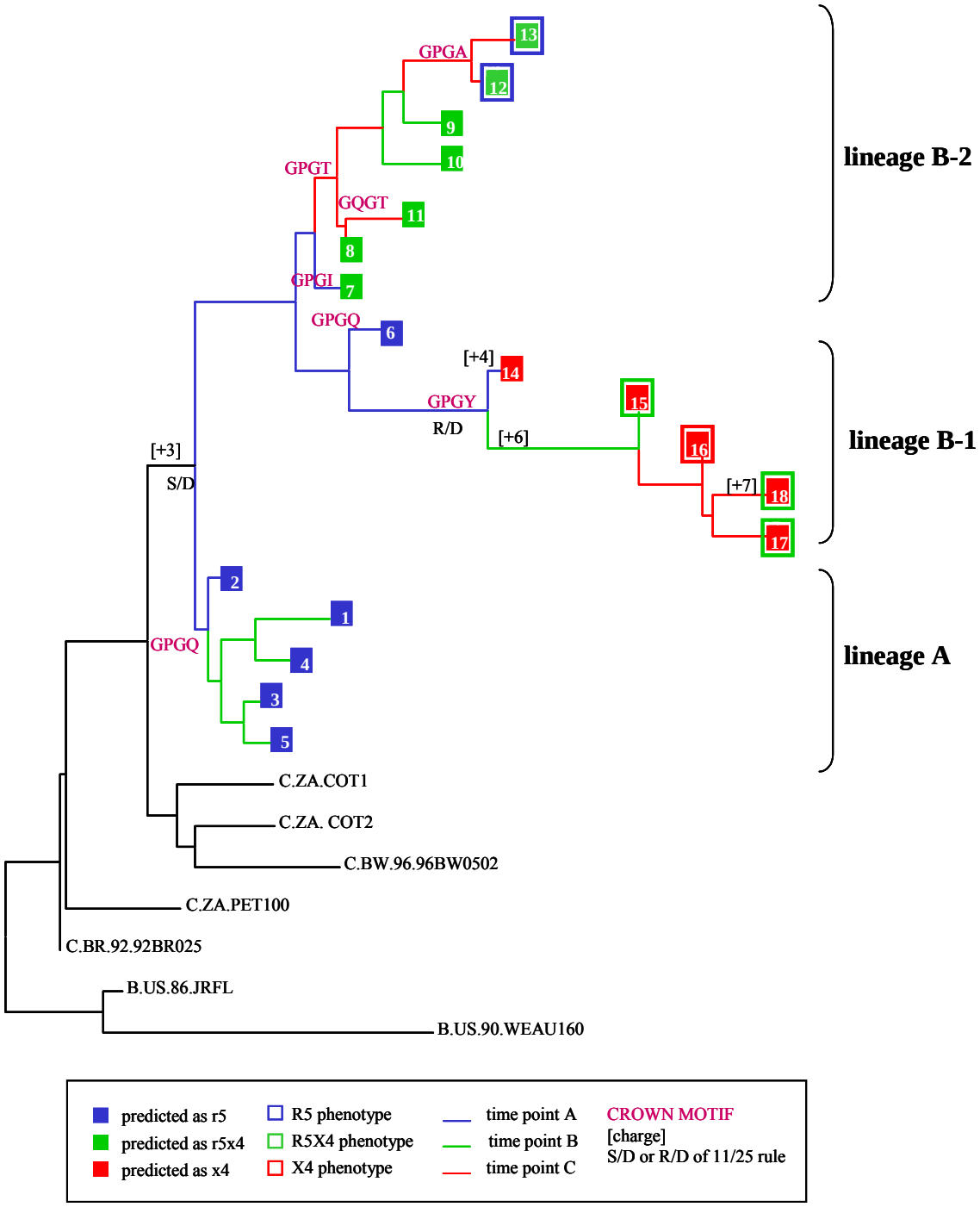


Figure 5.4: Schematic phylogenetic tree showing the evolution of CXCR4-using viruses. Evolutionary relationships between the 18 consensus sequences were used. Two lineages are shown, with lineage A R5 viruses and lineage B indicating two possible methods of CXCR4 development. Lineage B-1 switches from R5X4 to X4, whereas lineage B-2 remains mostly dualtropic.

they were shown to be R5. This suggested that although the V3 regions were possibly able to use CXCR4, other regions might have played a role in the preferential use of CCR5 *in vitro* of these late stage variants as has been shown previously.

5.3.7 Other regions contributing to coreceptor usage

While the V3 region is the major determinant of viral phenotype, various studies have indicated that other regions within the envelope contribute to phenotype. We therefore investigated the number of amino acids, the charge of the variable loops and number of predicted N-glycosylation sites within the V1-V2 and V4-V5 regions (Table 5.4). The V1-V2 length increased in CXCR4-using viruses, but were even longer in dualtropic viruses. The opposite was seen in the V4-V5 loop that shortened as viruses switched to X4. The median V1-V2 amino acid charge in R5 viruses were +4 but decreased to +1 in CXCR4-using viruses, whereas the V4-V5 loop charge stayed relatively unchanged. The number of glycosylation sites within V1-V2 and V4-V5 remained constant with more glycans in V1-V2. Phylogenetic trees were constructed for the V1-V2 and V4-V5 sequences showing similar clustering of CCR5- and CXCR4-using viruses as in the V3 tree (data not shown). These data suggest that the transition from R5 to X4 may require an increase in V1-V2 and a decrease in V4-V5 lengths.

Table 5.4: Variations within V1-V2 and V4-V5 that may contribute to phenotype switching of HIV-1 subtype C. These variations include the number of amino acids, amino acid charge and number of glycosylation sites.

Phenotype	Number of amino acids				Amino acid charge		Number of N-glycosylation sites	
	V1-V2	range	V4-V5	range	V1-V2	V4-V5	V1-V2	V4-V5
r5	65	61-73	84	78-87	4	1	6	4.5
r5x4	77.5	69-82	81	77-86	1	2.5	6	5
x4	73	72-85	74	74-84	1	1	6	3

5.4 DISCUSSION

The development of CXCR4 usage within a patient, that initially had CCR5-using viruses, provided the opportunity to study the evolution of X4 viruses within HIV-1 subtype C. Although this patient was infected by mother-to-child transmission, similar phenotypic changes have been reported in HIV-1 infected children as in adults (Scarlatti, Tresoldi *et al.* 1997). Furthermore, Nowak *et al.* (2002) has shown that although there are differences in the immune system between adults and children, this does not result in significant differences in the viral evolution rate as reflected within the V3 region. Thus, to determine which genetic changes were associated with coreceptor switching and whether X4 viruses were evolutionary variants of R5 viruses, the envelope V1-V5 region of molecular and biological clones from this patient were investigated.

The coreceptor switch from R5 to X4 in this patient was associated with a dramatically reduced CD4 count and viral load. Similar reductions in CD4 count have been shown in other patients undergoing coreceptor switching, suggesting that X4 viruses have increased pathogenicity (Sheppard, Celum *et al*. 2002). The loss of CD4 target cells probably account for the reduction in viral load. However there are still significant viral replication suggesting the virus gained the ability to use other target cells by broadening its coreceptor usage. Differences between the original isolate and biological clones suggested that some variants were lost or outgrown due to culture adaptations. Therefore, the molecular clones were included in the analysis as these were more representative of the circulation variants, but unfortunately lacked the experimentally determined coreceptor usage. For example, comparing the biological and molecular

clones within time point B showed that no R5 biological clones were isolated at that time point, but they were observed in the molecular clones. A possible reason for this might be that *in vivo* R5 viruses are better adapted, but *in vitro* X4 viruses are preferentially amplified (van 't Wout, Blaak *et al.* 1998; Koning, van Rij *et al.* 2002). By combining the molecular and biological clones we hoped to capture the full spectrum of variants present in this patient.

Phenotypes were predicted for the 71 clones (27 biological and 44 molecular) using various methods. These methods were developed to differentiate between r5 and x4 phenotypes based on the V3 genotype (Chapter 3). The C-PSSM was the most informative prediction method because in addition to predicting a phenotype, it also indicated sequences with intermediate scores (-27.45> x >-15.55). These intermediate scored-sequences suggested that the sample did not fit the criteria (score) of an r5 or x4 virus. It has previously been showed that samples with intermediate scores correlated well with R5X4 viruses (Jensen, Li *et al.* 2003) and therefore, in combination with the other prediction methods these variants were phenotyped as r5x4. Interestingly, the PSSM scored late stage R5 viruses (from time point C) as r5x4, indicating that although these viruses had CXCR4-using V3's, they used CCR5 as coreceptor. Genotypic characterisation of the different phenotypes suggested that there were two types of R5 viruses, isolated early in infection and from late infection. The early R5's had the typical low charge, S/D amino acids at positions 11/25 and GPGQ crown associated with CCR5 usage. The late stage R5's and X4 variants were very similar with increased charge, positive amino acid at position 11 and changes in the crown motif. The dualtropic

variants had a combination of these characteristics, such as changes in the crown, but low amino acid charge. Therefore, the ability of a virus to broaden coreceptor usage occurred in a step-wise manner, implying that an R5 virus acquires the ability to use CXCR4, while retaining the ability to use CCR5 as a coreceptor (Pantaleo, Graziosi *et al.* 1993). This is possible because HIV-1 interacts with different parts of CCR5 and CXCR4 during entry into the host cell (Rucker, Samson *et al.* 1996; Lu, Berson *et al.* 1997). Although only a few amino acid changes are needed for coreceptor switching as was shown in this study, it usually takes years, indicating that there might be obstacles to prevent phenotype switching. Pastore *et al.* (2004) suggested that this delay in switching might be due to loss of replication fitness, sub-optimal use of coreceptor and limited transitional pathways within these intermediate variants.

While R5 variants are responsible for persistent infection and found throughout disease, X4 variants are either a cause of or evolve in response to the progressive immune suppression (Schuitemaker, Koot *et al.* 1992; Tscherning, Alaeus *et al.* 1998). In this study there were two possible routes for variants to obtain the ability to use CXCR4. Firstly, changing within the crown to GPGY as well as the increase in V3 charge resulted in X4 viruses. In a second route, most variants gained the ability to use CXCR4 but could also use CCR5 (dualtropic). It has been suggested that a CCR5 variant needs certain characteristics to evolve to using CXCR4, but once it mutates beyond these characteristics, it is unable to develop the ability to use CXCR4 (van 't Wout, Blaak *et al.* 1998).

Studies have shown that other regions in the envelope contribute and/or influence coreceptor usage. This has also been observed in this study in particular with the dualtropic viruses, where some had an R5-like V3 loop, but were able to use CXCR4 in addition. Various studies have suggested that the V1-V2 region contribute to CXCR4 usage (Groenink, Fouchier *et al.* 1993; Fouchier, Broersen *et al.* 1995), the combination of changes within the V3 and V1-V2 increases the efficiency of CXCR4 usage (Cho, Lee *et al.* 1998). Others have suggested that this extension of the V2 loop might render a virus less fit or provide the basis for subsequent mutations in the V3 that are essential for phenotype switching (Fouchier, Groenink *et al*. 1992; Pastore, Ramos *et al*. 2004). Similar to these studies, the V1-V2 loop length within the X4 variants were longer than the R5 viruses but the R5X4 variants had the longest loops, further suggesting that this increased length might be necessary during the switch but also to maintain the X4 phenotype. The opposite was seen within the V4-V5 region, with shorter loops in CXCR4-using variants, possible as a compensatory response in order to maintain the overall structure of the envelope protein.

In conclusion the molecular and biological heterogeneity in sequential HIV-1 isolates from a patient that gained the ability to use CXCR4 showed that the X4 variants evolve from the R5 viruses in a step-wise manner with the R5X4 viruses intermediate in this process. Although changes within the V1-V2 region contributed, the V3 region of envelope was the major determinant of phenotype switching.

CHAPTER 6

THE IMPACT OF ACTIVE TUBERCULOSIS ON HIV-1 SUBTYPE C GENETIC DIVERSITY

6.1 INTRODUCTION

Opportunistic infections in HIV-1 infected patients are often associated with transient increases (viral bursts) in HIV-1 viral load due to immune activation (Sulkowski, Chaisson *et al.* 1998) and studies suggest that this impacts on viral diversity (Ostrowski, Krakauer *et al.* 1998; Collins, Mayanja-Kizza *et al.* 2000). Tuberculosis (TB) is the most common opportunistic infection in developing countries and is the leading cause of death among HIV positive people (www.unaids.org). Unlike most opportunistic infections, which are acute and usually occur late in HIV-1 infection, TB is a chronic condition, which can occur at all stages of HIV-1 infection. In a small study of seven patients by Goletti *et al.* (1996), active TB was shown to increase HIV-1 viral load, although this was not substantiated in a larger study that forms the basis of this investigation (Day, Grant *et al.* 2004). However, in this latter study an episode of TB did result in viral fluctuations and patients with higher viral levels were more prone to developing TB (Day, Grant *et al.* 2004). Furthermore, another study showed that the viral load increased marginally during the first month of TB treatment, but then declined to baseline levels (Morris, Martin *et al.* 2003). Collectively these data suggest that TB impact on HIV-1 replication in HIV/TB co-infected patients.

Viral bursts in patients with opportunistic infections may be due to non-specific amplification of existing populations, selected expansion of sub-populations or the emergence of new variants. In some instances this could lead to the appearance of variants with increased fitness as a result of viral escape from immune responses or

viruses with altered phenotypes. Opportunistic infections often occur late in disease where X4 variants are more likely to appear. However, since *Mycobacterium tuberculosis* infects macrophages, it is likely that active TB will promote the replication of macrophage-tropic viruses (which use CCR5). Furthermore, TB is associated with an increased pool of activated CD4 T cells that express CCR5 during co-infection, that could promote replication of these R5 viruses (Ostrowski, Krakauer *et al.* 1998). This was shown in a study where viral isolates from TB patients were found to use CCR5 and not CXCR4 (Morris, Cilliers *et al.* 2001).

The aim of this study was to determine the effect of active TB and TB treatment on HIV-1 heterogeneity. Plasma samples were obtained from HIV-1 infected patients who developed active TB and were then treated. Patients were selected if TB was the only AIDS-defining illness. Sequence analysis of viral quasispecies at different time points within an individual were performed to examine whether viral load fluctuations were associated with genetic shifts in viral populations.

6.2 MATERIALS AND METHODS

6.2.1 Patient information

Plasma samples were selected from a previously described cohort of asymptomatic HIV-1 infected individuals working at a gold-mining company in the Free State, South Africa who attended a TB Prevention Clinic between 2000-2002 (Day, Grant *et al.* 2004). Ethical clearances were obtained from the University of Witwatersrand Committee for Research on Human Subjects and London School of Hygiene and Tropical Medicine Ethics Committee (see Appendices A and B). These patients were seen routinely every six months, during which data were collected on symptoms and physical signs, as well as blood taken. Eighteen participants who had an episode of TB (and no other AIDS defining illness) were included in this study, six of whom were selected for in-depth analysis. Samples were labelled according to the weeks before or after the start of TB treatment, with pre-TB treatment samples denoted with (-), post-TB treatment labelled with (+) and the sample collected at the start of TB treatment labelled as (0). Five HIV-1 patients who did not have an episode of TB were included as controls. Levels of virus in plasma were measured using the Versant HIV-1 RNA 3.0 assay (bDNA from Bayer Nucleic Acid Diagnostics, with >0.5 logs as clinical significant difference) and CD4 counts were determined using a FACScount (Becton Dickinson, San Jose, CA).

6.2.2 Viral RNA isolation and RT-PCR

Viral RNA was extracted from plasma using a MagnaPure LC Isolation station and the Total Nucleic Acid isolation kit (Roche Applied Science, Penzberg, Germany). The

C2V3 *env* region was amplified using primers as described (Grobler, Gray *et al.* 2004). Briefly, RNA was reverse transcribed with Reverse Transcriptase, AMV (Roche Molecular Biochemicals) and amplified with Super-Therm Polymerase (Southern Cross Biotechnologies), using primers BF (5'-TAA CAC AAG CCT GTC CAA AGG-3') and BR (5'-AAT TCT AGG TCC CCT CCT GA-3'). PCR products were purified using the High Pure PCR Product Purification kit (Roche Diagnostics GmbH, Mannheim, Germany). Population-based PCR products were sequenced using an ABI PRISM 3100 genetic analyser with ABI PRISM BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems).

6.2.3 C2V3-HTA

A C2V3-HTA was used to screen the 18 patients with samples from various time points. HTA probe construction and labelling was done as previously described by Nelson *et al*. (1997) and Ping *et al*. (1999), using a plasmid with the C2V3 region originating from a subtype C R5 virus, very similar to the subtype C consensus (Grobler, Gray *et al.* 2004). Single stranded probe labelling was done by digesting plasmid DNA with *Bam*H1 (Amersham Pharmacia Biotech, UK), end-labelling at room temperature with a mixture containing 12.5µCi ^{35}S-dATP (Amersham Pharmacia Biotech, UK), unlabeled dGTP and Klenow DNA polymerase I (Amersham Pharmacia Biotech, UK). The probe was removed from the vector by digestion with *Spe*I (Amersham Pharmacia Biotech, UK) and purified using the High Pure PCR purification kit (Roche Diagnostics GmbH, Mannheim, Germany) into a final volume of 50 µl. Heteroduplexes were formed between the probe and PCR product in a 10 µl reaction containing 5 µl PCR product, 3 µl labelled probe, 1

µl annealing buffer (1M NaCl, 100mM Tris-HCL [pH7.5], 20mM EDTA) and 1 µM of the BR-primer denatured at 95°C for 2 minutes. The reactions were then cooled at room temperature for 10 minutes. The heteroduplexes were separated in non-denaturing 6% polyacrylamide gels as described by Grobler *et al.* (2004). Dried gels were exposed to autoradiograms (BioMax MR, Kodak).

6.2.4 Cloning

Purified PCR product was cloned into the pGEMTeasy vector (Promega, USA) and individual molecular clones were screened by C2V3-HTA to select all variants within a sample which were then sequenced as described above. Generally, 20 clones per patient were selected, except in PC0137 where 51 clones were available.

6.2.5 Subtyping

Sequences were aligned with ClustalX and predicted protein translations were performed using BioEdit. Phylogenetic analysis as well as genetic distances were determined using MEGA (version 2.1; Molecular evolution Genetic Analysis). To root the sequences, total population sequencing was performed on the 5 control patients who did not experience an episode of TB. The probe sequence of the C2V3 region (CONSENSUS C) was also included in the analysis.

6.2.6 Phenotype prediction

Viral phenotypes were predicted from the sequenced V3 region, using four prediction methods: V3 charge, 11/25 rule (Fouchier, Brouwer *et al.* 1995), observed changes in the crown motif [(Coetzer, Cilliers *et al.* 2005), Chapter 2], and the C-PSSM [(Jensen, Coetzer *et al.* 2005), Chapter 3]. A sample was predicted as r5 or x4 based on the results from the four methods. However, the C-PSSM score was used as the primary method for phenotype prediction. Those with intermediate C-PSSM scores were considered to have an r5x4 phenotype.

6.3 RESULTS

6.3.1 Characterisation of HIV-1 populations in HIV/TB patients

Samples were obtained over a 2-year study period from 18 HIV-1 infected patients at different time points during pre- and post-TB diagnosis and treatment. A single sample (close to the TB diagnosis) from each patient was selected to characterise HIV-1 present within this cohort. Population-based sequencing was performed on plasma samples in the C2V3 region to determine the genetic subtypes and perform phenotype predictions. All the samples, including the 5 control sample clustered with subtype C references (Figure 6.1), indicating that these patients were infected with HIV-1 subtype C.

The viral phenotype was predicted based on the V3 region using the following methods: V3 charge, 11/25 rule, GPGQ crown motif and C-PSSM (Table 6.1). The V3 charge was similar to other HIV-1 subtype C CCR5-using isolates (Chapter 2), with the exception of PC0074 which had a +5.5 charge, more commonly seen in CXCR4-using viruses. All samples had either negative (D/E) or neutral (A/G/S) amino acids at positions 11 and 25, except PC0074 and PC0261. These samples had an R at position 25 or 11 respectively, indicative of CXCR4-using viruses. All samples had a GPGQ crown motif, except for PC0284 that had a GPGR crown motif also more frequently found among CXCR4-using viruses. The C-PSSM program predicted one sample to be x4 (PC0261), but could not predict with confidence the phenotype of two other samples due to their intermediate scores (PC0074 and PC0284).

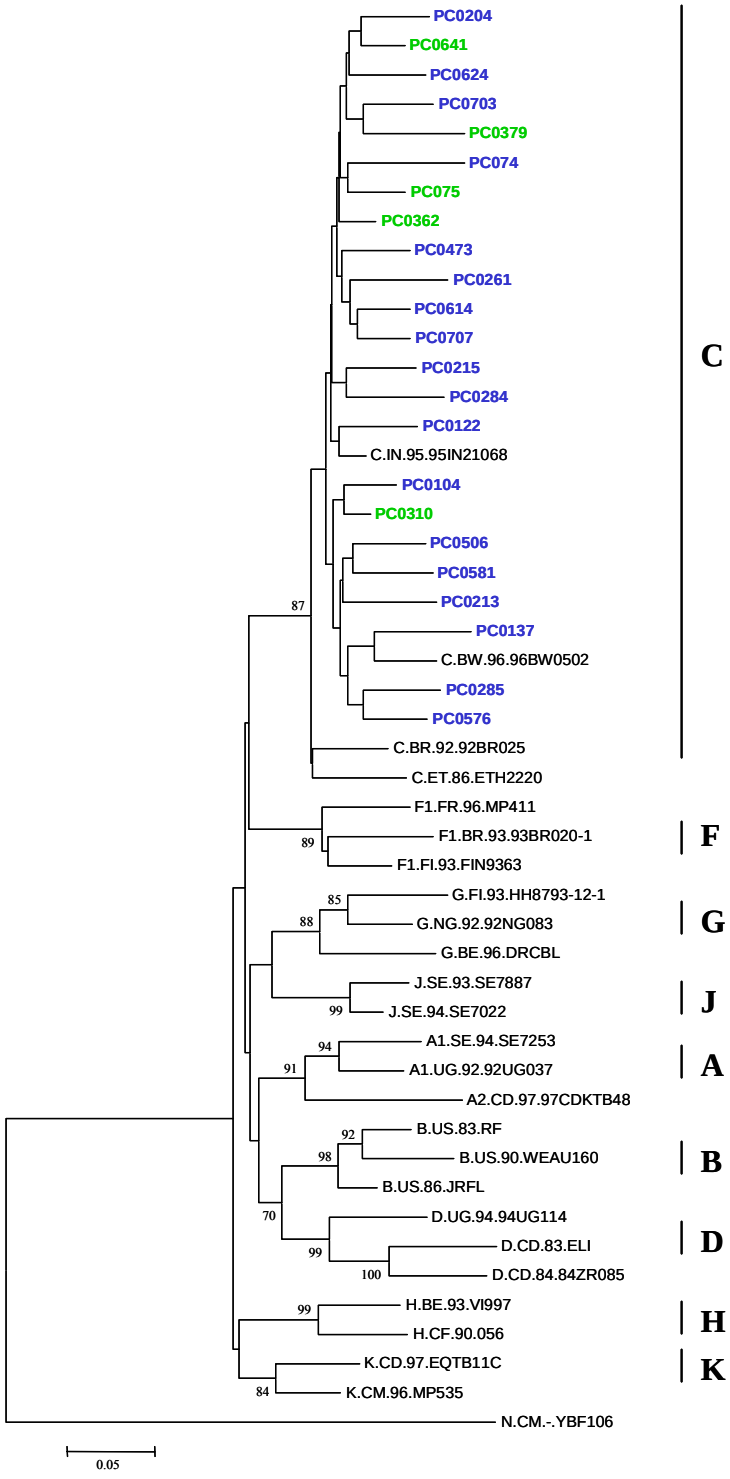


Figure 6.1: Phylogenetic analysis of 18 HIV/TB and 5 HIV infected patients. The HIV/TB patients shown in blue and HIV-1 infected control samples in green. Subtype clustering shown on right. Reference samples were obtained from the Los Alamos database (http://www.hiv.lanl.gov).

For 15 samples all four prediction methods predicted r5 variants, while three (PC0074, PC0261 and PC0284) were predicted as CXCR4-using based on two or more of the tools. PC0261 was predicted as x4 based on the C-PSSM score. However, whether or not these three samples are able to use CXCR4 as a coreceptor can only be determined experimentally.

Table 6.1: Phenotype prediction based on the V3 region of 18 HIV/TB patients. Viral phenotype was predicted using various methods based on V3 population sequencing of a sample close to the TB diagnosis. Red highlighted characteristics are associated with CXCR4 usage. Samples selected for further analysis are bolded.

Patient ID	Subtype	V3 Charge	11/25	Crown motif	C-PSSM score		Predicted phenotype	*
PC0074	C	5.5	G/R	GPGQ	(-18.93)	x	r5x4	3/4
PC0104	C	3.5	S/D	GPGQ		r5	r5	4/4
PC0122	C	3.5	S/D	GPGQ		r5	r5	4/4
PC0137	C	3	S/E	GPGQ		r5	r5	4/4
PC0204	C	3.5	S/D	GPGQ		r5	r5	4/4
PC0213	C	4.5	S/A	GPGQ		r5	r5	4/4
PC0215	C	2.5	S/D	GPGQ		r5	r5	4/4
PC0261	C	3.5	R/E	GPGQ		x4	x4	2/4
PC0284	C	3.5	S/D	GPGR	(-18.64)	x	r5x4	2/4
PC0285	C	3.5	S/D	GPGQ		r5	r5	4/4
PC0473	C	3.5	S/E	GPGQ		r5	r5	4/4
PC0506	C	3.5	S/D	GPGQ		r5	r5	4/4
PC0576	C	3.5	S/D	GPGQ		r5	r5	4/4
PC0581	C	4	S/A	GPGQ		r5	r5	4/4
PC0614	C	3.5	S/D	GPGQ		r5	r5	4/4
PC0624	C	3	S/D	GPGQ		r5	r5	4/4
PC0703	C	3	S/D	GPGQ		r5	r5	4/4
PC0707	C	3.5	S/D	GPGQ		r5	r5	4/4

C-PSSM: Scores <-27.45 were predicted as r5, scores >-15.55 were predicted as x4. Those with intermediate scores (x) (highlighted in grey) are assumed to be r5x4.
Predicted phenotype (*): phenotypes were predicted based on four methods, and the number of methods with concordant phenotype indicated.

6.3.2 Screening of HIV/TB patients for genetic diversity

The 18 patients were screened using a C2V3-HTA to investigate sample complexity within a patient (such as number of variants) present during and after a TB episode. Variation of populations within a sample was correlated to the shifts and number of bands present. The mobility ratios of these bands were classified as low or high shifts, and the number of bands as single or multiple bands (data not shown).

Seven patients had low mobility bands of which two had single populations present at all time points and five had multiple bands in one or more samples, suggesting a homogenous population not influenced by TB or TB treatment. Eleven patients had variants with higher mobility ratios and / or multiple bands, suggesting more complex variants. Six patients were selected for further analysis to characterise the variants present within these patients during an episode of TB and TB treatment.

6.3.3 Molecular clone analysis of viral populations

Six patients (PC0122, PC0137, PC0213, PC0261, PC0614 and PC0703) were selected for further analysis based on their C2V3-HTA profiles (Figure 6.2). Four patients showed expansions of variants (of these two had high mobility shifts, PC0137and PC0213). In one patient the variants did not change over time (PC0261) and in one the V3-HTA profiles suggested a selection of specific variants (PC0614).

Viral populations from at least three different time points from each patient were cloned and random clones were picked and screened with the C2V3-HTA. A median of 7 clones

was selected from each time point for each patient, representative of the different variants present in the sample. These clones were sequenced and used to calculate the genetic diversity within and between samples.

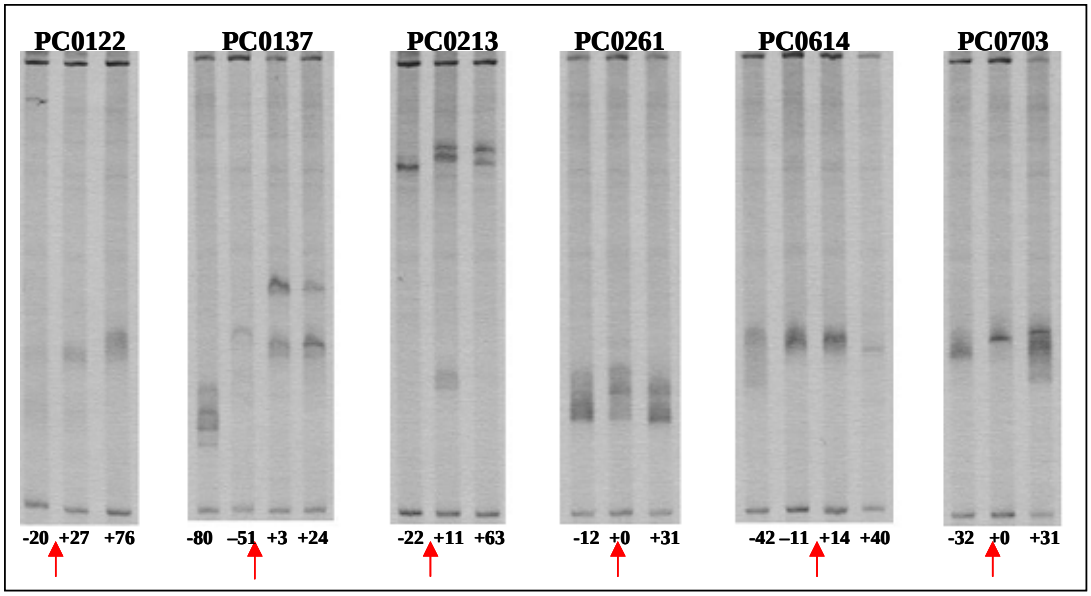


Figure 6.2: C2V3-HTA profiles of 6 HIV/TB patients over time. Weeks before and after TB treatment indicated at bottom. The red arrow depicts the start of TB treatment.

6.3.3.1 Genetic diversity within patient samples

The genetic diversity within each sample was compared to the CD4 count and viral load for each patient (Figure 6.3). In 5 of the 6 patients CD4 counts decreased and were lower compared to the CD4 counts before TB treatment. Two patients showed a decrease in viral load while four patients had higher viral loads than pre-TB treatment (PC0122, PC0213, PC0614 and PC0703). The degree of genetic diversity fluctuated over time within a patient. In four patients there was a trend towards an increase in genetic diversity, whereas two patients showed no changes (PC0122 and PC0703). Patient PC0137 showed the highest level of diversity, which was clearly evident on the C2V3-

HTA. Patients PC0213 and PC0614 had 3 - 4% diversity at the last time points. All other patients had less than 3% diversity at all the time points. There appeared to be no correlation between genetic diversity and viral load fluctuations.

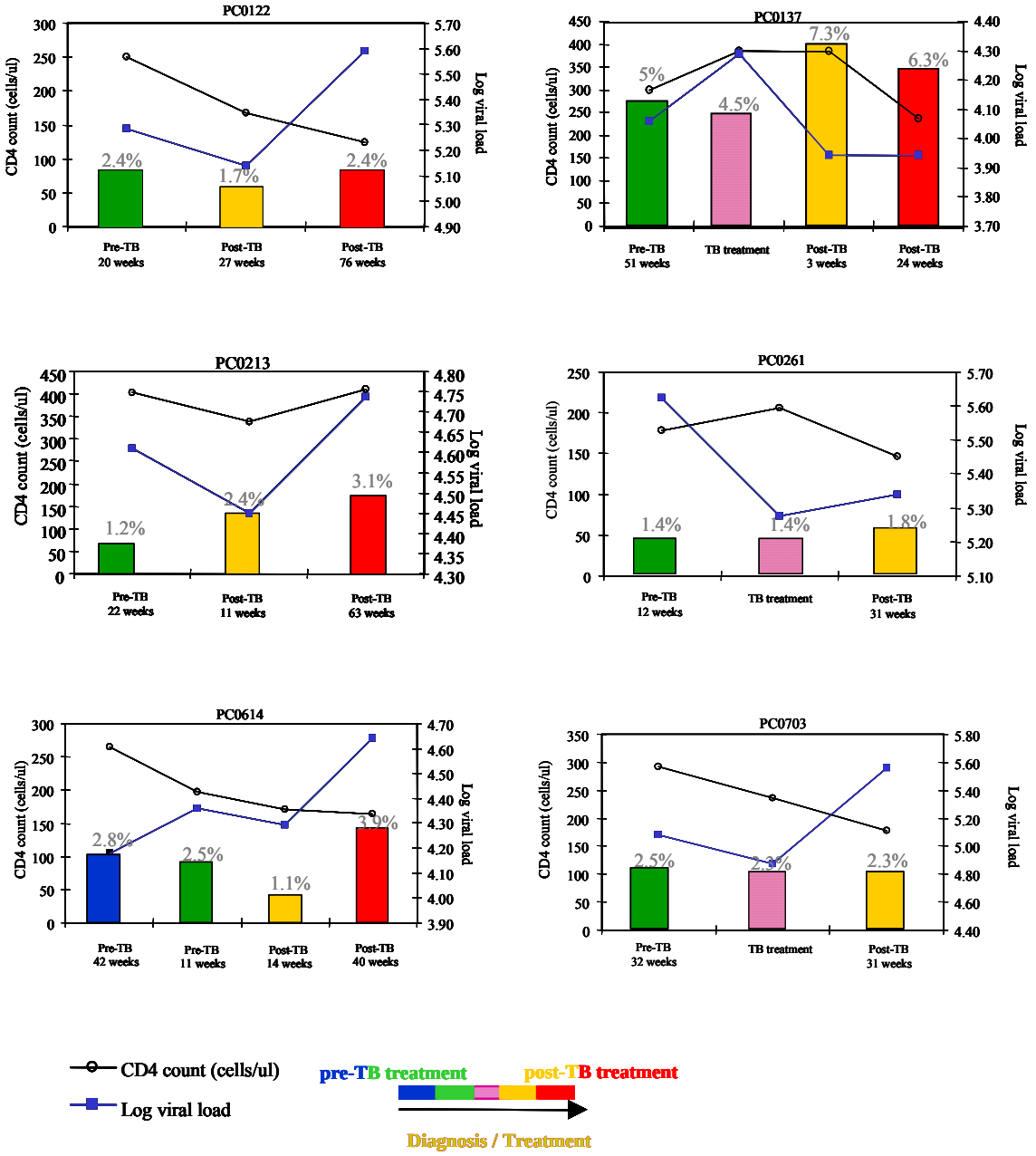


Figure 6.3: CD4 count and viral load compared to genetic diversity pre- and post-TB treatment. The percentage genetic diversity shown as coloured bars with pre-TB treatment in blue and green, start of TB treatment in pink and post-TB treatment in yellow and red, as indicated in the colour-time line.

6.3.3.2 Phylogenetic analysis within a patient

Phylogenetic trees based on the C2V3 region were drawn to reflect the relationship between the different variants present at each time point within a patient (Figure 6.4 A-F). Five control samples from patients that did not have TB were also included in the analysis, but only shown in Figure 6.4 A. In all the phylogenetic trees patient-variants clustered together in a monophyletic group compared to the control samples indicating that these variants were more closely related to each other than to other HIV-1 infected control samples.

The pairwise genetic distance was used to determine if the clustering of sequences within a phylogenetic tree were similar enough to classify them as a group of closely related variants, that could be differentiated from other variants within the tree. The pairwise genetic distance for each patient ranged from 0.017 - 0.071 [median 0.029] (Table 6.2). This reflects the amount of genetic diversity within a patient at all time points. If the overall genetic distance within a patient was less than the median of 0.029, it suggested that the variants were closely related (and any observed grouping of sequences was not significant) as seen in PC0122, PC0261 and PC0703. If the genetic distance was greater than 0.029 (such as in PC0137, PC0213 and PC0614), the genetic distance of the observed clusters within the tree were calculated separately. In these patients, variants could be differentiated into two groups (A and B) within a tree. The variation between groups A and B was >0.029, indicating significant variation between the groups.

Table 6.2: The pairwise genetic distance for 6 HIV/TB patients. In three patients genetic distance was >0.029, that separated variants into two distinct groups A and B.

Patient	Pairwise genetic distance	Group A	Group B	A vs B
122	0.026	-	-	-
137	0.071	0.027	0.023	0.12
213	0.03	0.011	0.026	0.039
261	0.017	-	-	-
614	0.03	0.014	0.03	0.039
703	0.028	-	-	-
Median	**0.029**			

Comparing the topology of the phylogenetic tree and pairwise genetic distance showed that the variants within PC0122, PC0261 and PC0703 were genetically similar (Figure 6.4 A-C). Although PC0122 and PC0703 showed some clustering of pre-TB and post-TB variants respectively in the phylogenetic trees, the overall genetic distance of variants within these patients were less than 0.029, indicating that the sequences of these variants were not significantly different. This suggested that there was amplification of existing variants in these patients, with no evidence of new populations emerging.

Phylogenetic analysis of the three remaining patients (PC0137, PC0213 and PC0614) showed clustering of variants from various time points (Figure 6.4 D-F). These clusters formed two groups within a tree. The genetic distance between the two groups (A vs B, Table 6.2) was > 0.029, suggesting that a group A had sufficient genetic differences from group B within the same patient. These groups consisted of variants from pre-TB and post-TB, pointing towards selective amplification of variants at various time points.

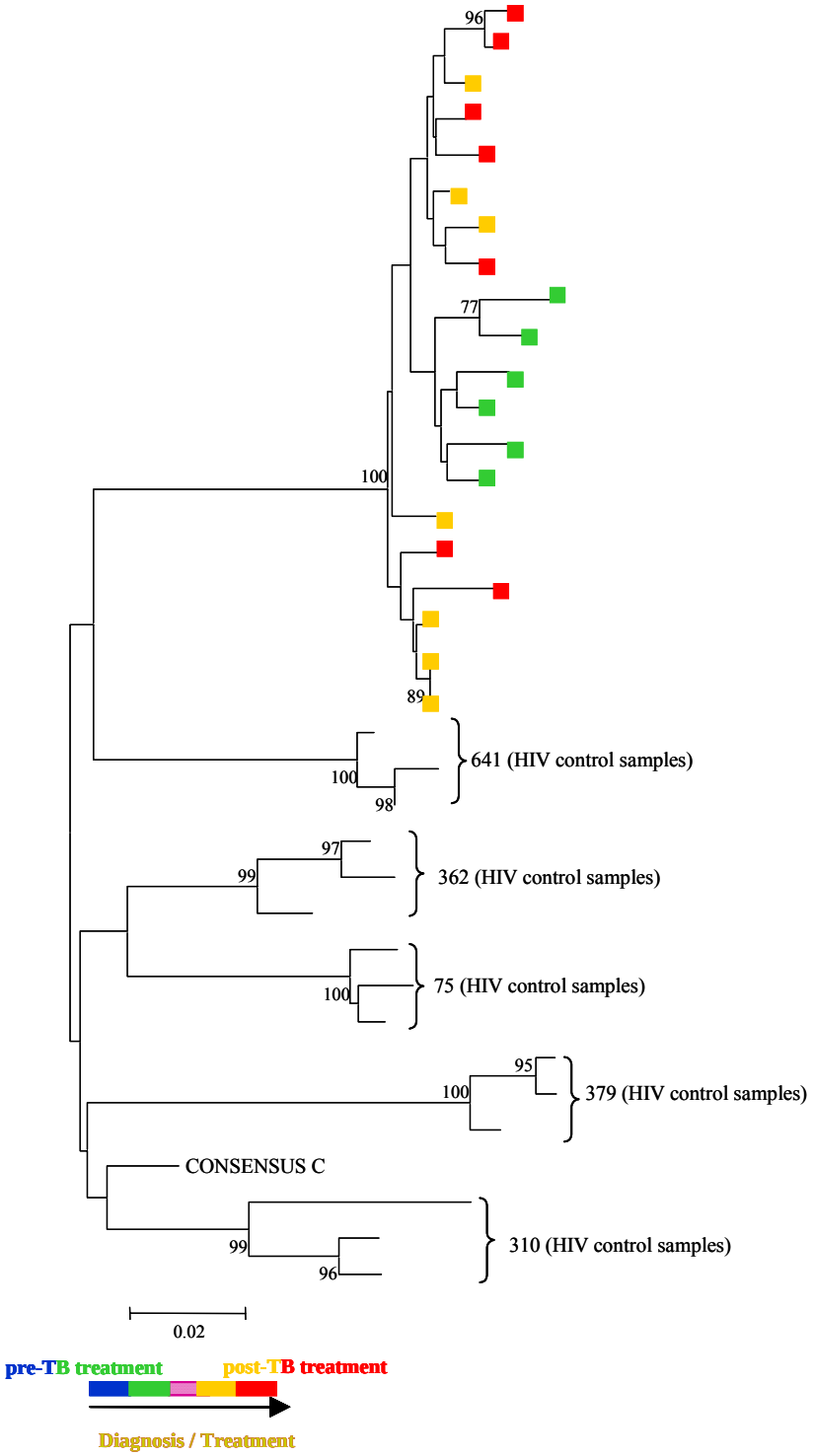


Figure 6.4 A: Phylogenetic tree of variants from PC0122 in the C2V3 region. The 5 HIV control samples and C2V3 probe consensus C sequence are indicated. The colour blocks reflect the time of samples, with blue and green pre-TB, pink diagnosis of TB and start of TB treatment, yellow and red post-TB treatment. Bootstrap values indicate confidence of tree topology and values above 75% shown on tree. Although pre-TB treatment (green) clustered closer together it was not significantly different from other variants in tree as intra patient genetic distance <0.029.

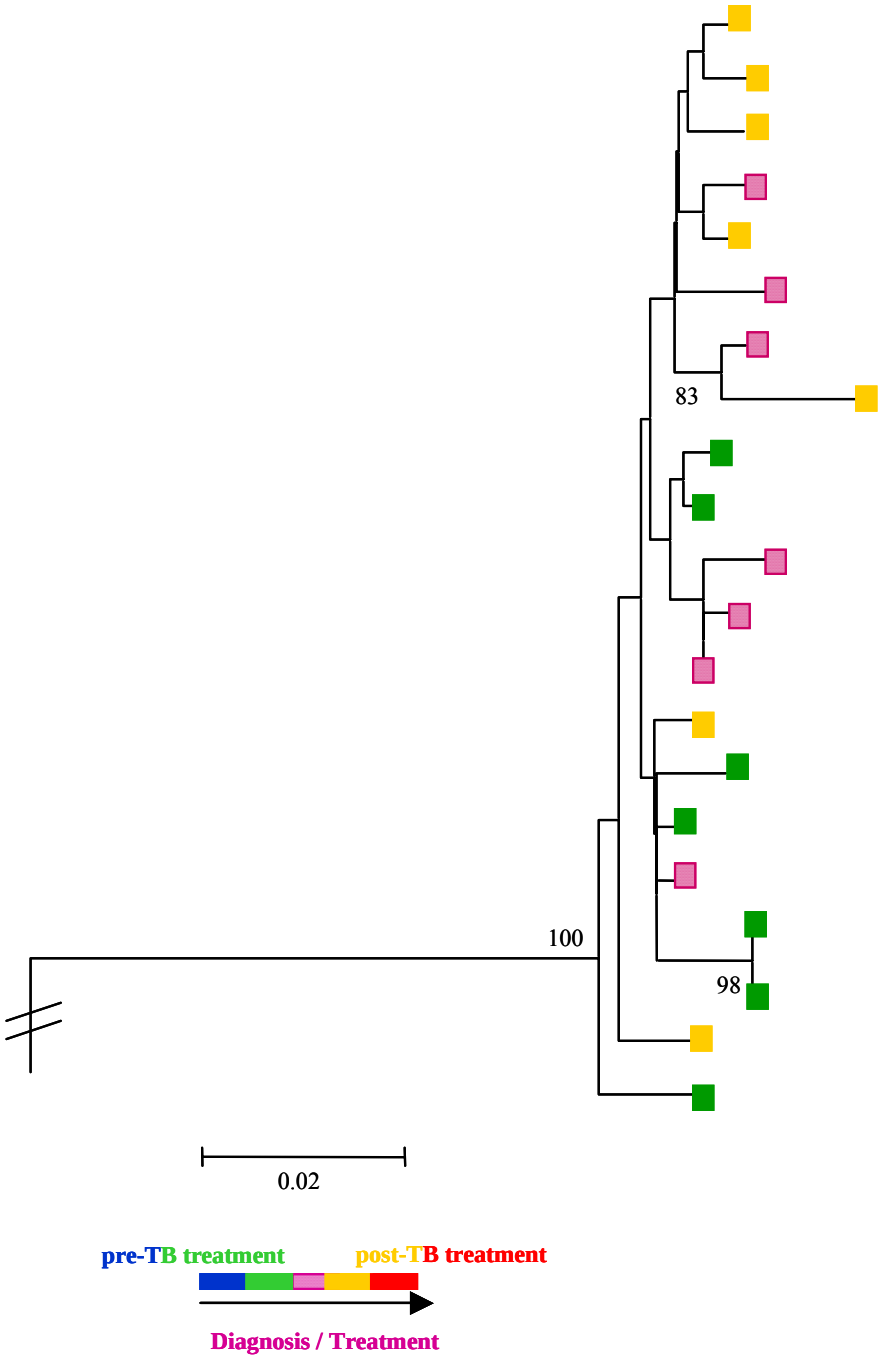


Figure 6.4 B: Phylogenetic tree of variants from PC0261 in the C2V3 region. The branch that linked control samples is indicated with (//). Variants within this patient were genetically similar, showing no clustering and therefore suggestive of non-specific amplification of variants.

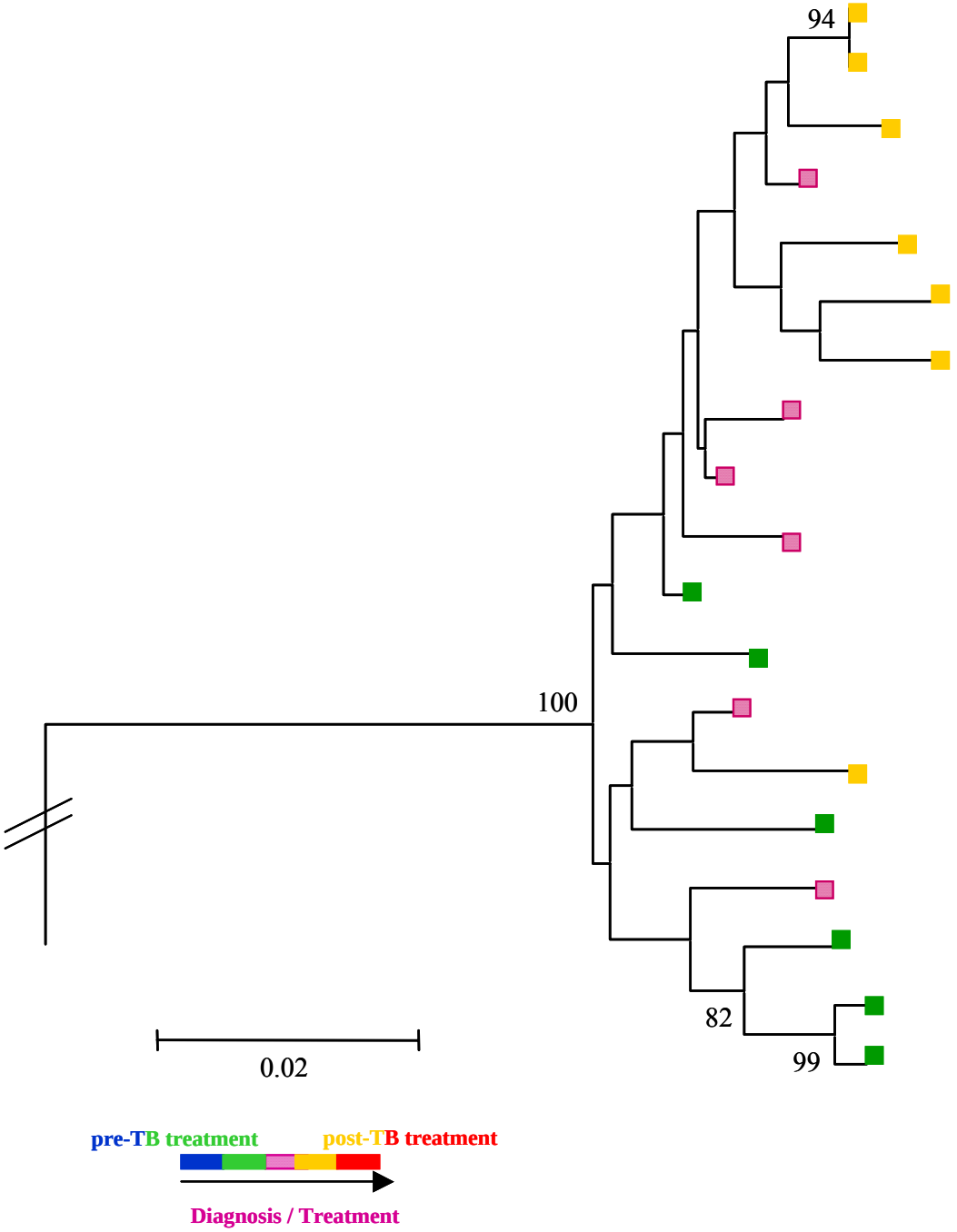


Figure 6.4 C: Phylogenetic tree of variants from PC0703 in the C2V3 region. The branch that linked control samples is indicated with (//). Variants within this patient show clustering within the tree topology but the overall pairwise genetic distance within this patient indicated that these groups were not significantly different.

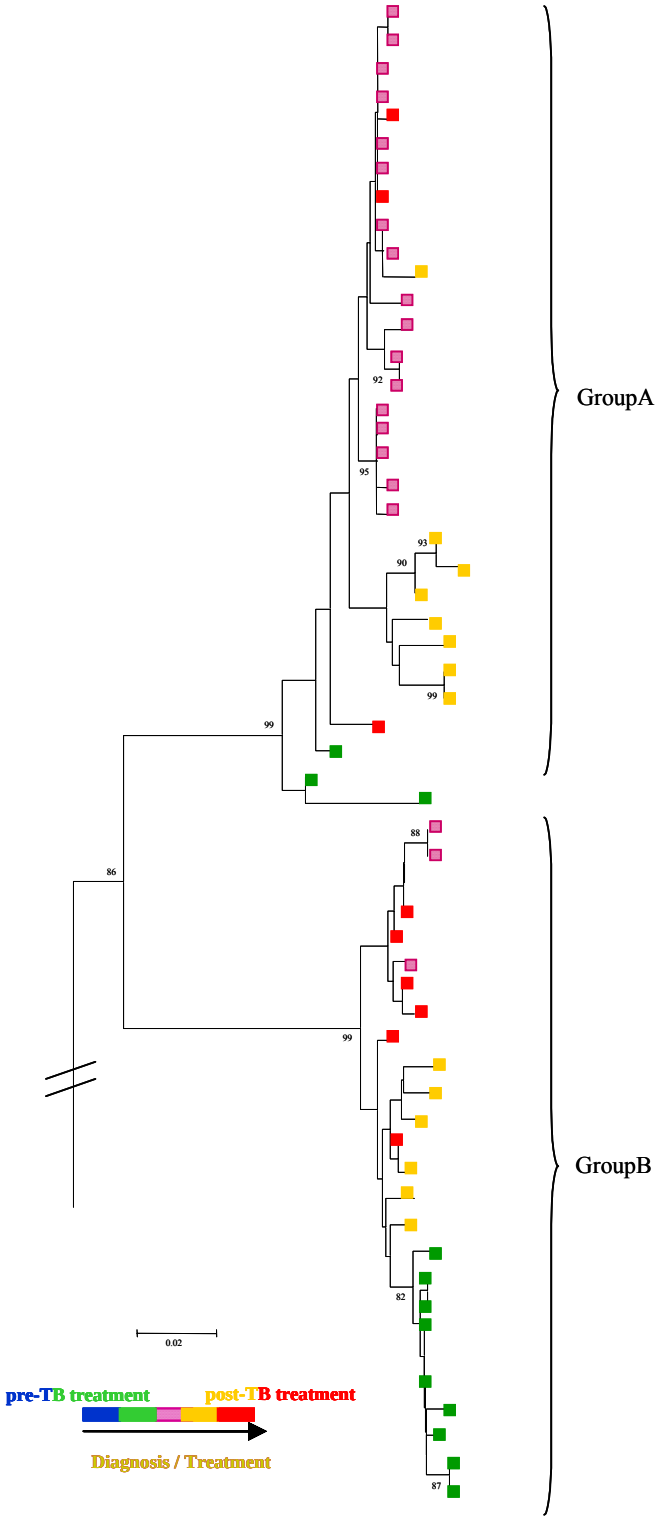


Figure 6.4 D: Phylogenetic tree of variants from PC0137 in the C2V3 region. Controls were included in analysis but they are not shown, hence the branch that linked the controls is indicated with (//). Variants within this patient clustered into 2 groups (A and B) and were genetically distinct.

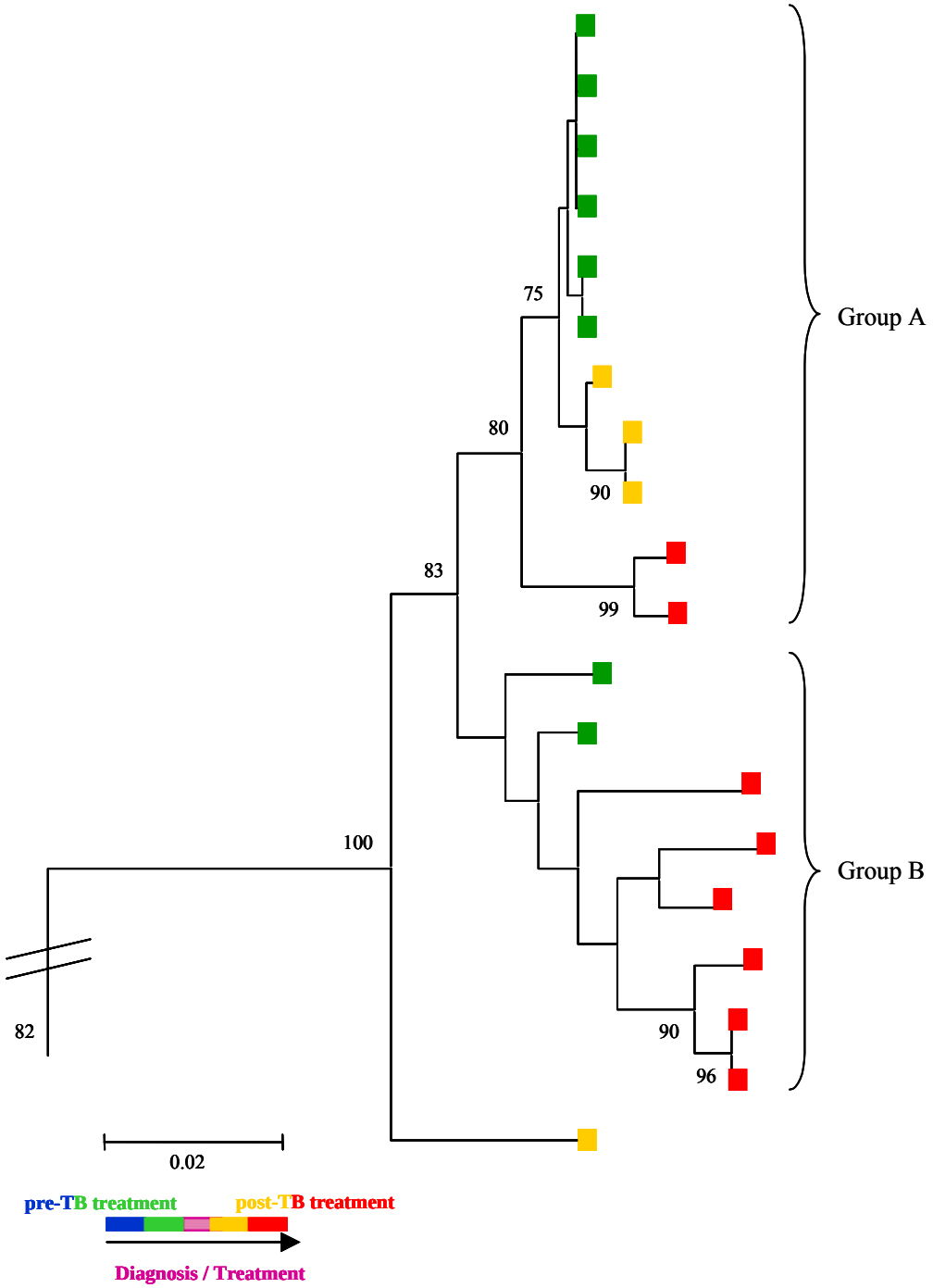


Figure 6.4 E: Phylogenetic tree of variants from PC0213 in the C2V3 region. The branch that linked control samples is indicated with (//). Variants within this patient clustered into 2 groups (A and B) pointing towards selective amplification of variants from pre-TB treatment. A single variant isolated post-TB treatment (yellow) did not cluster within these groups, suggestive of a new variant.

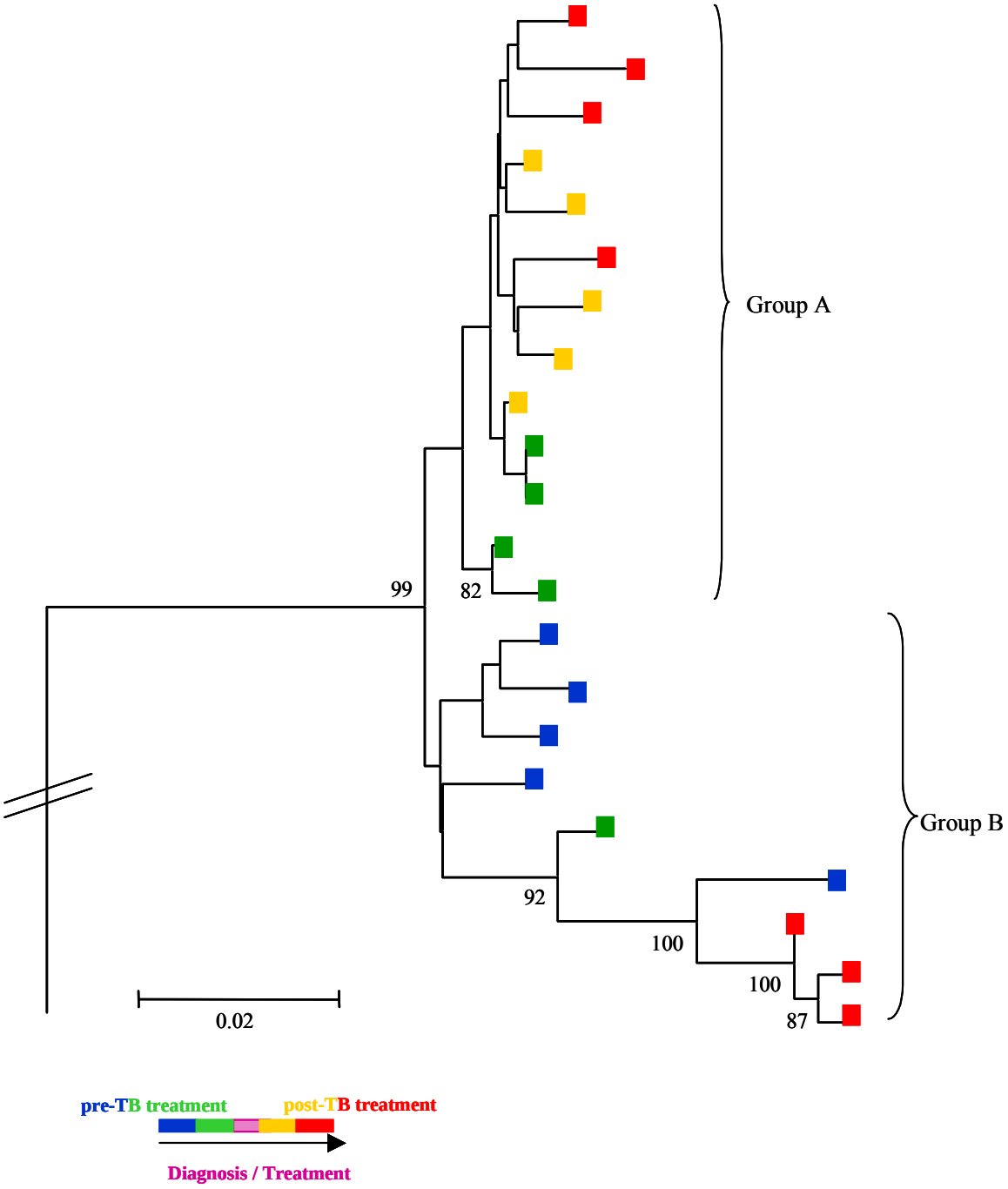
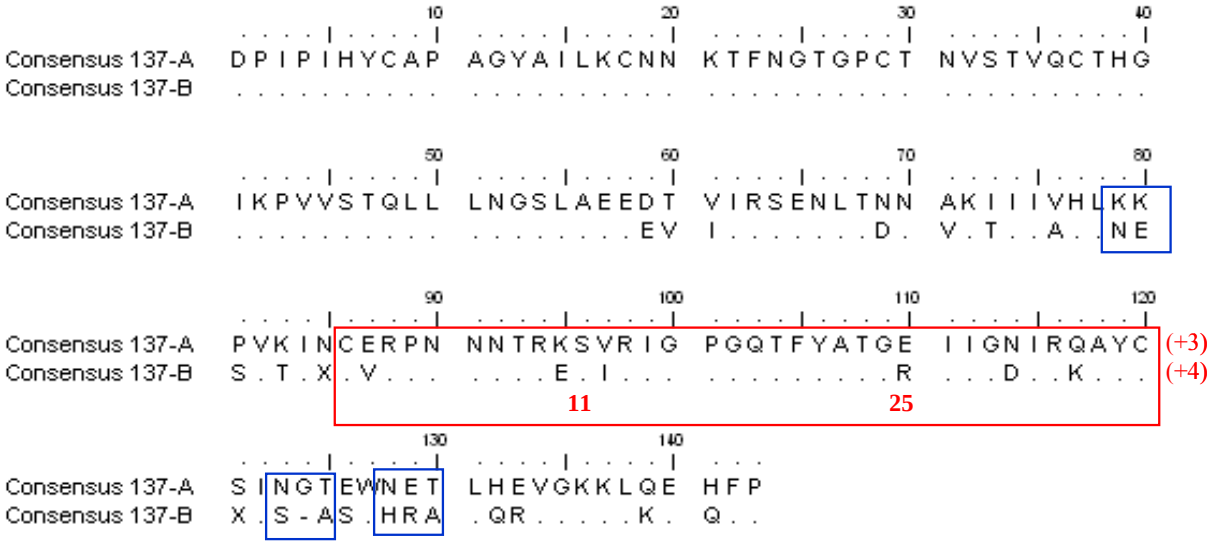


Figure 6.4 F: Phylogenetic tree of variants from PC0614 in the C2V3 region. The branch that linked control samples is indicated with (//). Variants within this patient clustered into 2 groups (A and B) suggesting selective amplification of variants.

6.3.3.3 Sequence analysis of patients with sub-populations

Consensus sequences were compiled for group A and B for each patient to investigate the genetic characteristics of these variants. The frequency of A and B variants was also calculated to quantify the presence of each group at different time points (Figure 6.5 A-C).

Analyses of variants from PC0137 showed that group A predominated at the time of TB treatment (time 0). Group B predominated during pre-TB and post-TB treatment, indicative of expansion of a minor population and then reversal to variants more similar to those present pre-TB. Consensus sequence analysis showed that group A had 2 potential glycosylation sites that were lacking in group B (at amino acid positions 123 and 128), and group B had one site not present in group A (amino acid position 79). Group B also had a deletion at amino acid position 124. There was a notable difference in the V3 charge between these groups; group A containing variants with charge of +3, whereas group B variants have a charge of +4 and an R at amino acid position 25. This further suggests 2 genetically different populations predominating at different time points within this patient. Hence, this patient experienced selection of a specific population during an episode of active TB infection, but with treatment the population profile returned to variants similar to those at pre-TB diagnosis.



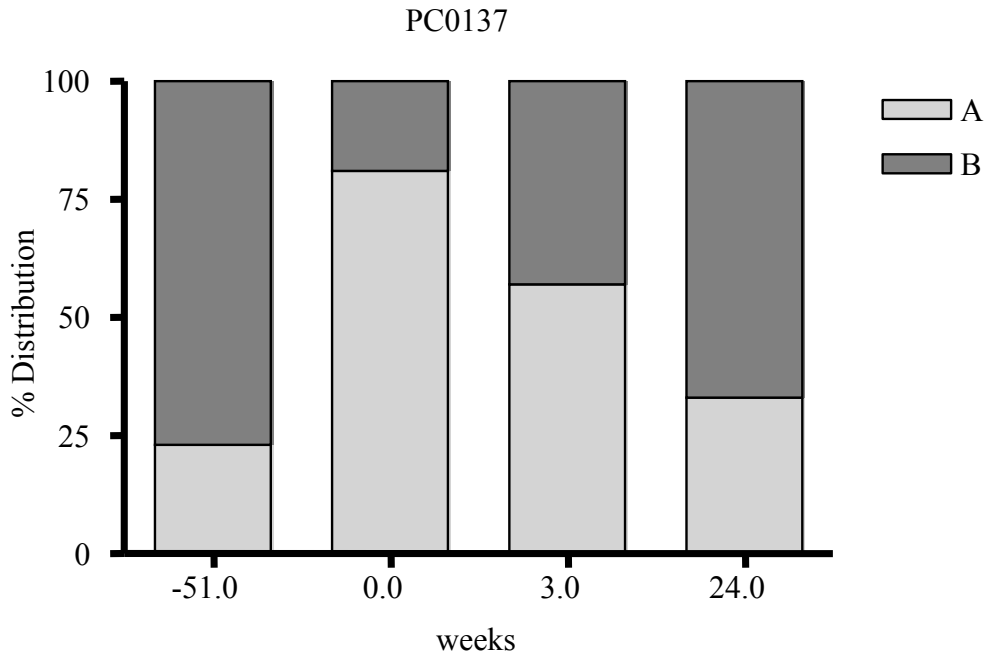


Figure 6.5 A: Consensus sequences and distribution of viral variants within PC0137 over time. Dots within amino acid sequence alignment indicate similar sequence and X depict variable amino acids. The V3 region is blocked in red, with position 11 and 25 indicated and the V3 charge in brackets. Potential glycosylation sites that are present in one consensus but not the other are boxed in blue. Distribution of these groups at the different time points are shown in the graph, indicating that group A expanded during TB treatment, but declines post-TB treatment.

The variants from PC0213 clustered into two sub-populations with most of the pre-TB treatment and 11 weeks post-TB treatment variants clustering in group A (Figure 6.5B). Variants at 63 weeks post-TB treatment clustered in group B with a few pre-TB variants, suggesting that the increased viral load post-TB treatment resulted in the amplification of a subpopulation that were not detected at 11 weeks post-TB treatment. There is also the possibility that the low viral levels at 11 weeks (Figure 6.3) might have resulted in group B not being sampled. Group A and B shared some potential glycosylation sites, but only group A had one at amino acid position 66. The consensus sequences for group A and B had similar V3 charges of +3. One clone representing a variant at 11 weeks post-TB did not cluster in either of these groups and had a V3 charge of 4.5. Sample mix-up or contamination was ruled out as this clone clustered with other sequences from this patient (Figure 6.4 E). This could be a new variant that developed at 11 weeks post-TB treatment, that was not seen at later time points.

The variants within patient PC0614 clustered into two significantly different groups A and B (Figure 6.5C, genetic distance 0.039), with eight amino acid sites of variability between them. The V3 charge was the same for both subpopulations, including a deletion at amino acid 109. Despite these few differences between the groups, group A variants were more predominant post-TB treatment. These analyses suggest selection of specific variants around TB diagnosis and treatment, but with the increased viral load earlier variants seen at pre-TB time points were also amplified.

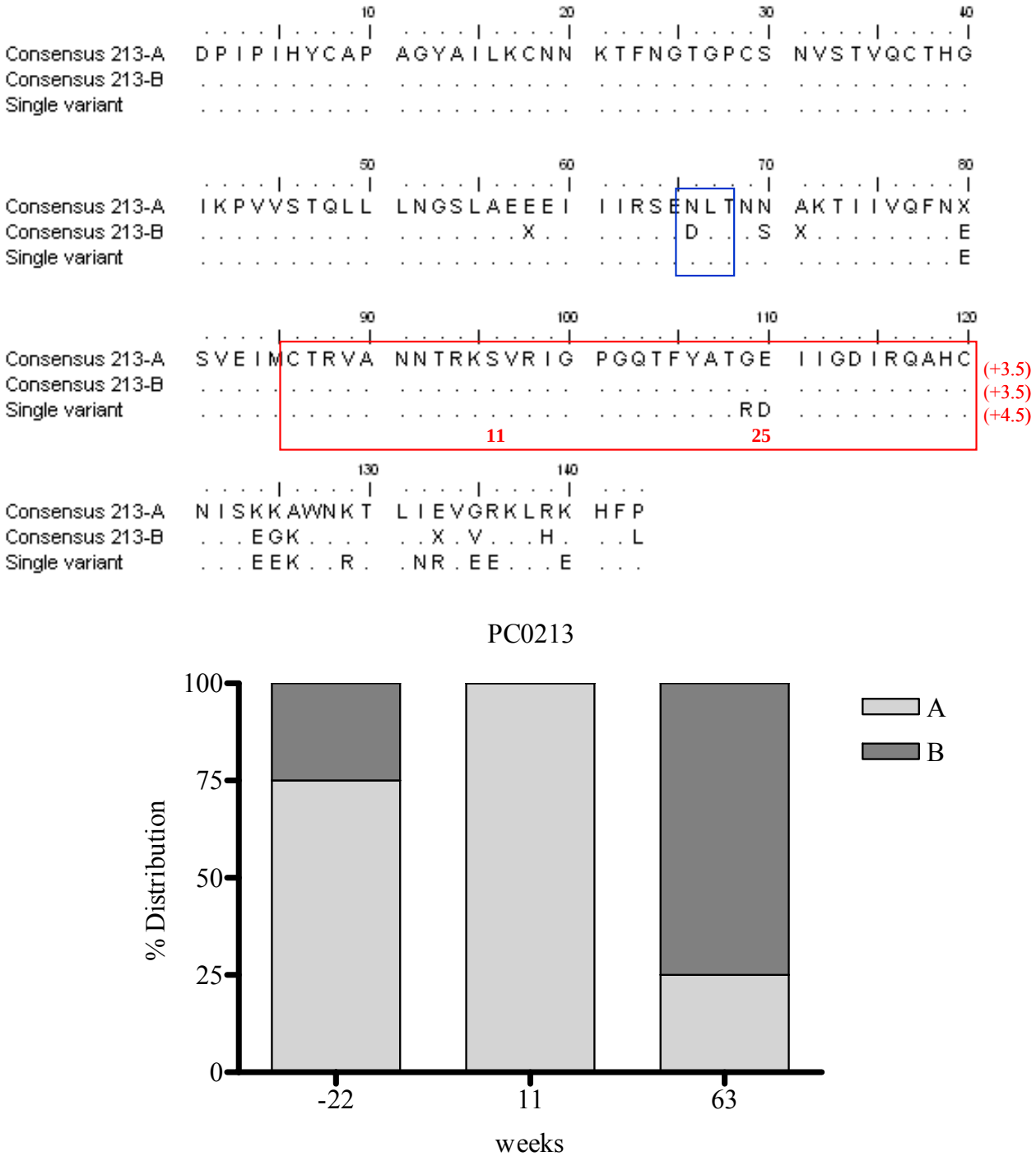
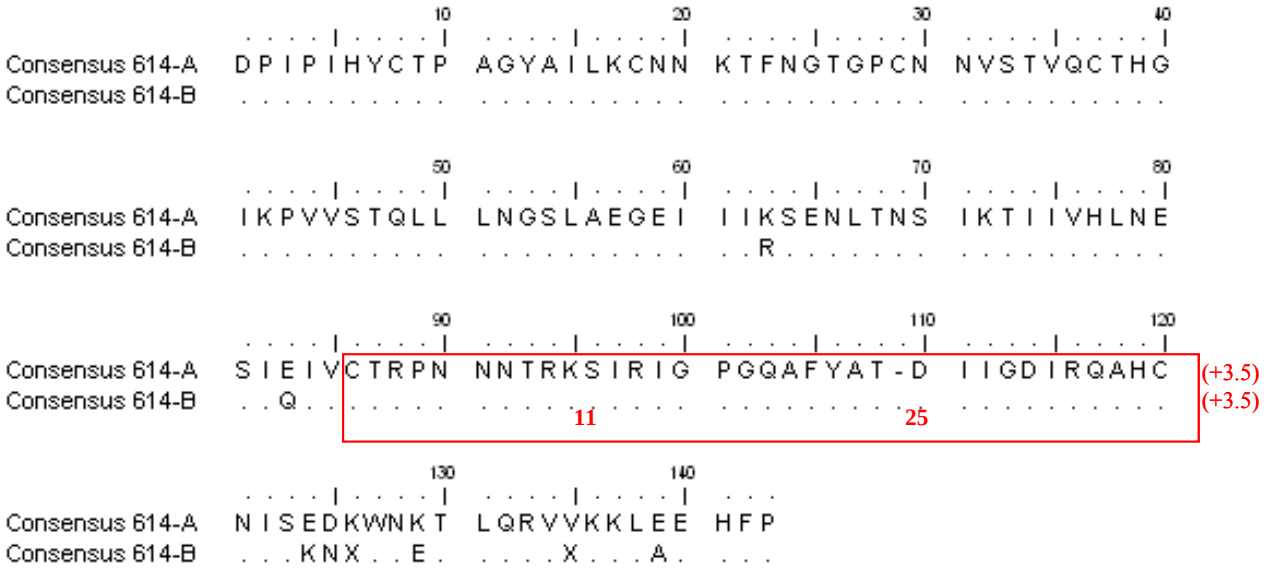


Figure 6.5 B: Consensus sequences and distribution of viral variants within PC0213 over time. The V3 region is blocked in red, with position 11 and 25 indicated and the V3 charge in brackets. Potential glycosylation sites that are present in one consensus but not the other are boxed in blue. Distribution of these groups at the different time points are shown in the graph, indicating that although 11 weeks post-TB treatment had only group B variants, variants from group A were present at 63 weeks post-TB treatment.



PC0614

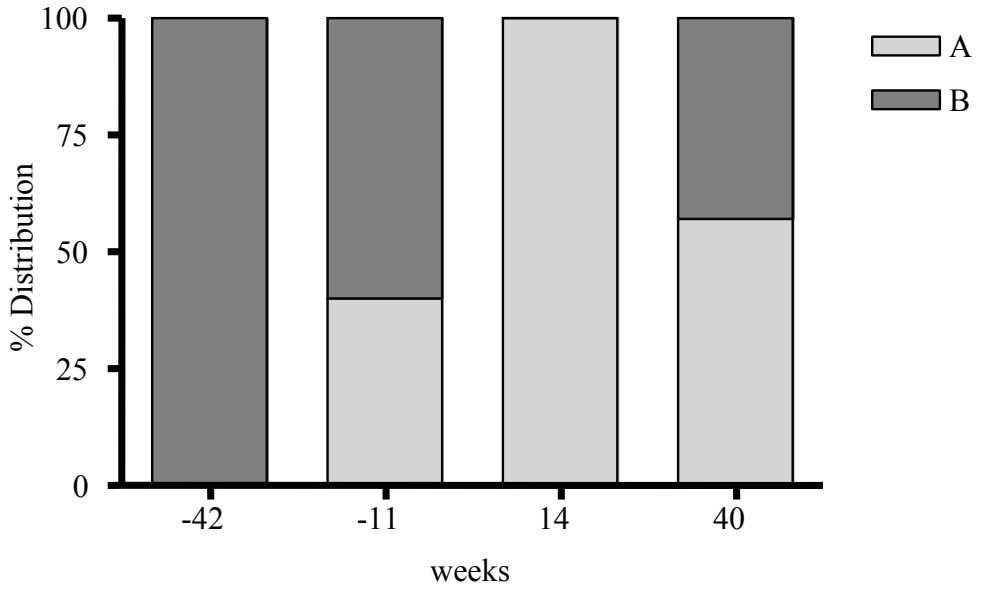


Figure 6.5C: Consensus sequences and distribution of viral variants within PC0614 over time. The V3 region is blocked in red, with position 11 and 25 indicated and the V3 charge in brackets. Potential glycosylation sites that are present in one consensus but not the other are boxed in blue. Distribution of these groups at the different time points are shown in the graph, indicating that group B were predominant at 42 week pre-TB whereas group A consist of variants present at 14 weeks post-TB treatment.

6.4 DISCUSSION

In this study we explored the effects of active TB on HIV-1 heterogeneity by determining the genetic characteristics of viral populations present over time in HIV-1 co-infected patients. Eighteen HIV-1 infected patients were followed for 2 years in which samples were obtained pre- and post-TB diagnosis and treatment. Six patients with a high degree of genetic diversity were selected for further analysis of which three showed an expansion of existing populations and three showed selection of a subpopulation, suggesting that TB can influence viral heterogeneity in some patients. A summary of the findings in this chapter is shown in Figure 6.6.

Genetic characterization of HIV-1 within this cohort showed that these patients were infected with HIV-1 subtype C strains. Phenotype prediction methods indicated that most of the samples contained R5 viruses. For three samples (PC0074, PC0261 and PC0284) the prediction methods gave discordant results. This included positively charged amino acids at positions 11/25, or mutations in the crown motif that resulted in variation in the GPGQ motif and intermediate PSSM scores suggestive of CXCR4 usage. However, these prediction discrepancies can only be addressed by determining the coreceptor usage experimentally with cell lines expressing either CCR5 or CXCR4. The finding that the majority of HIV/TB patients harbour CCR5-using variants support earlier studies on similar cohorts where all isolates were also reported to be R5 (Morris, Cilliers *et al.* 2001; Morris, Martin *et al.* 2003).

Figure 6.6: Summary of results from the 6 HIV/TB patients. The variability of variants present in the C2V3-HTA profiles of a patient over time is described. Increase, decrease or no change in CD4 count, viral load and genetic diversity are shown with arrows. If the variants within a patient formed subpopulations, the distribution of these populations at each time point are indicated, with the dominant group at that time point in capital.

Patient	C2V3-HTA profile	CD4 count	Viral load	Genetic diversity	Sub-populations	Distribution of populations		
						pre-TB	0	post-TB
PC0122	expansion	↓	↑	↔				
PC0137	high shifts, expansion	↓	↓	↑	A and B	aB	Ab	Ab aB
PC0213	high shifts, expansion	↔	↑	↑	A and B	Ab		A aB
PC0261	amplification of same	↓	↓	↑				
PC0614	selection	↓	↑	↑	A and B	B aB		A Ab
PC0703	expansion	↓	↑	↔				

In this study most of these patients had a decrease in CD4 count over time. Two patients had no change in genetic diversity although the viral load increased, suggestive of amplification of existing populations over time. Four patients had an increase in viral diversity of which two were associated with an increase in viral load. Two patients had a decrease in viral load while the genetic diversity increased, suggesting that minor populations were selectively amplified or there was selective amplification of a very diverse population. Previous studies have shown a correlation between increased fitness during disease with increased viral load and HIV-1 C2V3 diversity (Troyer, Collins *et al*. 2005). In this study the genetic diversity did increase over time suggestive of the selection of more fit variants, but it was not necessarily associated with increased viral load.

Phylogenetic and sequencing analyses were used to determine if increased replication resulted in amplification of non-specific variants, selection of a specific population or emergence of new variants. For three patients (PC0122, PC0261 and PC0703) there appeared to be non-specific amplification of existing variants as sequences from different time points were mixed in the tree topologies with no significant clustering of specific variants. Three patients had selective amplification of specific variants as reflected in the grouping of variants in the phylogenetic trees. Selection of a specific population usually created a cluster that was separated from other variants. These groups were determined based on the pairwise genetic distance within a patient as well as the distance between the selected groups. Therefore, PC0137, PC0213 and PC0614 had sufficient genetic diversity within their populations to suggest that (with some degree of confidence) a cluster of sequences was different from another as indicated by the increase genetic distance between group A and B within a patient.

Sequence analysis showed that some of these groups had different characteristics, such as different V3 amino acid charge and predicted glycosylation sites, further proving that they represent independent variants within a patient. The distribution of these groups also differed within a patient indicating that some groups were more adaptable at certain times e.g. pre-TB or post-TB treatment.

TB and HIV-1 infection are chronic conditions, which results in constant immune activation within a patient, and complicates studying the effect of TB on HIV-1. Nevertheless, in this study it was observed that different variants dominated at certain stages within some patients. To ascertain whether the selection of these specific variants were due to TB or TB treatment can only be determined with further investigations that include analysis of variants present in a control group of patients that experienced no TB infection. In conclusion, amplification of existing populations as well as selective amplification of variants were observed in this study, therefore TB and TB treatment did impact on viral heterogeneity but mostly by selective amplification of specific variants at certain time point and this was not necessarily due to the increased viral replication.

CHAPTER 7

CONCLUSION

7.1 CONCLUSION

In this study the *in vitro* and *in vivo* diversity of HIV-1 subtype C envelope proteins were investigated and correlated to the biological phenotype of viral isolates. During entry of HIV-1 into a target host cell, the virion binds to CD4 and a coreceptor. The major coreceptors are CCR5 and CXCR4 with CCR5 used more frequently. CXCR4 usage is usually seen in later stages of disease, but in subtype C only a few viruses have been isolated able to employ this coreceptor. The ability to use these different coreceptors is largely influenced by genetic changes in the envelope glycoprotein. Selective pressures due to the host immunological response, as well as the error prone reverse transcriptase contribute to variation within the envelope region and specifically the V3 region (Bebenek, Abbotts *et al*. 1989; Watkins, Reitz *et al*. 1993; Burns and Desrosiers 1994). This diversification lends HIV-1 the ability to expand its coreceptor utilisation during the course of HIV-1 infection, and provides the virus access to a spectrum of potential target cells within the host (Hu, Barry *et al*. 2000). Determining viral phenotype is important in many different research aspects, as coreceptor usage impacts on the pathogenicity, tissue tropism and transmissibility of a virus *in vivo*.

Understanding and identifying the differences in coreceptor utilisation of subtype C is important as this subtype predominates worldwide. Recently more focus has been placed on blocking the entry stage of the viral life cycle and preventing the sequelae of infection and provirus integration. Entry inhibitors, that bind to CCR5 or CXCR4 and prevent binding of the gp120, are being tested in human clinical trials and may prove particularly useful as part of a microbicide formulation to prevent infection at

mucosal sites. However, HIV-1 can escape from these inhibitors by mutation allowing gp120 to still bind in the presence of the inhibitor (Hartley, Klasse *et al*. 2005). Escape from CCR5 inhibitors could also select for CXCR4-using variants. Therefore, insight into the evolution of X4 viruses is crucial to understanding the interaction between the immune system and the virus, immune escape and drug resistance, as these will assist in better vaccine strategies, drug development and treatment.

The reason why X4 viruses emerge in some patients is uncertain. It is not clear, for instance, whether X4 viruses are the cause or consequence of disease progression, how common X4 viruses are *in vivo*, or whether there are virological and/or immunological constraints selecting against these viruses early in infection (Kuhmann, Pugach *et al*. 2004; Moore, Kitchen *et al*. 2004). The infrequency of X4 viruses in subtype C suggests virological differences compared to other subtypes or possibly host differences since subtypes are generally geographically distributed. Clearly, X4 viruses are not required for disease progression in subtype C infections as most patients with advanced AIDS do not develop them (Tscherning, Alaeus *et al*. 1998; Ping, Nelson *et al*. 1999; Cilliers, Nhlapo *et al*. 2003). This gives rise to important questions. Is pathogenesis in subtype C infections fundamentally different from that of subtype B infections? If not, does the rarity of X4 virus in progressive subtype C infection support the idea that X4 development in B infections is a by-product and not a cause of end-stage disease? It has been suggested that the low frequency of X4 viruses in subtype C infection has more to do with the later onset of the epidemic in developing countries, where subtype circulates; i.e. that the virus has not yet evolved CXCR4 usage. Since the time to AIDS is often shorter in individuals

residing in developing countries, it has also been suggested that X4 variants have simply not had sufficient time to evolve *in vivo* (Cilliers, Nhlapo *et al*. 2003). However this would not explain patients in Uganda infected with subtype D viruses where X4 viruses are frequently found (De Wolf, Hogervorst *et al*. 1994). This latter observation suggests that CXCR4 usage is mainly a virus-driven process.

To address these questions we investigated the development and characteristics of CXCR4 viruses in HIV-1 subtype C infection (Chapters 2, 4, and 5). Comparing the V3 region of CCR5 and CXCR4 using isolates it was clear that there were genetically distinct. R5 viruses were more conserved irrespective of disease stage, whereas the CXCR4-using viruses were variable in the stem and crown of the V3 region. Characteristics of CXCR4 usage included the loss of the N-glycosylation site, increased charge and specific amino acid changes in particular at positions 11 and 25 in the V3 region, similar to subtype B. The most noteworthy difference between the subtypes was the highly conserved crown motif GPGQ in subtype C R5 viruses, whereas X4 viruses had mutations in this region, particularly Q changing to R. The consensus crown motif in subtype B (for both CCR5 and CXCR4) is GPGR. Thus, we suggest that the conserved GPGQ crown motif of subtype C R5 viruses might restrict the development of X4 viruses. Although some subtype C CXCR4-using viruses have a GPGQ motif and have therefore overcome this restriction, mutations in the crown might be just one of the many ways for subtype C viruses to develop CXCR4 usage. The importance of the V3 crown in coreceptor binding has recently been shown in a study of the crystal structure of the core of gp120 containing the V3. The authors propose that the N-terminus of the coreceptor binds to the V3 base and the V3 crown interacts directly with the second extracellular loop of CCR5 (Huang,

Tang *et al*. 2005). Similar studies on subtype C envelopes and CXCR4-using envelopes need to be done. Such studies will help us in understanding how genetic variation in the crown affects interaction with HIV-1 coreceptors.

The genetic differences between R5 and X4 viruses within the small region of V3 (consisting of an average of 35 amino acids), has made it a suitable region for phenotype prediction methods based on genotype. Most prediction methods are based on subtype B sequences, as these are more readily available. In this study, we have provided to our knowledge, the largest number of unlinked subtype C sequences and this has provided the opportunity to test these prediction methods for subtype C. Although the 'gold standard' for phenotyping is experimentally determined coreceptor usage of viral isolates, *in vitro* culture probably skews the coreceptor profiles by selecting variants as was seen in Chapter 5. In particular CXCR4-using viruses are often better adapted for growth *in vitro*. Future studies should focus on cloning envelope genes from plasma RNA for accessing coreceptor usage and comparing this with genotypic predictors. Of the four different prediction methods used in this study, all had high specificity (correctly predicting R5 viruses) but low sensitive for X4 viruses (Chapter 4). We therefore developed a subtype C specific prediction method based on position specific scoring matrix (PSSM). This method was more sensitive in detecting X4 viruses as well highlighting that certain positions within the V3 contribute differently to phenotype in subtype B and subtype C.

Dualtropic viruses, seen as the intermediate step in the transition from R5 to X4, are of particular interest, as these viruses have the ability to use both receptors. Jensen and colleagues have shown that R5X4 viruses had intermediate PSSM scores, with

some viruses more R5-like, while others had a higher score and were more X4-like (Jensen, Li *et al*. 2003), and this was confirmed in this study. Better understanding of this process might clarify the changes that lead to a virus switching coreceptor usage. The rare occurrence of CXCR4-using viruses in subtype C makes it difficult to identify patients who have undergone a coreceptor switch. We identified and analysed three such patients (Chapters 4 and 5). In these studies it was shown that coreceptor switching happened in a step-wise manner dependent on multiple changes within the V3 region that differed between isolates. Two of the patients were dually infected suggesting that 'genetic leaps' as a result of recombination facilitated rapid emergence of X4 usage (Chapter 4). Reconstructing the development of X4 viruses within a single patient that has undergone a coreceptor switch indicated two routes to CXCR4 development (Chapter 5). In one pathway, viruses developed dualtropism and were characterised with changes in the crown motif, but low V3 charge and no positive amino acids at positions 11 or 25. The second route resulted in X4 viruses, which were associated with changes in the crown motif, as well as positive amino acids at position 11 that resulted in an increased charge. Thus, the development of CXCR4 usage within a patient can follow different mutational pathways and does not necessarily lead to X4 only viruses. Pastore *et al.* (2004) suggested that this mutational pathway not only differs between samples, but that the mutational distance (number of mutations necessary) from CCR5 to CXCR4 usage (although relatively small) differs between samples as well. Given that R5 to X4 switching requires so few changes it is not clear why these viruses don't occur more frequently. *In vitro* they are readily generated in weeks. Pastore *et al*. (2004) suggest that the changes required to undergo a coreceptor switch renders them unfit and sensitive to coreceptor inhibitors.

This study has shown that similar to subtype B, the V3 region is the major contributor of phenotypic expression of coreceptor tropism in subtype C. The same characteristics such as the loss of the N-glycosylation site, increased charge and specific amino acids contributed to the development of CXCR4 usage. In addition, this study has highlighted another route of CXCR4 development involving changes within the crown motif of subtype C viruses. Therefore, pathogenesis in subtype C infections resulting from the broadening of coreceptor usage to CXCR4 is not fundamentally different from that of subtype B infections. Although known X4 viruses in subtype C are rare, with the increased understanding of their development and improved detection methods, the likelihood of identifying these viruses will increase. The rarity of X4 virus in progressive subtype C infection suggests that X4 development in subtype B infections is a by-product and not a cause of end-stage disease. However, the appearance of X4 variants is associated with rapid depletion of circulating CD4 T-cells as shown in chapter 5 and so future studies are needed to explore the pathogenicity of CXCR4-using viruses. In conclusion, this study has contributed to a better understanding of the development and role of CXCR4 in HIV-1 subtype C infection.

APPENDICES

Appendix A: Ethical clearance

Appendix B: Additional ethical clearance

Appendix C: Amino acid abbreviations

Appendix D: Reagents and recipes

Appendix E: HTA probe sequences

Appendix F: Vector map

Appendix G: Determining coreceptor usage in transfected cell lines

Appendix A: Ethical clearance

Ethical clearance was obtained from the University of Witwatersrand Committee for Research on Human Subjects (Medical) protocol number M990327.

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL)
Ref: R14/49 Morris

CLEARANCE CERTIFICATE PROTOCOL NUMBER M990327

PROJECT Isolation, Characterisation & Neutralisation Of Human Immunodeficiency Viruses In South Africa

INVESTIGATORS Dr L Morris

DEPARTMENT Department of Virology, National Institute for Virology

DATE CONSIDERED 990326

DECISION OF THE COMMITTEE *

Approved unconditionally

DATE CHAIRMAN .....(Professor P E Cleaton-Jones)

* Guidelines for written "informed consent" attached where applicable.

c c Supervisor: Dr L Morris
Dept of Department of Virology, National Institute for Virology

==========

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and ONE COPY returned to the Secretary at Room 10001, 10th Floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee.

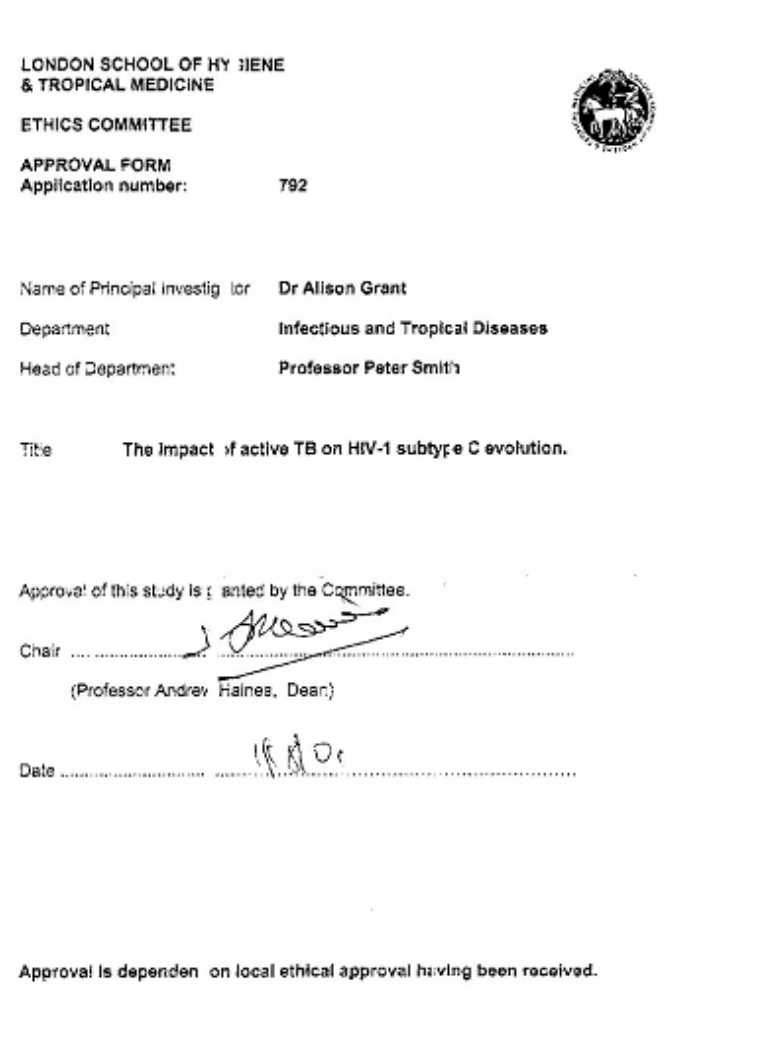
DATE 8 April 99 SIGNATURE .....

PROTOCOL NO.: M 990327

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix B: Additional ethical clearance

Ethical clearance was also obtained from the London School of Hygiene and Tropical Medicine Ethics Committee for research work done in Chapter 6, protocol number 792: The impact of active TB on HIV-1 subtype C evolution.



LONDON SCHOOL OF HY IIENE
& TROPICAL MEDICINE

ETHICS COMMITTEE

APPROVAL FORM
Application number: 792

Name of Principal investig tor Dr Alison Grant

Department Infectious and Tropical Diseases

Head of Department: Professor Peter Smith

Title The impact of active TB on HIV-1 subtype C evolution.

Approval of this study is granted by the Committee.

Chair .....

(Professor Andrew Haines, Dean)

Date .....

Approval is dependen on local ethical approval having been received.

Appendix C: Amino acid abbreviations

Amino acid	Abbreviation
Alanine	A
Arginine	R
Asparagine	N
Aspartic acid	D
Cysteine	C
Glutamic acid	E
Glutamine	Q
Glycine	G
Histidine	H
Isoleucine	I
Leucine	L
Lysine	K
Methionine	M
Phenylalanine	F
Proline	P
Serine	S
Threonine	T
Tryptophan	W
Tyrosine	Y
Valine	V
Any amino acid	X

Appendix D: Reagents and recipes

10XTBE

108g Tris-HCL

55g Boric acid

20ml 1.5M EDTA

Made to one litre with dH_2O.

HTA Annealing Buffer

1M NaCl

100mM Tris-HCL [pH7.5]

20mM EDTA

LB Broth

10g NaCl

5g Yeast extract

10g Tryptone

Made to one litre with dH_2O.

X-Gal, IPTG, Ampicillin LB agar plates

10g tryptone

5g yeast extract

10g NaCl

15g agar

Made up to one litre dH_2O.

Appendix E: HTA probe sequences

The V3-HTA and C2V3-HTA probes were obtained from HIV-1 subtype C R5 viruses.

>V3-HTA probe

TGTACAAGACCCAACAATAATACAAGAAAAAGTATGAGGATAGGACCAGG
ACAAACATTCTATGCAACAGGAGACATAATAGGAAACATAAGACAAGCAC
ATTGT

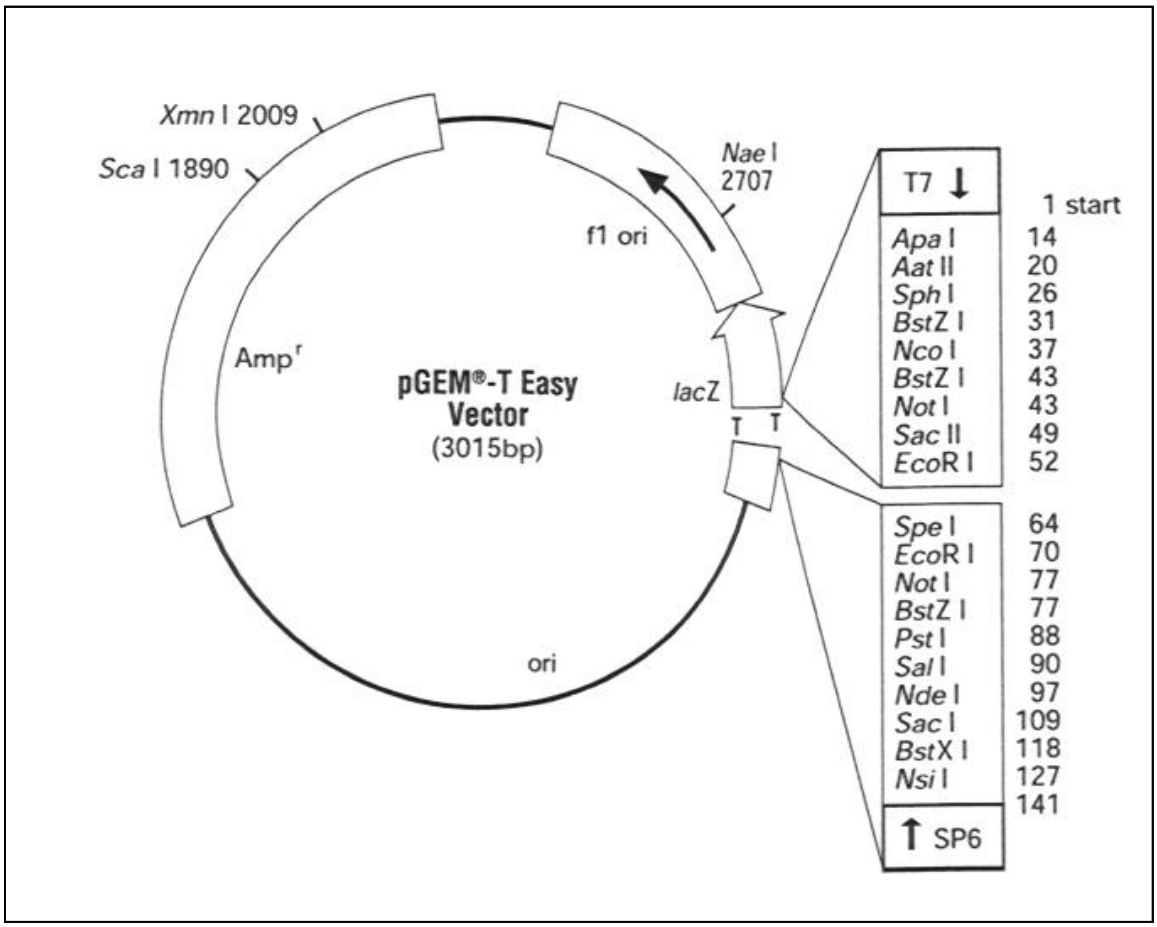
>C2V3-HTA probe

GGTTATGCGATTCTAAAGTGTAATAATAAGACATTCAATGGGACAGGACC
ATGCAATAATGTCAGCACAGTACAATGTACACATGGAATTAAGCCAGTGG
TATCAACTCAATTACTGTTAAATGGTAGCCTAGCAGAAGAAGAGATAATA
ATTAGATCTGAAAATCTGACAAACAATATCAAAACAATAATAGTCCACCT
TAATAAATCTGTAGAAATTGTGTGTACAAGACCCAACAATAATACAAGAA
AAAGTATAAGGATAGGACCAGGACAAACATTCTATGCAACAGATGCAAT
AATAGGAAACATAAGAGAAGCACATTGTAACATTAGTAAAAGTAACTGG
ACCAGTACTTTAGAACAGGTAAAGAAAAA

Appendix F: Vector map

pGEMTeasy vector map

pGEMTeasy vector map showing multiple cloning sites (Promega Corporation, USA). Amplicons were inserted between EcoRI and SpeI.



Appendix G: Determining coreceptor usage in transfected cell lines

Viral isolates were tested for their abilities to replicate in U87.CD4 cells transfected with either CCR5 or CXCR4. Cells were plated in 12-well plates 10^5 cells/well in 2 ml of selection medium [Dulbecco's modified eagle's medium (DMEM)] containing 10% FCS, antibiotics plus 500ug/ml G418 (Boehringer Mannheim GmbH) and 1ug/ml puromycin (ICN Biomedicals Inc., Ohio, USA). The following day, 1 000 $TCID_{50}$ (50% tissue culture infectious doses) of virus was added. After incubation overnight, the cultures were washed three times with DMEM plus 10% FCS to remove unbound virus, then monitored for syncytium formation and p24 antigen production on days 4, 8 and 12. Isolates that induced syncitia formation and generated increasing concentrations of p24 antigen were considered to be replication positive. (Performed by Tonie Cilliers and Melene Smith).

`

CHAPTER 8

REFERENCES

Abebe, A., D. Demissie, et al. (1999). "HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS." Aids **13**(11): 1305-11.

Allaway, G. P., K. L. Davis-Bruno, et al. (1995). "Expression and characterization of CD4-IgG2, a novel heterotetramer that neutralizes primary HIV type 1 isolates." AIDS Res Hum Retroviruses **11**(5): 533-9.

Baba, M., O. Nishimura, et al. (1999). "A small-molecule, nonpeptide CCR5 antagonist with highly potent and selective anti-HIV-1 activity." Proc Natl Acad Sci U S A **96**(10): 5698-703.

Barre-Sinoussi, F., J. C. Chermann, et al. (1983). "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)." Science **220**(4599): 868-71.

Batra, M., P. C. Tien, et al. (2000). "HIV type 1 envelope subtype C sequences from recent seroconverters in Zimbabwe." AIDS Res Hum Retroviruses **16**(10): 973-9.

Bebenek, K., J. Abbotts, et al. (1989). "Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase." J Biol Chem **264**(28): 16948-56.

Beerenwinkel, N., M. Daumer, et al. (2003). "Geno2pheno: Estimating phenotypic drug resistance from HIV-1 genotypes." Nucleic Acids Res **31**(13): 3850-5.

Berger, E. A. (1997). "HIV entry and tropism: the chemokine receptor connection." Aids **11**(Suppl A): S3-16.

Berger, E. A., R. W. Doms, et al. (1998). "A new classification for HIV-1." Nature **391**(6664): 240.

Berkowitz, R. D., S. Alexander, et al. (2000). "Causal relationships between HIV-1 coreceptor utilization, tropism, and pathogenesis in human thymus." AIDS Res Hum Retroviruses **16**(11): 1039-45.

Berson, J. F., D. Long, et al. (1996). "A seven-transmembrane domain receptor involved in fusion and entry of T-cell-tropic human immunodeficiency virus type 1 strains." J Virol **70**(9): 6288-95.

Binley, J. M., T. Wrin, et al. (2004). "Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies." J Virol **78**(23): 13232-52.

Bjorndal, A., A. Sonnerborg, et al. (1999). "Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients." AIDS Res Hum Retroviruses **15**(7): 647-53.

Bleul, C. C., M. Farzan, et al. (1996). "The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry." Nature **382**(6594): 829-33.

Briggs, D. R., D. L. Tuttle, et al. (2000). "Envelope V3 amino acid sequence predicts HIV-1 phenotype (co-receptor usage and tropism for macrophages)." Aids **14**(18): 2937-9.

Brumme, Z. L., W. W. Dong, et al. (2004). "Clinical and immunological impact of HIV envelope V3 sequence variation after starting initial triple antiretroviral therapy." Aids **18**(4): F1-9.

Bures, R., L. Morris, et al. (2002). "Regional clustering of shared neutralization determinants on primary isolates of clade C human immunodeficiency virus type 1 from South Africa." J Virol **76**(5): 2233-44.

Burns, D. P. and R. C. Desrosiers (1994). "Envelope sequence variation, neutralizing antibodies, and primate lentivirus persistence." Curr Top Microbiol Immunol **188**: 185-219.

Callahan, L. N., M. Phelan, et al. (1991). "Dextran sulfate blocks antibody binding to the principal neutralizing domain of human immunodeficiency virus type 1 without interfering with gp120-CD4 interactions." J Virol **65**(3): 1543-50.

Cao, J., L. Bergeron, et al. (1993). "Effects of amino acid changes in the extracellular domain of the human immunodeficiency virus type 1 gp41 envelope glycoprotein." J Virol **67**(5): 2747-55.

Carrillo, A. and L. Ratner (1996). "Human immunodeficiency virus type 1 tropism for T-lymphoid cell lines: role of the V3 loop and C4 envelope determinants." J Virol **70**(2): 1301-9.

Chan, D. C., D. Fass, et al. (1997). "Core structure of gp41 from the HIV envelope glycoprotein." Cell **89**(2): 263-73.

Cho, M. W., M. K. Lee, et al. (1998). "Identification of determinants on a dualtropic human immunodeficiency virus type 1 envelope glycoprotein that confer usage of CXCR4." J Virol **72**(3): 2509-15.

Choge, I., T. Cilliers, et al. (2005). "Genotypic and phenotypic characterization of viral isolates from HIV-1 subtype C infected children with slow and rapid disease progression." AIDS Res Hum Retroviruses.

Cilliers, T., J. Nhlapo, et al. (2003). "The CCR5 and CXCR4 coreceptors are both used by human immunodeficiency virus type 1 primary isolates from subtype C." J Virol **77**(7): 4449-56.

Cilliers, T., T. Patience, et al. (2004). "Sensitivity of HIV type 1 subtype C isolates to the entry inhibitor T-20." AIDS Res Hum Retroviruses **20**(5): 477-82.

Cilliers, T., S. Willey, et al. (2005). "Use of alternate coreceptors on primary cells by two HIV-1 isolates." Virology.

Clerici, M., S. Butto, et al. (2000). "Immune activation in africa is environmentally-driven and is associated with upregulation of CCR5. Italian-Ugandan AIDS Project." Aids **14**(14): 2083-92.

Cocchi, F., A. L. DeVico, et al. (1995). "Identification of RANTES, MIP-1 alpha, and MIP-1 beta as the major HIV-suppressive factors produced by CD8+ T cells." Science **270**(5243): 1811-5.

Coetzer, M., T. Cilliers, et al. (2005). "What genetic changes in V3 are associated with CXCR4 usage in HIV-1 subtype C isolates?" Submitted.

Collins, K. R., H. Mayanja-Kizza, et al. (2000). "Greater diversity of HIV-1 quasispecies in HIV-infected individuals with active tuberculosis." J Acquir Immune Defic Syndr **24**(5): 408-17.

Connor, R. I., K. E. Sheridan, et al. (1997). "Change in coreceptor use coreceptor use correlates with disease progression in HIV-1--infected individuals." J Exp Med **185**(4): 621-8.

Cooper, D. A., B. Tindall, et al. (1988). "Characterization of T lymphocyte responses during primary infection with human immunodeficiency virus." J Infect Dis **157**(5): 889-96.

Cormier, E. G. and T. Dragic (2002). "The crown and stem of the V3 loop play distinct roles in human immunodeficiency virus type 1 envelope glycoprotein interactions with the CCR5 coreceptor." J Virol **76**(17): 8953-7.

Cornelissen, M., G. Mulder-Kampinga, et al. (1995). "Syncytium-inducing (SI) phenotype suppression at seroconversion after intramuscular inoculation of a

non-syncytium-inducing/SI phenotypically mixed human immunodeficiency virus population." J Virol **69**(3): 1810-8.

Daar, E. S., T. Moudgil, et al. (1991). "Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection." N Engl J Med **324**(14): 961-4.

Day, J. H., A. D. Grant, et al. (2004). "Does tuberculosis increase HIV load?" J Infect Dis **190**(9): 1677-84.

De Jong, J. J., A. De Ronde, et al. (1992). "Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution." J Virol **66**(11): 6777-80.

de Roda Husman, A. M., R. P. van Rij, et al. (1999). "Adaptation to promiscuous usage of chemokine receptors is not a prerequisite for human immunodeficiency virus type 1 disease progression." J Infect Dis **180**(4): 1106-15.

De Wolf, F., E. Hogervorst, et al. (1994). "Syncytium-inducing and non-syncytium-inducing capacity of human immunodeficiency virus type 1 subtypes other than B: phenotypic and genotypic characteristics. WHO Network for HIV Isolation and Characterization." AIDS Res Hum Retroviruses **10**(11): 1387-400.

Del Amo, J., A. S. Malin, et al. (1999). "Does tuberculosis accelerate the progression of HIV disease? Evidence from basic science and epidemiology." Aids **13**(10): 1151-8.

Delobel, P., K. Sandres-Saune, et al. (2005). "R5 to X4 switch of the predominant HIV-1 population in cellular reservoirs during effective highly active antiretroviral therapy." J Acquir Immune Defic Syndr **38**(4): 382-92.

Delwart, E. L., H. Pan, et al. (1997). "Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS." J Virol **71**(10): 7498-508.

Deng, H., R. Liu, et al. (1996). "Identification of a major co-receptor for primary isolates of HIV-1." Nature **381**(6584): 661-6.

Derdeyn, C. A., J. M. Decker, et al. (2004). "Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission." Science **303**(5666): 2019-22.

Derdeyn, C. A., J. M. Decker, et al. (2000). "Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor specificity defined by the V3 loop of gp120." J Virol **74**(18): 8358-67.

Doranz, B. J., J. Rucker, et al. (1996). "A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors." Cell **85**(7): 1149-58.

Dragic, T., V. Litwin, et al. (1996). "HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5." Nature **381**(6584): 667-73.

Dragic, T., A. Trkola, et al. (1998). "Amino-terminal substitutions in the CCR5 coreceptor impair gp120 binding and human immunodeficiency virus type 1 entry." J Virol **72**(1): 279-85.

Drosopoulos, W. C., L. F. Rezende, et al. (1998). "Virtues of being faithful: can we limit the genetic variation in human immunodeficiency virus?" J Mol Med **76**(9): 604-12.

Dybul, M., M. Daucher, et al. (2003). "Genetic characterization of rebounding human immunodeficiency virus type 1 in plasma during multiple interruptions of highly active antiretroviral therapy." J Virol **77**(5): 3229-37.

Edinger, A. L., T. L. Hoffman, et al. (1998). "Use of GPR1, GPR15, and STRL33 as coreceptors by diverse human immunodeficiency virus type 1 and simian immunodeficiency virus envelope proteins." Virology **249**(2): 367-78.

Farzan, M., H. Choe, et al. (1998). "A tyrosine-rich region in the N terminus of CCR5 is important for human immunodeficiency virus type 1 entry and mediates an association between gp120 and CCR5." J Virol **72**(2): 1160-4.

Feng, Y., C. C. Broder, et al. (1996). "HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor [see comments]." Science **272**(5263): 872-7.

Fields, B. N. and D. M. Knipe, Eds. (1990). Fields Virology. New York, Raven Press.

Flynn, J. L. and J. Chan (2001). "Immunology of tuberculosis." Annu Rev Immunol **19**: 93-129.

Fouchier, R. A., S. M. Broersen, et al. (1995). "Temporal relationship between elongation of the HIV type 1 glycoprotein 120 V2 domain and the conversion toward a syncytium-inducing phenotype." AIDS Res Hum Retroviruses **11**(12): 1473-8.

Fouchier, R. A., M. Brouwer, et al. (1995). "Simple determination of human immunodeficiency virus type 1 syncytium-inducing V3 genotype by PCR." J Clin Microbiol **33**(4): 906-11.

Fouchier, R. A., M. Groenink, et al. (1992). "Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule." J Virol **66**(5): 3183-7.

Gao, F., E. Bailes, et al. (1999). "Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes." Nature **397**(6718): 436-41.

Gao, F., S. G. Morrison, et al. (1996). "Molecular cloning and analysis of functional envelope genes from human immunodeficiency virus type 1 sequence subtypes A through G. The WHO and NIAID Networks for HIV Isolation and Characterization." J Virol **70**(3): 1651-67.

Gao, F., L. Yue, et al. (1992). "Human infection by genetically diverse SIVSM-related HIV-2 in west Africa." Nature **358**(6386): 495-9.

Gervaix, A., J. Nicolas, et al. (2002). "Response to treatment and disease progression linked to CD4+ T cell surface CC chemokine receptor 5 density in human immunodeficiency virus type 1 vertical infection." J Infect Dis **185**(8): 1055-61.

Glushakova, S., J. C. Grivel, et al. (1998). "Evidence for the HIV-1 phenotype switch as a causal factor in acquired immunodeficiency." Nat Med **4**(3): 346-9.

Goletti, D., D. Weissman, et al. (1996). "Effect of Mycobacterium tuberculosis on HIV replication. Role of immune activation." J Immunol **157**(3): 1271-8.

Gottlieb, G. S., D. C. Nickle, et al. (2004). "Dual HIV-1 infection associated with rapid disease progression." Lancet **363**(9409): 619-22.

Gottlieb, M. S., R. Schroff, et al. (1981). "Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency." N Engl J Med **305**(24): 1425-31.

Goudsmit, J., C. Debouck, et al. (1988). "Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees." Proc Natl Acad Sci U S A **85**(12): 4478-82.

Grant, A. D. and K. M. De Cock (2001). "ABC of AIDS. HIV infection and AIDS in the developing world." Bmj **322**(7300): 1475-8.

Greene, W. C. (1993). "AIDS and the immune system." Sci Am **269**(3): 98-105.

Gribskov, M., A. D. McLachlan, et al. (1987). "Profile analysis: detection of distantly related proteins." Proc Natl Acad Sci U S A **84**(13): 4355-8.

Grobler, J., C. M. Gray, et al. (2004). "Incidence of HIV-1 dual infection and its association with increased viral load set point in a cohort of HIV-1 subtype C-infected female sex workers." J Infect Dis **190**(7): 1355-9.

Groenink, M., R. A. Fouchier, et al. (1993). "Relation of phenotype evolution of HIV-1 to envelope V2 configuration [see comments]." Science **260**(5113): 1513-6.

Guillon, C., M. E. van der Ende, et al. (1998). "Coreceptor usage of human immunodeficiency virus type 2 primary isolates and biological clones is broad and does not correlate with their syncytium-inducing capacities." J Virol **72**(7): 6260-3.

Hartley, O., P. J. Klasse, et al. (2005). "V3: HIV's switch-hitter." AIDS Res Hum Retroviruses **21**(2): 171-89.

Henikoff, S., J. C. Wallace, et al. (1990). "Finding protein similarities with nucleotide sequence databases." Methods Enzymol **183**: 111-32.

Hoffman, N. G., F. Seillier-Moiseiwitsch, et al. (2002). "Variability in the human immunodeficiency virus type 1 gp120 Env protein linked to phenotype-associated changes in the V3 loop." J Virol **76**(8): 3852-64.

Hogervorst, E., J. de Jong, et al. (1995). "Insertion of primary syncytium-inducing (SI) and non-SI envelope V3 loops in human immunodeficiency virus type 1 (HIV-1) LAI reduces neutralization sensitivity to autologous, but not heterologous, HIV-1 antibodies." J Virol **69**(10): 6342-51.

Hu, Q., J. O. Trent, et al. (2000). "Identification of ENV determinants in V3 that influence the molecular anatomy of CCR5 utilization." J Mol Biol **302**(2): 359-75.

Hu, Q. X., A. P. Barry, et al. (2000). "Evolution of the human immunodeficiency virus type 1 envelope during infection reveals molecular corollaries of specificity for coreceptor utilization and AIDS pathogenesis." J Virol **74**(24): 11858-72.

Huang, C. C., M. Tang, et al. (2005). "Structure of a V3-containing HIV-1 gp120 core." Science **310**(5750): 1025-8.

Jansson, M., E. Backstrom, et al. (2001). "Length variation of glycoprotein 120 V2 region in relation to biological phenotypes and coreceptor usage of primary HIV type 1 isolates." AIDS Res Hum Retroviruses **17**(15): 1405-14.

Jensen, M. A., M. Coetzer, et al. (2005). "A Reliable Phenotype Predictor for HIV-1 Subtype C Based on Envelope V3 Sequences." Submitted.

Jensen, M. A., F. S. Li, et al. (2003). "Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences." J Virol **77**(24): 13376-88.

Jensen, M. A. and A. B. van 't Wout (2003). "Predicting HIV-1 coreceptor usage with sequence analysis." AIDS Rev **5**(2): 104-12.

Johnston, E. R., L. S. Zijenah, et al. (2003). "High frequency of syncytium-inducing and CXCR4-tropic viruses among human immunodeficiency virus type 1 subtype C-infected patients receiving antiretroviral treatment." J Virol **77**(13): 7682-8.

Kajumo, F., D. A. Thompson, et al. (2000). "Entry of R5X4 and X4 human immunodeficiency virus type 1 strains is mediated by negatively charged and tyrosine residues in the amino-terminal domain and the second extracellular loop of CXCR4." Virology **271**(2): 240-7.

Kalpana, A., T. Srikanth, et al. (2004). "gp120 sequences from HIV type 1 subtype C early seroconverters in India." AIDS Res Hum Retroviruses **20**(8): 889-94.

Keulen, W., M. Nijhuis, et al. (1997). "Reverse transcriptase fidelity and HIV-1 variation." Science **275**(5297): 229; author reply 230-1.

Kilby, J. M., S. Hopkins, et al. (1998). "Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry." Nat Med **4**(11): 1302-7.

Koito, A., G. Harrowe, et al. (1994). "Functional role of the V1/V2 region of human immunodeficiency virus type 1 envelope glycoprotein gp120 in infection of primary macrophages and soluble CD4 neutralization." J Virol **68**(4): 2253-9.

Koning, F., R. van Rij, et al. (2002). Biological and Molecular Aspects of HIV-1 Coreceptor Usage. HIV Sequence Compendium 2002. C. Kuiken, B. Foley, E. Freedet al. Los Alamos, Theoretical Biology and Biophysics Group**:** pp24-42.

Koot, M., I. P. Keet, et al. (1993). "Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS." Ann Intern Med **118**(9): 681-8.

Korber, B., S. Wolinsky, et al. (1992). "HIV-1 intrapatient sequence diversity in the immunogenic V3 region." AIDS Res Hum Retroviruses **8**(8): 1461-5.

Koup, R. A., J. T. Safrit, et al. (1994). "Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome." J Virol **68**(7): 4650-5.

Kuhmann, S. E., P. Pugach, et al. (2004). "Genetic and phenotypic analyses of human immunodeficiency virus type 1 escape from a small-molecule CCR5 inhibitor." J Virol **78**(6): 2790-807.

Kuiken, C. L., J. J. de Jong, et al. (1992). "Evolution of the V3 envelope domain in proviral sequences and isolates of human immunodeficiency virus type 1 during transition of the viral biological phenotype." J Virol **66**(9): 5704.

Kumar, S., K. Tamura, et al. (2001). "MEGA2: molecular evolutionary genetics analysis software." Bioinformatics **17**(12): 1244-5.

Kwong, P. D., R. Wyatt, et al. (1998). "Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody." Nature **393**(6686): 648-59.

Labrosse, B., A. Brelot, et al. (1998). "Determinants for sensitivity of human immunodeficiency virus coreceptor CXCR4 to the bicyclam AMD3100." J Virol **72**(8): 6381-8.

Land, A. and I. Braakman (2001). "Folding of the human immunodeficiency virus type 1 envelope glycoprotein in the endoplasmic reticulum." Biochimie **83**(8): 783-90.

LaRosa, G. J., J. P. Davide, et al. (1990). "Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant." Science **249**(4971): 932-5.

Lawn, S. D., R. J. Shattock, et al. (1999). "Sustained plasma TNF-alpha and HIV-1 load despite resolution of other parameters of immune activation during treatment of tuberculosis in Africans." Aids **13**(16): 2231-7.

Lazzarin, A., B. Clotet, et al. (2003). "Efficacy of enfuvirtide in patients infected with drug-resistant HIV-1 in Europe and Australia." N Engl J Med **348**(22): 2186-95.

Leonard, C. K., M. W. Spellman, et al. (1990). "Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells." J Biol Chem **265**(18): 10373-82.

Loetscher, P., B. Moser, et al. (2000). "Chemokines and their receptors in lymphocyte traffic and HIV infection." Adv Immunol **74**: 127-80.

Lu, Z., J. F. Berson, et al. (1997). "Evolution of HIV-1 coreceptor usage through interactions with distinct CCR5 and CXCR4 domains." Proc Natl Acad Sci U S A **94**(12): 6426-31.

Lukashov, V. V. and J. Goudsmit (1998). "HIV heterogeneity and disease progression in AIDS: a model of continuous virus adaptation." Aids **12**(Suppl A): S43-52.

Masciotra, S., S. M. Owen, et al. (2002). "Temporal relationship between V1V2 variation, macrophage replication, and coreceptor adaptation during HIV-1 disease progression." Aids **16**(14): 1887-98.

McKnight, A., M. T. Dittmar, et al. (1998). "A broad range of chemokine receptors are used by primary isolates of human immunodeficiency virus type 2 as coreceptors with CD4." J Virol **72**(5): 4065-71.

McNearney, T., Z. Hornickova, et al. (1992). "Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease." Proc Natl Acad Sci U S A **89**(21): 10247-51.

Mellors, J. W., C. R. Rinaldo, Jr., et al. (1996). "Prognosis in HIV-1 infection predicted by the quantity of virus in plasma." Science **272**(5265): 1167-70.

Michael, N. L., G. Chang, et al. (1997). "The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression." Nat Med **3**(3): 338-40.

Milich, L., B. Margolin, et al. (1993). "V3 loop of the human immunodeficiency virus type 1 Env protein: interpreting sequence variability." J Virol **67**(9): 5623-34.

Milich, L., B. H. Margolin, et al. (1997). "Patterns of amino acid variability in NSI-like and SI-like V3 sequences and a linked change in the CD4-binding domain of the HIV-1 Env protein." Virology **239**(1): 108-18.

Moore, J. P., Y. Cao, et al. (1994). "Development of the anti-gp120 antibody response during seroconversion to human immunodeficiency virus type 1." J Virol **68**(8): 5142-55.

Moore, J. P. and R. W. Doms (2003). "The entry of entry inhibitors: a fusion of science and medicine." Proc Natl Acad Sci U S A **100**(19): 10598-602.

Moore, J. P., S. G. Kitchen, et al. (2004). "The CCR5 and CXCR4 coreceptors--central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection." AIDS Res Hum Retroviruses **20**(1): 111-26.

Moore, J. P., Q. J. Sattentau, et al. (1994). "Probing the structure of the human immunodeficiency virus surface glycoprotein gp120 with a panel of monoclonal antibodies." J Virol **68**(1): 469-84.

Moore, J. P. and J. Sodroski (1996). "Antibody cross-competition analysis of the human immunodeficiency virus type 1 gp120 exterior envelope glycoprotein." J Virol **70**(3): 1863-72.

Moore, J. P. and M. Stevenson (2000). "New targets for inhibitors of HIV-1 replication." Nat Rev Mol Cell Biol **1**(1): 40-9.

Morris, L., T. Cilliers, et al. (2001). "CCR5 is the major coreceptor used by HIV-1 subtype C isolates from patients with active tuberculosis." AIDS Res Hum Retroviruses **17**(8): 697-701.

Morris, L., D. J. Martin, et al. (2003). "Human immunodeficiency virus-1 RNA levels and CD4 lymphocyte counts, during treatment for active tuberculosis, in South African patients." J Infect Dis **187**(12): 1967-71.

Moulard, M., S. K. Phogat, et al. (2002). "Broadly cross-reactive HIV-1-neutralizing human monoclonal Fab selected for binding to gp120-CD4-CCR5 complexes." Proc Natl Acad Sci U S A **99**(10): 6913-8.

Munoz, A., M. C. Wang, et al. (1989). "Acquired immunodeficiency syndrome (AIDS)-free time after human immunodeficiency virus type 1 (HIV-1) seroconversion in homosexual men. Multicenter AIDS Cohort Study Group." Am J Epidemiol **130**(3): 530-9.

Muster, T., F. Steindl, et al. (1993). "A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1." J Virol **67**(11): 6642-7.

Nabatov, A. A., G. Pollakis, et al. (2004). "Intrapatient alterations in the human immunodeficiency virus type 1 gp120 V1V2 and V3 regions differentially modulate coreceptor usage, virus inhibition by CC/CXC chemokines, soluble CD4, and the b12 and 2G12 monoclonal antibodies." J Virol **78**(1): 524-30.

Nelson, J. A., F. Baribaud, et al. (2000). "Patterns of changes in human immunodeficiency virus type 1 V3 sequence populations late in infection." J Virol **74**(18): 8494-501.

Nelson, J. A., S. A. Fiscus, et al. (1997). "Evolutionary variants of the human immunodeficiency virus type 1 V3 region characterized by using a heteroduplex tracking assay." J Virol **71**(11): 8750-8.

Nicholson, J. K., S. W. Browning, et al. (2001). "CCR5 and CXCR4 expression on memory and naive T cells in HIV-1 infection and response to highly active antiretroviral therapy." J Acquir Immune Defic Syndr **27**(2): 105-15.

Nowak, P., A. C. Karlsson, et al. (2002). "The selection and evolution of viral quasispecies in HIV-1 infected children." HIV Med **3**(1): 1-11.

Oberlin, E., A. Amara, et al. (1996). "The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1." Nature **382**(6594): 833-5.

Ogert, R. A., M. K. Lee, et al. (2001). "N-linked glycosylation sites adjacent to and within the V1/V2 and the V3 loops of dualtropic human immunodeficiency virus type 1 isolate DH12 gp120 affect coreceptor usage and cellular tropism." J Virol **75**(13): 5998-6006.

Ostrowski, M. A., D. C. Krakauer, et al. (1998). "Effect of immune activation on the dynamics of human immunodeficiency virus replication and on the distribution of viral quasispecies." J Virol **72**(10): 7772-84.

Pantaleo, G., C. Graziosi, et al. (1993). "HIV infection is active and progressive in lymphoid tissue during the clinically latent stage of disease." Nature **362**(6418): 355-8.

Pastore, C., A. Ramos, et al. (2004). "Intrinsic obstacles to human immunodeficiency virus type 1 coreceptor switching." J Virol **78**(14): 7565-74.

Perelson, A. S., A. U. Neumann, et al. (1996). "HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time." Science **271**(5255): 1582-6.

Pierdominici, M., A. Giovannetti, et al. (2002). "Changes in CCR5 and CXCR4 expression and beta-chemokine production in HIV-1-infected patients treated

with highly active antiretroviral therapy." J Acquir Immune Defic Syndr **29**(2): 122-31.

Pillai, S., B. Good, et al. (2003). "A new perspective on V3 phenotype prediction." AIDS Res Hum Retroviruses **19**(2): 145-9.

Ping, L. H., J. A. Nelson, et al. (1999). "Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants." J Virol **73**(8): 6271-81.

Pohlmann, S., M. Krumbiegel, et al. (1999). "Coreceptor usage of BOB/GPR15 and Bonzo/STRL33 by primary isolates of human immunodeficiency virus type 1." J Gen Virol **80 (Pt 5)**: 1241-51.

Pollakis, G., A. Abebe, et al. (2004). "Phenotypic and genotypic comparisons of CCR5- and CXCR4-tropic human immunodeficiency virus type 1 biological clones isolated from subtype C-infected individuals." J Virol **78**(6): 2841-52.

Pollakis, G., S. Kang, et al. (2001). "N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization." J Biol Chem **276**(16): 13433-41.

Polzer, S., M. T. Dittmar, et al. (2002). "The N-linked glycan g15 within the V3 loop of the HIV-1 external glycoprotein gp120 affects coreceptor usage, cellular tropism, and neutralization." Virology **304**(1): 70-80.

Porco, T. C., P. M. Small, et al. (2001). "Amplification dynamics: predicting the effect of HIV on tuberculosis outbreaks." J Acquir Immune Defic Syndr **28**(5): 437-44.

Preston, B. D., B. J. Poiesz, et al. (1988). "Fidelity of HIV-1 reverse transcriptase." Science **242**(4882): 1168-71.

Rabut, G. E., J. A. Konner, et al. (1998). "Alanine substitutions of polar and nonpolar residues in the amino-terminal domain of CCR5 differently impair entry of macrophage- and dualtropic isolates of human immunodeficiency virus type 1." J Virol **72**(4): 3464-8.

Rambaut, A., D. Posada, et al. (2004). "The causes and consequences of HIV evolution." Nat Rev Genet **5**(1): 52-61.

Raviglione, M. C., A. D. Harries, et al. (1997). "Tuberculosis and HIV: current status in Africa." Aids **11 Suppl B**: S115-23.

Reeves, J. D., S. A. Gallo, et al. (2002). "Sensitivity of HIV-1 to entry inhibitors correlates with envelope/coreceptor affinity, receptor density, and fusion kinetics." Proc Natl Acad Sci U S A **99**(25): 16249-54.

Resch, W., N. Hoffman, et al. (2001). "Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks." Virology **288**(1): 51-62.

Richman, D. D. and S. A. Bozzette (1994). "The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression." J Infect Dis **169**(5): 968-74.

Richman, D. D., T. Wrin, et al. (2003). "Rapid evolution of the neutralizing antibody response to HIV type 1 infection." Proc Natl Acad Sci U S A **100**(7): 4144-9.

Rucker, J., M. Samson, et al. (1996). "Regions in beta-chemokine receptors CCR5 and CCR2b that determine HIV-1 cofactor specificity." Cell **87**(3): 437-46.

Russell, D. G. (2001). "TB comes to a sticky beginning." Nat Med **7**(8): 894-5.

Safrit, J. T., A. Y. Lee, et al. (1994). "A region of the third variable loop of HIV-1 gp120 is recognized by HLA-B7-restricted CTLs from two acute seroconversion patients." J Immunol **153**(8): 3822-30.

Sanders, R. W., M. Venturi, et al. (2002). "The mannose-dependent epitope for neutralizing antibody 2G12 on human immunodeficiency virus type 1 glycoprotein gp120." J Virol **76**(14): 7293-305.

Santiago, M. L., C. M. Rodenburg, et al. (2002). "SIVcpz in wild chimpanzees." Science **295**(5554): 465.

Sattentau, Q. J., J. P. Moore, et al. (1993). "Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding." J Virol **67**(12): 7383-93.

Sattentau, Q. J., S. Zolla-Pazner, et al. (1995). "Epitope exposure on functional, oligomeric HIV-1 gp41 molecules." Virology **206**(1): 713-7.

Scanlan, C. N., R. Pantophlet, et al. (2002). "The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of alpha1--->2 mannose residues on the outer face of gp120." J Virol **76**(14): 7306-21.

Scarlatti, G., E. Tresoldi, et al. (1997). "In vivo evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression." Nat Med **3**(11): 1259-65.

Schneider, T. D. and R. M. Stephens (1990). "Sequence logos: a new way to display consensus sequences." Nucleic Acids Res **18**(20): 6097-100.

Schuitemaker, H., M. Koot, et al. (1992). "Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population." J Virol **66**(3): 1354-60.

Shankarappa, R., J. B. Margolick, et al. (1999). "Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection." J Virol **73**(12): 10489-502.

Sheppard, H. W., C. Celum, et al. (2002). "HIV-1 infection in individuals with the CCR5-Delta32/Delta32 genotype: acquisition of syncytium-inducing virus at seroconversion." J Acquir Immune Defic Syndr **29**(3): 307-13.

Shimizu, N., Y. Haraguchi, et al. (1999). "Changes in and discrepancies between cell tropisms and coreceptor uses of human immunodeficiency virus type 1 induced by single point mutations at the V3 tip of the env protein." Virology **259**(2): 324-33.

Shioda, T., J. A. Levy, et al. (1991). "Macrophage and T cell-line tropisms of HIV-1 are determined by specific regions of the envelope gp120 gene." Nature **349**(6305): 167-9.

Shioda, T., J. A. Levy, et al. (1992). "Small amino acid changes in the V3 hypervariable region of gp120 can affect the T-cell-line and macrophage tropism of human immunodeficiency virus type 1." Proc Natl Acad Sci U S A **89**(20): 9434-8.

Shioda, T., S. Oka, et al. (1997). "In vivo sequence variability of human immunodeficiency virus type 1 envelope gp120: association of V2 extension with slow disease progression." J Virol **71**(7): 4871-81.

Sing, T., N. Beerenwinkel, et al. (2004). Learning mixtures of localized rules by maximizing the area under the ROC curve. 16th European Conference on Artificial Intelligence (ECAI).

Smyth, R. J., Y. Yi, et al. (1998). "Determinants of entry cofactor utilization and tropism in a dualtropic human immunodeficiency virus type 1 primary isolate." J Virol **72**(5): 4478-84.

Stiegler, G., R. Kunert, et al. (2001). "A potent cross-clade neutralizing human monoclonal antibody against a novel epitope on gp41 of human

immunodeficiency virus type 1." AIDS Res Hum Retroviruses **17**(18): 1757-65.

Sulkowski, M. (1998). "HIV and hepatitis C virus co-infection." Hopkins HIV Rep **10**(6): 8, 12.

Sulkowski, M. S., R. E. Chaisson, et al. (1998). "The effect of acute infectious illnesses on plasma human immunodeficiency virus (HIV) type 1 load and the expression of serologic markers of immune activation among HIV-infected adults." J Infect Dis **178**(6): 1642-8.

Suphaphiphat, P., A. Thitithanyanont, et al. (2003). "Effect of amino acid substitution of the V3 and bridging sheet residues in human immunodeficiency virus type 1 subtype C gp120 on CCR5 utilization." J Virol **77**(6): 3832-7.

Taylor, N. (2004). Neutralizing antibody levels in HIV-1 subtype C infected individuals. Manuscript towards M.Sc. Johannesburg, South Africa.

Thompson, J. D., T. J. Gibson, et al. (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Res **25**(24): 4876-82.

Toossi, Z., J. L. Johnson, et al. (2001). "Increased replication of HIV-1 at sites of Mycobacterium tuberculosis infection: potential mechanisms of viral activation." J Acquir Immune Defic Syndr **28**(1): 1-8.

Troyer, R. M., K. R. Collins, et al. (2005). "Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression." J Virol **79**(14): 9006-18.

Tscherning, C., A. Alaeus, et al. (1998). "Differences in chemokine coreceptor usage between genetic subtypes of HIV-1." Virology **241**(2): 181-8.

van 't Wout, A. B., H. Blaak, et al. (1998). "Evolution of syncytium-inducing and non-syncytium-inducing biological virus clones in relation to replication kinetics during the course of human immunodeficiency virus type 1 infection." J Virol **72**(6): 5099-107.

Wallis, R. S., M. Vjecha, et al. (1993). "Influence of tuberculosis on human immunodeficiency virus (HIV-1): enhanced cytokine expression and elevated beta 2-microglobulin in HIV-1- associated tuberculosis." J Infect Dis **167**(1): 43-8.

Watkins, B. A., M. S. Reitz, Jr., et al. (1993). "Immune escape by human immunodeficiency virus type 1 from neutralizing antibodies: evidence for multiple pathways." J Virol **67**(12): 7493-500.

Wei, X., J. M. Decker, et al. (2003). "Antibody neutralization and escape by HIV-1." Nature **422**(6929): 307-12.

Weissenhorn, W., A. Dessen, et al. (1997). "Atomic structure of the ectodomain from HIV-1 gp41." Nature **387**(6631): 426-30.

Whalen, C., C. R. Horsburgh, et al. (1995). "Accelerated course of human immunodeficiency virus infection after tuberculosis." Am J Respir Crit Care Med **151**(1): 129-35.

Wild, C. T., D. C. Shugars, et al. (1994). "Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection." Proc Natl Acad Sci U S A **91**(21): 9770-4.

Willey, S. J., J. D. Reeves, et al. (2003). "Identification of a subset of human immunodeficiency virus type 1 (HIV-1), HIV-2, and simian immunodeficiency virus strains able to exploit an alternative coreceptor on untransformed human brain and lymphoid cells." J Virol **77**(11): 6138-52.

Williamson, C., L. Morris, et al. (2003). "Characterization and selection of HIV-1 subtype C isolates for use in vaccine development." AIDS Res Hum Retroviruses **19**(2): 133-44.

Wolfs, T. F., G. Zwart, et al. (1991). "Naturally occurring mutations within HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino acid substitution." Virology **185**(1): 195-205.

Wolinsky, S. M., B. T. Korber, et al. (1996). "Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection." Science **272**(5261): 537-42.

Wu, L., G. LaRosa, et al. (1997). "Interaction of chemokine receptor CCR5 with its ligands: multiple domains for HIV-1 gp120 binding and a single domain for chemokine binding." J Exp Med **186**(8): 1373-81.

Wyatt, R., P. D. Kwong, et al. (1998). Structure of the core of the HIV-1 gp120 exterior envelope glycoprotein. Human Retroviruses and AIDS. B. Korber, C. Kuiken, B. Foleyet al. Los Alamos, Theoretical Biology and Biophysics Group**:** pp III-3-9.

Wyatt, R. and J. Sodroski (1998). "The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens." Science **280**(5371): 1884-8.

Ye, Y., Z. H. Si, et al. (2000). "Association of structural changes in the V2 and V3 loops of the gp120 envelope glycoprotein with acquisition of neutralization resistance in a simian-human immunodeficiency virus passaged in vivo." J Virol **74**(24): 11955-62.

Yi, Y., F. Shaheen, et al. (2005). "Preferential use of CXCR4 by R5X4 human immunodeficiency virus type 1 isolates for infection of primary lymphocytes." J Virol **79**(3): 1480-6.

Yoshida, K., M. Nakamura, et al. (1997). "Mutations of the HIV type 1 V3 loop under selection pressure with neutralizing monoclonal antibody NM-01." AIDS Res Hum Retroviruses **13**(15): 1283-90.

Zhang, L. Q., P. MacKenzie, et al. (1993). "Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection." J Virol **67**(6): 3345-56.

Zhang, M., B. Gaschen, et al. (2004). "Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin." Glycobiology **14**(12): 1229-46.

Zhang, Y. J., T. Dragic, et al. (1998). "Use of coreceptors other than CCR5 by non-syncytium-inducing adult and pediatric isolates of human immunodeficiency virus type 1 is rare in vitro." J Virol **72**(11): 9337-44.

Zhu, X., C. Borchers, et al. (2000). "Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells." Biochemistry **39**(37): 11194-204.

Zhuang, J., A. E. Jetzt, et al. (2002). "Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots." J Virol **76**(22): 11273-82.

Zolla-Pazner, S. (2004). "Identifying epitopes of HIV-1 that induce protective antibodies." Nat Rev Immunol **4**(3): 199-210.

Zwick, M. B., R. Jensen, et al. (2005). "Anti-human immunodeficiency virus type 1 (HIV-1) antibodies 2F5 and 4E10 require surprisingly few crucial residues in the membrane-proximal external region of glycoprotein gp41 to neutralize HIV-1." J Virol **79**(2): 1252-61.