

CHARACTERISATION OF THE HIV-1 SUBTYPE C ENV GENE AND THE EXPRESSION OF THE ENV PROTEIN FROM SELECTED ISOLATES IN MAMMALIAN CELLS

TANIA DE VILLIERS

Thesis presented in partial fulfilment of the requirements for the degree of Master
of Sciences in Medical Sciences (Medical Virology) at the Faculty of Health
Sciences, University of Stellenbosch.



PROMOTER: Prof. E Janse van Rensburg
CO-PROMOTER: Associate Prof. S Engelbrecht

APRIL 2003

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature:

Date:

SUMMARY

At the end of 2002, human immunodeficiency virus (HIV) had infected 42 million people worldwide. The morbidity and mortality rate, as well as the epidemic proportions of the disease have led to concentrated scientific efforts to reveal the disease's pathogenesis and develop effective preventative and treatment measures. Advances have been made to inhibit viral replication by suppressing the virus' ability to replicate by developing anti-retroviral treatments, although development of a safe and effective vaccine is the only way to stem the pandemic. Advances in vaccine design, animal models and clinical research have led to the creation of promising candidate vaccines to counter this rampage, but most of these vaccines entering phase I-III clinical trials are based mainly on subtype B genomes. HIV-1 subtype C is the most commonly transmitted subtype worldwide, and is the predominant subtype in India, China, East and Southern Africa. A subtype C vaccine is critical for the developing nations such as South Africa, where anti-retroviral therapies are largely unaffordable. The envelope gene (*env*) is an attractive target as immunogen to be included in a HIV vaccine. The envelope protein (Env) elicits neutralising antibodies and cytotoxic T-lymphocyte (CTL) responses. This protein will therefore be useful in creating a humoral and cellular immune response in the host. A shortage in characterised subtype C *env* gene sequences from South Africa was recognised, and this study focussed on the characterisation of generated sequences, as well as the expression of selected *env* genes. These immunogens were created for possible use in a prime-boost vaccine modality. The *env* genes from recent circulating strains in South Africa were amplified by polymerase chain reaction (PCR). The genes were then cloned for sequencing and expression purposes. Phylogenetic relationships were determined by comparing the sequences to reference subtype strains and subtype C strains. Expression of the genes was assessed by Western Blot in 293 cells with HIV-1 positive patient sera.

Sequence analysis showed a more conserved third variable (V3) loop in South African subtype C sequences, with a more variable region downstream from the loop. The crown sequence (GPGQ) and positions of uncharged or negatively charged residues in

the V3 loop indicated a non-syncytium-inducing (NSI) phenotype for the isolates. Phylogenetic analysis showed the sequences to all belong to the C subtype, and further that the sequences were not recombinant, which was confirmed by recombination analysis. The intersample diversity observed for strains from South Africa was significantly higher than distances observed to the subtype C consensus sequence. The South African sequences were distributed across several subclusters in a subtype C phylogenetic tree, highlighting the concept that these infections represent a more longstanding epidemic with multiple introductions from different geographic areas. Western Blot with HIV-1 positive patient sera showed the expression of uncleaved gp160 Env proteins, which were Rev dependent.

This study has generated much needed subtype C South African *env* gene sequences that can be used as basis for modification for use as immunogens in a South African vaccine.

OPSOMMING

Teen die einde van 2002 was 42 miljoen mense wêreldwyd geïnfekteer met die menslike immuniteitsgebrek virus (MIV). Die dode- en sterfte syfers, asook die skaal van die epidemie, het gelei tot 'n wetenskaplike poging om die siekte se patogene te openbaar en om effektiewe voorkomende en terapeutiese middels te ontwikkel. Vordering is reeds gemaak om die virus se replikasie te hinder deur die ontwerp van antivirale middels, alhoewel die ontwikkeling van 'n doeltreffende en veilige entstof die enigste manier is om die pandemie te stuit. As gevolg van die vordering in entstof ontwerp, diere modelle en kliniese navorsing is belowende kandidaat entstowwe wat die infeksie kan teenwerk ontwikkel, maar die meeste van hierdie entstowwe wat vir fase I-III kliniese proewe gebruik word is gebaseer op sub tipe B genome. MIV-sub tipe C is wêreldwye die algemeenste sub tipe wat oorgedra word en is die oorheersende sub tipe in lande soos Indië, China, oostelike en suidelike Afrika. 'n Sub tipe C entstof word dringend benodig in ontwikkelende lande soos Suid-Afrika waar antivirale middels onbekostigbaar is. Die membraangeen is 'n aanloklike teiken om as immunogeen in 'n MIV entstof te dien. Die membraanproteïen lok neutraliserende teenliggame en sitotoksiese T-limfosiet reaksies uit. Die proteïen sal dus 'n humorale en sellulêre immuunrespons in die gasheer ontlok. 'n Tekort aan gekarakteriseerde sub tipe C membraangeen volgordes van Suid-Afrika is opgemerk, en dus fokus hierdie studie op die karakterisering van gegenereerde volgordes, asook die uitdrukking van geselekteerde membraangene. Die immunogene is geskep om moontlik gebruik te word in 'n stimuleer-versterkingsstof toedieningstrategie. Die membraangene van onlangs sirkulerende virusstamme in Suid-Afrika was geamplifiseer deur polimerase kettingreaksie (PKR). Die gene is daarna gekloneer vir beide volgordebepalings en uitdrukkingdoeleindes. Filogenetiese verhoudings is uitgewerk deur die volgordes met verwysingsstamme en sub tipe C stamme te vergelyk. Uitdrukking van die gene is waargeneem in 293 selle deur die Westerse kladtegniek te gebruik met MIV-1 positiewe pasiëntsera as teenliggaam.

Volgorde-analise het aangetoon dat die derde varieerbare (V3) lus meer gekonserveer is, en dat die gedeelte wat op die lus volg meer varieerbaar is. Die kroonvolgorde (GPGQ) asook posisies van ongelaaide of negatief gelaaiide aminosure in die V3 lus het aangedui dat die isolate 'n nie-syncytia induserende fenotipe het. Filogenetiese analise het aangedui dat al die volgordes sub tipe C is en dat die volgordes nie rekombinant is nie. Dit is ook deur rekombinasie analise bewys. Die inter-monster diversiteit van die Suid-Afrikaanse volgordes was hoër as die waargenome afstand vanaf die sub tipe C konsensus volgorde. Die Suid-Afrikaanse volgordes is versprei oor verskeie subgroepe in 'n sub tipe C boom, wat die konsep dat hierdie infeksies 'n meer gevestigde epidemie voorstel waar veelvuldige infeksies met verskillende geografiese oorspronge plaasgevind het beklemtoon. Die Westerse klad het ongeprosesseerde gp160 membraanproteïene aangetoon wat Rev afhanklik was.

Hierdie studie het hoogs benodigde sub tipe C Suid-Afrikaanse volgordes van membraangene geproduseer. Die volgordes kan as basis dien om die gene te modifiseer sodat dit gebruik kan word as immunogene in 'n entstof vir Suid-Afrika.

ACKNOWLEDGEMENTS

My sincere thanks are extended to:

Prof. E Janse van Rensburg, my promoter, and Dr. S Engelbrecht, my co-promoter for assistance and encouragement during the course of the project and process of compiling my thesis.

Thomas J Scriba and Florette Treurnicht for aid, support and technical assistance.

The Poliomyelitis Research Foundation and the South African AIDS Vaccine Initiative (SAAVI) for financial support to complete the study.

Chiron Corporation (Emeryville, CA, USA) for supplying constructs TV001c8.2, pCMVLink TV001 o-gp140dv2 modified, and pCMVLink TV001 gp120dv2 modified.

Thomas J Scriba for supplying a *rev* construct.

My family and friends, especially André Jordaan, for their support and encouragement throughout the study period.

Doing something for the greater good of humankind, gives a sense of accomplishment and a feeling of content.

Tania de Villiers (2002)

PUBLICATIONS RESULTING FROM THIS STUDY

Engelbrecht,S., de Villiers,T., Sampson,C.C., zur Megede,J., Barnett,S.W., and van Rensburg,E.J. (2001). Genetic analysis of the complete *gag* and *env* genes of HIV type 1 subtype C primary isolates from South Africa. *AIDS Res. Hum. Retroviruses* 17, 1533-1547.

A copy of this article appears in Appendix C.

TABLE OF CONTENTS

	p. No.
LIST OF ABBREVIATIONS	1
CHAPTER 1 LITERATURE REVIEW	5
INTRODUCTION	6
1.1 DISCOVERY	7
1.2 EPIDEMIOLOGY	8
1.3 HIV-1 STRUCTURE AND REPLICATION	16
1.4 THE ENVELOPE GLYCOPROTEIN	19
1.5 HIV-1 VACCINE STRATEGIES AND DESIGN	33
1.6 GOAL	39
CHAPTER 2 MATERIALS AND METHODS	40
INTRODUCTION	41
2.1 VIRUS ISOLATES	41
2.2 BACTERIAL STRAINS	43
2.3 PLASMID VECTORS AND CONSTRUCTS	43
2.4 CELL LINES AND CELL CULTURE	44
2.5 DNA CONSTRUCTS	45
2.6 PCR AMPLIFICATION	46
2.7 CLONING EXPERIMENTS	48
2.8 EXPRESSION EXPERIMENTS	53
2.9 SEQUENCING OF <i>ENV</i> CONSTRUCTS	56
2.10 PHYLOGENETIC ANALYSIS	57

	p. No.
CHAPTER 3 RESULTS	63
INTRODUCTION	64
3.1 PCR AMPLIFICATION	64
3.2 CLONING EXPERIMENTS	65
3.3 EXPRESSION EXPERIMENTS	69
3.4 SEQUENCING OF <i>ENV</i> CONSTRUCTS	71
3.5 PHYLOGENETIC ANALYSIS	72
FIGURES	80
CHAPTER 4 DISCUSSION AND CONCLUSION	105
INTRODUCTION	106
1.1 SEQUENCE ANALYSIS OF <i>ENV</i> GENES	107
1.2 PHYLOGENETIC RELATIONSHIPS OF <i>ENV</i> SEQUENCES	109
1.3 IMPLICATIONS OF OBSERVED DIVERSITY FOR VACCINE DESIGN	110
1.4 PROTEIN EXPRESSION STUDIES	114
CONCLUSION	117
CHAPTER 5	118
REFERENCES	118
APPENDIX A	A - 1
SEQUENCES USED IN ALIGNMENTS	A - 2 to A - 11
APPENDIX B	B - 1
ENV PROTEIN ALIGNMENT	B - 2 to B -12

APPENDIX C

Copy of article: Engelbrecht,S., de Villiers,T., Sampson,C.C., zur Megede,J., Barnett,S.W., and van Rensburg,E.J. (2001). Genetic analysis of the complete *gag* and *env* genes of HIV type 1 subtype C primary isolates from South Africa. *AIDS Res. Hum. Retroviruses* 17, 1533-1547.

LIST OF ABBREVIATIONS

AIDS	Acquired Immune Deficiency Syndrome
ARV	AIDS-associated Retrovirus
ATV	Active Trypsin Versine
BGH	Bovine Growth Hormone
BSA	Bovine Serum Albumin
C1	First conserved region
C2	Second conserved region
C3	Third conserved region
C4	Fourth conserved region
C5	Fifth conserved region
CA	Major Capsid protein
CD4BS	CD4 binding site
CD4i	CD4 induced
CMV	Cytomegalovirus
cPEIA	Competitive peptide enzyme-linked immunosorbent assay
CRFs	Circulating Recombinant Forms
CTL	Cytotoxic T-lymphocyte
DME	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
<i>env</i>	Envelope gene
Env	Envelope protein
ER	Endoplasmic Reticulum

GNA	Galanthus Nivalis-agarose
gp	glycoprotein
HIV	Human Immunodeficiency Virus
HTLV	Human T-cell lymphotropic virus
IN	Integrase
LAV	Lymphadenopathy-associated Retrovirus
LB	Luria Bertani
LTR	Long Terminal Repeat
MA	Matrix protein
MIV	Menslike immunitetsgebrevirus
MMP	Methyl Mannose Pyranosidase
Mr	Molecular weight
mRNA	Messenger Ribonucleic acid
MVA	Modified Vaccinia Ankara
Nab	Neutralising antibodies
NC	Nucleocapsid protein
NIAID	National Institute of Allergy and Infectious diseases
NSI	Non-syncytium-inducing
o-gp140	oligomeric glycoprotein140
PBMCs	Peripheral Blood Mononuclear Cells
PBS	Phosphate Buffered Saline
PCP	<i>Pneumocystis carinii</i> pneumonia
PCR	Polymerase chain reaction
PKR	Polimerase kettingreaksie

RNA	Ribonucleic acid
RRE	Rev responsive element
PR	Protease
RT	Reverse Transcriptase
SDS-PAGE	Sodium Dodecyl-Sulphate Polyacrylamide Gel Electrophoresis
SD	Standard deviation
SE	Standard error
SHIV	Simian Human Immunodeficiency Virus
SI	Syncytium-inducing
SIV	Simian Immune Deficiency Virus
SNF	Supernatant Fluid
SU	Surface envelope protein
TCLA	T-cell line laboratory adapted
TM	Transmembrane protein
Tm	Melting temperature
V1	First variable
V2	Second variable
V3	Third variable/derde varieerbare
V4	Fourth variable
V5	Fifth variable
WHO	World Health Organisation
g/L	Grams per litre
°C	Degrees Celsius
g	Grams

M	Molar
mA	Milli Ampère
mg/L	Milligrams per litre
ml	Millilitre
mM	Milli Molar
pmol	Pico mol
V/cm	Volts per centimetre
µM	Micro Molar
µl	Micro litre
µg/ml	Micro grams per millilitre
µg	Micrograms
U	Unit

CHAPTER 1 LITERATURE REVIEW

	p. No.
INTRODUCTION	6
1.1 DISCOVERY	7
1.2 EPIDEMIOLOGY	8
1.2.1 Classification of HIV into a family and genus	8
1.2.2 Origin of the HIV virus	9
1.2.3 Causes and outcomes of diversity in the virus	11
1.2.4 The global epidemic	13
1.2.5 The South African epidemic	14
1.3 HIV-1 STRUCTURE AND REPLICATION	16
1.3.1 Virion structure and viral genome organisation	16
1.3.2 HIV-1 life cycle	19
1.4 THE ENVELOPE GLYCOPROTEIN	19
1.4.1 Synthesis and assembly of the envelope glycoprotein	19
1.4.2 The gp120 subunit or SU	22
1.4.3 The gp41 subunit or TM	23
1.4.4 The role of the envelope glycoprotein in HIV-1	25
1.4.4.1 CD4 binding	26
1.4.4.2 Tissue tropism and chemokine receptor usage	27
1.4.4.3 Membrane fusion	29
1.4.4.4 Neutralisation	30
1.4.5 Envelope cloning and protein expression in transfected cell-lines	32
1.5 HIV-1 VACCINE STRATEGIES AND DESIGN	33
1.6 GOAL	39
1.6.1 Objectives	39

INTRODUCTION

Twenty-one years have past since the discovery of the most devastating disease humankind has ever faced. The first clinical evidence of acquired immunodeficiency syndrome (AIDS) was reported in 1981 (Gottlieb *et al.*, 1981). Since the epidemic began, the virus has infected more than 65 million people. At the end of 2002, 42 million people were living with HIV globally. AIDS killed 2.4 million African people in 2002 and the estimated 3.5 million new infections in sub-Saharan Africa in the past year mean that 29.4 million Africans now live with HIV (UNAIDS and WHO, 2002). Of these infected patients, 4.74 million reside in South Africa, which translates to approximately one in every nine South Africans (Department of Health, 2001).

The extreme morbidity, mortality, and epidemic proportions of the disease have led to concentrated scientific efforts to reveal the disease's pathogenesis and develop effective preventative and treatment measures. Advances have been made to inhibit viral replication by suppressing its ability to replicate, and these nucleoside analogue reverse-transcriptase - (e.g. AZT and 3TC), non-nucleoside reverse-transcriptase - (e.g. nevirapine) and protease inhibitors (e.g. zidovudine and didanosine) (Kearney B, 1995) have shown to lower the death rate of HIV infected individuals in the USA by 42% in 1996-1997 (UNAIDS and WHO, 2001). South Africa as part of sub-Saharan Africa is a developing nation, where these therapies are largely unaffordable. Historically, vaccines are the only way to stop an epidemic, and a safe and effective vaccine could be the only way to halt the pandemic in developing nations such as South Africa.

As the outer shell of all viruses are considered to be the predominant target for neutralising antibodies (Nab) from the host, as well as being able to elicit CTL responses to specific epitopes, the envelope gene of HIV has been studied in order to include this gene in candidate vaccines. Most of these candidate vaccines being used in Phase I-III trials are based mainly on subtype B genomes. Only a few subtype C candidate vaccines are in pre-clinical development (Johnston and Flores, 2001). Research on subtype C envelope genes is necessary as this subtype is the predominant subtype

found in India, China, East and Southern Africa, and is the most commonly transmitted subtype worldwide (Esparza and Bhamarapavati, 2000; Osmanov *et al.*, 2002).

In this chapter the discovery and epidemiology of HIV-1 is described as background to the specific structure and function of the envelope gene in the life cycle of the virus. Recent advances in HIV-1 vaccine design incorporating the envelope gene are also highlighted in the Literature Review.

1.1 DISCOVERY

Unexplained cases of immunosuppression, accompanied by opportunistic infections in the early 1960s and 1970s, were the start of what would become the recognition of a new syndrome (Hummer *et al.*, 1987). In June of 1981, a short report in the Morbidity and Mortality Weekly Report of the United States Centers for Disease Control described five cases of *Pneumocystis Carinii* pneumonia (PCP) occurring in previously healthy men in the Los Angeles area (Centers for Disease Control, 1981). The first peer-reviewed article recognising AIDS was published 4 months later when Gottlieb *et al.* (1981) detected PCP in four previously healthy homosexual men. The clinical manifestations and studies of cellular immune function in the patients showed severe acquired T-cell deficiency. There was no indication of spontaneous recovery of cellular immunocompetence in the surviving patients, and all continued to have a severe wasting syndrome despite intensive supportive measures.

In 1983 and 1984, Barre-Sinoussi *et al.* (1983) and Montagnier *et al.* (1984) isolated retroviruses from the lymph node cells of a homosexual man who presented with signs and symptoms often preceding AIDS. Accordingly, this virus was designated lymphadenopathy-associated virus (LAV). After the report of LAV, Gallo *et al.* (1983 and 1984) described the isolation of a cytopathic T-lymphotropic retrovirus from peripheral blood lymphocytes of patients with AIDS. Montagnier

reported that the virus isolated by Gallo in 1983 was in fact given to him by their group from the Institute Pasteur laboratory (Montagnier, 1984). Levy *et al.* (1984) at the University of California in San Francisco cultured an AIDS-associated retrovirus (ARV) from peripheral blood mononuclear cells (PBMCs) of an AIDS patient. In the same year, Popovic *et al.* (1984) propagated Human T-cell lymphotropic virus (HTLV) isolated from patients with AIDS in an immortalised T-cell population. These cytopathic variants differed from HTLV-I and HTLV-II not only in their biological effects, but also in several immunological assays and in their morphology (Schupbach *et al.*, 1984). The new HTLV isolates were collectively designated HTLV-III (Popovic *et al.*, 1984; Schupbach *et al.*, 1984).

Rabson and Martin (1985) set out to determine the divergence between the sequences of LAV, HTLV-III and ARV. Molecular cloning (Alizon *et al.*, 1984; Luciw *et al.*, 1984; Hahn *et al.*, 1984) and nucleotide sequence analysis revealed that the genomes of LAV and HTLV-III were nearly identical (Wain-Hobson *et al.*, 1985; Muesing *et al.*, 1985; Ratner *et al.*, 1985; Sanchez-Pescador *et al.*, 1985), exhibiting 1.4% to 2.2% divergence. In contrast, the ARV envelope nucleotide sequence differed from the LAV sequence by 21% (Rabson and Martin, 1985).

1.2 EPIDEMIOLOGY

1.2.1 Classification of HIV into a family and genus

The family *Retroviridae* comprises animal viruses containing a ribonucleic acid (RNA) dependent deoxyribonucleic acid (DNA) polymerase or reverse transcriptase. The reverse transcriptase enzyme discovered by Temin and Mizutani (1970), as well as Baltimore (1970) works by synthesising a DNA representation of the viral RNA. These viruses integrate into the host genome. In a study conducted by Gonda *et al.* (1985) molecular hybridisation and heteroduplex analysis indicated a great extent of nucleotide sequence homology between HTLV-III and Visna virus; a

pathogenic lentivirus. The morphogenesis and fine structure of the two viruses also demonstrated distinct similarities, as determined by electron microscopy. Therefore, strong evidence was created for a close taxonomic and thus evolutionary relation between HTLV-III and the *Lentivirinae* genus belonging to the *Retroviridae* family. Comparison of genomic sequences supported the notion that the human retroviruses associated with AIDS were related to lentiviruses of other animals (Sonigo *et al.*, 1985).

The designation Human Immunodeficiency Virus was proposed for the lentivirus of humans by Coffin *et al.* (1986). The name conforms to common nomenclature for retroviruses, beginning with the host species ("human") ending with "virus" and containing a word that denotes a major pathogenic property of the prototypic members of the group ("immunodeficiency").

1.2.2 Origin of the HIV virus

Classifying HIV as a lentivirus was supported when evidence indicated that HIV is a zoonoses. This means that the virus crossed species from nonhuman primates, which are the natural host, to humans. Five lines of evidence have been used to substantiate the zoonotic origins of HIV. These include: (i) similarities in viral genome organisation, (ii) phylogenetic relatedness, (iii) prevalence in the natural host, (iv) geographic coincidence, and (v) plausible routes of transmission. HIV-2 originated from the cross-species transmission from sooty mangabeys (*Cercocebus atys*) to humans, and the five sets of criteria were met as follows: (i) HIV-2 and simian immunodeficiency virus (SIV)_{sm} share identical genome structure and both these viruses encode for an accessory protein called Vpx which has not been found in any other primate lentivirus, (ii) trees constructed from the sequences of the 2 viruses indicate that SIV_{sm} and HIV-2 strains cannot be separated into distinct phylogenetic lineages

according to their species of origin. Geographic and phylogenetic linkage of the strains has also been indicated at local level (Chen *et al.*, 1996), which means that sequences of the viruses from humans and primates were the most related in the same immediate geographical area, (iii) the sooty mangabey monkeys habitat includes many west African countries and the infection rate of SIV_{sm} reaches 22% in some groups, (iv) the habitat of the sooty mangabey is in close proximity to the epicentres of the HIV-2 epidemic in Senegal, Guinea-Bissau, Guinea “Conkary” and Côte d’Ivoire, and it overlaps with Sierra Leone and Liberia where the most divergent HIV-2 strains have been identified (Gao *et al.*, 1992; Schim van der Loeff and Aaby, 1999), and (v) sooty mangabeys are hunted for food and orphans are kept as pets (Chen *et al.*, 1996). There is therefore ample opportunity for frequent human contact with infected animals.

The criteria have also been met for transmission of HIV-1. The genome organisation of SIVcpz from the chimpanzee *Pan troglodytes troglodytes* is identical to the HIV-1 genome, containing the *vpu* gene not found in any other lentivirus. HIV-1 is closely related to SIVcpz at phylogenetic level (Gao *et al.*, 1999). *P. t. troglodytes* represents a natural host for the virus. The phylogenetic relatedness presented indirect evidence of a higher prevalence of natural SIVcpz infection based on the discovery of viral recombination between SIVcpz viruses of different lineages. A geographical linkage of human and chimpanzee viruses was found in Cameroon and its immediate vicinity (Corbet *et al.*, 2000a). Furthermore hunting and field dressing (a common practice in West Africa) is a plausible route of zoonotic transmission.

It is predicted that HIV-2 subtypes (A-F) arose from separate cross-species transmissions (Hahn *et al.*, 2000) and that HIV-1 groups M, N and O resulted from no fewer than 3 separate SIVcpz transmission events. The phylogeny exemplified by HIV-1 group M viruses arose as the

consequence of a single SIVcpz transmission event. This should help to estimate the timing of the onset of the pandemic by calculating the date of the last common ancestor of HIV-1 group M.

Earlier estimates placed the origin of HIV-1 group M around 1960 (Li *et al.*, 1988). This was contradicted when HIV-1 isolated from a 1959 plasma sample from a patient in Kinshasa was characterised as belonging to group M (Zhu *et al.*, 1998). Predictions indicating the estimated time of onset are now 1931 (range 1915 to 1941) (Korber *et al.*, 2000).

1.2.3 Causes and outcomes of diversity in the virus

Retroviruses possess the ability to rapidly alter their genomes by mutation and recombination and thus to change in response to altered environmental conditions. This alteration has led to very high levels of variation between strains of HIV-1 and HIV-2 (HIV Sequence Compendium, 2001). These two HIV types are genetically distinct, notably with respect to the regulatory genes *vpu* and *vpx*, which are unique to HIV-1 and HIV-2 respectively. The variation is due to the error prone reverse transcriptase enzyme, which lacks exonucleolytic (proof reading) activity (Bebenek *et al.*, 1989; Boyer *et al.*, 1992), recombination (Robertson *et al.*, 1995), as well as negative and positive selection pressure. Recombinant viruses are created when recombination occurs between viruses with different subtypes in individuals co-infected with multiple divergent HIV-1 strains. Both negative and positive selection pressures are directed against certain regions in the HIV genome to ensure infectivity and survivability. In the *env* gene, negative selection operates against amino acid changes on sites for receptor binding, therefore ensuring infectivity, and positive selection operates against amino acid changes at antigenic sites, therefore ensuring that the virus survives as it is able to escape recognition by the host immune system (Yamaguchi-Kabata and Gojobori, 2000). The *env*

gene is the most divergent among all the genes in the HIV genome (Rabson and Martin, 1985) and a nucleotide divergence of up to 35% has been recorded for these proteins (Gaschen *et al.*, 2002). This diversity has allowed the gene to be used to differentiate between different groups and subtypes in the nomenclature. HIV-2 found predominantly in West Africa (De Cock and Brun-Vezinet, 1989; De Cock *et al.*, 1991; De Cock *et al.*, 1993) is currently subdivided into subtypes A-F, and the HIV-1 sequences are divided into three main groups, M (major), O (outlier), and N (non-M and -O), which are defined as distinct clusters on phylogenetic trees (**Figure 1**) (HIV Sequence Compendium, 2001; Loussert-Ajaka *et al.*, 1995; Simon *et al.*, 1998).

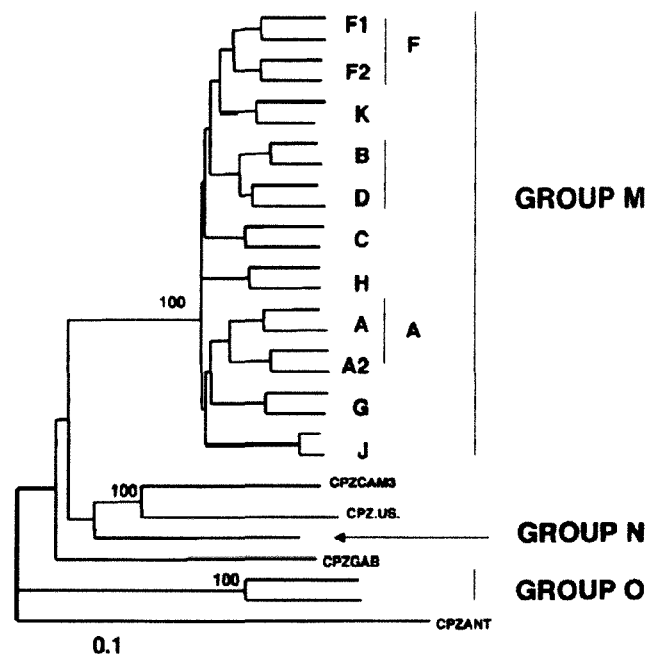


Figure 1. Evolutionary relationships among non-recombinant HIV strains from the HIV-1/SIVcpz lineage based on neighbour joining phylogenetic analysis of nearly full-length genome sequences. The phylogenetic tree shows the different HIV-1 groups, subtypes and sub-subtypes. Each of the internal branches defining a subtype or sub-subtype is supported by 100% of bootstraps (Peeters, 2000).

The viruses of group M are currently subdivided into subtypes A - D, F – H, J, K, and circulating recombinant forms (CRFs) (HIV Sequence Compendium, 2001). Subtype E is now referred to as CRF01_AE and subtype I has been dropped from the nomenclature (Carr *et al.*, 1996; Kostrikis *et al.*, 1995; Gao *et al.*, 1996b).

1.2.4 The global epidemic

The spread of HIV and subtype distribution worldwide is diverse (HIV Sequence Compendium, 2001). Subtype B is the most predominant subtype found in the Americas, Western Europe and Australia (**Figure 2**) (Osmanov *et al.*, 2002). North America had 980 000 people living with HIV/AIDS at the end of 2002, and in Latin America, 1.5 million people were infected (UNAIDS and WHO, 2002). The B subtype is predominantly transmitted by intravenous drug use and by men having sex with men.

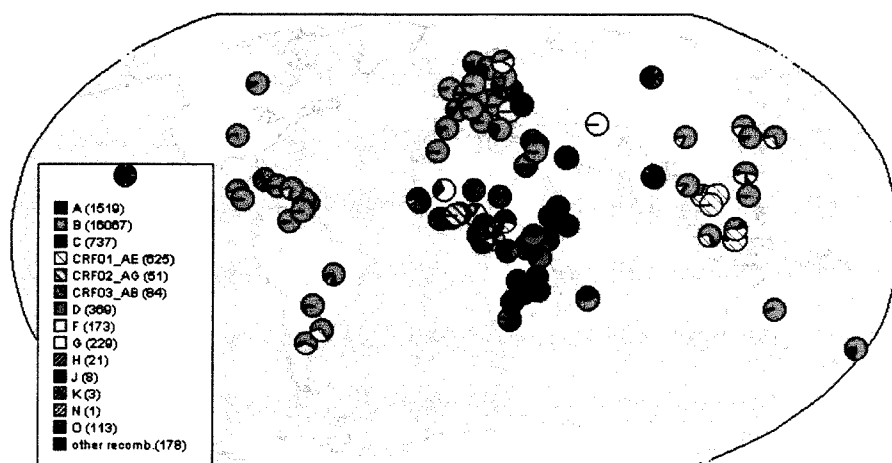


Figure 2. The map represents the distribution of subtypes for each country using a pie chart. The list of subtypes presented on the map are: A, B, C, D, F, G, H, J, K, CRF01_AE, CRF02_AG, CRF03_AB, CRF04_cpx, group N and O (HIV Sequence Compendium, 2001) (<http://hiv-web.lanl.gov/content/hivdb/geography/geography.comp>).

Countries along the central belt of Africa and some North African countries have a subtype A distribution. The leading mode of transmission of subtype A strains is predominantly heterosexual contact. Subtype A infections have also started to spread globally outside Africa and have already been reported in Eastern Europe, East Asia, and the Pacific. Eastern Europe, and East Asia are also affected by subtype B viruses (Osmanov *et al.*, 2002). The epidemic only started in Eastern Europe and Central Asia in the early 1990s and 1.2 million people were reported to have HIV/AIDS at the end of 2002. South and South-East Asia have a predominantly B subtype distribution. In these areas 6.0 million people are reported to be infected with HIV/AIDS (UNAIDS and WHO, 2002). Almost all the subtypes and CRFs are found in central Africa. Subtype C is the most predominant subtype in India (South Asia), East and Southern Africa (**Figure 3**). Subtype C also circulates in China together with CRF07_BC and CRF08_BC. Subtype C viruses are mostly spread by heterosexual contact (Williamson *et al.*, 1995; Osmanov *et al.*, 2002), and these viruses constitute 47.2% of all circulating viruses and are the most commonly transmitted subtype worldwide (Osmanov *et al.*, 2002; Esparza and Bhamarapavati, 2000).

1.2.5 The South African epidemic

The initial AIDS epidemic in South Africa in the early 1980s affected mainly the homosexual/bisexual population (Sher, 1989) and HIV-1 subtype B was predominant (Williamson *et al.*, 1995). The epidemic later shifted to a predominantly heterosexual orientation with HIV-1 subtype C being the most common subtype circulating the country (Engelbrecht *et al.*, 1999; Williamson *et al.*, 1995; van Harmelen *et al.*, 1999; van Harmelen *et al.*, 1997; Moodley *et al.*, 1998; Bredell *et al.*, 1998). The subtype B epidemic, although a lot smaller than the subtype C epidemic, is still observed in the

homosexual population. Subtype B isolates have recently been discovered in the heterosexual population as well (Janse van Rensburg *et al.*, 2002).

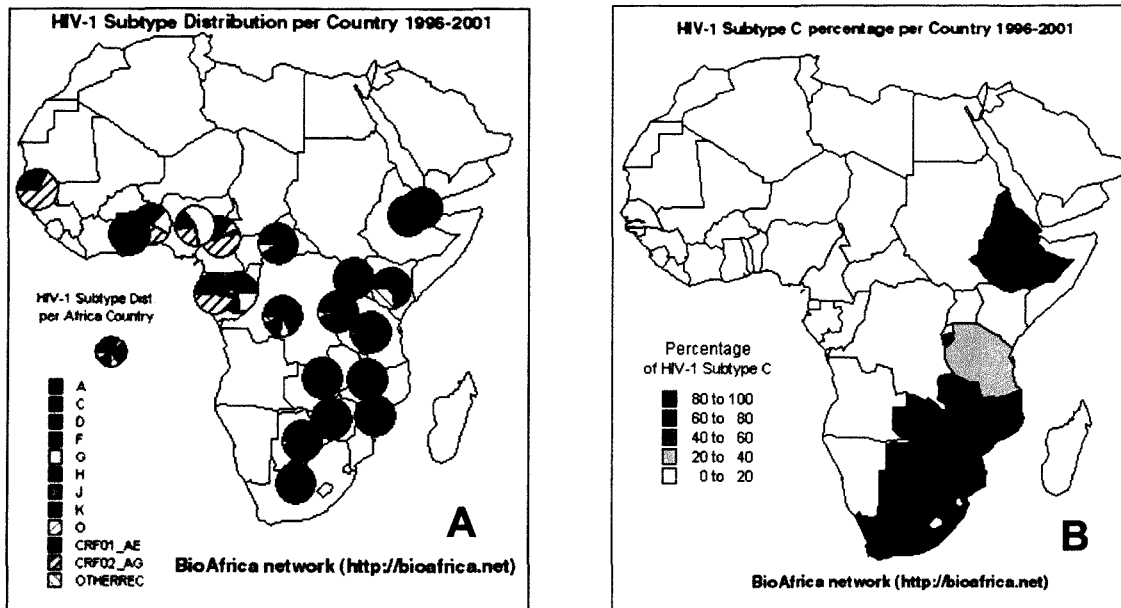


Figure 3. Maps of Africa representing the: A) subtype distribution per country indicated by a pie chart, and B) percentage of viruses in each country classified as belonging to subtype C (<http://bioafrica.net/subtype/index.html>).

Given the high levels of migrancy and the explosive HIV-1 epidemic in the Southern African region as well as the isolation of subtypes A, D, G and intersubtype recombinants in South Africa from 1995-2000 (Bredell *et al.*, 2000; Bredell *et al.*, 2002), indications are that the number of genetic subtypes in South Africa will increase with time.

Seventeen years after the isolation of the first HIV-1 virus in South Africa (Becker *et al.*, 1985), infection rates have reached epidemic proportions. The National HIV and Syphilis sero-prevalence survey of woman attending public antenatal clinics in South Africa indicate that 4.74 million people were infected with HIV-1 by the end of 2001 (Department of Health, 2001). Although the report indicates that a plateau has been reached, infection

rates of pregnant woman are still predicted to be as high as 24.8% (**Figure 4**).

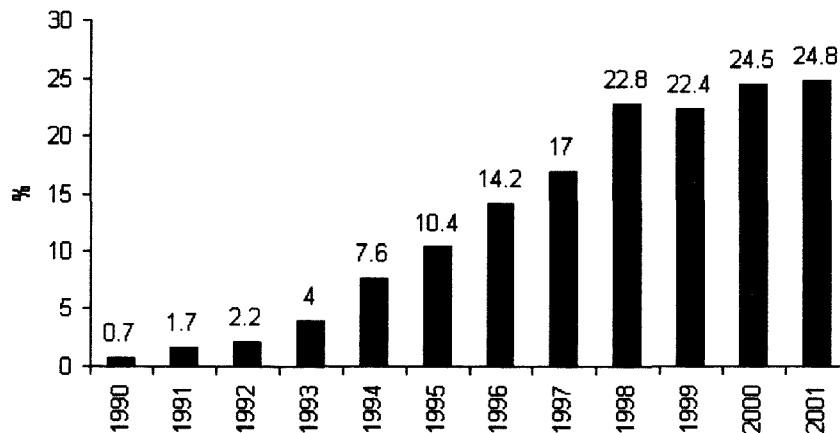


Figure 4. National HIV prevalence trends among antenatal clinic attendees in South Africa: 1990-2001 (Department of Health, 2001).

1.3 HIV-1 STRUCTURE AND REPLICATION

1.3.1 Virion structure and viral genome organisation

As reviewed (Luciw, 1996), HIV virions are spherical in shape and are about 110 nm in diameter (**Figure 5**). A lipid bilayer membrane or envelope surrounds the cone-shaped nucleocapsid, which spans the entire diameter of the virion with a wide free end and narrow end connected to the lipid bilayer. The nucleocapsid within each mature virion contains two molecules of viral single-stranded RNA, which are encapsulated by proteins proteolytically processed from the *gag* precursor polypeptide. These *gag* gene products are the matrix protein (MA or p17), the major capsid protein (CA or p24), the nucleocapsid protein (NC or p7), as well as p1, p2 and p6, which bind tightly to the viral genome.

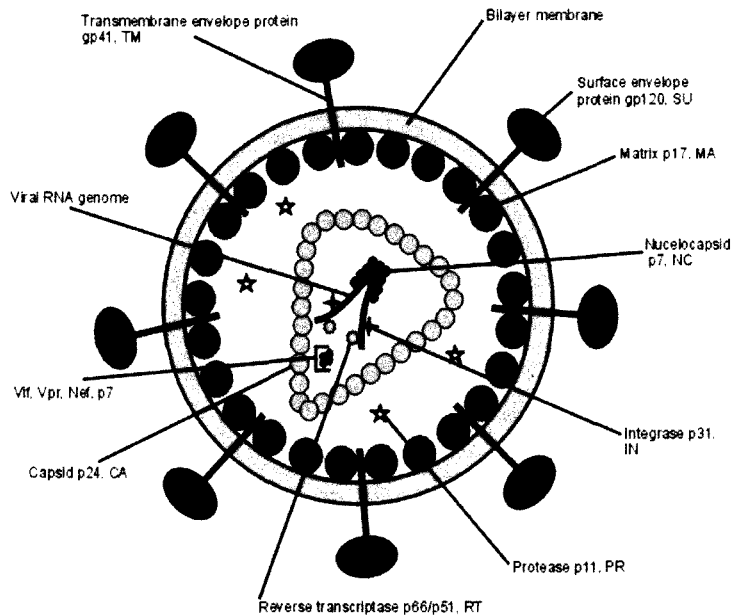


Figure 5. A cartoon depicting the structure of the HIV-1 virion. Adapted from Freed (1998). Structural and accessory gene products have been indicated.

Viral enzymes protease (PR), reverse transcriptase (RT) and integrase (IN) are derived from the *pol* gene precursor polypeptide and are packaged into the virions. RT is a heterodimer composed of a p51 and p66 subunit. An additional enzymatic activity, Rnase H (ribonuclease specific to RNA in RNA-DNA hybrids), is an independent domain in the RT heterodimer. RT, Rnase H and IN are contained within the nucleocapsid in mature virions. The *env* gene products comprise the seventy-two spikes (Ozel *et al.*, 1988) that protrude from the HIV virion in a triangular symmetry fashion. The spikes consist of a trimeric gp120 protein knob and are linked to the lipid bilayer by a gp41 protein stalk. The gp120 of HIV contains a domain that recognises and binds cell receptors and co-receptors vital to virus entry. The receptor for HIV-1 is the CD4 antigen, a cell surface protein on T-helper lymphocytes (Dalgleish *et al.*, 1984; Klatzmann *et al.*, 1984). The gp41 mediates fusion of viral and cellular membranes to allow viral

contents to enter the host cell. Together the *pol*, *env* and *gag* genes form the structural genes (**Figure 6**).

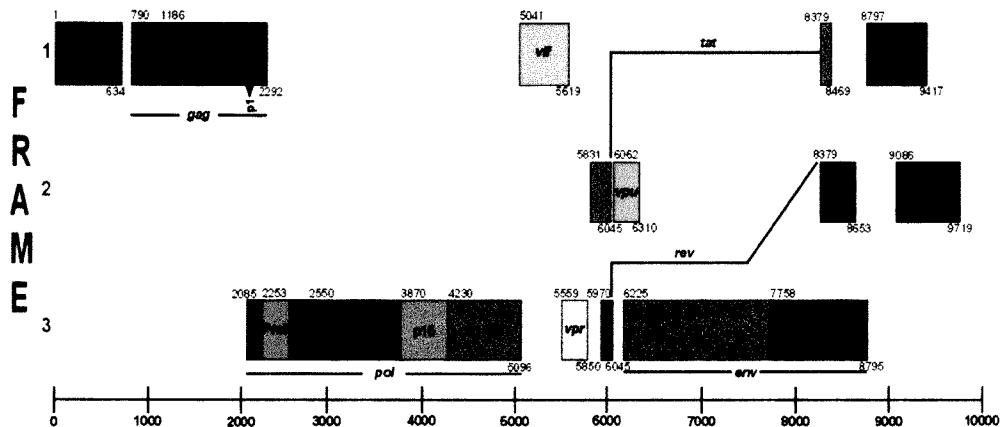


Figure 6. Diagram showing the structural genes *gag*, *pol* and *env* in the HIV-1 genome. Starting positions of genes relative to B.FR.HXB2 (subtype, followed by an abbreviation for the country of isolation, followed by the common name of the strain) have been indicated in the top left corners, and the end positions in the bottom right. Diagram adapted from the Los Alamos HIV sequence database (HIV Sequence Compendium, 2001).

As reviewed (HIV Sequence Compendium, 2001), the transcriptional activator (*tat*) and regulator of viral expression (*rev*) genes are each encoded by two overlapping exons and produce small non-virion proteins which are essential for viral replication. Tat acts as transactivator of HIV gene expression by binding to a TAR element and activating transcription initiation and elongation from the long terminal repeat (LTR) promoter. Rev is a phosphoprotein localised primarily in the nucleus of the host cell. Rev acts by binding to a Rev responsive element (RRE), which promotes the nuclear export, stabilisation and utilisation of viral messenger ribonucleic acid (mRNA) containing RRE. Rev is considered the most functionally conserved regulatory protein of lentiviruses. Rev rapidly cycles between the nucleus and cytoplasm. Accessory genes encoded by HIV-1 are *vif*,

vpr, *vpu* and *nef*. *Vif* is the viral infectivity factor, which promotes infectivity of the virus, but not the production of the virus. *Vpr* or Viral protein R targets the nuclear import of preintegration complexes, arrests cell growth, acts as transactivator of cellular genes and induces cellular differentiation. *Vpu* or Viral protein U degrades CD4 in the endoplasmic reticulum (ER) and enhances virion release from the plasma membrane of HIV-1 infected cells. *Vpu* is further involved in Env maturation. *Nef* down regulates CD4 and MHC class I molecules and increases viral infectivity.

1.3.2 HIV-1 life cycle

The HIV-1 life cycle can be considered as a sequence of events divided into early and late events as demonstrated in **Figure 7**.

1.4 THE ENVELOPE GLYCOPROTEIN

1.4.1 Synthesis and assembly of the envelope glycoprotein

The synthesis and processing of the Env glycoprotein occurs during the secretory pathway of the host cell. This pathway is also utilised for producing cellular secreted and membrane proteins as well as glycoprotein of other enveloped viruses (Doms *et al.*, 1993). The *env* mRNA is translated to envelope glycoprotein on membrane-bound ribosomes as 846- to 868-amino acid precursors. These precursors are only synthesised after the hydrophobic signal peptide is cleaved during translocation into the rough ER of an infected cell. These infected cells include T-cells, monocytes or macrophages, dendritic cells and brain microglia (Klatzmann *et al.*, 1984; Schnittman *et al.*, 1989; Koenig *et al.*, 1986; Wiley *et al.*, 1986; Nicholson *et al.*, 1986; Ho *et al.*, 1986; Gendelman *et al.*, 1985).

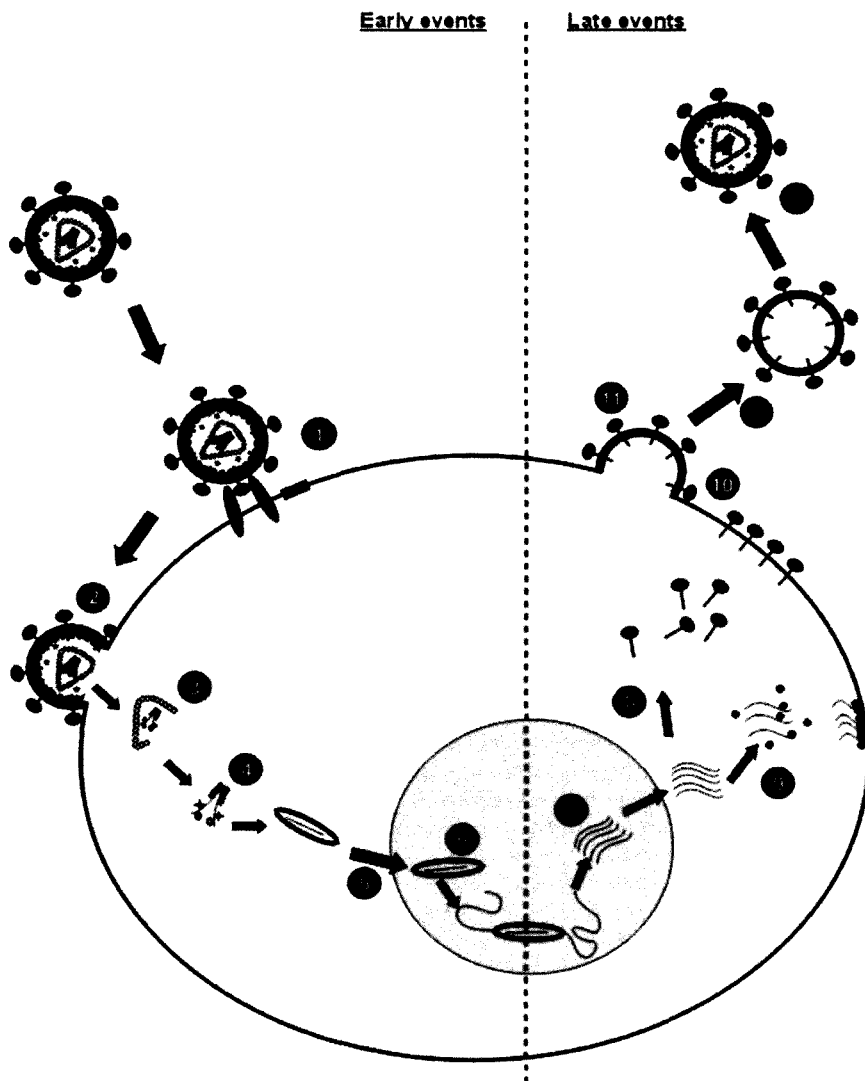


Figure 7. The HIV-1 viral life cycle is shown in this diagram. Adapted from Freed (1998). Steps in the early and late phase have been indicated by numbers 1 through 13: 1, binding of gp120; 2, fusion initiated by gp41; 3, uncoating; 4, reverse transcription of viral RNA; 5, transport of pre-integration complex; 6, integration by IN; 7, transcription; 8, synthesis and transport of Env; 9, synthesis and transport of Gag and Gag-pol precursor; 10, membrane curvature; 11, incorporation of Env into budding virions; 12, completion of budding; 13, cleavage of Gag and Gag-pol precursor and condensing of the core leads to maturation of virion.

This precursor is designated the gp160 (based on apparent molecular mass) after asparagine-linked, high-mannose sugars are added to the precursor (Robey *et al.*, 1985; Allan *et al.*, 1985). The gp160 monomers are then inserted into the lumen of the ER where they oligomerise into trimers (Sodroski, 1999; Weiss *et al.*, 1990; Chan *et al.*, 1997; Weissenhorn *et al.*, 1997). A group has also shown a tetrameric envelope structure (Pinter *et al.*, 1989). The protein is now competent to bind CD4 (Fennie and Lasky, 1989). The oligomerisation process is thought to be required for transport of the oligomeric protein to the *cis* compartment of the Golgi apparatus (Willey *et al.*, 1991), where high mannose ER-acquired N-linked oligosaccharide side chains are modified to more complex forms (Ratner, 1992; Dewar *et al.*, 1989; Stein and Engleman, 1990).

The gp160 is then proteolytically cleaved by a host furin or furin-like protease (Hallenberger *et al.*, 1992) at an Arg-X-Lys/Arg-Arg motif (where X is any amino acid) (Hosaka *et al.*, 1991), which is a highly conserved area among viral Env precursors (McCune *et al.*, 1988; Freed *et al.*, 1989).

The cleavage yields the mature exterior surface envelope protein (SU or gp120) (Allan *et al.*, 1985), and transmembrane glycoprotein (TM or gp41) (Veronese *et al.*, 1985). The gp120 and gp41 glycoproteins are maintained in an assembled trimer by noncovalent interactions (Helseth *et al.*, 1990) and these complexes are then transported to the cell surface. At the cell surface, the envelope protein is destined for one of three fates. First, due to the weak association of gp120 and gp41, some of the gp120 protein is released or shed for the virion (Helseth *et al.*, 1990) and capable of binding CD4 expressed on uninfected or infected lymphocytes (Siliciano *et al.*, 1988). Second, surface expressed gp120-gp41 complexes are capable of binding CD4 on uninfected lymphocytes, resulting in fusion of uninfected with infected lymphocytes (giant cell formation or syncytia) (Lifson *et al.*, 1986a; Lifson *et al.*, 1986b). Third, a fraction of these envelope

glycoproteins are incorporated into budding virus particles by the interaction of gp120-gp41 complexes with Gag and Gag-Pol precursor proteins (Ratner *et al.*, 1991; Wyatt and Sodroski, 1998; Freed, 1998).

1.4.2 The gp120 subunit or SU

The gp120 monomer is a highly glycosylated and variable hydrophobic protein which, after oligomerisation is positioned as a trimer (spike) on the external surface of the HIV-1 virion membrane. Sequence heterogeneity is a characteristic of the *env* gene and five variable domains (V1-V5) interspersed with more conserved regions (C1-C5) have been identified (**Figure 8**) (Starcich *et al.*, 1986). The first four variable regions (V1-V4) form surface-exposed loops that contain disulfide bonds at their bases. The disulfide bonds are generated between 18 cysteine residues, which are highly conserved in the glycoproteins of HIV-1. The conserved gp120 regions (C1-C5) form discontinuous structures important for the interaction with the gp41 ectodomain and with the viral receptors on the target cell (Kwong *et al.*, 1998). Both conserved and variable gp120 regions are extensively glycosylated. Approximately 50% of the apparent molecular mass of gp120 is carbohydrate and the number of N-linked glycosylation sites in gp120 ranges from 18-33, with a median of 25 (Korber *et al.*, 2001). Twenty-four potential sites for N-linked glycosylation (Asn-X-Ser/Thr) have been shown in HIV-1 gp120 of B.FR.HXB2 (HIV Sequence Compendium, 2001). The 24 sites are largely conserved between viral isolates. Leonard *et al.* (1990) showed that less fully processed (i.e. endoglycosidase H-susceptible) oligosaccharides of gp120 are found preferentially at the most conserved glycosylation sites and the remaining sites are relatively conserved although many of them occur in the hypervariable domains. The positions of these sites may shift or be deleted, but there is always one or more new site(s) within 5-10 residues of the reference HIV-1 III_B sequence.

Glycosylation appears to affect intracellular processing and influences both functional and immunological properties of the mature protein. Posttranslational modification governs the transport and eventual maturation of the glycoprotein (Ratner, 1992). Glycosylation also affects viral infectivity (Lee *et al.*, 1992). Most importantly for vaccine purposes, carbohydrate moieties may appear as 'self' to the immune system and the extensive glycosylation of the outer domain surfaces may render it less visible to immune surveillance (Wyatt *et al.*, 1998). Glycosylation also occludes peptide epitopes, therefore providing a mechanism for the virus to escape antiviral immune responses.

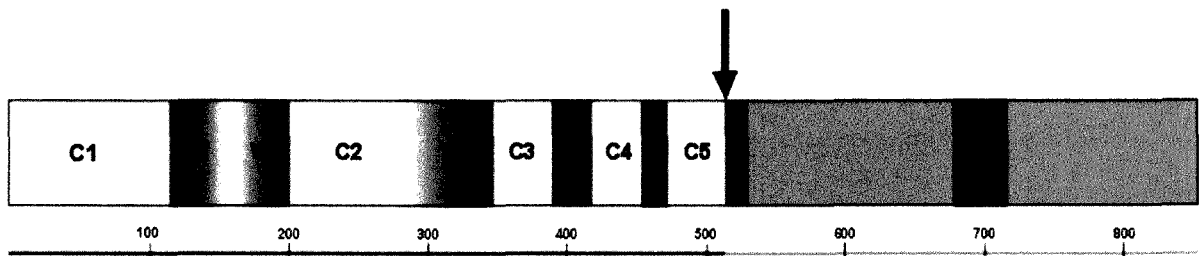


Figure 8. Structure of the gp160 precursor. The gp120 unit is shown in blue and grey, and the gp41 unit is shown in purple, olive and black. The arrow indicates the Arg-X-Lys/Arg-Arg cleavage site. Conserved regions in the gp120 are delineated as C1-C5 (grey), whereas the variable loops are indicated as V1-V5 (blue). The purple bar represents the fusion domain in the gp41 unit, and the black bar represents the transmembrane domain. This diagram was adapted from Helseth *et al.* (1991).

1.4.3 The gp41 subunit or TM

The N-terminus of the gp41 contains a hydrophobic fusion domain, and mutagenesis of this domain leads to fusion-incompetent virions (Freed *et al.*, 1990; Bergeron *et al.*, 1992). Within the hydrophobic fusion region there is a FLG tandem repeat beginning 7 or 8 residues from the N-terminus depending on the virus isolate, and it is hypothesised that this

repeat is the fusogenic site of gp41 (Gallaher, 1987; Schaal *et al.*, 1995). The fusion peptide region of the ectodomain is followed by two regions with a 4-3 hydrophobic (heptad) repeat, a sequence which is characteristic of coiled coils (Chambers *et al.*, 1990; Delwart *et al.*, 1990). These heptad repeats of leucine or isoleucine residues form a soluble, α -helical core consisting of a trimer of antiparallel dimers, and crystallographic analysis of this gp41 core confirmed that it is a six-helix bundle (Chan *et al.*, 1997; Weissenhorn *et al.*, 1997; Tan *et al.*, 1997). It is suggested that this six-helix bundle represents the fusion-active gp41 and plays an essential role in membrane fusion (Wang *et al.*, 2002; Sackett and Shai, 2002). This view has been confirmed by antibody-binding studies (Jiang *et al.*, 1998).

The RRE has been mapped to a region that comprises 240 nucleotides and has the propensity to form extensive secondary structure (Cochrane *et al.*, 1990; Dayton *et al.*, 1988). This region starts in the 3' end of the gp120, 48 nucleotides before the cleavage site of gp120 and gp41 (HIV Sequence Compendium, 2001). The interaction of Rev and the RRE has been shown to regulate the expression of the *env* gene at posttranscriptional level (Dayton *et al.*, 1988; Hadzopoulou-Cladaras *et al.*, 1989; Hammarskjold *et al.*, 1989). The Rev-mediated transport to the cytoplasm of unspliced and singly spliced mRNA possessing a RRE, is dependent on direct binding of Rev (Malim *et al.*, 1989).

A second hydrophobic region, the membrane-spanning region, is found at amino acids 684-706 (positions in the B.FR.HXB2 strain) (HIV Sequence Compendium, 2001) and this region stops translocation through the lipid bilayer of the ER and serves to anchor the protein during fusion of the viral and cellular membranes. A conserved tryptophan-rich motif in the region adjacent to the membrane-spanning region is predicted to form a α -helix, and deletion of this region abrogates the ability of the envelope

glycoprotein to mediate both cell-cell fusion and virus entry (Salzwedel *et al.*, 1999).

The next readily identifiable region is marked by two or three cysteine residues (Cys-598, Cys-604, and Cys-764) (positions in B.FR.HXB2), which are highly conserved in different HIV-1 isolates (HIV Sequence Compendium, 2001). Cys-764 is located downstream from the hydrophobic membrane anchor region of gp41 and presumed to be in the intracellular domain of gp41. In contrast, both Cys-598 and Cys-604 are located upstream from the membrane anchor region and are believed to be in the extracellular domain of gp41. Mutagenesis of the latter mentioned cysteine residues resulted in the loss of viral infectivity, which could be attributed to severe impairment in the processing of gp160 precursor to gp120 (Syu *et al.*, 1991). This region containing the Cys-598 and Cys-604 further comprises the principal epitope recognised by antibodies to the TM (Wang *et al.*, 1986; Gnann Jr. *et al.*, 1987), and is designated the immunodominant region (amino acids 588-605) (HIV Sequence Compendium, 2001). A conserved glycosylation site follows the immunodominant region, and is one of a possible 5-6 glycosylation sites in the gp41 (Gallaher *et al.*, 1989; HIV Sequence Compendium, 2001).

A long cytoplasmic tail of approximately 150 amino acids in length is found at the C-terminus of the gp41. Studies have indicated that the tail is required for infectivity (Dubay *et al.*, 1992; Yu *et al.*, 1993) as well as incorporation of envelope proteins into virions in a cell type-dependent manner (Murakami and Freed, 2000; Gabuzda *et al.*, 1992).

1.4.4 The role of the envelope glycoprotein in HIV-1

The envelope protein plays an essential role in the lifecycle of HIV-1. The protein is responsible for the entry of the virus into a target cell by the

action of the gp120 in its trimeric form, which binds to the CD4 receptor of the target cell. The gp120 further specifies the tropism of the virus, which determines the type of host cell, which will be infected. The gp41 is responsible for fusion of the viral and host membranes, therefore allowing the viral genome to enter the host cell. The envelope protein further elicits antibodies from the host as neutralising epitopes are exposed on the surface of the protein.

1.4.4.1 CD4 binding

The primary step in the HIV-1 infection involves the binding of the gp120 CD4 binding domain (Kowalski *et al.*, 1987; Lasky *et al.*, 1987) to the host cell surface molecule CD4, which serves as the major receptor for HIV-1 (Dalglish *et al.*, 1984; Klatzmann *et al.*, 1984; Maddon *et al.*, 1986). The more conserved regions in gp120 are folded into the core, which is composed of two domains: an inner and outer domain, and a β sheet (**Figure 9**) (Wyatt and Sodroski, 1998). Elements of both domains and the bridging (β) sheet contribute to CD4 binding. Genetic, biochemical and immunochemical studies have indicated regions in the gp120, which interact with the CD4 receptor. Site-specific mutations and the analysis thereof indicated a limited number of conserved amino acids in different regions of the gp120, which are required for efficient binding to the CD4 receptor. Trp427 and Val430-Pro438 in hydrophobic regions, as well as Asp368/Glu370 and Asp457 in hydrophilic regions are the contact amino acids for CD4 in the gp120 protein (HIV Sequence Compendium, 2001). In addition to binding CD4 on the cell surface, HIV-1 Env associates with CD4 intracellularly soon after gp160 synthesis in the ER by adopting a conformation suitable for CD4 binding (Fennie and Lasky, 1989). This leads to the down-regulation of CD4 expression from the cell surface of Env-expressing cells (Hoxie *et al.*, 1986; Kawamura *et al.*, 1989), which may reduce the ability of these cells to become infected with additional

virions. This is known as super-infection interference (Stevenson *et al.*, 1988).

The binding of gp120 to CD4 induces conformational changes in the gp120, which contributes to the exposure of the binding site for the chemokine receptors (Sattentau and Moore, 1991; Sattentau *et al.*, 1993).

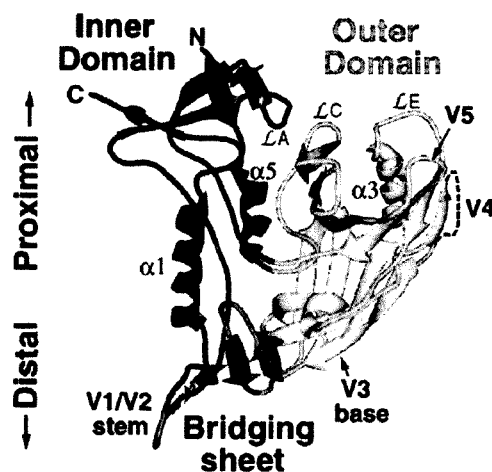


Figure 9. A ribbon diagram depicting the core gp120 structure. The inner and outer domain as well as the bridging sheet, which play a role in CD4 binding is indicated (Kwong *et al.*, 1998).

1.4.4.2 Tissue tropism and chemokine receptor usage

An additional function of the HIV-1 Env glycoprotein is to determine the cell-type specificity of virus infection. Most primary virus isolates of HIV-1 obtained from infected individuals during the early, asymptomatic phase of infection are NSI and macrophage-tropic. These isolates use the CCR5 chemokine receptor as co-receptor for entry into cells of monocyte or macrophage lineage (Deng *et al.*, 1996; Dragic *et al.*, 1996; Alkhatib *et al.*, 1996; Choe *et al.*, 1996; Wu *et al.*, 1996). HIV-1 isolates capable of

infecting T-cell lines tend to arise late in infection after the onset of AIDS-defining symptoms. These isolates are syncytium-inducing (SI) and often use other chemokine receptors, frequently CXCR4, in addition to CCR5 as co-receptors (Schuitemaker *et al.*, 1992; Connor *et al.*, 1997; Zhang *et al.*, 1996; Björndal *et al.*, 1997). There appears to be subtype-dependent differences in frequency of usage for certain chemokine receptors. A dual tropism for both CXCR4 and CCR5 has been more rare among subtype D isolates, and the CXCR4-positive phenotype was significantly more rare among subtype C isolates (Tscherning *et al.*, 1998). Subtype C isolates obtained from late-stage Ethiopian AIDS patients showed no CXCR4 usage in any of the isolates, unlike other subtypes, which have been indicated to start using CXCR4 in later stages of the disease progression (Björndal *et al.*, 1999). The discovery that subtype C isolates from patients in later stages of disease progression did not use CXCR4, but remained NSI and CCR5-tropic was also reflected in isolates obtained by our group in 1998 (Treurnicht *et al.*, 2002). Our group also studied the phenotype of more recent subtype C isolates (obtained in 2000-2001), and these studies indicate that subtype C viruses are now following the same trend as other subtypes by switching to CXCR4 usage (Janse van Rensburg *et al.*, 2002).

Studies involving chimeric envelope glycoproteins have demonstrated that the third variable loop (V3) in the gp120 plays a central role in determining cell tropism and chemokine receptor usage (Speck *et al.*, 1997; Chesebro *et al.*, 1991; Chesebro *et al.*, 1992; Cheng-Mayer *et al.*, 1990; Willey *et al.*, 1994). Versions of the gp120 protein where the V3 loop has been deleted do not bind CCR5, although CD4 binding occurs at wild-type levels. Antibodies to the V3 loop have also been shown to interfere with gp120-CCR5 binding (Wu *et al.*, 1996). Basic amino acids in one or more of positions 305, 321 or 327 confer a SI phenotype (positions in B.FR.HXB2), whereas hydrophobic amino acids in these positions correlate with a NSI phenotype (Björndal *et al.*, 1999). The net charge in the V3 loop of SI and

NSI variants generally have an overall charge of $>+5$ and $=<+5$ respectively. Therefore the position and charge of a particular residue influences the overall phenotype of a variant. A tetrapeptide crown sequence located at the apex of the V3 loop is highly conserved motif among HIV-1 isolates (Gly311-Pro312-Gly313-Arg314 for strain B.FR.HXB2) (HIV Sequence Compendium, 2001). The sequence of the crown is indicative of the viral phenotype as well. NSI sequences mostly have a GPGQ sequence, whereas SI viruses have a positively charged residue replacing the Q with either an H or R (Abebe *et al.*, 1999). Phenotypes are further influenced by sequences outside the V3 loop. A functional interaction between the V3 loop and V1/V2 loops has been detected, which is important for infectivity as well as syncytium formation and cell tropism (Koito *et al.*, 1994). Stamatatos and Cheng-Mayer (1993) suggest that the overall conformation of the envelope glycoprotein determines the phenotype of the virus.

1.4.4.3 Membrane fusion

The receptor binding-mediated activation of membrane fusion is postulated to involve a conformational change in the envelope protein complex from a native (nonfusogenic) to a fusion-active (fusogenic) state. The hydrophobic (heptad) repeats following the fusion domain form a six-helix bundle, which represents the fusion-active gp41 as described earlier. The bundle has a coiled-coil structure (Wild *et al.*, 1994), the identical structure which in an influenza virus system induces fusion in a pH-dependent process (Carr and Kim, 1993). In an HIV system, the transition from native to fusogenic by the formation of the coiled-coil is a result of conformational change after binding of the gp120 to CD4, and is pH-independent (Stein *et al.*, 1987; McClure *et al.*, 1988; McClure *et al.*, 1990). The conformational change facilitates the exposure of the hydrophobic, glycine-rich fusion peptide (Sattentau and Moore, 1991) and insertion of the peptide into the

membrane of the target cell (Wyatt and Sodroski, 1998). Viral and cellular membranes are then fused by an unknown mechanism allowing the virus' genetic material to enter the host cell. Mutagenic analysis (Cao *et al.*, 1993; Freed *et al.*, 1990) and crystal structures of the HIV-1 gp41 ectodomain support this model (Chan *et al.*, 1997; Weissenhorn *et al.*, 1997).

1.4.4.4 Neutralisation

The envelope protein is a prime target for the human immune system. The trimeric protein elicits several different groups of antibodies during the course of infection (**Figure 10**) (Wyatt, 2002). Early after infection occurs non-neutralising antibodies are created by the host, which are directed to the gp41 protein as well as the C1 and C5 domains in the gp120. It has been proposed that these antibodies are non-neutralising because the epitopes are either not exposed as a result of gp120 shedding from the virions (Helseth *et al.*, 1990), or the conformation of the protein doesn't allow the exposure of the epitopes at this stage of the infection (Wyatt, 2002).

The next group of antibodies raised in the host is strain-restricted Nab directed mainly at the V3 loop in the gp120. A few antibodies are also directed at the V2 loop. These antibodies are relatively potent, but the viruses aren't neutralised because escape mutants are created relatively easy by point mutations as a result of the error prone reverse transcriptase enzyme (Bebenek *et al.*, 1989; Boyer *et al.*, 1992), and antibodies therefore can't recognise the altered epitopes.

Potent broadly Nab are elicited much later in infection and 4 classes of antibodies have been identified so far. The first class comprises antibodies directed at the CD4 binding site (CD4BS). The CD4BS antibodies cross

compete with the gp120 protein to bind to the CD4 binding site expressed on the surface of host cells (Zwick *et al.*, 2001a).

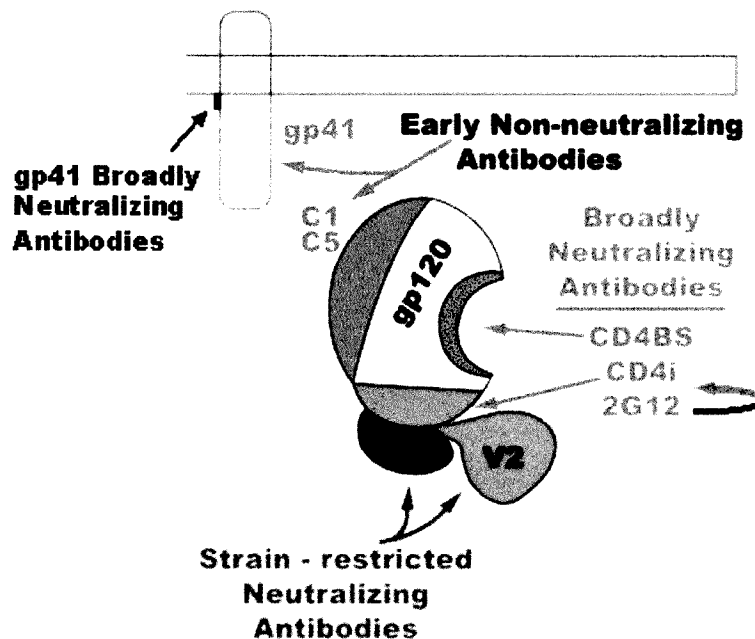


Figure 10. A cartoon depicting the different antibodies elicited by the envelope glycoproteins during infection (Wyatt, 2002).

More than 50 of these antibodies have been identified so far, but only the IgGb12 antibody first identified by Burton *et al.* (1994) has been shown to neutralise primary strains, and these are extremely rare in hosts (Kessler *et al.*, 1997). The second class contains the CD4 induced (CD4i) antibodies 17b, 48d, A32 and CG10 (Thali *et al.*, 1993; Weinberg *et al.*, 1997; Sullivan *et al.*, 1998), which bind to conserved regions in the C4 region and portions of the V3 loop. They are V3 dependent and seem to bind to these sites with a greater affinity after gp120 has attached to the CD4 molecule. These antibodies have been shown to neutralise T-cell line laboratory adapted (TCLA) strains, but aren't potent against primary strains. The very rare 2G12 antibodies recognise a complex discontinues

epitope involving the C3-C4 region of the gp120 (Trkola *et al.*, 1996) and are classified in the third class. The epitope is formed by mannose residues contributed by glycans attached to N295 and N332 and other glycans playing an indirect role in maintaining epitope conformation (Scanlan *et al.*, 2002), therefore making the antibodies carbohydrate dependent (Sanders *et al.*, 2002).

The fourth class contains antibodies directed at the gp41 region. They are 7-8 amino acids in length and are conformation dependent. This class contains the 2F5 antibodies (Muster *et al.*, 1993), as well as Z13 and 4E10 which recognise a region immediately C-terminal of the 2F5 epitope (Zwick *et al.*, 2001b; Buchacher *et al.*, 1994). The Z13 and 4E10 antibodies are able to neutralise selected primary isolates (Zwick *et al.*, 2001b).

Although the envelope contains epitopes to elicit all the above-mentioned antibodies, viruses aren't neutralised efficiently in the host. Spontaneous shedding of the gp120 from the virions are a likely mechanism of immune evasion as the envelope either doesn't present these epitopes after shedding, or exposes "decoy" epitopes to force the immune system to create the wrong non-neutralising antibodies (Wyatt, 2002). Potent and broadly Nab are relatively rare in infected individuals, and some are more subtype specific than others. 2F5 antibodies have been shown to neutralise subtype C viruses to a lesser extent than subtype B, and subtype C viruses have shown to resist neutralisation by 2G12 antibodies (Bures *et al.*, 2002). This poses a major challenge for HIV vaccine design.

1.4.5 Envelope cloning and protein expression in transfected cell-lines

Several envelope genes have previously been cloned and transiently expressed in cell-lines. Clones were generated for full-length gp120 and gp160 *env* genes from isolates obtained from the World Health

Organisation (WHO)- and National Institute of Allergy and Infectious Diseases (NIAID)-sponsored Networks for Virus isolation and Characterisation and represent viruses collected at potential vaccine evaluation sites. Env proteins were expressed after transient transfection into HeLa cells. Only unprocessed gp160 proteins were detected as a result of experimental conditions used, and the proteins varied considerably in size and carbohydrate content (Gao *et al.*, 1996a; Björndal *et al.*, 1997; WHO Network for HIV Isolation and Characterization, 1994).

1.5 HIV-1 VACCINE STRATEGIES AND DESIGN

A safe, highly effective, and affordable preventative vaccine is the only way to halt the HIV-1 pandemic worldwide. Although early host immune responses against HIV would ordinarily portend a successful quest for a vaccine, the quest is hampered by various factors, which provide the virus with the ability to escape from the immune system during the course of infection.

These include extensive glycosylation, conformational flexibility, variable domains and unstable subunit association, which causes shedding of gp120. Carbohydrate moieties render the virus less visible to the immune system and together with hypervariable loops camouflage neutralising epitopes providing a mechanism for the virus to escape antiviral immune responses (Wyatt and Sodroski, 1998; Wyatt, 2002). Furthermore, as a retrovirus, HIV-1 is capable of rapidly altering its genome by mutation and recombination and thus changes in response to altered environmental conditions. A vaccine would therefore have to be designed with all these limitations in mind.

Preventative vaccines establish immunological memory for antigenic structures presented by a pathogen or by infected cells, therefore including elements of the adaptive immune responses in the recipient (Graham, 2000). The basic cellular elements of the adaptive immunity include B-lymphocytes, which produce

antibodies (humoral immune response), and T lymphocytes. Cytolytic activity (cellular immune response) is mediated primarily by CD8⁺ T cells. Soluble factors produced by activated CD4⁺ and CD8⁺ T cells have anti-viral activity and can influence the differentiation, expansion and duration of T cell responses. Antibodies are the only component of the adaptive immune response that can neutralise a virus prior to infection of a cell, and CD8⁺ T cells are the principal effector mechanism to clear cells already infected with virus and therefore controls viral replication during early infection (Koup *et al.*, 1994). This suggests that a HIV-1 vaccine should stimulate both the cellular and humoral arms of the immune system.

In the natural history of HIV-1 (**Figure 11**), primary viremia lasts 3 months after the initial infection. Thereafter a set point is reached which is described as the steady state in which virus clearance matches virus production (Graham, 2000). Viral RNA levels increase after seroconversion at a rate of 0.11 log₁₀ copies/ml per year (Sabin *et al.*, 2000) until the patient develops AIDS (Henrard *et al.*, 1995). The level of CD4⁺ T cells initially decreases during primary viremia, but recovers and stabilises during the asymptomatic period of the disease only to deplete when AIDS develops. A CD8⁺ memory T-cell response develops after 3-6 months of infection, declines after 6-8 months and stabilises after 12 months (Musey *et al.*, 1997). Nab are produced after approximately 8 months of infection. If the production of CD8⁺ T-cell and Nab responses can be shifted towards the time point at which infection takes place, an adaptive immune response will most likely be able to clear virus before the infection can take place (Montefiori, D.C. (2002), personal communication, Duke University, Durham, North Carolina, USA).

Strategies to prevent infection have been directed toward the development of vaccine regimens that can elicit effector T-cell responses to eliminate viral infected cells and antibody responses to neutralise cell-free virus. As the envelope of HIV-1 contains both neutralising antibody binding sites (Chanh *et al.*, 1986) and CTL

epitopes (Takahashi *et al.*, 1988), the gene has been studied in order to be included in candidate vaccines.

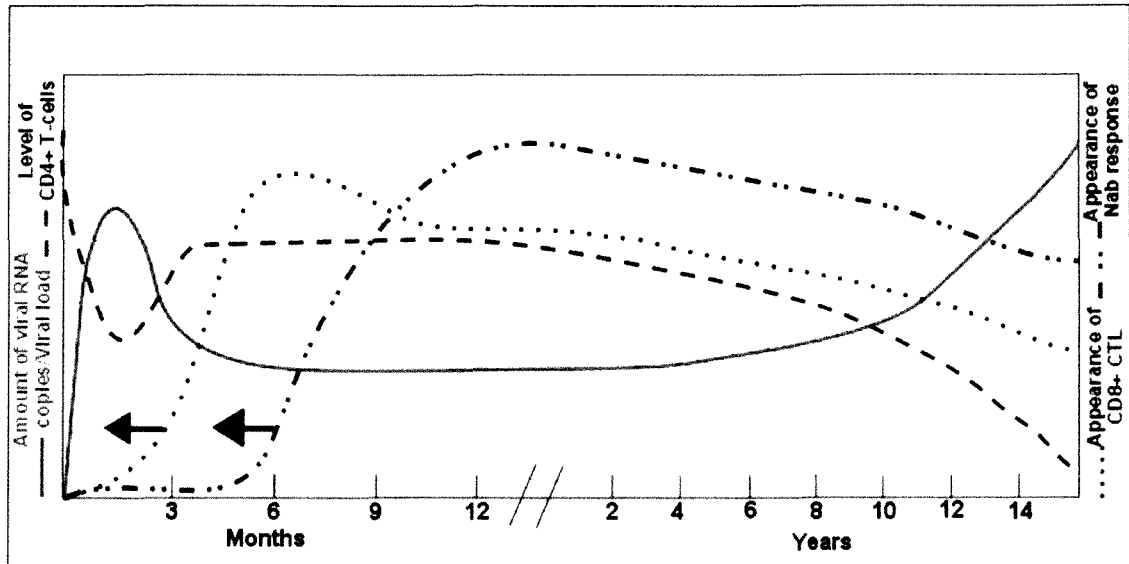


Figure 11. Schematic drawing showing the course of HIV-1 infection. The viral load is indicated by the solid blue line. CD4+ T-cell levels are shown by the dashed green line. CD8+ CTL responses develop after approximately 3-6 months of infection, and is indicated by the dotted red line. Nab only develop after 8 months of infection and is shown by the dashed dotted black line. The red and black arrows indicate that shifting of CTL and Nab responses may be able to prevent infection. Figure adapted from Mindel and Tenant-Flowers (2001), Sabin *et al.* (2000) and Musey *et al.* (1997).

Several adjuvant recombinant monovalent HIV envelope proteins (e.g. gp160 or mature gp120) based on TCLA isolates of subtype B HIV have shown to induce Nab in virtually all volunteers, but these antibodies exhibited little cross-reactivity against primary isolates of HIV (Mascola *et al.*, 1996). Subsequently, bivalent candidates containing two subtypes of gp120 (one from a TCLA subtype B, and one from a subtype B or E primary isolate) have been designed by VaxGen, and Phase III trial results are expected within the next 2-3 years (Esparza *et al.*, 2002; Johnston and Flores, 2001; Esparza and Bhamarapravati, 2000).

Live recombinant vectors expressing HIV genes are promising candidates for vaccines. Attenuated vaccinia expressing the HIV gp160 protein, and additional more complex vaccinia vectors expressing *env* and *gag-pol*, boosted with recombinant envelope proteins induced Nab in vaccine recipients. CTL activity could not be measured accurately in these trials because sensitive assays to detect CTL responses were not available at the time (Corey *et al.*, 1998; Johnston and Flores, 2001).

The risk associated with potential virulence of vaccinia in immune deficient individuals has directed attention to the design of recombinant viral vectors, such as the recombinant canarypox vector or ALVAC. This vector induces HIV-specific CTL responses in approximately one third of volunteers, and a boost with gp120 subunit vaccines further induces Nab and makes this prime-boost strategy a promising approach (Clements-Mann *et al.*, 1998).

A recombinant adenoviral vector has also been created to act as a boost in a prime-boost regimen where a plasmid encoding a codon-optimised subtype B *gag* serves as prime (Shiver *et al.*, 2002). The replication-defective adenoviral vector expressing the optimised Gag (identical to *gag* from plasmid vector) regimen has shown significant and persistent CD8+ and CD4+ cellular immune responses in monkeys, which mitigate simian human immunodeficiency virus (SHIV) infection following challenge. Preliminary data from Phase I safety and immunogenicity trials in HIV-uninfected individuals suggest that both the Gag expressing DNA plasmid and adenoviral vector are tolerated well, the Gag plasmid vector in saline can elicit a moderate Gag-specific cellular immune response and the Gag expressing adenoviral vector can elicit a persistent CD8+ and CD4+ response. The immunogenicity of the adenoviral vector may be compromised by high levels of pre-existing anti-adenoviral neutralising antibody titres in the host, but this can be overcome by increasing the dose of the vector. Another recombinant pox vector, modified vaccinia Ankara (MVA) expressing HIV Gag and a number of CTL

epitopes is currently being evaluated in a Phase I clinical trial in the United Kingdom and Kenya.

Peptide based vaccines comprising viral envelope or internal proteins have failed to elicit a high frequency and strength of Nab and CTL responses, but lipopeptides derived from Env, Gag and Nef proteins promise to improve immunogenicity (Gahéry-Ségard *et al.*, 2000). A peptide vaccine representing the variable regions of the envelope glycoprotein has been shown to induce HIV-1 Nab in mice, rabbits and rhesus macaques monkeys (Carlos *et al.*, 2000). The use of adjuvants, cytokines and co-stimulatory molecules which will augment vaccines, as well as peptides based on predictions of epitopes representing immunodominant, conserved epitopes are also under development (Evans *et al.*, 2001; Wilson *et al.*, 2001; Gherardi *et al.*, 2000).

The concept of DNA immunisation against viral diseases offers several advantages compared to the concept of conventional subunit vaccines. With DNA, both humoral and cell mediated immune responses can be generated, and the protein antigens of interest are synthesised in the host cell resulting in authentic glycosylation and processing (Donnelly *et al.*, 1997). This technology has shown that cytotoxic immunity is induced against various agents in experimental animals as well as humans (Boyer *et al.*, 2000). DNA vaccines based on corresponding recombinant gp120 proteins alone have failed to protect individuals against infection in Phase I/II trials (Bolognesi and Matthews, 1998). This has led to the design of different vaccine approaches where the immunogenicity of the envelope DNA is increased after an additional boost with subunit proteins.

Vaccination with HIV-1 gp120 DNA in experimental animals has been shown to induce antibody responses to gp120, which were elevated after the addition of a recombinant gp120 subunit boost, which also increased neutralising antibody titres. Furthermore, strong CTL responses against the Env V3 epitope were induced (Barnett *et al.*, 1997). Several groups further hope to increase the immunogenicity

of the envelope by constructing modified envelopes which are codon optimised (Corbet *et al.*, 2000b), and which reveal conserved conformational epitopes critical to HIV entry (Chakrabarti *et al.*, 2002; Chow *et al.*, 2002). Attempts have also been made to stabilise oligomeric envelopes in order to resemble the functional envelope glycoprotein on the surface of the virion (therefore creating an oligomeric gp140 or o-gp140) (Binley *et al.*, 2000; Zhang *et al.*, 2001; Stamatatos *et al.*, 2000). Biochemical, structural and immunological characterisations of the o-gp140 have indicated that the purified proteins are in a stable trimeric conformation with critical neutralising epitopes exposed. Preliminary vaccine studies with rabbits demonstrated that the o-gp140 protein was highly immunogenic and induced high-avidity and predominantly conformationally directed antibody responses (Srivastava *et al.*, 2002). Stabilised o-gp140 derived from primary isolates was able to induce Nab against primary isolates and TCLA HIV strains (Yang *et al.*, 2001).

Envelopes lacking V1 and V2 loops have been constructed (Stamatatos *et al.*, 1998), as it has been indicated that these loops mask the CD4 binding site and neutralising epitopes (Wyatt *et al.*, 1995). Deletion of the V2 loop renders a virus more susceptible to neutralisation (Cao *et al.*, 1997; Stamatatos *et al.*, 1998). Deletion of the V2 loop of a gp140 envelope resulted in a candidate vaccine which induced both envelope-specific T-cell lymphoproliferative responses and potent Nab in macaques monkeys (Cherpelis *et al.*, 2001a; Cherpelis *et al.*, 2001b). Immunogens with the V2 deletion were further more successful in eliciting Nab than its gp140 counterpart with V2 intact (Barnett *et al.*, 2001). Vaccination with these V2 deleted immunogens offered partial protection from pathogenic SHIV challenge in rhesus macaques monkeys (Cherpelis *et al.*, 2001b).

Several other approaches are also under investigation. Removing carbohydrates from the envelope has shown to strongly prime antibody responses (although non-neutralising), as well as T-cell responses (Bolmstedt *et al.*, 2001). Single-chain polypeptide analogues of the gp120-CD4 complex are also under investigation as targets for anti-HIV vaccines and therapies as these are intermediate structures

with distinct conformational, functional and antigenic features (Weinberg *et al.*, 1997; Sullivan *et al.*, 1998; Fouts *et al.*, 2000).

Genes such as *env* need to be modified to enhance their immunogenicity and basic molecular research is needed to gain knowledge on how to modify these genes to be used in vaccine strategies. As only one full-length subtype C envelope sequence from South Africa had been described in a near full-length molecular clone (Rodenburg *et al.*, 2001) at the start of this study, the goal was to generate envelope reference reagents for further use in vaccine research.

1.6 GOAL

To characterise and express *env* immunogens for possible use in a prime-boost vaccine modality.

In order to achieve this goal, the following two objectives were identified:

1.6.1 Objectives

1.6.1.1 Clone, sequence and phylogenetically analyse subtype C *env* genes from recent circulating strains in South Africa.

1.6.1.2 Express the proteins from selected isolates in mammalian cells to confirm intact genes and open reading frames.

CHAPTER 2 MATERIALS AND METHODS

	p. No.
INTRODUCTION	41
2.1 VIRUS ISOLATES	41
2.2 BACTERIAL STRAINS	43
2.3 PLASMID VECTORS AND CONSTRUCTS	43
2.3.1 pCR3.1	43
2.3.2 PCMVLink	43
2.3.3 Constructs received from outside sources	44
2.4 CELL LINES AND CELL CULTURE	44
2.5 DNA CONSTRUCTS	45
2.5.1 Constructs for sequencing purposes	45
2.5.2 Constructs for expression experiments	46
2.6 PCR AMPLIFICATION	46
2.6.1 Full-length gp160 <i>env</i> genes for pCR3.1 cloning	46
2.6.2 Native <i>env</i> genes for directional cloning	47
2.7 CLONING EXPERIMENTS	48
2.7.1 Ligation and transformation reactions	48
2.7.2 Screening of recombinant clones	51
2.8 EXPRESSION EXPERIMENTS	53
2.8.1 Transfection of <i>env</i> and <i>rev</i> constructs	53
2.8.2 Harvesting of SNF and lysates	53
2.8.3 Detection of expressed proteins	54
2.9 SEQUENCING OF <i>ENV</i> CONSTRUCTS	56
2.9.1 Cycle sequencing reactions	56
2.10 PHYLOGENETIC ANALYSIS	57
2.10.1 Sequence assembly	57
2.10.2 Sequence analysis	58
2.10.3 Phylogenetic relationships	59
2.10.4 Analysis for intersubtype recombinants	62

INTRODUCTION

This chapter describes the different techniques and methods that were used to characterise and express the *env* gene from selected isolates. The genes were amplified by PCR and cloned into bacterial and mammalian expression vectors. The bacterial plasmids were sequenced to determine the nucleotide sequence of the selected *env* genes. The acquired sequences were analysed phylogenetically. The expression plasmids were transiently transfected into mammalian cells to determine whether Env proteins were expressed, and further whether these proteins were expressed at higher levels when Rev was present.

2.1 VIRUS ISOLATES

HIV-1 isolates were from HIV positive patients attending the Infectious Disease Clinic at the Tygerberg hospital from 1998 to 1999. A summary showing the clinical data of the patients can be viewed in **Table 1**. Ethical aspects of the study were approved by the University of Stellenbosch under project number 98/158. The subtypes of the viruses were determined by serotyping using a V3 competitive peptide enzyme-linked immunosorbent assay (cPEIA) (Engelbrecht *et al.*, 1999). Fifteen of the 18 isolates were serotyped as subtype C. TV017 reacted to subtype B-specific peptides, and TV016 reacted to both B- and D-specific peptides. TV015 was non-reactive to any of the peptides used.

A total of 18 isolates were expanded to high concentrations of virus stock in our department (Claassen, 1999). Briefly, PBMCs were isolated from both HIV-seronegative donors, as well as the above-mentioned HIV-1 positive patients. The HIV-seronegative donor blood was received from the Western Province Blood Transfusion Services. The PBMCs were co-cultured and the HIV-1 virus was allowed to grow to high concentrations (Gartner and Popovic, 1990). High molecular weight DNA was isolated (Sambrook *et al.*, 1989) and sequencing of a 650 bp fragment in the gp120 envelope, including the V3 loop, genotyped TV015, TV016 and TV017 as subtype B. The remainder of the isolates was genotyped as subtype C (Treurnicht *et al.*, 2002). All the isolates were of the NSI phenotype as determined by MT-2 cell

assays, and used CCR5 as their principal coreceptor as determined in HOS-CD4 cell lines expressing different coreceptors (Treurnicht *et al.*, 2002).

Table 1. Patient demographics

Patient	Age	Gender ^a	Ethnicity ^b	Risk factor ^c	Clinical Symptoms	CD4 Count	Serotype	Genotype	Source of infection ^d
TV001	36	M	MR	Het	Dermatitis	196	C	C	CT, WCP, RSA
TV002	35	M	MR	Het	Oral Candidiosis	309	C	C	CT, WCP, RSA
TV003	38	M	BI	Het	Pulmonary TB	117	C	C	ECP / CT, WCP, RSA
TV004	38	F	BI	Het	TB adenitis	435	C	C	Unknown
TV005	36	F	MR	Het	Oral/tracheal Candidiosis	64	C	C	CT, WCP, RSA
TV006	26	F	MR	Het	TB pleuritis	116	C	C	CT, WCP, RSA
TV007	24	F	BI	Het	Dermatitis	363	C	C	Gauteng, RSA
TV008	31	F	MR	Het	Bladder infection	125	C	C	Zimbabwe
TV009	31	M	BI	Het	Pulmonary TB	171	C	C	CT, WCP, RSA
TV010	21	F	BI	Het	Lymphadenopathy	167	C	C	CT, WCP, RSA
TV012	32	M	Ca	Het	Oral/ nasopharyngeal Candidiosis, diarrhoea	3	C	C	CT, WCP, RSA
TV013	43	F	MR	Het	Pneumonia	765	C	C	CT, WCP, RSA
TV014	33	M	BI	Het	Pulmonary TB	Un	C	C	Namibia
TV015	28	M	Ca	Hom	Wasting, stomach ulcers	420	NR	B	Unknown
TV016	39	M	MR	Bi	Loose stools	421	BD	B	Unknown
TV017	22	M	Ca	Het	Glandular fever, Candidiosis, wasting	200	B	B	Unknown
TV018	56	M	BI	Het	Pulmonary TB	201	C	C	ECP / CT, WCP, RSA
TV019	37	F	BI	Het	Asymptomatic	100	C	C	CT, WCP, RSA

TV number indicates the Tygerberg Virology number issued to the specific patient. No TV011 was issued.

a: M = Male, F = Female; b: MR = Mixed Race, BI = Black, Ca = Caucasian; c: Het = Heterosexual, Hom = Homosexual, Bi = Bisexual, d: CT = Cape Town, WCP = Western Cape Province, RSA = Republic of South Africa, ECP = Eastern Cape Province

2.2 BACTERIAL STRAINS

Both One Shot TOP10 (Invitrogen Corporation, Carlsbad, CA, USA) and Max Efficiency STBL2 (Gibco, Invitrogen Corporation, Carlsbad, CA, USA) competent cells were used for transformation procedures. The genotype for the TOP10 cells is $F^- mcrA \Delta(mrr - hsdRMS - mcrBC) \phi80lacZ \Delta M15 \Delta lacX74 deoR recA1 araD139 \Delta(ara-leu) 7697 galU galK rpsL (Str^R) endA1 nupG$. The STBL2 cells are derived from JM109/J5 and have a genotype of $mcrA \Delta(mcrBC - hsdRMS - mrr) recA1$.

2.3 PLASMID VECTORS AND CONSTRUCTS

2.3.1 pCR3.1

The pCR3.1 vector (Invitrogen Corporation, Carlsbad, CA, USA) was used in cloning procedures, which was needed in subsequent sequencing of the full-length *env* gene. The 5060 bp vector contains a Cytomegalovirus (CMV) immediate early promoter, T7 promoter and priming site, TA cloning site, polylinker, bovine growth hormone (BGH) reverse priming site, BGH polyadenylation signal, SV40 early promoter, SV40 origin, neomycin/kanamycin resistance gene, ampicillin resistance gene, pUC origin and f1 origin.

2.3.2 pCMVLink

The pCMVLink mammalian expression vector (Chiron Corporation, Emeryville, CA, USA) (Chapman *et al.*, 1991) was used for directional cloning procedures in order to express Env proteins. The 4351 bp vector contains a CMV major immediate-early (IE1) promoter, an intron A which acts as an enhancer for expression of proteins, especially glycoproteins from HIV-1, a BGH polyadenylation signal for termination, a SV40 origin and a kanamycin resistance gene.

2.3.3 Constructs received from outside sources

Clones were generated for native *env* genes of isolates TV001 (TV001c8.2 and TV001c8.5), TV002 (TV002c12.1), TV006 (TV006c9.1 and TV006c9.2), TV008 (TV008c4.3 and TV008c4.4), TV012 (TV012c2.1), TV014 (TV014c6.3 and TV014c6.4), and TV019 (TV019c5) by Chiron Corporation (Emeryville, CA, USA). The full-length *env* gene sequences obtained from the clones were used in sequence analysis and in determining phylogenetic relationships.

Codon-optimised versions of TV001 where the V2 loop had been deleted were received from Chiron Corporation (Emeryville, CA, USA). These included an oligomeric gp140 clone (pCMVLink o-gp140dv2 modified) of which the cleavage site was mutated to prevent cleavage into a gp120 and gp41, and a gp120 clone (pCMVLink gp120dv2 modified). The native gp160 *env* clone from TV001 (TV001c8.2), as well as the modified/codon-optimised clones was used in expression experiments. The pCMVLink o-gp140dv2 modified and pCMVLink gp120dv2 modified constructs served as positive controls in expression studies, as these constructs had already been shown to express at high levels (Barnett, S. (2002), personal communication, Chiron Corporation, Emeryville, CA, USA).

A *rev* construct consisting of the pCMVKm2 vector (Chiron Corporation, Emeryville, CA, USA) and a TV010 *rev* gene was used for co-transfection experiments (received from Scriba, T.J., Department of Medical Virology, University of Stellenbosch).

2.4 CELL LINES AND CELL CULTURE

The 293 cell line (Cat no. 11631 from Gibco, Invitrogen Corporation, Carlsbad, CA, USA) was used as host cell line for transfection procedures. The cell line is a permanent line established from primary embryonal human kidney, which was transformed with sheared human adenovirus type 5 DNA. Cells were recovered

from cryogenically frozen vials for culturing following manufacturers instructions. The cells were maintained in Dulbecco's Modified Eagle Medium (DME) (Sigma-Aldrich, Atlasville, RSA) containing 4500 mg/L glucose and L-Glutamine supplemented with 3.7 g/L NaHCO₃, 0.109 g/L NaH₂PO₄ (anhydrous), 0.1 mM MEM Non-Essential Amino Acids (Gibco, Invitrogen Corporation, Carlsbad, CA, USA), 10% Foetal Bovine Serum (Delta Bioproducts, Kempton Park, RSA), 100 mg/L penicillin (Novo-Pen, Novo Nordisk (Pty) Ltd, Johannesburg, RSA), and 100 mg/L streptomycin (Novo-Pen, Novo Nordisk (Pty) Ltd, Johannesburg, RSA). The cultures were maintained in an Autoflow CO₂ water-jacketed incubator (NuAire Inc., Plymouth, MN, USA) at 37°C containing a humidified atmosphere and 5% CO₂. The T-75cm² disposable sterile T-flasks with 0.2 µm vented caps were used for culturing (Corning Incorporated, NY, USA). The media was changed every 3-4 days and cells were sub cultured when a confluence of 80-90% was reached (approximately once a week). Sub culturing or trypsinisation was accomplished by incubating the cells in a 10% active trypsin versine (ATV) solution (135 mM NaCl, 5 mM KCl, 5.6 mM Glucose, 7 mM NaHCO₃, 0.5% Trypsin, 0.5 mM EDTA) for 2 minutes at room temperature, or until the cells detached from the flask wall, where after DME was added to inactivate the ATV. The cell suspension was centrifuged at 1000 x g for 5-10 minutes and supernatant was removed and discarded. The cells were then resuspended in DME and placed into T-75cm² disposable sterile T-flasks with 0.2 µm vented caps and maintained in an Autoflow CO₂ water-jacketed incubator (NuAire Inc., Plymouth, MN, USA) as described.

2.5 DNA CONSTRUCTS

2.5.1 Constructs for sequencing purposes

Constructs of native gp160 *env* genes in pCR3.1 vectors of selected isolates (TV003, TV004, TV006, TV007, TV010, TV013, TV018) were generated. These plasmids were then sequenced in order to obtain the full-length native *env* gene sequences.

2.5.2 Constructs for expression experiments

Constructs containing native *env* genes in pCMVLink were generated (TV004, TV006 and TV018) for transfection studies in order to observe protein expression.

2.6 PCR AMPLIFICATION

2.6.1 Full-length gp160 *env* genes for pCR3.1 cloning

Near full-length 9 kb fragments of the HIV-1 genome from isolates TV003, TV004, TV006, TV007, TV010, TV013 and TV018 were received from Engelbrecht, S., (Department of Medical Virology, University of Stellenbosch). Briefly, the Expand Long Template Kit (Roche Molecular Biochemicals, Mannheim, Germany) was used to amplify the fragment from high molecular weight DNA together with the primer pair UP1A and LOW2 (Gao *et al.*, 1998). The QIAEX II Gel Extraction kit (QIAGEN, GmbH, Germany) was used to purify the fragments, which were excised after separation on an electrophoresis gel. The prepared 9 kb PCR products served as template for amplification of *env* gene fragments using the Expand High Fidelity PCR System (Roche Molecular Biochemicals, Mannheim, Germany). The Expand high fidelity PCR system was used for all amplifications, as the yield is increased two-fold and the fidelity of the enzyme (8.5×10^{-6}) is three times greater than the fidelity observed for *Taq* DNA polymerase (2.6×10^{-5}) (Barnes, 1994). The primer pair *Env-st*: GAAAGAGCAGAAGACAGTGGCAA and *Env-end*: CTTTTTGACCACT TGCCACCCAT was used. The PCR was carried out in a total volume of 100 μ l, containing 1.5 mM MgCl₂, 200 μ M deoxynucleoside triphosphates, 40 pmol of each primer, and 3.5 U of Expand High Fidelity PCR System enzyme mix. Samples were subjected to a denaturing step at 94°C for two minutes, followed by 40 cycles of a denaturing step at 94°C for 30 seconds, an annealing step at 50°C for

30 seconds, and an extension step at 72°C for 3 minutes. Extended elongation followed for 15 minutes at 72°C.

PCR products were visualised by agarose gel electrophoresis on a 0.8% agarose gel in TAE buffer (0.04 M Tris-acetate, 0.001 M EDTA) containing 5 µg/ml ethidium bromide in the agarose gel. A 1 kb DNA ladder (Promega, Madison, WI, USA) was included in the electrophoresis to compare DNA fragment sizes, unless stated otherwise.

2.6.2 Native *env* genes for directional cloning

The TV004cC300, TV006cE9 and TV018cF1027 pCR3.1 clones served as template for amplification of *env* gene fragments to generate directional clones to use in expression experiments using the Expand High Fidelity PCR System (Roche Molecular Biochemicals, Mannheim, Germany). Primer pairs for TV006cE9 and TV018cF1027 were designed, which contained restriction sites to enable directional cloning. Chiron Corporation (Emeryville, CA, USA) designed a primer pair for TV004cC300. The 5'- primers contained an *EcoRI* restriction site, and the 3'- primers contained a *XhoI* site. The primers are listed in **Table 2**. The PCR was standardised by performing a MgCl₂ concentration titration. Concentrations included 1.0 mM, 1.5 mM, 2.0 mM, 2.5 mM, 3.0 mM, 3.5 mM and 4.0 mM. The standardised PCR was carried out in a total volume of 100 µl, containing 4.0 mM MgCl₂, 200 µM deoxynucleoside triphosphates, 40 pmol of each primer, and 2.6 U of Expand High Fidelity PCR System enzyme mix.

Samples were subjected to a denaturing step at 94°C for two minutes, followed by 10 cycles of a denaturing step at 94°C for 30 seconds, an annealing step at 65°C for 30 seconds, and an extension step at 72°C for 2 minutes. Thirty cycles followed consisting of a denaturing step at 94°C for 30 seconds, an annealing step at 65°C for 30 seconds, and an

extension step at 72°C for 2 minutes with a 10 second increment per cycle. Extended elongation followed for 10 minutes at 72°C.

PCR products were visualised by agarose gel electrophoresis as described earlier.

Table 2. Primers used for PCR amplification

Template	5'- primer	3'- primer
TV004cC300	5'- AGCACGGAATTCATG AGAGTGAGGGAGATAC CGAGGAATTGGCAACA -3'	5'- AGCACGCTCGAGTTA TAGCAAAGCTGCTTCAA AGCCCTGTCTTATTC -3'
TV006cE9	5'- AGCACGGAATTCATG AGAGTGACGGGCATAC TGAGGAATTATCCACA -3'	5'- AGCACGCTCGAGTTA TAGTAAAGCAGTTTCAA AGCCCTGTCTTATTC -3'
TV018cF1027	5'- AGCACGGAATTCATG AGAGTGATGGGGATCA AGAGGAATTGTCAACA -3'	5'- AGCACGCTCGAGTTA TAGCAAAGCTGCTTCAA AGCCCCGTCTTATTC -3'

2.7 CLONING EXPERIMENTS

2.7.1 Ligation and transformation reactions

Plasmid vectors and inserts were prepared for ligation for both pCR3.1 and pCMVLink cloning. TA-cloning was used for cloning of PCR products into the pCR3.1 vectors, as the system eliminates any enzymatic modifications of the PCR product and does not require the use of PCR primers that contain restriction sites. Because the Expand high fidelity PCR system generates PCR products, which resemble a mixture of 3' single A-overhang products and blunt ended products analogous to *Taq* DNA polymerase, 3' A-overhangs were added post amplification to ensure that all the products contained a 3' A-overhang by adding 1 U of *Taq* polymerase (Promega, Madison, WI, USA) to the PCR tubes on ice. The samples were incubated at 72°C for 8 to 10 minutes. The DNA was purified using the QIAquick PCR Purification Kit

(QIAGEN, GmbH, Germany) following the manufacturer's protocol. The manufacturers of pCR3.1 (Invitrogen Corporation, Carlsbad, CA, USA) supply the plasmid vector in a linear form containing T-overhangs and no preparations were therefore needed before ligation.

The PCR products generated for pCMVLink cloning were separated by electrophoresis on a large 0.8% agarose gel. The gel was run at 30V/cm for 18 hours. The gel was stained with TAE buffer (0.04M Tris-acetate, 0.001M EDTA) containing 5 µg/ml ethidium bromide for 30 minutes. The 2.5 kb *env* gene PCR products were excised from the agarose and purified with the QIAGEN Gel extraction kit (QIAGEN, GmbH, Germany) following manufacturers instructions. Both the pCMVLink mammalian expression vector and the purified gel extracted PCR products were digested with *EcoRI* and *XhoI* in Buffer H (90 mM Tris-HCl, 10 mM MgCl₂, 50 mM NaCl, pH 7.5) for 2 hours at 37°C (Promega, Madison, WI, USA). The DNA was purified using the QIAquick PCR Purification Kit (QIAGEN, GmbH, Germany) following the manufacturers protocols.

The PCR products were ligated to the pCR3.1 and pCMVLink vectors at vector:insert ratios of 1:10 and 1:2 respectively using T4 DNA Ligase (Promega, Madison, WI, USA). The ligation reaction, with a total reaction volume of 10 µl, was incubated at 15°C in an incubator (Mettler, Schwabach, Germany) overnight for optimal ligation results.

The plasmids containing the ligated PCR products were transformed into both One shot TOP10 and Max Efficiency STBL2 competent cells. The TOP10 cells were originally used for the transformation procedures until it was found that, in DNA prepared from bacteria on a large scale, the gene was repeatedly excised or deleted from the plasmid. The excision of the gene generated bacteria resistant to the selective antibiotic present in the LB agar plates and liquid growth media. This reaction has been reported for palindromic sequences (Collins, 1981; Hagan and Warren, 1983) and sequences containing short direct

repeats (Collins *et al.*, 1982; Jones *et al.*, 1982) in *Escherichia coli* plasmids. Another group has reported deletions in the gp41 region extending into their cloning vector, suggesting instability of the recombinant plasmid during large-scale growth in bacteria (Gao *et al.*, 1994). The deletion of the gene was prevented by using the STBL2 strain for transformation procedures, as this strain was manufactured for the use of unstable retroviral sequences (Trinh *et al.*, 1994). The deletion was also prevented by reducing the growth temperature to 30°C when working with the TOP10 cells. Lowering the growth temperature has been shown to minimize the instability of plasmids containing HIV-1 proviral genomes (Joshi and Jeang, 1993).

One Shot TOP10 as well as Max Efficiency STBL2 competent cells was transformed with 2 µl of the ligation mixtures. Control reactions included a positive pUC19 ligation reaction from the competent cell kits, a negative reaction where linear vector was transformed into the cells, a second negative reaction where vector was ligated with T4 DNA ligase without adding prepared PCR product insert, and a cell only control reaction. Manufacturers instructions from the Eukaryotic TA Cloning Kit (Invitrogen Corporation, Carlsbad, CA, USA) were followed for transformation of the TOP10 cells with the following modification: the vials containing the cells and ligation mixture were incubated in Labcon shaking incubator (Labmark, Roodepoort, RSA) at 225 rpm and 30°C for one and a half hours after the heat shock treatment, instead of incubating the vials at 37°C for one hour. After adding SOC media (2% Tryptone, 0.5% Yeast extract, 0.05% NaCl, 2.5 mM KCl, 10 mM MgCl₂, 20 mM Glucose) to the one and half hour incubated vials, cells were pelleted by centrifugation at 6000 x g for 2 minutes. The supernatant was removed allowing approximately 50 µl to remain in the vials and the pellet was then resuspended in the 50 µl of supernatant and plated onto Luria-Bertani (LB) agar plates (10 g/L bacto-tryptone, 5 g/L bacto-yeast extract, 10 g/L NaCl, 15 g/L bacto-agar) containing 50 mg/ml ampicillin. The LB agar plates were incubated at 30°C overnight. The

STBL2 cells were transformed as per manufacturers instructions, but cells were also pelleted by centrifugation as described above before plating onto LB plates containing 50 mg/ml kanamycin. The cells were pelleted before plating to ensure that all possible transformed bacterial cells were included when plating the mixture onto LB plates.

2.7.2 Screening of recombinant clones

Colonies from the LB agar plates were inoculated into 2 ml LB media (10 g/L bacto-tryptone, 5 g/L bacto-yeast extract, 10 g/L NaCl) containing the appropriate antibiotic and incubated in a Labcon shaking incubator (Labmark, Roodepoort, RSA) at 225 rpm and 30°C for 16 hours. DNA was isolated on a small scale for screening purposes following the Small-scale preparations of plasmid DNA protocol (Sambrook *et al.*, 1989). The plasmid DNA was separated by gel electrophoresis on 0.8% agarose gels. A Supercoiled DNA ladder (Promega, Madison, WI, USA) was added as reference for fragment size. Fragments migrating at a slower pace compared to vector alone indicate possible positive clones and were digested with restriction enzymes to verify insert presence. Samples were separated by gel electrophoresis to visualise excised inserts.

The pCR3.1 plasmid DNA samples of these longer fragments were digested with the restriction enzyme *EcoRI* (Promega, Madison, WI, USA) in Buffer H (90 mM Tris-HCl, 10 mM MgCl₂, 50 mM NaCl, pH 7.5).

The pCMVLink plasmid DNA was digested with *EcoRI* and *XhoI*, (Promega, Madison, WI, USA) in Buffer H (90 mM Tris-HCl, 10 mM MgCl₂, 50 mM NaCl, pH 7.5).

Orientation of the genes present in the pCR3.1 vectors was determined by restriction mapping with *NdeI* (Promega, Madison, WI, USA) in Buffer D (6 mM Tris-HCl, 6 mM MgCl₂, 150 mM NaCl, 1 mM DTT, pH

7.9). Digestion with *NdeI* yields a distinct fragment pattern, which is indicative of insert orientation. The C.BW.96BW0502 subtype reference sequence (the subtype is followed by a 2 letter abbreviation of the isolation country, followed by the sequence name) (HIV Sequence Compendium, 2001) was restriction mapped with *NdeI* to determine which patterns would be observed for an insert in the correct as well as reversed orientation. The reference strain has a *NdeI* restriction site at position 177, while the pCR3.1 sequence has a *NdeI* restriction site at position 260. If the reference sequence were inserted into the pCR3.1 plasmid, 2 fragments of 655 and 6970 bp each would indicate that the gene was inserted into the vector in the correct orientation. Two fragments of 2869 and 4758 bp each would be indicative of the insert being inserted into the vector in a reversed orientation. Digested samples were separated by gel electrophoresis as described before.

DNA was isolated from bacteria on a larger scale after confirmation of *env* gene presence in the plasmid DNA using a QIAfilter Midiprep Kit (QIAGEN, GmbH, Germany). The manufacturer's instructions were followed. The concentration, as well as the purity of the plasmid DNA isolated was determined using the following calculations (Sambrook *et al.*, 1989):

$$\text{DNA concentration} = \frac{\text{OD}_{260}}{20} \times \text{dilution factor}$$

$$\text{DNA purity} = \frac{\text{OD}_{260}}{\text{OD}_{280}}$$

Glycerol stocks were prepared from all positive clones by adding 0.15 ml sterile glycerol to 0.85 ml bacterial culture (Sambrook *et al.*, 1989). The cultures were transferred to sterile screw cap cryogenic vials and stored at -70°C.

All constructs to be used in expression experiments were screened to confirm *env* insert presence. The TV001c8.2, LinkTV004cC300, LinkTV006cE9, LinkTV018cF1027, pCMVLink o-gp140dv2 modified and pCMVLink gp120dv2 modified constructs were digested with *EcoRI* and *XhoI*. The *rev* construct was digested with *EcoRI* and *Sall*. The restricted DNA was visualised on an agarose gel as described earlier.

2.8 EXPRESSION EXPERIMENTS

2.8.1 Transfection of *env* and *rev* constructs

The 293 cell line was recovered, cultured and maintained as described earlier for use in transfection using DNA constructs. The 293 cells were seeded at 7×10^5 cells/ml and 1.4×10^6 cells were plated per well in sterile cell culture cluster 6-well plates (Corning Incorporated, NY, USA) one day before transfections commenced. Transfections were performed with cells which were 70-80% confluent.

The experiment was set up to include 2 wells per transfection for each construct. The cells were transfected with 1 μ g plasmid DNA using the Transfast transfection reagent (Promega, Madison, WI, USA) according to manufacturers instructions. All the wild type native gp160 *env* constructs (TV001c8.2, LinkTV004cC300, LinkTV006cE9 and LinkTV018cF1027) were co-transfected with the TV010 *rev* construct to determine Rev dependence. These experiments included the transfection of 1 μ g plasmid DNA from the *env* and *rev* constructs each. Negative/mock reactions were included as controls where the cells were treated with the Transfast reagent without the addition of plasmid DNA constructs.

2.8.2 Harvesting of SNF and lysates

Supernatant fluid (SNF) was harvested from the culture 72 hours post transfection. The SNF was filtered through FP30/0.2 CA-5 disposable

filter units (Schleicher & Schuell, Germany) to remove cell debris. Proteins were isolated from the filtered SNF (4 ml from duplicate wells) by adding Galanthus Nivalis-agarose (GNA) (Vector Laboratories, CA, USA) and incubating the mixture on a Orbital shaker (Stuart Scientific, Bibbi, UK) at 4°C for a minimum of 3 hours (Srivastava *et al.*, 2002). Glycoproteins in the SNF bound to the agarose as a result of the affinity to the lectin. The proteins were eluted by the addition of Methyl Mannose Pyranoside (MMP) (10 mM NaH₂PO₄, 100 mM NaCl, pH 7.4). The MMP buffer is a sugar solution, which inhibits the binding of the lectin to the glycoproteins. The SNFs were stored in aliquots at -20°C until further use.

Cell lysates were prepared by washing the cells with phosphate buffered saline (PBS) (80 g/L NaCl, 2 g/L KCl, 9.1 g/L Na₂H₂PO₄ (anhydrous), 1.2 g/L KH₂PO₄). The wash mixture was transferred to a microcentrifuge tube and centrifuged at 6000 x g for 3 minutes. The supernatant was discarded. Lysis buffer (0.14 mM NaCl, 1.5 mM MgCl₂, 10 mM Tris-HCl at pH 7.4, Nonident-P40) was added to the cells and incubated for 30 minutes at room temperature. The lysis mixture was then centrifuged at 10 000 x g for 3 minutes to remove cell debris. The lysates were stored in aliquots at -70°C until further use.

2.8.3 Detection of expressed proteins

SNF and cell lysates were prepared for electrophoreses to separate proteins by sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) by adding NuPage LDS sample buffer (Invitrogen Corporation, Carlsbad, USA) and Nupage sample reducing agent (Invitrogen Corporation, Carlsbad, USA). The samples were denatured by boiling at 95°C for 10 minutes and then placed on ice. The samples were loaded onto 4-12% Bis-Tris polyacrylamide gels (Invitrogen Corporation, Carlsbad, USA) and proteins were separated for one and a half hours under denaturing conditions at 115 mA in

NuPage MOPS SDS running buffer supplemented with NuPage Antioxidant (Invitrogen Corporation, Carlsbad, USA). A positive control included the HIV-1 SF2 recombinant gp120 (Haigwood, 1990; Levy *et al.*, 1984; Sanchez-Pescador *et al.*, 1985; Scandella *et al.*, 1993) (received from the NIH AIDS Research & Reference Reagent Program) and negative controls comprised the negative transfection control SNF and cell lysates. A full range rainbow marker (Amersham Pharmacia Biotech Limited, Buckinghamshire, England) was included in the electrophoresis to compare protein fragment sizes.

The polyacrylamide gels were trimmed and soaked in Towbin transfer buffer (25mM Tris, 192mM Glycine, 10% methanol) for 5 minutes. The transfer electroblot sandwich from the Mini VE vertical electrophoresis system blot model (Hoeffer Scientific Instruments, San Francisco, CA, USA) was assembled according to manufacturers instructions. The proteins were transferred to Hybond ECL nitrocellulose membrane (Amersham Pharmacia Biotech Limited, Buckinghamshire, England) for 1 hour at 350 mA.

The membranes were blocked in PBS-T (20 mM Tris-Cl at pH 7.5, 150 mM NaCl, 0.5% Tween-20) containing 3% bovine serum albumin (BSA) for 16 hours at room temperature with gentle agitation on a STOVALL Belly dancer (Life Science incorporated, Greensboro, NC, USA). HIV-1 positive patient sera (primary antibody) were diluted 1:100 with PBS-T and added to the membranes for a two-hour incubation with gentle agitation on a STOVALL Belly dancer (Life Science incorporated, Greensboro, NC, USA). Washing commenced consisting of 3 washes of 3 minutes each with PBS-T. Goat Anti-Human IgG (H + L) Horseradish Peroxidase (Pierce, Rockford, USA) (secondary antibody) was diluted 1:2500 with PBS-T and added to the membranes for a one-hour incubation with gentle agitation on a STOVALL Belly dancer (Life Science incorporated, Greensboro, NC, USA). Washing commenced as described above.

Proteins were visualised by ECL chemiluminescence using ECL western blotting analysis system detection reagents (Amersham Pharmacia Biotech Limited, Buckinghamshire, England). Hyperfilm ECL chemiluminescence film (Amersham Pharmacia Biotech Limited, Buckinghamshire, England) was exposed for 1 minute and developed.

A log graph of molecular weight (Mr) versus relative electrophoretic mobility was drawn to determine detected proteins Mr. The distance travelled by fragments of known Mr (from the full range rainbow marker) were plotted on the log graph. Expressed Env protein Mr was then determined by plotting the distances travelled by the proteins on the graph constructed from known fragment Mr (Mathews and van Holde, 1990).

2.9 SEQUENCING OF *ENV* CONSTRUCTS

2.9.1 Cycle sequencing reactions

Sequencing was performed to acquire the nucleotide sequences of selected *env* genes, as well as to determine the orientation in which genes were inserted into the pCR3.1 plasmid vector. pCMVlink clones were sequenced to ensure that the *env* gene sequences were intact. Sequences were acquired using the ABI Prism™ 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Manufacturer's instructions of the ABI PRISM™ Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster city CA 94404) were followed with three modifications. Firstly, the kit instructions suggest the use of 320 ng DNA in a sequencing reaction. Instead, 1 µg plasmid DNA was used. Secondly, a few sequencing reactions were performed using 4 µl (half reaction) of the Terminator ready reaction mix (TTRM) from the ABI PRISM™ Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit instead of 8 µl (full reaction) so as to limit reagent use. Lastly, the reactions contained 5 or 10 pmol primer (the kit specifies using 5 pmol primer) and 5 x Sequencing Buffer (Applied Biosystems, Foster City,

CA). A total of 19 primers were used to amplify the full-length *env* gene (**Table 3**). The primers were designed by Sanders-Buell *et al.* (1995). The T7 forward and BGH reverse primers anneal in the pCR3.1 vector, and were used to sequence the ends of the genes, as well as to determine the orientation of the genes in the pCR3.1 plasmid vector. The pCMVseqF and pCMVseqR primers (Scriba T.J., unpublished) anneal in the pCMVLink vector, and were used to sequence the ends of the cloned genes to ensure intact restriction enzyme sites. Annealing temperatures for primers were 5-10°C lower than the melting temperatures (T_m) supplied by the manufacturers (Applied Biosystems, Foster City, CA or Integrated DNA Technologies Inc., Coralville, IA, USA).

Sequencing reactions were subjected to 25 cycles of a denaturing step at 96°C for 10 seconds, an annealing step at the annealing temperature indicated in **Table 3** for 30 seconds, and an extension step at 60°C for 4 minutes. The products were then purified using the DyeEx Spin protocol for Dye-Terminator Removal (QIAGEN, GmbH, Germany) before inserting them into the ABI Prism™ 310 Genetic Analyzer.

2.10 PHYLOGENETIC ANALYSIS

2.10.1 Sequence assembly

The *env* gene nucleotide sequences were assembled using the Auto Assembler 2.1 program (Applied Biosystems, Foster City, CA). Amino acid sequences were predicted in DNA Strider (Marck, 1988) and open reading frames were detected in the same program. The nucleotide sequences were submitted to GenBank under accession numbers AF391230 – AF391250 (Engelbrecht *et al.*, 2001). These sequences included the sequences from our laboratory, as well as sequences generated from clones created using the isolates listed in **Table 1** by Chiron Corporation (Emeryville, CA, USA).

2.10.2 Sequence analysis

An amino acid alignment was constructed to manually search for specific motifs and regions.

Table 3. Sequencing primers, their respective sequences and annealing temperatures used in cycle sequencing reactions.

Primer	Sequence	T _m	Annealing temp. (°C)
E00	5'- TAG AAA GAG CAG AAG ACA GTG GCA ATG A -3'	58.3°C	55°C
E05	5'- TAT TTG AGG GCT TCC CAC CCC C -3'	58.4°C	55°C
E15	5'- CTC TCT CTC CAC CTT CTT CTT C -3'	54.7°C	50°C
E20	5'- GGG CCA CAC ATG CCT GTG TAC CCA CAG -3'	65.7°C	60°C
E30	5'- GTG TAC CCA CAG ACC CCA GCC CAC AAG -3'	65.7°C	60°C
E45	5'- CCT GCC TAA CTC TAT TCA C -3'	48.7°C	45°C
E55	5'- GCC CCA GAC TGT GAG TTG CAA CAG ATG -3'	62.6°C	55°C
E60	5'- TAA TCA GTT TAT GGG ATC AAA GC -3'	49.8°C	45°C
E70	5'- GGG ATC AAA GCC TAA AGC CAT GTG TAA -3'	58.1°C	55°C
E80	5'- CCA ATT CCC ATA CAT TAT TGT G -3'	49.1°C	45°C
E120	5'- GTA GAA ATT AAT TGT ACA AGA CCC -3'	50.4°C	45°C
E160	5'- GTG GGA ATA GGA GCT GTG TTC CTT GGG -3'	62.6°C	55°C
E170	5'- AGC AGG AAG CAC TAT GGG -3'	50.1°C	45°C
E180	5'- GTC TGG TAT AGT GCA ACA GCA -3'	52.2°C	45°C
E260	5'- TTC AGC TAC CAC CGC TTG AGA GAC T -3'	59.2°C	53°C
T7 forward	5'- GTA ATA CGA CTC ACT ATA -3'	42.3°C	45°C
BGH reverse	5'- TAG AAG GCA CAG TCG AGG -3'	54.0°C	55°C
pCMV seqF	5'- AGT CTG AGC AGT ACT CGT TG -3'	54.9°C	50°C
pCMV seqR	5'- GCT GGC AAC TAG AAG GCA CA -3'	58.8°C	53°C

The alignment contained all the translated *env* gene sequences generated in this study, translated sequences generated by Chiron Corporation (Emeryville, CA, USA) for clones TV001c8.2, TV001c8.5, TV002c12.1, TV006c9.1, TV006c9.2, TV008c4.3, TV008c4.4, TV012c2.1, TV014c6.3, TV014c6.4 and TV019c5, translated sequences for the pCMVLink constructs, and the amino acid sequences for the B.FR.HXB2 and C.ZA.97ZA012 strains. The B.FR.HXB2 sequence was used to identify conserved glycosylation sites and cysteine residues, and the remaining sequences were compared to the C.ZA.97ZA012 sequence, as this was the first South African full-length *env* sequence described in a near full-length molecular clone (Rodenburg *et al.*, 2001). The gp120 and gp41 regions, signal peptide, variable loops (V1-V5), CD4 binding sites, fusion domain, immunodominant region, tryptophan-rich sequence, membrane spanning domain, and the intracellular cytoplasmic domain/tail were identified in the sequences. Further, the crown sequence within the V3 loop and the fusogenic site within the fusion domain were also identified. The phenotype of the sequences were predicted by locating amino acids with specific charges at positions 11, 25, and 32 in the V3 loop (Björndal *et al.*, 1999). The overall charges of the V3 regions were also determined (Björndal *et al.*, 1999; Treurnicht *et al.*, 2002). The tetrapeptide crown sequences were further analysed to predict the phenotype (Abebe *et al.*, 1999).

Hydrophobicity was determined for pCMVLink amino acid sequences in the cases where amino acid differences from the original pCR3.1 sequences were observed. This was performed by determining whether amino acids are hydrophilic or hydrophobic (Zubay, 1993).

2.10.3 Phylogenetic relationships

Phylogenetic relationships were determined for all *env* gene sequences, as well as the sequences generated by Chiron Corporation (Emeryville, CA, USA). The sequences from Chiron Corporation

included TV001c8.2, TV001c8.5, TV002c12.1, TV006c9.1, TV006c9.2, TV008c4.3, TV008c4.4, TV012c2.1, TV014c6.3, TV014c6.4 and TV019c5. All the *env* sequences generated, as well as the sequences from Chiron Corporation will collectively be referred to as the TV *env* sequences from here on.

The subtypes of the TV *env* sequences were determined by comparing the sequences to the reference strains from the Los Alamos HIV sequence database (HIV Sequence Compendium, 2001). Phylogenetic relationships of the TV *env* sequences were estimated from comparisons to previously reported subtype C nucleotide sequences. The sequences used to compare the TV *env* sequences to, as well as their accession numbers and references are listed in Appendix A. Nucleotide and amino acid sequences were aligned in ClustalX 1.81 (Higgins and Sharp, 1988; Thompson *et al.*, 1997). The alignment parameters were set to 15.00 for gap opening and 6.66 for gap extension and the IUB DNA weight matrix was used (default parameters of program) when doing the nucleotide multiple alignments. Amino acid alignments were performed using a gap opening of 10.00 and gap extension of 0.20 and the Gonnet series protein weight matrix was used (default parameters of program). Sequence alignments were manually adjusted in Genedoc version 2.6 (Nicholas *et al.*, 1997; Nicholas and Nicholas, 1997).

Pairwise evolutionary distances for nucleotide sequences were estimated using Kimura's two-parameter method (Kimura, 1980) in the DNADist program, and for amino acid sequences using the Dayhoff PAM distance matrix (Dayhoff *et al.*, 1979) in the ProtDist program, both from the PHYLIP package (Felsenstein, 1993; Felsenstein, 1989). Phylogenetic trees were constructed by the neighbor-joining method (Saitou and Nei, 1987) using the Neighbor program from the PHYLIP package (Felsenstein, 1993; Felsenstein, 1989). The reliability was estimated from 100 bootstrap replicates (Felsenstein, 1985).

Phylogenetic trees were viewed in TreeView version 1.5.2 (Page, 1996).

Distances of the TV *env* sequences to subtype C sequences from Botswana, India and previously published South African sequences were used in statistical analysis. Intersample diversity observed for strains from Botswana, India and previously published South African sequences combined with the TV *env* sequences, as well as TV *env* sequences on their own were compared to the variation seen when compared to the subtype C consensus sequence from the Los Alamos HIV sequence database (HIV Sequence Compendium, 2001). The consensus sequence was based on HIV database alignments. The intersample diversity for the TV *env* sequences was also compared to the variation detected when compared to a generated South African consensus. The South African consensus sequence was generated in Genedoc version 2.6 (Nicholas *et al.*, 1997; Nicholas and Nicholas, 1997) from an alignment containing all the published South African sequences in the Los Alamos HIV sequence database to date. The mean distances of the sequences were estimated by the sum of all the distances calculated in DNADist or ProtDist from the PHYLIP package (Felsenstein, 1993 and 1989) divided by the number of distances using Microsoft Excel 2000 software (Microsoft Corporation). As indicated (Altman, 1991), the median is a value, which comes halfway when the data are ranked in order. A percentile is the value below which a given percentage of the values occur. The 25th and 75th percentiles, together with the median divide the data into four equally populated subgroups. The standard deviation (SD) quantifies the variability based on the idea of averaging the distance each value is from the mean. A confidence interval (CI) is a range of values that can be confident includes the true value. The median, 25th and 75th percentiles, SD for a specific population (size, n) as well as CI of 95 and 99% were calculated in SigmaPlot 2002 for Windows, version 8.0 (SPSS Inc.) and used to construct box plots for visual presentation of observed diversity.

2.10.4 Analysis for intersubtype recombinants

The TV *env* sequences were analysed for any evidence of recombination using Simplot version 2.5 (Lole *et al.*, 1999). The program calculates and plots the percent identity of a query sequence to a panel of reference sequences in a sliding window. It was found that this method represents an excellent initial screening tool to identify recombinant viral sequences (Gao *et al.*, 1998). The percentage similarity between the TV *env* sequences and reference subtypes A-D, F-H, J and K (HIV Sequence Compendium, 2001) was determined by moving a window of 200 bp along the *env* gene alignment in 20 bp increments (default parameters of program).

CHAPTER 3 RESULTS

	p. No.
INTRODUCTION	64
3.1 PCR AMPLIFICATION	64
3.1.1 Full-length gp160 <i>env</i> genes for pCR3.1 cloning	64
3.1.2 Native <i>env</i> genes for directional cloning	64
3.2 CLONING EXPERIMENTS	65
3.2.1 Ligation and transformation reactions	65
3.2.2 Screening of recombinant clones	65
3.3 EXPRESSION EXPERIMENTS	69
3.3.1 Culture and transfection of cell lines	69
3.3.2 Detection of proteins	70
3.4 SEQUENCING OF <i>ENV</i> CONSTRUCTS	71
3.4.1 Cycle sequencing reactions	71
3.5 PHYLOGENETIC ANALYSIS	72
3.5.1 Sequence assembly	72
3.5.2 Sequence analysis	72
3.5.3 Phylogenetic relationships	76
3.5.4 Analysis for intersubtype recombinants	78
FIGURES	
Figures 12-36	80 - 104

INTRODUCTION

This chapter lists the results generated to characterise the *env* gene from selected isolates and shows the expression obtained by transient transfection. Clones were created, which were sequenced to obtain the full-length *env* gene sequences. The sequences were analysed and phylogenetic relationships were determined. Expressed proteins were visualised after transient transfection in the absence and presence of Rev. All figures created appear at the end of the chapter.

3.1 PCR AMPLIFICATION

3.1.1 Full-length gp160 *env* genes for pCR3.1 cloning

The *env* gene fragments from different isolates obtained by amplification using the 9 kb PCR product as template were run on a 0.8% agarose gel (**Figure 12**). All the PCRs were successful and products of approximately 2500 bp in size were generated, which correlates with the size of the *env* gene. The negative control containing all the PCR reagents and primers with no added template DNA showed only primer excess and no amplification product, indicating that there was no contamination in the PCR reactions.

3.1.2 Native *env* genes for directional cloning

The *env* gene fragments generated by amplification using the pCR3.1 clones TV004cC300, TV006cE9 and TV018cF1027 as templates were run on a 0.8% agarose gel (**Figure 13**). All the PCRs were successful and products of approximately 2500 bp in size were generated, which correlates with the size of the *env* gene. A few non-specific fragments were visible; therefore the PCR products were run on a large gel to enable the excision of

the *env* fragments (described below). The negative control containing all the PCR reagents and primers with no added template DNA showed no amplification product, indicating that there was no contamination in the PCR reactions.

3.2 CLONING EXPERIMENTS

3.2.1 Ligation and transformation reactions

The PCR products generated for pCMVLink constructs contained non-specific fragments, and the products were therefore run on a large agarose gel to allow the excision of the *env* fragments. **Figure 14** shows the large 0.8% agarose gel where the fragments were separated.

The pUC19 control LB agar plates with antibiotic contained >200 colonies, therefore indicating that cells used were competent. The transformation efficiency for the pCR3.1 cloning was 2×10^7 CFU/ μ g and 4×10^6 CFU/ μ g for the pCMVLink cloning. The LB agar plate with no antibiotic harboured more than 200 colonies for the cell only control, which indicated that the bacterial cells used were viable. The re-ligation of linear plasmid vector without the addition of insert was rare and showed a transformation efficiency of 2×10^4 CFU/ μ g.

3.2.2 Screening of recombinant clones

Small-scale prepared plasmid DNA was visualised on 0.8% agarose gels, and fragments migrating at a slower pace compared to vector alone indicate possible positive clones and were digested with restriction enzymes to verify insert presence.

An agarose gel showing pCR3.1 plasmid DNA is shown in **Figure 15**. Miniprep DNA where no *env* insert was present

displayed fragment sizes of approximately 5000 bp, which correlated with the vector's 5060 bp size. Possible positive clones were identified by fragments of approximately 7500 bp, which correlated with the size of the vector (5060 bp) and the addition of a 2500 bp insert.

An agarose gel showing pCMVLink plasmid DNA is shown in **Figure 16**. Preps where no *env* insert was present displayed fragment sizes of between 4000 - 4500 bp, which correlated with the vector's 4351 bp size. Possible positive clones were identified by fragments of between 6500 - 7000 bp, which correlated with the size of the vector (4351 bp) and the addition of a 2500 bp insert.

Figure 17 shows pCR3.1 plasmid DNA that was digested with *EcoRI*. Positive clones were identified by plasmid DNA showing the excision of an approximately 2500 bp fragment, which correlates with the size of the *env* gene. In total, 14 clones were generated in the pCR3.1 vector for 7 different isolates. These included 1 clone each for TV003 and TV004, 3 clones for TV006, 2 clones for TV007, 4 clones for TV010, 2 clones for TV013 and 1 clone for TV018.

Figure 18 shows pCMVLink plasmid DNA that was digested with *EcoRI* and *XhoI*. Positive clones were identified by plasmid DNA showing the excision of an approximately 2500 bp fragment, which correlates with the size of the *env* gene. One clone was selected each for TV004, TV006 and TV018 in expression experiments. **Table 4** gives a summary of all the clones generated.

Table 4. Summary of clones created.

Isolate	Clone generated	Plasmid vector
TV003	TV003cE260	pCR3.1
TV004	TV004cC300	pCR3.1
	LinkTV004cC300	pCMVLink
TV006	TV006cE5	pCR3.1
	TV006cE9	pCR3.1
	TV006cF3	pCR3.1
	LinkTV006cE9	pCMVLink
TV007	TV007cB104	pCR3.1
	TV007cB105	pCR3.1
TV010	TV010cC3	pCR3.1
	TV010cD7	pCR3.1
	TV010cD9	pCR3.1
	TV010cD10	pCR3.1
TV013	TV013cB20	pCR3.1
	TV013cH17	pCR3.1
TV018	TV018cF1027	pCR3.1
	LinkTV018cF1027	pCMVLink

The orientation of the genes present in the pCR3.1 vectors was determined by digesting the positive clone plasmid DNA with *Nde*I. **Figure 19** shows a 0.8% agarose gel where different fragment patterns are observed. Only the TV004cC300 clone showed the predicted fragment pattern of a reversed orientation, therefore yielding 2 fragments of approximately 2800 and 4700 bp each. The rest of the clones showed varying fragment patterns, as a result of the variability seen in *env* genes. The

variability was therefore displayed by sequences having *NdeI* restriction sites at different positions in the sequences, as well as a varying number of *NdeI* restriction sites. TV003cE260 showed 3 fragments of approximately 1300, 1500 and 4700 bp each. The restriction of TV006cE9, TV006cE5 and TV006cF3 yielded 3 fragments of approximately 650, 1500 and 5400 bp each. Both the TV007cB104 and TV007cB105 clones showed only one fragment of 2600 bp. The TV010cD7 clone showed 2 fragments of approximately 1400 and 6200 bp each. The TV013cB20 and TV013cH17 clones showed 3 fragments of 350, 2500 and 4700 bp each. The restriction of TV018cF1027 yielded 3 fragments of approximately 1300, 1500 and 4700 bp each. It was therefore decided to confirm the orientation of the genes by sequencing. Sequencing results indicated that all the clones showed an insert of reversed orientation, except the TV006cE9 sequence, which displayed the correct insert orientation. As the restriction of all the TV006 clones yielded the exact same restriction pattern, it can be assumed that all three of the clones (TV006cE9, TV006cE5 and TV006cF3) contained the *env* gene insert in the correct orientation.

The T7 promoter present in the pCR3.1 plasmids is constitutive, which would probably result in the translation of the *env* genes into proteins if the genes were cloned in frame. A possible reason for so few clones containing the *env* gene in the correct orientation may be that expressed proteins are toxic to the bacterial cells. If the proteins are indeed toxic, very few clone colonies will survive when grown on the agar plates, and therefore most of the colonies selected for screening would contain the gene in the reversed orientation. If this reaction was also taking place for the pCMVLink clones, it could explain why so few colonies grew on the agar plates, as the genes are forced to insert into the vectors in the correct orientation. This would imply that the CMV promoter could promote translation of

proteins in bacterial cells. It has been suggested that a cryptic *E.coli* promoter is present within the CMV promoter (Invitrogen Corporation (2002), personal communication, Carlsbad, CA, USA), which would therefore allow the translation of proteins in bacterial cells.

All constructs to be used in expression experiments were screened to confirm *env* or *rev* insert presence. The restricted DNA was visualised on an agarose gel and is shown in **Figure 20**. All the constructs showed the excision of an insert from the vector yielding two fragments; one fragment was between 4000 - 4500 bp in size correlating with the size of the pCMVLink and pCMVKm2 vectors (4351 bp), and another fragment displaying the size of the respective inserts cloned into the vectors. A 2500 bp fragment was excised from the TV001c8.2, LinkTV004cC300, LinkTV006cE9, and LinkTV018cF1027 constructs, confirming the presence of the *env* gene. The pCMVLink o-gp140dv2 modified construct showed an excised insert of 2000 bp. The pCMVLink gp120dv2 modified construct showed an excised insert of 1500 bp. A 300 bp insert was excised from the *rev* construct, which confirmed the presence of the *rev* gene.

3.3 EXPRESSION EXPERIMENTS

3.3.1 Culture and transfection of cell lines

Cultured 293 cells grew adherently and media was replaced every 3-4 days as the amount of cells in a visual field increased logarithmically from the day of subculture. The cells were subcultured once a week as a confluence of 80-90% was reached. In preparation for transfection, cells were plated into 6 well plates as described before. The cells reached a confluence of 70-80% the day after plating, and the cells were subsequently transfected.

A preliminary transfection was performed with the pCMVLink o-gp140dv2 modified and pCMVLink gp120dv2 modified constructs only to establish if the transfection method was successful. As this transfection showed expression of Env proteins, the cells were transfected with all the constructs mentioned in the materials and methods section in one experiment.

3.3.2 Detection of proteins

Harvested SNF and lysates were run on SDS-PAGE gels to visualise the expressed proteins. Protein Mr for the Env proteins was calculated by plotting the distance travelled by the proteins on a log graph of Mr versus relative electrophoretic mobility (Mathews and van Holde, 1990) of the Full Range Rainbow Marker (Roche Molecular Biochemicals, Mannheim, Germany). **Figures 21-24** show the developed film for glycoproteins detected in the SNF and prepared lysates, as well as the calculated Mr of the proteins in kDa respectively. **Figures 25-28** show the developed film for native gp160 proteins detected in the SNF and prepared lysates when co-transfected with the *rev* construct, and the calculated Mr of the proteins in kDa respectively.

The native Env proteins had a Mr of approximately 160 kDa in both the SNF and lysate preparations. The TV001 o-gp140 dv2 construct expressed a protein with a Mr of approximately 170 kDa in the SNF, and 145 kDa in the lysate. The TV001 gp120 dv2 construct showed a protein with a Mr of approximately 140 kDa in the SNF, and 105 kDa in the lysate.

Strong expression was detected for the TV001 o-gp140 dv2 and TV001 gp120 dv2 constructs in both SNF and lysates. The TV001c8.2 construct showed expression of a gp160 Env protein

in the presence as well as absence of Rev in both the SNF and lysate preparations. An Env protein was detected in the SNF from the LinkTV004cC300 construct in the presence of Rev only. The lysate showed very low expression of the Env protein for this construct in the absence of Rev, which was elevated in the presence of Rev. No Env protein was detected in the SNF of constructs LinkTV006cE9 and LinkTV018cF1027. The lysates of these constructs showed very low expression of the Env protein, which was slightly increased in the presence of Rev.

3.4 SEQUENCING OF *ENV* CONSTRUCTS

3.4.1 Cycle sequencing reactions

Cycle sequencing reactions were initially performed using 320 ng plasmid DNA, 8 µl of the Terminator ready reaction mix (TTRM) from the ABI PRISM™ Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit, 5 pmol primer and 5 x Sequencing Buffer. These reactions yielded unreadable sequences. The plasmid DNA was increased to 1 µg, which resulted in the readability of a high percentage of the sequences. The use of 4 µl of the TTRM didn't affect the readability of most of the sequences. All the primers did not yield very good sequences for all the genes. An approximate total of 11 primers out of the 19 listed were used to sequence the individual genes, as a few primers worked for some genes, and not for others. Sequencing reactions that yielded unreadable sequences were repeated using 1 µg of plasmid DNA, 8 µl TTRM, 10 pmol primer and 5 x Sequencing Buffer. These reactions increased the readability of some of the sequences.

3.5 PHYLOGENETIC ANALYSIS

3.5.1 Sequence assembly

After assembly of the sequences, the ends of the sequences, which contained parts of the pCR3.1 sequence, were used to determine the orientation of the genes in the pCR3.1 plasmid vectors. All the clones showed an insert of reversed orientation, except the TV006cE9 sequence, which displayed the correct insert orientation. All the pCMVLink constructs (LinkTV004cC300, LinkTV006cE9 and LinkTV018cF1027) showed intact restriction enzyme sites. The *env* gene nucleotide sequences from the pCR3.1 clones were submitted to GenBank (Engelbrecht *et al.*, 2001), and the sequence names and accession numbers are listed in **Table 5**.

Table 5. Summary of sequence names and accession numbers.

Sequence name	Accession number
TV003cE260	AF931233
TV004cC300	AF391234
TV006cE9	AF391237
TV007cB104	AF391238
TV007cB105	AF391239
TV010cD7	AF391242
TV013cB20	AF391245
TV013cH17	AF391246
TV018cF1027	AF391249

3.5.2 Sequence analysis

Nucleotide sequence analysis revealed that all except one clone contained *env* genes with open reading frames. The TV014c6.3

clone had a T2G nucleotide substitution resulting in a M1Q amino acid substitution. The clone also had a premature truncation in the gp41 region, lacking 5 amino acids as a result of a G2558T nucleotide substitution, which created a stop codon. Both clones from isolate TV007 (TV007cB104 and TV007cB105) also had a premature truncation in the gp41 region as a result of a C2410T nucleotide substitution, which created a stop codon. All the clones from TV006 (TV006c9.1, TV0069.2 and TV006cE9), as well as the TV010 clone (TV010cD7) had a deletion of 2 amino acids in the C-terminal of the signal peptide.

The pCMVLink clones created (LinkTV004cC300, LinkTV006cE9 and LinkTV018cF1027) were also sequenced to determine if the sequences varied from the original pCR3.1 clones. The LinkTV004cC300 had no nucleotide changes when compared with the original TV004cC300 clone. The LinkTV006cE9 sequence had a T2411C nucleotide change, which resulted in a W804R amino acid change. LinkTV018cF1027 had major sequence substitutions compared to the original TV018cF1027 clone. There was a C1400T substitution (resulting in a A467V amino acid change), a C1619T substitution (resulting in a P540L amino acid change), and a T1623G substitution, which resulted in a F541L amino acid change. In addition to these amino acid changes, the last 5 amino acids at the C-terminal of the gp41 (EAALL) were replaced with a SSFAI amino acid sequence, followed by an insertion of 54 amino acids before a stop codon was detected.

Noncovalent intermolecular forces determine the three-dimensional conformation of macromolecules (Zubay, 1993), and assessing the hydrophobicity is therefore important to determine if the conformations of proteins expressed by the pCMVLink constructs which had amino acid substitutions, would

be different from the conformations of proteins which would be expressed by the original pCR3.1 clones. As no amino acid changes were observed for the TV004cC300 clone, no effect on conformation or folding of the expressed protein is predicted. The LinkTV006cE9 clone had a W804R amino acid change. The tryptophan in the original sequence was hydrophobic, and the new arginine is hydrophilic. This should not have an effect on conformation of the protein, as the change in hydrophobicity involves only one residue. LinkTV018cF1027 had major sequence substitutions compared to the original TV018cF1027 clone. The three amino acid substitutions within the sequence (A467V, P540L, F541L) should have no effect on conformation, as alanine, valine, proline, leucine, and phenylalanine are all hydrophobic amino acids. In addition to these amino acid changes, the last 5 amino acids at the C-terminal of the gp41 (EAALL) were replaced with a SSFAI amino acid sequence. This change should not have any effect on conformation, as the sequence is still hydrophobic for the last 3 amino acids in this 5 amino acid sequence. The insertion of 54 amino acids has no pattern when looking at hydrophobicity. Hydrophobic amino acids are distributed in between hydrophilic amino acids. This insertion should have major effects on the conformation and folding of the protein.

The predicted amino acid sequence alignment is shown in Appendix B. The B.FR.HXB2 sequence contains a total of 21 cysteine residues in the full-length gp160 (HIV Sequence Compendium, 2001). Eighteen of these residues are found within the gp120, and the residual 3 in the gp41 region. The 21 cysteine residues were also detected in the full-length gp160 region from the TV *env* sequences. A near-perfect conservation of the 18 cysteine residues in the gp120 was observed. The 2 cysteine residues in the immunodominant region of gp41 were highly conserved, and the carboxyl-terminal gp41 cysteine

residue was replaced with an arginine in clone TV013cH17. The number of N-linked glycosylation sites in gp120 ranges from 18-33, with a median of 25 (Korber *et al.*, 2001). The gp120 of B.FR.HXB2 contains 24 possible N-(asparagine) linked glycosylation sites and an additional 6 possible glycosylation sites in the gp41 region (HIV Sequence Compendium, 2001). The positions of these sites may shift or be deleted, but there is always one or more new site(s) within 5-10 residues of the described sites. Twenty-three of the possible glycosylation sites were conserved in the TV *env* sequences. Three of these sites were shifted within 5-10 residues from the described sites. Two additional glycosylation sites were present in the TV *env* sequences. They were situated at positions 442 and adjacent to amino acid 459 (numbering according to B.FR.HXB2 strain). Four out of the possible 6 glycosylation sites were conserved in the gp41 protein sequence. The V1, V2, V4 and V5 loops were highly variable, with the V3 loop being more conserved with a more variable region downstream from the loop. The CD4 binding sites were highly conserved in all the sequences. The cleavage site for gp160 processing (Arg-X-Lys/Arg-Arg) was intact in all the sequences. All the sequences showed highly conserved tryptophan residues in the tryptophan-rich sequence. All the above findings indicate that the *env* genes in the pCR3.1 vectors should be functional (except clone TV014c6.3), although a previous study has reported that sequence analysis alone cannot be used to identify potentially functional *env* clones (Gao *et al.*, 1996). Results from the Gao *et al.* (1996) study indicate that the biological activity of envelope constructs have to be confirmed in the context of infectious virions. The tetrapeptide crown sequence in the V3 region was GPGQ for all except one of the sequences, where a RPGQ crown was observed for clone TV018cF1027. Uncharged or negatively charged amino acid residues were identified at positions 11, 25, and 32 of the V3

loop. The total charge observed for the V3 loop region in the sequences is shown in **Figure 29**.

3.5.3 Phylogenetic relationships

To determine the subtype classification of the TV *env* sequences, a phylogenetic tree was constructed from the full-length gp160 nucleotide sequences from an alignment of the TV *env* sequences and subtype reference strains from the Los Alamos HIV sequence database (HIV Sequence Compendium, 2001). The neighbor-joining tree in **Figure 30** shows major clusters containing sequences representative of the different reference subtypes and CRFs (A1, A2, B, C, D, F1, F2, G, H, J, K, and CRF01_AE, CRF02_AG, CRF03_AB, CRF04_CPX, CRF05_DF, CRF06_CPX, CRF07_BC, CRF08_BC, CRF10_CD, CRF11_CPX, CRF12_BF, CRF13_CPX, CRF14_BG, and group N and O). Bootstrap analysis strongly supported the observed branching orders. The TV *env* sequences appear to form a relatively tight, distinct cluster in the tree. This cluster can be attributed to a bias that was introduced by including a disproportionate number of sequences from southern Africa, as only four other subtype C sequences were used in the alignment. The group O sequences were used to root the tree. More than one sequence was used in rooting the tree to prevent possible bias tree topology as a result of identical sequence artefacts (Sharp, P.M. (2001), personal communication, University of Nottingham, United Kingdom). Using more than one sequence to root a tree also ensures that the node position of the root is more correct when sequences used to root the tree are not very divergent from the rest of the sequences in the tree. It is however common practise by HIV-1 researchers to use only one sequence to root a tree, as the node position of the root is not a critical aspect in determining phylogenetic relationships between sequences (Seoighe, C.

(2002), personal communication, SANBI, University of the Western Cape, RSA). All 21 *env* sequences from the 13 isolates clustered with the subtype C reference strains.

A neighbor-joining tree constructed from an alignment of the TV *env* sequences and 130 full-length subtype C *env* gene sequences is shown in **Figure 31**. The subtype B reference strains from the Los Alamos HIV sequence database (HIV Sequence Compendium, 2001) were used to root the tree. The TV *env* sequences and sequences from Botswana were distributed across several subclusters. All except one of the Indian strains (C.IN.94IN476, subtype followed by country of isolation and then sequence name) were grouped together. The sequences from Brazil were also grouped together.

Nucleotide distances of the TV *env* sequences to subtype C sequences from Botswana, India and previously published South African sequences are shown in **Figure 32**. The statistical values used to construct the box plots are also shown. The TV *env* sequences showed the least divergence from the residual South African strains in the Los Alamos Database (mean value of 11.48% with SD=1.27). The TV *env* sequences were further more similar to the strains from India (divergence mean of 11.81% with SD=1.26), than to strains from Botswana (divergence mean of 12.63% with SD=1.07).

Intersample diversity observed for strains from Botswana, India and previously published South African nucleotide sequences combined with the TV *env* sequences, compared to the variation observed when compared to the subtype C consensus sequence from the Los Alamos HIV sequence database (HIV Sequence Compendium, 2001) is shown in **Figures 33-35** respectively. The Botswana strains showed the highest intersample diversity (mean value of 12.64% with SD=1.27) and

showed the highest variation from the subtype C consensus sequence (mean value of 8.15% with SD=0.77). All the South African sequences in the Los Alamos HIV sequence database showed a mean variation from the subtype C consensus of 8.03% (SD=0.77) and an intersample diversity mean of 11.86% (SD=1.55). The Indian strains were the most similar to the subtype C consensus (mean diversity of 6.99% with SD=1.03) and showed an intersample diversity mean of 8.51% (SD=2.28). The nucleotide sequences of the TV *env* sequences showed an intersample divergence with a mean value of 12.00% (SD=1.66), while the protein sequences showed an intersample divergence with a mean value of 19.66% (SD=1.68).

The nucleotide distance observed between TV *env* sequences and the subtype C consensus, as well as the generated South African consensus is shown in **Figure 36**. Most of the TV *env* sequences were more similar to the generated South African consensus than to the subtype C consensus. The TV002c12.1, TV004cC300, TV012c2.1, TV012c2.2 and TV018cF1027 sequences were more similar to the subtype C consensus. The TV010cD7 sequence showed the closest similarity to both the subtype C consensus and generated South African consensus sequences.

The nucleotide distance observed for the TV *env* sequences and subtype reference strains (HIV Sequence Compendium, 2001) is summarised in **Table 6**. The different nucleotide sequences for *env* clones from isolates TV001, TV006, TV007, TV008, TV013, and TV014 varied between 0-4%. The different protein sequences for *env* clones from isolates TV001, TV006, TV007, TV008 and TV013 varied between 0-4%. The variation between clones from TV014 were not calculated, as the TV014c6.3 clone does not have an open reading frame, as described earlier.

2.5.4 Analysis for intersubtype recombination

The TV *env* sequences were analysed for any recombination. Simplot windows were generated for each sequence and none of the sequences showed intersubtype recombination and showed the highest similarity to the subtype C reference strains from the Los Alamos HIV sequence database (HIV Sequence Compendium, 2001) (data not shown).

Table 6. Distances observed for the TV *env* sequences and subtype reference strains and all subtype C strains. The number of sequences used for analysis is indicated by n.

Subtype	n	Minimum distance (%)	Maximum distance (%)	Mean (%)	SD
A1 and A2	8	15.63	22.77	20.30	1.79
B	4	17.77	23.27	20.87	1.41
C	130	8.26	18.02	12.51	1.93
D	4	17.48	23.87	21.12	1.65
F1 and F2	6	15.08	23.06	19.80	2.17
G	3	16.9	22.6	20.19	1.59
H	3	16.41	22.76	20.14	1.78
J	2	19.76	21.87	20.58	0.56
K	2	19.28	22.34	20.93	0.75
U	3	19.69	23.99	21.81	1.23
GROUP N	2	44.79	48.59	46.33	1.04
GROUP O	4	57.26	65.54	61.35	2.07
CPZ	5	44.72	68.12	50.53	7.56

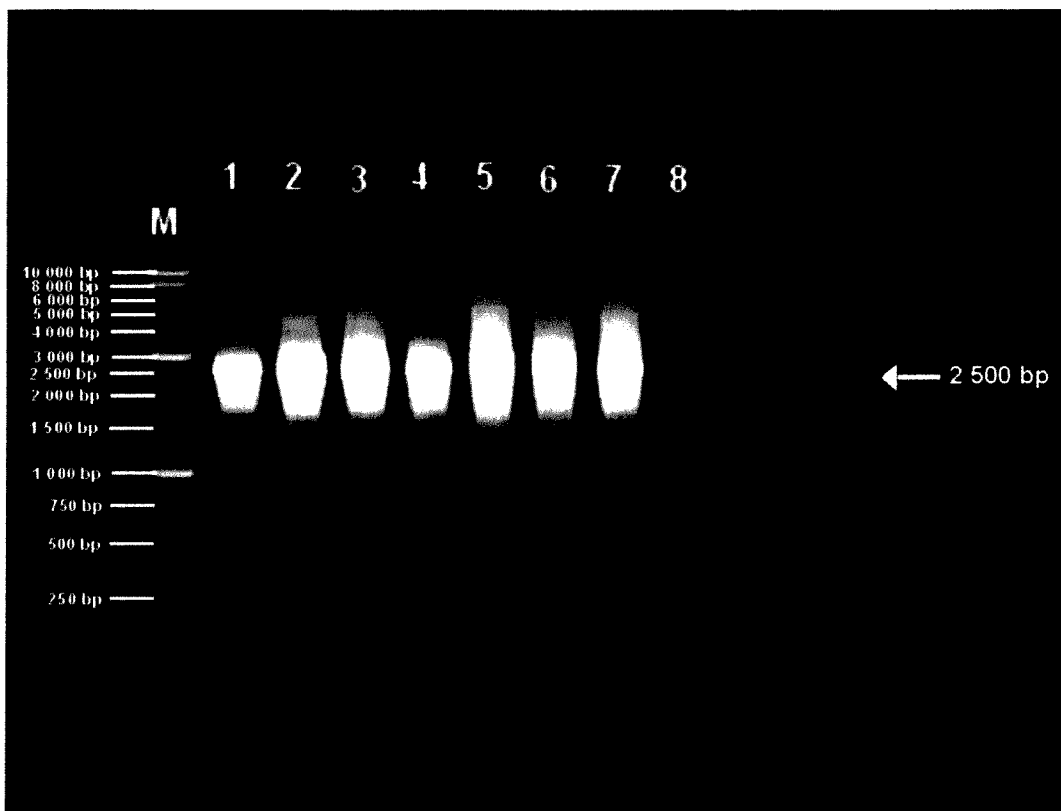
FIGURES

Figure 12. Photograph of an ethidium-bromide stained 0.8% agarose gel showing the PCR products generated. The products were run on the gel with 1 kb DNA Ladder as reference for fragment size (indicated by M). The fragment sizes are indicated to the left of the figure. The products are approximately 2500 bp in size (indicated by white arrow). Lanes 1-7 show the products for isolates TV003, TV004, TV006, TV007, TV010, TV013 and TV018 respectively. Lane 8 contains a negative control (all reagents except template DNA).

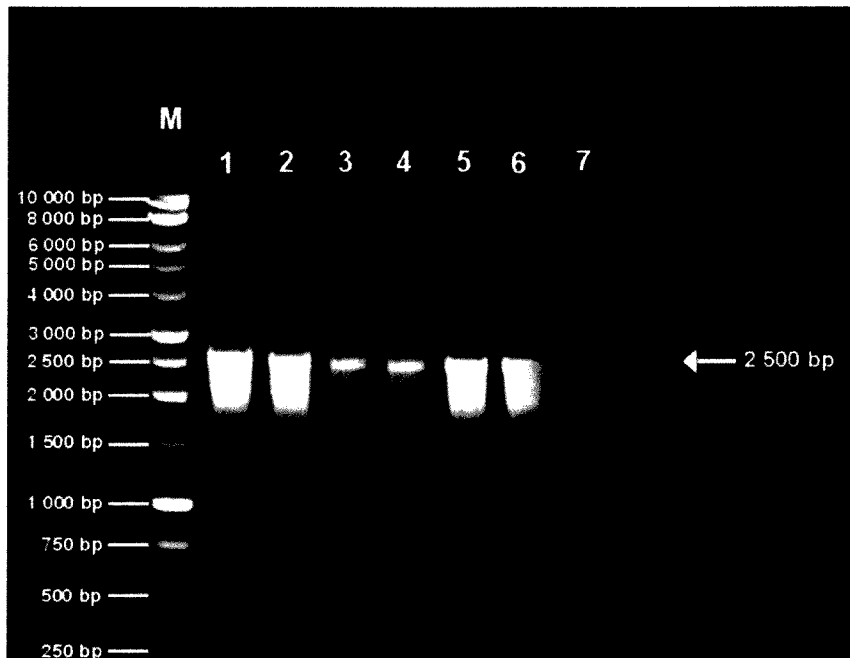


Figure 13. Photograph of an ethidium-bromide stained 0.8% agarose gel showing the PCR products generated. The 1 kb DNA Ladder (indicated by M) fragment sizes are shown to the left of the figure. The 2500 bp *env* fragments are indicated by the white arrow. Lanes 1-2 show the products for TV004, 3-4 show products for TV006, and lanes 5-6 show products for TV018. Lane 7 contains a negative control (all reagents except template DNA).

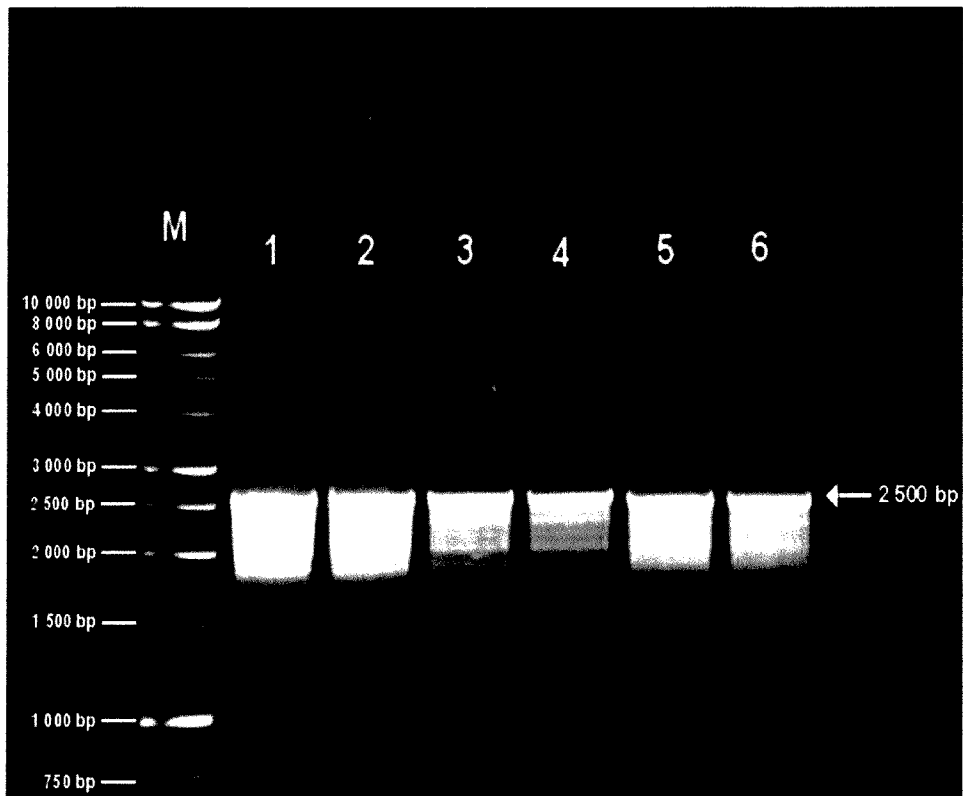


Figure 14. Photograph of a large ethidium-bromide stained 0.8% agarose gel showing the separation of non-specific PCR fragments. The 1 kb DNA Ladder (indicated by M) fragment sizes are shown to the left of the figure. The 2500 bp *env* fragments are indicated by the white arrow. Lanes 1-2 show the products for TV004, 3-4 show products for TV006, and lanes 5-6 show products for TV018.

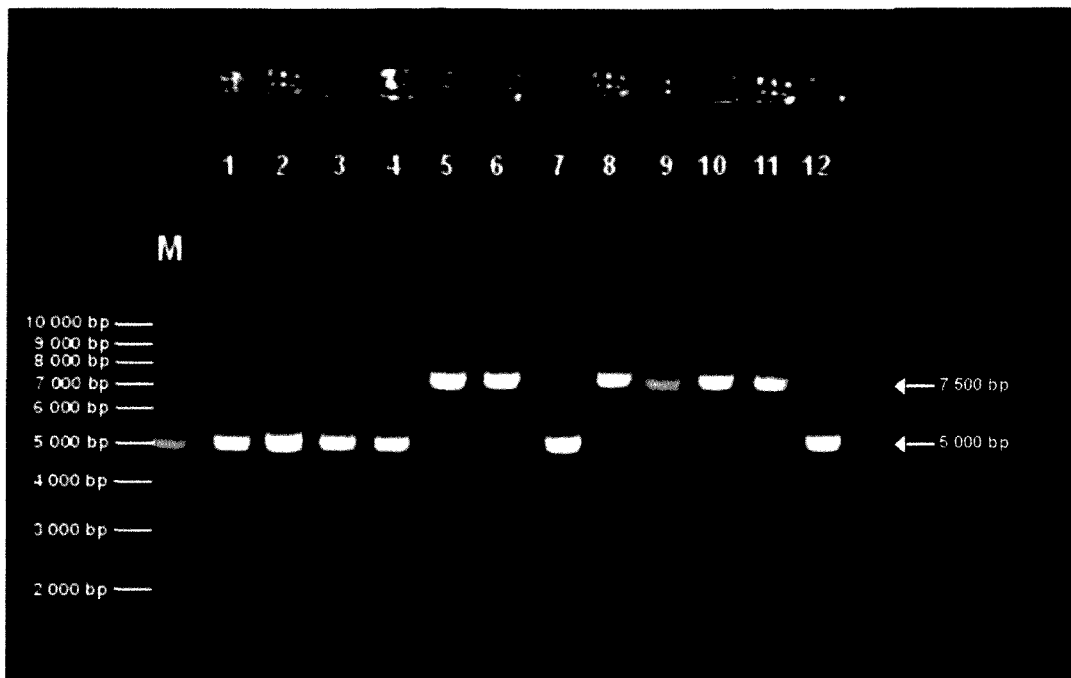


Figure 15. Photograph of an ethidium-bromide stained 0.8% agarose gel showing miniprep DNA from pCR3.1 cloning procedures. A Supercoiled DNA ladder was added as reference for fragment size (indicated by M). The fragment sizes are indicated to the left of the figure. Lanes 1-4, 7 and 12 indicate DNA preps where no *env* insert is present as the fragment sizes are approximately 5000 bp, which correlates with the vector's 5060 bp size (indicated by white arrow). Lanes 5-6, and 8-11 show DNA where an *env* insert might be present (indicated by white arrow), as the fragments are approximately 7500 bp in size (vector of \pm 5000 bp + insert of \pm 2500 bp = 7500 bp).

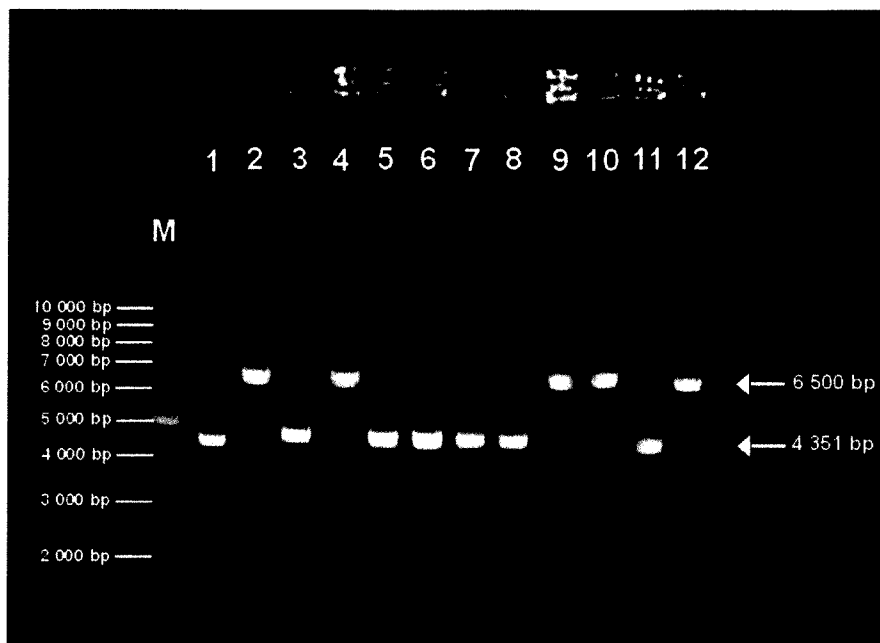


Figure 16. Photograph of an ethidium-bromide stained 0.8% agarose gel showing miniprep DNA from pCMVLink cloning procedures. A Supercoiled DNA ladder was added as reference for fragment size (indicated by M). The fragment sizes are indicated to the left of the figure. Lanes 1, 3, 5-8 and 11 indicate DNA preps where no *env* insert is present as the fragment sizes are between 4000 - 4500 bp, which correlates with the vector's 4351 bp size (indicated by white arrow). Lanes 2, 4, 9, 10, and 12 show DNA where an *env* insert might be present (indicated by white arrow), as the fragments are between 6500 - 7000 bp in size (vector of \pm 4000 to 4500 bp + insert of \pm 2500 bp = 6500 to 7000 bp).

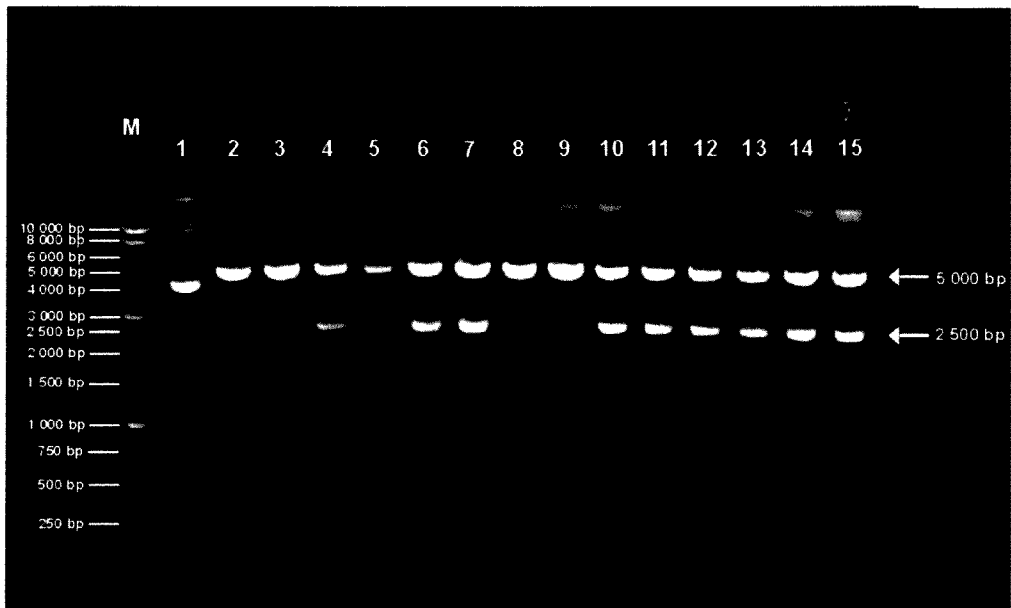


Figure 17. Photograph of an ethidium-bromide stained 0.8% agarose gel showing the *EcoRI* digested miniprep pCR3.1 plasmid DNA. M indicates the 1 kb DNA Ladder and fragment sizes are indicated to the left of the figure. Lane 1 contains undigested miniprep DNA. Lanes 2-3 and 8-9 show miniprep DNA that did not contain the *env* insert and shows a fragment of approximately 5000 bp, which correlates with the 5060 bp size of the vector (indicated by white arrow). Lanes 4-7 and 10-15 show DNA where the *env* insert was excised from vector. The excised fragments are 2500 bp in length (indicated by white arrow), which correlates with the length of the *env* gene.

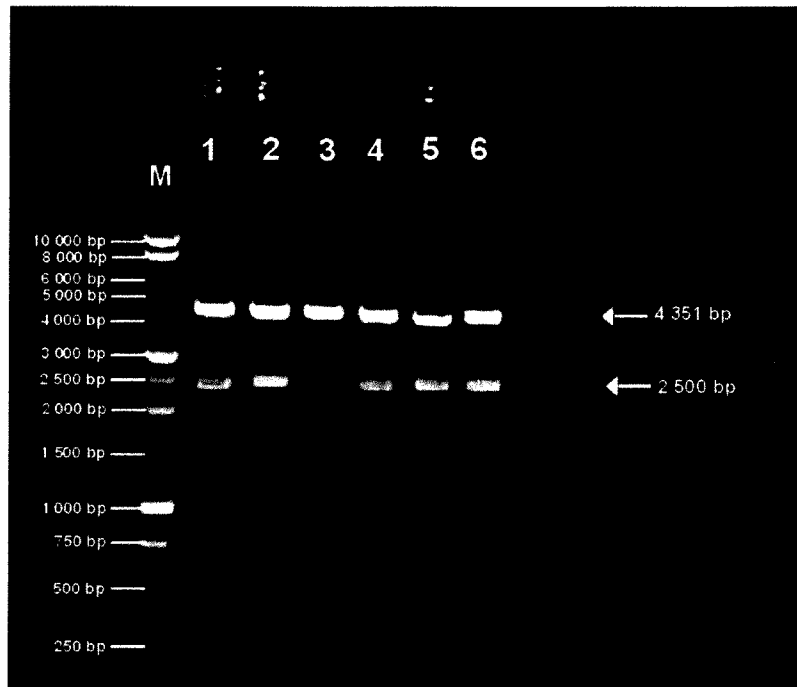


Figure 18. Photograph of an ethidium-bromide stained 0.8% agarose gel showing the *EcoRI* and *XhoI* digested miniprep pCMVLink plasmid DNA. M indicates the 1 kb DNA Ladder. Fragment sizes are indicated to the left of the figure. Lane 3 shows miniprep DNA which did not contain the *env* insert and shows only a fragment of between 4000 - 4500 bp (indicated by white arrow), which correlates with the size of the pCMVLink vector of 4351 bp. Lanes 1-2 and 4-6 show DNA where the *env* insert was excised from vector. The excised fragments are approximately 2500 bp in length (indicated by white arrow), which correlates with the length of the *env* gene.

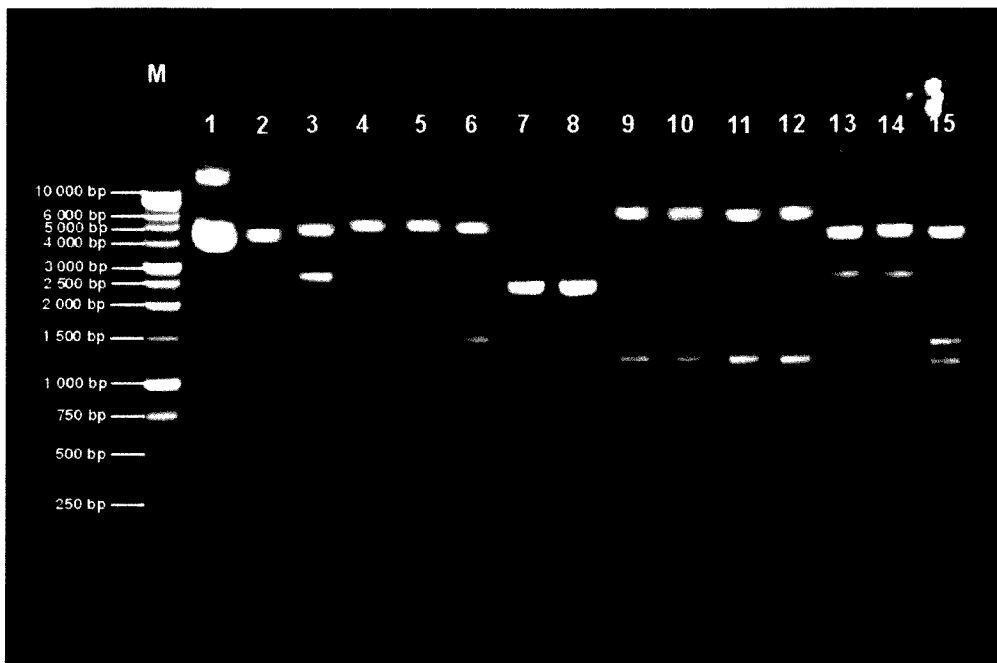


Figure 19. Photograph of an ethidium-bromide stained 0.8% agarose gel showing the *NdeI* digested miniprep pCR3.1 plasmid DNA. M indicates the 1kb DNA Ladder and fragment sizes are indicated to the left of the figure. Lane 1 contains undigested miniprep DNA. Lanes 2-15 shows digestion patterns for clones TV003cE260, TV004cC300, TV006cE9, TV006cE5, TV006cF3, TV007cB104, TV007cB105, TV010cC3, TV010cD7, TV010cD9, TV010cD10, TV013cB20, TV013cH17 and TV018cF1027 respectively.

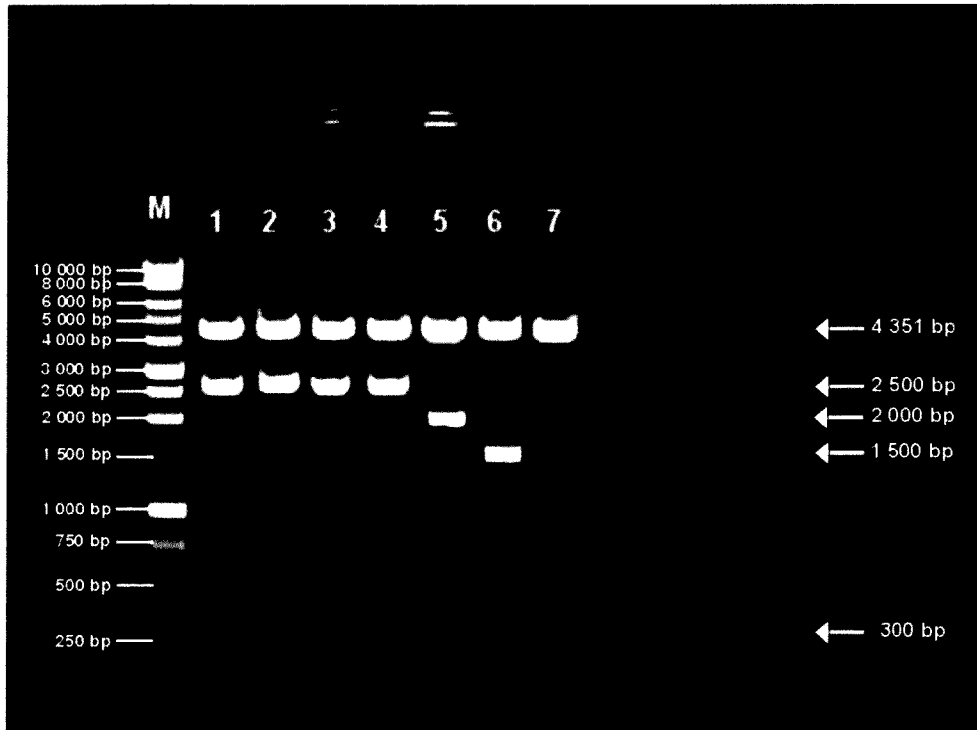


Figure 20. Photograph of an ethidium-bromide stained 0.8% agarose gel showing the digested construct plasmid DNA. M indicates the 1kb DNA Ladder and fragment sizes of the ladder are indicated to the left of the figure. Lanes 1-7 show TV001c8.2, LinkTV004cC300, LinkTV006cE9, LinkTV018cF1027, pCMVLink α -gp140dv2 modified, pCMVLink gp120dv2 modified and the TV010 *rev* construct respectively. The excised genes and plasmid vector sizes are indicated to the right of the figure.

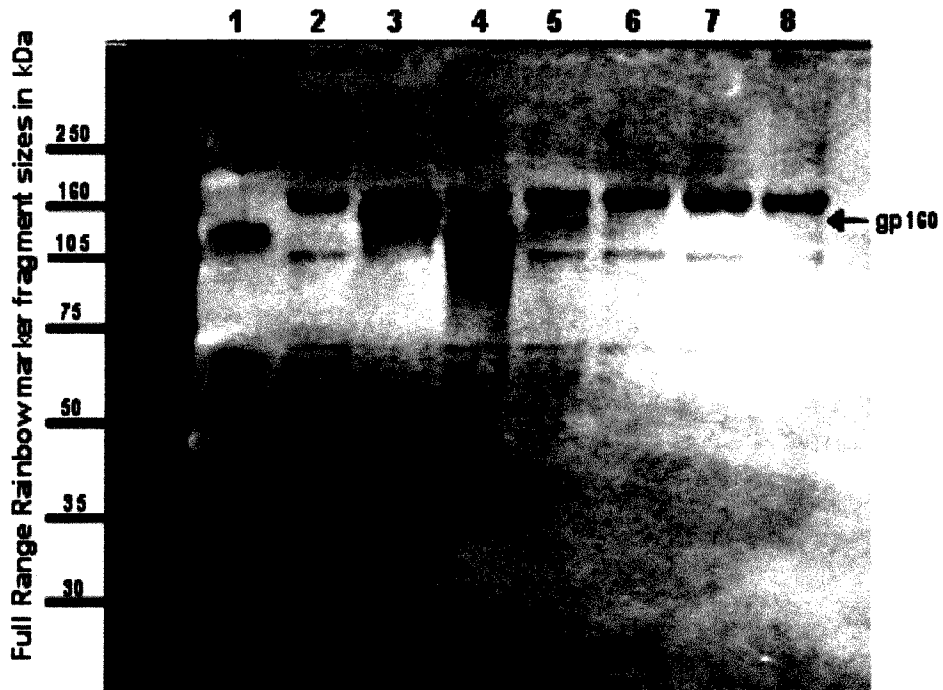
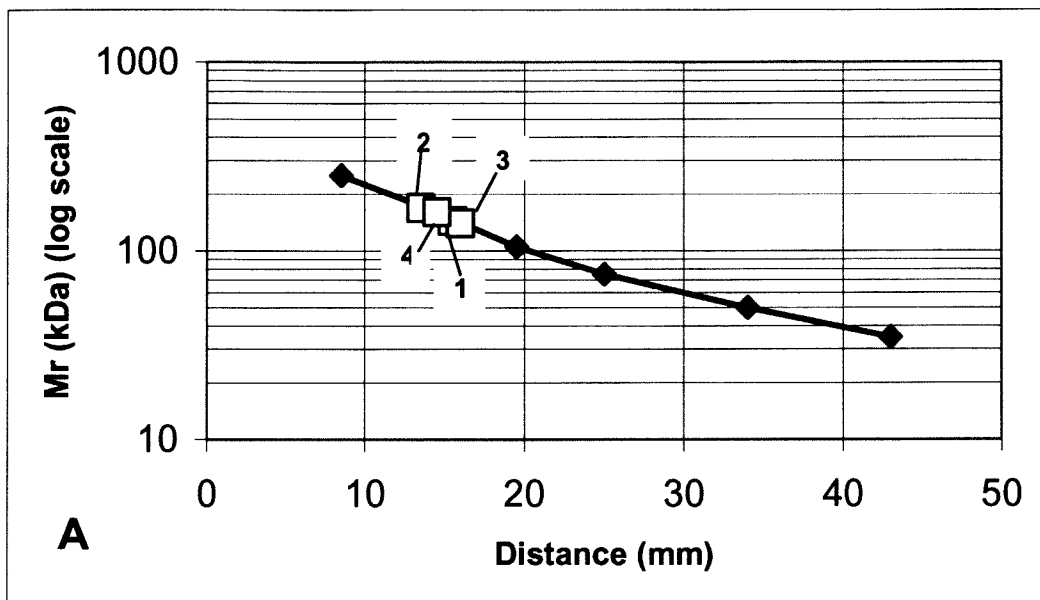


Figure 21. Proteins detected in the SNF of transiently transfected 293 cells. The Full range Rainbow Marker fragment sizes are indicated. Lane 1 contains the HIV-1 SF2 recombinant gp120 positive control. The mock/negative transfected SNF showing cellular glycoproteins appears in Lane 2. Lanes 3-4 contain pCMVLink o-gp140dv2 modified and pCMVLink gp120dv2 modified respectively. Lanes 5-8 show the SNF from 293 cells transfected with the native *env* gene constructs TV001c8.2, LinkTV004cC300, LinkTV006cE9, and LinkTV018cF1027 respectively. The black arrow indicates the distance travelled by the Env protein expressed from the TV001c8.2 construct in lane 5.



B Full range rainbow marker					Calculated
$M_r \times 1000$ (kDa)	Distance (mm)	#	Sample	Distance (mm)	$M_r \times 1000$ (kDa)
250	8.5	1	SF2 gp120 positive control	15.5	145
160	14.5	2	TV001 o-gp140dv2 modified	13.5	170
105	19.5	3	TV001 gp120dv2 modified	16.0	140
75	25.0	4	TV001c8.2 + Rev co-transfection	14.5	160
50	34.0				
35	43.0				

Figure 22. A shows a graph of the log M_r versus the relative distance travelled by the different fragments of the Full Range Rainbow Marker indicated by filled black diamonds. The distances travelled by the detected Env proteins are plotted on the graph as open white squares and are numbered. The numbers correspond with the respective proteins expressed from constructs listed in B. The distances used to calculate the protein M_r are also shown.

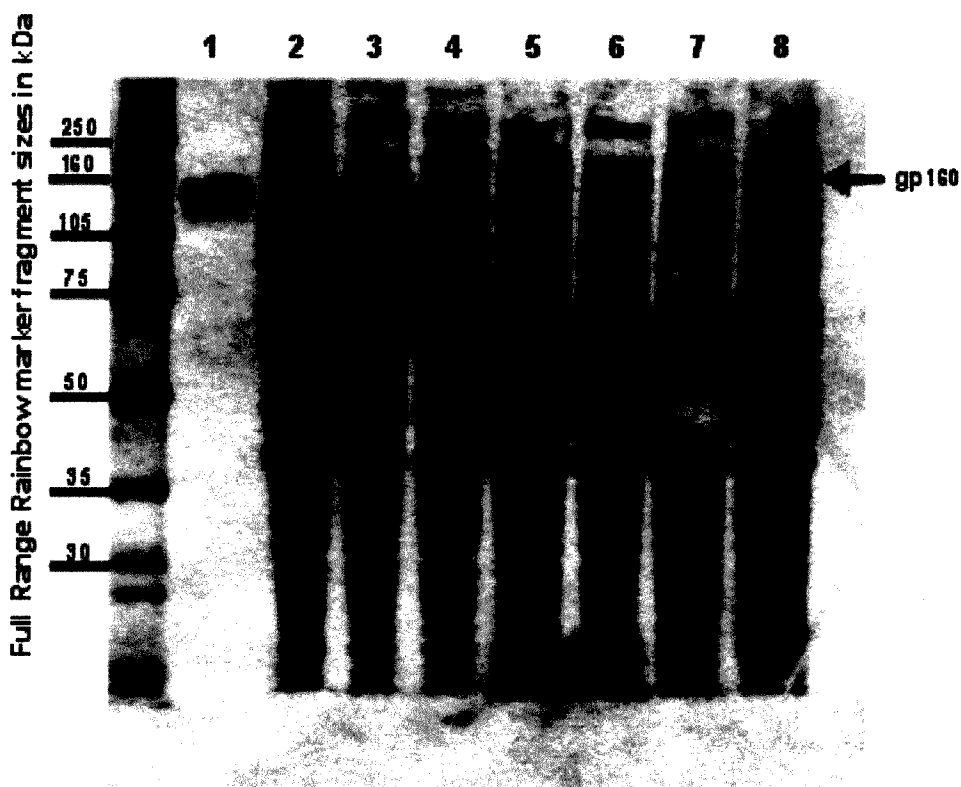
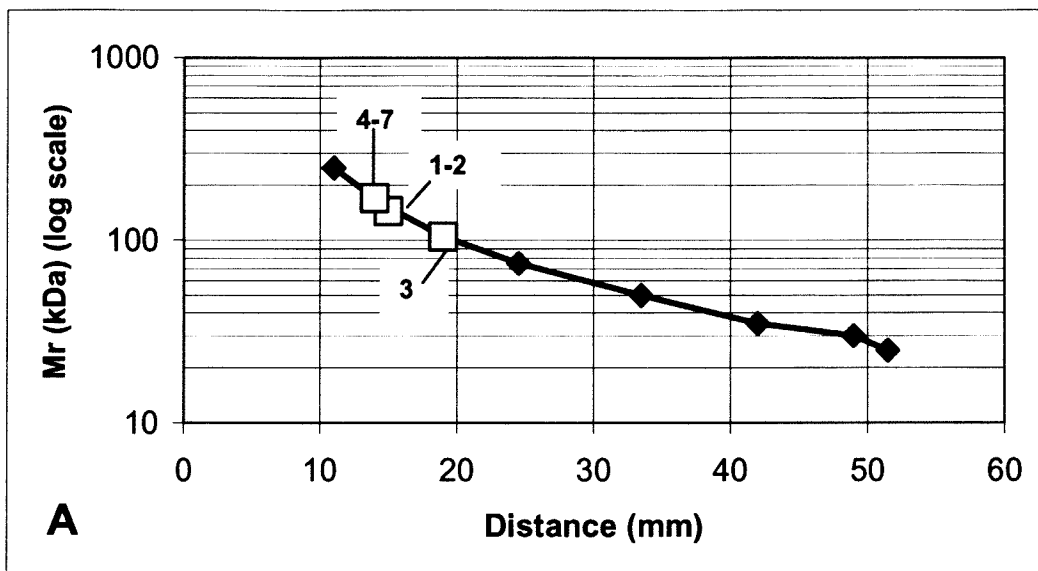


Figure 23. Proteins detected in prepared lysates from transfected 293 cells. The Full range Rainbow Marker fragment sizes are indicated. Lane 1 contains the HIV-1 SF2 recombinant gp120 positive control. The mock/negative transfected cell lysate showing cellular glycoproteins appears in Lane 2. Lanes 3-4 contain pCMVLink o-gp140dv2 modified and pCMVLink gp120dv2 modified respectively. Lanes 5-8 show the lysates from 293 cells transfected with the native *env* gene constructs TV001c8.2, LinkTV004cC300, LinkTV006cE9, and LinkTV018cF1027 respectively. The black arrow indicates the distance travelled on the gel by the gp160 proteins in lanes 5-8.



B Full range rainbow marker					Calculated
$M_r \times 1000$ (kDa)	Distance (mm)	#	Sample	Distance (mm)	$M_r \times 1000$ (kDa)
250	11.0	1	SF2 gp120 positive control	15.0	145
160	14.5	2	TV001 o-gp140dv2 modified	15.0	145
105	19.0	3	TV001 gp120dv2 modified	19.0	105
75	24.5	4	TV001c8.2	14.0	170
50	33.5	5	LinkTV004cC300	13.5	165
35	42.0	6	LinkTV006cE9	13.5	165
30	49.0	7	LinkTV018cF1027	13.5	165
25	51.5				

Figure 24. A shows a graph of the log M_r versus the relative distance travelled by the different fragments of the Full Range Rainbow Marker indicated by filled black diamonds. The distances travelled by the detected Env proteins are plotted on the graph as open white squares and are numbered. The numbers correspond with the respective proteins expressed from constructs listed in B. The distances used to calculate the protein M_r are also shown.

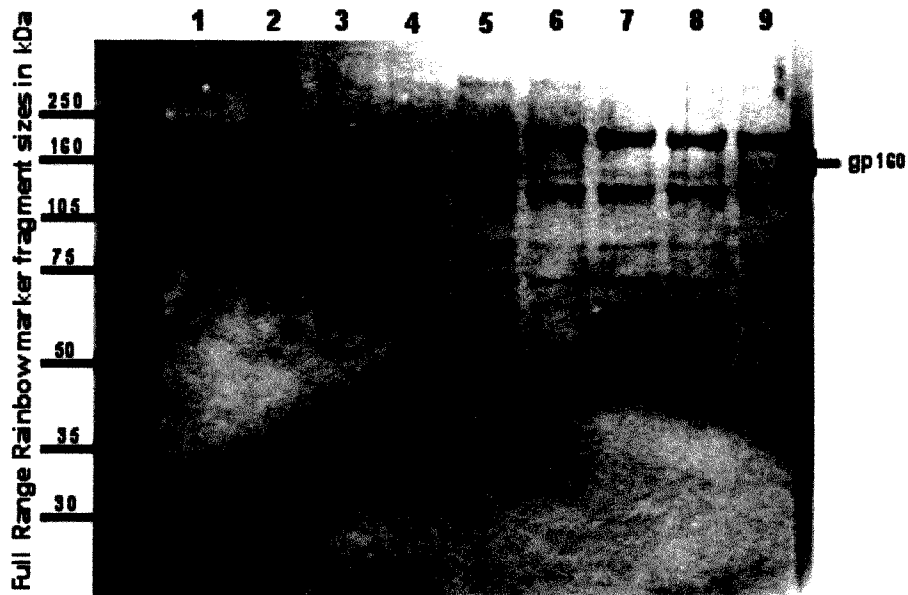
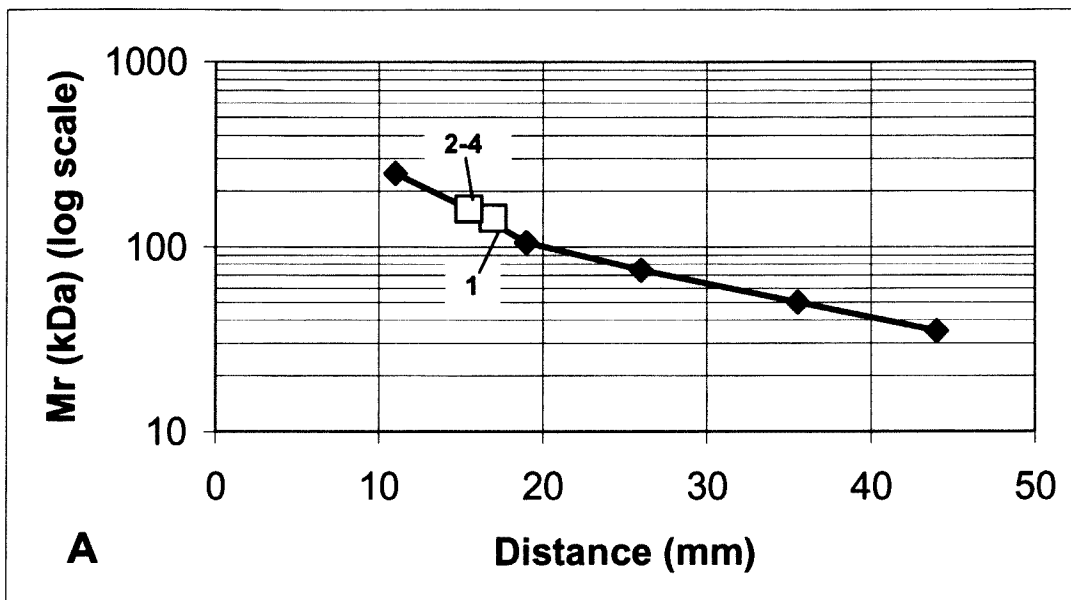


Figure 25. Proteins detected in the SNF of transiently transfected 293 cells when the native *env* constructs were co-transfected with a *rev* construct. The Full range Rainbow Marker fragment sizes are indicated. Lane 1 contains the HIV-1 SF2 recombinant gp120 positive control. Lanes 2, 4, 6 and 8 show the SNF from 293 cells transfected with the native *env* gene constructs TV001c8.2, LinkTV004cC300, LinkTV006cE9 and LinkTV018cF1027 respectively. Lanes 3, 5, 7 and 9 show the SNF from the co-transfection with the *rev* construct for constructs TV001c8.2, LinkTV004cC300, LinkTV006cE9 and LinkTV018cF1027 respectively. A black arrow shows the distance travelled by the detected gp160 proteins in lanes 2, 3 and 5.



B Full range rainbow marker					Calculated
$M_r \times 1000$ (kDa)	Distance (mm)	#	Sample	Distance (mm)	$M_r \times 1000$ (kDa)
250	11.0	1	SF2 gp120 positive control	17.0	143
160	15.5	2	TV001c8.2	15.5	160
105	19.0	3	TV001c8.2 + Rev co-transfection	15.5	160
75	26.0	4	LinkTV004cC300 + Rev co-transfection	15.5	160
50	35.5				
35	44.0				

Figure 26. **A** shows a graph of the log M_r versus the relative distance travelled by the different fragments of the Full Range Rainbow Marker indicated by filled black diamonds. The distances travelled by the detected Env proteins are plotted on the graph as open white squares and are numbered. The numbers correspond with the respective proteins expressed from constructs listed in **B**. The distances used to calculate the protein M_r are also shown.

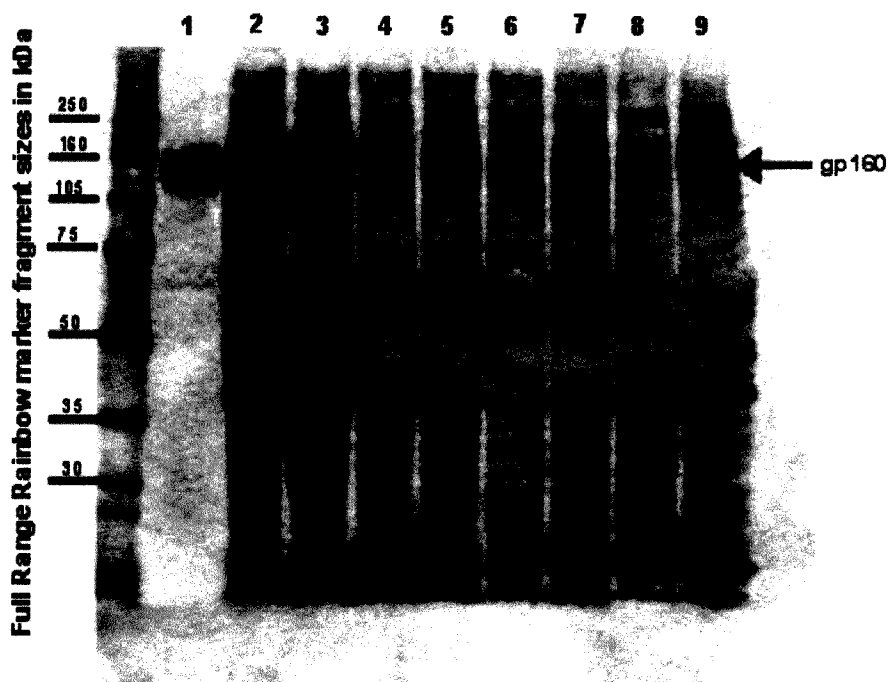
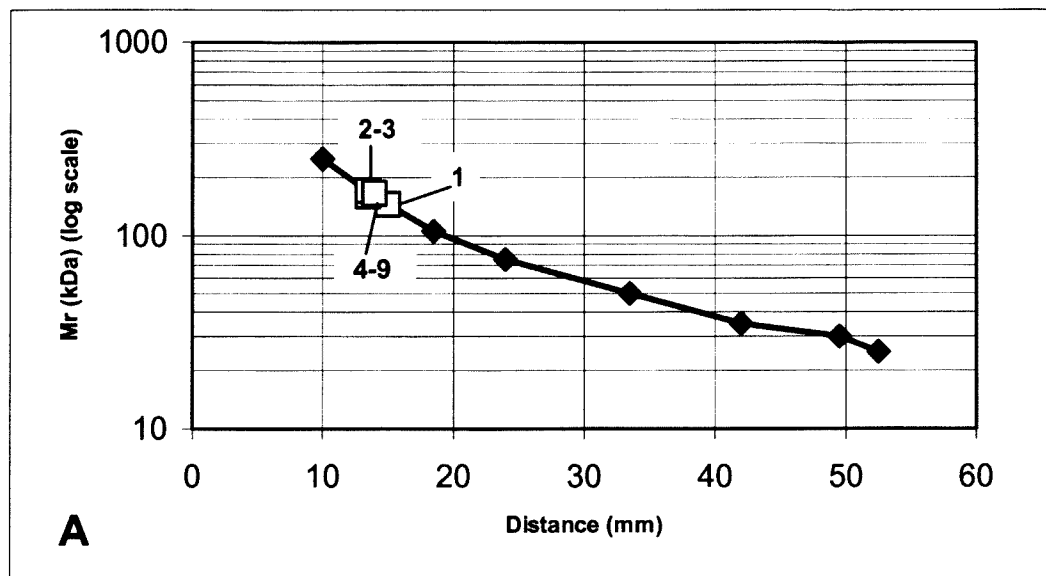


Figure 27. Proteins detected in the prepared cell lysates when the native *env* constructs were co-transfected with a *rev* construct. The Full range Rainbow Marker fragment sizes are indicated. Lane 1 contains the HIV-1 SF2 recombinant gp120 positive control. Lanes 2, 4, 6 and 8 show the lysates from 293 cells transfected with the native *env* gene constructs TV001c8.2, LinkTV004cC300, LinkTV006cE9 and LinkTV018cF1027 respectively. Lanes 3, 5, 7 and 9 show the lysates from the co-transfection with the *rev* construct for constructs TV001c8.2, LinkTV004cC300, LinkTV006cE9 and LinkTV018cF1027 respectively. A black arrow shows the distance travelled by the detected gp160 proteins in lanes 2-9.



B Full range rainbow marker					Calculated
$M_r \times 1000$ (kDa)	Distance (mm)	#	Sample	Distance (mm)	$M_r \times 1000$ (kDa)
250	10.0	1	SF2 gp120 positive control	15.0	145
160	14.0	2	TV001c8.2	13.5	170
105	18.5	3	TV001c8.2 + Rev co-transfection	13.5	170
75	24.0	4	LinkTV004cC300	14.0	165
50	33.5	5	LinkTV004cC300+ Rev co-transfection	14.0	165
35	42.0	6	LinkTV006cE9	14.0	165
30	49.5	7	LinkTV006cE9+ Rev co-transfection	14.0	165
25	52.5	8	LinkTV018cF1027	14.0	165
		9	LinkTV018cF1027+ Rev co-transfection	14.0	165

Figure 28. A shows a graph of the log M_r versus the relative distance travelled by the different fragments of the Full Range Rainbow Marker indicated by filled black diamonds. The distances travelled by the detected Env proteins are plotted on the graph as open white squares and are numbered. The numbers correspond with the respective proteins expressed from constructs listed in B. The distances used to calculate the protein M_r are also shown.

		Charge	Predicted Phenotype
98TV001c8.2	: CTRPNNNTRKSVRIGPGQAFYATNDVIGNIRQAHC	+5	NSI
98TV001c8.5	: CTRPNNNTRKSVRIGPGQAFYATNDVIGNIRQAHC	+5	NSI
98TV003cE260	: CTRPNNNTRKSIRIGPGQTFYATNGIIGNIRQAHC	+6	NSI
98TV007cB104	: CPRPNHNTRRSIRIGPGQAFYATGDIIGDIRQAHC	+5	NSI
98TV007cB105	: CPRPNHNTRRSIRIGPGQAFYATGDIIGDIRQAHC	+5	NSI
98TV013cB20	: CTRPNNNTRRSIRIGPGQAFYT-NDIIGDIRQAHC	+4	NSI
98TV013cH17	: CTRPNNNTRRSIRIGPGQAFYT-NDIIGDIRQAHC	+4	NSI
98TV004cC300	: CTRPNNNTRKSIRIGPGQTFYATGEIIGDIRQAHC	+4	NSI
98TV018cF1027	: CTRPNNNTRRSMRIRPGQTFYATGEIIGDIRQAYC	+4	NSI
98TV002c12.1	: CTRPGNNTRKSVRIGPGQAFYATGDIIGDIRQAHC	+4	NSI
98TV014c6.4	: CTRPGNNTRKSVRIGPGQTFYATGDIIGDIRQAHC	+4	NSI
98TV006c9.1	: CTRPGNNTRQSIRIGPGQTFYATGDIIGDIRQAHC	+3	NSI
98TV006c9.2	: CTRPGNNTRQSIRIGPGQTFYATGDIIGDIRQAHC	+3	NSI
98TV006cE9	: CTRPGNNTRQSIRIGPGQTFYATGDIIGDIRQAHC	+3	NSI
98TV012c2.1	: CTRPNNNTRKSMRIGPGQTFYATGDIIGDIRQAHC	+4	NSI
98TV012c2.2	: CTRPNNNTRKSMRIGPGQTFYATGDIIGDIRQAHC	+4	NSI
98TV010cD7	: CTRPNNNTRKSVRIGPGQTFYATGDIIGDIREAHC	+4	NSI
98TV008c4.3	: CTRPNNNTRKSIRIGPGQTFYATNDIIGDIRQAYC	+3	NSI
98TV008c4.4	: CTRPNNNTRKSIRIGPGQTFYATNDIIGDIRQAYC	+3	NSI
98TV019c5	: CTRPGNNTRKSIRIGPGQAFFATGAIIGDIRKAYC	+5	NSI

Figure 29. The charges calculated for the V3 sequences and the predicted phenotypes are indicated. The net charge was calculated by subtracting the number of negatively charged amino acids (D or E) from the number of positively charged amino acids (K, R or H). Amino acid charges were used as described by Klothe, the Biochemical Compounds Declarative Database at the Institute of Biomedical Computing, Washington University (<http://www.biocheminfo.org/klothe>).



Figure 30. Neighbor-joining phylogenetic tree analysis of full-length *env* sequences (~2500bp) from the reference subtype sequences (Appendix A) and TV *env* sequences. The TV *env* sequences are boxed in grey. An indication of the degree of sequence dissimilarity is shown on the horizontal axis and the subtype sequences on the vertical axis. The percentage of bootstrap trees out of a 100 replications supporting a particular phylogenetic group by more than 75% is placed alongside the node considered. Subtypes A-D, F-H, J-K and Groups O and N have been indicated to the right of the figure.

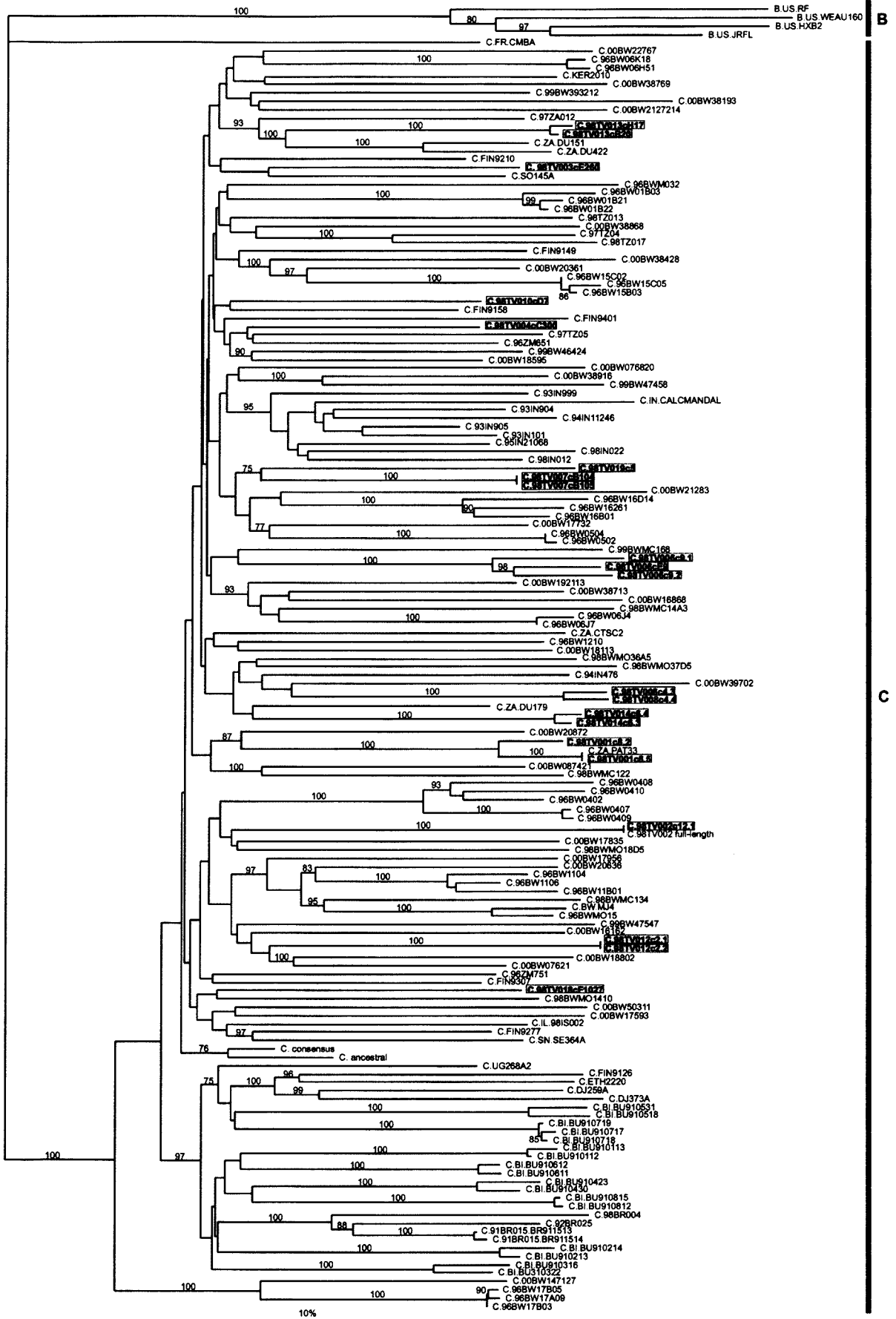
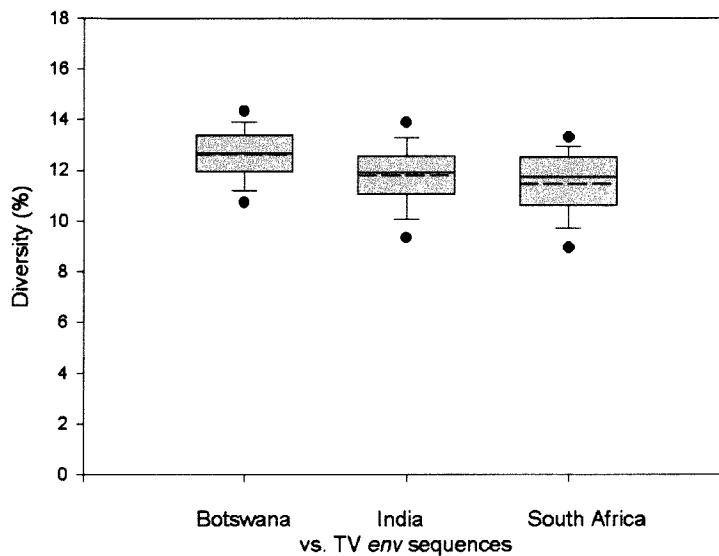
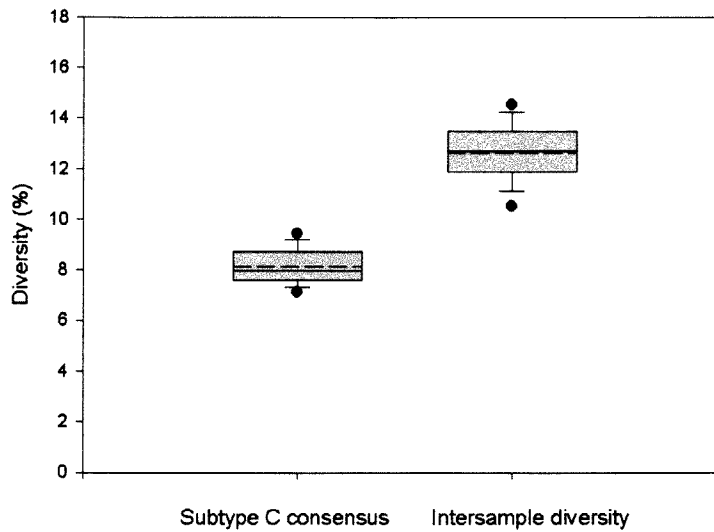


Figure 31. Neighbor-joining phylogenetic tree analysis of 130 full-length subtype C *env* nucleotide sequences (~2500bp) and TV *env* sequences. The TV *env* sequences are boxed in grey. An indication of the degree of sequence dissimilarity is shown on the horizontal axis and the sequence names on the vertical axis. The percentage of bootstrap trees out of a 100 replications supporting a particular phylogenetic group by more than 75% is placed alongside the node considered. Subtypes B and C have been indicated to the right of the figure.

A

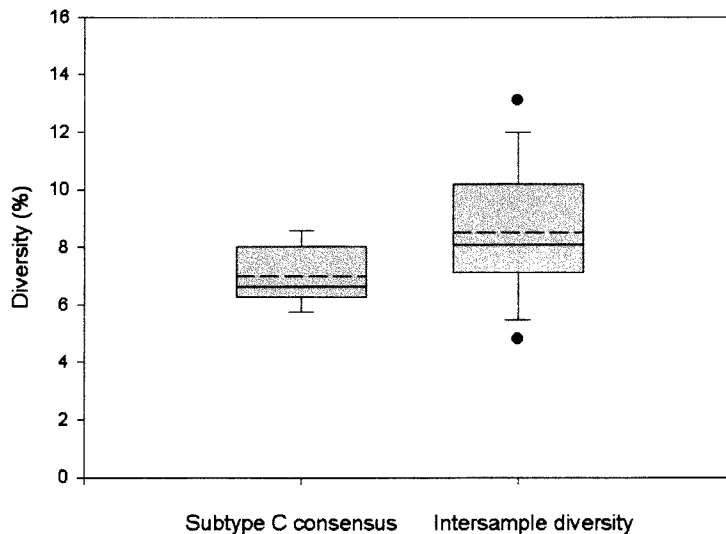
B	TV env sequences versus		
	Botswana	India	South Africa
Mean (%)	12.63	11.81	11.48
SD	1.07	1.26	1.27
SE	0.03	0.09	0.13
95% CI	(0.05700), 12.57-12.69	(0.1803), 11.63-11.99	(0.2524), 11.23-11.74
99% CI	(0.07490), 12.56-12.71	(0.2378), 11.57-12.05	(0.3341), 11.15-11.81
Size, n	1361	191	101
Minimum (%)	8.97	8.45	8.26
Maximum (%)	16.19	14.61	13.56

Figure 32. Nucleotide distance and corresponding statistics are shown in **A** and **B** respectively. Distances recorded between the TV *env* sequences and strains from Botswana, India, and South Africa are plotted as a percentage on the vertical axis. The boundary of the box closest to zero indicates the 25th percentile, a solid line within the box marks the median value, a dashed line within the box shows the arithmetic mean, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the boxes indicate the 10th and 90th percentiles. Points above and below the whiskers indicate the 5th and 95th percentiles. The calculated CI used to determine the range is shown in brackets. Size (n) indicates the number of distances analysed.

A

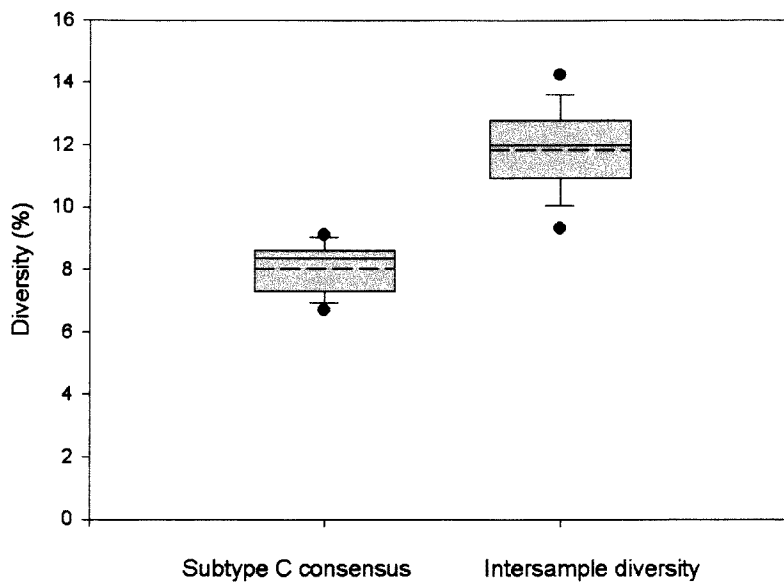
B Statistics	Botswana vs. subtype C consensus	Intersample diversity
Mean (%)	8.15	12.64
SD	0.77	1.27
SE	0.09	0.03
95% CI	(0.1844), 7.97 - 8.34	(0.0651), 12.58 - 12.71
99% CI	(0.2449), 7.91 - 8.40	(0.0855), 12.56 - 12.73
Size, n	71	1477
Minimum (%)	6.02	6.54
Maximum (%)	10.29	16.24

Figure 33. Nucleotide distance and corresponding statistics are shown in **A** and **B** respectively. Distances in percentage between Botswana strains are compared with distances to the subtype C consensus sequence. The boundary of the box closest to zero indicates the 25th percentile, a solid line within the box marks the median value, a dashed line within the box shows the arithmetic mean, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the boxes indicate the 10th and 90th percentiles. Points above and below the whiskers indicate the 5th and 95th percentiles. The calculated CI used to determine the range is shown in brackets. Size (n) indicates the number of distances analysed.

A

B Statistics	India vs. subtype C consensus	Intersample diversity
Mean (%)	6.99	8.51
SD	1.03	2.28
SE	0.32	0.34
95% CI	(0.7340), 6.26 - 7.73	(0.6857), 7.82 - 9.19
99% CI	(1.0545), 5.94 - 8.05	(0.9161), 7.59 - 9.42
Size, n	10	45
Minimum (%)	5.71	4.55
Maximum (%)	8.59	13.61

Figure 34. Nucleotide distance and corresponding statistics are shown in **A** and **B** respectively. Distances in percentage between India strains are compared with distances to the subtype C consensus sequence. The boundary of the box closest to zero indicates the 25th percentile, a solid line within the box marks the median value, a dashed line within the box shows the arithmetic mean, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the boxes indicate the 10th and 90th percentiles when the sample size permitted these calculations. Points above and below the whiskers indicate the 5th and 95th percentiles. The calculated CI used to determine the range is shown in brackets. Size (n) indicates the number of distances analysed.

A

B		
Statistics	South Africa and TV vs. subtype C consensus	Intersample diversity
Mean (%)	8.03	11.86
SD	0.77	1.55
SE	0.15	0.17
95% CI	(0.3183), 7.71 - 8.35	(0.3281), 11.53 - 12.19
99% CI	(0.4314), 7.60 - 8.46	(0.4348), 11.43 - 12.29
Size, n	25	88
Minimum (%)	6.61	4.38
Maximum (%)	9.12	14.67

Figure 35. Nucleotide distance and corresponding statistics are shown in **A** and **B** respectively. Distances in percentage between all South African strains are compared with distances to the subtype C consensus sequence. The boundary of the box closest to zero indicates the 25th percentile, a solid line within the box marks the median value, a dashed line within the box shows the arithmetic mean, and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the boxes indicate the 10th and 90th percentiles. Points above and below the whiskers indicate the 5th and 95th percentiles. The calculated CI used to determine the range is shown in brackets. Size (n) indicates the number of distances analysed.

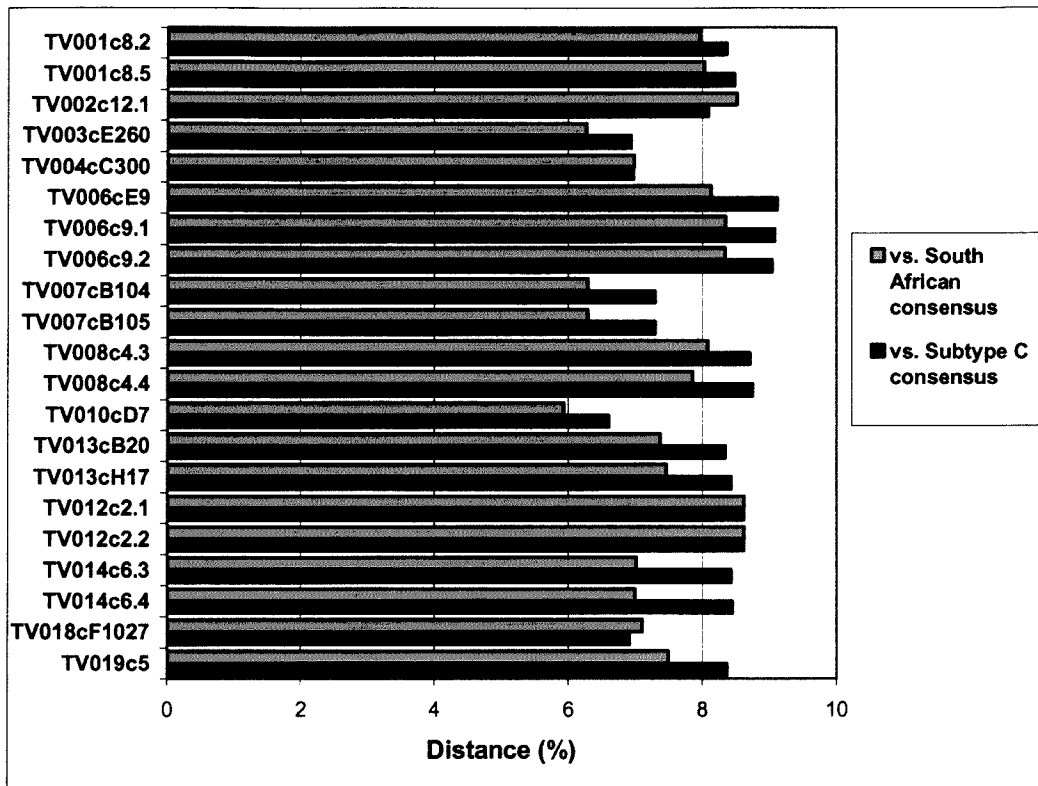


Figure 36. Nucleotide distances between TV *env* nucleotide sequences and the subtype C consensus sequence appearing in the Los Alamos HIV sequence Database (HIV Sequence Compendium, 2001), as well as the consensus generated from all the South African sequences. The distance is shown on the horizontal axis, and the different sequences on the vertical axis. The legend indicates which colour is used to show the calculated distance.

CHAPTER 4 DISCUSSION AND CONCLUSION

	p. No.
INTRODUCTION	106
1.1 SEQUENCE ANALYSIS OF <i>ENV</i> GENES	107
1.2 PHYLOGENETIC RELATIONSHIPS OF <i>ENV</i> SEQUENCES	109
1.3 IMPLICATIONS OF OBSERVED DIVERSITY FOR VACCINE DESIGN	110
1.4 PROTEIN EXPRESSION STUDIES	114
CONCLUSION	117

INTRODUCTION

At the end of 2002, 42 million people were infected with HIV-1, and this infection figure highlights the demand for the development of an effective vaccine to try and stem the pandemic. This is especially critical for a developing nation such as South Africa, where one out of every nine people is infected with HIV (UNAIDS and WHO, 2002).

The *env* gene is an attractive target as immunogen to be included in a HIV vaccine. The envelope elicits neutralising antibodies as it contains conformational neutralising epitopes (Chanh *et al.*, 1986). Effector T-cell responses are also generated by the CTL epitopes present in the Env protein (Takahashi *et al.*, 1988). This protein will therefore be useful in creating a humoral and cellular immune response in the host. A few obstacles need to be overcome if this gene is used as a vaccine. Glycosylation shields the virus from the immune system, and variable loops mask the neutralising antibody binding sites. The genes therefore need to be modified to make them more immunogenic. In order to generate codon-optimised oligomeric genes with their variable loops deleted to expose antibody-binding sites, the native genes first need to be characterised.

Currently, 130 full-length subtype C *env* gene sequences appear in the Los Alamos HIV sequence database (HIV Sequence Compendium, 2001). Excluding the sequences generated by this study and Chiron Corporation (Emeryville, CA, USA), only five subtype C *env* sequences from South Africa are listed (Rodenburg *et al.*, 2001; van Harmelen *et al.*, 2001). At the beginning of this study, only one full-length subtype C envelope gene sequence from South Africa had been published (Rodenburg *et al.*, 2001). A shortage in characterised *env* gene sequences from South Africa was therefore recognised, and this study focussed on the characterisation of generated sequences, as well as the expression of selected genes. These immunogens were created for possible use

in a prime-boost vaccine modality. The *env* genes from recent circulating strains in South Africa were cloned, sequenced and phylogenetically analysed. Selected *env* genes were also transiently expressed to confirm intact genes with open reading frames.

1.1 SEQUENCE ANALYSIS OF *ENV* GENES

It has been indicated that there are subtype-specific variations, which occur within genomic regions of known regulatory function (Gao *et al.*, 1998). Subtype C viruses are characterised by a premature truncation in their *rev* open reading frame, an enlarged Vpu protein, a highly conserved subtype C-specific LRLI motif in the Vpu protein, and three copies of a consensus NF- κ B binding site (Johansson *et al.*, 1995; McCormick-Davis *et al.*, 2000; Salminen *et al.*, 1996; Scriba *et al.*, 2001; Scriba *et al.*, 2002). Subtype B viruses show only two NF- κ B binding sites (HIV Sequence Compendium, 2001), whereas CRF01_AE viruses have only one consensus NF- κ B binding site (Gao *et al.*, 1996b). Reports have suggested that a more rapid transmission of subtype C in relation to the other subtypes can be due to higher levels of promoter activity (possibly induced by the 3rd NF- κ B binding site) resulting in upregulation of viral transcription (Naghavi *et al.*, 1999; Tong-Starksen *et al.*, 1990). Such changes in enhancer copy numbers and regulatory proteins may manifest themselves only after multiple rounds of replication *in vivo*. This could indicate that there are subtype-specific differences in virus biology.

Subtype C specific variations have also been detected in the Env proteins. The V3 loop, which is extremely variable in subtypes other than subtype C, is more conserved in subtype C sequences (Ping *et al.*, 1999). This conservation in the V3 loop was also detected in the TV *env* sequences, which displayed a more variable region downstream from the V3 loop.

There appear to be subtype-dependent differences in usage frequency for certain coreceptors. The CXCR4-positive phenotype is significantly more rare among subtype C isolates (Tscherning *et al.*, 1998), therefore indicating that subtype C viruses do not switch from an NSI phenotype to a SI phenotype as disease progresses. Subtype C isolates obtained from late-stage Ethiopian AIDS patients showed no CXCR4 usage in any of the isolates (Björndal *et al.*, 1999). This was also detected in South African sequences from subtype C isolates (Treurnicht *et al.*, 2002). The tetrapeptide crown sequence identified in most of the subtype C sequences is GPGQ. It has been indicated that this sequence is indicative of a NSI phenotype (Abebe *et al.*, 1999). Almost all of the TV *env* sequences displayed a GPGQ crown sequence. Although the net charge observed for the V3 loop was not $<+5$ for all the TV *env* sequences, uncharged or negatively charged residues were found at positions 11, 25 and 32 of the V3 loop. This also indicates that the sequences should have a NSI phenotype (Björndal *et al.*, 1999). The NSI phenotype was subsequently confirmed in cell culture experiments for all the isolates using the parental virus isolates (Treurnicht *et al.*, 2002). These results strengthen the observation that subtype C viruses isolated previously do not switch from a NSI to SI phenotype, as the isolates used for generating *env* sequences were from patients in all clinical disease stages determined according to the WHO staging system for HIV infection and disease (Lifson *et al.*, 1995). However, our group has indicated that subtype C viruses from more recent isolates have started following the same trend as other subtypes by switching to CXCR4 usage and a SI phenotype (Janse van Rensburg *et al.*, 2002). Another recent study reported the isolation of SI subtype C viruses from patients on antiretroviral therapy (Johnston *et al.*, 2002). This may indicate that the epidemic in South Africa is starting to mature (Janse van Rensburg *et al.*, 2002).

1.2 PHYLOGENETIC RELATIONSHIPS OF ENV SEQUENCES

Since HIV-1 began its expansion in humans roughly 70 years ago (Korber *et al.*, 2000), it has diversified rapidly. Phylogenetic analysis of globally circulating viral strains has identified three distinct groups of HIV-1. These include M (major), O (outlier), and N (non-M and –O) (HIV Sequence Compendium, 2001). Nine different subtypes have been proposed within the group M viruses (subtypes A - D, F – H, J, K) and CRFs have also been identified. The various subtypes differ in their geographic dissemination, and so the subtype designations are a powerful tool for tracking the course of the global epidemic (Hu *et al.*, 1996; WHO Network for HIV Isolation and Characterization, 1994). The epidemic in Thailand for example was believed to have resulted from a single introduction of HIV-1. Genetic analysis however revealed that there were two distinct epidemics of different origins. The intravenous drug users were infected with subtype B prevalent in the Americas, and the sex workers were infected with the CFR01_AE prevalent in Africa (Carr *et al.*, 1996; Gao *et al.*, 1996; Kalish *et al.*, 1995).

Tracking of the HIV-1 subtype C epidemic is critical, as this subtype constitutes 47.2% of all circulating HIV viruses, and is the most commonly transmitted subtype world-wide (Esparza and Bhamarapavati, 2000; Osmanov *et al.*, 2002). The subtype C strains are predominant in India (South Asia), East and Southern Africa, and also circulate in China. Subtype C sequences from the viruses circulating in these countries is essential for phylogenetic analysis and in creating possible immunogens for vaccine purposes. The diversity of the subtype C sequences also needs to be evaluated, as this information should prove to be valuable in the design of vaccine strategies.

Phylogenetic analysis of sequences generated in this study clearly showed that the sequences clustered with the subtype C reference strains with a maximum bootstrap value of 100%. This strongly suggests that the isolates used to generate the *env* genes are subtype C strains and not recombinant forms. Recombination analysis also indicated that the sequences generated in this study showed the closest similarity to subtype C reference strains over the full-length of the genes. This is supported by previous studies in which the accessory, regulatory and *gag* genes as well as the 5' long terminal repeat (LTR) regions of these isolates were characterised as being subtype C (Engelbrecht *et al.*, 2001; Scriba *et al.*, 2001; Scriba *et al.*, 2002). Full-length genome clone sequencing of TV001, TV002 and TV012 recently, further confirmed that the strains belong to subtype C and are not recombinant forms (zur Megede *et al.*, 2002).

It has been found that subtype C sequences from Botswana and South Africa intermingle (Groenink *et al.*, 1993). The sequences of these two countries were also distributed across several subclusters in this study. This highlights the concept that these infections represent a more longstanding epidemic with multiple introductions from different geographic areas (Abebe *et al.*, 2001; Novitsky *et al.*, 1999; van Harmelen *et al.*, 1999). A recent report contrasts with the finding that a starburst effect was seen for South African sequences (Novitsky *et al.*, 2002). The study showed that South African sequences segregated as a separate, distinct subcluster. This may have been due to a bias introduced by including a disproportionate number of Botswana sequences in their analysis, as only 5 South African sequences were used (zur Megede *et al.*, 2002).

1.3 IMPLICATIONS OF THE OBSERVED DIVERSITY FOR VACCINE DESIGN

HIV-1 diversity has important implications for diagnosis, treatment and development of vaccines. The sensitivity of serological and molecular

diagnostic tests may be affected by subtle changes in antigenic structure of HIV-1 variants (Peeters and Sharp, 2000). There are indications of variation in drug susceptibility within group M, as subtype F and G strains are less susceptible to non-nucleoside reverse transcriptase inhibitors and protease inhibitors respectively (Apetrei *et al.*, 1998; Descamps *et al.*, 1998). Diversity has been considered a major problem for the development of a vaccine. Currently, strains belonging to the same subtype can differ up to 20% in their Env proteins, and intersubtype distances can be as high as 35% (Gaschen *et al.*, 2002). The diversity is also continually increasing as a result of mutations generated by reverse transcription, accompanied by a high replication rate and rapid viral turnover (Perelson *et al.*, 1997), pressure for change coupled with tolerance of change (Korber *et al.*, 2001), as well as recombination. The variation is put into perspective when one considers that the influenza vaccine strain has to be replaced when less than 2% amino acid change occurs in the circulating strains, as this can cause a failure in cross-reactivity of the polyclonal response to the influenza vaccine strain (Gaschen *et al.*, 2002).

The development of genetic subtype-specific vaccines are commendable (Esparza and Bhamarapavati, 2000), as a vaccine homologous to the circulating strains in a region is likely to provide better breadth, strength, and durability of immune responses than a mismatched vaccine (Novitsky *et al.*, 2002). For this reason, vaccine candidates currently employ a particular viral strain isolated in a specific geographical region or clones generated from these isolates, with the hope that they will be sufficiently cross-reactive to protect against circulating viruses. However, this may be an optimistic view, as HIV-1 Env proteins can differ in more than 30% of their amino acids. To contend with diversity, country-specific vaccines are being considered, but evolutionary relationships may be more useful than regional considerations (Gaschen *et al.*, 2002). The lack of knowledge of the potential immunological relevance of HIV-1 genetic subtypes should be

addressed as well, as it has been shown that genetic subtypes do not seem to strictly correspond to immunotypes. More than one genetic subtype could share common protective epitopes, and it is possible that more than one immunotype is contained within a single genetic subtype (Esparza and Bhamarapravati, 2000). The type of immune response responsible for protection also needs to be confirmed. In general, neutralising antibodies seem to be more strain specific, whereas cell-mediated immune responses are more cross-reactive (Ferrari *et al.*, 1997).

If attempts are made to create a vaccine from an isolate or strain representative of the circulating viruses in a population or geographic area, certain difficulties will arise in areas such as countries in Africa, where multiple subtypes are circulating concurrently. An alternative would be to design cocktail vaccines containing multiple subtypes. These difficulties will be downscaled when efficacy trials are performed to explore the possibility of achieving cross-protection between different subtypes.

Recent studies indicate that either a consensus sequence or an ancestral sequence reconstructed on the basis of an evolutionary model will have the advantage of being central and most similar to currently circulating strains, and might have the enhanced potential for eliciting cross-reactive responses (Gaschen *et al.*, 2002; Korber *et al.*, 2001; Novitsky *et al.*, 2002). The data generated by Novitsky *et al.* (2002) illustrates that intersample diversity observed for strains from Botswana, India and South Africa is significantly higher than distances observed to the consensus sequences generated. The assumption is made that a consensus sequence will be a more extensive match to the circulating viruses in a specific region. The data generated in this study also shows that intersample diversity is higher compared to the distances seen to the subtype C consensus. As a concept, a consensus sequence approach to vaccine design and development may contain certain limitations. Substitutions of amino acids across HIV-1

proteins are not independent (Bickel *et al.*, 1996; Korber *et al.*, 1993), which indicates that artificial gene constructs might contain certain sequences that are not found in wild-type viruses. Further, the incorporation of multiple copies of numerous viral variants that correspond to minor amino acids in generated and extended consensus sequences would yield very large constructs, which would be unrealistic to generate for vaccine purposes. These limitations can be overcome by preventing the inclusion of residue combinations that rarely occur in virus populations, and construct sizes can be decreased by including only immunodominant regions. However, the concept of consensus-based vaccines still needs to be tested in regards to the correlates of immune protection.

If it is proved to be more efficacious to use strain specific isolates or clones rather than consensus sequences as immunogens, the TV *env* sequences could be used as a base to generate more immunogenic constructs for vaccine purposes. The intersample diversity observed for the amino acid sequences generated in this study (mean value of 19.66%, SD=1.68) is comparable to the variation recorded for the Env protein between all subtype C strains (mean value of 20.02, SD=2.28) (Novitsky *et al.*, 2002). This implies that any of the sequences generated in this study will be representative of circulating subtype C viruses in South Africa. Selecting a sequence with a short branch length relative to the common ancestor in subtype C might be advantageous, as it would tend to be most similar to the majority of sequences represented in the tree (Foley *et al.*, 2000). As the TV010cD7 clone showed the closest similarity to the subtype C consensus sequence and the generated South African consensus, and showed the shortest branch length to the subtype C ancestral sequence, this isolate should be analysed further for functionality and immunogenicity for possible use in vaccine development.

1.4 PROTEIN EXPRESSION STUDIES

The second objective of this study was to express the proteins from the selected isolates in mammalian cells to confirm intact genes and open reading frames. Expression of the *env* genes is dependent on many factors, which involve the synthesis of the proteins, regulation at posttranscriptional level and the cleavage of the gp160 into gp120 and gp41 subunits.

The gp160 precursor is proteolytically cleaved into a gp120 and gp41 protein by a host furin or furin-like protease (Hallenberger *et al.*, 1992) at an Arg-X-Lys/Arg-Arg motif (Hosaka *et al.*, 1991). Expression of envelope proteins in mammalian cell-lines should therefore be characterised by gp120 and gp41 protein fragments when analysed by SDS-PAGE. A group studying the expression of a gp140 protein after transient transfection into 293T cells discovered that the gp140 proteins weren't cleaved into gp120 and gp41 subunits (Binley *et al.*, 2000). They argued that cleavage didn't occur as a result of the large amount of gp140 expressed during the transient-transfection procedure. When exogenous furin was added by co-transfection, the proteins were completely processed into the gp120 and gp41 components. Cleavage of expressed envelope proteins was also reported to be inefficient when recombinant Env is expressed at high levels as full-length gp160 proteins (Binley *et al.*, 2002). The Env proteins expressed from the constructs designed in this study showed a Mr of approximately 160kDa. No gp120 and gp41 subunits were observed, therefore indicating that cleavage was inefficient as a result of cellular proteases being saturated by the recombinant Env proteins, which was expressed at high levels.

The expression of the *env* gene is regulated at posttranscriptional level by the presence of the Rev protein (Dayton *et al.*, 1988), as it contains a RRE.

The Rev-mediated transport to the cytoplasm of unspliced and singly spliced mRNAs possessing the RRE, is dependent on direct binding of Rev (Malim *et al.*, 1989). Most of the Env proteins expressed by the constructs designed in this study showed Rev dependence. It has been shown that expression of Rev dependent *env* genes in recombinant vectors can be increased when the genes are made Rev-independent (Corbet *et al.*, 2000b). This has also been shown to be effective for elevating expression levels of *gag* constructs (zur Megede *et al.*, 2000). Therefore, the expression levels of the constructs in this study can be improved in future studies if the genes are modified to be Rev-independent.

The ER has a 'quality control' system that ensures that proteins do not leave this compartment before they have reached their native state (Doms *et al.*, 1993; Hurtley and Helenius, 1989; Ou *et al.*, 1993). Failure to reach the native state results in the retention of the protein in the ER or in a *cis*-Golgi compartment (Crise *et al.*, 1990), and subsequently in the aggregation or proteolytic degradation in the cytosol (Kimura *et al.*, 1996; Klausner and Sitia, 1990; Lippincott-Schwartz *et al.*, 1988; Willey *et al.*, 1988; Yang *et al.*, 1998). Aggregation is a consequence of the poor solubility characteristics of folding intermediates and incorrectly folded proteins (Doms *et al.*, 1993). Studies have indicated that inefficient cleavage of the signal peptide of HIV-1 is directly responsible for the ER retention of the protein (Li *et al.*, 1996) as a result of aggregation and misfolding (Rottier *et al.*, 1987). A more recent study has also shown that cleavage of the signal peptide is a crucial step in determining the efficiency of subsequent folding and intracellular transport of viral glycoprotein (Li *et al.*, 2000). As the C-terminal of the LinkTV006cE9 construct has a deletion of 2 amino acids and an I25Y amino acid substitution, cleavage might be abrogated, resulting in misfolding and aggregation of the protein. Furthermore, there are several instances in which viral membrane proteins with mutations in the cytoplasmic domain are also retained in the ER (Doms *et al.*, 1988; Doyle *et al.*, 1985; Parks

and Lamb, 1990). The TV018cF1027 construct has a 54 amino acid insertion containing 4 cysteine residues at the C-terminal of gp41, which are further likely to induce misfolding, aberrant interchain disulfide bonding, and aggregation (Segal *et al.*, 1992; Wilcox *et al.*, 1988). Interchain disulfide bonds in the ectodomains of viral proteins are necessary not only for stability of the final protein, but also for folding and maturation (Doms *et al.*, 1993). Very low levels of expression were detected for the LinkTV006cE9 and LinkTV018cF1027 constructs in the lysate preparations, but no expression was seen in the SNF. It is therefore likely that the Env proteins expressed by both the LinkTV006cE9 and LinkTV018cF1027 constructs were retained in the ER or other early compartments of the secretory pathway. If this were the case, these proteins would not be secreted into the SNF and were probably being degraded in the cytosol. The degradation could explain why the levels of expression detected in the lysates were so low.

The Env precursor has a Mr of 160 kDa after asparagine-linked, high-mannose sugars are added to the precursor during synthesis of the protein (Allan *et al.*, 1985; Robey *et al.*, 1985). The precursor is subsequently transported to the *cis* compartment of the Golgi apparatus (Willey *et al.*, 1991), where high mannose ER-acquired amino (N)-linked oligosaccharide side chains are modified to more complex forms (Dewar *et al.*, 1989; Ratner, 1992; Stein and Engleman, 1990). Proteins detected in the lysates of both the TV001 o-gp140 dv2 modified and TV001 gp120 dv2 modified constructs were smaller (approximately 145 and 105kDa respectively) than proteins observed in the SNF (approximately 170 and 140kDa respectively). Asparagine-linked high-mannose sugars were probably still in the process of being added to the proteins during synthesis, or the sugars still had to be modified to more complex forms (Dewar *et al.*, 1989; Ratner, 1992; Stein and Engleman, 1990), which would increase the Mr of the proteins secreted into the SNF.

CONCLUSION

This study has shown that expression of selected *env* genes were Rev dependent and that cleavage into gp120 and gp41 subunits might have been abrogated as a result of cellular furin-like protease saturation. Further, this study has generated much needed subtype C South African *env* gene sequences. The sequences are not recombinant and represent sequences from isolates harbouring a NSI phenotype. Subtype C-specific variations were detected in the Env proteins, which included a more conserved V3 loop. Phylogenetic analysis showed that the sequences intermingled with the Botswana strains and that they were distributed across several subclusters in the phylogenetic tree. This highlights the concept that these infections represent a more longstanding epidemic with multiple introductions from different geographic areas. The sequences displayed a variation similar to the variation observed between all the different subtype C sequences. The *env* gene of the TV010 isolate may be valuable for vaccine design, as it is the most similar to the subtype C consensus compared to the other generated sequences, and shows the shortest branch length to the subtype C ancestral sequence.

CHAPTER 5 REFERENCES

1. Abebe, A., Demissie, D., Goudsmit, J., Brouwer, M., Kuiken, C.L., Pollakis, G., Schuitemaker, H., Fontanet, A.L., and Rinke de Wit, T.F. (1999). HIV-1 subtype C syncytium- and non-syncytium-inducing phenotypes and coreceptor usage among Ethiopian patients with AIDS. *AIDS* 13, 1305-1311.
2. Abebe, A., Lukashov, V.V., Pollakis, G., Kliphuis, A., Fontanet, A.L., Goudsmit, J., and de Wit, T.F. (2001). Timing of the HIV-1 subtype C epidemic in Ethiopia based on early virus strains and subsequent virus diversification. *AIDS* 15, 1555-1561.
3. Alizon, M., Sonigo, P., Barre-Sinoussi, F., Chermann, J.C., Tiollais, P., Montagnier, L., and Wain-Hobson, S. (1984). Molecular cloning of lymphadenopathy-associated virus. *Nature* 312, 757-760.
4. Alkhatib, G., Combadiere, C., Broder, C.C., Feng, Y., Kennedy, P.E., Murphy, P.M., and Berger, E.A. (1996). CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. *Science* 272, 1955-1958.
5. Allan, J.S., Coligan, J.E., Barin, F., McLane, M.F., Sodroski, J.G., Rosen, C.A., Haseltine, W.A., Lee, T.H., and Essex, M. (1985). Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science* 228, 1091-1094.
6. Altman, D.G. (1991). Describing data, p. 19-47. *In* Altman, D.G. ed., *Practical Statistics for medical research*, 1st edition. Published by Chapman & Hall, London, United Kingdom.
7. Apetrei, C., Descamps, D., Collin, G., Loussert-Ajaka, I., Damond, F., Duca, M., Simon, F., and Brun-Vezinet, F. (1998). Human immunodeficiency virus type 1 subtype F reverse transcriptase sequence and drug susceptibility. *J. Virol.* 72, 3534-3538.
8. Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 226, 1209-1211.
9. Barnes, W.M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. U. S. A* 91, 2216-2220.

10. Barnett, S.W., Lu, S., Srivastava, I., Cherpelis, S., Gettie, A., Blanchard, J., Wang, S., Mboudjeka, I., Leung, L., Lian, Y., Fong, A., Buckner, C., Ly, A., Hilt, S., Ulmer, J., Wild, C.T., Mascola, J.R., and Stamatatos, L. (2001). The ability of an oligomeric human immunodeficiency virus type 1 (HIV- 1) envelope antigen to elicit neutralizing antibodies against primary HIV-1 isolates is improved following partial deletion of the second hypervariable region. *J. Virol.* 75, 5526-5540.
11. Barnett, S.W., Rajasekar, S., Legg, H., Doe, B., Fuller, D.H., Haynes, J.R., Walker, C.M., and Steimer, K.S. (1997). Vaccination with HIV-1 gp120 DNA induces immune responses that are boosted by a recombinant gp120 protein subunit. *Vaccine* 15, 869-873.
12. Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220, 868-871.
13. Bebenek, K., Abbotts, J., Roberts, J.D., Wilson, S.H., and Kunkel, T.A. (1989). Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J. Biol. Chem.* 264, 16948-16956.
14. Becker, M.L., Spracklen, F.H., and Becker, W.B. (1985). Isolation of a lymphadenopathy-associated virus from a patient with the acquired immune deficiency syndrome. *S. Afr. Med. J.* 68, 144-147.
15. Bergeron, L., Sullivan, N., and Sodroski, J. (1992). Target cell-specific determinants of membrane fusion within the human immunodeficiency virus type 1 gp120 third variable region and gp41 amino terminus. *J. Virol.* 66, 2389-2397.
16. Bickel, P.J., Cosman, P.C., Olshen, R.A., Spector, P.C., Rodrigo, A.G., and Mullins, J.I. (1996). Covariability of V3 loop amino acids. *AIDS Res. Hum. Retroviruses* 12, 1401-1411.
17. Binley, J.M., Sanders, R.W., Clas, B., Schuelke, N., Master, A., Guo, Y., Kajumo, F., Anselma, D.J., Maddon, P.J., Olson, W.C., and Moore, J.P. (2000). A recombinant human immunodeficiency virus type 1 envelope glycoprotein complex stabilized by an intermolecular disulfide bond between the gp120 and gp41 subunits is an antigenic mimic of the trimeric virion-associated structure. *J. Virol.* 74, 627-643.

18. Binley, J.M., Sanders, R.W., Master, A., Cayan, C.S., Wiley, C.L., Schiffner, L., Travis, B., Kuhmann, S., Burton, D.R., Hu, S.L., Olson, W.C., and Moore, J.P. (2002). Enhancing the proteolytic maturation of human immunodeficiency virus type 1 envelope glycoproteins. *J. Virol.* 76, 2606-2616.
19. Björndal, Å., Deng, H., Jansson, M., Fiore, J.R., Colognesi, C., Karlsson, A., Albert, J., Scarlatti, G., Littman, D.R., and Fenyö, E.M. (1997). Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype. *J. Virol.* 71, 7478-7487.
20. Björndal, Å., Sönnernborg, A., Tscherning, C., Albert, J., and Fenyö, E.M. (1999). Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. *AIDS Res. Hum. Retroviruses* 15, 647-653.
21. Bolmstedt, A., Hinkula, J., Rowcliffe, E., Biller, M., Wahren, B., and Olofsson, S. (2001). Enhanced immunogenicity of a human immunodeficiency virus type 1 env DNA vaccine by manipulating N-glycosylation signals. Effects of elimination of the V3 N306 glycan. *Vaccine* 20, 397-405.
22. Bolognesi, D.P. and Matthews, T.J. (1998). HIV vaccines. Viral envelope fails to deliver? *Nature* 391, 638-639.
23. Boyer, J.C., Bebenek, K., and Kunkel, T.A. (1992). Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proc. Natl. Acad. Sci. U. S. A* 89, 6919-6923.
24. Boyer, J.D., Cohen, A.D., Vogt, S., Schumann, K., Nath, B., Ahn, L., Lacy, K., Bagarazzi, M.L., Higgins, T.J., Baine, Y., Ciccarelli, R.B., Ginsberg, R.S., MacGregor, R.R., and Weiner, D.B. (2000). Vaccination of seronegative volunteers with a human immunodeficiency virus type 1 env/rev DNA vaccine induces antigen-specific proliferation and lymphocyte production of beta-chemokines. *J. Infect. Dis.* 181, 476-483.
25. Bredell, H., Hunt, G., Casteling, A., Cilliers, T., Rademeyer, C., Coetzer, M., Miller, S., Johnson, D., Tiemessen, C.T., Martin, D.J., Williamson, C., and Morris, L. (2002). HIV-1 Subtype A, D, G, AG and unclassified sequences identified in South Africa. *AIDS Res. Hum. Retroviruses* 18, 681-683.

26. Bredell, H., Hunt, G., Morgan, B., Tiemessen, C.T., Martin, D.J., and Morris, L. (2000). Identification of HIV type 1 intersubtype recombinants in South Africa using env and gag heteroduplex mobility assays. *AIDS Res. Hum. Retroviruses* 16, 493-497.
27. Bredell, H., Williamson, C., Sonnenberg, P., Martin, D.J., and Morris, L. (1998). Genetic characterization of HIV type 1 from migrant workers in three South African gold mines. *AIDS Res. Hum. Retroviruses* 14, 677-684.
28. Buchacher, A., Predl, R., Strutzenberger, K., Steinfellner, W., Trkola, A., Purtscher, M., Gruber, G., Tauer, C., Steindl, F., Jungbauer, A., and Katinger, H. (1994). Generation of human monoclonal antibodies against HIV-1 proteins; electrofusion and Epstein-Barr virus transformation for peripheral blood lymphocyte immortalization. *AIDS Res. Hum. Retroviruses* 10, 359-369.
29. Bures, R., Morris, L., Williamson, C., Ramjee, G., Deers, M., Fiscus, S.A., Abdool-Karim, S., and Montefiori, D.C. (2002). Regional clustering of shared neutralization determinants on primary isolates of clade C human immunodeficiency virus type 1 from South Africa. *J. Virol.* 76, 2233-2244.
30. Burton, D.R., Pyati, J., Koduri, R., Sharp, S.J., Thornton, G.B., Parren, P.W., Sawyer, L.S., Hendry, R.M., Dunlop, N., Nara, P.L., Lamacchia, M., Garratty, E., Stiehm, E.R., Bryson, Y.J., Cao, Y., Moore, J.P., Ho, D.D., and Barbas, C.F. (III) (1994). Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* 266, 1024-1027.
31. Cao, J., Bergeron, L., Helseth, E., Thali, M., Repke, H., and Sodroski, J. (1993). Effects of amino acid changes in the extracellular domain of the human immunodeficiency virus type 1 gp41 envelope glycoprotein. *J. Virol.* 67, 2747-2755.
32. Cao, J., Sullivan, N., Desjardin, E., Parolin, C., Robinson, J., Wyatt, R., and Sodroski, J. (1997). Replication and neutralization of human immunodeficiency virus type 1 lacking the V1 and V2 variable loops of the gp120 envelope glycoprotein. *J. Virol.* 71, 9808-9812.
33. Carlos, M.P., Anderson, D.E., Gardner, M.B., and Torres, J.V. (2000). Immunogenicity of a vaccine preparation representing the variable regions of the HIV type 1 envelope glycoprotein. *AIDS Res. Hum. Retroviruses* 16, 153-161.

34. Carr, C.M. and Kim, P.S. (1993). A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* 73, 823-832.
35. Carr, J.K., Salminen, M.O., Koch, C., Gotte, D., Artenstein, A.W., Hegerich, P.A., St Louis, D., Burke, D.S., and McCutchan, F.E. (1996). Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. *J. Virol.* 70, 5935-5943.
36. Centers for Disease Control (1981). Pneumocystis pneumonia-Los Angeles. *Morbidity and Mortality Weekly Report* 30, 250-252.
37. Chakrabarti, B.K., Kong, W.P., Wu, B.Y., Yang, Z.Y., Friberg, J., Ling, X., King, S.R., Montefiori, D.C., and Nabel, G.J. (2002). Modifications of the human immunodeficiency virus envelope glycoprotein enhance immunogenicity for genetic immunization. *J. Virol.* 76, 5357-5368.
38. Chambers, P., Pringle, C.R., and Easton, A.J. (1990). Heptad repeat sequences are located adjacent to hydrophobic regions in several types of virus fusion glycoproteins. *J. Gen. Virol.* 71 (Pt 12), 3075-3080.
39. Chan, D.C., Fass, D., Berger, J.M., and Kim, P.S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 89, 263-273.
40. Chanh, T.C., Dreesman, G.R., Kanda, P., Linette, G.P., Sparrow, J.T., Ho, D.D., and Kennedy, R.C. (1986). Induction of anti-HIV neutralizing antibodies by synthetic peptides. *EMBO J.* 5, 3065-3071.
41. Chapman, B.S., Thayer, R.M., Vincent, K.A., and Haigwood, N.L. (1991). Effect of intron A from human cytomegalovirus (Towne) immediate-early gene on heterologous expression in mammalian cells. *Nucleic Acids Res.* 19, 3979-3986.
42. Chen, Z., Telfier, P., Gettie, A., Reed, P., Zhang, L., Ho, D.D., and Marx, P.A. (1996). Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. *J. Virol.* 70, 3617-3627.

43. Cheng-Mayer, C., Quiroga, M., Tung, J.W., Dina, D., and Levy, J.A. (1990). Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation. *J. Virol.* **64**, 4390-4398.
44. Cherpelis, S., Jin, X., Gettie, A., Ho, D.D., Barnett, S.W., Shrivastava, I., and Stamatatos, L. (2001a). DNA-immunization with a V2 deleted HIV-1 envelope elicits protective antibodies in macaques. *Immunol. Lett.* **79**, 47-55.
45. Cherpelis, S., Shrivastava, I., Gettie, A., Jin, X., Ho, D.D., Barnett, S.W., and Stamatatos, L. (2001b). DNA vaccination with the human immunodeficiency virus type 1 SF162DeltaV2 envelope elicits immune responses that offer partial protection from simian/human immunodeficiency virus infection to CD8(+) T-cell-depleted rhesus macaques. *J. Virol.* **75**, 1547-1550.
46. Chesebro, B., Nishio, J., Perryman, S., Cann, A., O'Brien, W., Chen, I.S., and Wehrly, K. (1991). Identification of human immunodeficiency virus envelope gene sequences influencing viral entry into CD4-positive HeLa cells, T-leukemia cells, and macrophages. *J. Virol.* **65**, 5782-5789.
47. Chesebro, B., Wehrly, K., Nishio, J., and Perryman, S. (1992). Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. *J. Virol.* **66**, 6547-6554.
48. Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P.D., Wu, L., Mackay, C.R., LaRosa, G., Newman, W., Gerard, N., Gerard, C., and Sodroski, J. (1996). The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. *Cell* **85**, 1135-1148.
49. Chow, Y.H., Wei, O.L., Phogat, S., Sidorov, I.A., Fouts, T.R., Broder, C.C., and Dimitrov, D.S. (2002). Conserved structures exposed in HIV-1 envelope glycoproteins stabilized by flexible linkers as potent entry inhibitors and potential immunogens. *Biochemistry* **41**, 7176-7182.
50. Claassen, M. (1999). The growth of HIV-1 C subtype isolates in large volumes at high concentrations for further studies. Honours Thesis, University of Stellenbosch.

51. Clements-Mann, M.L., Weinhold, K., Matthews, T.J., Graham, B.S., Gorse, G.J., Keefer, M.C., McElrath, M.J., Hsieh, R.H., Mestecky, J., Zolla-Pazner, S., Mascola, J., Schwartz, D., Siliciano, R., Corey, L., Wright, P.F., Belshe, R., Dolin, R., Jackson, S., Xu, S., Fast, P., Walker, M.C., Stablein, D., Excler, J.L., Tartaglia, J., Paoletti, E., and the NIAID AIDS Vaccine Evaluation Group (1998). Immune responses to human immunodeficiency virus (HIV) type 1 induced by canarypox expressing HIV-1MN gp120, HIV-1SF2 recombinant gp120, or both vaccines in seronegative adults. *J. Infect. Dis.* *177*, 1230-1246.
52. Cochrane, A.W., Chen, C.H., and Rosen, C.A. (1990). Specific interaction of the human immunodeficiency virus Rev protein with a structured region in the env mRNA. *Proc. Natl. Acad. Sci. U. S. A* *87*, 1198-1202.
53. Coffin, J., Haase, A., Levy, J.A., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P., and Weiss, R. (1986). Human immunodeficiency viruses. *Science* *232*, 697.
54. Collins, J. (1981). Instability of palindromic DNA in *Escherichia coli*. *Cold Spring Harb. Symp. Quant. Biol.* *45 Pt 1*, 409-416.
55. Collins, J., Volckaert, G., and Nevers, P. (1982). Precise and nearly-precise excision of the symmetrical inverted repeats of Tn5; common features of recA-independent deletion events in *Escherichia coli*. *Gene* *19*, 139-146.
56. Connor, R.I., Sheridan, K.E., Ceradini, D., Choe, S., and Landau, N.R. (1997). Change in coreceptor use correlates with disease progression in HIV-1-infected individuals. *J. Exp. Med.* *185*, 621-628.
57. Corbet, S., Muller-Trutwin, M.C., Versmisse, P., Delarue, S., Ayouba, A., Lewis, J., Brunak, S., Martin, P., Brun-Vezinet, F., Simon, F., Barre-Sinoussi, F., and Mauclore, P. (2000a). Env sequences of simian immunodeficiency viruses from chimpanzees in Cameroon are strongly related to those of human immunodeficiency virus group N from the same geographic area. *J. Virol.* *74*, 529-534.
58. Corbet, S., Vinner, L., Hougaard, D.M., Bryder, K., Nielsen, H.V., Nielsen, C., and Fomsgaard, A. (2000b). Construction, biological activity, and immunogenicity of synthetic envelope DNA vaccines based on a primary, CCR5-tropic, early HIV type 1 isolate (BX08) with human codons. *AIDS Res. Hum. Retroviruses* *16*, 1997-2008.

59. Corey, L., McElrath, M.J., Weinhold, K., Matthews, T., Stablein, D., Graham, B., Keefer, M., Schwartz, D., and Gorse, G. (1998). Cytotoxic T cell and neutralizing antibody responses to human immunodeficiency virus type 1 envelope with a combination vaccine regimen. *AIDS Vaccine Evaluation Group. J. Infect. Dis.* 177, 301-309.
60. Crise, B., Buonocore, L., and Rose, J.K. (1990). CD4 is retained in the endoplasmic reticulum by the human immunodeficiency virus type 1 glycoprotein precursor. *J. Virol.* 64, 5585-5593.
61. Dalgleish, A.G., Beverley, P.C., Clapham, P.R., Crawford, D.H., Greaves, M.F., and Weiss, R.A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* 312, 763-767.
62. Dayhoff, M.O., Schwartz, R.M., and Orcutt, B.C. (1979). A model of evolutionary change in proteins, p. 345-352. *In* Dayhoff, M.O. ed., *Atlas of Protein Sequence and Structure*, volume 5, supplement 3. Published by National Biomedical Research Foundation, Silver Spring, Maryland.
63. Dayton, A.I., Terwilliger, E.F., Potz, J., Kowalski, M., Sodroski, J.G., and Haseltine, W.A. (1988). Cis-acting sequences responsive to the rev gene product of the human immunodeficiency virus. *J. Acquir. Immune. Defic. Syndr.* 1, 441-452.
64. De Cock, K.M., Adjorlolo, G., Ekpini, E., Sibailly, T., Kouadio, J., Maran, M., Brattegaard, K., Vetter, K.M., Doorly, R., and Gayle, H.D. (1993). Epidemiology and transmission of HIV-2. Why there is no HIV-2 pandemic. *JAMA* 270, 2083-2086.
65. De Cock, K.M. and Brun-Vezinet, F. (1989). Epidemiology of HIV-2 infection. *AIDS* 3 *Suppl* 1, S89-S95.
66. De Cock, K.M., Brun-Vezinet, F., and Soro, B. (1991). HIV-1 and HIV-2 infections and AIDS in West Africa. *AIDS* 5 *Suppl* 1, S21-S28.
67. Delwart, E.L., Mosialos, G., and Gilmore, T. (1990). Retroviral envelope glycoproteins contain a "leucine zipper"-like repeat. *AIDS Res. Hum. Retroviruses* 6, 703-706.
68. Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R.E., Hill, C.M., Davis, C.B., Peiper, S.C., Schall, T.J., Littman, D.R.,

- and Landau, N.R. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature* **381**, 661-666.
69. Department of Health (2001). Summary Report: National HIV and Syphilis sero-prevalence survey of woman attending antenatal clinics in South Africa (12th in the series). Directorate Health Systems Research, Department of Health, Pretoria, RSA. Available online from <http://www.gov.za/reports/2002/hivsurvey01.pdf>.
70. Descamps, D., Apetrei, C., Collin, G., Damond, F., Simon, F., and Brun-Vezinet, F. (1998). Naturally occurring decreased susceptibility of HIV-1 subtype G to protease inhibitors. *AIDS* **12**, 1109-1111.
71. Dewar, R.L., Vasudevachari, M.B., Natarajan, V., and Salzman, N.P. (1989). Biosynthesis and processing of human immunodeficiency virus type 1 envelope glycoproteins: effects of monensin on glycosylation and transport. *J. Virol.* **63**, 2452-2456.
72. Doms, R.W., Lamb, R.A., Rose, J.K., and Helenius, A. (1993). Folding and assembly of viral membrane proteins. *Virology* **193**, 545-562.
73. Doms, R.W., Ruusala, A., Machamer, C., Helenius, J., Helenius, A., and Rose, J.K. (1988). Differential effects of mutations in three domains on folding, quaternary structure, and intracellular transport of vesicular stomatitis virus G protein. *J. Cell Biol.* **107**, 89-99.
74. Donnelly, J.J., Ulmer, J.B., Shiver, J.W., and Liu, M.A. (1997). DNA vaccines. *Annu. Rev. Immunol.* **15**, 617-648.
75. Doyle, C., Roth, M.G., Sambrook, J., and Gething, M.J. (1985). Mutations in the cytoplasmic domain of the influenza virus hemagglutinin affect different stages of intracellular transport. *J. Cell Biol.* **100**, 704-714.
76. Dragic, T., Litwin, V., Allaway, G.P., Martin, S.R., Huang, Y., Nagashima, K.A., Cayanan, C., Maddon, P.J., Koup, R.A., Moore, J.P., and Paxton, W.A. (1996). HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC- CKR-5. *Nature* **381**, 667-673.
77. Dubay, J.W., Roberts, S.J., Hahn, B.H., and Hunter, E. (1992). Truncation of the human immunodeficiency virus type 1 transmembrane glycoprotein cytoplasmic domain blocks virus infectivity. *J. Virol.* **66**, 6616-6625.

78. Engelbrecht, S., de Villiers, T., Sampson, C.C., zur Megede, J., Barnett, S.W., and Janse van Rensburg, E. (2001). Genetic analysis of the complete gag and env genes of HIV type 1 subtype C primary isolates from South Africa. *AIDS Res. Hum. Retroviruses* 17, 1533-1547.
79. Engelbrecht, S., Smith, T-L., Kasper, P., Faatz, E., Zeier, M., Moodley, D., Clay, C.G., and Janse van Rensburg, E. (1999). HIV type 1 V3 domain serotyping and genotyping in Gauteng, Mpumalanga, KwaZulu-Natal, and Western Cape Provinces of South Africa. *AIDS Res. Hum. Retroviruses* 15, 325-328.
80. Esparza, J. and Bhamarapravati, N. (2000). Accelerating the development and future availability of HIV-1 vaccines: why, when, where, and how? *Lancet* 355, 2061-2066.
81. Esparza, J., Osmanov, S., Pattou-Markovic, C., Toure, C., Chang, M.L., and Nixon, S. (2002). Past, present and future of HIV vaccine trials in developing countries. *Vaccine* 20, 1897-1898.
82. Evans, T.G., McElrath, M.J., Matthews, T., Montefiori, D., Weinhold, K., Wolff, M., Keefer, M.C., Kallas, E.G., Corey, L., Gorse, G.J., Belshe, R., Graham, B.S., Spearman, P.W., Schwartz, D., Mulligan, M.J., Goepfert, P., Fast, P., Berman, P., Powell, M., and Francis, D. (2001). QS-21 promotes an adjuvant effect allowing for reduced antigen dose during HIV-1 envelope subunit immunization in humans. *Vaccine* 19, 2080-2091.
83. Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39, 783-791.
84. Felsenstein, J. (1993). PHYLIP (Phylogeny Inference package) version 3.5c. Distributed by author. Department of Genetics, University of Washington, Seattle, USA. Available online from <http://evolution.genetics.washington.edu/phylip.html>.
85. Felsenstein, J. (1989). PHYLIP (Phylogeny Inference package) version 3.2. *Cladistics* 5, 164-166.
86. Fennie, C. and Lasky, L.A. (1989). Model for intracellular folding of the human immunodeficiency virus type 1 gp120. *J. Virol.* 63, 639-646.

87. Ferrari, G., Humphrey, W., McElrath, M.J., Excler, J.L., Duliege, A.M., Clements, M.L., Corey, L.C., Bolognesi, D.P., and Weinhold, K.J. (1997). Clade B-based HIV-1 vaccines elicit cross-clade cytotoxic T lymphocyte reactivities in uninfected volunteers. *Proc. Natl. Acad. Sci. U. S. A* *94*, 1396-1401.
88. Foley, B., Pan, H., Buchbinder, S., and Delwart, E.L. (2000). Apparent founder effect during the early years of the San Francisco HIV type 1 epidemic (1978-1979). *AIDS Res. Hum. Retroviruses* *16*, 1463-1469.
89. Fouts, T.R., Tuskan, R., Godfrey, K., Reitz, M., Hone, D., Lewis, G.K., and DeVico, A.L. (2000). Expression and characterization of a single-chain polypeptide analogue of the human immunodeficiency virus type 1 gp120-CD4 receptor complex. *J. Virol.* *74*, 11427-11436.
90. Freed, E.O. (1998). HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology* *251*, 1-15.
91. Freed, E.O., Myers, D.J., and Risser, R. (1989). Mutational analysis of the cleavage sequence of the human immunodeficiency virus type 1 envelope glycoprotein precursor gp160. *J. Virol.* *63*, 4670-4675.
92. Freed, E.O., Myers, D.J., and Risser, R. (1990). Characterization of the fusion domain of the human immunodeficiency virus type 1 envelope glycoprotein gp41. *Proc. Natl. Acad. Sci. U. S. A* *87*, 4650-4654.
93. Gabuzda, D.H., Lever, A., Terwilliger, E., and Sodroski, J. (1992). Effects of deletions in the cytoplasmic domain on biological functions of human immunodeficiency virus type 1 envelope glycoproteins. *J. Virol.* *66*, 3306-3315.
94. Gahéry-Ségard, H., Pialoux, G., Charmeteau, B., Sernet, S., Poncelet, H., Raux, M., Tartar, A., L'évy, J-P., Gras-Masse, H., and Guillet, J-G. (2000). Multiepitopic B- and T-cell responses induced in humans by a human immunodeficiency virus type 1 lipopeptide vaccine. *J. Virol.* *74*, 1694-1703.
95. Gallaher, W.R. (1987). Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. *Cell* *50*, 327-328.

96. Gallaher, W.R., Ball, J.M., Garry, R.F., Griffin, M.C., and Montelaro, R.C. (1989). A general model for the transmembrane proteins of HIV and other retroviruses. *AIDS Res. Hum. Retroviruses* 5, 431-440.
97. Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P., and Markham, P.D. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* 224, 500-503.
98. Gallo, R.C., Sarin, P.S., Gelmann, E.P., Robert-Guroff, M., Richardson, E., Kalyanaraman, V.S., Mann, D., Sidhu, G.D., Stahl, R.E., Zolla-Pazner, S., Leibowitch, J., and Popovic, M. (1983). Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science* 220, 865-867.
99. Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M., and Hahn, B.H. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes*. *Nature* 397, 436-441.
100. Gao, F., Morrison, S.G., Robertson, D.L., Thornton, C.L., Craig, S., Karlsson, G., Sodroski, J., Morgado, M., Galvao-Castro, B., and von Briesen, H. (1996a). 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, 1651-1667.
101. Gao, F., Robertson, D.L., Carruthers, C.D., Morrison, S.G., Jian, B., Chen, Y., Barre-Sinoussi, F., Girard, M., Srinivasan, A., Abimiku, A.G., Shaw, G.M., Sharp, P.M., and Hahn, B.H. (1998). A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J. Virol.* 72, 5680-5698.
102. Gao, F., Robertson, D.L., Morrison, S.G., Hui, H., Craig, S., Decker, J., Fultz, P.N., Girard, M., Shaw, G.M., Hahn, B.H., and Sharp, P.M. (1996b). The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J. Virol.* 70, 7013-7029.
103. Gao, F., Yue, L., Craig, S., Thornton, C.L., Robertson, D.L., McCutchan, F.E., Bradac, J.A., Sharp, P.M., and Hahn, B.H. (1994). Genetic variation of HIV type 1 in four World

- Health Organization- sponsored vaccine evaluation sites: generation of functional envelope (glycoprotein 160) clones representative of sequence subtypes A, B, C, and E. WHO Network for HIV Isolation and Characterization. *AIDS Res. Hum. Retroviruses* 10, 1359-1368.
104. Gao, F., Yue, L., White, A.T., Pappas, P.G., Barchue, J., Hanson, A.P., Greene, B.M., Sharp, P.M., Shaw, G.M., and Hahn, B.H. (1992). Human infection by genetically diverse SIV_{SM}-related HIV-2 in west Africa. *Nature* 358, 495-499.
105. Gartner, S. and Popovic, M. (1990). Virus Isolation and Production, p.53-70. *In* Aldovini, A. and Walker, B.D., eds., *Techniques in HIV Researsch*. Published by Stockton Press, New York, NY; and Canada.
106. Gaschen, B., Taylor, J., Yusim, K., Foley, B., Gao, F., Lang, D., Novitsky, V., Haynes, B., Hahn, B.H., Bhattacharya, T., and Korber, B. (2002). Diversity considerations in HIV-1 vaccine selection. *Science* 296, 2354-2360.
107. Gendelman, H.E., Narayan, O., Molineaux, S., Clements, J.E., and Ghotbi, Z. (1985). Slow, persistent replication of lentiviruses: role of tissue macrophages and macrophage precursors in bone marrow. *Proc. Natl. Acad. Sci. U. S. A* 82, 7086-7090.
108. Gherardi, M.M., Ramírez, J.C., and Esteban, M. (2000). Interleukin-12 (IL-12) enhancement of the cellular immune response against human immunodeficiency virus type 1 env antigen in a DNA prime/vaccinia virus boost vaccine regimen is time and dose dependent: suppressive effects of IL-12 boost are mediated by nitric oxide. *J. Virol.* 74, 6278-6286.
109. Gnann, J.W. (Jr.), Nelson, J.A., and Oldstone, M.B. (1987). Fine mapping of an immunodominant domain in the transmembrane glycoprotein of human immunodeficiency virus. *J. Virol.* 61, 2639-2641.
110. Gonda, M.A., Wong-Staal, F., Gallo, R.C., Clements, J.E., Narayan, O., and Gilden, R.V. (1985). Sequence homology and morphologic similarity of HTLV-III and visna virus, a pathogenic lentivirus. *Science* 227, 173-177.
111. Gottlieb, M.S., Schroff, R., Schanker, H.M., Weisman, J.D., Fan, P.T., Wolf, R.A., and Saxon, A. (1981). *Pneumocystis carinii* pneumonia and mucosal candidiasis in previously

healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N. Engl. J. Med.* 305, 1425-1431.

112. Graham, B.S. (2000). Clinical trials of HIV vaccines, p. 120-138. *In* Kuiken, C., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Mullins, J., Wolinsky, S, and Korber, B., eds., *Human Retroviruses and AIDS 2000: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.
113. Groenink, M., Fouchier, R.A., Broersen, S., Baker, C.H., Koot, M., van't Wout, A.B., Huisman, H.G., Miedema, F., Tersmette, M., and Schuitemaker, H. (1993). Relation of phenotype evolution of HIV-1 to envelope V2 configuration. *Science* 260, 1513-1516.
114. Hadzopoulou-Cladaras, M., Felber, B.K., Cladaras, C., Athanassopoulos, A., Tse, A., and Pavlakis, G.N. (1989). The rev (trs/art) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a cis-acting sequence in the env region. *J. Virol.* 63, 1265-1274.
115. Hagan, C.E. and Warren, G.J. (1983). Viability of palindromic DNA is restored by deletions occurring at low but variable frequency in plasmids of *Escherichia coli*. *Gene* 24, 317-326.
116. Hahn, B.H., Shaw, G.M., Arya, S.K., Popovic, M., Gallo, R.C., and Wong-Staal, F. (1984). Molecular cloning and characterization of the HTLV-III virus associated with AIDS. *Nature* 312, 166-169.
117. Hahn, B.H., Shaw, G.M., De Cock, K.M., and Sharp, P.M. (2000). AIDS as a zoonosis: scientific and public health implications. *Science* 287, 607-614.
118. Haigwood, N.L. (1990). Evidence for neutralizing antibodies directed against conformational epitopes of HIV-1 gp120, p. 313-320. *In* *Vaccines 90*. Published by Cold Spring Harbor Laboratory Press, New York, NY.
119. Hallenberger, S., Bosch, V., Angliker, H., Shaw, E., Klenk, H.D., and Garten, W. (1992). Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature* 360, 358-361.

120. Hammarskjold, M.L., Heimer, J., Hammarskjold, B., Sangwan, I., Albert, L., and Rekosh, D. (1989). Regulation of human immunodeficiency virus env expression by the rev gene product. *J. Virol.* 63, 1959-1966.
121. Helseth, E., Olshevsky, U., Furman, C., and Sodroski, J. (1991). Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. *J. Virol.* 65, 2119-2123.
122. Helseth, E., Olshevsky, U., Gabuzda, D., Ardman, B., Haseltine, W., and Sodroski, J. (1990). Changes in the transmembrane region of the human immunodeficiency virus type 1 gp41 envelope glycoprotein affect membrane fusion. *J. Virol.* 64, 6314-6318.
123. Henrard, D.R., Phillips, J.F., Muenz, L.R., Blattner, W.A., Wiesner, D., Eyster, M.E., and Goedert, J.J. (1995). Natural history of HIV-1 cell-free viremia. *JAMA* 274, 554-558.
124. Higgins, D.G. and Sharp, P.M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* 73, 237-244.
125. HIV Sequence Compendium (2001). Kuiken, C., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Mullins, J., Wolinsky, S., and Korber, B., eds. Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.
126. Ho, D.D., Rota, T.R., and Hirsch, M.S. (1986). Infection of monocyte/macrophages by human T lymphotropic virus type III. *J. Clin. Invest* 77, 1712-1715.
127. Hosaka, M., Nagahama, M., Kim, W.S., Watanabe, T., Hatsuzawa, K., Ikemizu, J., Murakami, K., and Nakayama, K. (1991). Arg-X-Lys/Arg-Arg motif as a signal for precursor cleavage catalyzed by furin within the constitutive secretory pathway. *J. Biol. Chem.* 266, 12127-12130.
128. Hoxie, J.A., Alpers, J.D., Rackowski, J.L., Huebner, K., Haggarty, B.S., Cedarbaum, A.J., and Reed, J.C. (1986). Alterations in T4 (CD4) protein and mRNA synthesis in cells infected with HIV. *Science* 234, 1123-1127.
129. Hu, D.J., Dondero, T.J., Rayfield, M.A., George, J.R., Schochetman, G., Jaffe, H.W., Luo, C.C., Kalish, M.L., Weniger, B.G., Pau, C.P., Schable, C.A., and Curran, J.W. (1996). The

emerging genetic diversity of HIV. The importance of global surveillance for diagnostics, research, and prevention. *JAMA* 275, 210-216.

130. Huminer, D., Rosenfeld, J.B., and Pitlik, S.D. (1987). AIDS in the pre-AIDS era. *Rev. Infect. Dis.* 9, 1102-1108.
131. Hurtley, S.M. and Helenius, A. (1989). Protein oligomerization in the endoplasmic reticulum. *Annu. Rev. Cell Biol.* 5, 277-307.
132. Janse van Rensburg, E., Smith, T-L., Zeier, M., Robson, B., Sampson, C., Treurnicht, F., and Engelbrecht, S. (2002). Change in co-receptor usage of current South African HIV-1 subtype C primary isolates. *AIDS* 16, 2479.
133. Jiang, S., Lin, K., and Lu, M. (1998). A conformation-specific monoclonal antibody reacting with fusion-active gp41 from the human immunodeficiency virus type 1 envelope glycoprotein. *J. Virol.* 72, 10213-10217.
134. Johansson, B., Sherefa, K., and Sönnernborg, A. (1995). Multiple enhancer motifs in HIV type 1 strains from Ethiopia. *AIDS Res. Hum. Retroviruses* 11, 761-764.
135. Johnston, E., Mutetwa, S., Zijenah, L., von Lieven, A., Israelski, D., Kittinunvorakoon, C., Kantor, R., and Katzenstein, D. (2002). Presence of CXCR4-tropic viruses among subtype C HIV-Infected individuals. 9th Conference on Retroviruses and Opportunistic Infections, Seattle, WA (Abstract no. 137-M).
136. Johnston, M.I. and Flores, J. (2001). Progress in HIV vaccine development. *Current Opinion in Pharmacology* 1, 504-510.
137. Jones, I.M., Primrose, S.B., and Ehrlich, S.D. (1982). Recombination between short direct repeats in a recA host. *Mol. Gen. Genet.* 188, 486-489.
138. Joshi, A. and Jeang, K.T. (1993). Reduction in growth temperature minimizes instability of large plasmids containing HIV-1 proviral genomes. *Biotechniques* 14, 880, 884-886.
139. Kalish, M.L., Baldwin, A., Raktham, S., Wasi, C., Luo, C.C., Schochetman, G., Mastro, T.D., Young, N., Vanichseni, S., Rübnsamen-Waigmann, H., von Briesen, H., Mullins, J.I., Delwart, E., Herring, B., Esparza, J., Heyward, W.L., and Osmanov, S. (1995). The

evolving molecular epidemiology of HIV-1 envelope subtypes in injecting drug users in Bangkok, Thailand: implications for HIV vaccine trials. *AIDS* 9, 851-857.

140. Kawamura, I., Koga, Y., Oh-Hori, N., Onodera, K., Kimura, G., and Nomoto, K. (1989). Depletion of the surface CD4 molecule by the envelope protein of human immunodeficiency virus expressed in a human CD4⁺ monocytoid cell line. *J. Virol.* 63, 3748-3754.
141. Kearney, B. (1995). Antiviral drugs, p.298-344. *In* Petrow S, ed., *The HIV Drug Book*. Published by Pocket Books, New York, NY.
142. Kessler, J.A., McKenna, P.M., Emini, E.A., Chan, C.P., Patel, M.D., Gupta, S.K., Mark, G.E. (III), Barbas, C.F. (III), Burton, D.R., and Conley, A.J. (1997). Recombinant human monoclonal antibody IgG1b12 neutralizes diverse human immunodeficiency virus type 1 primary isolates. *AIDS Res. Hum. Retroviruses* 13, 575-582.
143. Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111-120.
144. Kimura, T., Nishikawa, M., and Fujisawa, J. (1996). Uncleaved env gp160 of human immunodeficiency virus type 1 is degraded within the Golgi apparatus but not lysosomes in COS-1 cells. *FEBS Lett.* 390, 15-20.
145. Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J.C., and Montagnier, L. (1984). T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* 312, 767-768.
146. Klausner, R.D. and Sitia, R. (1990). Protein degradation in the endoplasmic reticulum. *Cell* 62, 611-614.
147. Koenig, S., Gendelman, H.E., Orenstein, J.M., Dal Canto, M.C., Pezeshkpour, G.H., Yungbluth, M., Janotta, F., Aksamit, A., Martin, M.A., and Fauci, A.S. (1986). Detection of AIDS virus in macrophages in brain tissue from AIDS patients with encephalopathy. *Science* 233, 1089-1093.

148. Koito, A., Harrowe, G., Levy, J.A., and Cheng-Mayer, C. (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, 2253-2259.
149. Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., and Detours, V. (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* 58, 19-42.
150. Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B.H., Wolinsky, S., and Bhattacharya, T. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science* 288, 1789-1796.
151. Korber, B.T., Farber, R.M., Wolpert, D.H., and Lapedes, A.S. (1993). Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. *Proc. Natl. Acad. Sci. U. S. A* 90, 7176-7180.
152. Kostrikis, L.G., Bagdades, E., Cao, Y., Zhang, L., Dimitriou, D., and Ho, D.D. (1995). Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J. Virol.* 69, 6122-6130.
153. Koup, R.A., Safrit, J.T., Cao, Y., Andrews, C.A., McLeod, G., Borkowsky, W., Farthing, C., and Ho, D.D. (1994). Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J. Virol.* 68, 4650-4655.
154. Kowalski, M., Potz, J., Basiripour, L., Dorfman, T., Goh, W.C., Terwilliger, E., Dayton, A., Rosen, C., Haseltine, W., and Sodroski, J. (1987). Functional regions of the envelope glycoprotein of human immunodeficiency virus type 1. *Science* 237, 1351-1355.
155. Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393, 648-659.
156. Lasky, L.A., Nakamura, G., Smith, D.H., Fennie, C., Shimasaki, C., Patzer, E., Berman, P., Gregory, T., and Capon, D.J. (1987). Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. *Cell* 50, 975-985.

157. Lee, W.R., Syu, W.J., Du, B., Matsuda, M., Tan, S., Wolf, A., Essex, M., and Lee, T.H. (1992). Nonrandom distribution of gp120 N-linked glycosylation sites important for infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U. S. A* 89, 2213-2217.
158. Leonard, C.K., Spellman, M.W., Riddle, L., Harris, R.J., Thomas, J.N., and Gregory, T.J. (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, 10373-10382.
159. Levy, J.A., Hoffman, A.D., Kramer, S.M., Landis, J.A., Shimabukuro, J.M., and Oshiro, L.S. (1984). Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* 225, 840-842.
160. Li, W.H., Tanimura, M., and Sharp, P.M. (1988). Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol. Biol. Evol.* 5, 313-330.
161. Li, Y., Bergeron, J.J., Luo, L., Ou, W.J., Thomas, D.Y., and Kang, C.Y. (1996). Effects of inefficient cleavage of the signal sequence of HIV-1 gp 120 on its association with calnexin, folding, and intracellular transport. *Proc. Natl. Acad. Sci. U. S. A* 93, 9606-9611.
162. Li, Y., Luo, L., Thomas, D.Y., and Kang, C.Y. (2000). The HIV-1 Env protein signal sequence retards its cleavage and down-regulates the glycoprotein folding. *Virology* 272, 417-428.
163. Lifson, A.R., Allen, S., Wolf, W., Serufilira, A., Kantarama, G., Lindan, C.P., Hudes, E.S., Nsengumuremyi, F., Taelman, H., and Batungwanayo, J. (1995). Classification of HIV infection and disease in women from Rwanda. Evaluation of the World Health Organization HIV staging system and recommended modifications. *Ann. Intern. Med.* 122, 262-270.
164. Lifson, J.D., Feinberg, M.B., Reyes, G.R., Rabin, L., Banapour, B., Chakrabarti, S., Moss, B., Wong-Staal, F., Steimer, K.S., and Engleman, E.G. (1986a). Induction of CD4-dependent cell fusion by the HTLV-III/LAV envelope glycoprotein. *Nature* 323, 725-728.

165. Lifson, J.D., Reyes, G.R., McGrath, M.S., Stein, B.S., and Engleman, E.G. (1986b). AIDS retrovirus induced cytopathology: giant cell formation and involvement of CD4 antigen. *Science* 232, 1123-1127.
166. Lippincott-Schwartz, J., Bonifacino, J.S., Yuan, L.C., and Klausner, R.D. (1988). Degradation from the endoplasmic reticulum: disposing of newly synthesized proteins. *Cell* 54, 209-220.
167. Loussert-Ajaka, I., Chaix, M.L., Korber, B., Letourneur, F., Gomas, E., Allen, E., Ly, T.D., Brun-Vezinet, F., Simon, F., and Saragosti, S. (1995). Variability of human immunodeficiency virus type 1 group O strains isolated from Cameroonian patients living in France. *J. Virol.* 69, 5640-5649.
168. Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., and Ray, S.C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* 73, 152-160.
169. Luciw, P.A. (1996). Human Immunodeficiency Viruses and their replication, p. 1881-1952. *In* Fields, B.N., Knipe, D.M., and Howley, P.M., eds., *Fields Virology* 3rd edition, volume 2. Published by Lippincott-Raven Publishers, Philadelphia, PA.
170. Luciw, P.A., Potter, S.J., Steimer, K., Dina, D., and Levy, J.A. (1984). Molecular cloning of AIDS-associated retrovirus. *Nature* 312, 760-763.
171. Maddon, P.J., Dalgleish, A.G., McDougal, J.S., Clapham, P.R., Weiss, R.A., and Axel, R. (1986). The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* 47, 333-348.
172. Malim, M.H., Hauber, J., Le, S.Y., Maizel, J.V., and Cullen, B.R. (1989). The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* 338, 254-257.
173. Marck, C. (1988). 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res.* 16, 1829-1836.

174. Mascola, J.R., Snyder, S.W., Weislow, O.S., Belay, S.M., Belshe, R.B., Schwartz, D.H., Clements, M.L., Dolin, R., Graham, B.S., Gorse, G.J., Keefer, M.C., McElrath, M.J., Walker, M.C., Wagner, K.F., McNeil, J.G., McCutchan, F.E., and Burke, D.S. (1996). Immunization with envelope subunit vaccine products elicits neutralizing antibodies against laboratory-adapted but not primary isolates of human immunodeficiency virus type 1. The National Institute of Allergy and Infectious Diseases AIDS Vaccine Evaluation Group. *J. Infect. Dis.* **173**, 340-348.
175. Mathews, C.K. and van Holde, K.E. (1990). Determining molecular weights and the number of subunits in a protein molecule, p.213-215. *In* Mathews, C.K. and van Holde, K.E., eds., *Biochemistry*. Published by The Benjamin/Cummings Publishing Company Inc., Redwood City, CA.
176. McClure, M.O., Marsh, M., and Weiss, R.A. (1988). Human immunodeficiency virus infection of CD4-bearing cells occurs by a pH-independent mechanism. *EMBO J.* **7**, 513-518.
177. McClure, M.O., Sommerfelt, M.A., Marsh, M., and Weiss, R.A. (1990). The pH independence of mammalian retrovirus infection. *J. Gen. Virol.* **71 (Pt 4)**, 767-773.
178. McCormick-Davis, C., Dalton, S.B., Singh, D.K., and Stephens, E.B. (2000). Comparison of Vpu sequences from diverse geographical isolates of HIV type 1 identifies the presence of highly variable domains, additional invariant amino acids, and a signature sequence motif common to subtype C isolates. *AIDS Res. Hum. Retroviruses* **16**, 1089-1095.
179. McCune, J.M., Rabin, L.B., Feinberg, M.B., Lieberman, M., Kosek, J.C., Reyes, G.R., and Weissman, I.L. (1988). Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus. *Cell* **53**, 55-67.
180. Mindel, A. and Tenant-Flowers, M. (2001). ABC of AIDS: Natural history and management of early HIV infection. *BMJ* **322**, 1290-1293.
181. Montagnier, L. (1984). AIDS priority. *Nature* **310**, 446.
182. Montagnier, L., Chermann, J.C., Barre-Sinoussi, F., Chamaret, S., Gruest, J., Nugeyre, M.T., Rey, F., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Saimot, G-A., Rozenbaum, W., Gluckman, J.C., Klarmann, G.J., Vilmer, E., Griscelli, C., Foyer-Gazenge,

- C.I., and Brunet, J.B. (1984). A new human T-lymphotropic retrovirus: Characterization and possible role in Lymphadenopathy and Acquired Immune Deficiency Syndromes, p.363-379. *In* Gallo, R.C., Essex, M.E., and Gross, L, eds., Human T-cell Leukemia/Lymphoma virus. Published by Cold Spring Harbor Laboratory, New York, NY.
183. Moodley, D., Smith, T-L., Janse van Rensburg, E., Moodley, J., and Engelbrecht, S. (1998). HIV type 1 V3 region subtyping in KwaZulu-Natal, a high-seroprevalence South African region. *AIDS Res. Hum. Retroviruses* 14, 1015-1018.
184. Muesing, M.A., Smith, D.H., Cabradilla, C.D., Benton, C.V., Lasky, L.A., and Capon, D.J. (1985). Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature* 313, 450-458.
185. Murakami, T. and Freed, E.O. (2000). The long cytoplasmic tail of gp41 is required in a cell type-dependent manner for HIV-1 envelope glycoprotein incorporation into virions. *Proc. Natl. Acad. Sci. U. S. A* 97, 343-348.
186. Musey, L., Hughes, J., Schacker, T., Shea, T., Corey, L., and McElrath, M.J. (1997). Cytotoxic-T-cell responses, viral load, and disease progression in early human immunodeficiency virus type 1 infection. *N. Engl. J. Med.* 337, 1267-1274.
187. Muster, T., Steindl, F., Purtscher, M., Trkola, A., Klima, A., Himmler, G., Ruker, F., and Katinger, H. (1993). A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *J. Virol.* 67, 6642-6647.
188. Naghavi, M.H., Salminen, M.O., Sönnnerborg, A., and Vahlne, A. (1999). DNA sequence of the long terminal repeat of human immunodeficiency virus type 1 subtype A through G. *AIDS Res. Hum. Retroviruses* 15, 485-488.
189. Nicholas, K.B., Nicholas, H.B. (Jr.), and Deerfield, D.W. II. (1997). Genedoc: Analysis and visualization of genetic variation, available online at <http://www.cris.com/~Ketchup/genedoc.shtml>.
190. Nicholas, K.B. and Nicholas, H.B. (Jr.) (1997). Genedoc: Analysis and visualization of genetic variation. *EMBNEW.NEWS* 4, 14

191. Nicholson, J.K., Cross, G.D., Callaway, C.S., and McDougal, J.S. (1986). In vitro infection of human monocytes with human T lymphotropic virus type III/lymphadenopathy-associated virus (HTLV-III/LAV). *J. Immunol.* 137, 323-329.
192. Novitsky, V., Smith, U.R., Gilbert, P., McLane, M.F., Chigwedere, P., Williamson, C., Ndung'u, T., Klein, I., Chang, S.Y., Peter, T., Thior, I., Foley, B.T., Gaolekwe, S., Rybak, N., Gaseitsiwe, S., Vannberg, F., Marlink, R., Lee, T.H., and Essex, M. (2002). Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? *J. Virol.* 76, 5435-5451.
193. Novitsky, V.A., Montano, M.A., McLane, M.F., Renjifo, B., Vannberg, F., Foley, B.T., Ndung'u, T.P., Rahman, M., Makhema, M.J., Marlink, R., and Essex, M. (1999). Molecular cloning and phylogenetic analysis of human immunodeficiency virus type 1 subtype C: a set of 23 full-length clones from Botswana. *J. Virol.* 73, 4427-4432.
194. Osmanov, S., Pattou, C., Walker, N., Schwardlander, B., and Esparza, J. (2002). Estimated global distribution and regional spread of HIV-1 genetic subtypes in the year 2000. *J. Acquir. Immune. Defic. Syndr.* 29, 184-190.
195. Ou, W.J., Cameron, P.H., Thomas, D.Y., and Bergeron, J.J. (1993). Association of folding intermediates of glycoproteins with calnexin during protein maturation. *Nature* 364, 771-776.
196. Ozel, M., Pauli, G., and Gelderblom, H.R. (1988). The organization of the envelope projections on the surface of HIV. *Arch. Virol.* 100, 255-266.
197. Page, R.D.M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *CABIOS* 12, 357-358.
198. Parks, G.D. and Lamb, R.A. (1990). Defective assembly and intracellular transport of mutant paramyxovirus hemagglutinin-neuraminidase proteins containing altered cytoplasmic domains. *J. Virol.* 64, 3605-3616.
199. Peeters, M. (2000). Recombinant HIV sequences: their role in the global epidemic, p54-72. *In* Kuiken, C., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Mullins, J., Wolinsky, S., and Korber, B., eds., *Human Retroviruses and AIDS 2000: A Compilation and*

Analysis of Nucleic Acid and Amino Acid Sequences. Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.

200. Peeters, M. and Sharp, P.M. (2000). Genetic diversity of HIV-1: the moving target. *AIDS 14 Suppl 3*, S129-S140.
201. Perelson, A.S., Essunger, P., Cao, Y., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M., and Ho, D.D. (1997). Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature 387*, 188-191.
202. Ping, L.H., Nelson, J.A., Hoffman, I.F., Schock, J., Lamers, S.L., Goodman, M., Vernazza, P., Kazembe, P., Maida, M., Zimba, D., Goodenow, M.M., Eron, J.J. (Jr.), Fiscus, S.A., Cohen, M.S., and Swanstrom, R. (1999). Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: underrepresentation of X4 variants. *J. Virol.* 73, 6271-6281.
203. Pinter, A., Honnen, W.J., Tilley, S.A., Bona, C., Zaghoulani, H., Gorny, M.K., and Zolla-Pazner, S. (1989). Oligomeric structure of gp41, the transmembrane protein of human immunodeficiency virus type 1. *J. Virol.* 63, 2674-2679.
204. Popovic, M., Sarngadharan, M.G., Read, E., and Gallo, R.C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science* 224, 497-500.
205. Rabson, A.B. and Martin, M.A. (1985). Molecular organization of the AIDS retrovirus. *Cell* 40, 477-480.
206. Ratner, L. (1992). Glucosidase inhibitors for treatment of HIV-1 infection. *AIDS Res. Hum. Retroviruses* 8, 165-173.
207. Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R. (Jr.), Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghayeb, J., Chang, N.T., Gallo, R.C., and Wong-Staal, F. (1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* 313, 277-284.

208. Ratner, L., Vander Heyden, N., Garcia, J., Polinsky, M., Westervelt, P., and Becich, M. (1991). Formation of noninfectious HIV-1 virus particles lacking a full-length envelope protein. *AIDS Res. Hum. Retroviruses* 7, 287-294.
209. Robertson, D.L., Sharp, P.M., McCutchan, F.E., and Hahn, B.H. (1995). Recombination in HIV-1. *Nature* 374, 124-126.
210. Robey, W.G., Safai, B., Oroszlan, S., Arthur, L.O., Gonda, M.A., Gallo, R.C., and Fischinger, P.J. (1985). Characterization of envelope and core structural gene products of HTLV- III with sera from AIDS patients. *Science* 228, 593-595.
211. Rodenburg, C.M., Li, Y., Trask, S.A., Chen, Y., Decker, J., Robertson, D.L., Kalish, M.L., Shaw, G.M., Allen, S., Hahn, B.H., and Gao, F. (2001). Near full-length clones and reference sequences for subtype C isolates of HIV type 1 from three different continents. *AIDS Res. Hum. Retroviruses* 17, 161-168.
212. Rottier, P.J., Florkiewicz, R.Z., Shaw, A.S., and Rose, J.K. (1987). An internalized amino-terminal signal sequence retains full activity in vivo but not in vitro. *J. Biol. Chem.* 262, 8889-8895.
213. Sabin, C.A., Devereux, H., Phillips, A.N., Hill, A., Janossy, G., Lee, C.A., and Loveday, C. (2000). Course of viral load throughout HIV-1 infection. *J. Acquir. Immune. Defic. Syndr.* 23, 172-177.
214. Sackett, K. and Shai, Y. (2002). The HIV-1 gp41 N-terminal heptad repeat plays an essential role in membrane fusion. *Biochemistry* 41, 4678-4685.
215. Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-425.
216. Salminen, M.O., Johansson, B., Sönnernborg, A., Ayehunie, S., Gotte, D., Leinikki, P., Burke, D.S., and McCutchan, F.E. (1996). Full-length sequence of an ethiopian human immunodeficiency virus type 1 (HIV-1) isolate of genetic subtype C. *AIDS Res. Hum. Retroviruses* 12, 1329-1339.

217. Salzwedel, K., West, J.T., and Hunter, E. (1999). A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity. *J. Virol.* 73, 2469-2480.
218. Sambrook, S., Fritsch, E.F., and Maniatis, T., eds. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edition, p. 9.14 - 9.23, 1.25 - 1.28, E5 and A5. Published by Cold Spring Harbor Laboratory Press, New York, NY.
219. Sanchez-Pescador, R., Power, M.D., Barr, P.J., Steimer, K.S., Stempien, M.M., Brown-Shimer, S.L., Gee, W.W., Renard, A., Randolph, A., Levy, J.A., Dina, D., and Luciw, P.A. (1985). Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). *Science* 227, 484-492.
220. Sanders, R.W., Venturi, M., Schiffner, L., Kalyanaraman, R., Katinger, H., Lloyd, K.O., Kwong, P.D., and Moore, J.P. (2002). The mannose-dependent epitope for neutralizing antibody 2G12 on human immunodeficiency virus type 1 glycoprotein gp120. *J. Virol.* 76, 7293-7305.
221. Sanders-Buell, E., Salminen, M., and McCutchan, F. (1995). Sequencing primers for HIV-1, p. III-15 to III-21. *In* Kuiken, C., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J., Mullins, J., Wolinsky, S, and Korber, B., eds., *Human Retroviruses and AIDS 1995: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.
222. Sattentau, Q.J. and Moore, J.P. (1991). Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding. *J. Exp. Med.* 174, 407-415.
223. Sattentau, Q.J., Moore, J.P., Vignaux, F., Traincard, F., and Poignard, P. (1993). Conformational changes induced in the envelope glycoproteins of the human and simian immunodeficiency viruses by soluble receptor binding. *J. Virol.* 67, 7383-7393.
224. Scandella, C.J., Kilpatrick, J., Lidster, W., Parker, C., Moore, J.P., Moore, G.K., Mann, K.A., Brown, P., Coates, S., Chapman, B., Masiarz, F.R., Steimer, K.S., and Haigwood, N.L. (1993). Nonaffinity purification of recombinant gp120 for use in AIDS vaccine development. *AIDS Res. Hum. Retroviruses* 9, 1233-1244.

225. Scanlan, C.N., Pantophlet, R., Wormald, M.R., Ollmann, S.E., Stanfield, R., Wilson, I.A., Katinger, H., Dwek, R.A., Rudd, P.M., and Burton, D.R. (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, 7306-7321.
226. Schaal, H., Klein, M., Gehrmann, P., Adams, O., and Scheid, A. (1995). Requirement of N-terminal amino acid residues of gp41 for human immunodeficiency virus type 1-mediated cell fusion. *J. Virol.* 69, 3308-3314.
227. Schim van der Loeff, M.F. and Aaby, P. (1999). Towards a better understanding of the epidemiology of HIV-2. *AIDS 13 Suppl A*, S69-S84.
228. Schnittman, S.M., Psallidopoulos, M.C., Lane, H.C., Thompson, L., Baseler, M., Massari, F., Fox, C.H., Salzman, N.P., and Fauci, A.S. (1989). The reservoir for HIV-1 in human peripheral blood is a T cell that maintains expression of CD4. *Science* 245, 305-308.
229. Schuitemaker, H., Koot, M., Kootstra, N.A., Dercksen, M.W., de Goede, R.E., van Steenwijk, R.P., Lange, J.M., Schattenkerk, J.K., Miedema, F., and Tersmette, M. (1992). Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytoprotropic to T-cell-tropic virus population. *J. Virol.* 66, 1354-1360.
230. Schupbach, J., Popovic, M., Gilden, R.V., Gonda, M.A., Sarngadharan, M.G., and Gallo, R.C. (1984). Serological analysis of a subgroup of human T-lymphotropic retroviruses (HTLV-III) associated with AIDS. *Science* 224, 503-505.
231. Scriba, T.J., de Villiers, T., Treurnicht, F.K., zur Megede, J., Barnett, S.W., Engelbrecht, S., and Janse van Rensburg, E. (2002). Characterization of the South African HIV type 1 subtype C complete 5' long terminal repeat, nef, and regulatory genes. *AIDS Res. Hum. Retroviruses* 18, 149-159.
232. Scriba, T.J., Treurnicht, F.K., Zeier, M., Engelbrecht, S., and Janse van Rensburg, E. (2001). Characterization and phylogenetic analysis of South African HIV-1 subtype C accessory genes. *AIDS Res. Hum. Retroviruses* 17, 775-781.
233. Segal, M.S., Bye, J.M., Sambrook, J.F., and Gething, M.J. (1992). Disulfide bond formation during the folding of influenza virus hemagglutinin. *J. Cell Biol.* 118, 227-244.

234. Sher, R. (1989). HIV infection in South Africa, 1982-1988-a review. *S. Afr. Med. J.* 76, 314-318.
235. Shiver, J.W., Fu, T.M., Chen, L., Casimiro, D.R., Davies, M.E., Evans, R.K., Zhang, Z.Q., Simon, A.J., Trigona, W.L., Dubey, S.A., Huang, L., Harris, V.A., Long, R.S., Liang, X., Handt, L., Schleif, W.A., Zhu, L., Freed, D.C., Persaud, N.V., Guan, L., Punt, K.S., Tang, A., Chen, M., Wilson, K.A., Collins, K.B., Heidecker, G.J., Fernandez, V.R., Perry, H.C., Joyce, J.G., Grimm, K.M., Cook, J.C., Keller, P.M., Kresock, D.S., Mach, H., Troutman, R.D., Isopi, L.A., Williams, D.M., Xu, Z., Bohannon, K.E., Volkin, D.B., Montefiori, D.C., Miura, A., Krivulka, G.R., Lifton, M.A., Kuroda, M.J., Schmitz, J.E., Letvin, N.L., Caulfield, M.J., Bett, A.J., Youil, R., Kaslow, D.C., and Emini, E.A. (2002). Replication-incompetent adenoviral vaccine vector elicits effective anti-immunodeficiency-virus immunity. *Nature* 415, 331-335.
236. Siliciano, R.F., Lawton, T., Knall, C., Karr, R.W., Berman, P., Gregory, T., and Reinherz, E.L. (1988). Analysis of host-virus interactions in AIDS with anti-gp120 T cell clones: effect of HIV sequence variation and a mechanism for CD4+ cell depletion. *Cell* 54, 561-575.
237. Simon, F., Mauclore, P., Roques, P., Loussert-Ajaka, I., Muller-Trutwin, M.C., Saragosti, S., Georges-Courbot, M.C., Barre-Sinoussi, F., and Brun-Vezinet, F. (1998). Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. *Nat. Med.* 4, 1032-1037.
238. Sodroski, J.G. (1999). HIV-1 entry inhibitors in the side pocket. *Cell* 99, 243-246.
239. Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A., and Wain-Hobson, S. (1985). Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. *Cell* 42, 369-382.
240. Speck, R.F., Wehrly, K., Platt, E.J., Atchison, R.E., Charo, I.F., Kabat, D., Chesebro, B., and Goldsmith, M.A. (1997). Selective employment of chemokine receptors as human immunodeficiency virus type 1 coreceptors determined by individual amino acids within the envelope V3 loop. *J. Virol.* 71, 7136-7139.
241. Srivastava, I.K., Stamatatos, L., Legg, H., Kan, E., Fong, A., Coates, S.R., Leung, L., Wininger, M., Donnelly, J.J., Ulmer, J.B., and Barnett, S.W. (2002). Purification and

characterization of oligomeric envelope glycoprotein from a primary R5 subtype B human immunodeficiency virus. *J. Virol.* 76, 2835-2847.

242. Stamatatos, L. and Cheng-Mayer, C. (1993). Evidence that the structural conformation of envelope gp120 affects human immunodeficiency virus type 1 infectivity, host range, and syncytium-forming ability. *J. Virol.* 67, 5635-5639.
243. Stamatatos, L., Lim, M., and Cheng-Mayer, C. (2000). Generation and structural analysis of soluble oligomeric gp140 envelope proteins derived from neutralization-resistant and neutralization-susceptible primary HIV type 1 isolates. *AIDS Res. Hum. Retroviruses* 16, 981-994.
244. Stamatatos, L., Wiskerchen, M., and Cheng-Mayer, C. (1998). Effect of major deletions in the V1 and V2 loops of a macrophage-tropic HIV type 1 isolate on viral envelope structure, cell entry, and replication. *AIDS Res. Hum. Retroviruses* 14, 1129-1139.
245. Starcich, B.R., Hahn, B.H., Shaw, G.M., McNeely, P.D., Modrow, S., Wolf, H., Parks, E.S., Parks, W.P., Josephs, S.F., Gallo, R.C., and Wong-Staal, F. (1986). Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* 45, 637-648.
246. Stein, B.S. and Engleman, E.G. (1990). Intracellular processing of the gp160 HIV-1 envelope precursor. Endoproteolytic cleavage occurs in a cis or medial compartment of the Golgi complex. *J. Biol. Chem.* 265, 2640-2649.
247. Stein, B.S., Gowda, S.D., Lifson, J.D., Penhallow, R.C., Bensch, K.G., and Engleman, E.G. (1987). pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. *Cell* 49, 659-668.
248. Stevenson, M., Meier, C., Mann, A.M., Chapman, N., and Wasiak, A. (1988). Envelope glycoprotein of HIV induces interference and cytolysis resistance in CD4+ cells: mechanism for persistence in AIDS. *Cell* 53, 483-496.
249. Sullivan, N., Sun, Y., Sattentau, Q., Thali, M., Wu, D., Denisova, G., Gershoni, J., Robinson, J., Moore, J., and Sodroski, J. (1998). CD4-Induced conformational changes in the human immunodeficiency virus type 1 gp120 glycoprotein: consequences for virus entry and neutralization. *J. Virol.* 72, 4694-4703.

250. Syu, W.J., Lee, W.R., Du, B., Yu, Q.C., Essex, M., and Lee, T.H. (1991). Role of conserved gp41 cysteine residues in the processing of human immunodeficiency virus envelope precursor and viral infectivity. *J. Virol.* *65*, 6349-6352.
251. Takahashi, H., Cohen, J., Hosmalin, A., Cease, K.B., Houghten, R., Cornette, J.L., DeLisi, C., Moss, B., Germain, R.N., and Berzofsky, J.A. (1988). An immunodominant epitope of the human immunodeficiency virus envelope glycoprotein gp160 recognized by class I major histocompatibility complex molecule-restricted murine cytotoxic T lymphocytes. *Proc. Natl. Acad. Sci. U. S. A* *85*, 3105-3109.
252. Tan, K., Liu, J., Wang, J., Shen, S., and Lu, M. (1997). Atomic structure of a thermostable subdomain of HIV-1 gp41. *Proc. Natl. Acad. Sci. U. S. A* *94*, 12303-12308.
253. Temin, H.M. and Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* *226*, 1211-1213.
254. Thali, M., Moore, J.P., Furman, C., Charles, M., Ho, D.D., Robinson, J., and Sodroski, J. (1993). Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J. Virol.* *67*, 3978-3988.
255. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* *25*, 4876-4882.
256. Tong-Starksen, S.E., Welsh, T.M., and Peterlin, B.M. (1990). Differences in transcriptional enhancers of HIV-1 and HIV-2. Response to T cell activation signals. *J. Immunol.* *145*, 4348-4354.
257. Treurnicht, F.K., Smith, T-L., Engelbrecht, S., Claassen, M., Robson, B.A., Zeier, M., and Janse van Rensburg, E. (2002). Genotypic and phenotypic analysis of the env gene from South African HIV-1 subtype B and C isolates. *J. Med. Virol.* *68*, 141-146.
258. Trinh, T., Jessee, J, Bloom, F.R., and Hirsch, V. (1994). STBL2: An Escherichia coli strain for the stable propagation of retroviral clones and direct repeat sequences. *Focus* *16*, 78-80.

259. Trkola, A., Purtscher, M., Muster, T., Ballaun, C., Buchacher, A., Sullivan, N., Srinivasan, K., Sodroski, J., Moore, J.P., and Katinger, H. (1996). Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J. Virol.* *70*, 1100-1108.
260. Tscherning, C., Alaeus, A., Fredriksson, R., Björndal, Å., Deng, H., Littman, D.R., Fenyö, E.M., and Albert, J. (1998). Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. *Virology* *241*, 181-188.
261. UNAIDS and WHO. UNAIDS/WHO Report 2001: Report on the global HIV/AIDS epidemic. (2001). Published by World Health Organisation, Geneva, Switzerland.
262. UNAIDS and WHO. UNAIDS/WHO Report 2002: Report on the global HIV/AIDS epidemic. (2002). Published by World Health Organisation, Geneva, Switzerland.
263. van Harmelen, J., Williamson, C., Kim, B., Morris, L., Carr, J., Karim, S.S., and McCutchan, F. (2001). Characterization of full-length HIV type 1 subtype C sequences from South Africa. *AIDS Res. Hum. Retroviruses* *17*, 1527-1531.
264. van Harmelen, J., Wood, R., Lambrick, M., Rybicki, E.P., Williamson, A.L., and Williamson, C. (1997). An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. *AIDS* *11*, 81-87.
265. van Harmelen, J.H., Van der Ryst, E., Loubser, A.S., York, D., Madurai, S., Lyons, S., Wood, R., and Williamson, C. (1999). A predominantly HIV type 1 subtype C-restricted epidemic in South African urban populations. *AIDS Res. Hum. Retroviruses* *15*, 395-398.
266. Veronese, F.D., DeVico, A.L., Copeland, T.D., Oroszlan, S., Gallo, R.C., and Sarngadharan, M.G. (1985). Characterization of gp41 as the transmembrane protein coded by the HTLV- III/LAV envelope gene. *Science* *229*, 1402-1405.
267. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., and Alizon, M. (1985). Nucleotide sequence of the AIDS virus, LAV. *Cell* *40*, 9-17.
268. Wang, J.J., Steel, S., Wisniewolski, R., and Wang, C.Y. (1986). Detection of antibodies to human T-lymphotropic virus type III by using a synthetic peptide of 21 amino acid residues

- corresponding to a highly antigenic segment of gp41 envelope protein. *Proc. Natl. Acad. Sci. U. S. A* *83*, 6159-6163.
269. Wang, S., York, J., Shu, W., Stooler, M.O., Nunberg, J.H., and Lu, M. (2002). Interhelical interactions in the gp41 core: implications for activation of HIV-1 membrane fusion. *Biochemistry* *41*, 7283-7292.
270. Weinberg, J., Liao, H.X., Torres, J.V., Matthews, T.J., Robinson, J., and Haynes, B.F. (1997). Identification of a synthetic peptide that mimics an HIV glycoprotein 120 envelope conformational determinant exposed following ligation of glycoprotein 120 by CD4. *AIDS Res. Hum. Retroviruses* *13*, 657-664.
271. Weiss, C.D., Levy, J.A., and White, J.M. (1990). Oligomeric organization of gp120 on infectious human immunodeficiency virus type 1 particles. *J. Virol.* *64*, 5674-5677.
272. Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J., and Wiley, D.C. (1997). Atomic structure of the ectodomain from HIV-1 gp41. *Nature* *387*, 426-430.
273. WHO Network for HIV Isolation and Characterization (1994). HIV type 1 variation in World Health Organization-sponsored vaccine evaluation sites: genetic screening, sequence analysis, and preliminary biological characterization of selected viral strains. *AIDS Res. Hum. Retroviruses* *10*, 1327-1343.
274. Wilcox, W.C., Long, D., Sodora, D.L., Eisenberg, R.J., and Cohen, G.H. (1988). The contribution of cysteine residues to antigenicity and extent of processing of herpes simplex virus type 1 glycoprotein D. *J. Virol.* *62*, 1941-1947.
275. Wild, C., Dubay, J.W., Greenwell, T., Baird, T. (Jr.), Oas, T.G., McDanal, C., Hunter, E., and Matthews, T. (1994). Propensity for a leucine zipper-like domain of human immunodeficiency virus type 1 gp41 to form oligomers correlates with a role in virus-induced fusion rather than assembly of the glycoprotein complex. *Proc. Natl. Acad. Sci. U. S. A* *91*, 12676-12680.
276. Wiley, C.A., Schrier, R.D., Nelson, J.A., Lampert, P.W., and Oldstone, M.B. (1986). Cellular localization of human immunodeficiency virus infection within the brains of acquired immune deficiency syndrome patients. *Proc. Natl. Acad. Sci. U. S. A* *83*, 7089-7093.

277. Willey, R.L., Bonifacino, J.S., Potts, B.J., Martin, M.A., and Klausner, R.D. (1988). Biosynthesis, cleavage, and degradation of the human immunodeficiency virus 1 envelope glycoprotein gp160. *Proc. Natl. Acad. Sci. U. S. A* 85, 9580-9584.
278. Willey, R.L., Klimkait, T., Frucht, D.M., Bonifacino, J.S., and Martin, M.A. (1991). Mutations within the human immunodeficiency virus type 1 gp160 envelope glycoprotein alter its intracellular transport and processing. *Virology* 184, 319-329.
279. Willey, R.L., Theodore, T.S., and Martin, M.A. (1994). Amino acid substitutions in the human immunodeficiency virus type 1 gp120 V3 loop that change viral tropism also alter physical and functional properties of the virion envelope. *J. Virol.* 68, 4409-4419.
280. Williamson, C., Engelbrecht, S., Lambrick, M., Janse van Rensburg, E., Wood, R., Bredell, W., and Williamson, A.L. (1995). HIV-1 subtypes in different risk groups in South Africa. *Lancet* 346, 782.
281. Wilson, C.C., Palmer, B., Southwood, S., Sidney, J., Higashimoto, Y., Appella, E., Chesnut, R., Sette, A., and Livingston, B.D. (2001). Identification and antigenicity of broadly cross-reactive and conserved human immunodeficiency virus type 1-derived helper T-lymphocyte epitopes. *J. Virol.* 75, 4195-4207.
282. Wu, L., Gerard, N.P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A.A., Desjardin, E., Newman, W., Gerard, C., and Sodroski, J. (1996). CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature* 384, 179-183.
283. Wyatt, R. (2002). Analysis of HIV-1 envelope glycoproteins and neutralizing antibodies. 9th Conference on Retroviruses and Opportunistic Infections, Seattle, WA (Abstract no. L4).
284. Wyatt, R., Kwong, P.D., Desjardins, E., Sweet, R.W., Robinson, J., Hendrickson, W.A., and Sodroski, J.G. (1998). The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* 393, 705-711.
285. Wyatt, R., Moore, J., Accola, M., Desjardin, E., Robinson, J., and Sodroski, J. (1995). Involvement of the V1/V2 variable loop structure in the exposure of human

- immunodeficiency virus type 1 gp120 epitopes induced by receptor binding. *J. Virol.* *69*, 5723-5733.
286. Wyatt, R. and Sodroski, J. (1998). The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* *280*, 1884-1888.
287. Yamaguchi-Kabata, Y. and Gojobori, T. (2000). Reevaluation of amino acid variability of the human immunodeficiency virus type 1 gp120 envelope glycoprotein and prediction of new discontinuous epitopes. *J. Virol.* *74*, 4335-4350.
288. Yang, M., Omura, S., Bonifacino, J.S., and Weissman, A.M. (1998). Novel aspects of degradation of T cell receptor subunits from the endoplasmic reticulum (ER) in T cells: importance of oligosaccharide processing, ubiquitination, and proteasome-dependent removal from ER membranes. *J. Exp. Med.* *187*, 835-846.
289. Yang, X., Wyatt, R., and Sodroski, J. (2001). Improved elicitation of neutralizing antibodies against primary human immunodeficiency viruses by soluble stabilized envelope glycoprotein trimers. *J. Virol.* *75*, 1165-1171.
290. Yu, X., Yuan, X., McLane, M.F., Lee, T.H., and Essex, M. (1993). Mutations in the cytoplasmic domain of human immunodeficiency virus type 1 transmembrane protein impair the incorporation of Env proteins into mature virions. *J. Virol.* *67*, 213-221.
291. Zhang, C.W., Chishti, Y., Hussey, R.E., and Reinherz, E.L. (2001). Expression, purification, and characterization of recombinant HIV gp140. The gp41 ectodomain of HIV or simian immunodeficiency virus is sufficient to maintain the retroviral envelope glycoprotein as a trimer. *J. Biol. Chem.* *276*, 39577-39585.
292. Zhang, L., Huang, Y., He, T., Cao, Y., and Ho, D.D. (1996). HIV-1 subtype and second-receptor use. *Nature* *383*, 768.
293. Zhu, T., Korber, B.T., Nahmias, A.J., Hooper, E., Sharp, P.M., and Ho, D.D. (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* *391*, 594-597.

294. Zubay, G. (1993). Noncovalent intermolecular forces determine the three-dimensional folding pattern of macromolecules, p.12-13. *In* Zubay, G. ed., *Biochemistry* 3rd edition. Published by Wm. C. Brown Publishers, Dubuque, Iowa.
295. zur Megede, J., Chen, M.C., Doe, B., Schaefer, M., Greer, C.E., Selby, M., Otten, G.R., and Barnett, S.W. (2000). Increased expression and immunogenicity of sequence-modified human immunodeficiency virus type 1 gag gene. *J. Virol.* 74, 2628-2635.
296. zur Megede, J., Engelbrecht, S., de Oliveira, T., Cassol, S., Scriba, T.J., Janse van Rensburg, E., and Barnett, S.W. (2002). Novel evolutionary analyses of full-length HIV type 1 subtype C molecular clones from Cape Town, South Africa. *AIDS Res. Hum. Retroviruses* 18, 1327-1332.
297. Zwick, M.B., Bonnycastle, L.L., Menendez, A., Irving, M.B., Barbas, C.F. (III), Parren, P.W., Burton, D.R., and Scott, J.K. (2001a). Identification and characterization of a peptide that specifically binds the human, broadly neutralizing anti-human immunodeficiency virus type 1 antibody b12. *J. Virol.* 75, 6692-6699.
298. Zwick, M.B., Labrijn, A.F., Wang, M., Spenlehauer, C., Saphire, E.O., Binley, J.M., Moore, J.P., Stiegler, G., Katinger, H., Burton, D.R., and Parren, P.W. (2001b). Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41. *J. Virol.* 75, 10892-10905.

APPENDIX A

This section shows the sequences used in the alignments. The sequence common names, accession numbers, subtype of the sequences, sampling country of the sequences and the articles in which the sequences were first described are listed.

Sequences with references used in alignments. Table adapted from the Los Alamos Database (HIV Sequence Compendium, 2001) and <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>.

Subtype	Accession	Common name	Sample Country	First author	Reference
A1	AF004885	A1.KE.Q23-17	Kenya	Poss, M	J Virol 73(5): 4393-4403 (1999)
A1	AF069670	A1.SE.SE7253	Sweden	Carr, JK	AIDS 13(14): 1819-1826 (1999)
A1	M62320	A1.UG.U455	Uganda	Oram, JD	ARHR 6(9): 1073-1078 (1990)
A1	U51190	A1.UG.92UG037	Uganda	Gao, F	J Virol 70(3): 1651-1657 (1996)
A2	AF286238	A2.CD.97.CDKTB48	Democratic republic of Congo	Gao, F	ARHR 17(8): 675-688 (2001)
A2	AF286237	A2.CY.94CY017-41	Cyprus	Gao, F	ARHR 17(8): 675-688 (2001)
A2	AF286241	A2.CD.97CDKS10	Democratic republic of Congo	Gao, F	ARHR 17(8): 675-688 (2001)
A2	AF286240	A2.CD.97CDKFE4	Democratic republic of Congo	Gao, F	ARHR 17(8): 675-688 (2001)
B	K03455	B.FR.HXB2	France	Ratner, L	Nature 313(6000): 277-284 (1985)
B	M17451	B.US.RF	United States	Starcich, BR	Cell 45(5): 637-648 (1986)
B	U63632	B.US.JRFL	United States	O'Brien, WA	Nature 348(6296): 69-73 (1990)
B	U21135	B.US.WEAU160	United States	Ghosh, S	Unpublished
C	AB023804	C.93IN101	India	Mochizuki, N	ARHR 15(14): 1321-1324 (1999)
C	AF067154	C.93IN999	India	Lole,KS	J Virol 73(1): 152-160 (1999)
C	AF067155	C.95IN21068	India	Lole,KS	J Virol 73(1): 152-160 (1999)
C	AF067157	C.93IN904	India	Lole,KS	J Virol 73(1): 152-160 (1999)
C	AF067158	C.93IN905	India	Lole,KS	J Virol 73(1): 152-160 (1999)
C	AF067159	C.94IN11246	India	Lole,KS	J Virol 73(1): 152-160 (1999)
C	AF110959	C.96BW01B03	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110960	C.96BW01B21	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110961	C.96BW01B22	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)

C	AF110962	C.96BW0402	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110963	C.96BW0407	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110964	C.96BW0408	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110965	C.96BW0409	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110966	C.96BW0410	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110967	C.96BW0502	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110968	C.96BW0504	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110969	C.96BW1104	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110970	C.96BW1106	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110971	C.96BW11B01	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110972	C.96BW1210	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110973	C.96BW15B03	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110974	C.96BW15C02	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110975	C.96BW15C05	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110976	C.96BW16B01	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110977	C.96BW16D14	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110978	C.96BW1626	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110979	C.96BW17A09	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110980	C.96BW17B03	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF110981	C.96BW17B05	Botswana	Novitsky, VA	J Virol 73(5): 4427-4432 (1999)
C	AF219262	C.FIN9126	Finland	Liitsola, K	Scand J Infect Dis 32(5): 475-480 (2000)
C	AF219263	C.FIN9149	Finland	Liitsola, K	Scand J Infect Dis 32(5): 475-480 (2000)
C	AF219264	C.FIN9158	Finland	Liitsola, K	Scand J Infect Dis 32(5): 475-480 (2000)
C	AF219266	C.FIN9210	Finland	Liitsola, K	Scand J Infect Dis 32(5): 475-480 (2000)
C	AF219269	C.FIN9277	Finland	Liitsola, K	Scand J Infect Dis 32(5): 475-480 (2000)
C	AF219270	C.FIN9307	Finland	Liitsola, K	Scand J Infect Dis 32(5): 475-480 (2000)

C	AF219274	C.FIN9401	Finland	Liitsola, K	Scand J Infect Dis 32(5): 475-480 (2000)
C	AF286223	C.94IN476	India	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
C	AF286224	C.96ZM651	Zambia	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
C	AF286225	C.96ZM751	Zambia	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
C	AF286227	C.97ZA012	South Africa	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
C	AF286228	C.98BR004	Brazil	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
C	AF286231	C.98IN012	India	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
C	AF286232	C.98IN022	India	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
C	AF286233	C.98IS002	Israel	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
C	AF286234	C.98TZ013	Tanzania	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
C	AF286235	C.98TZ017	Tanzania	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
C	AF290027	C.96BW06H51	Botswana	Ndung'u, T	Virology 278(2): 390-399 (2000)
C	AF290028	C.96BW06J4	Botswana	Ndung'u, T	Virology 278(2): 390-399 (2000)
C	AF290029	C.96BW06J7	Botswana	Ndung'u, T	Virology 278(2): 390-399 (2000)
C	AF290030	C.96BW06K18	Botswana	Ndung'u, T	Virology 278(2): 390-399 (2000)
C	AF321081	C.FR.CMBA	France	Lemiale, F	Unpublished
C	AF321523	C.BW.MJ4	Botswana	Ndung'u, T	J Virol 75(11): 4964-4972 (2001)
C	AF361874	C.97TZ04	Tanzania	Hoelscher, M	AIDS 15(12): 1461-1470 (2001)
C	AF361875	C.97TZ05	Tanzania	Hoelscher, M	AIDS 15(12): 1461-1470 (2001)
C	AF391230	C.98TV001c8.2	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391231	C.98TV001c8.5	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391232	C.98TV002c12.1	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391233	C.98TV003cE260	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391234	C.98TV004Cc300	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391235	C.98TV006c9.1	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391236	C.98TV006c9.2	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)

C	AF391237	C.98TV006cE9	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391238	C.98TV007cB104	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391239	C.98TV007cB105	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391240	C.98TV008c4.3	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391241	C.98TV008c4.4	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391242	C.98TV010cD7	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391243	C.98TV012c2.1	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391244	C.98TV012c2.2	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391245	C.98TV013cB20	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391246	C.98TV013cH17	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391247	C.98TV014c6.3	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391248	C.98TV014c6.4	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391249	C.98TV018cF1027	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF391250	C.98TV019c5	South Africa	Engelbrecht, S	ARHR 17(16): 1533-1547 (2001)
C	AF443074	C.96BWMO15	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443075	C.96BWM032	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443076	C.98BWMC122	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443077	C.98BWMC134	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443078	C.98BWMC14A3	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443079	C.98BWMO1410	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443080	C.98BWMO18D5	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443081	C.98BWMO36A5	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443082	C.98BWMO37D5	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443083	C.99BW393212	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443084	C.99BW46424	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443085	C.99BW47458	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)

C	AF443086	C.99BW47547	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443087	C.99BWMC168	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443088	C.00BW07621	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443089	C.00BW076820	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443090	C.00BW087421	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443091	C.00BW147127	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443092	C.00BW16162	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443093	C.00BW16868	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443094	C.00BW17593	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443095	C.00BW17732	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443096	C.00BW17835	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443097	C.00BW17956	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443098	C.00BW18113	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443099	C.00BW18595	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443100	C.00BW18802	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443101	C.00BW192113	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443102	C.00BW20361	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443103	C.00BW20636	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443104	C.00BW20872	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443105	C.00BW2127214	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443106	C.00BW21283	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443107	C.00BW22767	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443108	C.00BW38193	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443109	C.00BW38428	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443110	C.00BW38713	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443111	C.00BW38769	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)

C	AF443112	C.00BW38868	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443113	C.00BW38916	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443114	C.00BW39702	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF443115	C.00BW50311	Botswana	Novitsky, V	J Virol 76(11): 5435-5451 (2002)
C	AF457054	C.KER2010	Kenya	Dowling, WE	AIDS 16(13): 1809-1820 (2002)
C	AJ276221	C.CALCMANDAL	India	Mandal, D	Unpublished
C	AX455917	C.PAT33	South Africa	zur Megede, J	Patent WO 0204493-A 33 17-Jan-2002
C	AX455929	C.98TV002 full-length	South Africa	zur Megede, J	Patent WO 0204493-A 33 17-Jan-2002
C	AY043173	C.ZA.DU151	South Africa	van Harmelen, J	ARHR 17(16): 1527-1531 (2001)
C	AY043174	C.ZA.DU179	South Africa	van Harmelen, J	ARHR 17(16): 1527-1531 (2001)
C	AY043175	C.ZA.DU422	South Africa	van Harmelen, J	ARHR 17(16): 1527-1531 (2001)
C	AY043176	C.ZA.CTSC2	South Africa	van Harmelen, J	ARHR 17(16): 1527-1531 (2001)
C	L22940	C.DJ259A	Djibouti	Louwagie, J	J Virol 69(1): 263-271 (1995)
C	L22944	C.SE364A	Sweden	Louwagie, J	J Virol 69(1): 263-271 (1995)
C	L22946	C.SO145A	Somalia	Louwagie, J	J Virol 69(1): 263-271 (1995)
C	L22948	C.UG268A2	Uganda	Louwagie, J	J Virol 69(1): 263-271 (1995)
C	L23065	C.DJ373A	Djibouti	Louwagie, J	J Virol 69(1): 263-271 (1995)
C	U39233	C.BI.BU310112	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39234	C.91BR015.BR911513	Brazil	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39237	C.BI.BU310213	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39238	C.91BR015.BR911514	Brazil	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39239	C.BI.BU910316	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39240	C.BI.BU310518	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39241	C.BI.BU910430	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39242	C.BI.BU910322	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39243	C.BI.BU910423	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)

C	U39244	C.BI.BU910612	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39245	C.BI.BU910717	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39246	C.BI.BU910611	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39247	C.BI.BU910718	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39248	C.BI.BU910214	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39249	C.BI.BU910719	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39250	C.BI.BU910815	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39251	C.BI.BU910812	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39252	C.BI.BU910113	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U39257	C.BI.BU910531	Burundi	Penny, MA	ARHR 12(8): 741-747 (1996)
C	U46016	C.ETH2220	Ethiopia	Salminen, MO	ARHR 12(14): 1329-1339 (1996)
C	U52953	C.92BR025	Brazil	Gao, F	J Virol 70(3): 1651-1667 (1996)
D	K03454	D.CD.ELI	Democratic republic of Congo	Alizon, M	Cell 46(1): 63-74 (1986)
D	M27323	D.CD.NDK	Democratic republic of Congo	Spire, B	Gene 81(2): 275-284 (1989)
D	U88822	D.CD.84ZR085	Democratic republic of Congo	Gao, F	J Virol 72(7): 5680-5698 (1998)
D	U88824	D.UG.94UG1141	Uganda	Gao, F	J Virol 72(7): 5680-5698 (1998)
F1	AF077336	F1.BE.VI850	Belgium	Laukkanen, T	Virology 269(1): 95-104 (2000)
F1	AF005494	F1.BR.93BR020-1	Brazil	Gao, F	J Virol 72(7): 5680-5698 (1998)
F1	AF075703	F1.FI.FIN9363	Finland	Laukkanen, T	Virology 269(1): 95-104 (2000)
F1	AJ249238	F1.FR.MP411	France	Triques, K	ARHR 16(2): 139-151 (2000)
F2	AJ249236	F2.CM.MP255	Cameroon	Triques, K	ARHR 16(2): 139-151 (2000)
F2	AJ249237	F2.CM.MP257	Cameroon	Triques, K	ARHR 16(2): 139-151 (2000)
G	AF084936	G.BE.DRCBL	Belgium	Oelrichs, RB	ARHR 15(6): 585-589 (1999)
G	AF061641	G.FI.HH8793-12-1	Finland	Salminen, MO	ARHR 8(9): 1733-1742 (1992)

G	U88826	G.NG.92NG083	Nigeria	Gao, F	J Virol 72(7): 5680-5698 (1998)
G	AF061642	G.SE.SE6165	Sweden	Carr, JK	Virology 247(1): 22-31 (1998)
H	AF190127	H.BE.VI991	Belgium	Janssens, W	AIDS 14(11): 1533-1543 (2000)
H	AF190128	H.BE.VI997	Belgium	Janssens, W	AIDS 14(11): 1533-1543 (2000)
H	AF005496	H.CF.90CF056	Central African Republic	Gao, F	J Virol 72(7): 5680-5698 (1998)
J	AF082394	J.SE.SE7887	Sweden	Laukkanen, T	ARHR 15(3): 293-297 (1999)
J	AF082395	J.SE.SE7022	Sweden	Laukkanen, T	ARHR 15(3): 293-297 (1999)
K	AJ249235	K.CD.EQTB11C	Democratic republic of Congo	Triques, K	ARHR 16(2): 139-151 (2000)
K	AJ249239	K.CM.MP535	Cameroon	Triques, K	ARHR 16(2): 139-151 (2000)
01_AE	AF197340	01_AE.CF.90CF11697	Central African Republic	Anderson, JP	J Virol 74(22): 10752-10765 (2000)
01_AE	U51188	01_AE.CF.90CF402	Central African Republic	Gao, F	J Virol 70(10): 7013-7029 (1996)
01_AE	AF197341	01_AE.CF.90CF4071	Central African Republic	Anderson, JP	J Virol 74(22): 10752-10765 (2000)
01_AE	U54771	01_AE.TH.CM240	Thailand	Carr, JK	J Virol 70(9): 5935-5943 (1996)
02_AG	AJ286133	02_AG.CM.97CM-MP807	Cameroon	Montavon, C	J Acquir Immune Defic Syndr 23(5): 363-374 (2000)
02_AG	AF063224	02_AG.FR.DJ264	France	Carr, JK	Virology 247(1): 22-31 (1998)
02_AG	L39106	02_AG.NG.IBNG	Nigeria	Howard, TM	ARHR 10(12): 1755-1757 (1994)
02_AG	AF107770	02_AG.SE.SE7812	Senegal	Laukkanen, T	Unpublished
03_AB	AF193276	03_AB.RU.KAL153-2	Russia	Liitsola, K	ARHR 16(11): 1047-1053 (2000)
03_AB	AF193277	03_AB.RU.RU98001	Russia	Liitsola, K	ARHR 16(11): 1047-1053 (2000)
04_CPX	AF049337	04_CPX.CY.94CY032-3	Cyprus	Gao, F	J Virol 72(12): 10234-10241 (1998)
04_CPX	AF119820	04_CPX.GR.97PVCH	Greece	Nasioulas, G	ARHR 15(8): 745-758 (1999)
04_CPX	AF119819	04_CPX.GR.97PVMY	Greece	Nasioulas, G	ARHR 15(8): 745-758 (1999)

05_DF	AF193253	05_DF.BE.VI1310	Belgium	Laukkanen, T	Virology 269(1): 95-104 (2000)
05_DF	AF076998	05_DF.BE.VI961	Belgium	Laukkanen, T	Virology 269(1): 95-104 (2000)
06_CPX	AF064699	06_CPX.AU.BFP90	Australia	Oelrichs, RB	ARHR 14(16): 1495-1500 (1998)
06_CPX	AJ288982	06_CPX.ML.95ML127	Mali	Montavon, C	Unpublished
06_CPX	AJ245481	06_CPX.ML.95ML84	Mali	Montavon, C	ARHR 15(18): 1707-1712 (1999)
06_CPX	AJ288981	06_CPX.SN.97SE1078	Senegal	Montavon, C	Unpublished
07_BC	AF286226	07_BC.CN.97CN001	China	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
07_BC	AX149647	07_BC.CN.C54A	China	Shao, Y	Patent: WO 0136614-A 1 25-May-2001
07_BC	AF286230	07_BC.CN.98CN009	China	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
08_BC	AY008715	08_BC.CN.97CNGX6F	China	Piyasirisilp, S	J Virol 74(23): 11286-11295 (2000)
08_BC	AY008716	08_BC.CN.97CNGX7F	China	Piyasirisilp, S	J Virol 74(23): 11286-11295 (2000)
08_BC	AY008717	08_BC.CN.97CNGX9F	China	Piyasirisilp, S	J Virol 74(23): 11286-11295 (2000)
08_BC	AF286229	08_BC.CN.98CN006	China	Rodenburg, CM	ARHR 17(2): 161-168 (2001)
10_CD	AF289548	10_CD.TZ.96TZBF061	Tanzania	Kouliniska, IN	ARHR 17(5): 423-431 (2001)
10_CD	AF289549	10_CD.TZ.96TZBF071	Tanzania	Kouliniska, IN	ARHR 17(5): 423-431 (2001)
10_CD	AF289550	10_CD.TZ.96TXBF110	Tanzania	Kouliniska, IN	ARHR 17(5): 423-431 (2001)
11_CPX	AJ291718	11_CPX.CM.MP818	Cameroon	Montavon, C	ARHR 18(3): 231-236 (2002)
11_CPX	AJ291719	11_CPX.FR.MP1298	France	Montavon, C	ARHR 18(3): 231-236 (2002)
11_CPX	AJ291720	11_CPX.FR.MP1307	France	Montavon, C	ARHR 18(3): 231-236 (2002)
11_CPX	AF179368	11_CPX.GR.GR17	Greece	Paraskevis, D	ARHR 16(9): 845-855 (2000)
12_BF	AF385936	12_BF.AR.ARMA159	Argentina	Carr, JK	AIDS 15(15): F41-F47 (2001)
12_BF	AF385934	12_BF.UY.URTR23	Uruguay	Carr, JK	AIDS 15(15): F41-F47 (2001)
12_BF	AF385935	12_BF.UY.URTR35	Uruguay	Carr, JK	AIDS 15(15): F41-F47 (2001)
12_BF	AF408630	12_BF.AR.A32989	Argentina	Thomson, MM	J Gen Virol 83(Pt 1): 107-119 (2002)
13_CPX	AF460974	13_CPX.CM.96CM4164	Cameroon	Wilbe, K	ARHR 18(12): 849-856 (2002)
13_CPX	AF460972	13_CPX.CM.96CM1849	Cameroon	Wilbe, K	ARHR 18(12): 849-856 (2002)

14_BG	AF423758	14_BG.ES.X475	Spain	Delgado, E	J Acquir Immune Defic Syndr 29(5): 536-543 (2002)
14_BG	AF423756	14_BG.ES.X397	Spain	Delgado, E	J Acquir Immune Defic Syndr 29(5): 536-543 (2002)
14_BG	AF450097	14_BG.ES.X623	Spain	Delgado, E	J Acquir Immune Defic Syndr 29(5): 536-543 (2002)
14_BG	AF423759	14_BG.ES.X477	Spain	Delgado, E	J Acquir Immune Defic Syndr 29(5): 536-543 (2002)
U	AF457101	U.CD.90CD121E12	Democratic republic of Congo	Mokili, JL	ARHR 18(11): 817-823 (2002)
U	AY046058	U.GR.GR303	Greece	Paraskevis, D	J Gen Virol 82(Pt 10): 2509-2514 (2001)
U	AF286236	U.CD.83CD003	Democratic republic of Congo	Gao, F	ARHR 17(12): 1217-1222 (2001)
O	L20587	O.CM.ANT70	Cameroon	Vanden Haesevelde, M	J Virol 68(3): 1586-1596 (1994)
O	L20571	O.CM.MVP5180	Cameroon	Gurtler, LG	J Virol 68(3): 1581-1585 (1994)
O	AJ302646	O.SN.SEMP1299	Senegal	Toure-Kane, C	ARHR 17(12): 1211-1216 (2001)
O	AJ302647	O.SN.SEMP1300	Senegal	Toure-Kane, C	ARHR 17(12): 1211-1216 (2001)
N	AJ271370	N.CM.YBF106	Cameroon	Souquiere, S	Unpublished
N	AJ006022	N.CM.YBF30	Cameroon	Simon, F	Nature Med 4(9): 1032-1037 (1998)
CPZ	U42720	CPZ.CD.CPZANT	Democratic republic of Congo	Vanden Haesevelde, M	Virology 221(2): 346-350 (1996)
CPZ	AF115393	CPZ.CM.CAM3	Cameroon	Corbet, S	J Virol 74(1): 529-534 (2000)
CPZ	AJ271369	CPZ.CM.CAM5	Cameroon	Souquiere, S	Unpublished
CPZ	X52154	CPZ.GA.CPZGAB	Gabon	Huet, T	Nature 345(6273): 356-359 (1990)
CPZ	AF103818	CPZ.US.CPZUS	United States	Gao, F	Nature 397(6718): 436-441 (1999)

APPENDIX B

An alignment of the complete Env protein from TV Env sequences is shown in this section. The 4 pCMVLink clones are also shown in the alignment. The TV env sequences are compared to the C.ZA.97ZA012 sequence (Also see Appendix A). The B.FR.HXB2 sequence was used to identify conserved glycosylation sites and cysteine residues. The subtype C reference strains are also added (HIV Sequence Compendium, 2001). Dots denote sequence identity with the C.ZA.97ZA012 sequence, while dashes represent gaps introduced to optimise the alignment. Carets indicate glycosylation sites previously identified in the B.FR.HXB2 sequence. The signal peptide, V1, V2, V3, V4, and V5 variable loops, CD4-binding residues and sites, fusion domain/peptide, immunodominant region, tryptophan-rich sequence, membrane-spanning domain and intracellular cytoplasmic domain/tail are indicated above the sequences. A blue box indicates the deletion observed at the C-terminal of the signal peptide for the TV006 sequences (TV006cE9, TV006c9.1, TV006c9.2 and LinkTV006cE9). Yellow boxes indicate the glycosylation sites observed in the sequences. A purple box indicates the crown of the V3 loop. The fusogenic site of the fusion peptide is indicated in a green box. Amino acid changes observed in the LinkTV006cE9 and LinkTV018cF1027 sequences are shown in open boxes. Triangles at positions 11, 25, and 35 of the V3 loop indicate amino acids assessed for SI/NSI phenotype.



B.FR.HXB2	:	MRVKEKYQHLWRWGWRWGTMLLGLMLMI [*] CSATEKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHNWATHACVPTDPNPQEVVLLV ^{^^^} NTENFNM	:	95
C.ZA.97ZA012	:	MRVRGIPRNWPQW-WMWGILGFWMI [*] ICRVVGNM [*] VTVYYGVPVWTDAKTTLFCASDTKAYDRE [*] VHNWATHACVPTDPNPQEI ^{^^^} VLENVTENFNM	:	94
C.IN.95IN21068	:L...Q...-.....N.....KE.N.....A...E.....	:	94
C.BW.96BW0502	:	...M..L...Q...-.....SS.....KE.....T..A...ET.....	:	94
C.BR.92BR025	:	...E..Q...K...-.....YN.R.....KE.....A...A.....	:	94
C.ET.ETH2220	:	...M..Q..CQ...-.....NG.....K..SP.....A...T.....G.F.....S...G.....	:	94
TV007cB104	:L...Q...-...AS.....SYN.....KE.....A.G.E.....	:	94
TV007cB105	:L...Q...-...AS.....SYN.....KE.....A.G.E.....	:	94
TV014c6.4	:Q...Q...-.....N.....KE.....A...E.....G.....	:	94
TV014c6.3	:	-G...Q...Q...-.....N.....KE.I.....A...E.....G.....	:	93
TV008c4.3	:	...M..L..C...-.....S.....N.G.K.....KE.....A...E.....	:	94
TV008c4.4	:	...M..L..C...-.....S.....N.G.K.....KE.....A...E.....	:	94
TV019c5	:	...Q..L..CQ...-T.....T.N.....KE.....A...E.....	:	94
TV013cB20	:	...E.Q.....-.....SG.....KE.T.....A...E.....H...	:	94
TV013CH17	:	...E.Q.....-.....L.....SG.....KE.T.....A...E.....H...	:	94
TV006cE9	:	...T..L.....-...V.....Y N.G.....KE.....A.....D.....	:	92
LinkTV006cE9	:	...T..L.....-...V.....Y N.G.....KE.....A.....D.....	:	92
TV006c9.1	:	...T..L.....-...V...-...Y N.G.-.....KE.....A.....	:	90
TV006c9.2	:	...L.....-...V.....Y N.G.....K.....A.....	:	92
TV004cC300	:	...E...Q...-.....N.....KE.....A...EN.....	:	94
LinkTV004cC300	:	...E...Q...-.....N.....KE.....A...EN.....	:	94
TV003cE260	:L...Q...-.....YS.....KE.....A...E.....G.....	:	94
TV012c2.1	:	...M.SR...Q...-.....GN.....KE..A.....A...E.....	:	94
TV012c2.2	:	...M.SR...Q...-.....GN.....KE..A.....A...E.....	:	94
TV018cF1027	:	...M..K..CQ...-.....N.....RE.....W..A...E.....	:	94
LinkTV018cF1027	:	...M..K..CQ...-.....N.....RE.....W..A...E.....	:	94
TV001c8.2	:	...M.TQ..CQ...-.....NTED-.....R.....A...ET.....G.....	:	93
TV001c8.5	:	...M.TQ..CQ...-.....NTED-.....RE.....A...ET.....G.....	:	93
TV002c12.1	:	..A...L...RH.-.....N.K.-.....GRE.....A...E.....G.....	:	93
TV010cD7	:	...M..L..CQ...-.....N.....KE.....A.G.E.....	:	92

B.FR.HXB2 : WKNDMVEQM^{*}HEDI^{*}I^{*}SLWDQSLK-PCVKLTPLCVSLKCTDLKND^{AAA}TNT^{AAA}-----NSSSSGRMIME-----KGEIK^{AAA}NCSFN^{AAA}IST^{AAA}SIRGKVQ : 170
 C.ZA.97ZA012 : WKNDMVDQM^{*}HEDI^{*}I^{*}SLWDQSLK-PCVKLTPLCVTLHCT---NAT^{AAA}FKN---NVTN-----DMNKEIRNCSFN^{AAA}ITTEIRDKKQ : 163
 C.IN.95IN21068 :E.RN-V.S.G-----G.HS--KTYNES-----K-.....A.V..... : 168
 C.BW.96BW0502 :K.RN-V...NNI----.SMIDN-----SNKG.....V..... : 167
 C.BR.92BR025 : .E...E.....R.IDYN---R.....D.GG.....M..... : 164
 C.ET.ETH2220 :E.....G...N.NA-----IK--.N.KVT---NNSIN-----SA.D.....I.....R : 167
 TV007cB104 :R.VN-V...S-----A.SSSSATS^{AAA}SDN-----P.G.....V.....K : 171
 TV007cB105 :R.VN-V...S-----A.SSSSATS^{AAA}SDN-----P.G.....V.....K : 171
 TV014c6.4 :E.G----VKAT--.N-----SATNSSNVT-NKD.....A.....H : 169
 TV014c6.3 :E.G----VKAT--.N-----SATNSSNVT-N.D.....A.....H : 168
 TV008c4.3 :N...--DVIPS--...NTTVTHN---NIT---DKE.....I...T...S : 170
 TV008c4.4 :N...--DVIPS--...VTHNNIMD-----KE.....I...T...S : 167
 TV019c5 :N...--DV-DK--.S.-EM^{AAA}YRKT^{AAA}TNDNG-NDT-----A.D.Q...T : 173
 TV013cB20 :N...--NAINT--.A.-STTTTSATATSTIATSTY.N.G..... : 179
 TV013CH17 :N...--NAINT--.A.-STTTTSATATSTIATSTY.N.G..... : 179
 TV006cE9 :N.KA--.V.V.TNA--...VN^{AAA}TT-NF^{AAA}ND^{AAA}S-----IE.....I..... : 169
 LinkTV006cE9 :N.KA--.V.V.TNA--...VN^{AAA}TT-NF^{AAA}ND^{AAA}S-M-----IE.....I..... : 168
 TV006c9.1 :N.KA-----VN^{AAA}TT-NF^{AAA}ND^{AAA}S-----IE.....I..... : 158
 TV006c9.2 :N.D-----VN^{AAA}TTSLTVSPTV---NIT--E.....I.....K : 167
 TV004cC300 :Q.....N..T-----VT.S.VN^{AAA}TRGEMR-----M..... : 161
 LinkTV004cC300 :Q.....N..T-----VT.S.VN^{AAA}TRGEMR-----M..... : 161
 TV003cE260 :E.NN-----VN^{AAA}TTN-EMT-----A.D..... : 155
 TV012c2.1 :N...-----NNVKNAT--.SM.....I.....K : 160
 TV012c2.2 :N...-----NNVKNAT--.SM.....I.....K : 160
 TV018cF1027 : .E.....N.R--.V.VTTNNDN...YNNS-----PE.....I.....K : 169
 LinkTV018cF1027 : .E.....N.R--.V.VTTNNDN...YNNS-----PE.....I.....K : 169
 TV001c8.2 :A.....N..D-----T...GNRTVTGNSTNNTNGT-GIYNIE.....A.....H : 175
 TV001c8.5 :A.....N..D-----T...GNRTVTGNT-NDTNIANATYKYE.....A.....H : 175
 TV002c12.1 :Q.....N...-----NATVNY---NNT-S.D.....YV.....K : 159
 TV010cD7 :AM.KVDPT..HFKLYK-----CYYLHST-----GD.....V....T : 158



```

B.FR.HXB2      : KEYAFFYKLDIIPID-----AAANDT*AAA*--TSYKLTSCN*TSVITQACPKVSFEPIPIHYCAPAGFAILKCN*AAA*NKTF*AAA*NGTGPCTNV*STVQCTHGIRP : 253

C.ZA.97ZA012  : QGYALFYRPDIVLLKENR--NNS--NNS---EYILINCMANSTITQACPKVNFDPPIPIHYCAPAGYAILKCNNNKTFSSGKGPCNNVSTVQCTHGIRP : 251

C.IN.95IN21068 : TV.....L...P.DNEEQE...--S.--GY.R.....T.A.....T.....N.T...H..... : 259
C.BW.96BW0502  : .VH.....L...P.Q-----G...--N..R.....T.A.....S.....T.....Q..N.T.....A..... : 251
C.BR.92BR025   : KVH.....L...P.----K.E.-S.T.--GD.R.....T.A.....S.....N.T.....T.. : 251
C.ET.ETH2220   : KA.....L...P.N-----G.--TD.R.....T.....SL.....R.....T..H..... : 250

TV007cB104     : .V.....S.D-----S.--T.R.....T.....T.....N.T...T..... : 253
TV007cB105     : .V.....S.D-----S.--T.R.....T.....T.....N.T...T..... : 253
TV014c6.4      : KE.....L...P.NNGNPS--EG.S--EK.R.T..T.....S.....T.....N.T...H..... : 259
TV014c6.3      : KE.....L...P.NNGNPS--EG.S--EK.R.....T.....S.....T.....N.T...H..... : 258
TV008c4.3      : KE.....L...P.N.-----KD.K.--T.CR.....T.....S.E.....N.T..... : 256
TV008c4.4      : KE.....L...P.N.-----KD.K.--T..R.....T.....S.E.....N.T..... : 253
TV019c5        : GV.....L...P.-----NDT...--R..R.....T.....S.....T.....T..... : 258
TV013cB20      : NT.....S...P.N-----R.-----T.....S.....N.T...Q..... : 261
TV013CH17     : NT.....S...P.N-----R.-----T.....S.....N.T...Q..... : 261
TV006cE9       : -KKKQV.ALFYR.DIVQFD---ND.S----R.....T.A.....T.....N.T...S..G..... : 254
LinkTV006cE9   : -KKKQV.ALFYR.DIVQFD---ND.S----R.....T.A.....T.....N.T...S..G..... : 254
TV006c9.1      : -KKKQV.ALFYK.DIIQLD---ND.S.DNSG.R.....T.A.....T.....N.T...S..... : 248
TV006c9.2      : .V.....L...QFD-----S.----R.....T.A.....T.....N.T...S..... : 250
TV004cC300     : .V.....L...P.N.-NNS.S--F.E--R.....T.A.....T.....N.T...S..... : 250
LinkTV004cC300 : .V.....L...P.N.-NNS.S--F.E--R.....T.A.....T.....N.T...S..... : 250
TV003cE260     : RVS.-----FFY.L.IVPL.ENN.E--QK.R..S..T.....T.....T.....N.T...H..... : 241
TV012c2.1     : .E.....L...P.E--E-----S.---K.R.....T.A.....S.....N.T..... : 243
TV012c2.2     : .E.....L...P.E--E-----S.---K.R.....T.A.....S.....N.T..... : 243
TV018cF1027   : IE.....LG..P.-----ENKL.S---R.....T.A.....S.....N.T..... : 255
LinkTV018cF1027 : IE.....LG..P.-----ENKL.S---R.....T.A.....S.....N.T..... : 255
TV001c8.2     : KE.....L...P.N..SD-----F.---R.....T.....S.....N.T...Y..... : 260
TV001c8.5     : KE.....L...P.N..SN-----F.---R.....T.....S.....D.....N.T...Y..... : 260
TV002c12.1    : KEN.....L...P.NNR.NG.INN-----R.....T.A.....S.....P.....K.N.I..... : 246
TV010cD7      : KVH.....L...P.-----NGS.S---R.....T.....S.....N.T...Q..... : 242
    
```



B.FR.HXB2 : VVSTQ^{AAA}LL^{AAA}NGSLAEEEVVIRSVN^{AAA}FTD^{AAA}NAKTIIVQLN^{AAA}TSVEIN^{AAA}CTRPN^{AAA}NNT^{AAA}TRKRIRI^{AAA}QRGPGRA^{Crown}FTIGKIGNMRQAHCNISRAKWN^{AAAA}N^{AAA}TLKQIASK : 348

C.ZA.97ZA012 : VVSTQ^{AAA}LL^{AAA}NGSLAEKEIIRSEN^{AAA}LT^{AAA}DNVKTIIVHLN^{AAA}KSV^{AAA}EIVCTRPN^{AAA}NNT^{AAA}TRKSMRIG^{AAA}PGQ-TFYATGDIIGDIRQAYCNISGSKWNETLKR^{AAA}VKEK : 345

C.IN.95IN21068 :GG.....QP.....H...ED.....QN.SK. : 353

C.BW.96BW0502 :KG.....A.I..Q..P.K..V.....E.....I.NK.E..S..QG.SK. : 345

C.BR.92BR025 :E.....K.....E..N.....-A.....E.....H...R.A..K..QE.GK. : 345

C.ET.ETH2220 :G.T..F.....A.I..Q..E...T...S...E.....-.....H...EE..K..Q..... : 344

TV007cB104 :K.....AQ.....E...R.P..H.....-A.....H...E...K..Q..SK. : 347

TV007cB105 :K.....AQ.....E...R.P..H.....-A.....H...E...K..Q..SK. : 347

TV014c6.4 :E.....A.I..Q.....G.....H...ED.....QW.GK. : 353

TV014c6.3 :E.....A.I..Q.....G.....H...ED.....QW.GK. : 352

TV008c4.3 :A.....E..R.....-N.....KDD..K..Q..A.. : 350

TV008c4.4 :EG.....A.....E..R.....-N.....KDD..K..Q..A.. : 347

TV019c5 :K.....A.....E..A.....G.....-A...A...K...EGE..R..Q..GR. : 352

TV013cB20 :ED.R.....E.....Q..K.....-A..TN-.....H...RAE..N..A.... : 354

TV013CH17 :ED.R.....S.E.....Q..K.....-A..TN-.....H...RAE..N..A.... : 354

TV006cE9 :GD.....A.I..E.....G...Q.....-.....H...AG.....SK. : 348

LinkTV006cE9 :GD.....A.I..E.....G...Q.....-.....H...AG.....SK. : 348

TV006c9.1 :GD.....A.I..E.....G...Q.....-.....H...AG.....SK. : 342

TV006c9.2 :GD.....A.I..E.....G...Q.....-.....H...AG.....SK. : 344

TV004cC300 :E.....R.....-E.....H.T..SQN..R..Q..S.. : 344

LinkTV004cC300 :E.....R.....-E.....H.T..SQN..R..Q..S.. : 344

TV003cE260 :E.....A.I..E.....-NG.....H...EE...K..QQ.GK. : 335

TV012c2.1 :E.....A.I..E...T.....-.....H...EKQ.....Y..S.. : 337

TV012c2.2 :E.....A.I..E...T.....-.....H...EKQ.....Y..S.. : 337

TV018cF1027 :E..K.....E.....R...-.....E.....K..EDQ..K....S.. : 349

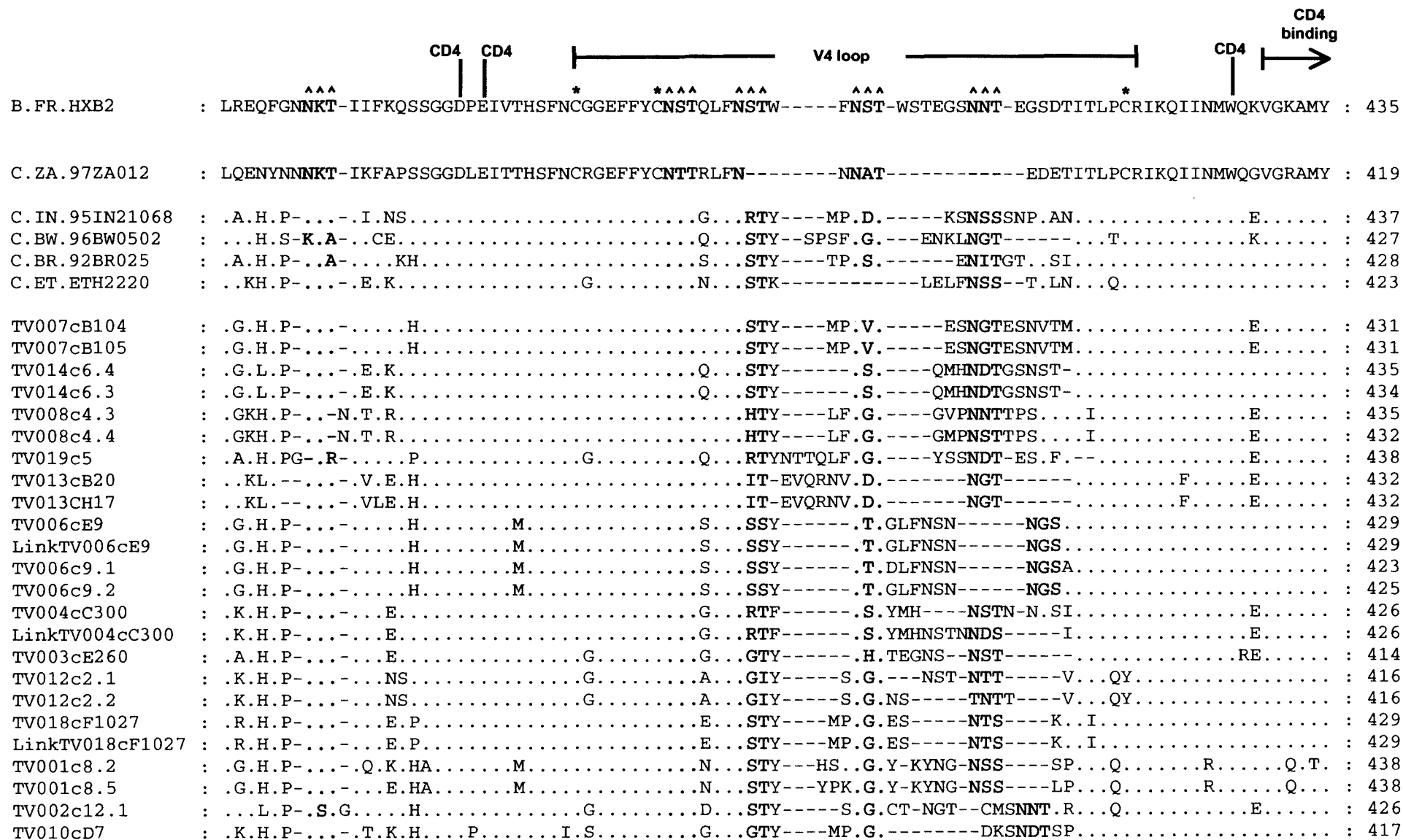
LinkTV018cF1027 :E..K..S.....E.....R...-.....E.....K..EDQ..K....S.. : 349

TV001c8.2 :EG.....E.T.....E..N.....-A..N.....H...TD...K..QQ.MK. : 354

TV001c8.5 :EG.....E.T.....E..N.....-A..N.....H...TD...K..QQ.MK. : 354

TV002c12.1 :E.....E...K...G.....-A.....H...KNE..T..Q..S.. : 340

TV010cD7 :G.....A.....F.Q.....-.....H...KE...N..QE.SK. : 336





```

B.FR.HXB2      : APPISGQIRCSSNIT*GAAALLLRDGG-----NSNAAANES-EIFRPGGGMDRDNWRSELYKYKVVKIEPLGVAPTAKRRRVVQREKRAVGIGALFLGFLG : 524

C.ZA.97ZA012  : APPIAGNITCKSNITGAAALLLRDGG-----EDNKAAATPEEIFRPGGGNMKDNWRSELYKYKVIELKPLGIAPTAAAGAKRRVVEREKRAVGIGAVFLGFLG : 508

C.IN.95IN21068 : ...E.K.....EDKNNT.T...T.....T..... : 531
C.BW.96BW0502  : .....E.....T.....KTGP.D.....N.....E.....C..... : 517
C.BR.92BR025   : ...E.I.....T.....TGMHDT.....E.....K..... : 517
C.ET.ETH2220   : ...E.I.M.....T..A----K.PHS.K.....E.....KP.....A-..... : 512

TV007cB104     : .....T.....T.....T.....N.....T.....A.....A.....L..... : 520
TV007cB105     : .....T.....T.....T.....N.....T.....A.....A.....L..... : 520
TV014c6.4      : .....NTNDT--G.....E.....L..... : 526
TV014c6.3      : .....NTNDT--G.-G.....E.....G.....L..... : 525
TV008c4.3      : .....T.....NSGK.T.....A..... : 526
TV008c4.4      : ...T.....T.....NSG.T.....A..... : 523
TV019c5        : ...K.....N.....EDNNT-E.D.-T..... : 530
TV013cB20      : .....T.....NIT--E.-TS.....E..... : 521
TV013CH17      : .....T.....NIT--E.-T.....E..... : 521
TV006cE9       : .....N.S.....NIT--S.-E.....N.....T..... : 518
LinkTV006cE9   : .....N.S.....NIT--S.-E.....N.....T..... : 518
TV006c9.1      : .....N.S.....NTT--S.-E.....S.....N.....T..... : 512
TV006c9.2      : .....N.S.....NTT--S.-T.E.....N.....T..... : 514
TV004cC300     : .....ENGT--N.....NK..... : 516
LinkTV004cC300 : .....ENGT--N.....NK..... : 516
TV003cE260     : .....ESNDS--NNI.....R..... : 506
TV012c2.1      : .....T.....EGNGTN--D.....A.....Q.....K.....A-..... : 506
TV012c2.2      : .....T.....EGNGTN--D.....A.....Q.....K.....A-..... : 506
TV018cF1027    : .....Q.....T.....EES---KSNGT.....A.....E..... : 520
LinkTV018cF1027 : .....Q.....T.....EES---KSNGT.....V.....E..... : 520
TV001c8.2      : .....T.....FNNT--NNT.T.....K..... : 528
TV001c8.5      : .....T.....F--NNTN-.D..T.....K.....K..... : 530
TV002c12.1     : .....T.....D--NNT.T----.T.....A..... : 515
TV010cD7       : .....T.....E-----...N.T.....T..... : 506
    
```



Immunodominant region

B. FR. HXB2	:	AAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEAQOHLQLTVWGIKQLQARILAVERYLKDQQLLGIWGC SGKLICTTAVPW [*] [*] ^{AAA} ^{AAA} NASWS-NKS	:	618
C. ZA. 97ZA012	:	AAGSTMGAASLTLTVQARQLLSIVQQQSNLLRAIEAQOHLQLTVWGIKQLQTRVLAIERYLKDQQLLGIWGC SGKLICTTNVPWNSSWS-NKS	:	602
C. IN. 95IN21068	:G.....	:	625
C. BW. 96BW0502	:L..G.....N.....	:	611
C. BR. 92BR025	:V..G.....	:	611
C. ET. ETH2220	:G.....H.....	:	606
TV007cB104	:G.....	:	614
TV007cB105	:G.....	:	614
TV014c6.4	:G.....Y.....A.....Q.....	:	620
TV014c6.3	:G.....A.....Q.....	:	619
TV008c4.3	:G.....	:	620
TV008c4.4	:G.....	:	617
TV019c5	:G.....	:	624
TV013cB20	:G.....A.....	:	615
TV013CH17	:G.....A.....	:	615
TV006cE9	:A.....G.....A.....	:	612
LinkTV006cE9	:A.....G.....A.....	:	612
TV006c9.1	:A.....G.....T.....A.....	:	606
TV006c9.2	:A..A..R..G.....A.....	:	608
TV004cC300	:A..A..G.....P.A.....	:	610
LinkTV004cC300	:A..A..G.....P.A.....	:	610
TV003cE260	:G.....	:	600
TV012c2.1	:G.....	:	600
TV012c2.2	:G.....	:	600
TV018cF1027	:PF.G.....	:	614
LinkTV018cF1027	:G.....	:	614
TV001c8.2	:G.....A.....	:	622
TV001c8.5	:G.....A.....	:	624
TV002c12.1	:G.....A.....Q.....L.....	:	609
TV010cD7	:	M.....G.....K.....A.....K.....A.....N.....	:	601



B.FR.HXB2	: LEQIWN ^{AAA} HTTWMEWDREIN ^{AAA} NYTSLIHS ^{AAA} SLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKLFIMIVGGLVGLRIVFAVLSIVNRVRQGYS : 713
C.ZA.97ZA012	: QTDIWN ^{AAA} HTTWMEWDREIS ^{AAA} NYTDTIYRLL ^{AAA} EDSQTQQEKNEKDLLALDSWKNLWSWFDISNWLWYIKIFIMIVGGLIGLRIIFAVLSIVNRVRQGYS : 697
C.IN.95IN21068	: .KE.....N.....E.N..E.....N..K..... : 720
C.BW.96BW0502	: HDE.....N.....E.N.....Q..N..S..... : 706
C.BR.92BR025	: .E.....N.....Q..K.Q.....G.....K..... : 706
C.ET.ETH2220	: .EE.....I..N..V..N..D.....K.E..N..... : 701
TV007cB104	:E.N..N.....Q..... : 709
TV007cB105	:E.N..N.....Q..... : 709
TV014c6.4	: L.....N..T..Q..K..I..Q.....K.Q..N..S..Q..... : 715
TV014c6.3	: L.....N..T..Q..K..I..Q.....K.Q..N..S..Q..... : 714
TV008c4.3	: LG.....N.....E.....Q.....K.Q..... : 715
TV008c4.4	: L.....N..S.....E.....Q.....K.Q..... : 712
TV019c5	: .G.....A.N..Q.....SN.Q..... : 719
TV013cB20	:K.....GI.....Q.....K.....LG.....K..... : 710
TV013CH17	:GI.....Q.....K.....LG.....K..... : 710
TV006cE9	: .K..E.....I..N..V..I..Q.....N..K..... : 707
LinkTV006cE9	: .K..E.....I..N..V..I..Q.....N..K..... : 707
TV006c9.1	: .K..E.....I..N..V..I..Q.K.....I.....S..... : 701
TV006c9.2	: .K..E.....I..N..V..I..Q.....S.N..K..... : 703
TV004cC300	: .E..G.....S.....N..E..R.....E..NF.....K..... : 705
LinkTV004cC300	: .E..G.....S.....N..E..R.....E..NF.....K..... : 705
TV003cE260	: V.....Q.....N..N..K..... : 695
TV012c2.1	: .DY..G.....N.....G.A.N..E..E..E..R.G..N..S..... : 695
TV012c2.2	: .DY..G.....N.....G.A.N..E..E..E..R.G..N..S..... : 695
TV018cF1027	:E.....N..Q.....N..K..... : 709
LinkTV018cF1027	:E.....N..Q.....N..K..... : 709
TV001c8.2	: .K.....GL.N.....N.....E.K.N..N.....P..... : 717
TV001c8.5	: .A.....N..E.....N.....E.K.N..N..... : 719
TV002c12.1	:R.N..N..S..... : 704
TV010cD7	: ..E.....E.N.....N..N..S..K.....A..... : 696



*

B.FR.HXB2 : PLSFQTHLPTPRGPDRPEGIEEEGGGERDRDRSIRLVNGSLALIWDDLRSCLFSYHRLRDLLIVTRIVELLGR-----RGWEALKYWNNLLQ : 801

C.ZA.97ZA012 : PLSFQTLTPNPREDRLGRIIEEEGGEQDRDRSIRLVNGFLALAWDDLRSCLCFIYHQLRDFILLTARAVELLGRSSLRGLQRGWEALKYLGNLVQ : 792

C.IN.95IN21068 :GGP.....S.....F.....N....S..R.....A..V.....R.....S... : 815

C.BW.96BW0502 : ..P.....P.....G.....S.....S..R.....A..V.....Q-----S... : 794

C.BR.92BR025 :GP...G.....S.....S..R...L..A.....I.....G... : 801

C.ET.ETH2220 :I.H..GP...G.....G.....F.....S..R...L..A..T.....T.....S... : 796

TV007cB104 : ..LS.....GP.....S...S.....S..R...L..VV.....G..- : 803

TV007cB105 : ..LS.....GP.....S...S.....S..R...L..VV.....G..- : 803

TV014c6.4 :GP.....E.....S..FS.....N....S..R.....VT.V.....T.....S.G. : 803

TV014c6.3 :GP.....E.....S..FS.....N....S..R.....AT.V.....T.....S.G. : 802

TV008c4.3 :I...GP..R.....S.....S..L.....GV..V.....K.....S... : 810

TV008c4.4 :I...GP..R.....S.....C.....S..L.....GV..V.....K.....S... : 807

TV019c5 : ...L.....GP..RG.....S.....S.....V.....S... : 807

TV013cB20 :I...P..RG.....S.....S.....V.....Q..... : 805

TV013CH17 :I...P..RG.....S.....R.....S.....V.....Q..... : 805

TV006cE9 :I.....R.....S..R.....A..A..H.....I.....S.A. : 802

LinkTV006cE9 :I.....R.....S..R.....A..A..H.....I.....S.A. : 802

TV006c9.1 :A.....Q.....S..R.....A..A..HN.....I.....S.A. : 796

TV006c9.2 :I.....Q.....S..R.....A..A..H.....I.....S.A. : 798

TV004cC300 :S.GGP.....S..R.....S... : 800

LinkTV004cC300 :S.GGP.....S..R.....S... : 800

TV003cE260 :I.S..GP.....S..R.....AT.....S... : 790

TV012c2.1 :AQ...GP...T.....S.....E..N.....R.....I.....S... : 790

TV012c2.2 :AQ...GP...T.....S.....E..N.....R.....I.....S... : 790

TV018cF1027 : ...L.....GP.....V.....S.....L..V.....N.....T.....L : 804

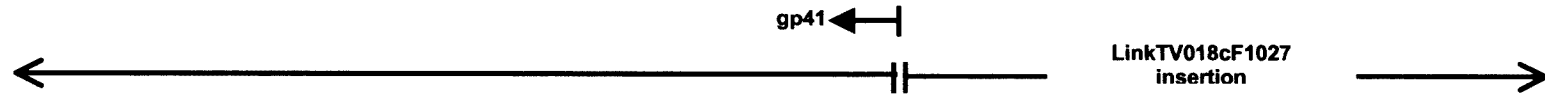
LinkTV018cF1027 : ...L.....GP.....V.....S.....L..V.....N.....T.....L : 804

TV001c8.2 :S..G...G.....S..S.....N....S..R.....AV.....H.....I.....S... : 812

TV001c8.5 :S..G...G.....S..S.....S..R.....AV.....H.....I.....S... : 814

TV002c12.1 : ...L..I...GP...G.....SS.....S..T.....C..R.....VV.....H.....GT.....S... : 799

TV010cD7 : ...L..I...GP..P.....S.....C..R.....A..V.....S... : 791



B.FR.HXB2	: YWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL-----	: 856
C.ZA.97ZA012	: YWGLELKKSAISLLDPIAITVAEGTDRIIEVVQRICRAIRNIPRRIRQGFEEALP-----	: 847
C.IN.95IN21068	:N...R...A.....Q-----	: 870
C.BW.96BW0502	:A.....F.....Q-----	: 849
C.BR.92BR025	: ..S.....F...A.....G.W...C.....Q-----	: 856
C.ET.ETH2220	:N...T...V.G...F.....W..FC.....L...Q-----	: 851
TV007cB104	: -----	: -
TV007cB105	: -----	: -
TV014c6.4	:A..A.V.....F.....T.....Q-----	: 858
TV014c6.3	:A..A.....F.....T...H...-----	: 856
TV008c4.3	:E.....G.....F.....H...Q-----	: 865
TV008c4.4	:G.....F.....Q-----	: 862
TV019c5	:N.....A.....LG.G...C.....Q-----	: 862
TV013cB20	:A.....L-----	: 860
TV013CH17	:A.....L-----	: 860
TV006cE9	:IT..A.....W.....T.....T..L-----	: 857
LinkTV006cE9	: [R].....IT..A.....W.....T.....T..L-----	: 857
TV006c9.1	:A.....W.....T.....Q-----	: 851
TV006c9.2	:IT..A.....W.....T.....Q-----	: 853
TV004cC300	:T.....F...LH.....L-----	: 855
LinkTV004cC300	:T.....F...LH.....L-----	: 855
TV003cE260	:V.....T.....T..L-----	: 845
TV012c2.1	:A.....F..G.G...Y.....Q-----	: 845
TV012c2.2	:A.....F..G.G...Y.....Q-----	: 845
TV018cF1027	:T..V.....F.A.C.....R.....L-----	: 859
LinkTV018cF1027	:T..V.....F.A.C.....R..SSFAITRASLESHGYRIHYALELADQPRCLLVATICCLPLPRAF	: 899
TV001c8.2	:L.....L-----	: 867
TV001c8.5	:P...A.....L.....L-----	: 869
TV002c12.1	:N...A.....F..N..G.....Q-----	: 854
TV010cD7	:AT.....G.G...Y.....Q-----	: 846



B.FR.HXB2	:	-----	:	-
C.ZA.97ZA012	:	-----	:	-
C.IN.95IN21068	:	-----	:	-
C.BW.96BW0502	:	-----	:	-
C.BR.92BR025	:	-----	:	-
C.ET.ETH2220	:	-----	:	-
TV007cB104	:	-----	:	-
TV007cB105	:	-----	:	-
TV014c6.4	:	-----	:	-
TV014c6.3	:	-----	:	-
TV008c4.3	:	-----	:	-
TV008c4.4	:	-----	:	-
TV019c5	:	-----	:	-
TV013cB20	:	-----	:	-
TV013CH17	:	-----	:	-
TV006cE9	:	-----	:	-
LinkTV006cE9	:	-----	:	-
TV006c9.1	:	-----	:	-
TV006c9.2	:	-----	:	-
TV004cC300	:	-----	:	-
LinkTV004cC300	:	-----	:	-
TV003cE260	:	-----	:	-
TV012c2.1	:	-----	:	-
TV012c2.2	:	-----	:	-
TV018cF1027	:	-----	:	-
LinkTV018cF1027	:	LDPGSPLPLSFPNKGNCIALSEVSF	:	924
TV001c8.2	:	-----	:	-
TV001c8.5	:	-----	:	-
TV002c12.1	:	-----	:	-
TV010cD7	:	-----	:	-

APPENDIX C

This section contains the article that resulted from this study.

Sequence Note

Genetic Analysis of the Complete *gag* and *env* Genes of HIV Type 1 Subtype C Primary Isolates from South Africa

SUSAN ENGELBRECHT,¹ TANIA DE VILLIERS,¹ CANDICE C. SAMPSON,¹ JAN ZUR MEGEDE,²
SUSAN W. BARNETT,² and ESTRELITA JANSE VAN RENSBURG¹

ABSTRACT

South Africa has one of the fastest growing HIV-1 epidemics, with an estimated 4.7 million people infected. To better understand the genetic diversity of this epidemic and its potential impact on vaccine development, we have cloned and sequenced the complete *gag* and *env* genes of 13 primary virus isolates. Phylogenetic analysis of our sequences and 69 complete *env* genes from the Los Alamos and GenBank databases revealed multiple subclusters within subtype C. The V3 loop region was relatively conserved in all our strains when compared with other subtypes, but the region immediately downstream was highly variable. No intersubtype recombinant forms were observed when comparing the *gag* and *env* sequences. Characterization of the complete *gag* and *env* genes enabled us to select specific strains for further vaccine development.

MORE THAN 36.1 MILLION PEOPLE are infected with HIV, with 25.3 million living in sub-Saharan Africa.¹ Currently South Africa has one of the fastest growing HIV-1 epidemics, with an estimated 4.7 million people infected out of a total population of 40.6 million. The prevalence rate varies from region to region, with the highest rate in KwaZulu-Natal (36.2%) and the lowest in the Western Cape Province (8.7%).² A safe and effective HIV vaccine, which can be used for both therapeutic and preventative immunization, is required to stop the pandemic. The Env glycoprotein contains the major functional domains for the two most important functions in the viral life cycle: viral attachment and fusion to target cells. The diversity of HIV-1 envelope proteins and its complex interaction with the immune system poses one of the major obstacles to HIV vaccine development. The more conserved *gag* gene encodes a major structural protein, which contains several cytotoxic T cell (CTL) epitopes. Gag-specific CTL responses are important in controlling viral load during acute infection and asymptomatic stages of the infection.

Phylogenetic analysis has shown that HIV-1 can be classified into three major groups: M, N, and O. The M group viruses

are responsible for the majority of HIV-1 infections and can be subdivided into subtypes A to K and circulating recombinant forms (CRFs), in which the *env* gene exhibits an average nucleotide divergence of 25%.³ Subtype C viruses constitute 56% of all circulating viruses and are the most commonly transmitted subtype worldwide.⁴ Although many partial *env* and *gag* gene sequences from South Africa have been published,⁵⁻⁸ no complete *env* and *gag* sequences have been described, other than in the publication of one nearly full-length molecular clone.⁹ The objective of this study was to amplify, clone, and sequence the complete *env* and *gag* genes of 13 HIV-1 subtype C primary isolates to generate a set of reference reagents for studies of virus-induced pathogenesis and for vaccine development.

Samples from HIV-1-infected patients were obtained during 1998 and 1999 from Tygerberg Hospital in the Western Cape. Viruses were isolated from patient peripheral blood lymphocytes (PBMCs) and cocultured with cord blood lymphocytes.^{5,6} All the isolates were of the non-syncytium-inducing (NSI) phenotype and used CCR5 as their principal co-receptor.¹⁰

Polymerase chain reaction (PCR) amplification of the *gag*

¹Department of Medical Virology, University of Stellenbosch and Tygerberg Hospital, Tygerberg 7505, South Africa.

²Chiron Corporation, Emeryville, California 94608.

and *env* gene fragments was done by using proviral DNA as template with the following primer pairs:

GagF: GCTAGAAGGTCTAGAATGGGTGCGAGAGCG
 GagR: AGTTGCCCCGAATTCTTATTGTGACGAGG¹¹
 Env-st: GAAAGAGCAGAAGACAGTGGCAA
 Env-end: CTTTTTGACCACTTGCCACCCAT

using the Expand high-fidelity PCR system (Roche Molecular Biochemicals, Mannheim, Germany). The *gag* and *env* PCR products were cloned either into pCR 3.1, using the Eukaryotic TA cloning kit (Invitrogen, Carlsbad, CA), or into pGEM-T-Easy. The pCMVlink (Chiron, Emeryville, CA) or pcDNA3.1(-) (Invitrogen) vector was used for directional cloning. Sequencing reactions were done with the ABI Prism BigDye Terminator Cycle Sequencing Ready reaction kit (Applied Biosystems, Foster City, CA) and run on an ABI 310 genetic analyzer (Applied Biosystems).

Sequence assembly was performed with either DNAMAN version 4.0 (Lynnon BioSoft, Vaudreuil-Dorion, Quebec, Canada) or Auto Assembler (Applied Biosystems). Nucleotide sequences were aligned with reference sequences from the Los Alamos HIV sequence database and phylogenetic trees were constructed as previously described.^{5,6,10} Sequence analyses revealed that all except one clone of isolate TV014 had intact open reading frames, suggesting the presence of functional *env* genes. Both clones of isolate TV007 had a premature truncation of the gp41 region. All except one clone of TV006 had intact *gag* open reading frames.

In the trees constructed from the *gag* and *env* nucleic acid sequences (Fig. 1A and B), major clusters can be observed containing sequences representative of the different subtypes and CRFs (A1, B, C, D, F1, F2, G, K, and H and CRF01_AE, CRF02_AG, CRF03_AB, CRF05_DF, and CRF04_CPX, and CRF06_CPX). Bootstrap analysis strongly supported these branching orders. All 21 *env* sequences from the 13 patients clustered with the subtype C reference strains (Fig. 1A). Intersubtype nucleotide distances for our isolates and all other subtypes varied between 21 and 30% (data not shown). The clone sequences for isolates TV001, TV002, TV003, TV004, TV007, TV008, TV010, TV012, TV013, TV014, TV018, and TV019 varied between 1–4 and 4–6% for TV006.

The *env* and *gag* sequences of the isolates clustered with the subtype C reference strains, strongly suggesting that our isolates are subtype C strains and not recombinant forms. This is supported by a previous study in which the accessory genes of these isolates were also characterized as being subtype C.¹⁰ The diversity among our isolates was comparable to the diversity

seen among the C subtypes as a whole. The subtype C *gag* sequences varied between 6 and 13%, while the intersubtype variation within the M group ranged between 8 and 19%. The highest variation within *gag* was seen in the p17 matrix protein (8–25%).

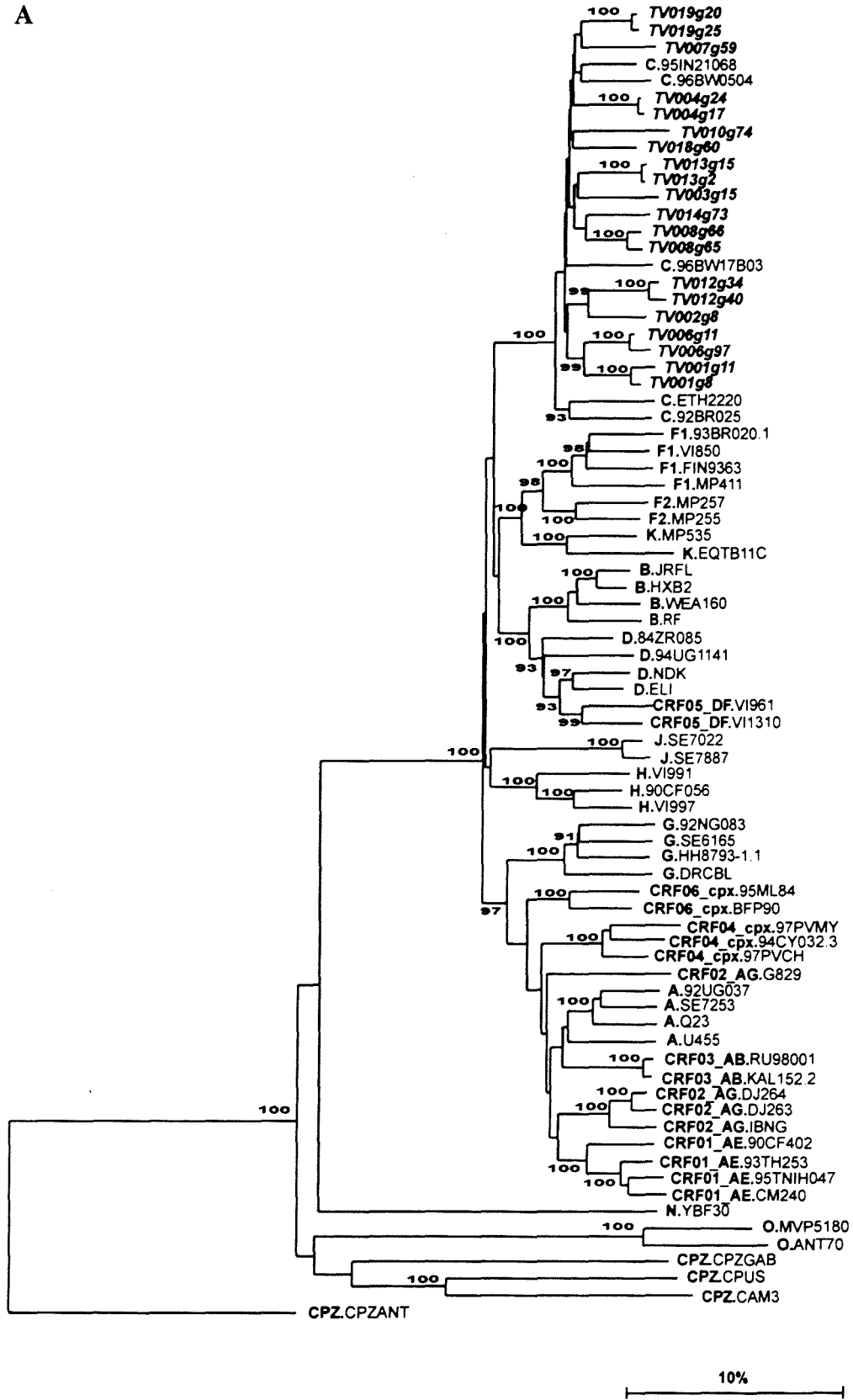
Phylogenetic analysis of 90 complete subtype C *env* sequences from South Africa, Botswana, India, China, Somalia, Zambia, Burundi, Ethiopia, Djibouti, Uganda, and Brazil indicated several subclusters within subtype C (Fig. 2). A distinct starburst effect was observed with a high degree of divergence between the South African isolates. Unlike the Indian and Chinese isolates, the South African and Botswana isolates were distributed across several subclusters, suggesting multiple introductions into the region. The *env* sequences of our isolates varied between 10 and 14%, while the levels of nucleotide divergence of the Botswana and Indian subsets were 13–17 and 6–20%, respectively. The calculated average intrasubtype differences between South African subtype C sequences and those of Indian and Botswana origin were in the range of 12–21 and 11–20%, respectively. Results from earlier studies suggested a close similarity between strains from India and southern Africa.^{12,13} Our results suggest that the Indian epidemic may have originated from southern Africa. Geographic clustering was observed for sequences from several adjacent East African countries: Djibouti, Ethiopia, Uganda, Somalia, and Burundi.

The V1 to V5 loops are highly variable regions of the viral genome.³ Compared with sequencing results of other subtypes, the V3 loop of our strains was more conserved, but with a more variable region downstream of the loop (Fig. 3). This was also found by another group.¹⁴ The highly conserved N-linked glycosylation site proximal to the first cysteine residue, observed in all HIV-1 subtypes except subtype C,³ was also absent in the V3 regions of our strains. The tetrapeptide crown sequence of the loop was GPGQ for all our C subtypes. Uncharged or negatively charged amino acid residues at positions 11, 25, and 32 were identified in all our isolates, suggesting an NSI phenotype,¹⁵ which was confirmed subsequently in cell culture experiments using the parental virus isolates (F.K. Treurnicht *et al.*, submitted).

As with the accessory genes,¹⁰ phylogenetic analyses of our *env* and *gag* sequences revealed no recombinant forms, and multiple subclusters within subtype C with a high degree of diversity, the latter indicating multiple independent introductions of strains of this subtype into southern Africa. The molecular characterization of these functional *env* and *gag* genes derived from primary South African HIV-1 isolates will enable us to select candidate subtype C antigens for further vaccine development.

FIG. 1. Phylogenetic tree analysis of South African HIV-1 sequences and Los Alamos HIV database reference sequences. The South African Tygerberg Virology (TV) sequences are indicated in boldface italic lettering. The subtype reference designation is indicated in boldface roman lettering. (A and B) The neighbor-joining method for tree construction is based on the complete (A) *gag* gene (~1500 bp) and (B) *env* gene (~2500 bp). An indication of the degree of sequence dissimilarity is shown on the horizontal axis, and the subtypes are indicated on the vertical axis. The number of bootstrap trees out of 1000 replications supporting a particular phylogenetic group by more than 90% is placed alongside the node considered.

A



B

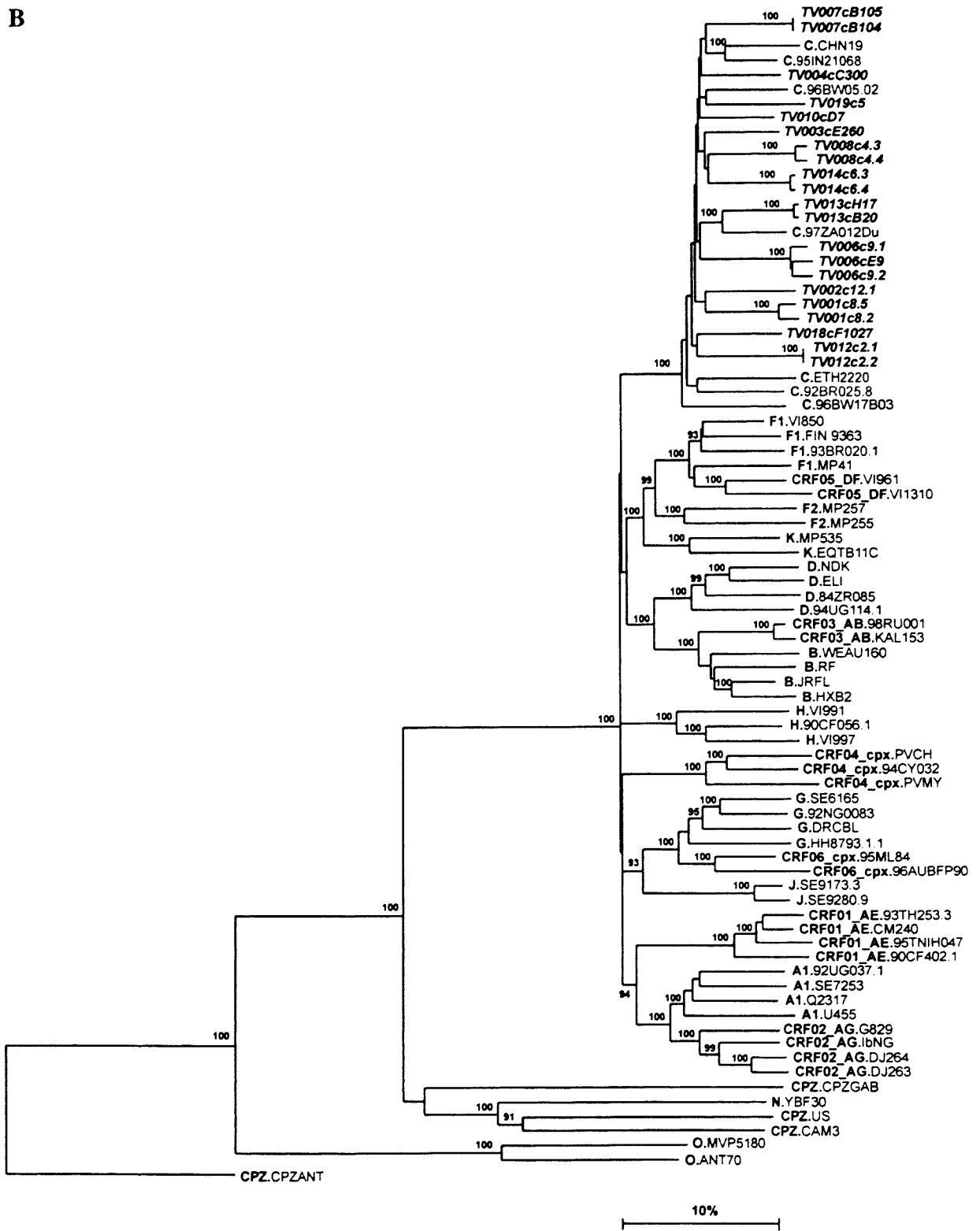


FIG. 1. Continued.

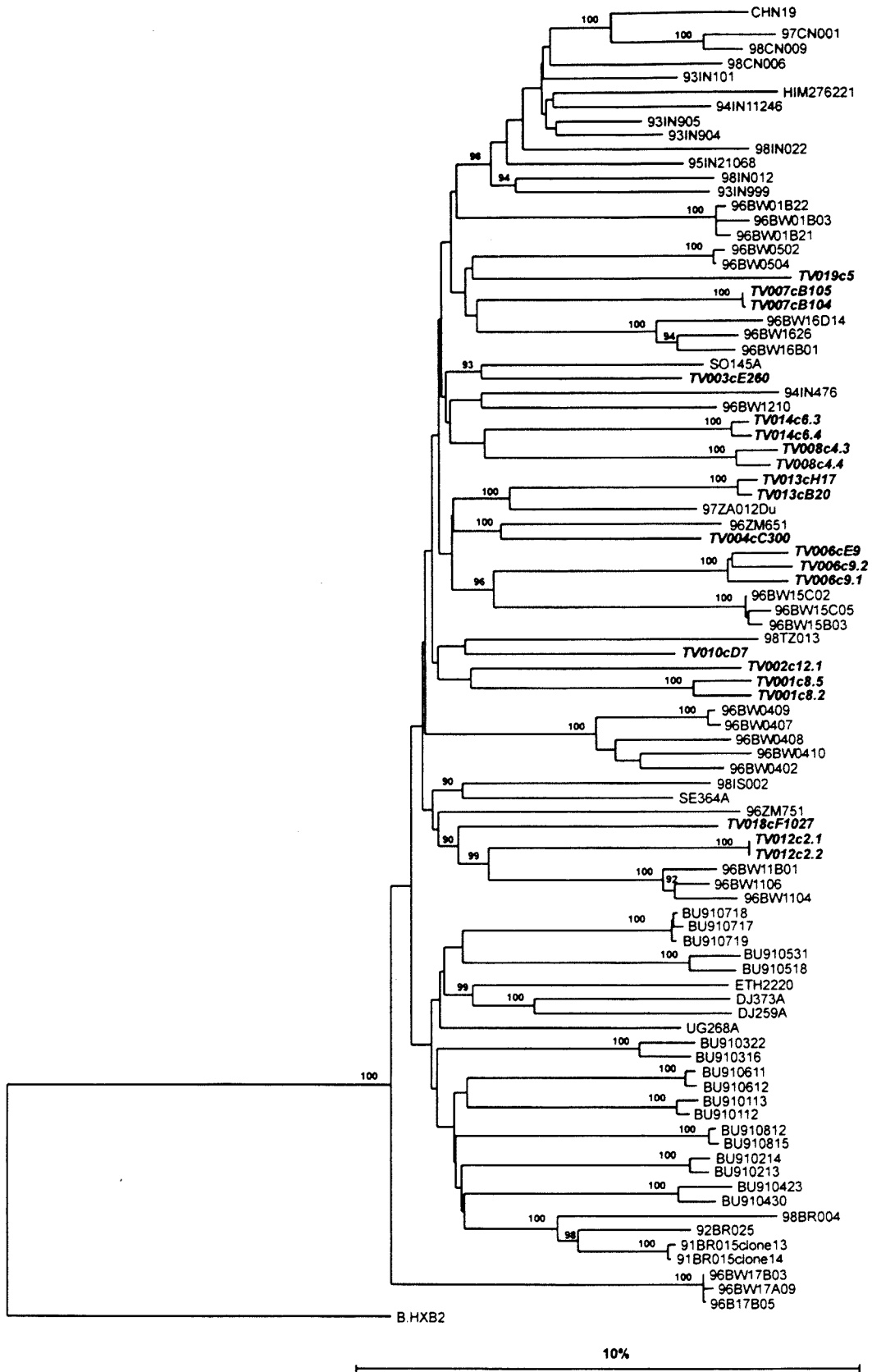


FIG. 2. Phylogenetic tree analyses of all the HIV-1 subtype C complete *env* gene sequences. The neighbor-joining method for tree construction is based on the complete *env* gene (~2500 bp). An indication of the degree of sequence dissimilarity is shown on the horizontal axis. The number of bootstrap trees out of 1000 replications supporting a particular phylogenetic group by more than 90% is placed alongside the node considered.

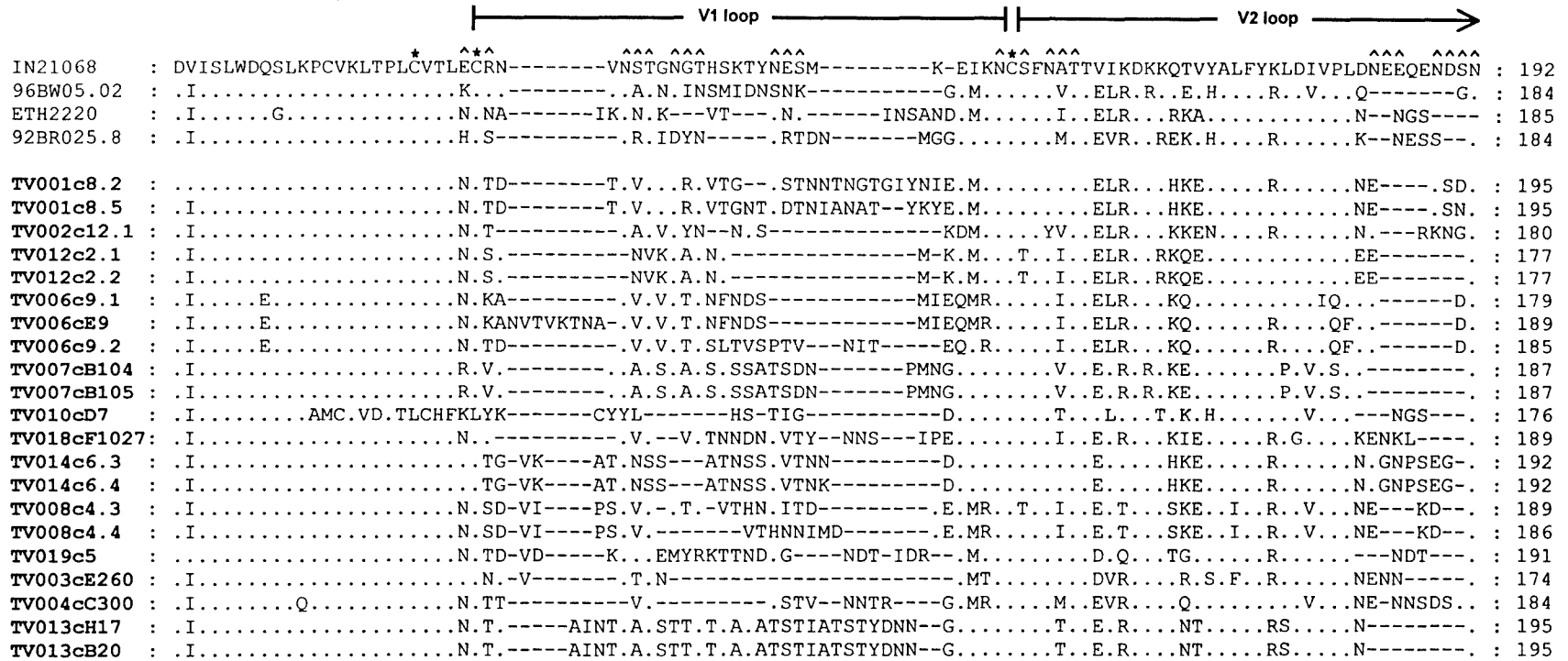


FIG. 3. Continued.

←

```

IN21068 : ^^  SS---GYRLINCNTSALTQACPKVTFDPIPIHYCAPAGYAILKCNKTFNGTGPCNVSTVQCTHGIKPVVSTQLLNGSLAEGGIIIRSENLTNNVKTIIIVHL : 294
96BW05.02 : N.---NE.....I.....S.....T.....Q.....N...S...A.....V.K.E.....A.I...Q. : 286
ETH2220 : ----TD.....TI.....SL.....RD...T.....I...ET...F....A.I...Q. : 285
92BR025.8 : T.---D.....I.....S.....N...I....T.....EE....K...D..... : 286

TV001c8.2 : FT-----TI.....S.....Y.....E.....E.T..... : 295
TV001c8.5 : FT-----TI.....S.....D.....Y.....E.....E.T..... : 295
TV002c12.1 : ----INN.....I.....S.....P.....K...I...D.....EE..... : 281
TV012c2.1 : ..---K.....I.....S.....N.....EE.V....M...A.I.... : 278
TV012c2.2 : ..---K.....I.....S.....N.....EE.V....M...A.I.... : 278
TV006c9.1 : ..DNS-G.....I.....S.....D....Q....A.I.... : 283
TV006cE9 : ..----I.....S.G.....D....Q....A.I.... : 289
TV006c9.2 : ..----I.....S.....D....Q....A.I.... : 285
TV007cB104 : ..---T.....T.....T.....KE...K.K...AQ..... : 288
TV007cB105 : ..---T.....T.....T.....KE...K.K...AQ..... : 288
TV010cD7 : ..---E.....TI.....S.....Q.....I.....E.M.....A....F : 277
TV018cF1027 : ..---E.....I.....S.....N.....EE....K.M..... : 290
TV014c6.3 : ..---EK...T.....T.....S.....T.....EE.....A.I...Q. : 294
TV014c6.4 : ..---EK...T.....T.....S.....T.....EE.....A.I...Q. : 294
TV008c4.3 : K.---TEC.....TV.....S.E.....N...I.....KE.....D.A..... : 291
TV008c4.4 : K.---TE.....TV.....S.E.....N...I.....I.E.....D.A..... : 288
TV019c5 : N.---RE.....TM.....S.....T.....S...N.....KE....K...D.A..... : 293
TV003cE260 : E.---QK...S...TI.....T.....EE.....A.I.... : 276
TV004cC300 : F.---E...M.....S.....KE..... : 285
TV013cH17 : R.---E.I.....TI.....S.....F.....Q.....ED.R...S.E...I.... : 296
TV013cB20 : R.---E.I.....TI.....S.....F.....Q.....ED.R...E...I...V... : 296
    
```

FIG. 3. Continued.

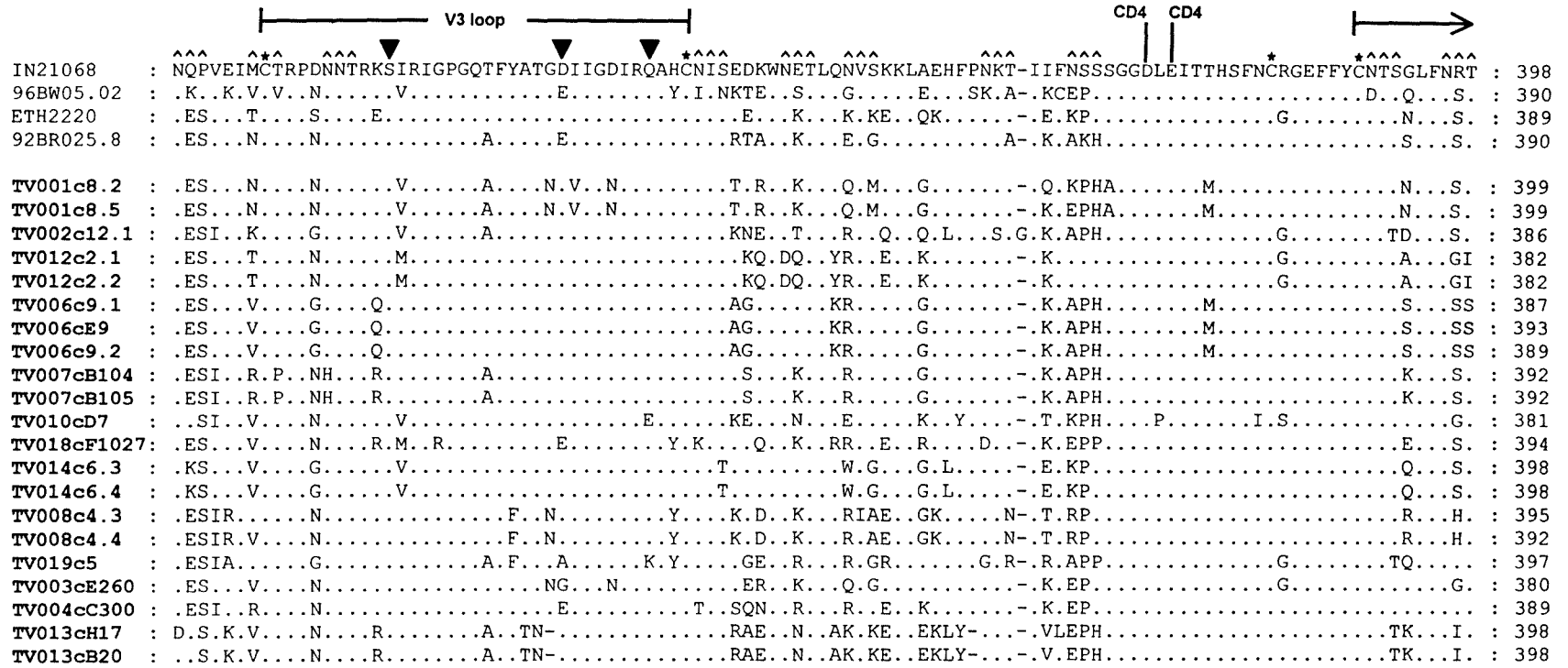


FIG. 3. Continued.



IN21068	:	LYKYKVVVEVKPLGVAPTTAKRRVVEREKRAVGIGAVFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQSNLLRAIEAQQHLLQLTVWGIKQLQTRVLAIERYL	:	594
96BW05.02	:I.....E.....C.....L.....N.....I..V....	:	580
ETH2220	:I.....KP.....A-L..L.....K.....M.....H.....	:	575
92BR025.8	:I...I...K.....V.....M.....	:	580
TV001c8.2	:I...I...K.....Q.....K.....M.....A.....	:	591
TV001c8.5	:I...I...K.....Q.K.....K.....M.....A.....	:	593
TV002c12.1	:I.....A.....M.....A.....	:	578
TV012c2.1	:IQ.....K.....A-L.....V.....M.....	:	569
TV012c2.2	:IQ.....K.....A-L.....V.....M.....	:	569
TV006c9.1	:	S.....I...I...N.....TL..M.....A.....T...M.....A.....	:	575
TV006cE9	:I...I...N.....TL..M.....A.....V.....M.....A.....	:	581
TV006c9.2	:I...I...N.....TL..M.....A..A..R.....K.....M.....A.....	:	577
TV007cB104	:I...I.....A.....A.L..L.....K.....V.....	:	583
TV007cB105	:I...I.....A.....A.L..L.....K.....V.....	:	583
TV010cD7	:I...I.....M.....M.....K.....M.....A.....K.....	:	569
TV018cF1027	:	..R.....I.....E.....PF.....M.....V.....	:	583
TV014c6.3	:I...I...E.....G.....L.....M.....A.....	:	589
TV014c6.4	:I...I...E.....L.....YM.....A.....	:	589
TV008c4.3	:I...I...A.....M.....	:	589
TV008c4.4	:I...I...A.....M.....	:	586
TV019c5	:I...I...G.....M.....	:	593
TV003cE260	:I...I...G.....L..M.....K.....M.....	:	569
TV004cC300	:I.....NK.....A..A..V.....K.....	:	579
TV013cH17	:IR...I...E.....A.....	:	584
TV013cB20	:I...I...E.....A.....	:	584

FIG. 3. Continued.

	Immunodominant region		
IN21068		: KDQQLLGIWGC [*] SGKLICTTAVPWNSSWS-N ^{^^^} R ^{^^^} TQKEIWDNMTWMQWDREINNYTNTIYRLL ^{^^^} EESQ ^{^^^} QEEENEKDLLALDSWKNLWNWFDITKWLWYIKIFIIIVGG	: 698
96BW05.02		:SHD.....D.....K.....Q.....S.N.....M.....	: 684
ETH2220		: R.....-KS.E.....S...DI..N..V.....DK.....K.E.....N..N.....M.....	: 679
92BR025.8		: R.....-S.ED..N.....S.....D.....K..Q.....K.Q...T.G..N.....K....	: 684
TV001c8.2		:R.....-KSE.D.....S...GL..N...D.....K.....E..K.N.....SN.P.....M.....	: 695
TV001c8.5		:R.....-KSEAD.....E.F...D.....K.....E..K.N.....SN.....M.....	: 697
TV002c12.1		: Q.....L.....N.L...-K..SD.....S.....D..S...R.....R.N.....S..N.....M.....	: 682
TV012c2.1		:N...-KS.DY.G.....K.....D.....GDA.....E.E.R.G.....M.S.....MVI..	: 673
TV012c2.2		:N...-KS.DY.G.....K.....D.....GDA.....E.E.R.G.....M.S.....MVI..	: 673
TV006c9.1		:-K.E.D.E.....S...DI..N..V..I..Q.K.....I..S...SS.....R..M.....	: 679
TV006cE9		:-SK.E.D.E.....S...DI..N..V..I..Q.....S.....M.....	: 685
TV006c9.2		:-DK.E.D.E.....S...DI..N..V..I..Q.....S.....S.....M.....	: 681
TV007cB104		:-S.TD.N.....S...D..K.....N.....N..Q..S.N.N.....R..M.....	: 687
TV007cB105		:-S.TD.N.....S...D..K.....N.....N..Q..S.N.N.....R..M.....	: 687
TV010cD7		:L.....N.KS.T.....S..S.....K.....N.....S.....R.....	: 674
TV018cF1027		:L.....-KS.TD.....S...E..K..D...Q.....M.....	: 687
TV014c6.3		: Q.....L.....T...-KSLTD.....T..Q..K..I..Q.....K.Q.....S..Q.....M.....	: 693
TV014c6.4		: Q.....L.....T...-KSLTD.....T..Q..K..I..Q.....K.Q.....S..Q.....M.....	: 693
TV008c4.3		: R.....N...A...-KSLGD.....T..Q.....K.Q...S..N..N.....M.....	: 693
TV008c4.4		:N...T...-KSLSD.....I.....S.....T..Q.....K.Q...S..N..N.....M.....	: 690
TV019c5		:-EGD..N.L.....S..SD.....A.....Q.....SN.Q...S..N..SN.....R..M.....	: 697
TV003cE260		:L.....-KSVD.....S.....D..T..Q.....N.....N.....MM.....	: 673
TV004cC300		:I..P.....-KS.ED.G.....SS.....N.....R..K.....E..NF..S.....M.....	: 683
TV013cH17		:L.....-KS.TD.....S...GI...D..T..Q.....D..T..S.....R..M.....	: 688
TV013cB20		:L.....-KS.TD.....K.S...GI...D..T..Q.....T..S.....R..MM.....	: 688

FIG. 3. Continued.

IN21068	:	LIGLRIFAVISIVNRVRQGYSPLSFQTLTPNPGGPDRLGRIEEEEGEQDKDRSIRLVSGFLALFWDDLRLNL [*] CLFSYHRLRDFILVAARVLELLGRRSLRGLQRG	:	803
96BW05.02	:L.....P.....RE.....RG.....A.....S.....I.....Q-----	:	782
ETH2220	:	V.....L.....I.H.R.....G.....GR.....N...I.....S.....L.I..TV....S..K.....	:	784
92BR025.8	:L.....R.....G.....R.....A.....S.....L.I..AV....S...I...	:	789
TV001c8.2	:L.....S.R.L...G.....R.....S.A.....I.V.AV...HS.....	:	800
TV001c8.5	:L.....S.R.L...G.....R.....S.A...S.....I.V.AV...HS.....	:	802
TV002c12.1	:L.L.....L..I..R...G.....SS.....T.A...S...C.....IVV.AV...HS.....	:	787
TV012c2.1	:L.....AQ..R.....T.....R.....A.E...I.L...V..T..AV....S.....	:	778
TV012c2.2	:L.....AQ..R.....T.....R.....A.E...I.L...V..T..AV....S.....	:	778
TV006c9.1	:S.L.....A..REL.....R.....Q...A...S.....I..KAA...HN.....	:	784
TV006cE9	:L.....I..REL.....R.....R...A...S.....I..AA...HS.....	:	790
TV006c9.2	:L.....I..REL.....R.....Q...A...S.....N..I..AA...HS.....	:	786
TV007cB104	:LS.....R.....S.A...S.....L.IVV.AV...S.....	:	792
TV007cB105	:LS.....R.....S.A...S.....L.IVV.AV...S.....	:	792
TV010cD7	:A.....L..I.D.R...P.....R.....A...S...C.....I...V...S.....	:	779
TV018cF1027	:L..I.....L.....R.....R..V..N...V...S.....Q...L..IV..AV.V..N.....T	:	792
TV014c6.3	:L.....R.....RE.....FS.A.....I.T.V...-----	:	791
TV014c6.4	:L.....R.....RE.....FS.A.....IVT..V...-----	:	791
TV008c4.3	:L.....I.D.R...R.....V.....A...S...L.....GV..V...S...K....	:	798
TV008c4.4	:L.....I.D.R...R.....V.....A...C...L.....GV..V...S...K....	:	795
TV019c5	:L.L..K.....L.....R...RG.....R..V.....A...S.....Q.....IV..AV.I...-----	:	795
TV003cE260	:L.V.....I.S.R.....R..V..N...IA...S.....I.T.AV...S.....	:	778
TV004cC300	:L.....S.....R..V..N...A...S.....T..AV...S.....	:	788
TV013cH17	:LG.L..K.....I..RE..RG.....A...S.R...Q.....IV..AV...QS.....	:	793
TV013cB20	:LG.L..K.....I..RE..RG.....A...S.....Q.....IV..AV...QS.....	:	793

FIG. 3. *Continued.*

gp41 ←

IN21068	:	WEALKYLGSLVQYWGLELKKSAINLLDRIAIAVAEGTDRILELVQRICRAIRNIPRRIRQGFEEAALQ	:	870
96BW05.02	:S...T.....I.FI.....	:	849
ETH2220	:	..T.....NTT..V.G....FI..I..W..FC.....L.....	:	851
92BR025.8	:	..I.....G.....S.....S.F.T.....I.VI.G.W...C.....	:	856
TV001c8.2	:	..I.....S...T...T.....I.....L.....L	:	867
TV001c8.5	:	..I.....SP..T.....I.....L.....L	:	869
TV002c12.1	:	.GT.....T.....FI.NL..G...V.....	:	854
TV012c2.1	:	..I.....VS...SL.....I.FL.G.G...Y.....	:	845
TV012c2.2	:	..I.....VS...SL.....I.FL.G.G...Y.....	:	845
TV006c9.1	:	..I.....A.....S...T.....I..I..W.....T.....	:	851
TV006cE9	:	..I.....A.....S...IT.....I..I..W.....T.....T..L	:	857
TV006c9.2	:	..I.....A.....R..S...IT.....I..I..W.....T.....	:	853
TV007cB104	:G.....	:	803
TV007cB105	:G.....	:	803
TV010cD7	:S...T...T.....I..I.G.G...Y.....	:	846
TV018cF1027	:N..L.....R..S...TT..V.....F.AIC.....R.....L	:	859
TV014c6.3	:	..T.....G.....S..A.....I.FI.....T.....H..-	:	857
TV014c6.4	:	..T.....G.....R..S...A...V.....I.FI.....T.....	:	858
TV008c4.3	:E...S...T...T..G...I.FL.....L.....H.....	:	865
TV008c4.4	:S...T...T..G...I.FL.....L.....	:	862
TV019c5	:T.....I..ILGLG...C.....	:	862
TV003cE260	:VS...TV..V.....I...V.....T.....T..L	:	845
TV004cC300	:TS...T...T.....I..I..F...LH.....L	:	855
TV013cH17	:N.....S...T.....V.II.....L	:	860
TV013cB20	:N.....S...T.....I.II.....L	:	860

FIG. 3. Continued.

ACKNOWLEDGMENTS

This work was supported by grants from the Poliomyelitis Research Foundation (PRF), the South African AIDS Vaccine Initiative (SAAVI), and the Harry Crossley Foundation. Excellent technical assistance from Brenda Robson and Annette Laten is acknowledged. We also thank Thomas J. Scriba and Florette K. Treurnicht for helpful discussions and critical reading of the manuscript.

SEQUENCE DATA

Nucleotide sequences were submitted to GenBank under accession numbers AF391230–AF391270.

REFERENCES

- UNAIDS/WHO Working Group on Global HIV/AIDS and STD Surveillance: Report on the global HIV/AIDS epidemic—December 2000. UNAIDS, Geneva, 2000. (http://www.unaids.org/epidemic_update/report/index.html).
- Department of Health/Directorate Health Systems Research: *Seventh National HIV Survey of Women Attending Antenatal Clinics of the Public Health Services in South Africa*. October/November 2000. Directorate Health Systems Research. Department of Health, Pretoria, South Africa, 2001.
- Kuiken CL, Foley B, Hahn BH, *et al.*: *Human Retroviruses and AIDS 1999: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Theoretical Biology and Biophysics Group. Los Alamos National Laboratory, Los Alamos, New Mexico, 1999.
- Esparza J and Bhamarapravati N: Accelerating the development and future availability of HIV-1 vaccines: Why, when, where and how? *Lancet* 2000;355:2061–2066.
- Engelbrecht S, Laten JD, Smith T-L, and Janse van Rensburg E: Identification of *env* subtypes in fourteen HIV type 1 isolates from South Africa. *AIDS Res Hum Retroviruses* 1995;11:1269–1271.
- Engelbrecht S, Smith T-L, Kasper P, *et al.*: HIV type 1 V3 domain serotyping and genotyping in Gauteng, Mpumalanga, KwaZulu-Natal, and Western Cape Provinces of South Africa. *AIDS Res Hum Retroviruses* 1999;15:325–328.
- Bredell H, Williamson C, Sonnenberg P, Martin DJ, and Morris L: Genetic characterization of HIV type 1 from migrant workers in three South African gold mines. *AIDS Res Hum Retroviruses* 1998;14:677–684.
- Van Harmelen JH, Van der Ryst E, Loubser AS, *et al.*: A predominantly HIV type 1 subtype C-restricted epidemic in South African urban populations. *AIDS Res Hum Retroviruses* 1999;15:395–398.
- Rodenburg CM, Li Y, Trask SA, *et al.*: Near full-length clones and reference sequences for subtype C isolates of HIV type 1 from three different continents. *AIDS Res Human Retroviruses* 2001;17:161–168.
- Scriba TJ, Treurnicht FK, Zeier M, Engelbrecht S, and Janse van Rensburg E: Characterization and phylogenetic analysis of South African HIV-1 subtype C accessory genes. *AIDS Res Human Retroviruses* 2001;17:775–782.
- Qiu J-T, Song R, Dettenhofer M, *et al.*: Evaluation of novel human immunodeficiency virus type 1 *gag* DNA vaccines for protein expression in mammalian cells and induction of immune responses. *J Virol* 1999;73:9145–9152.

12. Dietrich U, Grez M, von Briesen H, *et al.*: HIV-1 strains from India are highly divergent from prototypic African and US/European strains, but are linked to a South African isolate. *AIDS* 1993;7: 23–27.
13. Lole KS, Bollinger RC, Paranjape RS, *et al.*: Full-length HIV type 1 subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 1999;73:152–160.
14. Ping L, Nelson J, Hoffman I, *et al.*: Characterization of V3 sequence heterogeneity in subtype C human immunodeficiency virus type 1 isolates from Malawi: Under representation of X4 variants. *J Virol* 1999;73:6271–6281.
15. Björndal Å, Sönnnerborg A, Tscherning C, Albert J, and Fenyö EM: Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. *AIDS Res Hum Retroviruses* 1999;15:647–653.

Address reprint requests to:
Susan Engelbrecht
Department of Medical Virology
P.O. Box 19063
Tygerberg 7505, South Africa

E-mail: se1@sun.ac.za