THE EXPRESSION, PURIFICATION AND CHARACTERISATION OF RECOMBINANT HIV-1 SUBTYPE C GP120

Katherine Laura Michler

A dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, in fulfillment of the requirements for the degree of Master of Science.

Johannesburg, July 2007

DECLARATION

I, Katherine Laura Michler declare that this dissertation is my own work. It is being submitted for the degree of Master of Science in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.

Katherine Laura Michler

July 2007

This work is dedicated to my parents. Thank you for giving me this opportunity, for encouraging me to take it, and for supporting me in every way possible.

To the rest of my family, thank you for believing in me and supporting me.

To Richard, thank you for holding my hand.

"A philosopher once said "It is necessary for the very existence of science that the same conditions always produce the same results." Well, they do not."

Richard Feynman

The most beautiful thing we can experience is the mysterious. It is the source of all true art and all science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed.

Albert Einstein

PUBLICATIONS AND PRESENTATIONS ARISING FROM THIS STUDY

Bridgette J Connell, **Katherine L Michler**, Alexio Capovilla, Francois Venter, Wendy S Stevens, and Maria Papathanasopoulos (2006). *Emergence of X4 usage among HIV-1 Subtype C: evidence for an evolving epidemic in South Africa*, Conference on Retroviruses and Opportunistic Infections (CROI), Denver CO, USA

Bridgette J Connell, **Katherine L Michler**, Alexio Capovilla, Francois Venter, Wendy S Stevens, and Maria Papathanasopoulos (2006). *Emergence of X4 usage among HIV-1 Subtype C: evidence for an evolving epidemic in South Africa*,Aids, 2008. **22** (7): p. 896-9.

Katherine L Michler, Wendy S. Stevens, Maria A. Papathanasopoulos, Alexio Capovilla (2008). *Molecular Characterization of Recombinant CCR5- and CXCR4-utilizing HIV-1 Subtype C gp120.* Conference on Retroviruses and Opportunistic Infections (CROI), Boston MA, USA

Katherine L Michler, Bridgette J. Connell, Wendy S. Stevens, Maria A. Papathanasopoulos, Alexio Capovilla. *Genetic characterization and comparison of HIV-1 subtype C full-length env that can utilize CCR5 and/or CXCR4*. (In press)

ABSTRACT

HIV-1, the virus that causes AIDS, is spreading at an alarming rate. Subtype C, which accounts for approximately 50% of infections worldwide, and 98% of infections in Southern Africa, is by far the most prevalent form of the virus. Most molecular and biochemical studies have been performed on HIV-1 subtype B isolates and products, however, and there is a relative scarcity of corresponding data on subtype C. It is therefore of crucial importance to study subtype C HIV-1 strains in order to understand their characteristic pathogenic effects and to develop effective treatment strategies. The aim of research in our laboratory is the development of novel treatment strategies, with particular focus on identifying novel Subtype C Env-binding peptide ligands. This necessitates the development of reagents for use in the discovery and testing of these compounds. In line with this, the aim of this project was the production and characterisation of recombinant Subtype C gp120s generated from a recently compiled HIV-1 virus cohort. To this end, the gp160-coding regions of 20 South African Subtype C HIV-1 strains isolated from AIDS patients presenting at the Johannesburg General hospital in 2005 were amplified by PCR and sequenced. The gp160 amplicons were used to amplify and clone the gp120-encoding regions of these isolates. Two clones, pTriEx-FV3 and pTriEx-FV5, originating from CXCR4- and CCR5-utilising strains respectively, were selected for further use. These clones were cotransfected into insect cells together with a baculoviral DNA backbone in order to generate gp120-expressing baculoviruses by homologous recombination. Recombinant baculoviruses were used to infect Sf9 insect cell cultures for expression of recombinant gp120, which was then purified using a combination of lectin affinity chromatography and ion exchange chromatography. In

order to determine the functionality and conformational integrity of the recombinant gp120, the ability of these purified gp120s to bind CD4 and a panel of well-characterised monoclonal antibodies against various epitopes on gp120 (F425 A1g8, 2G12, F425 B4a1, F425 B4e8, 48d, 17b, IgG1 b12, 5F7, 4G10, 9301, ID6, Chessie 13-39.1, 654-30D and 670-30D) was assessed. Gp120 from the CXCR4-using isolate, FV3, appeared to have an intact, functional CD4 binding site as measured by its ability to bind to CD4 and the CD4 binding site antibody 654-30D. It showed low binding to the monoclonal antibody 654-30D, moderate binding to 2G12, Chessie 13-39.1 and 9301, and high binding to ID6, but did not show binding to any of the other antibodies used in the recognition profile. Gp120 from the CCR5-using isolate, FV5, showed low binding to the monoclonal antibodies F425 B4a1 and Chessie 13-39.1, moderate binding to 2G12, and showed good binding to 9301and ID6. FV5 gp120 could not, however, bind to CD4. This is likely to be related to a D368G substitution, a mutation affecting a critical structural determinant of CD4 binding. The lack of CD4-binding activity of this gp120 highlights the importance of Asp368 for CD4 binding and hints at a region vulnerable for therapeutic targeting. Our results also highlight the challenges of developing broadly therapeutic drugs for HIV-1, as well as the importance of investigating the specific biochemical and pathogenic properties associated with subtype C HIV-1.

ACKNOWLEDGEMENTS

To my supervisor, Dr Alexio Capovilla. For giving me the opportunity to do this work and supporting me the entire way through with a seemingly endless supply of ideas, enthusiasm, patience and optimism. For always having an open door, for always challenging me, and for always making it a primary goal to instill in me an appreciation and passion for science.

To my co-supervisor, Dr Maria Papathanasopolous, for being so generous with both her time and expertise in all areas of this project, but in particular for the huge amount of time spent with me during the sequencing phase. Dr Papathanasopolous also performed the phylogenetic analysis and generated the phylogenetic tree for the 20 isolates sequenced.

To Bridgette Connell, who provided the HIV-infected PBMCs and information on the phenotypic studies she performed with regards to the coreceptor usage of the isolates. Also for general help, support and friendship throughout the last couple of years.

To Nichole Cerruti, for always being so willing to discuss work and give me experienced opinions, and for continuous support and help throughout the last year.

To my other present and past colleagues in the HIV pathogenesis research laboratory, Emma Jamieson, Dael Williamson, Catherine Bell and Tlhogi Mncube for vast amounts of help, ideas and moral support.

To Kuben Naidoo and Sonja Lauterbach of the Red Cell Membrane Unit for extensive help and advice.

To Lauren Lopes, for friendship, support, and help with some of the graphics for this dissertation.

To the PRF, NRF, Elevation Biotech, and the University of the Witwatersrand for financial support.

To my boss, Dr Adrian Burstein, for not allowing me to resign, but instead cheerfully putting up with the most part-time part-time employee anyone ever had. I could not have remained a full time student without this support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS USED

AIDS	Acquired Immune Deficiency Syndrome
Anti-human	ECL™ Anti-Human IgG, HRP linked whole antibody (from sheep)
Anti-mouse	ECL™ Anti-Mouse IgG, HRP linked whole antibody (from sheep)
BSA	Bovine Serum Albumin
CCR5	Chemokine (C-C motif) Receptor 5
CD4	Cluster of Differentiation 4
CD4+	CD4 positive
CD4i	CD4-induced
CRF	Circulating recombinant form
CXCR4	Chemokine (C-X-C motif) Receptor 4
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribose Nucleic Acid
dNTPs	Deoxynucleotide triphosphates
ECL	Enhanced Chemiluminescence
ELISA	Enzyme-Linked Immunosorption Assay
Fab	Antigen binding fragment
GAPDH	Glyceraldehyde Phosphate Dehydrogenase
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Virus

HRP	Horseradish Peroxidase
HTLV	Human T-cell Leukemia Virus
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilobase pairs
kDa	Kilodaltons
LB	Luria Broth
MMP	Methyl α-D-mannopyranoside
MOI	Multiplicity of infection
MW	Molecular weight
NIAID	National Institute of Allergy and Infectious Diseases
NIH	National Institute of Health
NRTI	Nucleoside Reverse Transcriptase Inhibitor
NSI	Non syncytium-inducing
PBMC	Peripheral Blood Mononuclear Cells
PBS	Phosphate Buffered Saline
PBS-T	Phosphate Buffered Saline containing 0.05% Tween-20
PIC	Preintegration complex
PCR	Polymerase Chain Reaction
PI	Protease Inhibitor
R5	CCR5
RTI	Reverse Transcriptase Inhibitor
sCD4	Soluble CD4
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SI	Syncytium-Inducing
SIV	Simian Immunodeficiency Virus
T-TBS	Tris-Buffered Saline containing 0.1% Tween-20
$TCID_{50}$	50% tissue culture infectious dose
UV	Ultraviolet
X4	CXCR4

The IUPAC-IUBMB three and one letter codes for amino acids are used

1.Introduction

1.1. AIDS and HIV

Acquired Immune Deficiency Syndrome (AIDS) is a disease characterised by a decline in CD4-positive (CD4+) T-lymphocytes, which play a central role in humoral and cell-mediated immunity[1, 2]. This decline results in patients being unable to mount an immune response and therefore vulnerable to opportunistic infections and malignancies[3].

AIDS was first recognized as a disease in 1981. In May of that year, the possibility of a cellular immune disorder related to a common exposure and predisposing individuals to certain infections was first noted in a United States Centers for Disease Control and Prevention report of a number of cases of *Pneumocystis* pneumonia in Los Angeles[4]. Following this the number of reports of unusual infections and tumours in homosexual men increased, resulting in speculation about the possibility of a sexually transmitted acquired immune disorder[1, 2, 5, 6]. While the disease was first noted in homosexual men, it soon became evident that it was not limited to these individuals. By 1983, AIDS had also been identified in intravenous drug abusers, in female partners of drug-abusing males with AIDS, in individuals receiving blood products, and among Haitians in the United states with no reported history of either drug abuse or homosexuality[7-9].

Soon after AIDS was identified as a disease, the search for a causative agent began. The fact that AIDS was characterised by a decrease in CD4+ T-cells suggested an agent specifically targeting these cells, such as the then recently discovered Human T-cell lymphotropic retrovirus (HTLV) family[10]. Researchers began looking for a retrovirus in

AIDS patients, finally isolating a cytopathic, T-lymphotropic retrovirus postulated to be the cause of AIDS[11-13]. The link between this virus and AIDS became stronger as the virus was detected frequently in patients with AIDS, AIDS-related complex or in healthy people at high risk for AIDS, such as homosexual men, but never in healthy heterosexuals not considered at risk for AIDS[3, 14]. The virus, originally given different names by the various research groups involved, was ultimately, although somewhat controversially, named Human Immunodeficiency Virus (HIV)[15, 16]. Later, this virus became known as HIV-1 as the existence of another, less virulent variant of the virus (HIV-2) was discovered[17, 18].

The latest available figures estimate that in 2006, 39.5 million people worldwide were living with HIV, 2,9 million lost their lives to AIDS, and 4.3 million people became newly infected with HIV-1[19]. In South Africa, it is estimated that 5.5 million people are living with HIV and that 1.2 million children (aged 0-17) have been orphaned by AIDS[19, 20].

1.2. HIV-1 structure, genome and life cycle

1.2.1. Viral structure and genomic organisation

HIV-1 is a lentivirus, more specifically, a T-lymphotropic retrovirus[11]. The mature virion is encapsulated by a lipid bilayer derived from the host cell membrane, which includes a

number of host cell proteins such as the human lymphocyte antigen alpha and beta chains, actin and ubiquitin[21]. Also embedded in this bilayer are the viral surface envelope glycoproteins gp120 and gp41. Beneath the membrane is the viral matrix, with a conical capsid core in the centre containing two copies of the unspliced RNA genome of the virus in a ribonucleoprotein complex. Also found in the core are the viral protease, integrase and reverse transcriptase, as well as the accessory proteins nef, vif and vpr.[22] (Figure 1.1).

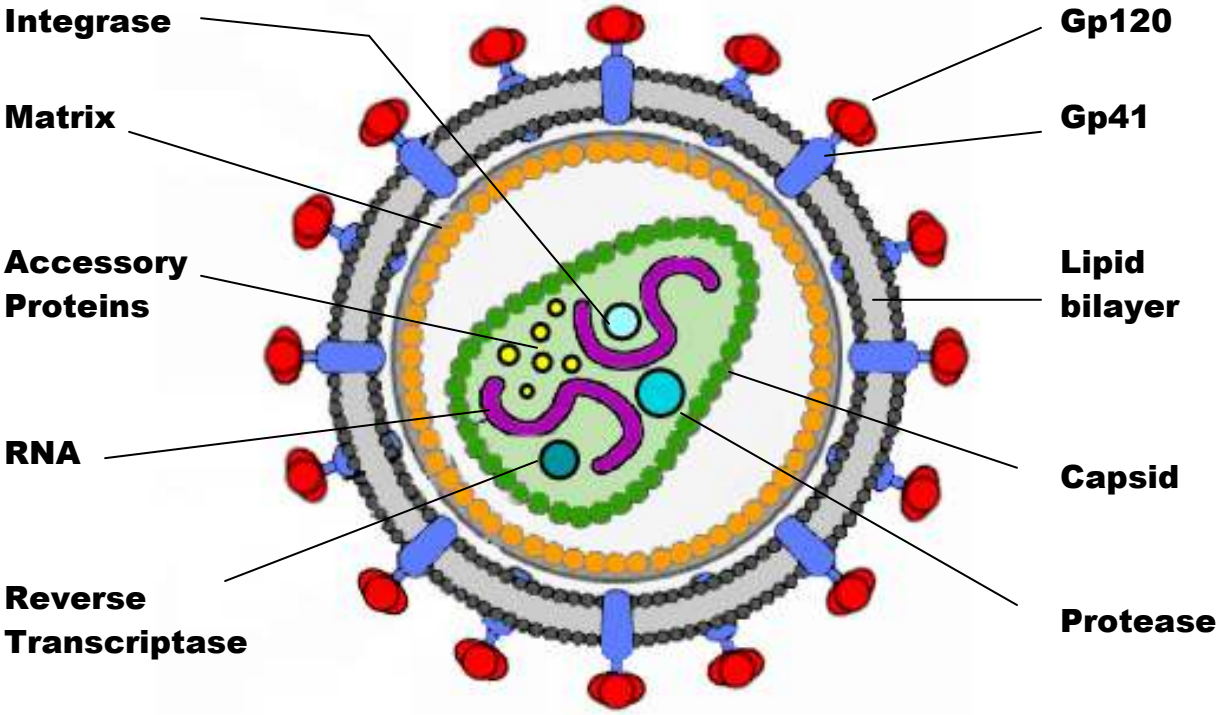


Figure 1.1: Structure of the HIV-1 Virion. Figure adapted from Sierra et al.[23]

The HIV genome (Figure 1.2) is approximately 9,5 kb in length and consists of 9 open reading frames, namely *gag*, *pol*, *env tat*, *rev*, *nef*, *vpr*, *vpu* and *vif* [24]. *Gag* encodes the viral matrix, capsid and nucleocapsid proteins. *Pol* codes for the protease, reverse transcriptase, RNase H and integrase proteins, while *env* encodes the surface glycoprotein gp120 and the transmembrane glycoprotein gp41.

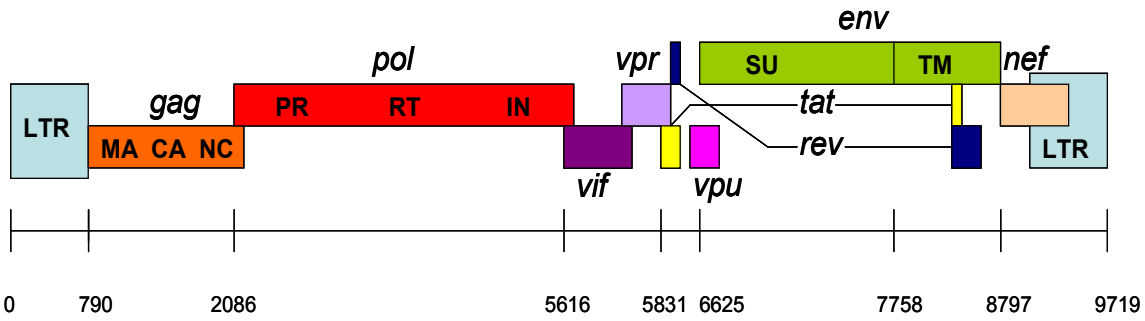


Figure 1.2: Genomic Organisation of HIV-1. Each open reading frame is shown in a different colour. Positions of some major genomic elements are shown (according to numbering of the HIV-1 strain, HXB2). MA: matrix protein, CA: capsid protein, NC: nucleocapsid protein, PR: protease, RT: reverse transcriptase, IN: integrase, SU: surface glycoprotein gp120, TM: transmembrane glycoprotein gp41, LTR: long terminal repeat. Diagram adapted[23, 25].

1.2.1.1. Genetic Variability of HIV-1

HIV-1 is a genetically diverse organism. There is a large amount of sequence variability across the entire genome between viral isolates obtained from different patients as well as between different isolates obtained from the same patient (quasi-species)[26-29]. This can be attributed to the error-prone retroviral reverse transcriptase (which introduces mutations

into the reverse-transcribed genome at a rate of 3.4 x 10^{-5} per base pair per cycle)[30-32], the high replication rate of HIV-1 (10^{8}-10^{10} virions per patient per day)[33-35], and viral recombination events[36].

Phylogenetic analysis has allowed for HIV-1 to be divided into three distinct groups. These are the major group (M), which is responsible for much of the current pandemic, and two much rarer groups, an outlier group (O) and a non-M/non-O group (N), both found predominantly in Africa and in particular Cameroon[37-40]. Group M HIV-1 viruses have been further divided into distinct subtypes[41], of which there are currently nine, designated by the letters A-D, F-H, J and K[42]. The envelopes of viruses within the same subtype can differ from each other by up to 20% in their amino acid sequences, while those of viruses belonging to different subtypes can differ from each other by up to 35%[43]. Recombination between viruses of different subtypes within the same patient has also resulted in the emergence of circulating recombinant forms or CRFs. These viruses contain clearly defined sections derived from multiple subtypes and to be defined as a CRF, such a recombinant virus must form the basis of at least three infections that are not epidemiologically linked[42, 44]. At least 34 CRFs have been described to date[45].

The HIV-1 subtypes and CRFs are found with varying prevalence in different regions worldwide. Statistics published recently by the WHO/UNAIDS[46] showed that Subtype C was the most prevalent subtype of HIV-1, accounting for approximately 50% of all HIV-1 infections worldwide. Subtypes A, B, D and G followed, accounting for 12%, 10%, 3% and 6% respectively. Together, subtypes F, H, J and K were responsible for 0.94% of

infections. Recombinant forms of HIV-1 accounted for 18% of infections worldwide, with CRF01_AE and CRF02_AG being responsible for 5% of cases each and CRF03_AB being responsible for 0.1%. Other recombinant forms made up the remaining 8%. With regards to the global distribution of HIV-1, Sub-Saharan Africa accounted for 64% of all HIV-1 infections worldwide, with 56% of these infections being subtype C. In southern Africa, which accounted for 30% of the world's HIV-1 positive individuals, 98% of infections were caused by subtype C. These figures, together with the genetic variability of HIV-1, highlight the importance of studying subtype C in order to elucidate similarities and differences between subtypes that may be crucial in disease control.

1.2.2. The HIV-1 life cycle

The life cycle of HIV-1 (figure 1.3) begins with the release of the viral core into CD4-positive host cells. Once the viral core has been released, the capsid disassembles in a process known as uncoating, releasing the ribonucleoprotein complex and associated proteins into the cell[47,48]. Reverse transcription of the viral RNA into cDNA is initiated at this stage, possibly even before the uncoating process takes place[48, 49]. Full length, double-stranded cDNA is transported via the cell's dynein and microtubule network as part of a preintegration complex (PIC) to the nucleus of the host cell for integration[49]. The PIC is imported into the nucleus and the cDNA is integrated into the host genome by the viral integrase. Interestingly, HIV appears to target sites for integration selectively, favouring active genes and in particular those activated by HIV-1 infection of the cell[50, 51]. Once integration of the viral cDNA has taken place, spliced and unspliced forms of the

viral RNA transcripts are synthesized and transported out of the nucleus. Spliced mRNA is translated into the viral accessory proteins tat, nef and rev. Rev is responsible for transport of mRNA out of the nucleus into the cytoplasm of the cell for translation of the gag, the gag-pol precursor protein and env. Full length mRNA binds to the gag protein and is packaged into new virions. As new viral particles assemble, they bud from the host cell membrane and the viral protease cleaves the viral polyproteins to yield functional viral enzymes and structural components, a process known as maturation. Mature virions are then able to infect new host cells[22].

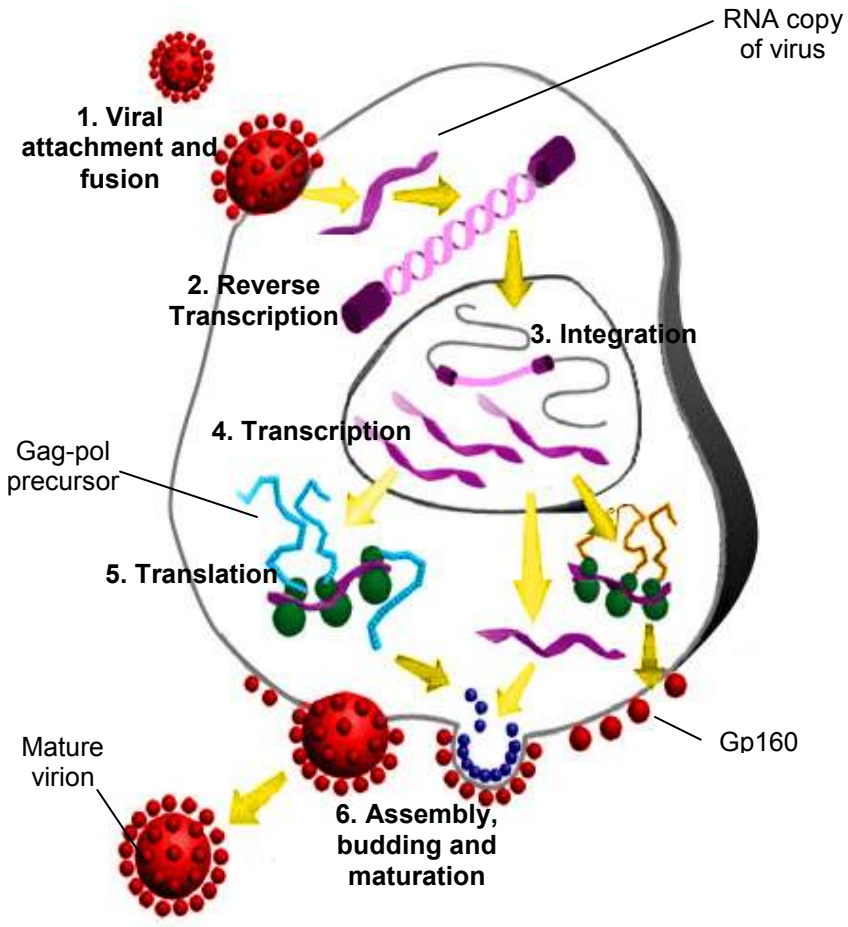


Figure 1.3: The life-cycle of HIV-1. Figure adapted from Sierra et al.[23]

1.3. HIV-1 entry into host cells

1.3.1. The HIV-1 Envelope Protein

Entry of HIV-1 into host cells is mediated by the HIV-1 envelope glycoprotein (env/gp160), a precursor protein that is transferred into the endoplasmic reticulum during translation for folding, disulfide bond formation and glycosylation, before transportation into the Golgi complex where it is cleaved to yield gp120 and gp41, the surface and transmembrane subunits of the envelope respectively[52-57]. Cleavage of the gp160 precursor is catalysed by furin, a protein convertase located in the Golgi complex[58-61], and results in non-covalently associated gp120/gp41 complexes that are transported to the cell membrane, where they assemble as trimers on the surface of budding HIV-1 virions [55, 62]. Investigation of the surface features of HIV-1 using cryo-electron microscopy has revealed that although there is much variation in the number of these trimeric spikes on each virion (between 4 and 35 per virion), in general there are approximately 14 such spikes on the surface of HIV-1, and that these may be clustered[63].

Studies of the amino acid sequence of gp160 reveal a hydrophobic leader or signal peptide sequence at the N-terminus of the protein (approximately 30-40 amino acids long), which is cleaved during maturation[64, 65]. This region is immediately followed by the gp120 and gp41 regions, with an arginine-rich hydrophobic area at the cleavage site between the two[65]. A number of studies comparing sequences of the envelope regions of different viruses have been performed[65-67]. These analyses found that the envelope region

of HIV-1 contains relatively conserved regions, as well as regions of extreme variability. In a computer-assisted analysis, 6 conserved regions, designated C1-C6 were elucidated (although it is now generally accepted that there are 5 conserved regions, C1-C5), together with five variable regions designated V1-V5[66]. Strikingly, all of the cysteine residues in gp120 were absolutely conserved[65]. In a study of the disulphide bonds present in the protein, gp120 was also found to contain 9 highly conserved disulphide bonds, four of which serve to render the hypervariable regions V1-V4 as well-defined loops[68] (figure 1.4).

Examination of gp160 sequences also revealed the presence of a number of potential glycosylation sites on gp120[64, 65]. Approximately half the molecular mass of the HIV-1 envelope glycoprotein is attributed to N-linked glycans[64, 69]. The typical envelope glycoprotein has approximately 24 N-linked glycosylation sites in gp120[68], as well as three or four sites in gp41[65]. Experimental data has revealed the presence of both complex-type, as well as high mannose- or hybrid- type carbohydrates on gp120, with all of the potential glycosylation sites being used[68] (figure 1.4). It has been suggested that N-linked glycosylation plays an essential role in neutralisation escape by HIV-1[70]. In a study of viruses obtained from acutely-infected HIV-1 patients over time, a population of virus that was sensitive to neutralising antibodies became slowly resistant to these antibodies and developed further resistance over time as the formation of new neutralising antibodies occurred[70]. When the envelope genes of these viruses were sequenced, many of the changes present were related to potential N-linked glycosylation sites, involving the addition, deletion or repositioning of glycosylation sites in

comparison to earlier populations, with some changes becoming fixed and others constantly changing over time in response to new antibodies. These findings led to the proposal of an "evolving glycan shield", able to inhibit the binding of neutralising antibodies to gp120 sterically, without affecting viral fitness. In a more recent study of HIV-1 transmission between heterosexual couples, viruses from newly-infected partners were sensitive to neutralization by antibodies from the transmitting partner, suggesting that the antibody resistance acquired during chronic infection is accompanied by a compromise in viral fitness[71].

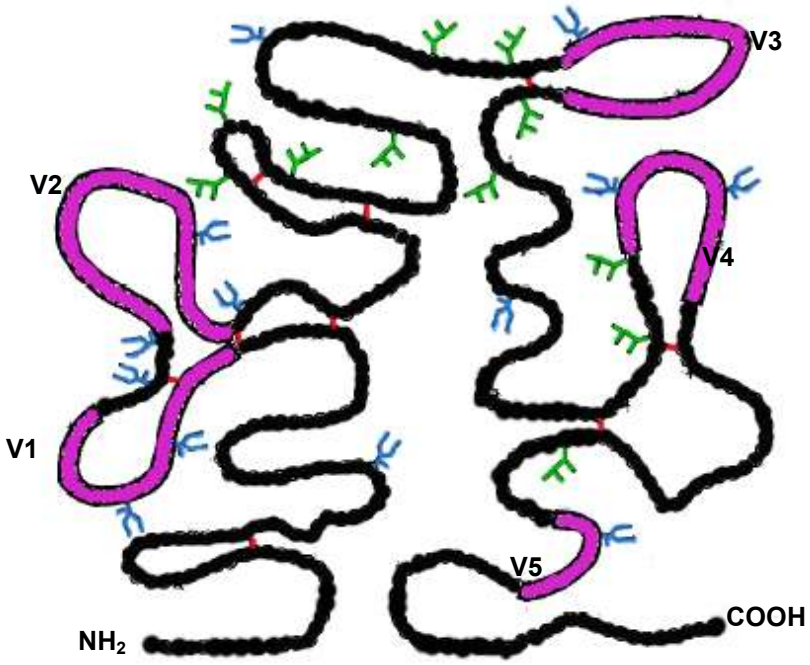


Figure 1.4: Schematic representation of gp120. The peptide backbone is coloured black, with the variable regions shown in purple and labelled V1-V5. High-mannose and/or hybrid-type glycans are shown in green, and complex glycans in blue. Disulfide bonds are indicated in red. Figure adapted from Leonard *et al.*[68]

1.3.2. Gp120

1.3.2.1. Structure of gp120

Determination of the structure of the envelope glycoprotein of HIV has been important for understanding of viral pathogenesis, as well as for the design of HIV inhibitors. Gp120 exhibits a large amount of conformational flexibility, and is extensively glycosylated, making it difficult to produce crystals that diffract well[72]. In order to crystallize gp120, it was therefore necessary to modify the protein, maintaining the biologically important properties such as CD4-binding, but in a more ordered structure[72]. For the first crystal structure of gp120, this was achieved by extensive deglycosylation of gp120, as well as truncation of gp120 by removal of the disordered and flexible variable loops[72, 73]. The conformational flexibility of gp120 was also restricted by crystallising it in complex with CD4 and a monoclonal antibody. Thus, the structure of the HIV-1 gp120 core (i.e. gp120 without the variable regions and N and C terminal segments) has been determined in complex with CD4 and the antigen-binding fragment (Fab) of the anti-gp120 antibody 17b[73, 74].

The structure of an unliganded simian immunodeficiency virus (SIV) gp120 core was determined later[75]. Interestingly, in the case of the SIV core, researchers found that the glycosylation of gp120 helped, rather than hindered the crystallization process[75]. More recently, the structure of an HIV-1 V3-loop-containing gp120 core has also been determined[76].

The first available structure of gp120[73] revealed the deglycosylated gp120 core to be a folded into two major domains (figure 1.5). These are an inner domain (with respect to the N- and C- termini, blue in figure 1.5), from which the V1/V2 loop extends, and a stacked double-barrel outer domain (green in figure 1.5), which includes the V4 and V5 loops. These two domains are linked by a four-stranded bridging sheet (orange in figure 1.5). HIV-1 entry into cells requires that gp120 binds to the host-cell surface receptor CD4[77, 78]. HIV-1 entry also requires that gp120 binds to a coreceptor, one of a family of G-coupled seven-transmembrane domain chemokine receptors, principally CXCR4 or CCR5 [79-84]. In this structure, a gp120-CD4-17b complex, CD4 was bound at the interface of the inner and outer domains and the bridging sheet, with its Phe43 residue "plugging" a cavity in gp120 at this interface (figure 1.6). In this structure, it was also noted that the binding site for the coreceptor CCR5 is geometrically directed to the cellular membrane when gp120 is bound to CD4.

Oligomeric modelling[74] (figure 1.7) suggested that all the variable loops of gp120, as well as the CD4 and neutralising antibody binding sites, are exposed on the gp120 trimer, and revealed the C-terminus of gp120 poised to interact with gp41. In this model, the coreceptor binding site is aimed directly at the target cell membrane and is not glycosylated. This model further showed that CD4 binds to gp120 at an angle, in a position far from the axis of the trimer, allowing for three CD4 molecules to bind to the gp120 trimer without sterically hindering one another. This also allows for the binding of chemokine receptors together with CD4, as these interact with gp120 close to the trimer axis.

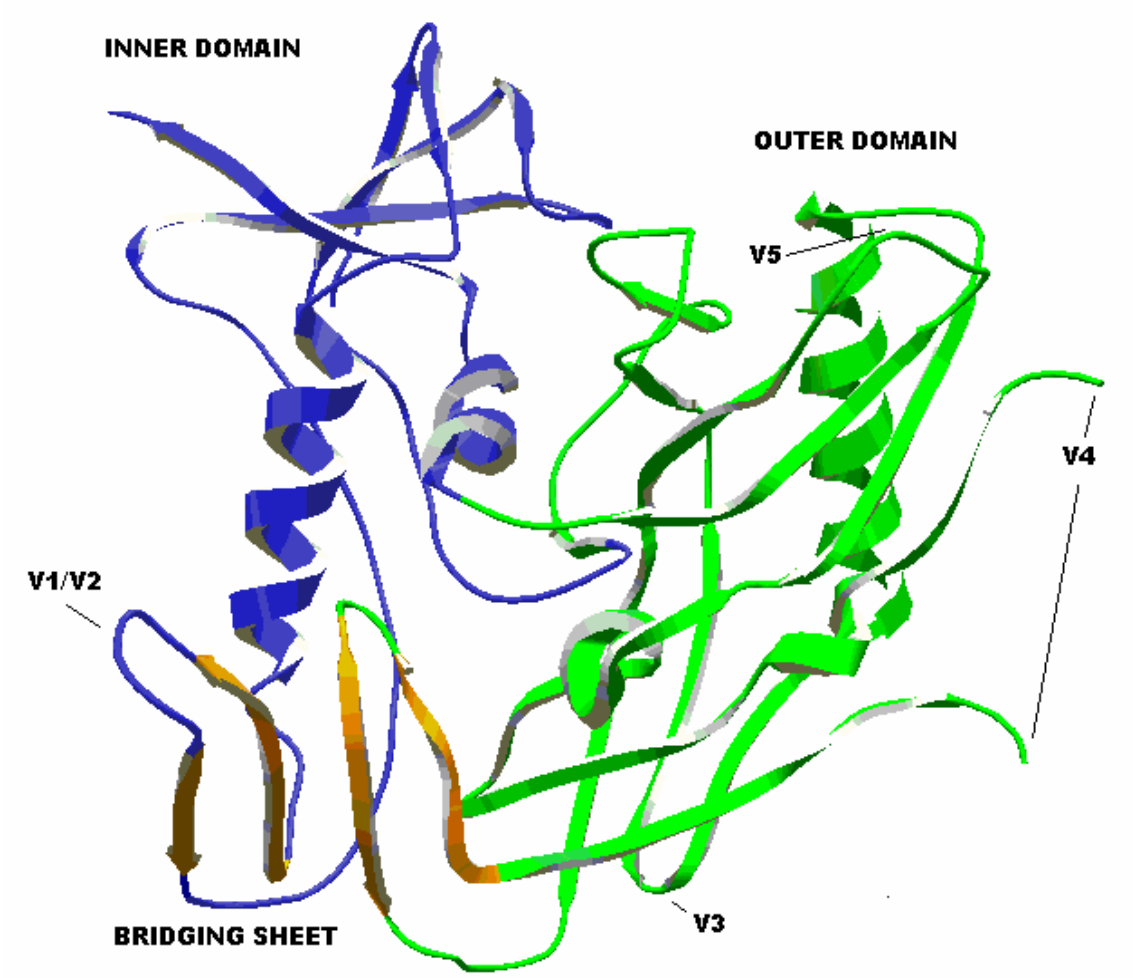


Figure 1.5: Ribbon diagram of gp120 in the CD4-bound conformation coloured to show the main structural domains. Adapted from Kwong *et al.*[73]. Crystal structure downloaded from the Protein Data Bank[85], PDB ID 1gc1, and altered using Swiss-Pdb Viewer 3.7 (SP5)[86] (http://www.expasy.org/spdbv/). The inner domain is blue, the outer domain is green, and the bridging sheet orange. Although the variable loops are missing in these structures, their positions are indicated in this diagram.

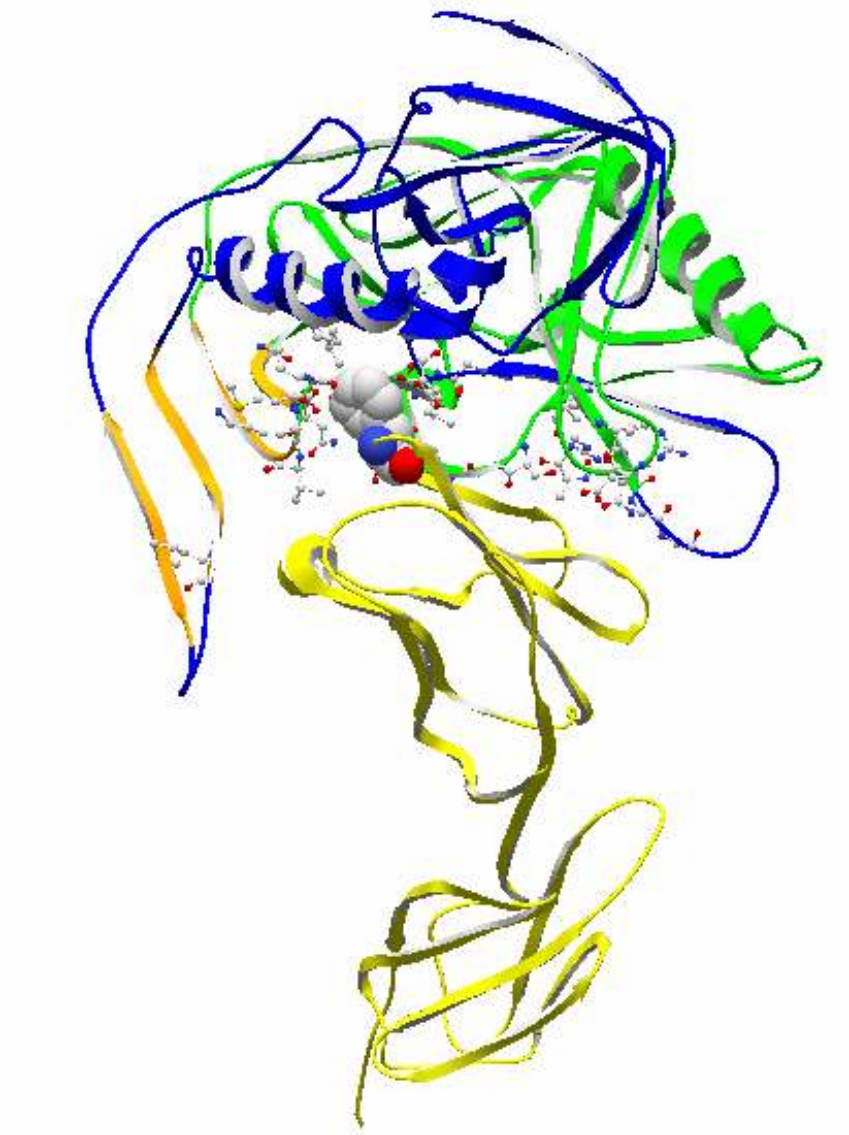
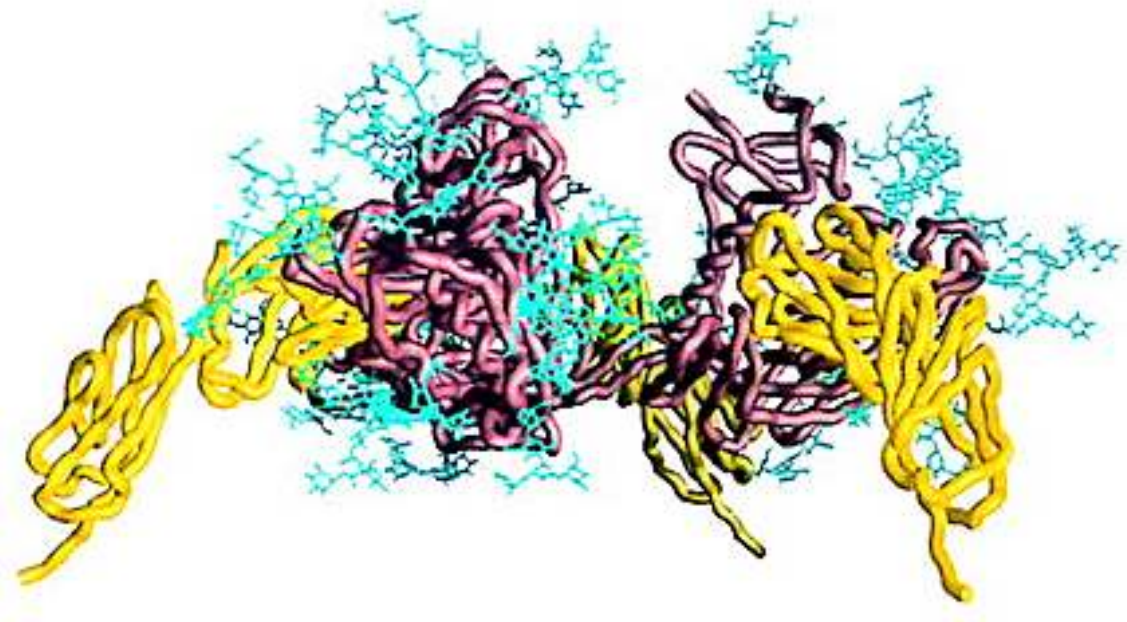


Figure 1.6: A ribbon diagram showing the interaction of CD4 and gp120. Adapted from Kwong *et al.*[73]. Crystal structure downloaded from the Protein Data Bank[85], PDB ID 1gc1, and altered using Swiss-Pdb Viewer 3.7 (SP5)[86] (http://www.expasy.org/spdbv/). The inner domain is blue; the outer domain is green, and the bridging sheet orange. CD4 is yellow. The residues of gp120 that are in direct contact with CD4 are coloured in red, with their amino acid side-chains showing. Phe43 of CD4 is shown as a space-filled model in order to highlight its position in the Phe43 cavity of gp120.

VIRAL MEMBRANE



HOST CELL MEMBRANE

Figure 1.7: Model of gp120 oligomer as viewed from the side with the viral membrane above and the target membrane below. The gp120 core is purple, while CD4 is yellow and carbohydrate structures are blue. In this orientation, the CCR5 binding site is on the target membrane side, aimed directly at the target membrane. Picture taken from Kwong *et al*.[74]

When compared to the structure of a non-glycosylated CD4-bound HIV-1 gp120, the structure of a fully glycosylated, unliganded SIV gp120 core revealed a number of remarkable differences[75] (figure 1.8). In this structure, the inner and outer domains of gp120 were in different positions relative to each other as compared to the liganded

structure, the bridging sheet was no longer an individual domain and the CD4 and coreceptor binding sites were not properly formed. This led to the postulation that unliganded gp120 has a more relaxed conformation than liganded gp120, and that the bridging sheet folds on CD4 binding to form the coreceptor binding site as seen in the original CD4-bound structure of Kwong *et al.*[73]. This postulation is supported by thermodynamic data[87], which indicates a great amount of structural flexibility within gp120. Large thermodynamic changes are seen on CD4 binding to gp120, which implies significant conformational rearrangement of the gp120 core. These changes can be accounted for by the folding of the bridging sheet, as well as the pulling together of the bridging sheet and inner and outer domains of gp120, leading to a loss of surface area and ultimately the exposure of the coreceptor binding site[87].

None of the aforementioned structures, however, included the third variable region (V3 loop) of gp120, a notable absence as this is a region critical for determination of coreceptor usage of HIV-1[88]. It has only recently been possible to elucidate the core structure of HIV-1 gp120 with the V3 loop included, providing fascinating insight into the role of the V3 loop during coreceptor binding[76]. The crystal structure of the V3-containing gp120 core revealed that the V3 loop, 50Å from base to tip, extends from the outer domain of gp120 (figure 1.9). It consists of a conserved base with a disulfide bridge, a flexible stem and a β-hairpin tip. The structure and position of this loop led to the proposal that the N-terminus of the coreceptor reaches up and binds to the gp120 core and V3 base, while the V3 tip reaches down, and acts as a "molecular hook", interacting with the second extracellular loop of the coreceptor.

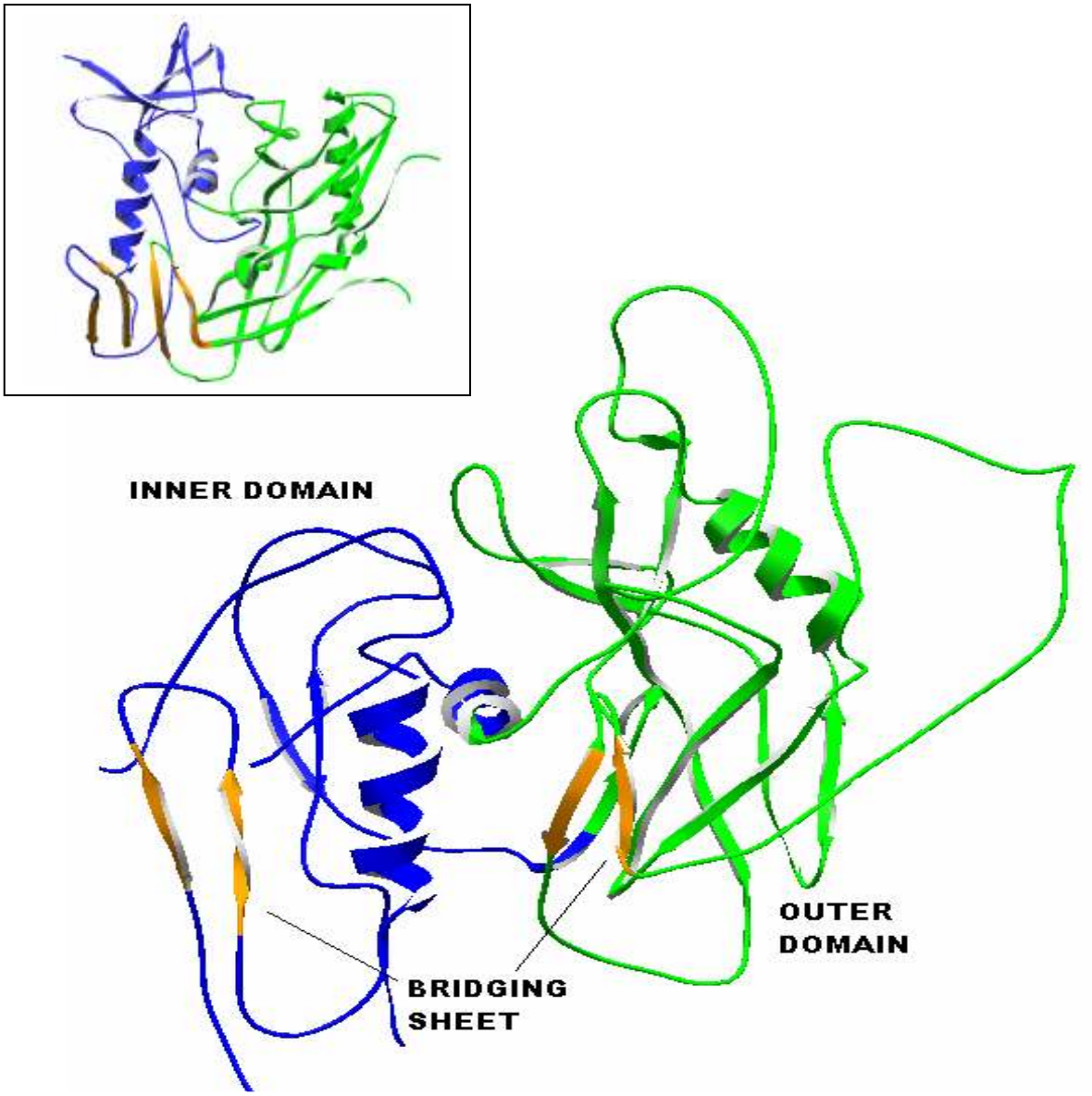


Figure 1.8: Crystal Structure of uliganded SIV gp120. Adapted from Chen *et al.*[75] Crystal structure downloaded from The Protein Data Bank[85], PDB ID 2bf1 , and altered using Swiss-Pdb Viewer 3.7 (SP5)[86] http://www.expasy.org/spdbv/. Ribbon diagram of gp120 coloured to show the main structural domains. Inset for comparison purposes is figure 1.5, the CD4-bound crystal structure of gp120. For both figures, the inner domain is blue, the outer domain is green, and the bridging sheet orange. Glycans are not shown.

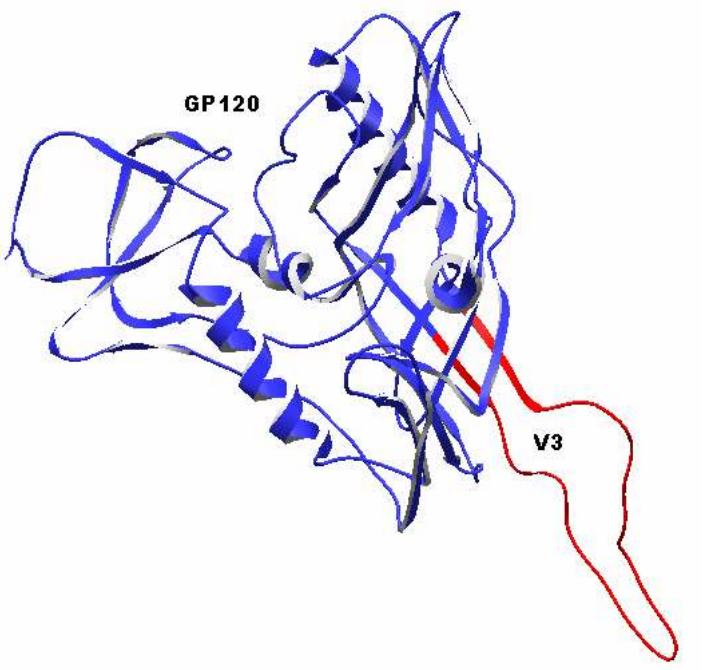


Figure 1.9: Crystal Structure of gp120 in a CD4-bound conformation, including the V3 loop. Adapted from Huang *et al*.[76] Crystal structure downloaded from the Protein Data Bank[85], PDB ID 2b4c , and altered using Swiss-Pdb Viewer 3.7 (SP5)[86] http://www.expasy.org/spdbv/. Gp120 is coloured in blue, with the V3-loop portion coloured in red. Glycans are not shown.

1.3.2.2. The role of gp120, its receptors and coreceptors in HIV-1 entry into host cells

Entry of HIV-1 into target cells is initiated by the interaction of gp120 and the host cell surface receptor CD4[77, 78]. This causes large conformational changes in gp120, resulting in exposure of previously masked gp41 epitopes and, *in vitro*, dissociation of gp120 from gp41 [87,89]. Conformational changes also lead to exposure of the V3 loop, as well as movement of the V1/V2 loop, which results in exposure of the coreceptor binding site of

gp120[89, 90]. The coreceptors for HIV entry are principally CXCR4 or CCR5 [79-84]. A number of other members of the G-protein-coupled receptor family have also been implicated as coreceptors for HIV-1 although to a much lesser extent than CCR5 and CXCR4. These include CCR3[81], CCR2b[83], GPR15[91], STRL33[91, 92], GPR1[91], CCR8[93], US28[94], V28[93], APJ[95] and ChemR23[96]. Viral tropism (i.e. Macrophage- or T-cell tropism) is linked to coreceptor usage, with CXCR4(X4)-using viruses being T-tropic and syncytium-inducing (SI), and CCR5(R5)-using viruses being M-Tropic and non syncytium-inducing (NSI)[79-84, 97, 98]. Coreceptor binding and specificity of the virus appears to be principally determined by the V3 loop, which is exposed by CD4-induced conformational changes, although the V1/V2 loops, also exposed upon binding of CD4, have also been implicated[81, 88, 99-104].

Coreceptor usage appears to be an important factor in HIV pathogenesis. Transmitted HIV-1 is NSI and these CCR5-using viruses have been shown to predominate in early or asymptomatic HIV-infection[105-107]. In later infection, however, CXCR4-utilising isolates often emerge and extensive analysis has shown this shift from a primarily CCR5-using to a primarily CXCR4-using virus population to be closely correlated with increased CD4+ T-cell decline and progression to AIDS[106-112]. When compared to CCR5-using viruses, CXCR4-using viruses also show increased cytopathicity[113, 114], which may account for the link between coreceptor switching and increased pathogenicity of HIV.

Coreceptor usage and switching has been studied across the various HIV-1 subtypes. In a study comprising HIV-1 isolates across all nine clades, CXCR4-using viruses were rarely

found amongst subtype C isolates[115], which account for 50% of HIV infections worldwide[46]. Numerous other studies have found CXCR4-using isolates to be extremely rare in subtype C cohorts[116-122]. Recently, however, a maturing of the HIV-1 subtype C epidemic has been suggested as higher percentages of CXCR4-utilising HIV-1 have been found amongst clade C isolates[123, 124], a disturbing finding when taking into account the association between X4 use and disease progression and pathogenesis.

1.3.3. Gp41

1.3.3.1. Structure of gp41

The structure of the transmembrane protein of HIV-1, gp41, has also been extensively studied[125-128]. Early protein dissection studies[128] resulted in two peptides designated N-51 and C-43, from the N- and C-terminal regions of gp41 respectively. When mixed, these peptides associated to a trimer with an interior, parallel coiled coil consisting of three N-51 helices, and with three C-43 helices packed around it in an antiparallel fashion. The stability of this structure led to it being proposed as the fusion active conformation of the HIV-1 envelope.

Crystal structures of the relevant regions of gp41 later confirmed the presence of a six-helix bundle within this molecule[125-127] (figure 1.10). All of these structures indicated the presence of a central, parallel, trimeric coiled coil consisting of three N-terminal helices with three C-terminal helices packed in the reverse direction into hydrophobic grooves on

its surface, juxtaposing the amino and carboxy termini at one end of the rod-shaped structure.

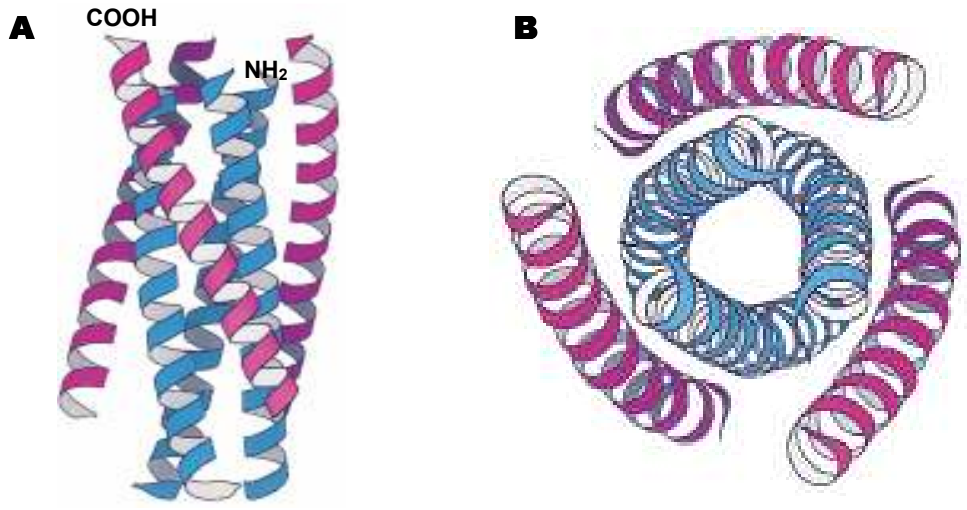


Figure 1.10: Representation of the coiled coil structure of gp41. N-terminal helices are coloured blue and C-terminal helices are purple. (A) Side-on view. (B) View from above. Figure taken from Chan *et al.*[125]

1.3.3.2. Gp41 and the HIV-1 fusion mechanism

The resemblance of gp41 to other viral membrane fusion proteins, including the influenza virus haemagglutinin protein, has long been recognised[129]. This similarity, taken together with experimental data, has led to a model for HIV-1 fusion similar to the influenza virus haemagglutinin model[127, 130-132] (figure 1.11). Based on this model, it has been suggested that on CD4 and coreceptor binding, the gp120/gp41 association is considerably

weakened, leading to exposure of a gp41 N-terminal fusion peptide, which penetrates the host-cell membrane.

It has been suggested that the conformational changes that result in exposure of the fusion peptide may be caused by the binding of multiple coreceptors to gp120[74]. The binding site for gp120 coreceptors is close to the axis of the gp120 trimer and thus the binding of multiple coreceptors to gp120 may result in steric strain. Movement of the gp120 subunit away from the axis of the trimer in order to relieve this strain may result in conformational changes in gp41, exposing the fusion peptide and allowing for its insertion into the host-cell membrane[74].

After exposure of the fusion peptide and penetration of the host cell membrane, gp41 trimers enter a fusion-active conformation by folding back on themselves to form a six-helix bundle, bringing the N-terminal fusion peptide, which is inserted into the host-cell membrane, into close proximity with the C-terminal transmembrane region fixed in the viral membrane. This action brings the viral and host-cell membranes together. Around the time of six-helix bundle formation, a fusion pore or channel connecting the aqueous compartments of both host cell and virus forms, allowing the passage of the viral core into the cell[133, 134]. It has been suggested that the formation of six helix bundles not only brings together the viral and host cell membranes, but also stabilizes the fusion pore for efficient transfer of the viral core[133].

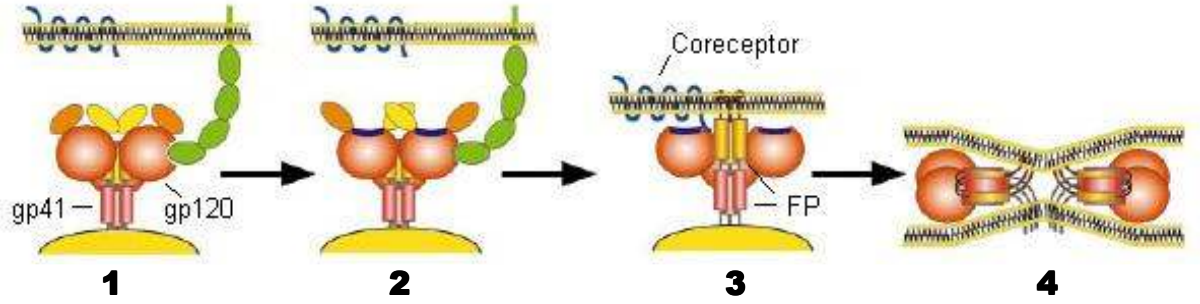


Figure 1.11: Stages of HIV Entry into host cells. Figure adapted from Doms and Moore[135]. 1. CD4 binds to gp120. 2. CD4 binding causes a conformational rearrangement in gp120 that results in exposure of the coreceptor binding site and a number of gp41 epitopes. 3. Upon coreceptor binding, further conformational changes result in the weakening of the gp120-gp41 interaction, and exposure of the N-terminal fusion peptide (FP), which then inserts into the host cell membrane. 4. Gp41 then folds back on itself to form a six-helix bundle, drawing the viral and host cell membranes together, and allowing for entry of the virus into the host cell.

1.4. HIV-1 drug targets

The development of drug therapies for HIV-1 has been targeted at many aspects of the HIV-1 life cycle, and a number of drug classes exist. These include protease inhibitors and reverse transcriptase inhibitors, which form the basis of the currently most effective HIV-treatment strategy - highly active antiretroviral therapy (HAART). HAART typically combines the use of two nucleoside reverse transcriptase inhibitors (NRTIs) with one protease inhibitor (PI) or two NRTIs with one non-nucleoside reverse

transcriptase inhibitor in order to inhibit viral replication[136]. HAART, particularly with the inclusion of a protease inhibitor, has been associated with a significant decline in morbidity and mortality rates linked with advanced HIV[137], but is hampered by the toxicity of the drugs, as well as the rapid emergence of drug resistant species of the virus. These problems have necessitated the development of novel therapeutic strategies against HIV. One class of drugs that has recently attracted a lot of attention is the fusion or entry inhibitors, which target various steps and proteins involved in the HIV viral entry process.

1.4.1. Therapeutic targeting of gp120 interactions

The pivotal role of gp120 in HIV-1 entry renders it an attractive target for drug design, and a number of strategies aimed at inhibition of HIV-1 replication by disrupting the interactions of gp120 have been attempted.

1.4.1.1. Large molecule inhibitors

The potential for using soluble CD4 (sCD4) as a competitor to block the gp120-CD4 interaction was explored early in the HIV epidemic. Initially, these studies yielded promising results[138, 139], however it later became clear that sCD4 was only effective against laboratory-adapted strains of HIV-1, with primary isolates being relatively resistant to challenge by this molecule[140].

Another approach to preventing the gp120-CD4 interaction involves the use of a recombinant CD4-based fusion protein, Pro542[141, 142]. This molecule comprises human IgG2, in which the Fv portions of both heavy and light chains have been replaced by the V1 and V2 domains of human CD4. Due to its tetravalent structure, Pro542 can bind Env with higher avidity than sCD4. Pro542 has been used with promising results in patients[143], but development of this drug has been halted as the manufacturer has refocused its efforts towards development of a more potent drug, Pro140. Pro140 is a monoclonal antibody directed against the HIV-1 coreceptor CCR5 that inhibits HIV entry at concentrations that do not affect its chemokine receptor activity[144]. In Phase 1 studies, this drug was well tolerated in healthy volunteers[145] and is currently under further development. Another monoclonal antibody that shows potential as an entry inhibitor is TNX-355. TNX-355 (previously called Hu5A8) is a humanized IgG4 monoclonal antibody that binds to CD4 domain 2 and does not prevent CD4 binding to gp120, but instead appears to prevent further conformational changes necessary for viral entry[146, 147]. Clinical trails using this antibody have shown some success[148, 149] and larger trials are currently underway.

1.4.1.2. Small molecule inhibitors

The promising results obtained from studies involving entry inhibitors have led to an increased interest in molecules that disrupt the interactions between gp120 and its human cellular receptors. The focus, however, has shifted onto smaller molecules than those described in the preceding section. Reduced size would allow greater access of the

molecule to conserved, functionally important domains of gp120 and would possibly have the advantage of cost-effectiveness as well as easier administration (i.e. oral) than bulkier recombinant proteins and antibodies (which are administered via injection).

One such molecule, identified from a whole cell infection screen against HIV-$1_{JRFL.}$ is BMS-378806[150-152]. This compound inhibits the gp120-CD4 interaction by binding gp120 in the CD4-binding pocket, thereby blocking the attachment of CD4. It has been shown capable of neutralising both laboratory-adapted and primary virus isolates, including those resistant to protease and reverse transcriptase inhibitors. Initial studies showed no significant cytotoxicity and supported an oral formulation of the compound for man[151]. Further studies, however, revealed that BMS-378806 shows a decreased activity against subtype C, as well as other HIV-1 subtypes when compared to its efficacy against subtype B strains[151], probably due to the inherent variability between HIV-1 envelope proteins and the use of a subtype B virus in the screening process. This drug is currently under investigation as a microbicide and has shown some promise in this regard[153]. A number of small molecule coreceptor antagonists are also currently under development. AMD070 is an orally bioavailable CXCR4 antagonist that has shown promising results in early clinical trials[154, 155]. Maraviroc, an imidazopyridine small molecule CCR5 antagonist, was discovered through screening of the Pfizer compound file and used with positive results in clinical trials, resulting in approval for its use from the FDA[156-158].

Other small molecules that have attracted attention as HIV entry inhibitors are peptides. In fact, the first FDA-approved entry inhibitor, Enfuvirtide, is a peptide. Enfuvirtide (T-

20/DP-178), which received FDA approval in 2003 as salvage therapy for AIDS patients failing antiretroviral treatment, is a 36 mer synthetic peptide derived from the sequence of HR2 in gp41[159]. It exerts antiviral activity by interacting with the viral sequence within HR1 and inhibiting its association with the viral HR2, thereby preventing gp41-mediated fusion[159, 160]. Peptide inhibitors have also been developed against gp120. Peptide-T is an octapeptide with the sequence ASTTTNYT that was designed based on an eight amino acid sequence in the V2 loop of gp120, and which has been demonstrating antiviral activity since 1986[161]. It was more recently discovered that this peptide works by blocking gp120/CCR5 chemokine receptor-mediated chemotaxis[162]. When administered to patients, Peptide-T shows both antiviral activity and immunological benefits and is currently under further investigation[163].

Peptides interfering with the gp120-CD4 interaction also have great potential as entry inhibitors. Aside from rational design, a successful approach for discovery of ligands that interact specifically with gp120 and which may inhibit its interaction with CD4 is that of selecting such ligands from phage display libraries[164, 165]. Ferrer and Harrison[164] used M13 bacteriophage libraries expressing random 12- and 7- mer peptides to select peptides that bind to gp120 with high affinity. Competition ELISA assays on peptides selected by this method showed that these reduced binding of gp120 to CD4. One specific peptide selected during this study, with the sequence RINNIPWSEAMM (designated 12p1), was found not only to inhibit the interaction of gp120 and CD4, but also the interaction of gp120 with monoclonal antibody (MAb) 17b (an antibody that recognises an epitope overlapping the CCR5 binding site). This is significant, as compounds that inhibit CD4

binding to gp120 may cause conformational changes similar to those caused by CD4, thereby activating gp120 to a state conducive to coreceptor binding and so enhancing, rather than inhibiting, virus entry[164, 166]. Rather than functioning as a competitive inhibitor and binding the CD4 binding site in place of CD4, 12p1 appears to function via an allosteric inhibitory mechanism, binding gp120 at a site distinct from that of CD4 and stabilising gp120 in a non-productive intermediate conformation, which is unable to bind either CD4 or a coreceptor[166]. It is important to note, however, that this study was done with recombinant subtype B gp120 clones (HXB2, SF2 and ADA) and the ability of the peptide to disrupt interactions between CD4 and gp120 cloned from other clades was not established. The efficacy of the compound against primary HIV-1 isolates in cell culture was also not established.

The peptides isolated by this method have also served as a starting point for the rational design of stronger inhibitors[167]. 12p1 was modified using click conjugation in order to generate the inhibitory molecule HNG-105[167, 168]. In order to establish whether or not this molecule has broad-range specificity, it was tested against a panel of gp120s, as well as viral isolates of different subtypes[167]. These studies revealed that HNG-105 shows inhibitory action against a variety of subtypes (A, B, C and D), although the panel used only included a maximum of two viruses from each non-B subtype. An additional point to note is that the inhibition observed was highly variable between viruses, both between different subtypes and within the subtypes themselves. Therefore, while these studies indicate that bio-panning is a powerful tool for the selection of gp120 antagonists and support the search for small molecule inhibitors of the gp120-CD4 interaction, they also

highlight the importance of using a locally relevant clone of gp120 in order to establish a compound that will be effective against the subtype C viruses found in Southern Africa.

1.5. Project objectives

One goal of the work in our laboratory is the identification of small 7-mer subtype C-specific gp120-binding peptides by phage display techniques. The large amount of gp120 that will be required for this necessitates the development of technologies for the expression and purification of recombinant gp120 from locally predominant strains of HIV. To this end, the aim of this study was to express, purify and characterise recombinant subtype C gp120. In other studies, the HIV-1 envelope protein was successfully expressed by recombinant baculoviruses in insect cells[169-171], as well as in mammalian cells[69, 172]. A common approach used to purify gp120 is that of capturing recombinant gp120 from cell culture supernatants on affinity chromatography columns containing immobilised lentil lectin (*Lens culinaris*), followed by ion-exchange chromatography[172]. This study describes similar work performed in our laboratory to produce, purify and characterise two new recombinant Subtype C gp120s.

The purified gp120 produced by procedures established in our laboratory during this study will be tested for its sensitivity to existing gp120 antagonists. Ultimately, it will also be used to select compounds with high affinity for subtype C gp120 from combinatorial phage-display libraries. The ability of these compounds to disrupt binding

of gp120 to CD4 and to inhibit the replication of a variety of HIV-1 isolates in cell culture will be assessed. This will allow a direct comparison of relative inhibitory effects on HIV-1 replication *in vitro*, of gp120 binding peptides identified in this study and those developed by selection with non-C gp120 clones. It is hoped that the ultimate outcome of the study will be a lead compound that may be used to augment the local currently available therapies.

2. Materials and Methods

2.1. Patient Samples

HIV-infected PBMCs were a kind gift of Ms B Connell, who processed blood samples collected in 5ml EDTA tubes from advanced AIDS patients presenting at the Johannesburg Hospital AIDS clinic in 2005. HIV-infected PBMCs were cocultured according to standard coculture techniques, with PBMCs obtained from healthy donors from the South African National Blood Service. Ethics clearance for the acquisition of the HIV-infected samples was obtained from the University Ethics Committee (Protocol No: M041002). The viral samples collected were designated 05ZAFV1-05ZAFV32, hereafter referred to as FV1-FV32

2.2. Viral Nucleic Acid Isolation

After cell culture, proviral DNA from the corresponding PBMC cocultures was isolated (from 1 ml PBMCs) using the High Pure PCR Template Preparation Kit (Roche; Mannheim, Germany) according to the manufacturer's protocol.

2.3. PCR Amplification of HIV-1 Subtype C gp160 coding regions

2.3.1. Primer Design

The basic *env* primers described by Gao *et al*[173] were adapted for subtype C specificity by comparison of 52 complete HIV-1 subtype C genome sequences retrieved from the National Library of Medicine Nucleotide Database[174]. The adapted primers, EnvA-1, EnvB-1, EnvM-1, and EnvN-1 are described in table 2.1, with the changes from the primers of Gao *et al*. underlined and in bold.

2.3.2. Nested PCR

Nested PCR was performed directly on proviral DNA, using the outer primers EnvA-1 and EnvN-1 and the inner primers EnvB-1 and EnvM-1 (Table 2.1). PCR was performed using the High Fidelity Expand^{PLUS} PCR System (Roche ; Mannheim, Germany). Proviral DNA isolated from PBMCs was diluted 100 times, and 5 µl (approximately 250 ng) of the dilution was used in 50 µl PCR reactions, which contained 1 x Expand HiFi^{PLUS} Reaction buffer with 1.5 mM $MgCl_2$, 200 µM dNTPs, 0.4 µM of each EnvA-1 and EnvN-1 primers and 2.5 U Expand HiFi^{PLUS} Enzyme Blend. All PCR reactions were done on a Bio-Rad MyCycler^{TM} (Bio-Rad; Hercules, CA). PCR began with a 3 minute hot start, followed by 30 cycles consisting of a 15 second denaturation at 94°C, a 30

second annealing step at 61°C, and a 3 minute elongation step at 72°C. This was followed by a single 10 minute elongation step at 72°C, where after the reactions were cooled to 4°C indefinitely.

The second round of amplification of the gp160 coding region was set up as for the first round, containing 0.4 μM each of EnvB-1 and EnvM-1 instead of EnvA-1 and EnvN-1. PCR began with 3 minute hot start, followed by 30 cycles consisting of a 30 second denaturation at 94°C, a 45 second annealing step at 55°C, and a 3 minute elongation step at 72°C. This was followed by a single 10 minute elongation step at 72°C, where after the reactions were cooled to 4°C indefinitely.

PCR products were run on a 1% agarose (Sigma; St Louis, MO) gel together with 0.5 μg of Quick-Load™ 1 kb DNA Ladder (New England Biolabs; Ipswich, MA). Products from successful PCR reactions were purified using a HighPure PCR Product Purification Kit (Roche ; Mannheim, Germany) according the manufacturer's recommended protocol and eluted in 50-75 μl of Elution buffer for every 100 μl of PCR product cleaned.

2.4. Population-based sequencing of the gp160 coding regions

Purified second round gp160 amplicons of 20 HIV-1 viral isolates were sequenced with 15 primers (table 2.1) in order to obtain double stranded sequence in both the forward and reverse direction for the entire amplicon. Sequencing reactions were set up in a total volume of 21 μl, using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied

BioSystems; Foster City, CA). Each reaction contained 20-50 ng of DNA, 4.0 µl of Terminator Ready Reaction Mix, 1 x BigDye Sequencing buffer, and 1 µl of primer at 3.2 pmol/µl (a final concentration of 0.15 µM). Sequencing reactions were placed in a Bio-Rad MyCycler™ (Bio-Rad; Hercules, CA); undergoing 25 cycles consisting of 10 seconds at 94°C , 5 seconds at 50°C and 4 minutes at 60°C. This was followed by an incubation of at least 7 minutes at 4°C.

After the cycle sequencing reaction, extension products were purified by isopropanol precipitation. Samples were transferred to a sequencing plate and 80 µl an of 80% isopropanol solution was added to each well and reverse pipetted in order to mix. The plate was then placed in the dark at room temperature for 15 minutes before a 45 minute centrifugation (2000 x g) at room temperature. Within one minute of the end of the centrifugation step, the plate was removed from the centrifuge and inverted onto paper toweling. The plate was then centrifuged in this position (inverted) for 1 minute (700 x g). The plate was allowed to air dry for a few minutes, after which, 10 µl Hi-Di™ Formamide (Applied BioSystems; Foster City, CA) was added to each well. Sequences were run in an ABI 3100 automated sequencer (Applied BioSystems; Foster City, CA). Full length gp160 sequences were then assembled and edited using Sequencher v4.6 (GeneCodes, Ann Arbor, MI).

Phylogenetic analysis was performed by Dr M Papathansopoulos. A phylogenetic tree was constructed for the 20 isolates. Briefly, a multiple alignment of the gp160 regions with references from HIV-1 subtypes A to K, CRF01_AE and CRF02_AG[45] was

generated in Clustal X and aligned sequences were converted to MEGA V3.0 format, and used in phylogenetic and molecular evolutionary analyses. A phylogeny reconstruction of each *env* gene was performed by Neighbor-Joining using the Kimura two-parameter distance matrix. The stability of the nodes was assessed by bootstrap analysis (1000 replicates) and bootstrap values greater than 70% were considered significant.

2.5. Amplification of gp120 coding regions

Gp120-encoding DNA was amplified from second round gp160 amplicons using a High Fidelity Expand^{PLUS} PCR System (Roche ; Mannheim, Germany) and the primers gp120(F) and gp120(R) (Table 2.1). These primers were designed according to consensus gp120 5' and 3' sequences identified by comparison of 26 complete HIV-1 subtype C *env* sequences published on the NLM nucleotide database[174]. The primers additionally included 5'-terminal restriction endonuclease recognition sequences to facilitate subsequent cloning procedures. In gp120(F), a restriction endonuclease recognition sequence for *Pag* I (*Bsp* HI) (T/CATGA) was included, while in gp120(R), the site included was for *Xho* I (C/TCGAG).

Second round gp160 amplicons were diluted 100 times and 5 µl of this dilution (approximately 300 ng) was used in 50 µl reactions containing 1 x Expand HiFi^{PLUS} Reaction buffer with 1.5 mM $MgCl_2$, 200 µM dNTPs, 0.4 µM each of gp120(F) and gp120(R), and 2.5 U Expand HiFi^{PLUS} Enzyme Blend. PCR began with a 5 minute hot start at 94°C, followed by 30 cycles consisting of 30 seconds at 94°C, 30 seconds at

55°C, and 2 minutes at 72°C. This was followed by a single 10 minute elongation step at 72°C, and a 4°C hold. This reaction yielded a product of approximately 1500 bp, which was resolved on a 0.8% agarose gel together with 0.5 μg of Quick-Load™ 1 kb DNA Ladder (New England Biolabs; Ipswich, MA) and visualised under UV light.

Products from successful PCR reactions were purified using the HighPure PCR Product Purification Kit (Roche ; Mannhein, Germany) according the manufacturer's recommended protocol, and eluted in 50 μl Elution buffer for every 50 μl of PCR product added.

2.6. Generation of Recombinant gp120-Expression Vectors

Gp120-encoding DNA was cloned into the protein expression vector pTriEx-3 (Novagen; Darmstadt, Germany). This vector incorporates three different promoters allowing the expression of protein from sequences inserted into the pTriEx multiple cloning site in mammalian, insect and bacterial cells (see Appendix 2 for pTriEx map). A CMV immediate early enhancer and promoter allows for protein expression in vertebrate cells and a T7 *lac* promoter regulates expression in *E. coli*. pTriEx-3 also contains baculovirus sequences which permit the generation of recombinant baculoviruses by homologous recombination for the infection of insect cells, where the vector's p10 promoter controls protein expression in this host.

2.6.1. Digestion and ligation of pTriEx-3 and gp120-encoding DNA

Approximately 2.5 µg of each purified gp120 amplicon was digested in a 50 µl reaction containing 20 U of *Pag* I (Fermentas, Ontario, Canada) and 20 U of *Xho* I (Fermentas; Ontario, Canada) in 1 x Fermentas buffer O (Fermentas; Ontario, Canada). 3.4 µg of pTriEx-3 stock was also digested in a 50 µl reaction containing 20 U of *Nco* I (Promega; Madison, WI), 20 U of *Xho* I (Promega; Madison, WI) and 1 x Promega buffer D (Promega; Madison, WI). All digestion reactions were incubated for 3 hours in a water bath set at 37˚C.

Digested gp120 inserts were purified with a GFX^{TM} PCR DNA and Gel Band Purification kit (Amersham Biosciences; Freiburg, Germany) according to the manufacturer's protocol and the DNA was eluted in 10-30 µl of dH_2O. The pTriEx digestions were resolved on a 0.8% preparative agarose gel. The linearised plasmid was visualised under ultraviolet light, excised from the gel with a clean razor blade, and the DNA recovered using the Amersham GFX^{TM} PCR DNA and Gel Band Purification kit (Amersham Biosciences; Freiburg, Germany) according to the manufacturer's protocol. pTriEx DNA was eluted in 50 µl pure H_2O. Concentrations of the purified restriction products were determined by UV spectrophotometry.

Ligation reactions were set up using a Quick Ligation^{TM} Kit (New England Biolabs; Ipswich, MA). 50 ng of linearised pTriEx-3 DNA were combined with a 3-fold molar excess of gp120-encoding fragments, and the volume adjusted to 10 µl with dH_2O. 10 µl of 2 x Quick Ligation buffer was added to the mixture, followed by the addition of 1 µl of

Quick T4 DNA Ligase. The reaction was centrifuged briefly and left at room temperature for 10-15 minutes. 5 µl of the ligation mix was then used to transform competent *E.coli* DH5α according to standard protocols (Appendix 1).

2.6.2. Colony Screening Procedures

2.6.2.1. PCR Screening

Initial screening of colonies present on positive ligation plates was accomplished by PCR using Promega PCR Master Mix (Promega; Madison, WI). Colonies were picked and cultured overnight in a shaking incubator at 37˚C in 1 ml Luria Broth containing 0.1 mg/ml ampicillin (Appendix 1). Overnight cultures were diluted 10 times and 2 µl of diluted culture was used in a 20 µl PCR reaction containing 1 x Promega PCR Master Mix, and 0.5 µM each of gp120(F) and gp120(R) primers. PCR conditions were as described in section 2.5 for the amplification of gp120-encoding DNA. PCR products were resolved on a 1% agarose gel together with 0.5 µg of Quick-Load™ 1 kb DNA Ladder (New England Biolabs; Ipswich, MA). The presence of an approximately 1.5 kb product was indicative of the presence of gp120 in the clone.

2.6.2.2. Restriction Mapping

Cultures of clones positive for gp120 by PCR were grown in a shaking incubator overnight at 37˚C in 8 ml of Luria Broth containing 0.1 mg/ml ampicillin. 5 µl of these

cultures were used to make replica plates, and the rest of the culture was used for isolation of recombinant plasmids. Plasmids were isolated using a Sigma GenElute™ Plasmid Miniprep Kit (Sigma; St. Louis, MO) according to the manufacturer's protocol. The presence of plasmid DNA in the elution was confirmed by resolving the samples on a 1% agarose gel and visualizing the DNA under UV light.

The integrity of the clones was investigated using restriction digestion analysis. Each plasmid was linearised using *Xho* I (Fermentas; Ontario, Canada) in order to get an accurate estimation of the size of the plasmid containing the insert (approximately 6.5 kb for successful clones). This was done in 20 µl reactions containing 5 U *Xho* I in 1 x Fermentas Buffer R. As a further control, the gp120 fragment was excised from the plasmid using *Xba* I (New England Biolabs; Ipswich, MA) and *Xho* I (Promega; Madison WI). These reactions were set up as double digests in a volume of 20 µl containing 10 U *Xba* I and 5 U *Xho* I with 100 µg/ml Bovine Serum Albumin (New Englad Biolabs (NEB); Ipswich, MA) in 1 x NEB Buffer 2. All digestion reactions for screening purposes were incubated in a water bath for one hour at 37˚C. The products of these reactions were resolved on a 0.8% agarose gel together with 0.5 µg of Quick-Load™ 1 kb DNA Ladder (New England Biolabs; Ipswich, MA) and visualised under UV light.

2.6.2.3. Sequencing and clone selection

From a panel of 17 recombinant pTriEx-gp120 clones, two representative plasmids were selected for further analysis. These were selected according to *in vitro* viral culture

studies done in our laboratory. Clones originating from one CXCR4-utilising and one CCR5-utilising isolate as determined by phenotypic studies were chosen. The clones selected for further propagation were those originating from 05ZAFV3 (CXCR4-utilising) and 05ZAFV5 (CCR5-utilizing). These clones are hereafter referred to as pTriEx-FV3 and pTriEx-FV5 respectively. Clones of these isolates that had tested positive for the presence of the gp120 insert by PCR, and that showed correct orientation on restriction digestion analysis were selected for sequencing in order to further confirm the integrity and frame of the insert. Sequencing was done by Inqaba Biotech (Pretoria, South Africa) using primers PTF, PTR, PF(2), PR(2), PF(3) and PR(3) as described in table 2.1. The inner primers (PF(2), PR(2), PF(3) and PR(3)) were designed as sequencing progressed, according to sequences generated by the preceding sequencing reactions. Full length gp120 sequences were then assembled and edited using Sequencher v4.6 (GeneCodes, Ann Arbor, MI). Sequences were then inspected for in-frame start and stop codons.

Table 2.1: PCR and cycle sequencing primers

Primer	Sequence
EnvA-1[§]	5'-GGCTTAGGCAT**T**TCCTATGGCAGGAAGAA-3'
EnvB-1[§]	5'- **G**GAAAGAGCAGAAGACAGTGGCAATGA -3'
EnvM-1[§]	5'- TA**A**CCC**A**TCCAGTCCCCCCTTTTCTTTTA –3'
EnvN-1[§]	5'- **T**TGCCAATCAGGGAAG**A**AGCCTTGTGT -3'
gp120-5'[‡ Ψ]	5'- AGA GCA GAA GAC AGT GGC AAT GA-3'
TU-A[‡ Ψ]	5'- TAT TAT GGG GTT CCT GTG TGG-3'
ES21[‡ Ψ]	5'- ACA CAT GCC TGT GTA CCC ACA G-3'
ED5[‡ Ψ]	5'-ATG GGA TCA AAG CCT AAA GCC ATG TG-3'
A590[‡ Ψ]	5'-AAT CGC GAA ACC AGC CGG CGC ACA AT-3'
Env1f[‡ Ψ]	5'-CCA TAA CAC AAG CCT GTC CAA AGG-3'
JL74[‡ Ψ]	5'-ACA TGT GGA AAA ATA ACA TGG TAG AAC AG-3'
Envf*[‡ Ψ]	5'CTG TAG AAA TTG TGT GTA CAAGAC CC-3'
AES6[‡ Ψ]	5'-GGA CAA GCA TTC TAT GCA ACA GGT G-3'
AA1570*[‡ Ψ]	5'-GGA GCA GCA GGA AGC ACT ATG GGC-3'
Env3F[‡ Ψ]	5'GGA AGC ACT ATG GGC GCG GC-3'
JL103[‡ Ψ]	5-TAA CAA ATT GGC TGT GGT ATA TAA-3'
ZM184D[‡ Ψ]	5'CCA CTC AGC TAC TGC TAT TGC TAT GGT-3'
Env4f[‡ Ψ]	5-GAG TTA GGC AGG GAT ACT CAC C-3'
TU-J[‡ Ψ]	5'-GTT AGG CAG GGA TAC TCA CC-3'
gp120(F)[¥]	5'-TCG ATC TCA TGA GAG TGA TGG GGA TAC AGA GG–3'
gp120(R)[¥]	5'-CCT CGA GTT ATC TTT TTT CTC TCT CCA CCA CTC TCC T-3'
PTF[¶]	5'- GTT ATT GTG CTG TCT CAT C-3'
PTR[¶]	5'- TCG ATC TCA GTG GTA TTT GTG-3'
PF(2)[¥]	5'-AAA TTG CTC TTT CAA TGT AAC CA-3'
PR(2)[¥]	5'-TCT ACA ATT AAA GCT ATG TGT TG-3'
PF(3)[¥]	5'- ACA CAT AGC TTT AAT TGT AGA G-3'
PR(3)[¥]	5'- AAG TGA CAC AGA GTG GG-3'
RTF[¥]	5'-ATG GGG CAT CTT AGG CTT TTG G-3'
RTR[¥]	5'-GTG GGT ACA CAG GCA TGT GT-3'

[§]Adapted from Gao *et al*[173]; changes from the original primers are bold and underlined

[‡] F. McCutchan, personal communication with Dr. M Papathanasopolous

[¥] Designed during this study

[¶]Source: Novagen, Darmstadt Germany

*These primers were only used on isolates 05ZAFV8, 05ZAFV9, 05ZAFV27, 05ZAFV28

[Ψ]These primers were used for sequencing of gp160-coding regions

2.7. Gp120 expression in bacteria

2.7.1. Transformation of *E. coli* BL21

As a pilot experiment designed to further confirm the presence a functional open reading frame in the clones, expression of gp120 in a bacterial system was assessed. Competent *E.coli* BL21 (Stratagene; La Jolla, CA) (Appendix 1) were transformed with approximately 0.075 µg pTriEx-FV3 according to standard protocols (Appendix 1). Single colonies were selected for inoculation of Luria Broth containing both 0.1 mg/ml chloramphenicol (Sigma; St Louis, MO) (Appendix 1) and 0.1 mg/ml ampicillin. This culture was grown overnight in a shaking incubator at 37˚C, and 5 µl was used to construct BL21-pTriEx-FV3 replica plates containing 0.1 mg/ml each of ampicillin and chloramphenicol.

2.7.2. Induction of FV3 gp120 expression in *E. coli* BL21.

For protein expression experiments, 8 ml of Luria Broth with 0.1 mg/ml each of ampicillin and chloramphenicol was inoculated with a single colony from the BL21-pTriEx-FV3 replica plate and grown overnight in a shaking incubator at 37˚C. As a negative control, untransformed *E. coli BL21* were grown up overnight in 8 ml Luria Broth containing 0.1 mg/ml chloramphenicol. The following morning, the cultures were split between two tubes, diluted 10 times in Luria Broth containing antibiotics, and grown for a further hour in a shaking incubator at 37˚C. A 1 ml sample was then taken from

each tube (time = 0). Isopropyl-β-D-thiogalactopyranoside (IPTG) (Roche; Mannheim, Germany) was then added to one tube at a final concentration of 1 mM in order to induce protein expression by the T7 promoter. IPTG was omitted in negative control cultures. 1 ml samples were taken from each tube on an hourly basis for 4 hours. Bacterial culture samples were centrifuged for 1 minute at 13 000 rpm in a Biofuge Pico (Heraeus; Hanau, Germany), and the pellets resuspended in 75 µl of PBS. Pellets were analysed for the presence of gp120 by SDS-PAGE and Western Blotting analysis as described in section 2.8.

2.8. Gp120 Assays

2.8.1. Antibodies

The following reagents were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 IIIB gp160 Monoclonal Antibody (Chessie 13-39.1) from Dr. George Lewis[175]; HIV-1 gp120 Monoclonal Antibody (ID6) from Dr. Kenneth Ugen and Dr. David Weiner;[176, 177]. NEA 9201 and NEA 9301 were obtained from Perkin Elmer (Wellesley, MA). ECL™ Anti-Human IgG, HRP linked whole antibody (from sheep) (Anti-human) and ECL™ Anti-Mouse IgG, HRP linked whole antibody (from sheep) (Anti-mouse) were obtained from Amersham Biosciences (Freiburg, Germany)

2.8.2. Sample preparation and SDS-PAGE

Clarified cell culture supernatants (or resuspended pellets in the case of bacterial expression) were added to equal volumes of 2 x treatment buffer (Appendix 1), and boiled for five minutes. Samples were resolved on 8% SDS-polyacrylamide gels according to a standard protocol (Appendix 1). Gels were then either stained with Coomassie® Brilliant Blue R-250 (Sigma; St Louis, MO), silver stained, or used for Western Blotting.

2.8.3. Staining of SDS-PAGE gels

Gels were fixed for an hour in a 40% Methanol, 10% Acetic Acid solution. Following this, gels were either silver stained using the Bio-Rad Silver Stain Kit (Bio-Rad; Hercules, CA) according to the manufactirer's recommended protocol, or stained with Coomassie® Brilliant Blue R-250 (Appendix 1).

2.8.4. Western Blot analysis of gp120 expression

Western blotting of gels was performed according to a standard protocol (described in Appendix 1).

A number of alternatives were used throughout this study as the primary antibody solution in order to detect expressed FV3 and FV5 gp120. The first alternative was using

human plasma as a primary antibody solution. Human plasma from HIV-1 positive individuals was obtained from the South African National Blood Service and heat-inactivated at 60°C for one hour before use. Membranes were shaken gently for 1 hour at room temperature in 20 ml of a 1:1000 dilution of plasma. A second alternative was the use of the monoclonal antibody 9201 diluted in 2 ml T-TBS to a concentration of 1:1000 and placed on top of the nitrocellulose membrane for 1 hour at room temperature. 9201 is a mouse monoclonal antibody directed against an epitope near the C-Terminus of gp120. It was raised against the peptide MRDNWRSELYKY[178], although the epitope has also been noted as the slightly overlapping sequence GGGDMRDNWRSE[179] . The third alternative for detection of gp120 was the use of a cocktail of three mouse monoclonal antibodies, ID6, Chessie13-39.1, and 9301 each at a concentration of 1:500, made up in 2 ml T-TBS and incubated on the membrane for 1 hour at room temperature. 9301, which replaced 9201 when it was discontinued, is directed against the epitope MRDNWRSELYKY[180]. Chessie 13-39.1 is directed against LAI gp120 amino acids 252-273: RVVSTQLLLNGSLAEEEVVIR[175], while ID6 is directed against the first 204 amino acids of LAV-1 gp160[176, 177].

2.9. gp120 Expression in Sf9 insect cells

2.9.1. Insect cell culture

Once the presence of a functional open reading frame in the pTriEx-gp120 clones was confirmed, preparations were made for the expression of fully glycosylated recombinant gp120 in Sf9 insect cells.

2.9.1.1. Insect cell culture in monolayers

Recommended cell lines for Insect cell expression of recombinant protein using pTriEx are the Ready-Plaque™ Sf9 and TriEx Sf9 cells (Novagen; Darmstadt, Germany). Ready-Plaque™ Sf9 cells were maintained in serum-free BacVector Insect Cell medium (Novagen, Darmstadt, Germany) with no additives. TriEx Sf9 cells were maintained in serum-free TriEx Insect Cell Medium (Novagen, Darmstadt, Germany) with no additives. Both TriEx and BacVector Insect cell media are specifically optimized for cell growth and protein expression in these cell lines. Cells were maintained in monolayers at 28°C and split approximately once every seven days when confluent. The splitting procedure involved pouring off the old medium, washing the cells once in PBS (Sigma; St Louis, MO) and flushing them off the side of the flask using a pipette and fresh medium. One tenth of these cells were then added to a new flask in a suitable volume of fresh medium.

2.9.1.2. Insect cell suspension culture

TriEx Sf9 cells were grown until confluent as a monolayer in a 175 cm^2 tissue culture flask (Nunc; Roskilde, Denmark). Cells were then washed with PBS and resuspended in 100 ml of TriEx medium in a conical flask at a concentration of 5 x 10^5 cells/ml. Cells were incubated at 28°C, shaking at 150 rpm.

2.9.2. Liquid Overlay Transfection Procedure

Liquid overlay transfections were done based on the manufacturer's recommended protocol. The basis for the production of recombinant baculoviruses is homologous recombination between sequences in the backbone baculoviral DNA supplied in trans by cotransfection and recombinant pTriEx-3 plasmids. For each transfection, a 25 cm^2 flask was seeded with total of 2.5 x 10^6 Ready Plaque™ Sf9 cells in 5 ml of BacVector Medium. Cells were left to attach to the flasks for at least 20 minutes. For each transfection, the following components were assembled in a sterile 6 ml polystyrene tube (in order): BacVector medium (25 µl less the amount of DNA added to the reaction), 100 ng of of BacVector®-3000 Triple Cut Virus DNA (Novagen, Darmstadt, Germany) and 500 ng of recombinant pTriEx-FV3/FV5. The negative control was set up in an identical manner, substituting BacVector medium for the recombinant plasmid. For each transfection, 5 µl of Eufectin™ Transfection Reagent (Novagen, Darmstadt, Germany) was added to 20 µl of Nuclease-free Water in a separate sterile polystyrene tube. The entire 25 µl DNA mixture was immediately added to this tube, swirled to mix and incubated at room temperature for 15 minutes. During the incubation, cells in the 25 cm^2

flasks were washed twice with BacVector medium. After the final wash, 1 ml of medium was added to each flask.

After the DNA/Eufectin incubation was complete, 450 µl BacVector medium was added to the mixture and gently mixed. 100 µl of this mixture was then added to 400 µl medium in a separate tube to constitute a 1/5 dilution. The DNA mixtures were then added to the 1 ml of medium covering the cells in the flasks and the flasks were then incubated for 1 hr at room temperature. During this hour, even and consistent distribution of the transfection mixture was achieved by gently swirling the flask every 15-20 minutes. After one hour, 6 ml of BacVector medium containing 5% Foetal Bovine Serum (FBS) (Invitrogen, Carlsbad, CA) was added to the flasks, which were then incubated at 28°C for 5 days. After 5 days, the medium was harvested from the cells, centrifuged at 300 x g for 5 minutes in an Eppendorf 5810R centrifuge (Eppendorf; Hamburg, Germany) and recombinant baculovirus (Bac-FV3/Bac-FV5)-containing supernatants were used to reinfect Ready Plaque^{TM} Sf9 cells.

2.9.3. Re-infection using the transfection supernatant

Successive rounds of reinfection with Bac-FV3 and Bac-FV5 were performed in order to increase baculoviral titers. 6-well culture dishes (Nunc, Roskilde, Denmark) were seeded with 1 x 10^6 Ready Plaque^{TM} Sf9 cells in 4 ml of BacVector medium. 1 ml of each transfection supernatant was added to each well and plates were incubated at 28°C. After 5 days, the medium was harvested from the cells and centrifuged at 300 x g for 5 minutes

in an Eppendorf 5810R centrifuge (Eppendorf; Hamburg, Germany). The supernatants were then assayed for protein expression as described in section 2.8. Successive rounds of reinfection were performed until supernatants were positive for gp120 by Western Blot analysis (2-3 rounds of reinfection were usually required to obtain a sufficient viral titer to result in detectable protein expression).

2.9.4. Plaque purification and amplification of baculoviral clones

2.9.4.1. Plaque purification

Plaque assays were performed according to a protocol adapted from the manufacturer (Novagen, Darmstadt, Germany). 6-well culture plates were seeded with 1 x 10^6 Ready Plaque^{TM} Sf9 cells per well (in 2 ml BacVector medium) and incubated at room temperature for at least 20 minutes in order to allow the cells to adhere. During this incubation period, serial dilutions of the virus-containing supernatant were made. Plaque assays for purification purposes required baculoviral dilutions of 5, 50, 500, 5 000 and 50 000 times in BacVector medium. Once the dilutions were set up, the medium was aspirated from the cells and 900 µl of each dilution was added to its allocated well. 900 µl of fresh medium was added to the sixth well of the plate to serve as a negative control. Plates were then incubated at room temperature for 1 hour. During the incubation period, the agarose overlay was prepared from an autoclaved 3% BacPlaque Agarose (Novagen, Darmstadt, Germany) stock. BacVector medium containing 5% FBS was added to the melted BacPlaque Agarose stock to a final agarose concentration of 1%. The agarose was

kept at 37°C in a water bath until further use. After the 1 hour incubation, the baculoviral dilutions were aspirated from the cell monolayer and 1 ml of the agarose overlay was added gently to the side of each well and allowed to set. 1 ml of BacVector medium containing 5% FBS was then added to each well. Plates were incubated at 28°C on damp tissue paper in a closed container for 2-5 days, until plaques were visible.

2.9.4.2. Plaque Staining

For plaque staining, a 0.33% (w/v) stock solution of Neutral Red (kind gift of Dr Clem Penny) was diluted 1:13 with sterile PBS just before use. The liquid overlay was aspirated from the plates, and 1 ml of diluted neutral red was added to the centre of each well and incubated at 28°C for 2 hours. After 2 hours, the neutral red was aspirated and plates were left at room temperature to allow the cells to take up the stain overnight. Only live cells are stained and thus plaques can be visualised under a microscope as grey clearings in a pinkish red background.

2.9.4.3. Plaque picking and amplification

Several Bac-FV3 and Bac-FV5 plaques visualised by neutral red staining were removed using a sterile glass Pasteur pipette. Plaque-containing agarose plugs were placed in 1 ml of BacVector medium and incubated at 4°C overnight in order to allow baculoviruses to diffuse out of the agarose plug and into the medium. 400 µl of this medium was then used in successive rounds of reinfection in order to amplify each virus as described in section

2.9.3. The plaque assay was then repeated in order to refine the selection of single baculoviral clones.

2.9.4.4. Generation of high-titer baculoviral stocks

After the second round of purification, high-titer stocks of Bac-FV3 and Bac-FV5 were generated. Single, well-defined plaques were selected after the second plaque purification step by withdrawal in a sterile glass Pasteur Pipette. These agarose plugs were then each placed in 1 ml medium and incubated at 4°C overnight in order to allow the baculoviruses to diffuse out (Bac-FV3' and Bac-FV5'). The following morning, 25 cm^2 flasks were seeded with 5 ml Ready-Plaque™ Sf9 cells at a density of 5 x 10^5 cells/ml and left for 20 minutes in order to allow the cells to attach. Following attachment, the medium was aspirated from the cells and 300 µl of Bac-FV3' or Bac-FV5'was added to the cells, which were allowed to incubate at room temperature for 1 hour. The medium was swirled over the cells twice during this time. After 1 hour, 5 ml fresh BacVector medium was added to the cells, which were then left for 5 days at 28ºC.

After 5 days, supernatants (Bac-FV3'' and Bac-FV5'') were harvested from these cells and clarified by centrifugation. To expand the baculoviral stock, 75 cm^2 flasks were seeded with 10 ml of Ready-Plaque™ Sf9 cells at a density of 2 x 10^6 cells/ml. After allowing the cells to attach for 20 minutes, the medium was aspirated and 500 µl Bac-FV3'' or Bac-FV5'' were added to the flasks. The flasks were incubated for 1 hour, with the liquid being swirled over the monolayers twice during this time. 10 ml of fresh

BacVector medium was then added to the cells, which were left for 5 days at 28ºC. This generated high titer Bac-FV3 and Bac-FV5 stocks to be used for protein expression studies.

2.9.5. Baculoviral titering

2.9.5.1. Plaque assay

Plaque assays were set up in order to determine Bac-FV3 and Bac-FV5 titers. These were performed as described in section 2.9.4.1. Dilutions of baculovirus stocks were optimized for each individual experiment, and ranged from 1 in 1000 to 1 in 10 000 000.

2.9.5.2. Real time PCR

A quantitative real-time PCR assay was established as an alternative method for quantification of baculoviral titers. Baculoviral DNA was extracted from 200 µl of clarified supernatant by automated extraction procedures using a MagNA Pure LC (Roche; Mannheim, Germany) and a MagNA Pure LC Total Nucleic Acid Isolation Kit (Roche, Mannheim, Germany) according to the manufacturer's instructions.

Baculoviral DNA was amplified using the gp120-specific primers PTR and PTF (table 2.1), which amplify an approximately 180 bp segment in the C1 region of gp120. These primers were designed based on the sequences of the clones FV3 and FV5. A standard

curve was set up for each clone, using pTriEx-FV3/FV5 at known concentrations ranging from 10^2 to 10^7 copies/µl.

PCR reactions were set up in a final volume of 20 µl containing 10 µl of SYBR® Green JumpStart Taq Readymix™ (Sigma; St. Louis, MO), 2 µl of extracted DNA or diluted plasmid and 0.5 µM of each primer (PTR and PTF). PCR was performed on a LightCycler® platform (version 1.2) (Roche Applied Sciences ; Mannheim, Germany).

PCR began with a 30 second hotstart at 95°C, followed by 40 cycles consisting of 10 seconds at 95°C, 10 seconds at 57°C and 15 seconds at 72°C, with a single acquisition of fluorescence signals after each 72°C extension. The rate of temperature change was 20°C/second. After the PCR program, a melting curve was established by raising the temperature of the reaction to 95°C for 5 seconds, dropping it to 40°C for 30 seconds and then raising the temperature gradually back to 95°C at a rate of 0.15°C/second. Fluorescence data was acquired continuously throughout this program. Fluorescence was detected using channel F1 on the LightCycler® instrument. Quantitative analysis (by the Second Derivative Maximum Method), including the construction of standard curves was done using the LightCycler® Software version 3.5.3.

2.9.6. Expression of recombinant gp120

Bac-FV3 and Bac-FV5-containing supernatant was added to suspension cultures of TriEx Sf9 cells (2 x 10^6 cells/ml) at an MOI of 3 gp120 copies/cell. Suspension cultures were

incubated for three days before the protein-enriched supernatants were harvested. Cultures were centrifuged at 300 x g for 8 minutes. Complete Protease Inhibitor Cocktail tablets (Roche; Mannheim, Germany) were added to the clarified supernatant (1 tablet/50 ml supernatant), which was then assessed for the presence of gp120 by SDS-PAGE and Western Blotting analysis as described in section 2.8.

2.10. Protein Purification and concentration

Gp120-containing supernatants were purified using the Bio-Rad BioLogic LP chromatography system. Absorbance data was captured by the LP Data View™ v1.03 software (Bio-Rad; Hercules, CA)

2.10.1. Purification of recombinant gp120

In this section, "PBS" refers to Dulbecco's Phosphate Buffered Saline without calcium chloride and magnesium chloride (Sigma; St Louis, MO).

2.10.1.1. Lectin-Affinity Purification

Columns containing lectin from *Galanthus Nivalis* immobilized on cross-linked 4% beaded agarose (Sigma; St Louis, MO) were equilibrated with PBS containing 0.5 M NaCl. Clarified supernatants containing recombinant FV3 or FV5 gp120 were loaded

onto columns at a flow rate of 0.2 ml/minute. Columns were then washed with 15 column volumes each of PBS containing 0.65 M NaCl, PBS, PBS containing 1 M NaCl and PBS again. 15 column volumes of elution buffer (20 mM sodium acetate, 0.1 mM EDTA, 0.5 M methyl α-D-mannopyranoside (MMP) (Sigma; St Louis, MO), pH 5.0) were then passed through the columns at 2.0 ml/minute in order to elute the protein. The eluates were collected in 1 ml fractions and fractions corresponding to A_{280} peaks as recorded by the LP Data View™ v1.03 were assayed for the presence of gp120 as described in section 2.8. Fractions that tested positive for the presence of gp120 by Western Blot analysis were pooled and dialysed three times against 50 volumes of 50 mM sodium acetate buffer, pH 5.0 for 5 hours at 4ºC.

2.10.1.2. Ion Exchange Chromatography

Following dialysis of affinity-purified gp120s, ion exchange chromatography was performed. Briefly, ion exchange columns were poured using SP Sepharose™ Fast Flow (Amersham Biosciences; Freiburg, Germany) and equilibrated with 50 mM sodium acetate, pH 5.0. Dialysed gp120-containing fractions eluted from the lectin column were loaded onto columns at a flow rate of 0.1 ml/minute. Columns were then washed sequentially with 15 column volumes each of 50 mM Sodium Acetate buffer, first at pH 6.0 and then at pH 7.0. Bound gp120 was then eluted using a NaCl gradient increasing from 0 to 400 mM in a 50 mM Sodium Acetate buffer, pH 7.0 over a volume of 40 ml. Fractions corresponding to A_{280} peaks as recorded by the LP Data View™ v1.03 were assayed for gp120 as described in section 2.8. Fractions positive for gp120 by Western

Blot analysis were pooled and dialysed three times for 5 hours each against 50 volumes of PBS (pH 7.4) at 4°C.

2.10.2. Sample Concentration

Dialysed gp120-containing fractions were concentrated using Amicon® Ultra-15 Centrifugal Filter devices (Molecular Weight cut off 50 kDA) (Millipore; Billerica, MA). Briefly, the columns were prewashed by adding 15 ml of PBS (pH 7.4) and centrifuging at 3200 x g for 5 minutes. The dialysed eluate was then added and, if necessary, the volume was brought to 15 ml with PBS. Columns were then centrifuged at 3200 x g for 5 minutes. Where the original volume of dialysed eluate was greater than 15 ml, this step was repeated in the same column until all of the eluate had been passed through. The concentrated retentate was then recovered from the filter unit. Protein was quantitated using the BCA^{TM} Protein Assay Kit (Pierce; Rockford, IL), using a slightly adapted version of the manufacturer's recommended protocol (described in appendix 1) and stored at -70°C.

2.11. Protein Characterisation

2.11.1. Silver Staining and Coomassie® Staining

In order to determine the level of purity of the concentrated protein, 1 µg and 1.5 µg of concentrated gp120 were loaded onto two 1 mm thick 8% SDS-PAGE gels, which were

resolved as described in appendix 1. Following this, the gel loaded with 1 µg gp120 was silver stained and the gel loaded with 1.5 µg gp120 was stained with Coomassie® Brilliant Blue R-25, both as described in section 2.8. The Coomassie®-stained gel was then analysed using a Bio-Rad Chemidoc XRS and Bio-Rad Quantity One 1-D Analysis Software for densitometric analysis in order to determine the percentage purity of the protein.

2.11.2. Monoclonal antibody binding to recombinant gp120

2.11.2.1. Materials

The following reagents were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: HIV-1 IIIB gp160 Monoclonal Antibody (Chessie 13-39.1) from Dr. George Lewis[175]; HIV-1 gp120 Monoclonal Antibody (ID6) from Dr. Kenneth Ugen and Dr. David Weiner;[176, 177] HIV-1 gp120 Monoclonal Antibody (654-30D) from Dr. Susan Zolla-Pazner[181-184]; HIV-1_{LAI} (BRU) V3 Monoclonal Antibody (4G10) from Dr. Albrecht von Brunn, Courtesy of the MRC AIDS Directed Programme[185, 186]; HIV-1 gp120 Monoclonal Antibody (F425 A1g8) from Dr. Marshall Posner and Dr. Lisa Cavacini; HIV-1 gp120 Monoclonal Antibody (670-30D) from Dr. Susan Zolla-Pazner[184, 187]; HIV-1 gp120 Monoclonal Antibody (F425 B4e8) from Dr. Marshall Posner and Dr. Lisa Cavacini; HIV-1 gp120 Monoclonal Antibody (48d) from Dr. James Robinson[188, 189]; HIV-1_{LAI} (BRU) V3 Monoclonal Antibody (5F7) from Dr. Albrecht von Brunn, Courtesy of the MRC AIDS Directed

Programme[185, 186]; HIV-1 gp120 Monoclonal Antibody (2G12) from Dr. Hermann Katinger[190-194]; HIV-1 gp120 Monoclonal Antibody (F015) from Dr. Marshall Posner and Dr. Lisa Cavacini[195-198]; HIV-1 gp120 Monoclonal Antibody (F425 B4a1) from Dr. Marshall Posner and Dr. Lisa Cavacini; HIV-1 gp120 Monoclonal Antibody (17b) from Dr. James E. Robinson[73, 103, 189, 199-201]; HIV-1 gp120 Monoclonal Antibody (IgG1 b12) from Dr. Dennis Burton and Carlos Barbas[202-205]; HIV-1_{BaL} gp120 from DAIDS, NIAID; HIV-1 gp120 CM; HIV-$1_{96ZM651}$ gp120 from DAIDS, NIAID. NEA 9301 was obtained from Perkin Elmer (Wellesley, MA). sCD4 was obtained from Progenics (Tarrytown, NY).

2.11.2.2. Characterisation of antibody-binding properties

All ELISA's were done in F96 Maxisorp Nunc-Immunoplates (Nunc; Roskilde, Denmark). All washes were done five times with 300 µl of PBS-T (0.05% Tween-20 in PBS) per well per wash on a EL_X50 Auto Strip Washer (Bio-Tek; Winooski, VT). Wells were coated with 100 ng of gp120 (Bal, CM, 96ZM651, FV3 or FV5) diluted in PBS (pH 7.4), and incubated for three hours at room temperature in the dark. Gp120 was then aspirated from the wells and 300 µl of PBS-TB (PBS with 0.05% Tween-20 and 10 mg/ml Bovine Serum Albumin Fraktion V (BSA)) was added to each well. Plates were incubated at 4°C overnight. Wells were then washed and 100 µl of primary antibody diluted in PBS-T was added to each well. The epitope for each antibody, as well as the concentration each was used at is described in table 2.2. After addition of the primary antibody, plates were incubated for 2 hours at 4°C and then washed. 100 µl of secondary

antibody diluted in PBS-T was then added to each well at a concentration of 1:2000. ECL™ Anti-Human IgG, HRP linked whole antibody (Amersham Biosciences; Freiburg, Germany) was used for the human antibodies F105, 2G12, 654-30D, F425 A1g8, F425 B4a1, F425 B4e8, 670-30D, 17b, IgG1 b12 and 48d. ECL™ Anti-Mouse IgG, HRP linked whole antibody (Amersham Biosciences; Freiburg, Germany) was used for the mouse antibodies 4E10, 5F7, Chessie 13-39.1 and ID6. Plates were incubated at 4°C with the secondary antibody for one hour, and then washed. 100 µl of 1-Step™ Ultra-TMB-ELISA Substrate (Pierce; Rockford, IL) was then added to each well, and the plate was incubated for 30 minutes at room temperature to allow a colour reaction to develop in the wells. After 30 minutes, the reaction was stopped by the addition of 2 M Sulphuric Acid and the absorbance of the wells at 450 nm was read on a Bio-Rad Model 680 Microplate Reader (Bio-Rad; Hercules, CA).

2.11.3 Characterisation of CD4-binding properties

Wells of a 96-well micro-titer plate were coated for 3 hours at room temperature with 100 ng of sCD4 diluted in PBS. sCD4 was then removed by automated aspiration and 300 µl of PBS-TB was added to each well before plates were incubated at 4°C overnight. Wells were then washed and 150 ng of gp120 (Bal, CM, 96ZM651, FV3 or FV5) diluted in PBS-T was added to each well in a volume of 100 µl. The plate was incubated for two hours at room temperature and then washed. 100 µl of primary antibody diluted in PBS-T was then added to each well as follows: 2G12-1:2 000 (38 nM), 48d-1:100, F425 A1g8-1:1000 (13.4 nM), 17b-1:100 (21 nM). After two hours at 4°C, plates were washed and a

anti-human secondary antibody, diluted 1:2000 in PBS-T was then added. Plates were incubated at 4°C with the secondary antibody for one hour, washed and developed with 1-Step™ Ultra-TMB-ELISA Substrate as described in section 2.11.2.2.

Table 2.2: Table of antibodies used for protein characterisation, the concentration and molarity antibodies were used at (where available), and their epitopes

Antibody	Concentration	Molarity	Epitope
F425 A1g8	1:2000	6.7 nM	CD4i, binding enhanced by sCD4
F425 B4e8	1:250 000	0.04 nM	V3, I309, R313, F315 of $gp120_{JRCSF}$
F425 B4a1	1:250 000	0.12 nM	V3 loop
2G12	1:200 000	0.38 nM	Glycan-dependent
17b	1:10 000	0.21 nM	Overlaps binding site for CCR5
IgG1 b12	1:40 000	0.23 nM	Overlaps CD4 binding site
48d	1:100	TCS	Complex, binding enhanced by CD4
654-30D	1:2 000	0.16 nM	Intact tertiary structure, blocks cd4 binding
670-30D	1:200	0.18 nM	C5 region, PTKAKRR
F105	1:50 000	0.21 nM	Conformational, overlaps CD4 binding site
ID6	1:1000	TCS	$gp120_{MN}$ Aa86-100, VNVTENFNMWKNNMV
Chessie 13-39.1	1:1000	TCS	C2 region, RVVSTQLLLNGSLAEEEVVIR
NEA9301	1:1000	TCS	C5 region, MRDNWRSELYKY
4G10	1:100	TCS	V3, TRKSIRIQRGPGRAFVTIGKIGNMR
5F7	1:100	TCS	V3, TRKSIRIQRGPGRAFVTIGKIGNMR

TCS: Antibody provided at undetermined concentration in a tissue culture supernatant.

3. Results

3.1. PCR amplification of Env-encoding regions

The *Env* regions of 20 viral isolates were amplified, sequenced and analysed in order to obtain background information about this cohort and enable the selection of two representative isolates for cloning and production of recombinant gp120. *Env*-encoding DNA was amplified from proviral DNA isolated from PBMCs infected with primary isolates of HIV-1 (figure 3.1a). After several rounds of optimization, gp160 was successfully amplified by nested PCR from 20 isolates, namely FV2, FV3, FV5-FV15, FV20, FV22, FV23 and FV25-FV28. These PCR products were used for sequencing, as well as in PCR reactions to amplify the gp120-encoding regions of *Env* for cloning purposes (figure 3.1b). 19 gp120-coding regions were amplified from these products. We were unable to amplify FV28, and this is likely related to significant differences observed between the reverse primer and the 3' gp120 primer region of this isolate.

3.2. Population-based sequencing of the gp160 coding regions

Sequence data was generated for all 20 full length gp160s. The 20 *env* sequences were submitted to GenBank using Sequin V5.35[206] and are available under the accession numbers DQ382361 to DQ382380. Nucleotide sequence analysis showed that the full-length *env* genes for the 20 isolates ranged from 2526 to 2634 bp and revealed that all 20 isolates had intact open reading frames, confirming the presence of functional *env* genes.

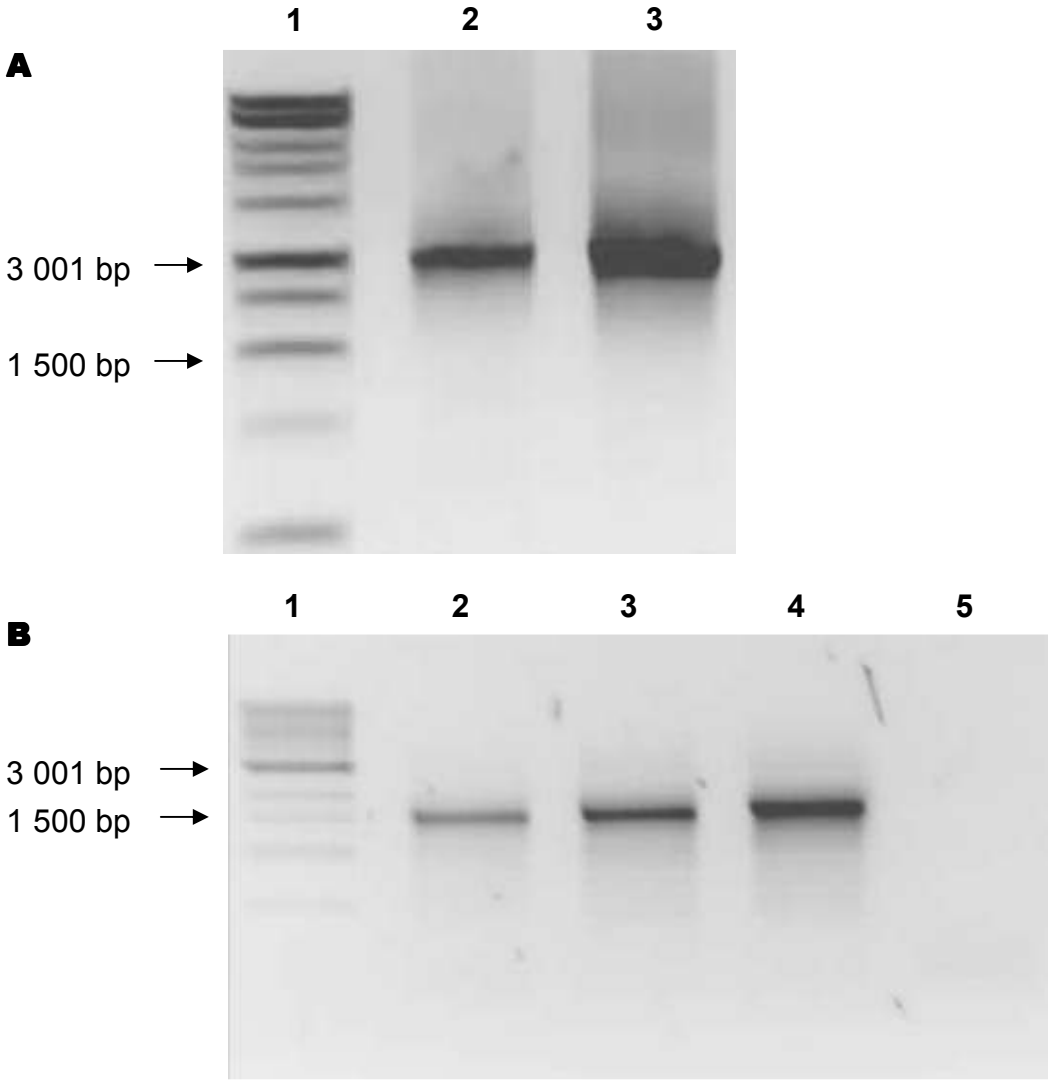


Figure 3.1: Agarose gel electrophoresis of *Env* PCR products. PCR products were resolved on 0.8% agarose gels with the Molecular Weight (MW) Marker (Quick-Load 1 kb DNA Ladder (New England Biolabs, Ipswich, MA) and visualised under UV light. **(A) Representative gp160 PCR.**: Lane 1: MW marker, Lane 2: gp160 (FV6) PCR product before purification, Lane 3: gp160 (FV6) PCR product after purification with a GFX^{TM} PCR DNA and Gel Band Purification kit. **(B) Representative gp120 PCR.** Lane 1: MW marker, Lane 2: gp120 amplicon (FV6), Lane 3: gp120 amplicon (FV9), Lane 4: gp120 amplicon, (FV13), Lane 5: Negative Control. Similar results were obtained for all other *Env* PCR reactions.

Phylogenetic analysis of the 20 full-length *env* nucleotide sequences, together with reference viruses from the major subtypes revealed that 19 isolates clustered within HIV-1 subtype C with a bootstrap value of 100%, while one isolate (05ZAFV10) showed evidence of recombination (Figure 3.2). The FV10 sequence was further analyzed using the RIP 2.0 Program[45], which showed that it is a unique CD recombinant. There are two recombination breakpoints at approximate nucleotide positions of 160 and 1850. Segments within positions 1-160 and 1850 to 2559 were identified as subtype C, while the segment between positions 160 and 1850 was subtype D. Phylogenetic analysis also indicated that there were no differences in clustering between the CCR5-using isolates and those able to utilize CXCR4. Coreceptor usage of the 20 isolates was determined phenotypically in our laboratory by Ms. B Connell. This was performed by measuring viral replication on U87.CD4, MT-2 cells and Cf2Th-CD4 cells expressing either the CCR5 or CXCR4 coreceptors independently[124]. FV3, FV10, FV14 and FV26 were CXCR4-using, FV15 and FV27 were dual-tropic and the rest of the isolates used CCR5 as a means of gaining entry into host cells.

The amino acid sequences were derived and aligned using GeneDoc[207] and compared to the consensus C and ancestral C sequences from Los Alamos[45]. The full length amino acid alignment can be seen in figure 3.3. The predicted proteins ranged from 842 to 878 amino acids in length, with the variation being primarily due to insertions/deletions in the variable regions (V1 toV5) of gp120, and deletions in the gp41 cytoplasmic tail region.

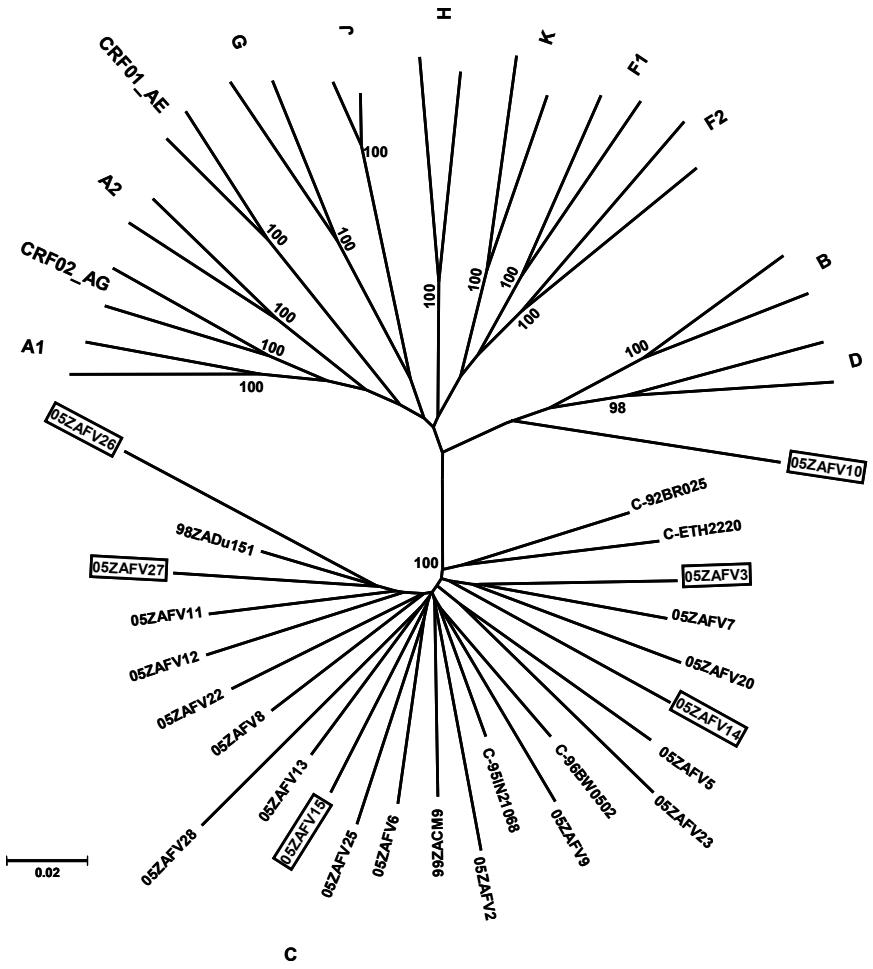
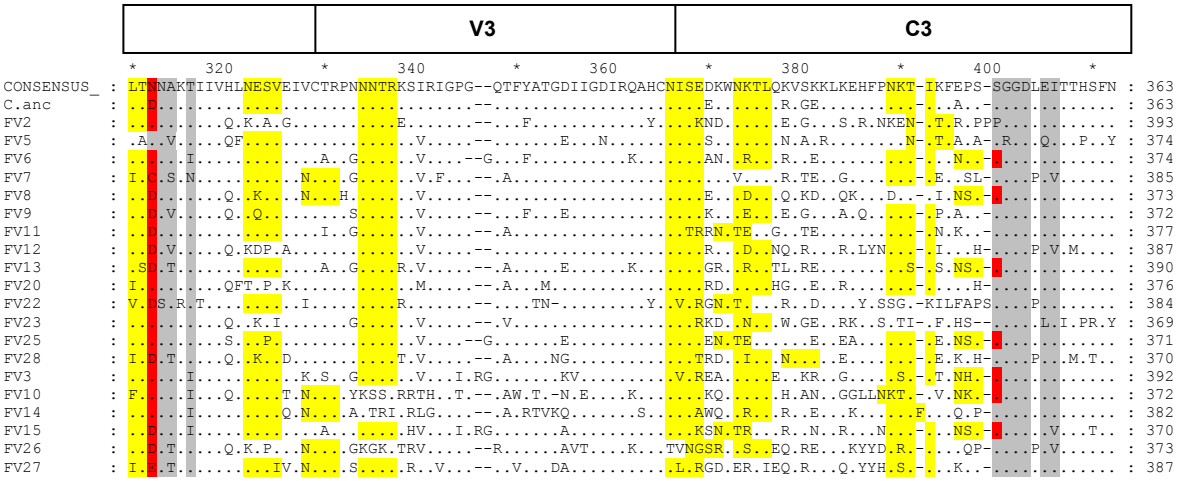
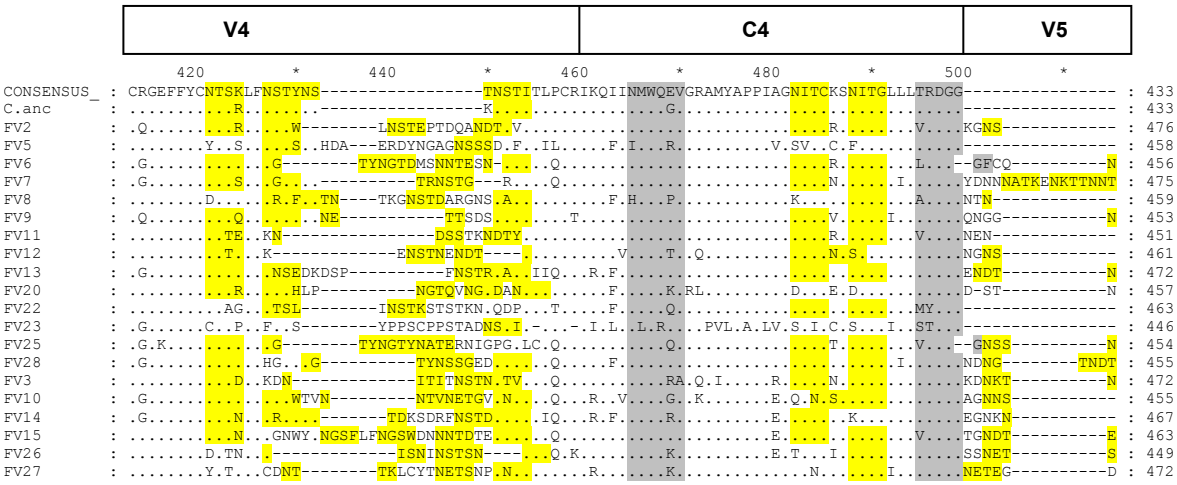


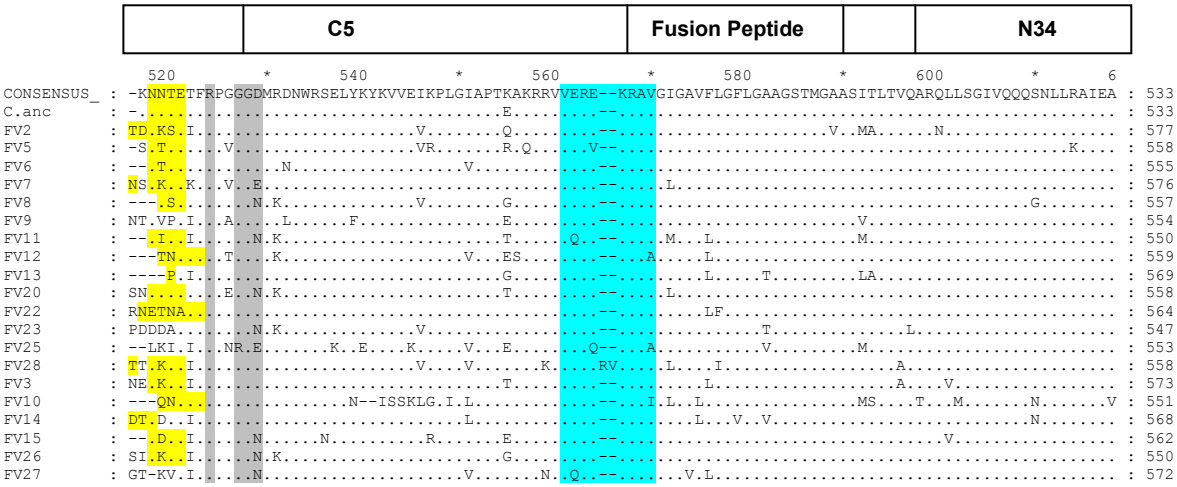
Figure 3.2: Phylogenetic relationships of the full-length gp160s from 20 South African subtype C isolates with HIV-1 subtype reference sequences from the Los Alamos database[45]. The 20 isolates are named 05ZAFV$_n$, and sequences from isolates able to use the CXCR4 coreceptor are boxed. The phylogenetic tree was constructed from nucleotide sequences, using the neighbour-joining method with maximum parsimony bootstrap values to estimate the stability of the nodes. Only bootstrap values of 70% or higher are shown. Figure courtesy of Dr M Papathanasopoulos.


Signal Peptide
C1
V1
V2
C2





end gp120, start gp41



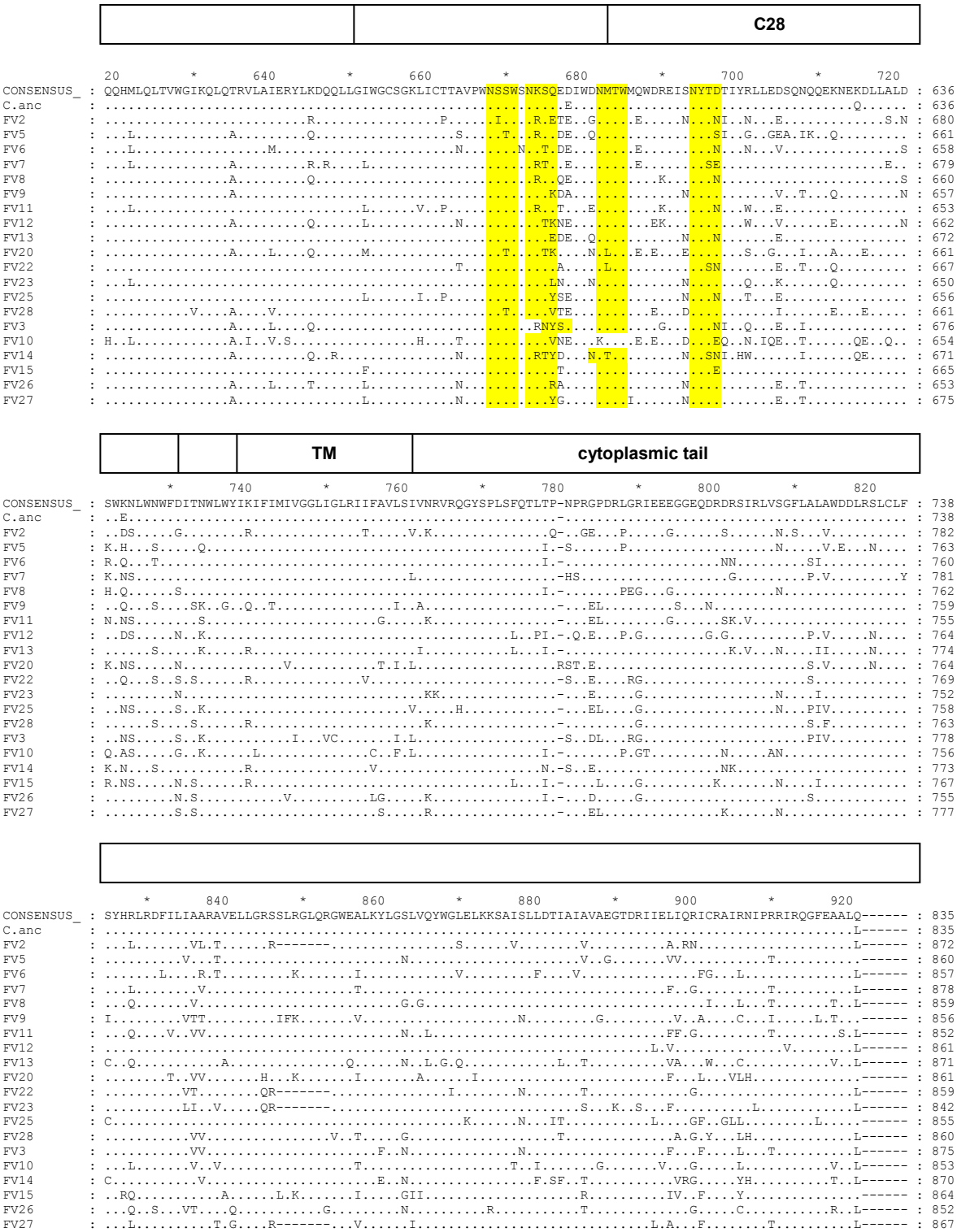


Figure 3.3: Legend overleaf.

Figure 3.3: Alignment of the predicted envelope glycoprotein amino acid sequences of 20 South African HIV-1 subtype C isolates. Dots indicate identity to the consensus subtype C sequence and dashes indicate insertions or deletions. Potential N-linked glycosylation sites are shaded in yellow. Residues that have direct contact with CD4 are shaded grey, and red indicates an overlap between a potential N-linked glycosylation site and a residue in direct contact with CD4. The gp120/gp41 cleavage site is shaded blue. Approximate locations of the various envelope regions are indicated in the bars above the sequence alignment. C1-C5: gp120 constant regions, V1-V5: gp120 variable regions, N34 and C28 are peptides derived from N51 and C43 (see section 1.3.3.1) that form the six-helix bundle of gp41 and which were used to resolve it's crystal structure. TM: Transmembrane domain.

Potential N-linked glycosylation sites in the Env sequences of the 20 newly characterized isolates, as well as the consensus C and ancestral C sequences were predicted using Prosite[208].These are highlighted in yellow in figure 3.3. Overall positive charge in the V3 loop of each isolate was also calculated (table 3.1), by starting at zero and adding 1 for the positively charged amino acids lysine and arginine, adding 0.5 for histidine, and subtracting 1 for the negatively charged amino acids aspartic acid and glutamic acid. The V3 loops of the isolates able to use CXCR4 had a consistently higher charge than those exclusively utilizing CCR5, and the V3 loops of the CXCR4-using isolates tended to be longer than those of the CCR5-using isolates.

Table 3.1: Table of charges and lengths of the V3 regions of the 20 isolates

Isolate	Charge of V3 loop	Length of V3 loop
Isolates exclusively using CCR5		
FV2	+1	35
FV5	+4.5	35
FV6	+4.5	35
FV7	+3.5	35
FV8	+4	35
FV9	+3.5	35
FV11	+3.5	35
FV12	+3.5	35
FV13	+4.5	35
FV20	+3.5	35
FV22	+3	34
FV23	+3.5	35
FV25	+3.5	35
FV28	+3.5	35
Isolates able to use CXCR4		
FV3	+6.5	37
FV10	+6	34
FV14	+6.5	35
FV15	+6	37
FV26	+8.5	35
FV27	+4.5	35

C: Charge, L: Length

3.3. Generation of Recombinant gp120-Expression Vectors

Gp120-encoding DNA amplified from 19 of the 20 full-length gp160 products was inserted into the multiple cloning site of pTriEx. 17 of the 20 isolates that were sequenced were successfully cloned. We were unable to generate clones for FV11, FV27 and FV28. In order to confirm the presence of the insert within the vector, PCR was performed on colonies from the ligation plates that were cultured overnight (figure 3.4a). Where PCR confirmed the presence of the gp120 insert within the vector, plasmids were isolated from the overnight culture and their integrity confirmed by restriction digestion analysis (figure 3.4b). The restriction endonuclease recognition site for *Xba* I is upstream of the point of insertion of the 5' end of the gp120-encoding DNA and the *Xho* I site is reconstituted at the 3' end of the insert upon ligation. Thus a single, intact and correctly positioned insert in pTriEx would yield two DNA products of approximately 4.8kb (plasmid DNA) and 1.7kb (insert) (For map of pTriEx-gp120, see appendix 2, figure S1).

Plasmids containing the correct size inserts of FV3 and FV5 gp120 coding regions were sequenced, which confirmed the presence of start and stop codons in the correct reading frame of the plasmid. Selection of clones for further propagation was based on phenotypic determination of coreceptor use (i.e. one CXCR4-using and one CCR5-using clone were required). As FV3 (CXCR4-using) and FV5 (CCR5-using) were performing well in culture, these two clones were selected for expression of gp120 as this presented an opportunity for interfacing biochemical data obtained from recombinant proteins with data obtained from cell culture studies.

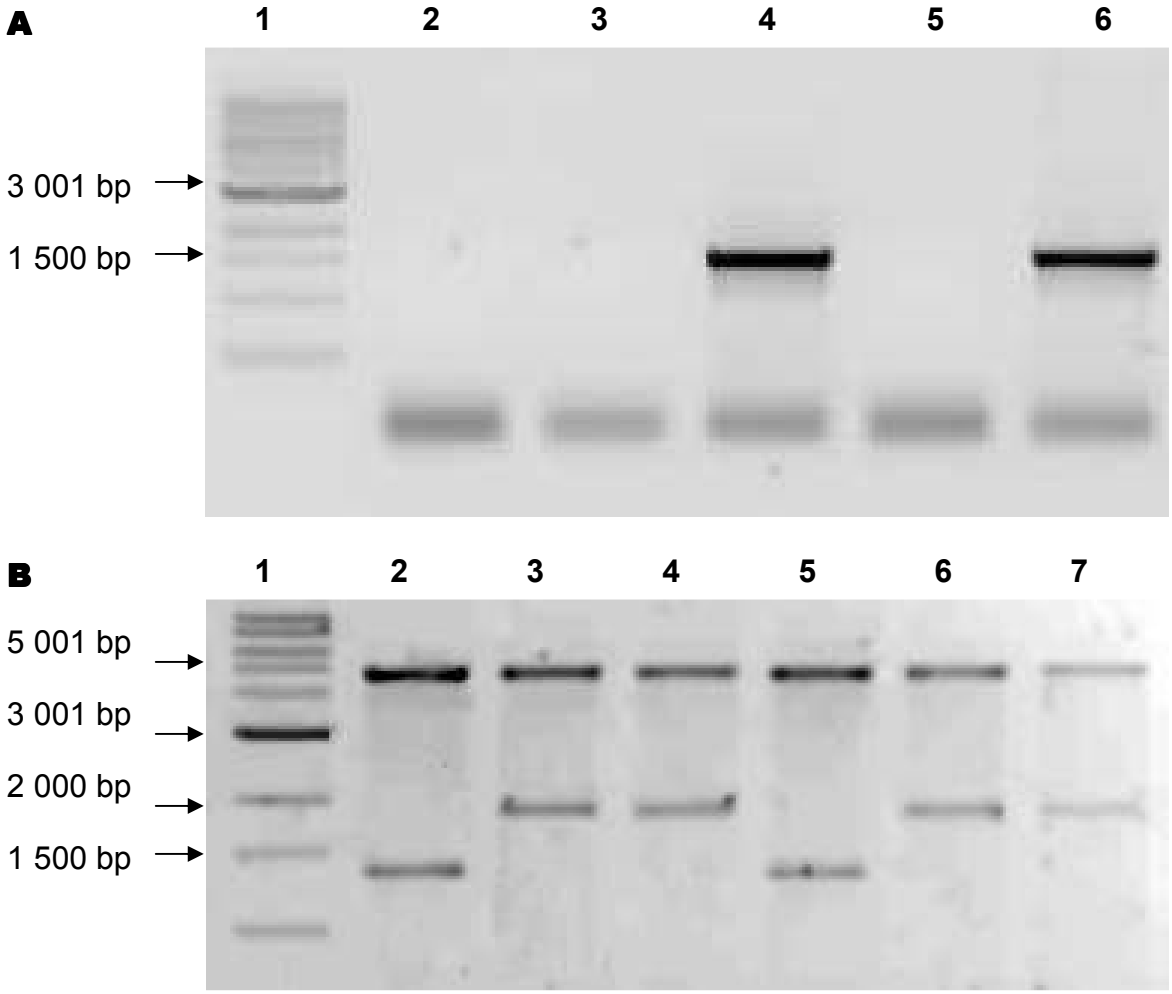


Figure 3.4: Screening for gp120 clones. Representative agarose gels showing **(A) PCR screening.** Colonies selected from positive ligation plates were cultured overnight and subjected to PCR with gp120-specific primers. Generation of a 1.5 kb fragment implied positive ligation between pTriEx and gp120-encoding inserts. Lane 1: MW marker (Quick-Load 1 kb DNA Ladder). Lane 2: Negative control. Lanes 3,5: gp120-negative colonies. Lanes 4,6: gp120-positive colonies (FV5). **(B) Restriction digestion analysis of FV3:** After PCR, clones with a gp120 insert were digested with *Xho*I and *Xba*I. 1:MW Marker (Quick-Load 1 kb DNA Ladder). Lanes 3,4,6,7: Clones containing the correct size insert. Lanes 2,5: Clones containing an aberrant insert. Similar results were obtained for all clones.

At the time of clone generation, the population based sequences of the gp160 amplicons were not available for comparison with the cloned sequences. Later, once population-based sequencing data became available, sequences of the cloned gp120's were compared to the gp120 portion of the population-based sequences of their gp160 counterparts (figure 3.5). FV3 (523 amino acids in length) was cloned with only one amino acid difference between the population-based gp160 and the gp120 insert (99.8% sequence similarity). This difference occurred at amino acid 6 (I6T), in the signal peptide region of gp120, which is cleaved before the protein reaches maturity, and thus the actual protein sequence is exactly the same as that of the predominant viral population in the patient. FV5 (508 amino acids in length) was cloned with 36 differences between the population-based amino acid sequence and the clone (93% sequence similarity). These differences are spread throughout the entire gp120 sequence, with the FV5 clone having two more potential glycosylation sites than its gp160 counterpart in the V4 region, one at position 383 and another at position 395 (highlighted yellow in figure 3.5).

The sequences of FV3 and FV5 were also aligned and compared (figure 3.6). FV3 was 523 amino acids long and contained 29 potential N-linked glycosylation sites, while FV5 was 508 amino acids long and contained 23 potential glycosylation sites (highlighted in yellow in figure 3.6). Comparison of the variable loops of each sequence reveals much shorter V1 and V5 loops in FV5, while FV3 has a shorter V4 region (figure 3.6).

A

```
                *        20         *        40         *        60         *        80         *       1
FV3gp120 : MRVMGIQRNCQQWWIWGILGFWMLMICNGGNLWVTVYYGVPVWKEAKTTLFCASDAKAYEKEVHNVWATHACVPTDPNPQEMKLRNVTENFNMWKNDM :  98
FV3gp160 : .....T.............................................................................................. :  98

          00         *       120         *       140         *       160         *       180         *
FV3gp120 : VDQMNEDIISLWDESLKPCVKLTPLCVTLNCSDVTYNATNATNNTTTTHNTTETTPYAKISNITDDMKNCSFNVTTGLRDKRKQESALFYRLDIIPL : 196
FV3gp160 : ..................................................................................................... : 196

           200         *       220         *       240         *       260         *       280         *
FV3gp120 : NGNKENSSEYRLINCNTSTIRQACPKVSFDPIPIHYCAPAGFAILKCNDKTFNGTGPCHDVSTVQCTHGIKPVVSTQLLLNGSLAEEEIVIRSENLTN : 294
FV3gp160 : ..................................................................................................... : 294

              300         *       320         *       340         *       360         *       380         *
FV3gp120 : NAKIIIVHLNESVEIKCSRPGNNTRKSVRIGIGRGQTFYATGKVIGDIRQAHCNVSREAWNKTLEKVKRKLGEHFPNSTITFNHSSGGDLEITTHSFN : 392
FV3gp160 : ..................................................................................................... : 392

                 400         *       420         *       440         *       460         *       480         *
FV3gp120 : CRGEFFYCNTSDLFKDNITITNSTNNTVITLQCRIKQIINMWQRAGQAIYAPPIRGNITCNSNITGLLLTRDGGKDNKTNNENKTEIFRPGGGDMRDN : 490
FV3gp160 : ..................................................................................................... : 490

                    500         *       520
FV3gp120 : WRSELYKYKVVEIKPLGIAPTTAKRRVVEREKR : 523
FV3gp160 : ................................... : 523
```

B

```
                *        20         *        40         *        60         *        80         *       1
FV5gp120 : MRVMGIQRSCQRWWIWGLLGFWMLLICNGTENLWVTVYYGVPVWREAKTTLFCASDAKSYEREVHNVWATHACVPTDPNPQEMVLENVTENFNMWKNDMV : 100
FV5gp160 : ...R..L....H............M.......................................................................... : 100

                *       120         *       140         *       160         *       180         *       200
FV5gp120 : DQMHEDIISLWDQSLKPCVKLTPLCVTLNCTDFDPLDNGTDAMGEIKNCSFKATTELKDKKRKEQALFYRLDIVSIDKVSNSSNRNFSMYRLINCNTSAI : 200
FV5gp160 : ..............................................................T.........YD.CY.H.......D...... : 200

                *       220         *       240         *       260         *       280         *       300
FV5gp120 : TQACPKVSFDPIPIHYCAPAGYAILKCNDKTFNGTGPCNNVSTVQCTHGIKPVVSAQLLLNGSLAEGEIIIRPENLANNVKTIIVQFNESVEIVCTRPNN : 300
FV5gp160 : ..........V.....................................T.............S.......................... : 300

                *       320         *       340         *       360         *       380         *       400
FV5gp120 : NTRKSVRIGPGQTFYATGEIIGNIRQAHCNISESKWNKTLQNVAKRLKEHFPNKNITFAPASGGGLEITTHSFNCRGEFFYCNTSKLFNSTYNSNDTERD : 400
FV5gp160 : ..........................................................R.D.Q...P..Y........Y..S.....S..H.A... : 400

                *       420         *       440         *       460         *       480         *       500
FV5gp120 : YNGTGNNSTDSIITIQCRIKQFINMWQRVGQAMYAPPIEGNVTCKSIITGLLLTRDGGSNTTETFRPVGGDMRDNWRSELYKYKVVEVRPLGIAPTRAQR : 500
FV5gp160 : ...A..S.S..F...L.......I......R.......V.S...C.F.................................................. : 500


FV5gp120 : RVVEREKR* : 508
FV5gp160 : .....V..  : 508
```

Figure 3.5: Alignments of cloned gp120 sequences with their population-based gp160 counterparts. Dots indicate identity to the top sequence and dashes indicate insertion/deletions. Where potential N-linked glycosylation differs between the gp120 clone sequence and the gp160 sequence, sites are highlighted in yellow. **(A) FV3 (B) FV5**.

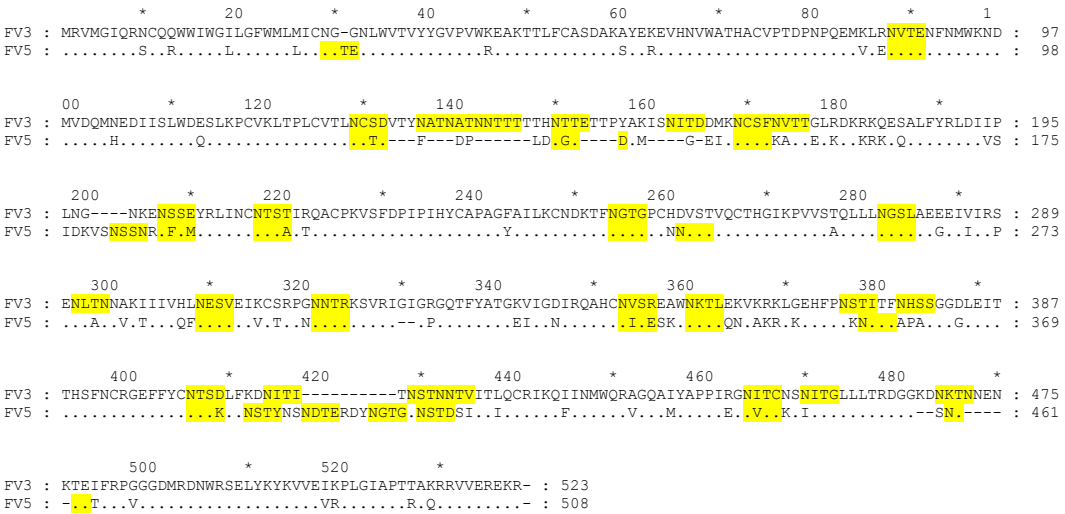


Figure 3.6: Alignment of cloned FV3 and FV5 sequences. Dots indicate identity to the top sequence and dashes indicate insertion/deletions. Potential N-linked glycosylation sites are shaded in yellow.

3.4. Bacterial expression of gp120

In order to evaluate whether gp120 fragments had ligated with pTriEx to produce functional open reading frames, *E. coli BL21* were transformed with the pTriEx-FV3 clones in a pilot expression induction experiment. Since *E. coli* BL21 do not glycosylate post-translationally, an activity essential for the generation of fully functional gp120, the objective of these experiments was simply to determine whether or not clones were capable of producing protein. Western Blots using gp120-specific antibodies confirmed the presence of two different sized gp120 bands (approximately 60 kDa and 50 kDa) (figure 3.7b), possibly formed by differential processing of gp120 in *E. coli*. These bands were absent in samples from untransformed cultures.

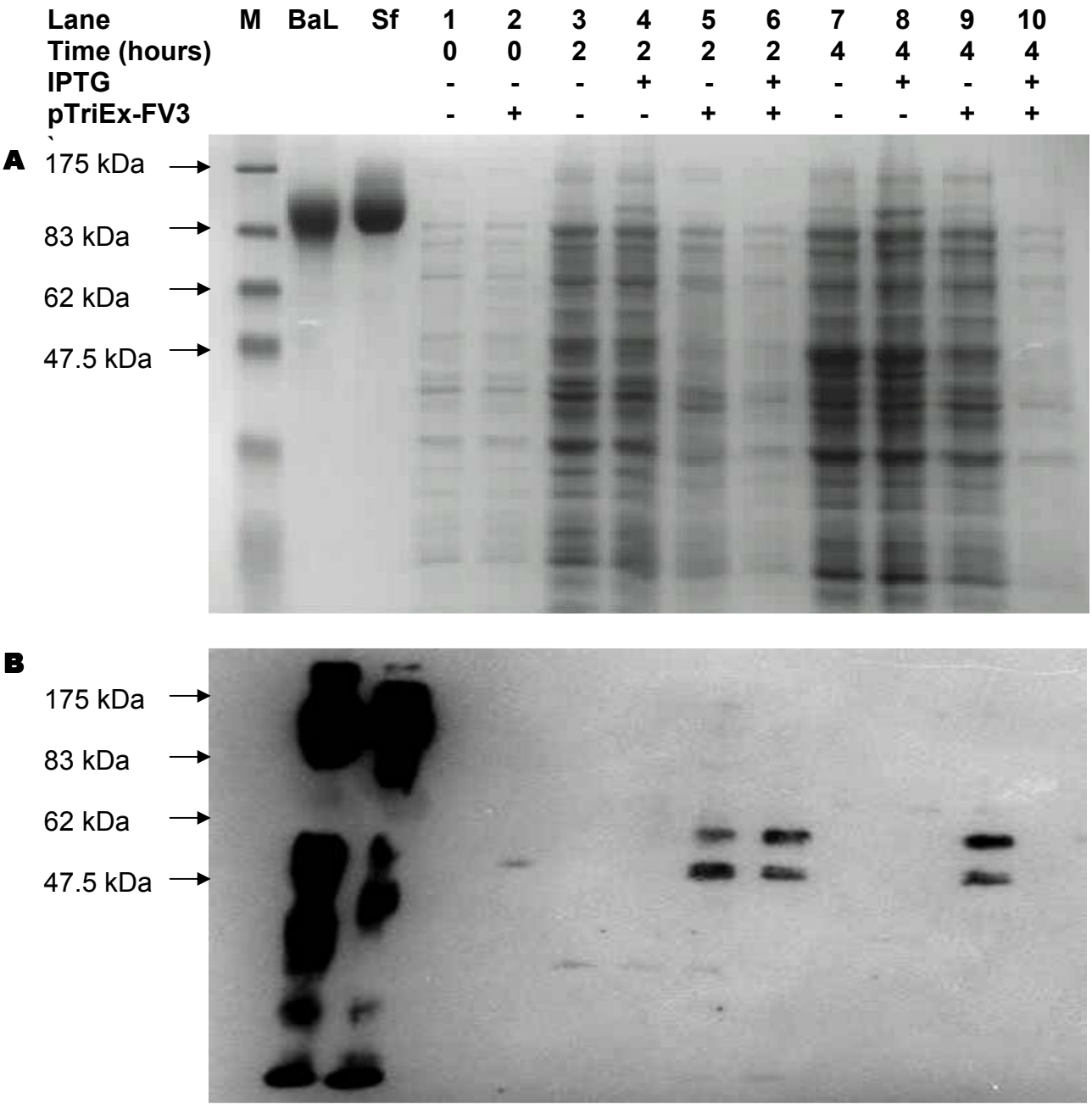


Figure 3.7: Expression of gp120 in *E. coli* BL21. BL21-pTriEx-FV3 cells, together with untransformed controls, were induced with 1 mM IPTG. Samples taken hourly were resolved by SDS-PAGE. Lane labeling is shared for both A and B: M: Prestained Protein Marker, Broad Range (New England Biolabs), Sf: Sf162. **A: Coomassie®-stained gel. (B) Western blot (approximate positions of molecular weight markers are shown).**

It was also noted that pTriEx-FV3-transformed *E. coli* BL21 showed continuous low-level protein expression, regardless of whether treated with IPTG (lane 2, figure 3.7). No gp120 was seen in the induced BL21-pTriEx-FV3 culture at four hours (lane 10, figure 3.7b), and it thus appears that high levels of expression of gp120 in *E. coli* BL21 may be toxic to the cells, resulting in cell death and degradation of protein.

3.5. Expression of gp120 in Insect cells

3.5.1. Transfections and Plaque Purification

3.5.1.1. Generation of recombinant baculoviruses

Transfection was achieved using a liquid overlay method, which involves adding a transfection mixture directly to the culture medium on a cell monolayer. The transfection mixture consisted of triple-cut baculoviral DNA and pTriEx-FV3/FV5, together with the transfection reagent. The promoter and cloning regions of pTriEx are flanked on either side by regions of baculovirus genomic DNA. This facilitates the generation of baculovirus recombinants by homologous recombination at the viral *polh* locus.

Supernatants obtained from transfection experiments were negative for gp120 when tested by Western Blotting analysis (figure 3.8a), but were nonetheless used to reinfect a cell monolayer in a five day infection. The resulting supernatant was used in a further

round of infection, after which supernatants were testing positive for the presence of gp120 by Western Blot (figure 3.8b).

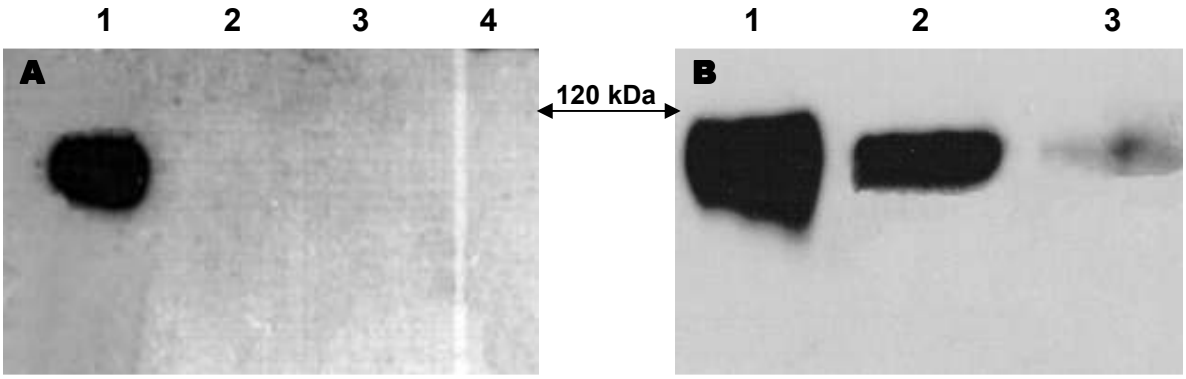


Figure 3.8: Western Blots representative of baculoviral transfection and amplification results. Transfection and amplification supernatants were resolved on 8% SDS-PAGE gels and gp120 was detected by Western Blotting analysis. **(A) Transfection supernatants:** Lane 1: CM gp120 (positive control), Lane 2: negative transfection. Lane 3: FV3 Undiluted transfection, Lane 4: FV3 transfection (1/5 dilution of transfection mixture). **(B) Amplified transfection supernatants** (obtained by amplification of transfection supernatants, after 2 rounds of amplification): Lane 1: CM gp120 (positive control), Lane 2: FV3 from undiluted transfection, Lane 3: FV3 from1/5 dilution of transfection mixture.

3.5.1.2. Plaque selection and generation of high-titer baculoviral stocks

Baculovirus-containing supernatants required purification in order to select single baculoviral clones to be used for protein expression. This was accomplished by two rounds of plaque assays, each time picking a single plaque for further use. Plaque-derived

viruses were amplified by two successive rounds of reinfection between plaque assays. After two rounds of plaque purification, selected clones were used to generate high-titer stocks of recombinant baculoviruses (Bac-FV3 and Bac-FV5).

3.5.1.3. Viral Titering

Once high-titer baculoviral stocks of Bac-FV3 and Bac-FV5 had been obtained, it was necessary to calculate corresponding viral titers in order to determine the optimal multiplicity of infection (MOI) required for consistent and efficient protein expression. Plaque assays are the recommended method for calculating viral titers, however we found these to be challenging and subjective and the failure of a number of such assays due to difficulties in visualizing and counting plaques necessitated the development of a different assay for the titering of baculoviral DNA. To this end, we designed a quantitative real-time PCR assay for quantifying the number of baculoviral DNA copies in Bac-FV3 and Bac-FV5 stock solutions. An approximately 180 bp target sequence within both FV3 and FV5 gp120-encoding genes was amplified using pTriEx-FV3 and pTriEx-FV5 as templates in order to generate external standard curves describing DNA concentration as a function of the amplification curve crossing point. DNA was extracted from baculoviral supernatants using automated extraction techniques and, using the generated standard curves externally, the number of copies of gp120-encoding DNA (and therefore the number of baculoviruses containing a gp120 insert) within each sample was quantified. Viral nucleic acid titers obtained ranged from 0.8×10^9 to 1.0×10^9 gp120

copies/ml of gp120. See appendix 2, figures S2 and S3 for standard curves and amplification curves.

3.5.2. Protein expression studies

Expression of protein in insect cells presented a number of challenges. Consistently low protein expression was obtained using monolayer cell cultures, and before methods for viral titering by real time PCR had been established. Using the baculoviral titers calculated by real time PCR, conditions for protein expression were optimized. Proteins were expressed in both suspension and monolayer cultures and monitored each day for three to five days for cell viability in the culture and protein expression (figure 3.9). In suspension cultures, while protein expression increased over time, lower molecular weight products, probably a combination of degraded and incompletely glycosylated proteins, also began to be produced at around day 2 or 3 (figure 3.9b, days 2 and 3).

After several rounds of optimization, it was determined that maximal expression of intact recombinant FV3 and FV5 gp120 was obtained in suspension cultures and required cells that had undergone fewer than 20 passages be infected at a density of 2.0×10^6 cells/ml. For both FV3 and FV5, the optimal MOI (as determined by real-time PCR) was 3 gp120-encoding copies/cell. Suspension cultures were grown for three days until cell viability (as determined by trypan blue staining) was less than 50%. These conditions resulted in a significant amount of protein with minimal degradation due to proteases and other enzymes released from dying cells.

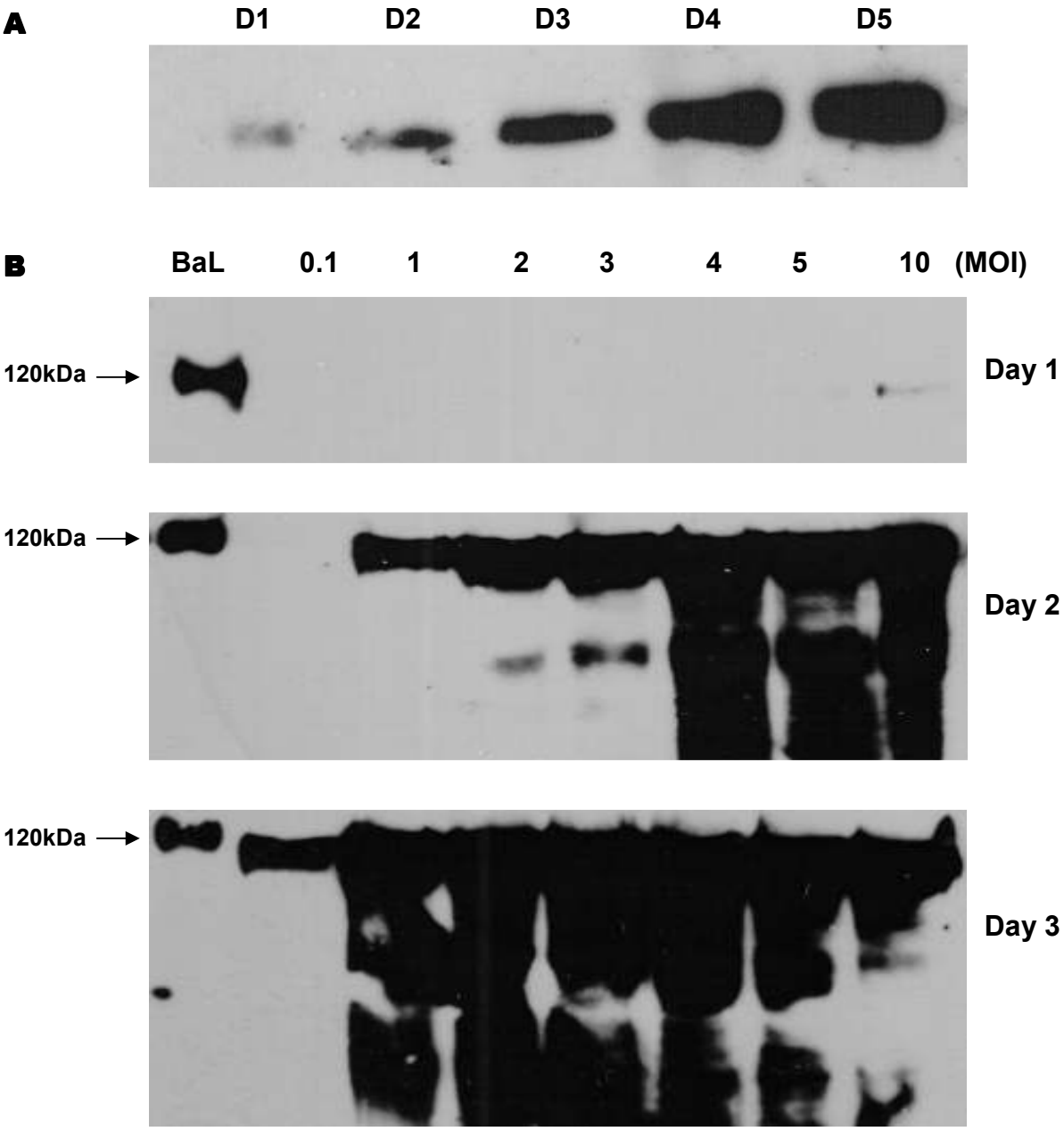


Figure 3.9: Gp120 expression in monolayers and suspension culture. Monolayers as well as suspension cultures were infected with supernatants containing Bac-FV3 and Bac FV5 and monitored for 3-5 days. (**A) Representative monolayer culture (FV3) at unknown MOI, days 1-5** (D: Day) **(B) Representative suspension culture infection, days 1-3** The MOI is written above each lane in the top image and corresponds to the same lane in all three days.

3.6. Protein Purification

A number of strategies for purification of gp120 were attempted, including lectin-affinity chromatography, size-exclusion chromatography and ion-exchange chromatography. While attempts at size-exclusion chromatography were never met with any appreciable amount of success, the final protocol for protein purification involved both lectin purification and ion-exchange chromatography. Purification of protein-containing supernatants using lectin from *Galanthus Nivalis* involved binding of protein to lectin columns, followed by extensive washing with NaCl and elution with MMP. This resulted in purification to a level of approximately 60-65% (by visual analysis of silver stains). The purity of these eluates was increased by subjecting them to ion-exchange chromatography using a pH and NaCl gradient. After dialysis of the ion-exchange eluate against PBS and concentration of protein samples, the yield of FV3 and FV5 gp120 obtained per 200 ml starting volume of culture supernatant was approximately 50 µg.

Analysis of protein purity by silver staining showed FV3 and FV5 purified to near homogeneity (figure 3.10). Silver staining of FV5 revealed a number of lower molecular weight bands, suggesting incomplete glycosylation or degradation of gp120. It was also noted that FV3 is smaller than FV5. This is clearly seen in Western Blot analysis (figure 3.10). The sizes of each protein were estimated using the molecular weight marker and Bio-Rad Quantity One 1-D Analysis Software. According to this analysis, FV3 was estimated to be 112 kDa and FV5 83 kDa (figure 3.10).

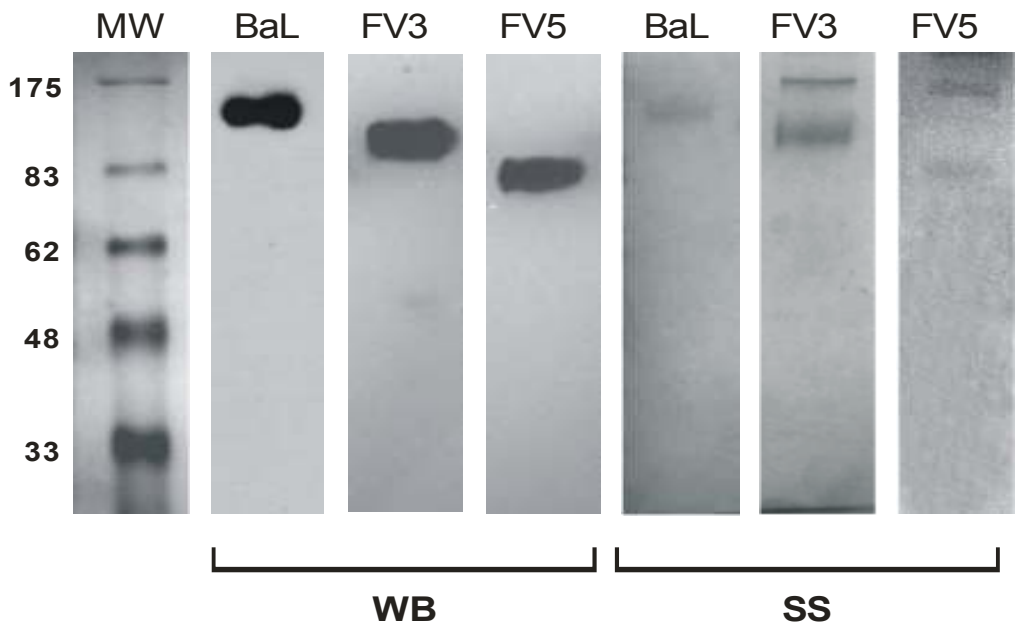


Figure 3.10: Analysis of purified gp120. Purified recombinant gp120 was analysed by Western Blot and Silver Stain (with BaL gp120 as a positive control). WB: Western Blot, SS: Silver Stain, MW: Molecular Weight Marker (Prestained Protein Marker, Broad Range (New England Biolabs)).

3.7 Protein Characterisation

3.7.1. Characterisation of binding of recombinant gp120 to a panel of monoclonal antibodies.

In order to determine epitope exposure and conformational integrity of the purified gp120s, proteins were concentrated, quantified and their antibody-binding profiles compared to each other, and to those of three reference gp120 proteins obtained from the NIH, by direct ELISA. The control proteins were the Subtype B $gp120_{BaL}$ (BaL), Subtype C $gp120_{96ZM651}$(96ZM651) (both produced in mammalian cells), and the insect-cell produced Subtype E (CRF01_AE) $gp120_{CM}$ (CM). The reactivity of these five proteins against a panel of antibodies was determined by ELISA in two experiments, performed each time in triplicate, with the same results each time. Due to the variability of the reactivity of the gp120s with the various antibodies, the A_{450} value was set to 3.5 when it was above the readable amount. The results of the antibody binding studies can be seen in figure 3.11. Graphs including experimental standard deviations can be seen in supplementary figure S4 in appendix 2.

In these experiments, and in comparison to the control panel of proteins, as well as a BSA blank, FV3 had low binding to the monoclonal antibody 654-30d, moderate binding to 2G12, Chessie 13-39.1 and 9301, and high binding to ID6. FV5 had low binding with the monoclonal antibodies F425 B4a1 and Chessie 13-39.1, moderate binding with 2G12, and good binding to 9301 and ID6 (figure 3.11).

	BaL	CM	96ZM651	FV3	FV5	Blank
F105	3.04	0.04	0.08	0.05	0.04	0.04
F425 A1g8	1.92	0.04	0.08	0.04	0.04	0.04
F425 B4a1	3.19	0.24	0.04	0.04	0.18	0.04
F425 B4e8	3.50	0.05	1.37	0.04	0.06	0.04
670-30D	2.31	0.18	0.07	0.05	0.03	0.03
654-30D	1.72	0.04	1.01	0.36	0.04	0.04
17b	0.04	0.04	0.03	0.04	0.04	0.04
2G12	3.50	0.76	0.06	1.08	0.71	0.04
ID6	2.19	0.19	2.57	2.74	2.00	0.04
Chessie 13-39.1	0.94	1.87	0.26	0.58	0.27	0.04
5F7	0.04	0.04	0.04	0.04	0.04	0.04
4G10	0.04	0.04	0.04	0.04	0.04	0.04
9301	2.73	2.89	3.49	1.70	1.64	0.07
IgG1 b12	1.41	0.07	0.05	0.04	0.04	0.04
48d	0.04	0.04	0.04	0.04	0.04	0.04

Figure 3.11: Comparison of antibody recognition characteristics of the panel of five gp120's studied. A_{450} values equal or lower to the highest blank value (0.07) were classified as having no binding. A_{450} values between 0.07 and 0.5 were classified as having low binding. A_{450} values between 0.5 and 1.5 were classified as showing moderate binding. A_{450} values between 1.5 and 2.5 were classified as showing good binding. A_{450} values between 2.5 and 3.5 were classified as showing high binding. Data was obtained from a triplicate experiment. Blank values were obtained by testing antibody binding to wells coated with BSA.

KEY	
Binding	Colour
No	
Low	
Moderate	
Good	
High	

3.7.2. CD4-binding capability of recombinant gp120

The presence of intact and functional CD4 binding sites in the expressed recombinant gp120 was also assessed by ELISA. ELISA plates were coated with CD4 and after a blocking step, gp120 was added to the coated plates and allowed to bind. CD4-bound gp120 was then detected using a selection of anti-gp120 monoclonal antibodies (figure 3.12). These analyses showed that FV3 reacted with CD4, as measured by its binding to 2G12. FV5 however, did not show any evidence of being able to react with CD4.

In order to quantify how well FV3 binds to CD4 relative to BaL, the value obtained for BaL binding to 2G12 in the antibody recognition experiments was normalized to 100%, and the percentage of FV3 binding to 2G12 relative to this value was calculated. These calculations indicated that FV3 binds to 2G12 with an efficiency that is 31% of the BaL binding efficiency to 2G12. This value was then used to normalize the values obtained for BaL and FV3 binding to 2G12 from the CD4-binding data. By these calculations, it is reasonable to infer that FV3 binds to CD4 with approximately 22.5% efficiency relative to BaL.

The CD4-binding data were also aligned next to the CD4-unbound data (figure 3.13). In this figure, a remarkable increase in binding to some gp120s by the antibodies targeted to the CD4-induced epitope (17b, 48d, and F425 A1g8) is seen in the presence of CD4.

	BaL	CM	ZM	FV3	FV5	Blank
2G12	3.50	0.21	0.05	0.26	0.04	0.04
17b	3.50	1.30	3.19	0.07	0.04	0.04
48d	1.26	0.09	0.04	0.04	0.04	0.04
F425 A1g8	3.50	0.77	1.14	0.04	0.05	0.07

KEY	
Binding	Colour
No	
Low	
Moderate	
Good	
High	

Figure 3.12: CD4 binding activity of a panel of gp120s. ELISA plates were coated with sCD4, after which gp120 was bound and detected by a number of monoclonal antibodies. A_{450} values were classified as follows: those equal to or lower than the highest blank value (0.07) showed no binding, low binding values were between 0.07 and 0.5, moderate binding values were between 0.5 and 1.5, good binding values were between 1.5 and 2.5 and high binding values were between 2.5 and 3.5. Data obtained from a triplicate experiment. ZM: 96ZM651. Blank values were obtained by testing antibody binding to wells coated with BSA.

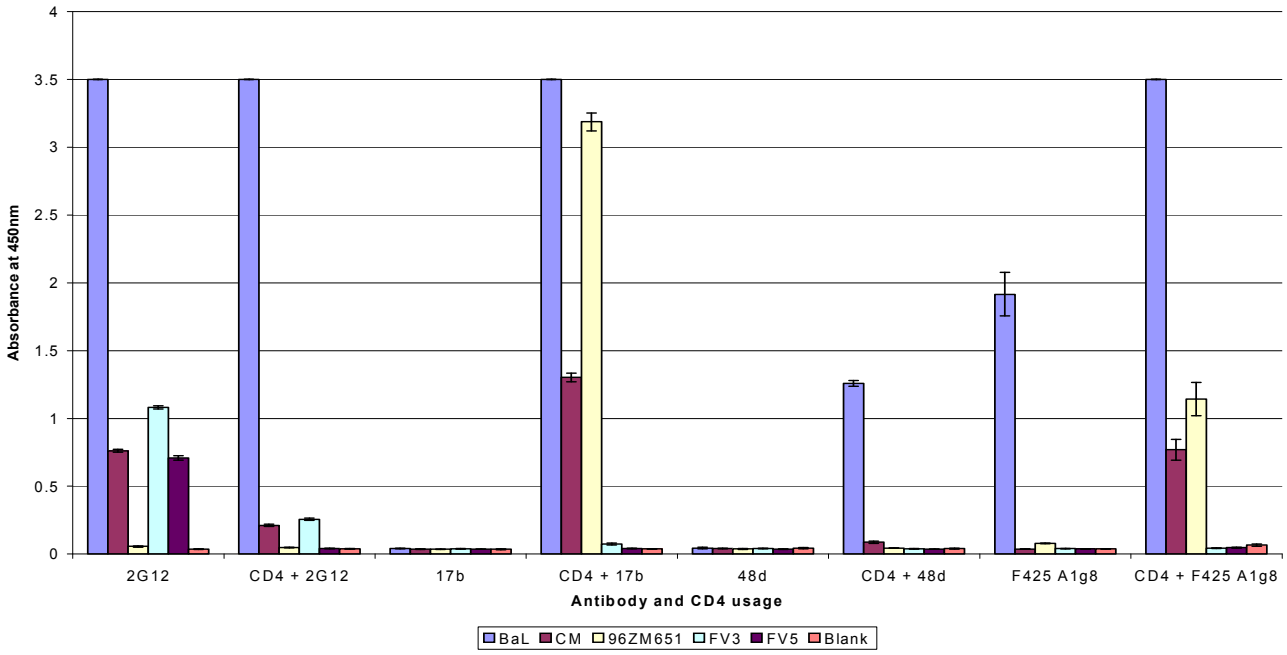


Figure 3.13: Comparison of monoclonal antibody binding to gp120 in the presence and absence of CD4. Error bars represent standard deviations.

4. Discussion

The aim of this project was the expression, purification and characterization of recombinant gp120 from recently obtained South African HIV-1 subtype C isolates. To this end, gp120-coding sequences were cloned from locally relevant isolates, expressed in insect cells and purified using a combination of lectin-affinity and ion-exchange chromatography. Purified recombinant FV3 and FV5 gp120s (CXCR4-utilising and CCR5-utilising respectively) were characterised according to their ability to bind CD4, as well as their ability to bind to a number of monoclonal antibodies. Following is a discussion of the issues surrounding this work.

4.1. PCR amplification of Env-encoding regions

PCR amplification of gp160 coding regions was successful when using proviral DNA as a template. Amplification of gp120 coding regions from gp160 products was also successful, except in the case of FV28. Inspection of the nucleotide sequence of this isolate revealed an 18 nucleotide insertion in this sequence where the binding site for the reverse primer for gp120 amplification is located. This insertion results in a complete mismatch between primer and proviral DNA at the 3' end, thereby preventing primer binding and successful PCR. Should it become of specific interest to clone this gp120, the problem can be easily rectified by the design of a reverse primer specific to this isolate.

4.2. Population-based sequencing of the gp160 coding regions

4.2.1. Amino acid sequences

Population based sequencing provided interesting insights into this cohort of AIDS patients. Due to the high mutation rate of HIV-1, patients infected with the virus usually contain a number of related viral quasispecies in the circulating viral population. Population based sequencing is based on PCR of all proviral DNA present in the patient's blood sample, and thus does not represent a single virus in the patient, but rather a sequence that is reflective of the entire viral population, usually the most predominant sequence.

All of the isolates sequenced in this study were shown to be subtype C, except for one isolate, FV10. This virus is a CD recombinant. Subtype D is typically found in West/West Central Africa, while subtype C is more common in Southern Africa. Thus, this strain could have originated in West Central Africa, or it could represent a recombination event between a West Central African and South African strain.

The amino acid sequences of the *env* regions of the 20 isolates were extensively analysed and compared. The variable loops of gp120, particularly the V1/V2 and V3 loops have been directly implicated in coreceptor usage[81, 99-104]. Coreceptor usage amongst these isolates was determined phenotypically in our laboratory, using a MT-2 fusion assay. The sequencing results confirmed the phenotypic findings as the typical features of the V3

loop that differentiate X4 from R5 isolates were present. The V3 loop of the R5 isolates was more conserved compared to that of the X4 isolates. The tetrapeptide crown sequence in the V3 loop of the R5 subtype C isolates was GPGQ (with the exception of FV6 and FV25, which was GPGG) and these V3 loops carried a net charge of +1 to +4.5 (table 3.1). A neutral serine residue was present at position 11 (corresponding to position 339), and amino acids with a negatively charged side chain (either D or E) were present at position 25 (corresponding to position 355) of the V3 loop, except for FV28. This pattern of amino acids at these positions has been shown to correspond to R5 usage among HIV-1 isolates[209]. In contrast, 4 of the 6 CXCR4-using isolates exhibited changes in the V3 tip tetramer motif and two of these isolates each had two amino acid insertions within this motif. The overall positive charges of the V3 loops for the X4 isolates range from +4.5 to +8.5 (table 3.1). The X4 phenotype has been associated with the presence of at least one basic or positive amino acid substitution at position 11 or 25[209]. Interestingly, all of the CXCR4-using isolates, except for FV3 had a positively charged residue at position 11 of the V3 loop and all, except FV10, had either a neutral or positively charged residue at position 25. Until recently, X4 usage amongst subtype C viruses was reported as extremely rare[115-122]. Recently, however, reports of higher percentages of CXCR4-using isolates in subtype C populations have emerged[123, 124], and the determinants of CXCR4-usage within subtype C populations have been extensively studied[210, 211]. The predominance of subtype C infection makes it of paramount importance as a focus of research efforts. The rarity of subtype C CXCR4-using isolates makes this cohort a valuable tool to study subtype C pathogenesis and CXCR4-usage within this clade, as well for the design of drugs specific to this clade.

Analysis of the residues involved in formation of the CD4 binding site[73] (shaded in grey/red in figure 3.3) revealed that most were conserved, although some differences were noted and these sequences were thus compared with those of other isolates found in the Los Alamos database[45]. Comparison of sequences revealed that amino acids can occur in various combinations around the CD4 binding site. It is interesting to note, however, that residues of gp120 crucial for CD4 binding, D368, E370 and W427 of the HIV-1 reference sequence HXBc2 (residues 403, 405 and 467 in figure 3.3 respectively)[73], are highly conserved among the 20 isolates. Only FV5 and FV23 have a different amino acid at position E370, a residue interacting with Phe43 of CD4[73]. Both of these isolates have been grown in culture, however, with FV5 replicating efficiently in both PBMC cultures and in lymphocytic cell lines following adaptation, suggesting that substitutions at this position do not altogether abrogate CD4 binding. It has indeed been found that substitutions at this site compromise, but do not altogether abrogate binding of CD4[212]. It should also be noted that the gp120 clone of FV5 did not contain this mutation (Figure 3.5b).

All 20 isolates, except FV28, had the consensus motif -R/K-X-R/K-R at the gp120/gp41 cleavage site (shaded blue in figure 3.3). Proteolytic activation of the HIV-1 gp160 precursors occurs at the carboxyl side of the consensus motif in this region[213]. This region of the precursor is the main determinant for cleavage, which is catalyzed by a cellular subtilisin-like protein convertase located in the Golgi complex[58]. It is thus expected that the gp160 precursor of these isolates would be efficiently cleaved into the gp120 and gp41 subunits with no differences between the R5 and X4 variants. In the case of FV28,

cleavage may be affected due to the two amino acid insertion at position 565, which abolishes the consensus cleavage site. HIV does, however, contain a secondary cleavage site, corresponding to position 555, with the sequence KAKRR[214]. This sequence is conserved in FV28, and thus cleavage of the envelope may occur here, although it is likely that the corresponding gp41 will be a monomer and therefore not biologically active[214].

The gp41 sequences of the 20 isolates are highly conserved relative to those of gp120, with fewer potential glycosylation sites, and no notable differences between the SI and NSI isolates. It was also noted that four of the 20 new isolates (3 CCR5-using and 1 CXCR4-using) contained 5-7 amino acid deletions in the gp41 cytoplasmic tail (figure 3.3).

4.2.2. Glycosylation

Potential glycosylation of the 20 amino acid sequences was also analysed. Overall, there was no drastic alteration in the number of potential glycosylation sites, and the number and positioning of sites appeared to be mostly conserved between the R5 and X4 isolates (figure 3.3). Interestingly, all the CCR5-utilizing sequences contained an N-linked glycosylation motif within the V3 loop (positions 334 to 337), consistent with recent findings highlighting the close association of this site with CCR5 usage[215]. Furthermore, the X4 isolates were likely to have an extra glycosylation site within the V4 loop and

further investigations could reveal whether glycosylation at this site could mask any immunologically relevant epitopes.

Glycosylation also plays a key role in the efficacy of antibody neutralisation[70]. 2G12 is a neutralizing antibody which relies on the presence of certain glycans for its activity. One postulation of the epitope for 2G12 suggests that two N-linked glycans (positions 295 and 332 of $gp120_{HxBc2}$, corresponding to positions 328 and 366 in figure 3.3) form the core epitope for 2G12, which is maintained conformationally by other N-linked glycans (at positions 339, 386 and 392 of $gp120_{HxBc2}$, corresponding to positions 373, 421 and 427 in figure 3.3)[216]. The high mannose glycan of residue 295 was absent in 14 of the 20 isolates, suggesting that they may be resistant to neutralization by the monoclonal antibody 2G12. It is interesting to note, however, that N295 was present in 3 of the 5 subtype C CXCR4-utilizing envelope glycoproteins, perhaps suggesting different evolution of glycosylation patterns in these isolates.

4.3. Generation of gp120-encoding vectors

Integrity of the gp120-encoding vectors was confirmed by sequencing. FV3 had 99.8% similarity when compared to the corresponding population-based gp160 sequence, while FV5 only had 93% sequence similarity. The reasons for these differences are unclear. One possibility is that the errors were obtained during PCR. This seems unlikely however, as a high fidelity polymerase was used to perform the PCR reactions, and the percentage of differences from the population based sequence are very different for each

clone. A more probable reason stems from the presence of a number of viral quasispecies in each patient. All of the viral quasispecies present in the patient would have been amplified by PCR and although the population-based sequence would represent the predominant species, it is possible that a less predominant quasispecies of the virus could have been the one incorporated into the vector during cloning.

4.4. Protein Expression

4.4.1. Bacterial Protein Expression

Expression of gp120 from pTriEx-FV3 was induced in transformed *E. coli* (BL21) in order to confirm the competence of the cloned open reading frames for protein expression. Culture lysates from transformed *E. coli* BL21 tested positive for the presence of gp120 in time course induction experiments by Western Blot analysis (figure 3.7). It is interesting to note that two different sized bands were detected by probing with the anti-gp120 antibody 9201. Gp120 has previously been expressed in *E.coli*, also producing two bands upon Western blot analysis[217]. These bands were approximately 45 kDa and 55 kDa, with the smaller band corresponding to mature, secreted gp120 without the signal peptide sequence, and the bigger band corresponding to gp120 including the signal peptide sequence[217]. In our gp120 clones, the HIV-1 signal peptide sequence was left intact, while Morikawa *et al.*[217] replaced the HIV-1 signal sequence with a bacterial signal sequence. While there is no evidence of the HIV-1 envelope leader sequence being processed in *E. coli*, it may be possible that the bigger and smaller bands seen in our cell extracts correspond to immature unprocessed gp120 with the signal sequence and mature,

processed gp120 without the signal sequence respectively. It is estimated that the bands are a similar size to those obtained by Morikawa *et al*. The estimated sizes of the bands (approximately 50 kDa and 60 kDa) is consistent with unglycosylated gp120, as would be produced in *E.coli*, which do not have glycosylation machinery[64, 69].

4.4.2. Insect Cell Protein Expresssion

While production of recombinant baculoviruses was relatively simple, protein expression and viral titering presented a number of challenges. Before viral titering procedures were in place, protein expression in insect cells was consistently low, regardless of whether studies were done using monolayer or suspension cell cultures. Viral titering itself was also problematic. In our hands, plaque assays, the recommended method for calculating viral titers, were poorly reproducible and defining titers in this manner was problematic. Plaque assays involve infecting a cell monolayer with varying dilutions of baculoviruses and counting the number of plaques (or defined areas of dead cells) that form as the baculovirus multiplies and begins to destroy the cells surrounding the originator infected cell. It was often difficult, however, to distinguish between plaques, cells destroyed during the process of setting up the assay, and gaps in the monolayer. Cells also tended to detach or overgrow very quickly, making plaques even harder to discern. Thus, while it was relatively straightforward to isolate a single plaque for plaque purification purposes, we experienced difficulties in obtaining accurate plaque counts for viral titering purposes.

Because of these problems we developed a real-time PCR assay in order to quantitate baculoviral DNA in the supernatants of infected cultures. This seemed a promising approach as, during infection, recombinant baculoviruses are released into the tissue culture supernatant. Once the supernatant has been clarified, any gp120 DNA present should originate from recombinant baculoviruses in the supernatant. Thus, if this DNA could be quantitated, it would give an idea of how many recombinant baculoviruses are present in the supernatant.

DNA was isolated from insect-cell culture supernatants using automated extraction procedures in order to obtain a uniform yield. The real-time assay proved to be extremely robust and allowed for quantification of the number of copies of the gp120 insert (and therefore the number of gp120-containing baculoviruses) per ml of supernatant. While this number does not distinguish between infectious and non-infectious baculoviruses, it provides a platform for addition of uniform amounts of virus to insect cell cultures. Real-time PCR has been used previously in order to quantitate baculoviral DNA[218]. In that study, which used a similar protocol, the baculoviral titers obtained from real-time PCR correlated closely with titers detected from 50% tissue culture infectious doses ($TCID_{50}$s) of baculovirus. It is therefore likely that the number of gp120 copies per ml calculated by real-time PCR approximates the number of infectious baculoviruses present. This supposition is supported by the fact the baculoviral titers calculated by real-time PCR were within the range that should be present as specified in the manufacturer's suggested protocols. In any event, the assay provided a basis for infecting cultures with uniform amounts of baculoviruses, and enabled optimization of required innocula.

Once a platform for viral titering was in place, it was possible to optimize the conditions for protein expression in insect cultures. During this process, a number of factors were found to influence the expression of protein in insect cells. Cells generally supported superior recombinant protein expression the fewer passages they had undergone. Suspension cultures exceeded monolayers in amount of protein produced (figure 3.9). The use of suspension cultures was also important in order to obtain high volumes of gp120-containing supernatants. Most gp120 expression systems employed involve secretion of recombinant protein into the supernatant media of mammalian or insect cell cultures, enabling purification of secreted protein by simple chromatographic methods[219-221]. In this study, experiments using Bac-FV3 and Bac-FV5 showed comparable levels of cell-associated and secreted gp120s. In order to simplify the purification protocol, and consistent with other successfully employed purification methods, we decided to focus exclusively on purifying soluble, secreted gp120.

Incubation times after infection also proved important for maximal production of intact protein. It has been reported that the best expression of completely glycosylated gp120 in insect cells occurs approximately 24 hours post-infection[170]. In that study, after approximately 30 hours, while gp120 production continued, it was produced in what is proposed to be a non-glycosylated form, as it is postulated that Sf9 cells are unable to process large amounts of glycosylated proteins.[170]. These experiments were performed on cell extracts, however, rather than on the cell-culture supernatant. In another study, while peak expression of gp120 occurred at approximately 48 hours in the cells themselves, peak levels of gp120 were only present in the supernatant 72 hours post-infection[169]. In

our study, peak expression of gp120 in the cell-culture supernatant occurred at approximately 72 hours post-infection (figure 3.9). Between 72 and 96 hours post-infection, additional low molecular weight bands were detected by Western Blot analysis. These bands were not distinct, however and possibly correspond to incompletely-glycosylated proteins, rather than completely non-glycosylated ones.

Possibly the most important factor for optimal protein production, however, was the amount of baculovirus used to infect insect cells. Before a platform for baculoviral quantification was in place, the amount of baculovirus being added to cultures (an estimated MOI of 5 gp120-encoding copies/cell) was estimated based on the amount of baculovirus the manufacturer suggested should be present after plaque purification and amplification of recombinant baculoviruses. Under these circumstances, protein expression in both monolayer culture and suspension was extremely poor. Once the supernatants being used were analysed by the real-time assay, however, it became clear that the amount of recombinant baculovirus being used to infect the infect cells was far below the required amount. Once the viral titer was accounted for, and the MOI increased to 3 gp120-encoding copies/cell, protein production dramatically improved (figure 3.9).

Thus, in our hands, consistently efficient protein expression could be obtained by infecting fresh suspension cultures at a cell density of 2 x 10^6 cells/ml and an MOI of approximately 3 gp120-encoding copies/cell. The formation of degraded and incompletely glycosylated proteins was minimized by restricting the time of incubation of infected insect cell cultures to a maximum of three days.

4.5 Protein Purification

Lectin-affinity chromatography is a common method used for the purification of HIV-1 envelope glycoproteins[172, 222-225] because of the affinity of lectins for carbohydrate residues. We attempted to use both *Galanthus Nivalis* lectin, which is specific for α-D-glucose and α-D-mannose and *Lens Culinaris* lectin, which is specific for branched mannoses with fucose-linked α(1,6) to N-acetyl glucosamine. Recombinant gp120 bound to both lectins with similar affinity and produced an end product of similar purity. Going forward, however, *Galanthus Nivalis* lectin was used to purify gp120 due to availability. Lectin-affinity chromatography is generally used in conjunction with other purification procedures, however, in order to obtain protein of higher purity[172, 225]. When oligomeric gp140 was produced by Jeffs *et al*.[224] in Chinese Hamster Ovary cells and purified by *Galanthus Nivalis* lectin alone, the purity of the protein obtained was only between 50 and 70%. This is consistent with our experience during the purification of FV3 and FV5 gp120.

Interestingly, in other experiments being performed in our laboratory, a single purification step with *Galanthus Nivalis* lectin was required in order to obtain recombinant gp120 (96ZM651) of greater than 90% purity. This protein, however, was expressed in mammalian cells (293T). Supernatants obtained from these transfected mammalian cells appeared to require far less intensive purification steps than supernatants obtained from infected insect cell cultures. This may be due to insect cell proteins forming complexes with the recombinant gp120, which co-purify on the lectin column. Insect cells may also produce and secrete a number of glycoproteins absent from

the mammalian cells used, or else it is possible that many more cells die in insect cell culture than in mammalian cell culture due to the baculoviral infection, and therefore release a large amount of extra proteins into the supernatant. Nevertheless, with the addition of an ion-exchange chromatography step to the purification protocol, the purity of the protein was dramatically increased.

4.6 Protein Characterisation

4.6.1. Protein Size and Purity

FV3 was estimated to be 112 kDa and FV5 83 kDa (figure 3.10). Examination of the clones' sequences (figure 3.6) provides the reason for the smaller size of FV5 when compared to FV3, as FV5 is not only 15 amino acids smaller than FV3, but it also contains six less potential glycosylation sites than FV3.

It was also interesting to note that FV3 expression appeared to be associated with the presence of fewer lower molecular weight products than FV5 (figure 3.10). This is suggestive of higher levels of protein degradation, as well as incomplete glycosylation of FV5. The presence of proteases released from dying cells in the cell culture supernatant may be the reason for the degradation observed. Decreasing the incubation time of infected insect cell cultures, however, did not solve the problem, which is also not seen with FV3. It seems possible, therefore, that there may be something intrinsically toxic about FV5, which destroyed the cells almost as soon as the protein was produced. Analysis of cell viability during insect cell infections, however, does not support this

theory (see supplementary figure S5 in appendix 2) as, in general, cells infected with FV5 appeared to remain viable for longer than those infected with FV3. It may be possible that FV5, having fewer glycosylation sites, is simply more accessible to cellular proteases than FV3, which may explain why it was more prone to proteolytic activity.

Another interesting observation regarding gp120 size is the difference between insect-cell and mammalian-cell produced proteins. While a number of attempts were made to produce gp120 in mammalian cells, these did not meet with a large amount of success. A small amount of His-tagged FV3 was produced in 293T cells in other experiments in our laboratory however, and when this gp120 was compared to its insect-cell produced counterpart, a notable difference in size was seen between the two (supplementary figure S6 in appendix 2). This size difference is common when comparing proteins produced in mammalian cells to those produced in insect cells. It has been widely reported that while insect cells have the ability to add N-glycans to proteins, the glycosylation pathways are less complex than those in mammalian cells[226-228]. There was no evidence of addition of complex carbohydrates to proteins or sialylation in insect cells in these studies. Rather, it is likely that the carbohydrates added were of the high-mannose variety, which are smaller and less complex than the sugars added to proteins in mammalian cells. In studies of gp120 produced in insect cells, these findings were confirmed, as the carbohydrates present were all of the high-mannose variety[229]. These differences account for the slightly lower molecular weight of insect-cell produced gp120 (approximately 110 kDa), when compared to mammalian-cell produced gp120[169].

Despite the differences in carbohydrate structure, gp120 has been widely produced in baculoviral systems, reportedly with full biological activity[169, 170]. Recombinant baculoviral gp160 has been used as an immunogen in order to produce monoclonal antibodies against gp120[176, 230] . Monoclonal antibodies, such as ID6, produced in such a manner are able to bind to both mammalian and insect cell gp120[176]. Baculoviral gp120 has also been used to study the mechanisms of the pathogenesis and neutralization of HIV-1[231].

4.6.2. Recognition of monoclonal antibodies by recombinant gp120

4.6.2.1. Antibodies to the CD4-binding site of gp120

4.6.2.1.1. F105

F105 is a human monoclonal antibody, which binds to a conformational epitope on gp120 overlapping with component residues of the CD4 binding site[198, 232] (shaded grey in figure 4.1). In early studies, F105 inhibited the binding of HIV-1 virions to CD4 positive cells, and its binding to gp120 was competitively inhibited by soluble CD4[198]. Resolution of the crystal structure of F105 revealed the presence of an extended loop with a Phe residue at its tip[232]. Computer modeling supports the hypothesis that this residue binds to gp120 in the same cavity as Phe43 of CD4[232]. F105 did not bind well to the subtype C proteins, only showing low binding to 96ZM651 and none to FV3 or FV5. The reasons for this are unclear. It may be possible however, as this antibody was isolated from an

HIV patient in the USA[198] where subtype B HIV predominates, that some structural difference between the gp120s from different clades is responsible for the lack of binding of F105 to clade C isolates.

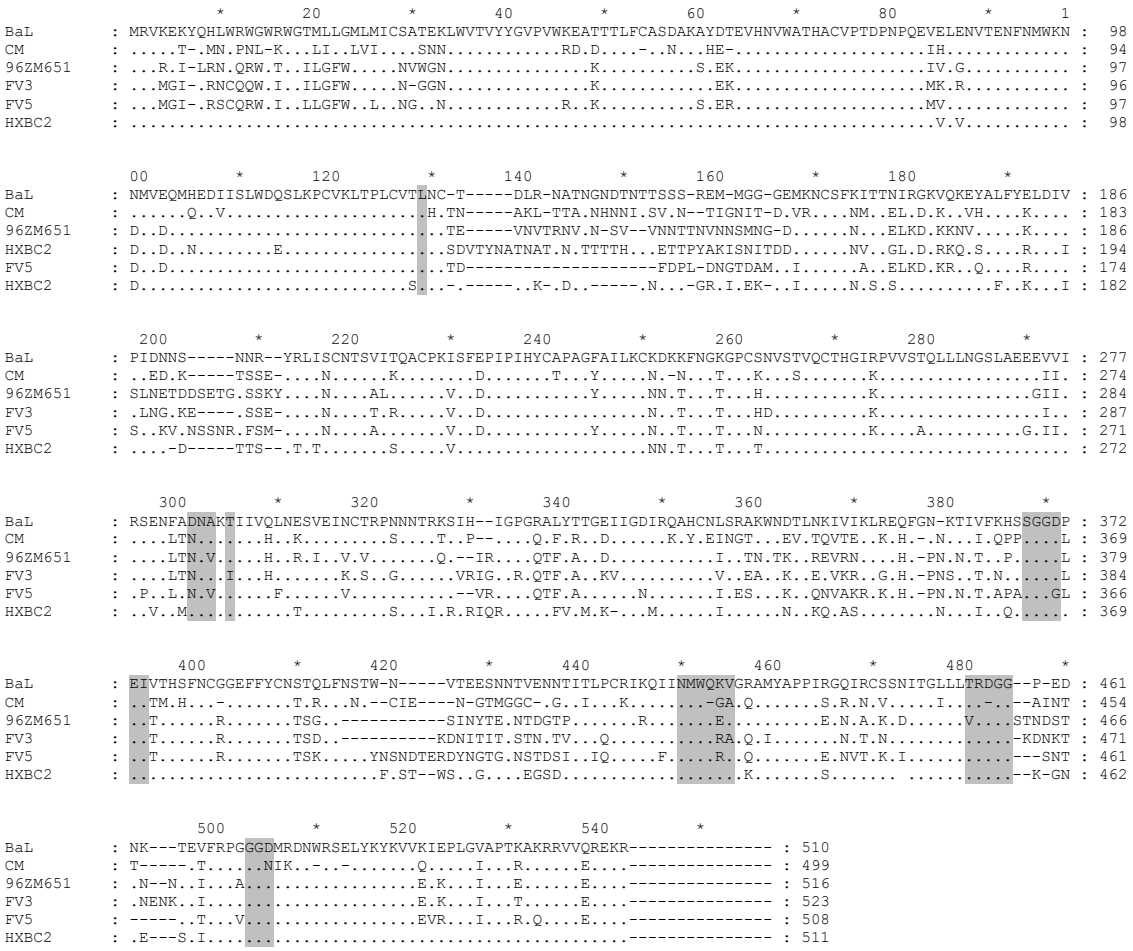


Figure 4.1: Amino Acid Alignment of the gp120s characterised by ELISA, including HXBC2 as a reference sequence, with the CD4 binding site highlighted. Amino acids in direct contact with CD4 are shaded grey.

4.6.2.1.2. 654-30D

654-30D is a monoclonal antibody that recognizes a discontinuous epitope in the CD4 binding domain of gp120[183] (shaded grey in figure 4.1). Its reactivity with gp120 is strictly conformation dependent and thus reactivity with gp120 is a good indicator of a conformationally intact protein. BaL showed good binding with this antibody, while 96ZM651 showed moderate binding and FV3 showed low binding. This suggests that FV3 has a conformationally intact CD4-binding site, and this was subsequently further confirmed in the CD4 binding studies. CM and FV5 showed no binding, however. It is known that CM has an intact CD4 binding site, as is it distributed as an intact, functional gp120 from the NIH, and therefore the lack of reactivity of this protein with 654-30D is probably due to protein-specific conformational or sequence differences. In the case of FV5, the reason for the lack of activity of this antibody is less clear. These analyses were performed to assess the conformation of the recombinant gp120, and thus the lack of binding of FV5 to 654-30D cannot simply be attributed to conformational differences as for the control gp120, CM. Although it is possible that the lack of binding reactivity may be due to protein-specific differences, it may also be possible that the CD4-binding site of FV5 is not conformationally intact. This, in fact, appears to be case, as a mutation was found in a critical CD4-binding residue of FV5 (discussed in section 4.6.3).

4.6.2.1.3. IgG1 b12

The monoclonal antibody IgG1 b12 was developed using combinatorial phage display technology, and binds to gp120 in or near the CD4 binding site[202, 203]. Resolution of the

crystal structure of IgG1 b12 resulted in the suggestion that IgG1 b12 and gp120 interact by the insertion of a tryptophan residue (W100) of IgG1 b12 into the same binding pocket on gp120 that is used by Phe43 of CD4[233]. The residues of gp120 important for this interaction are S365, D368, I371, Y384 and V430 of HXBc2 (corresponding to consensus residues 388, 391, 394, 407 and 455 in figure 4.1 respectively)[233]. While these residues are relatively highly conserved amongst the panel, with only a single substitution each in CM, FV3 and FV5, the only gp120 that showed any reactivity with IgG1 b12 was BaL, which had moderate reactivity. The differences may not be attributable to subtype differences as IgG1 b12 has previously been shown to bind well to subtype C envelopes [171] and generally shown good cross-clade reactivity with gp120, as well as cross-clade neutralising activity[187, 204, 234, 235]. Binding of IgG1 b12 to gp120, however, is particularly sensitive to changes within the V1/V2 region[205]. Although this was determined by experiments involving inhibition of IgG1 b12 binding by deletion of the V1/V2 region, it is possible that differences in these regions between the five proteins examined may contribute to the poor recognition of four of them by IgG1 b12.

4.6.2.2. Antibodies to the CD4-induced conformation of gp120

4.6.2.2.1. F425 A1g8

F425 A1g8 is a human MAb that appears to bind preferentially to the CD4-induced conformation of gp120[236], although it does show some binding to gp120 in the absence of CD4[236, 237]. In these experiments, A1g8 showed good binding to BaL and low binding to

96ZM651. It did not show binding activity with any other isolates. In the presence of sCD4, F425 A1g8 showed high binding to BaL and moderate binding to 96ZM651 and CM protein. No binding was seen for FV3 and FV5 in the presence of sCD4, however. Inspection of the sequences does not reveal an obvious reason for these differences as the amino acids implicated in the CD4i epitope (shaded green in figure 4.2) are well conserved between the entire panel, with the only exception being two substitutions in FV5. The difference in binding to A1g8 between the subtype C proteins is therefore possibly due to conformational variation between these isolates.

4.6.2.2.2. 48d

This monoclonal antibody recognises a complex discontinuous epitope in gp120, which includes amino acids in all of the constant regions of gp120, and which shows enhanced binding to gp120 in the presence of CD4[189]. In our experiments, none of the gp120 panel analysed showed binding to 48d in the absence of CD4. In the presence of CD4, however, moderate binding was seen in the case of BaL, and low binding was seen for CM. Analysis of the amino acids shown to important for binding of 48d to gp120[189] (shaded green in figure 4.2) revealed that only BaL contained all of these with no substitutions. All of the other isolates had at least one substitution. In a cross-clade study, 48d did not show good cross-clade reactivity, and demonstrated particularly poor reactivity with subtype C isolates[238]. Taken together, these results indicate that the poor performance of 48d in these assays is to be expected.

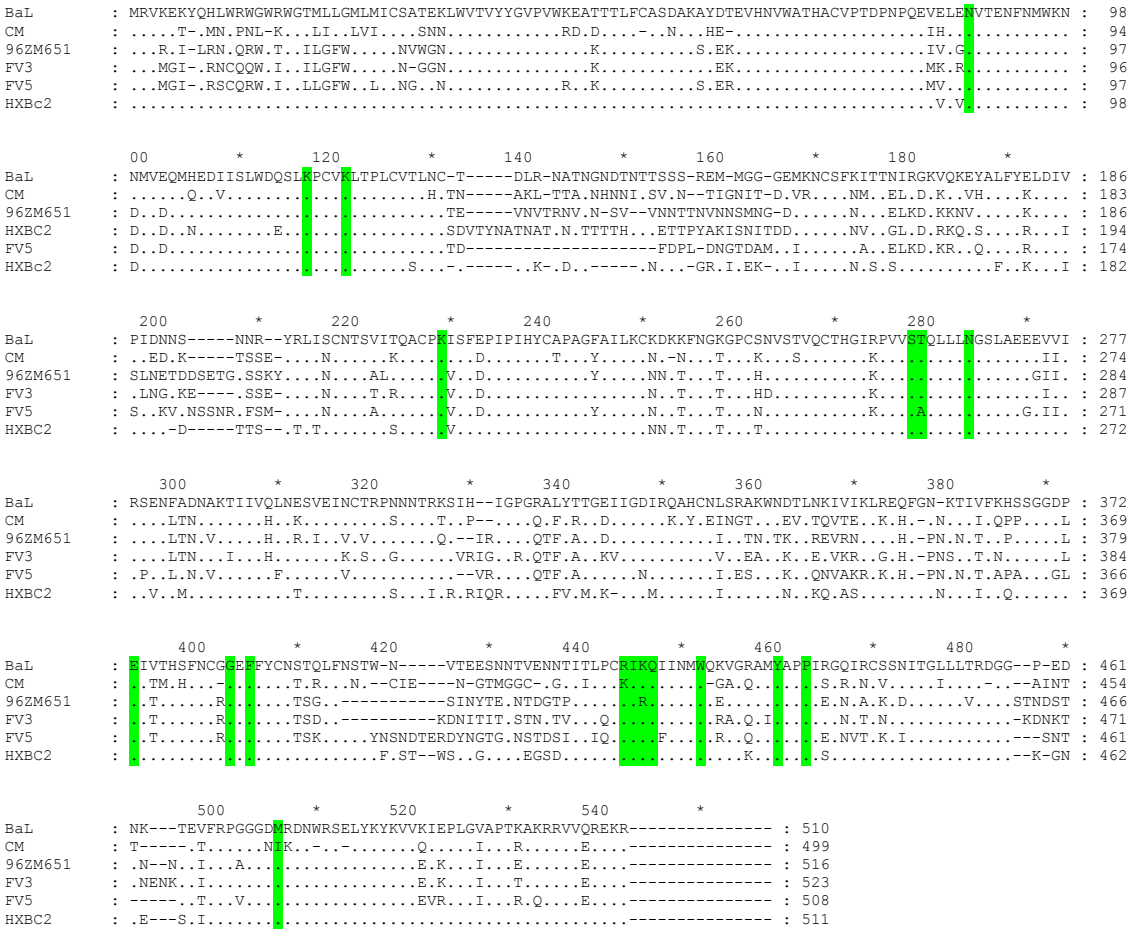


Figure 4.2: Amino Acid Alignment of the gp120s characterised by ELISA, including HXBC2 as a reference sequence, with amino acids implicated in the CD4-induced (CD4i) epitope[189] shaded green.

4.6.2.2.3. 17b

17b binds to an epitope on gp120 that is very similar to that of 48d[189]. The epitope, overlapping the CCR5 binding site of gp120, is well defined as the crystal structure of gp120 was first defined in complex with CD4 and17b[73]. Residues that have been defined as important for binding of 17b to gp120 (HXBc2 numbering, consensus numbering from figure 4.2 in brackets) are K121 (121), R419 (444), I420 (445), K421 (446), Q422 (447),

I423 (448) and Y435 (260)[73, 239, 240]. These residues are well conserved among the panel of gp120s under investigation, except for a single residue each in FV5 and CM. None of the isolates were reactive with 17b in the absence of CD4. In the presence of CD4, however, BaL and 96ZM651 showed high binding to 17b and CM showed moderate binding. Neither FV3, nor FV5 showed binding to 17b in the presence of CD4. In the case of FV5, the presence of Phe instead of Ile at position 445 may help to explain the lack of reactivity with 17b, as the aromatic side chain of phenylalanine may alter or prevent important gp120/17b interactions. FV3 has no mutations in the amino acids essential for binding of 17b, but it may be worth noting that FV3 has at least three extra potential glycosylation sites in the V1 loop in comparison to the other gp120s. It has been suggested that the V1/V2 loop shields the epitope for 17b and that binding of CD4 induces conformational changes that unmask the epitope in order to allow the antibody to bind[201]. The presence of extra glycans in the V1 loop of FV3 may supply extra bulk that cannot be moved completely out of the way by CD4-induced conformational changes, thus hindering the binding of 17b. It is also interesting to note that FV3 is a CXCR4-using isolate and, inaccessibility of the 17b (CCR5) epitope would not affect viral fitness.

4.6.2.3. Anti-V3 loop antibodies

4.6.2.3.1. F425 B4a1

MAb F425 B4a1 (B4a1) reacts with the V3 loop of clade B laboratory isolates, as well as primary isolates with diverse V3 loop sequences, and its binding to gp120 is enhanced after incubation with CD4[236]. B4a1 reacts more efficiently with dual-tropic isolates, than

solely CCR5-using isolates[236]. In our experiments, high binding to B4a1 by BaL was seen, as well as moderate binding by CM and low binding by FV5. Interestingly, FV3 and 96ZM651, which do not bind B4a1, have a mutation in the second residue of their V3 loops (amino acids 318-354, shaded red in figure 4.3), when compared to BaL, while CM and FV5 do not. This is the only residue in this epitope which is substituted in FV3 and 96ZM651. It is therefore tempting to speculate that this residue may be important for the binding of B4a1 to gp120.

4.6.2.3.2. F425 B4e8

Human MAb F425 B4e8 (B4e8) reacts with the crown/tip of the V3 loop of gp120 in X4, R5 and dual-tropic viruses[237]. In early experiments, B4e8 reacted well against a number of clade B HIV-1 isolates, but showed no reactivity with HXB2[237]. More recent studies indicate that B4e8 effectively neutralises clade B viruses as well as some clade C and D primary isolates[241]. These studies also clarified the B4e8 epitope[241], which incorporates, minimally, the amino acids I331, R337, F339 in figure 4.3 (highlighted blue) . Interestingly, BaL, which showed high binding to B4e8, has a Leu residue at position 339, rather than a Phe. This substitution is clearly well tolerated by the antibody. This finding is consistent with recently published structural data showing this antibody in complex with a V3 peptide, which revealed that the interaction of B4e8 with gp120 does not center on this residue, but rather the Ile and Arg residues implicated in the previously mentioned study[242]. In support of these findings, an Arg to Gln substitution at position 337, present in the rest of the panel of gp120s, appears to compromise antibody binding, as 96ZM651 showed only moderate binding and CM, FV3 and FV5 showed none.

Interestingly, the amino acid residues at the three critical positions in FV3 and 96ZM651 are identical, and yet FV3 does not bind to B4e8 with the same avidity as 96ZM651. FV3 does, however, have a 2 amino acid insertion in that region of the V3 loop. Thus, according to our studies, it seems that R337 is the most important of the three mentioned by Pantophlet *et al*[241] for B4e8 binding to gp120 to gp120, and that the epitope may be affected by the proximity of these three residues.

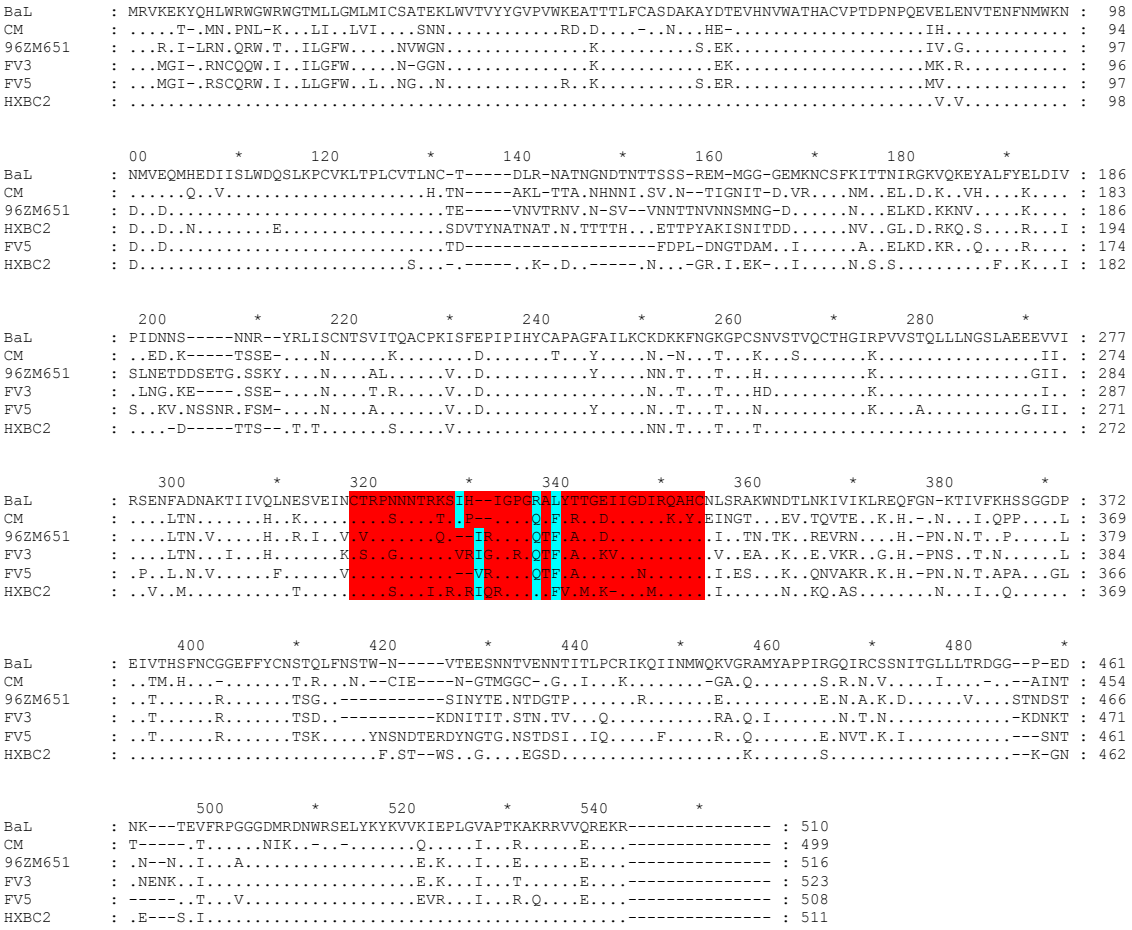


Figure 4.3: Amino Acid Alignment of the gp120s characterised by ELISA, including HXBC2 as a reference sequence, with the V3 loop highlighted. The V3 loop is shaded red, with amino acids implicated in the F425 B4e8 epitope in blue.

4.6.2.3.3. 5F7 and 4G10

The monoclonal antibodies 5F7 and 4G10 were raised against a 25 amino acid sequence derived from the V3 loop of the HIV-1 isolate BH10[185]. This sequence, TRKSIRIQRGPGRAFVTIGKIGNMR, is not highly conserved, however, and comparison with the V3 loops of the proteins being analysed (red in figure 4.3) reveals that the sequences are completely different, with the most conserved region being the GPGR motif found at the tip of V3. It is therefore not surprising that none of the proteins being studied showed any reactivity with either of these antibodies.

4.6.2.4. Antibodies to the constant regions of gp120

4.6.2.4.1. ID6

ID6 was raised against a subtype B HIV-1 isolate but showed binding activity with clade E and clade A/E isolates[176]. This is probably due to the fact that the epitope is found in a highly conserved region of C1[177]. This epitope, amino acids 86 to 100 of the HIV-1_{MN} gp120 sequence (consensus amino acids 87 to 101 in figure 4.4, shaded green) is well conserved between all gp120 variants analysed in these experiments. It is therefore not surprising that each of the gp120s showed binding with this antibody, with FV3 and 96ZM651 showing high binding, FV5 and BaL showing good binding, and CM showing low binding. Interestingly, CM gp120 which, together with BaL, contains the epitope for ID6 with no substitutions showed the poorest reactivity with ID6. This is intriguing and examination of the amino acid sequence around the ID6 epitope provides no clear

explanation for the difference in binding to ID6 between the two antibodies. CM does, however, have a histidine residue at position 85. It is tempting to speculate that the bulky aromatic ring in the histidine side-chain somehow hinders the interaction of CM and ID6 by its proximity to the epitope for ID6.

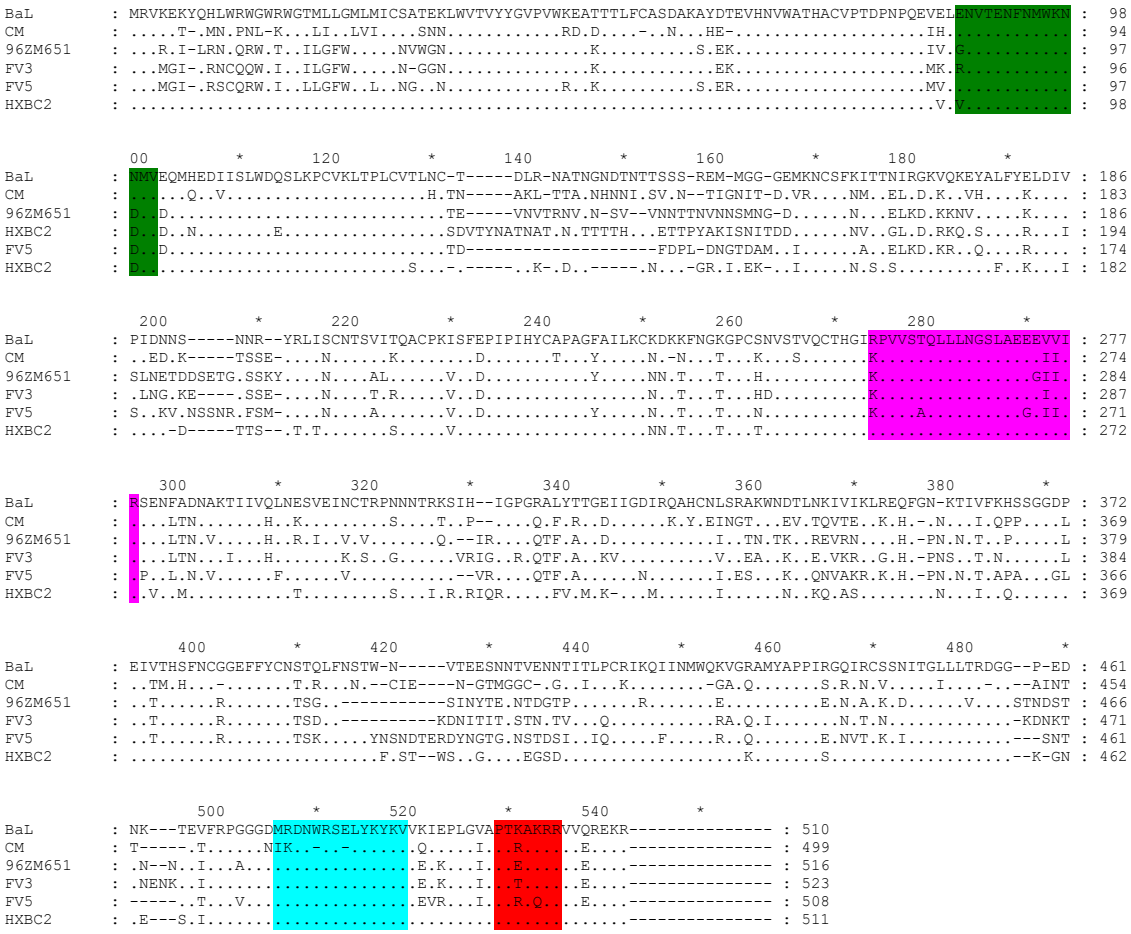


Figure 4.4: Amino Acid Alignment of the gp120s characterised by ELISA, including HXBC2 as a reference sequence, with the epitopes for antibodies against the constant regions of gp120 highlighted. Epitopes are shaded as follows: ID6 in green, Chessie 13-39.1 in pink, 9301 in blue and 670-30D in red.

4.6.2.4.2. Chessie 13-39.1

The monoclonal antibody, Chessie 13-39.1 recognizes a linear epitope in the C2 region of gp120. This epitope, corresponding to the sequence RVVSTQLLLNGSLAEEVVIR[175] (amino acids 274-295, highlighted in pink in figure 4.4) is relatively well conserved between the five proteins analysed, of which BaL and FV3 showed moderate binding, CM showed good binding and FV5 and 96ZM651 showed low binding affinity for this antibody. Interestingly, although the epitope is less well conserved in CM than in BaL, CM shows better reactivity with Chessie 13-39.1. There is no clear reason for this in the amino acid sequence, but taken together with other intriguing results obtained for the CM gp120, such as its lack of reactivity with ID6 and 654-30d, it seems possible that the conformation of the CM protein may be fundamentally different to the conformations of the other proteins studied, resulting in altered antibody epitope accessibility. Our CD4 binding data for CM (figure 4.4) suggests, however, that the fundamental structural nature of this protein is retained.

4.6.2.4.3. 670-30d

670-30d binds to an epitope in the C5 region of gp120, PTKAKRR (amino acids 529-535, shaded red in figure 4.4) and shows good cross-clade reactivity[184, 243]. In our isolates, 670-30d showed good reactivity with BaL, where the epitope is completely conserved. The only other gp120 that bound to this antibody was CM, which showed low reactivity. It is interesting to note that although there is a substitution in the third position of the epitope in CM (lysine to arginine), the charge of the region is retained as both lysine and

arginine are positively charged amino acids. Substitutions are also present in the epitope for the other gp120s, however, these substitutions result in a change in the overall charge of the epitope. Taken together with the lack of reactivity of these proteins with 670-30D, this suggests that the charge of the region may directly affect antibody affinity for gp120 or else play a role in gp120 conformation and therefore the accessibility of the antibody epitope.

4.6.2.4.4. MAb 9301

All gp120 variants tested reacted well with 9301 (epitope is shaded blue in figure 4.4, amino acids 506-519), with BaL, CM and 96ZM651 showing high binding and FV3 and FV5 showing good binding. This is not surprising as the C5 epitope is highly conserved between isolates, although changes to the epitope appear to be well tolerated as shown by the activity of 9301 with CM gp120. Although the binding site for 9301 in CM contains two substitutions and two deletions, the ability of the antibody to bind CM is retained. CM is the only isolate with an amino acid sequence differing from that recognised by 9301, and the changes to the epitope do not appear to significantly alter antibody binding.

4.6.2.5. Antibodies recognising the glycan structure of gp120

4.6.2.5.1. 2G12

The exact location of the epitope of the broadly neutralising antibody 2G12 is controversial. Research has led to the two slightly different models for 2G12 binding. The first suggests that the core epitope for 2G12 is made up entirely of the N-linked glycans present at positions 295 (320), 332 (355), and 392 (415) and supported peripherally by glycans at positions 386 (409) and 448 (473), with 2G12 having no contact with the peptide backbone[244] (numbering in this section refers to the HXBc2 consensus sequence, with numbers in brackets indicating the corresponding amino acid positions in figure 4.5; all significant glycosylation sites are highlighted in yellow in figure 4.5). The second model for binding of 2G12 suggests that the core epitope is formed by N-linked glycans at positions 295 (320) and 332 (355) and supported by glycans at positions 339 (362), 386 (409) and 392 (415)[216]. Resolution of the crystal structure of 2G12, however, led to the suggestion that 2G12 and gp120 make critical contacts at positions 332 (355) and 392 (415), with a potential contact at position 339 (362)[245]. In this model, the glycan at position 295 (320) is postulated to play an indirect role in 2G12 binding by shielding the glycan at 332 (355) from further processing into a complex carbohydrate. This is significant as 2G12 recognises high mannose and/or hybrid carbohydrates, rather than complex ones[216, 244]. 2G12 has been hailed as the most predictable antibody with regard to the presence of the residues important for its binding to gp120 and its ability to neutralize HIV[234].

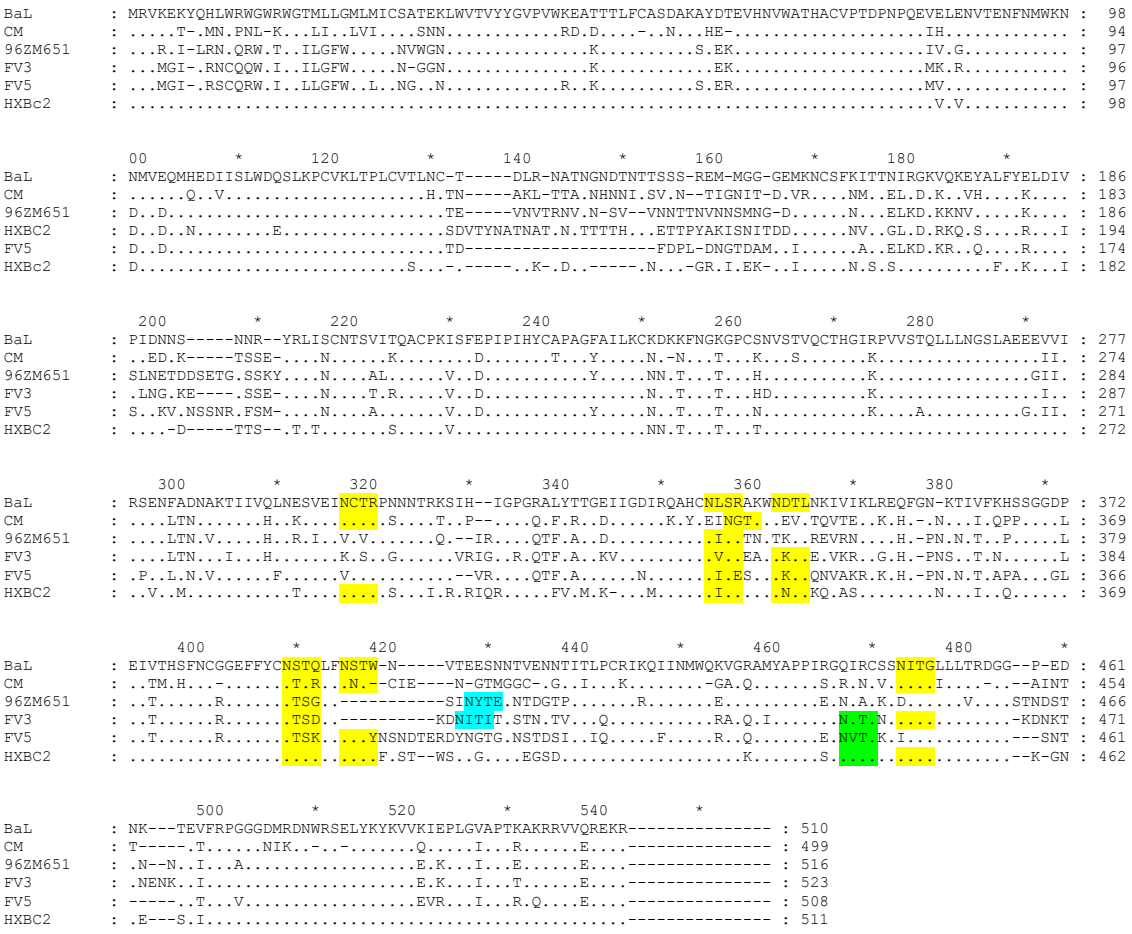


Figure 4.5: Amino Acid Alignment of the gp120s characterised by ELISA, including HXBC2 as a reference sequence, with glycans that have been implicated in 2G12 binding to gp120 highlighted. Glycans implicated in the literature as important are shaded yellow, those in slightly different positions to those cited in the literature are blue. Novel glycan implicated for 2G12 binding in subtype C viruses[246] is shaded green.

In our panel, all the glycosylation sites implicated in having direct or supportive roles in 2G12 binding were exactly conserved only in BaL, the gp120 with the highest reactivity with 2G12 (high). No reactivity was seen with 96ZM651, which is missing glycosylation

sites at positions 295 (320), 339 (362), 392 (415) and 448 (473), although there is a glycosylation sites present two amino acids downstream of position 392 (415) in this isolate (shaded blue in figure 4.5). Interestingly, a moderate amount of reactivity was still noted for FV3, FV5 and CM, even though CM is missing one glycosylation site (position 339 (362)) and both FV3 and FV5 are missing two sites each (position 295 (339) in both FV3 and FV5, position 392 (415) in FV3 and position 448 (473) in FV5).

It is remarkable that FV3 and FV5 showed binding to 2G12 in the absence of a glycan at position 295 (320), as this glycan has been postulated to be crucial for the interaction of gp120 and 2G12[216, 231, 244, 246, 247]. Furthermore, 2G12 has been shown to have consistently low binding to HIV-1 Subtype C recombinant gp120s[171] and neutralizing activity against HIV-1 Subtype C virions[234, 235, 246, 247]. FV3 gp120 binds to sCD4 as well as to the conformation-dependent monoclonal antibody 654-30D suggesting that the recombinant protein is conformationally intact. A possible explanation for the binding of our isolates to 2G12 is derived from examining the crystal structure of this antibody, where it is suggested that the glycan at position 295 (320) plays only an indirect role in formation of the 2G12 epitope by shielding the glycan at position 332 (355) from further processing[245]. As these proteins were produced in insect cells, they will not contain complex carbohydrates in their structure [229]. Thus, it has been suggested, that the glycan at position 295 would not be absolutely necessary for formation of the 2G12 epitope in insect-cell derived proteins[231]. This may explain the ability of FV3 and FV5 to bind 2G12 despite the absence of a glycosylation site at this position. The absence of this glycan, however, has still apparently affected the binding of FV3 and FV5 to 2G12, as the binding seen in

these assays was only moderate. Consistent with this, a study where the glycosylation sites missing at positions 295 and 392 of a subtype C isolate were reconstituted, binding of 2G12 to gp120 was significantly increased, particularly in the case of the glycan at position 295[231]. It is likely that binding of 2G12 to FV3 and FV5 gp120 would be significantly higher were glycans present at position 295 (320) in these recombinant proteins. Another possible explanation for the surprising results is that 2G12 may have bound non-specifically to another component of the gp120 preparation, giving a false positive result. While this cannot be ruled out, the differing levels of 2G12 binding between FV3 and FV5, as well as the complete absence of 2G12 binding to FV5 in the CD4-binding studies, suggest that this is not the case. Preliminary experiments in culture also suggest that FV3 may be susceptible to 2G12 neutralization to some extent. To elaborate on the role played by N295 in 2G12 binding, we aim to evaluate the binding of this antibody to FV3 gp120 expressed in mammalian cells further, as well as its ability to neutralize HIV-1 FV3 in cell culture.

Our results differ from others where binding of gp120 to 2G12 did not occur in the absence of a glycan at position 295[171, 231]. Thus, while the glycan on gp120 at N295 plays a major role in 2G12 binding, our results suggest there are other important determinants for this interaction. To gain further insight into these, we aligned the population-based sequences of the isolates used in these studies (CN54, Du151, Du179, SW7, CM9) with the cloned FV3 and FV5 sequences (appendix 2, figure S7). While all these sequences lacked a glycosylation site at position 295 (position 319 in figure S7), other key glycans important for 2G12 binding were conserved and located either in constant regions (332

and 339 in C4 (357 and 364 in figure S7), and 448 in C5 (474 in figure S7) or at the base of the variable loops (386 in V4 (410 in figure S7). However, the glycan at position 392 (~416 in figure S7), which has been shown to be critical for 2G12 binding[244, 245] occurs within the V4 loop, the length of which varies between these isolates. The position of this glycan within the loop is thus different for each isolate. This may affect accessibility of this glycan to 2G12 and possibly explains the binding of our recombinant subtype C gp120 to this antibody.

Another interesting observation with regards to these sequences is the presence of a conserved potential glycosylation site at position 442 (467). This has recently been implicated in the formation of the 2G12 epitope in subtype C HIV-1 isolates[246]. This glycosylation site is present in both FV3 and FV5 (shaded green in figure 4.5), but not in 96ZM651. A glycan present at this position may thus also play a role in the binding of FV3 and FV5 to 2G12.

Taken together, these results reinforce the important role played by glycans in formation of the epitope for 2G12, and highlight the complexity of the determinants of its binding to gp120.

4.6.3. CD4 Binding Studies

The CD4-gp120 binding data were difficult to interpret as they were influenced not only by the ability of the gp120 to bind to CD4, but also by the avidity of the monoclonal

antibodies used for detection of bound gp120. The orientation of the antibody epitopes with respect to the CD4 binding site was also an important factor, as gp120 was detected in the CD4-bound state. 2G12 was chosen as a principal detector of the interaction based on its ability to bind to all proteins used, except 96ZM651, in gp120-antibody binding studies (figures 3.11, S4). Experiments using 2G12 indicated the presence of a functional CD4 binding site in BaL, CM and FV3 (figures 3.12, 3.13). Further experiments using 17b, 48d, and A1g8, antibodies specific to the CD4-induced conformation of gp120, confirmed the presence of an intact CD4 binding site in these proteins, and further, in 96ZM651 (figures 3.12, 3.13). FV5 was consistently unable to bind to CD4 in these experiments.

In order to quantify the relative binding affinity of FV3 for CD4 as compared to BaL, the CD4 binding data were normalised against the data obtained from antibody recognition experiments. Results from the antibody 2G12 were used for normalisation as the binding of this antibody to gp120 is not affected by the presence of CD4. After normalisation, results indicated that, as compared to BaL, CD4 bound to FV3 with 22% efficiency and to CM with 28% efficiency. While CM and FV3 may have a lower affinity for CD4 than BaL, it is also possible that these values are an under-representation of actual binding. 2G12 binds extremely well to BaL, but has lower affinity for the other proteins studied. Thus, in both antibody recognition and CD4-binding studies, when the A_{450} readings for the other proteins were within the readable range, BaL absorbance readings were outside it and thus set to 3.5. It is likely that the absorbance readings for the BaL-2G12 interaction in the antibody recognition studies were higher than those in the CD4-binding

studies, but as both values were out of range and set to 3.5, this could not be determined and thus the binding efficiency of both FV3 and CM may be higher than the calculated values. We hope to more accurately determine the affinity of FV3 gp120 for CD4 by using Surface Plasmon Resonance. The apparent lesser affinity of FV3 and CM for CD4 as compared to BaL may be due to structural or glycosylation differences, as inspection of residues in contact with CD4 (figure 4.1) does not present an alternative explanation.

The lack of reactivity between CD4 and FV5 may be explained upon examination of the crystal structure of gp120 in complex with CD4[73]. Two residues of CD4, Arg59 and Phe43, form crucial interactions between gp120 and CD4. Arg59 interacts with Asp368 and Val430 of gp120, while Phe43 interacts with Asp368, Glu370, Ile371, Asn425, Met426, Trp427 and Gly473[73]. Inspection of the sequence of FV5 reveals an Asp to Gly substitution at Asp368. This mutation is not present in any of the other isolates. Mutations in Asp368 have serious consequences for CD4 binding[73]. Asp368 forms hydrogen bonds with Arg59 and Leu44 of CD4, as well as with the Phe43 aromatic ring. Absence of this residue, therefore, would seriously hinder binding of CD4 to gp120. Modeling of this substitution using Swiss-Pdb Viewer 3.7 (SP5)[86] indicates that a glycine residue at position 368 in FV5 does, indeed, prevent a number of hydrogen bonds forming at this position (figure 4.6). Interestingly, this mutation is not found in the population based sequence of FV5. It seems unlikely that the substitutions originated during PCR. It appears, therefore, that the FV5 gp120 encoded by pTriEx-FV5 represents a clone of a minor quasispecies present in the heterogenous pool of viral sequences sampled.

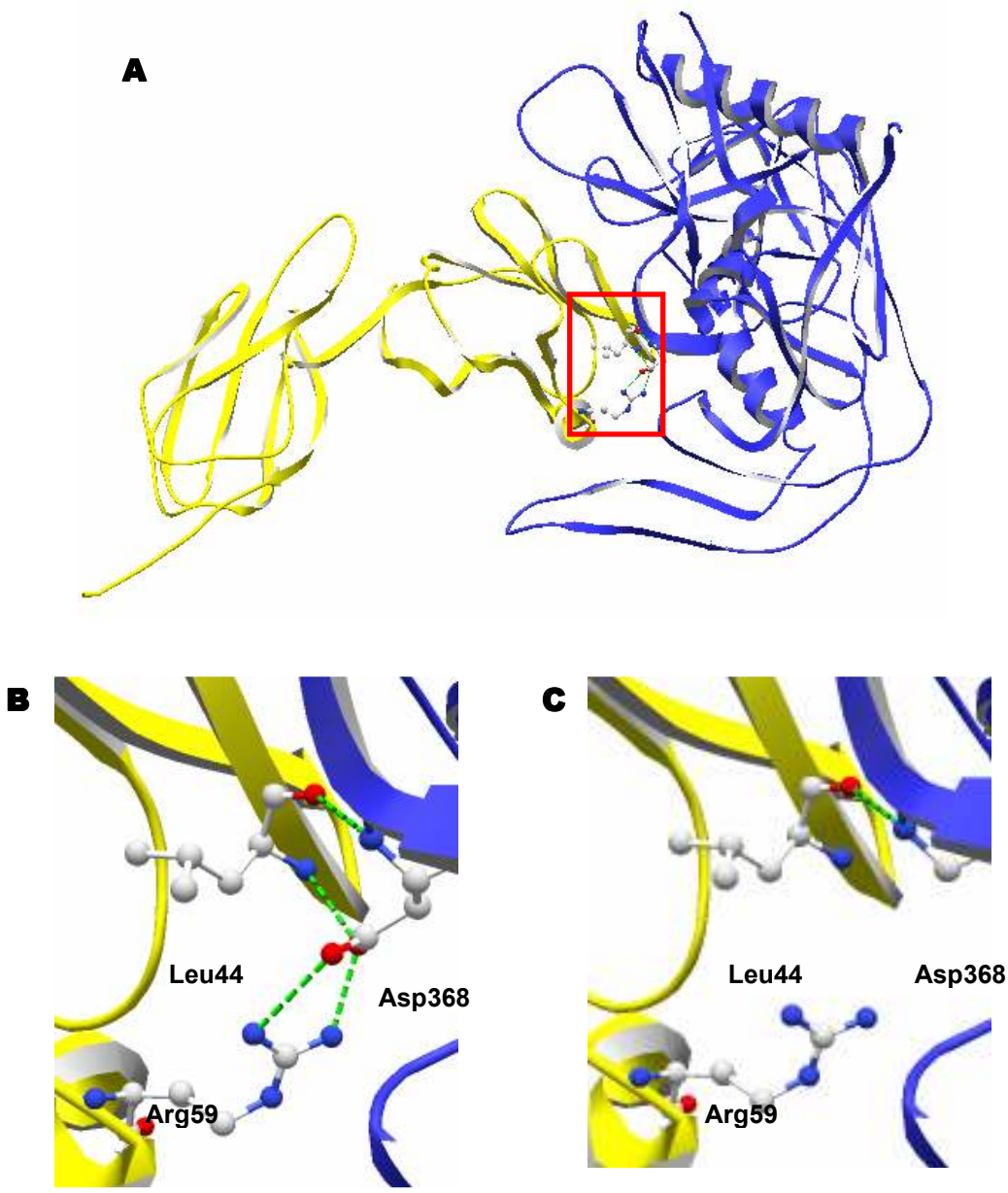


Figure 4.6: Substitution of Asp for Gly at gp120 position 368. Crystal structure downloaded from the Protein Data Bank[85], PDB ID: 1gc1, and altered using Swiss-Pdb Viewer 3.7 (SP5)[86] (http://www.expasy.org/spdbv/). Blue ribbon is gp120, yellow ribbon is CD4. **(A) Overview of gp120-CD4 binding.** Red block surrounds area zoomed in on in B and C. **(B) When Asp is present at position 368**. Dashed green lines indicate hydrogen bonds. **(B) When Asp is mutated to Gly at position 368.** Three hydrogen bonds at this position can no longer occur.

4.7. Perspective

CD4 and antibody binding studies provide valuable information about the nature of the newly synthesized subtype C recombinant glycoproteins FV3 and FV5. Antibodies were used that bind to varying linear regions along the length of gp120, as well as several that are dependent on intact conformation. The antibodies with linear epitopes confirmed the presence of a generally intact amino acid sequence, while the conformation dependent antibodies suggest canonical gp120 folding, at least in the case of FV3. Both FV3 and FV5 appear to have a unique antigenic structure, different to that of gp120s from other clades and from within subtype C.

It is important to note that analysis of the antibody binding studies was done on the presumption that the recombinant proteins, in particular FV3, were conformationally intact. This may not, however, be the case. The lack of binding of the recombinant proteins to many of the antibodies studied, some of which have been able to bind to Subtype C isolates in many other studies, may simply be a reflection of degraded or conformationally compromised proteins. Indeed, this is a plausible explanation and, based on this data, cannot be excluded. A number of factors contest this, however. Firstly, FV3 does show some binding activity with the conformationally dependent antibodies 2G12 and 654-30D. Secondly, FV5, which carries a mutation in a residue known to be critical for CD4 binding (D368G), does not react with CD4 or any antibodies requiring an intact CD4-binding site. The fact that this mutation has the expected effect on the ability of FV5 to bind CD4 indirectly supports the assumption that the proteins are conformationally intact.

The lack of binding to many of the antibodies studied may also be attributed to the fact that they were produced against clade B viruses, highlighting the fundamental challenge of broad HIV neutralisation. Minor changes in amino acid sequence can have profound effects on antibody recognition[188]. Major sequence differences between clades may also affect antibody recognition. Variation in amino acid sequence and resulting differences in glycosylation patterns between the different clades may result in conformational differences between proteins that extend outside the epitopes for each antibody. This may hinder the accessibility of these epitopes. Thus, even where epitopes are conserved between proteins from different clades, larger conformational differences may result in compromised antibody activity, or in antibodies binding to gp120 from other clades with much lesser affinity than to clade B proteins.

The glycosylation of these proteins is an issue that is also not completely resolved within the context of this project. This is important, as glycans may affect the binding of monoclonal antibodies by steric hindrance, or by affecting protein folding. In previous studies, Env proteins have been digested using trypsin and the resulting mixtures either directly analysed or deglycosylated and then analysed using mass spectrometry and/or matrix-assisted laser desorption ionization analysis techniques[68, 248].This has made it possible to determine both the extent and the nature of gp120 glycosylation. While such analysis would have been useful in order to further understand the antibody binding characteristics of FV3 and FV5, it was not feasible in this study. Nevertheless, different Env proteins vary in their glycosylation patterns and evidence suggests that glycosylation patterns even vary when expressing a single env[248]. These proteins are presumably no

different. Thus, while knowing the exact glycosylation status of these proteins may aid in understanding the antibody binding profiles, it is perhaps not directly important in this study. The presence of a uniform band of gp120 that is close to 120kDa when analysed by SDS-PAGE, and the ability of the gp120 to bind to lectin confirms that the gp120 is glycosylated to a reasonable and relatively uniform extent. The smaller bands, indicating incompletely glycosylated or partially degraded gp120, are a natural result of baculoviral infection, as cell membranes are damaged and the cytoplasm and incompletely processed proteins are released into the supernatant. The MOI and length of infection can be manipulated, however, in order to decrease the presence of such products. These products are also less evident when transfecting mammalian cells, as this procedure is less damaging to the cells than baculoviral infection, and in the long term it may also be useful to focus on expression of this protein in mammalian systems. Expression of mammalian proteins would also be a useful tool in order to investigate the effects of the different glycan processing machinery present in each system, and to determine whether this may be a cause of the different binding patterns seen.

FV3 and FV5 show variant antibody binding patterns to the other subtype C protein, 96ZM651. While 96ZM651 binds very well to B4e8 and 654-30D in the absence of CD4 as well as to 17b and A1g8 in the presence of CD4, FV3 and FV5 show little or no reactivity with these antibodies. In contrast, FV3 and FV5 show enhanced reactivity with 2G12 when compared 96ZM651. FV3 and FV5 therefore appear to be unique, even amongst subtype C isolates. These results, however, are based simply on two recombinant proteins, and it would be interesting to study subtype C proteins further,

both from within the same cohort as FV3 and FV5 and from others in order to determine the extent of variability between subtype C proteins and to elaborate on the determinants of monoclonal antibody binding. To this end, plans are in place to produce the mammalian counterpart of FV3 as well as to produce and characterise recombinant gp120 from other subtype C isolates. It would also be valuable to characterise these proteins further by characterizing their binding to other antibodies and to further determine, using more sophisticated techniques, the ability of these proteins to bind to CD4 and confirm their conformational integrity.

In the context of HIV therapeutics, it is of concern that FV3 did not show a large amount of reactivity with either of the two broadly neutralising antibodies, 2G12 and IgG1 b12, included in the panel. IgG1 b12 has previously shown some reactivity with subtype C isolates, as well as neutralising activity against them[187, 234, 235]. It does not, however, show reactivity with FV3 (or 96ZM651) and it has previously been shown that where antibodies have no reactivity with a virus, they have correspondingly low neutralizing activity[187].

Subtype C isolates either require huge concentrations of 2G12 for neutralization[249], or are altogether resistant to neutralization by this antibody[234, 235]. It is tempting to speculate that because FV3 shows some reactivity with 2G12 it would be susceptible to neutralisation by this antibody, although this may not be the case. Firstly, the ability of an antibody to bind to a virus does not necessarily correlate with its ability to neutralize it[187, 249, 250]. Secondly, it has also been noted, specifically in the case of 2G12, that weak binding of

the antibody to the protein, as in the case of FV3, results in weak neutralization ability of 2G12[249]. To this end, experiments are underway in our laboratory to evaluate the neutralizing potential of 2G12 in culture against these subtype C isolates. Interestingly, preliminary results show some degree of FV3 neutralization by 2G12, and this is being investigated further.

5. Conclusion

Advances in understanding the pathogenesis and molecular characteristics of HIV-1 assist in the development of strategies to combat the epidemic. Research has traditionally focused on subtype B HIV-1, leaving the other subtypes, including subtype C HIV-1 understudied. The prevalence and alarming spread of subtype C, however, highlights the importance of augmenting knowledge on the molecular biology and pathogenesis of this HIV-1 clade. This project was aimed at establishing protocols in our laboratory for the production of subtype C gp120 to be used for such research.

Once an effective protocol for production and purification of gp120 was in place, the antibody recognition profiles of the proteins were established and compared to other well-characterised recombinant gp120 proteins. Very few of the monoclonal anti-gp120 antibodies available showed any reactivity with the subtype C proteins. It was also interesting to note the highly variable profile of epitope presentation, even between closely related strains of HIV-1 (i.e. between the different subtype C proteins). While this could indicate that the recombinant proteins are conformationally compromised, there is evidence to suggest otherwise. Although these finding require further investigation, should they hold, they highlight both the importance of studying subtype C and comparing it to other subtypes, and the challenge of developing broadly-neutralising vaccine strategies. In this context, the importance of antibody-binding studies is paramount as these contribute towards elucidating the determinants of antibody recognition, essential information for the design of novel immunogens.

The FV3 gp120 produced in this study, along with a mammalian-expressed counterpart will be used for the development of entry inhibitors. The efficacy of existing gp120 antagonists to this protein will be evaluated and it will be used, together with phage display techniques in order to generate novel compounds. The sequence and antibody recognition information will contribute to the wider understanding of HIV-1, specifically in a subtype C context. While FV5 cannot be used for further drug development or other studies, the lack of binding of this protein to CD4 highlights the importance of a single key residue in gp120 for CD4 binding, confirming the potential of this region of gp120 as a profoundly important drug target.

Thus, a novel, and valuable subtype C recombinant gp120 has been produced and protocols are in place for continuous production of this protein, as well as other subtype C proteins of interest. Such proteins will used in the development of therapeutic strategies specific to South African HIV-1 strains, and will contribute, in the long term, to efficacious and cost-effective anti-HIV-1 approaches.

6. References

1. Gottlieb, M.S., et al., *Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency.* N Engl J Med, 1981. **305**(24): p. 1425-31.
2. Siegal, F.P., et al., *Severe acquired immunodeficiency in male homosexuals, manifested by chronic perianal ulcerative herpes simplex lesions.* N Engl J Med, 1981. **305**(24): p. 1439-44.
3. Gallo, R.C., et al., *Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS.* Science, 1984. **224**(4648): p. 500-3.
4. *Pneumocystis pneumonia--Los Angeles.* MMWR Morb Mortal Wkly Rep, 1981. **30**(21): p. 250-2.
5. Hymes, K.B., et al., *Kaposi's sarcoma in homosexual men-a report of eight cases.* Lancet, 1981. **2**(8247): p. 598-600.
6. Masur, H., et al., *An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction.* N Engl J Med, 1981. **305**(24): p. 1431-8.
7. *Update on acquired immune deficiency syndrome (AIDS)--United States.* MMWR Morb Mortal Wkly Rep, 1982. **31**(37): p. 507-8, 513-4.
8. *Pneumocystis carinii pneumonia among persons with hemophilia A.* MMWR Morb Mortal Wkly Rep, 1982. **31**(27): p. 365-7.
9. Harris, C., et al., *Immunodeficiency in female sexual partners of men with the acquired immunodeficiency syndrome.* N Engl J Med, 1983. **308**(20): p. 1181-4.
10. Gallo, R.C. and L. Montagnier, *The discovery of HIV as the cause of AIDS.* N Engl J Med, 2003. **349**(24): p. 2283-5.
11. Barre-Sinoussi, F., et al., *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).* Science, 1983. **220**(4599): p. 868-71.
12. Levy, J.A., et al., *Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS.* Science, 1984. **225**(4664): p. 840-2.

13. Popovic, M., et al., *Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS.* Science, 1984. **224**(4648): p. 497-500.
14. Salahuddin, S.Z., et al., *Isolation of infectious human T-cell leukemia/lymphotropic virus type III (HTLV-III) from patients with acquired immunodeficiency syndrome (AIDS) or AIDS-related complex (ARC) and from healthy carriers: a study of risk groups and tissue sources.* Proc Natl Acad Sci U S A, 1985. **82**(16): p. 5530-4.
15. Coffin, J., et al., *What to call the AIDS virus?* Nature, 1986. **321**(6065): p. 10.
16. Palca, J., *Virus nomenclature: controversy over AIDS virus extends to name.* Nature, 1986. **321**(6065): p. 3.
17. Clavel, F., et al., *Isolation of a new human retrovirus from West African patients with AIDS.* Science, 1986. **233**(4761): p. 343-6.
18. Marlink, R., et al., *Reduced rate of disease development after HIV-2 infection as compared to HIV-1.* Science, 1994. **265**(5178): p. 1587-90.
19. UNAIDS, *UNAIDS/WHO AIDS Epidemic Update: December 2006.* Geneva: UNAIDS, 2006.
20. UNAIDS, *2006 Report on the global AIDS epidemic.* Geneva: UNAIDS, 2006.
21. Arthur, L.O., et al., *Cellular proteins bound to immunodeficiency viruses: implications for pathogenesis and vaccines.* 1992. **v258**(n5090): p. p1935(4).
22. Turner, B.G. and M.F. Summers, *Structural biology of HIV*. J Mol Biol, 1999. **285**(1): p. 1-32.
23. Sierra, S., B. Kupfer, and R. Kaiser, *Basics of the virology of HIV-1 and its replication.* Journal of Clinical Virology, 2005. **34**: p. 233-44.
24. Leitner, T., et al., eds. *HIV Sequence Compendium 2005*. 2005, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, LA-UR number 06-0680.
25. Frankel, A.D. and J.A. Young, *HIV-1: fifteen proteins and an RNA.* Annu Rev Biochem, 1998. **67**: p. 1-25.
26. Saag, M.S., et al., *Extensive variation of human immunodeficiency virus type-1 in vivo.* Nature, 1988. **334**(6181): p. 440-4.

27. Shaw, G.M., et al., *Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome.* Science, 1984. **226**(4679): p. 1165-71.
28. Wong-Staal, F., et al., *Genomic diversity of human T-lymphotropic virus type III (HTLV-III).* Science, 1985. **229**(4715): p. 759-62.
29. Alizon, M., et al., *Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients.* Cell, 1986. **46**(1): p. 63-74.
30. Mansky, L.M. and H.M. Temin, *Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase.* J Virol, 1995. **69**(8): p. 5087-94.
31. Preston, B.D., B.J. Poiesz, and L.A. Loeb, *Fidelity of HIV-1 reverse transcriptase.* Science, 1988. **242**(4882): p. 1168-71.
32. Roberts, J.D., K. Bebenek, and T.A. Kunkel, *The accuracy of reverse transcriptase from HIV-1.* Science, 1988. **242**(4882): p. 1171-3.
33. Ho, D.D., et al., *Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection.* Nature, 1995. **373**(6510): p. 123-6.
34. Wei, X., et al., *Viral dynamics in human immunodeficiency virus type 1 infection.* Nature, 1995. **373**(6510): p. 117-22.
35. Perelson, A.S., et al., *HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time.* Science, 1996. **271**(5255): p. 1582-6.
36. Robertson, D.L., et al., *Recombination in HIV-1.* Nature, 1995. **374**(6518): p. 124-6.
37. Charneau, P., et al., *Isolation and envelope sequence of a highly divergent HIV-1 isolate: definition of a new HIV-1 group.* Virology, 1994. **205**(1): p. 247-53.
38. Mauclere, P., et al., *Serological and virological characterization of HIV-1 group O infection in Cameroon.* Aids, 1997. **11**(4): p. 445-53.
39. Peeters, M., et al., *Geographical distribution of HIV-1 group O viruses in Africa.* Aids, 1997. **11**(4): p. 493-8.
40. Simon, F., et al., *Identification of a new human immunodeficiency virus type 1 distinct from group M and group O.* Nat Med, 1998. **4**(9): p. 1032-7.

41. Louwagie, J., et al., *Phylogenetic analysis of gag genes from 70 international HIV-1 isolates provides evidence for multiple genotypes.* Aids, 1993. **7**(6): p. 769-80.
42. Robertson, D.L., et al., *HIV-1 nomenclature proposal.* Science, 2000. **288**(5463): p. 55-6.
43. Gaschen, B., et al., *Diversity considerations in HIV-1 vaccine selection.* Science, 2002. **296**(5577): p. 2354-60.
44. Korber, B., et al., *Evolutionary and immunological implications of contemporary HIV-1 variation.* Br Med Bull, 2001. **58**: p. 19-42.
45. *Los Alamos HIV Sequence Database* <http://www.hiv.lanl.gov/content/hiv-db/mainpage.html>, [Accessed Jan 2005 - May 2007].
46. Hemelaar, J., et al., *Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004.* Aids, 2006. **20**(16): p. W13-23.
47. Grewe, C., A. Beck, and H.R. Gelderblom, *HIV: early virus-cell interactions.* J Acquir Immune Defic Syndr, 1990. **3**(10): p. 965-74.
48. Fassati, A. and S.P. Goff, *Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1.* J Virol, 2001. **75**(8): p. 3626-35.
49. McDonald, D., et al., *Visualization of the intracellular behavior of HIV in living cells.* J Cell Biol, 2002. **159**(3): p. 441-52.
50. Mitchell, R.S., et al., *Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences.* PLoS Biol, 2004. **2**(8): p. E234.
51. Schroder, A.R., et al., *HIV-1 integration in the human genome favors active genes and local hotspots.* Cell, 2002. **110**(4): p. 521-9.
52. Veronese, F.D., et al., *Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene.* Science, 1985. **229**(4720): p. 1402-5.
53. Willey, R.L., et al., *Biosynthesis, cleavage, and degradation of the human immunodeficiency virus 1 envelope glycoprotein gp160.* Proc Natl Acad Sci U S A, 1988. **85**(24): p. 9580-4.

54. Earl, P.L., B. Moss, and R.W. Doms, *Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type 1 envelope protein.* J Virol, 1991. **65**(4): p. 2047-55.

55. Stein, B.S. and E.G. Engleman, *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, 1990. **265**(5): p. 2640-9.

56. Dettenhofer, M. and X.F. Yu, *Characterization of the biosynthesis of human immunodeficiency virus type 1 Env from infected T-cells and the effects of glucose trimming of Env on virion infectivity.* J Biol Chem, 2001. **276**(8): p. 5985-91.

57. McCune, J.M., et al., *Endoproteolytic cleavage of gp160 is required for the activation of human immunodeficiency virus.* Cell, 1988. **53**(1): p. 55-67.

58. Hallenberger, S., et al., *The role of eukaryotic subtilisin-like endoproteases for the activation of human immunodeficiency virus glycoproteins in natural host cells.* J Virol, 1997. **71**(2): p. 1036-45.

59. Bresnahan, P.A., et al., *Human fur gene encodes a yeast KEX2-like endoprotease that cleaves pro-beta-NGF in vivo.* J Cell Biol, 1990. **111**(6 Pt 2): p. 2851-9.

60. Wise, R.J., et al., *Expression of a human proprotein processing enzyme: correct cleavage of the von Willebrand factor precursor at a paired basic amino acid site.* Proc Natl Acad Sci U S A, 1990. **87**(23): p. 9378-82.

61. Hallenberger, S., et al., *Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160.* Nature, 1992. **360**(6402): p. 358-61.

62. Center, R.J., et al., *Oligomeric structure of the human immunodeficiency virus type 1 envelope protein on the virion surface.* J Virol, 2002. **76**(15): p. 7863-7.

63. Zhu, P., et al., *Distribution and three-dimensional structure of AIDS virus envelope spikes.* Nature, 2006. **441**(7095): p. 847-52.

64. Allan, J.S., et al., *Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III.* Science, 1985. **228**(4703): p. 1091-4.

65. Starcich, B.R., et al., *Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS.* Cell, 1986. **45**(5): p. 637-48.

66. Modrow, S., et al., *Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions.* J Virol, 1987. **61**(2): p. 570-8.
67. Willey, R.L., et al., *Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus.* Proc Natl Acad Sci U S A, 1986. **83**(14): p. 5038-42.
68. Leonard, C.K., et al., *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, 1990. **265**(18): p. 10373-82.
69. Lasky, L.A., et al., *Neutralization of the AIDS retrovirus by antibodies to a recombinant envelope glycoprotein.* Science, 1986. **233**(4760): p. 209-12.
70. Wei, X., et al., *Antibody neutralization and escape by HIV-1.* Nature, 2003. **422**(6929): p. 307-12.
71. Derdeyn, C.A., et al., *Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission.* Science, 2004. **303**(5666): p. 2019-22.
72. Kwong, P.D., et al., *Probability Analysis of Variational Crystallization and Its Application to gp120, The Exterior Envelope Glycoprotein of Type 1 Human Immunodeficiency Virus (HIV-1).* J Biol Chem, 1999. **274**: p. 4115-23.
73. Kwong, P.D., et al., *Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody.* Nature, 1998. **393**(6686): p. 648-59.
74. Kwong, P.D., et al., *Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of human immunodeficiency virus.* J Virol, 2000. **74**(4): p. 1961-72.
75. Chen, B., et al., *Structure of an unliganded simian immunodeficiency virus gp120 core.* Nature, 2005. **433**(7028): p. 834-41.
76. Huang, C.C., et al., *Structure of a V3-containing HIV-1 gp120 core.* Science, 2005. **310**(5750): p. 1025-8.
77. Dalgleish, A.G., et al., *The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus.* Nature, 1984. **312**(5996): p. 763-7.

78. Klatzmann, D., et al., *T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV*. Nature, 1984. **312**(5996): p. 767-8.
79. Feng, Y., et al., *HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor.* Science, 1996. **272**(5263): p. 872-7.
80. Alkhatib, G., et al., *CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1.* Science, 1996. **272**(5270): p. 1955-8.
81. Choe, H., et al., *The beta-chemokine receptors CCR3 and CCR5 facilitate infection by primary HIV-1 isolates.* Cell, 1996. **85**(7): p. 1135-48.
82. Deng, H., et al., *Identification of a major co-receptor for primary isolates of HIV-1.* Nature, 1996. **381**(6584): p. 661-6.
83. Doranz, B.J., et al., *A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors.* Cell, 1996. **85**(7): p. 1149-58.
84. Dragic, T., et al., *HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5.* Nature, 1996. **381**(6584): p. 667-73.
85. Berman, H.M., et al., *The Protein Data Bank.* Nucl. Acids Res., 2000. **28**(1): p. 235-242.
86. Guex, N. and M.C. Peitsch, *SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling.* . Electrophoresis, 1997. **18**: p. 2714-23.
87. Myszka, D.G., et al., *Energetics of the HIV gp120-CD4 binding reaction.* Proc Natl Acad Sci U S A, 2000. **97**(16): p. 9026-31.
88. Hwang, S.S., et al., *Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1.* Science, 1991. **253**(5015): p. 71-4.
89. Sattentau, Q.J. and J.P. Moore, *Conformational changes induced in the human immunodeficiency virus envelope glycoprotein by soluble CD4 binding.* J Exp Med, 1991. **174**(2): p. 407-15.
90. Kolchinsky, P., et al., *Loss of a single N-linked glycan allows CD4-independent human immunodeficiency virus type 1 infection by altering the position of the gp120 V1/V2 variable loops.* J Virol, 2001. **75**(7): p. 3435-43.

91. Edinger, A.L., et al., *Use of GPR1, GPR15, and STRL33 as coreceptors by diverse human immunodeficiency virus type 1 and simian immunodeficiency virus envelope proteins.* Virology, 1998. **249**(2): p. 367-78.
92. Liao, F., et al., *STRL33, A novel chemokine receptor-like protein, functions as a fusion cofactor for both macrophage-tropic and T cell line-tropic HIV-1.* J Exp Med, 1997. **185**(11): p. 2015-23.
93. Rucker, J., et al., *Utilization of chemokine receptors, orphan receptors, and herpesvirus-encoded receptors by diverse human and simian immunodeficiency viruses.* J Virol, 1997. **71**(12): p. 8999-9007.
94. Pleskoff, O., et al., *Identification of a chemokine receptor encoded by human cytomegalovirus as a cofactor for HIV-1 entry.* Science, 1997. **276**(5320): p. 1874-8.
95. Choe, H., et al., *The orphan seven-transmembrane receptor apj supports the entry of primary T-cell-line-tropic and dualtropic human immunodeficiency virus type 1.* J Virol, 1998. **72**(7): p. 6113-8.
96. Samson, M., et al., *ChemR23, a putative chemoattractant receptor, is expressed in monocyte-derived dendritic cells and macrophages and is a coreceptor for SIV and some primary HIV-1 strains.* Eur J Immunol, 1998. **28**(5): p. 1689-700.
97. Bjorndal, A., et al., *Coreceptor usage of primary human immunodeficiency virus type 1 isolates varies according to biological phenotype.* J Virol, 1997. **71**(10): p. 7478-87.
98. de Roda Husman, A.M., et al., *Adaptation to promiscuous usage of chemokine receptors is not a prerequisite for human immunodeficiency virus type 1 disease progression.* J Infect Dis, 1999. **180**(4): p. 1106-15.
99. Boyd, M.T., et al., *A single amino acid substitution in the V1 loop of human immunodeficiency virus type 1 gp120 alters cellular tropism.* J Virol, 1993. **67**(6): p. 3649-52.
100. Cho, M.W., et al., *Identification of determinants on a dualtropic human immunodeficiency virus type 1 envelope glycoprotein that confer usage of CXCR4.* J Virol, 1998. **72**(3): p. 2509-15.

101. Kim, F.M., et al., *V3-independent determinants of macrophage tropism in a primary human immunodeficiency virus type 1 isolate.* J Virol, 1995. **69**(3): p. 1755-61.

102. Koito, A., et al., *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, 1994. **68**(4): p. 2253-9.

103. Trkola, A., et al., *CD4-dependent, antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5.* Nature, 1996. **384**(6605): p. 184-7.

104. Wu, L., et al., *CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5.* Nature, 1996. **384**(6605): p. 179-83.

105. Zhu, T., et al., *Genotypic and phenotypic characterization of HIV-1 patients with primary infection.* Science, 1993. **261**(5125): p. 1179-81.

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

107. Connor, R.I. and D.D. Ho, *Human immunodeficiency virus type 1 variants with increased replicative capacity develop during the asymptomatic stage before disease progression.* J Virol, 1994. **68**(7): p. 4400-8.

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

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

110. Schramm, B., et al., *Viral entry through CXCR4 is a pathogenic factor and therapeutic target in human immunodeficiency virus type 1 disease.* J Virol, 2000. **74**(1): p. 184-92.

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

112. Tersmette, M., et al., *Association between biological properties of human immunodeficiency virus variants and risk for AIDS and AIDS mortality.* Lancet, 1989. **1**(8645): p. 983-5.

113. Kaneshima, H., et al., *Rapid-high, syncytium-inducing isolates of human immunodeficiency virus type 1 induce cytopathicity in the human thymus of the SCID-hu mouse.* J Virol, 1994. **68**(12): p. 8188-92.

114. Kreisberg, J.F., et al., *Cytopathicity of human immunodeficiency virus type 1 primary isolates depends on coreceptor usage and not patient disease status.* J Virol, 2001. **75**(18): p. 8842-7.

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

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

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

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

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

120. Peeters, M., et al., *Evidence for differences in MT2 cell tropism according to genetic subtypes of HIV-1: syncytium-inducing variants seem rare among subtype C HIV-1 viruses.* J Acquir Immune Defic Syndr Hum Retrovirol, 1999. **20**(2): p. 115-21.

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

122. Tien, P.C., et al., *Primary subtype C HIV-1 infection in Harare, Zimbabwe.* J Acquir Immune Defic Syndr Hum Retrovirol, 1999. **20**(2): p. 147-53.
123. van Rensburg, E.J., et al., *Change in co-receptor usage of current South African HIV-1 subtype C primary isolates.* Aids, 2002. **16**(18): p. 2479-80.
124. Connell, B.J., et al., *Emergence of X4 usage among HIV-1 subtype C: evidence for an evolving epidemic in South Africa.* Aids, 2008. **22**(7): p. 896-9.
125. Chan, D.C., et al., *Core structure of gp41 from the HIV envelope glycoprotein.* Cell, 1997. **89**(2): p. 263-73.
126. Tan, K., et al., *Atomic structure of a thermostable subdomain of HIV-1 gp41.* Proc Natl Acad Sci U S A, 1997. **94**(23): p. 12303-8.
127. Weissenhorn, W., et al., *Atomic structure of the ectodomain from HIV-1 gp41.* Nature, 1997. **387**(6631): p. 426-30.
128. Lu, M., S.C. Blacklow, and P.S. Kim, *A trimeric structural domain of the HIV-1 transmembrane glycoprotein.* Nat Struct Biol, 1995. **2**(12): p. 1075-82.
129. Gallaher, W.R., et al., *A general model for the transmembrane proteins of HIV and other retroviruses.* AIDS Res Hum Retroviruses, 1989. **5**(4): p. 431-40.
130. Colman, P.M. and M.C. Lawrence, *The structural biology of type I viral membrane fusion.* Nat Rev Mol Cell Biol, 2003. **4**(4): p. 309-19.
131. Eckert, D.M. and P.S. Kim, *Mechanisms of viral membrane fusion and its inhibition.* Annu Rev Biochem, 2001. **70**: p. 777-810.
132. Harrison, S.C., *Mechanism of membrane fusion by viral envelope proteins.* Adv Virus Res, 2005. **64**: p. 231-61.
133. Markosyan, R.M., F.S. Cohen, and G.B. Melikyan, *HIV-1 envelope proteins complete their folding into six-helix bundles immediately after fusion pore formation.* Mol Biol Cell, 2003. **14**(3): p. 926-38.
134. Melikyan, G.B., et al., *Evidence that the transition of HIV-1 gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion.* J Cell Biol, 2000. **151**(2): p. 413-23.
135. Doms, R.W. and J.P. Moore, *HIV-1 membrane fusion: targets of opportunity.* J Cell Biol, 2000. **151**(2): p. F9-14.

136. Turpin, J.A., *The next generation of HIV/AIDS drugs: novel and developmental antiHIV drugs and targets.* Expert Rev Anti Infect Ther, 2003. **1**(1): p. 97-128.

137. Palella, F.J., Jr., et al., *Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators.* N Engl J Med, 1998. **338**(13): p. 853-60.

138. Smith, D.H., et al., *Blocking of HIV-1 infectivity by a soluble, secreted form of the CD4 antigen.* Science, 1987. **238**(4834): p. 1704-7.

139. Deen, K.C., et al., *A soluble form of CD4 (T4) protein inhibits AIDS virus infection.* Nature, 1988. **331**(6151): p. 82-4.

140. Daar, E.S., et al., *High concentrations of recombinant soluble CD4 are required to neutralize primary human immunodeficiency virus type 1 isolates.* Proc Natl Acad Sci U S A, 1990. **87**(17): p. 6574-8.

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

142. Zhu, P., W.C. Olson, and K.H. Roux, *Structural flexibility and functional valence of CD4-IgG2 (PRO 542): potential for cross-linking human immunodeficiency virus type 1 envelope spikes.* J Virol, 2001. **75**(14): p. 6682-6.

143. Jacobson, J.M., et al., *Treatment of advanced human immunodeficiency virus type 1 disease with the viral entry inhibitor PRO 542.* Antimicrob Agents Chemother, 2004. **48**(2): p. 423-9.

144. Trkola, A., et al., *Potent, broad-spectrum inhibition of human immunodeficiency virus type 1 by the CCR5 monoclonal antibody PRO 140.* J Virol, 2001. **75**(2): p. 579-88.

145. Olson, W.C., et al. *Prolonged Coating of CCR5 Lymphocytes by PRO 140, a Humanized CCR5 Monoclonal Antibody for HIV-1 Therapy, abstract 515*. in *13th Conference on Retroviruses and Opportunistic Infections, San Francisco, CA*. 2006. http://www.retroconference.org/2006/home.htm.

146. Burkly, L.C., et al., *Inhibition of HIV infection by a novel CD4 domain 2-specific monoclonal antibody. Dissecting the basis for its inhibitory effect on HIV-induced cell fusion.* J Immunol, 1992. **149**(5): p. 1779-87.

147. Moore, J.P., et al., *A monoclonal antibody to CD4 domain 2 blocks soluble CD4-induced conformational changes in the envelope glycoproteins of human immunodeficiency virus type 1 (HIV-1) and HIV-1 infection of CD4+ cells.* J Virol, 1992. **66**(8): p. 4784-93.

148. Jacobson, J.M., et al. *Phase 1b Study of the Anti-CD4 Monoclonal Antibody TNX-355 in HIV-1-infected Subjects: Safety and Antiretroviral Activity of Multiple Doses, abstract 536*. in *11th Conference on Retroviruses and Opportunistic Infections, San Francisco, CA*. 2004. http://www.retroconference.org/2004/home.htm.

149. Kuritzkes, D.R., et al., *Antiretroviral activity of the anti-CD4 monoclonal antibody TNX-355 in patients infected with HIV type 1.* J Infect Dis, 2004. **189**(2): p. 286-91.

150. Wang, T., et al., *Discovery of 4-benzoyl-1-[(4-methoxy-1H- pyrrolo[2,3-b]pyridin-3-yl)oxoacetyl]-2- (R)-methylpiperazine (BMS-378806): a novel HIV-1 attachment inhibitor that interferes with CD4-gp120 interactions.* J Med Chem, 2003. **46**(20): p. 4236-9.

151. Lin, P.F., et al., *A small molecule HIV-1 inhibitor that targets the HIV-1 envelope and inhibits CD4 receptor binding.* Proc Natl Acad Sci U S A, 2003. **100**(19): p. 11013-8.

152. Guo, Q., et al., *Biochemical and genetic characterizations of a novel human immunodeficiency virus type 1 inhibitor that blocks gp120-CD4 interactions.* J Virol, 2003. **77**(19): p. 10528-36.

153. Veazey, R.S., et al., *Protection of macaques from vaginal SHIV challenge by vaginally delivered inhibitors of virus-cell fusion.* Nature, 2005. **438**(7064): p. 99-102.

154. Moyle, G., et al. *CXCR4 Antagonism: Proof of Activity with AMD11070, abstract 511*. in *14th Conference on Retroviruses and Opportunistic Infections, San Francisco, CA*. 2007. http://www.retroconference.org/2007/home.htm.

155. Schols, D., et al. *Anti-HIV Activity Profile of AMD070, an Orally Bioavailable CXCR4 Antagonist H, abstract 563*. in *10th Conference on Retroviruses and*

Opportunistic Infections, San Francisco, CA. 2003. http://www.retroconference.org/2003/home.htm.

156. Dorr, P., et al., *Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity.* Antimicrob Agents Chemother, 2005. **49**(11): p. 4721-32.

157. Lalezari, J., et al. *Efficacy and Safety of Maraviroc plus Optimized Background Therapy in Viremic ART-experienced Patients Infected with CCR5-tropic HIV-1: 24-Week Results of a Phase 2b/3 Study in the US and Canada, abstract 104bLB*. in *14th Conference on Retroviruses and Opportunistic Infections, San Francisco, CA*. 2007. http://www.retroconference.org/2007/home.htm.

158. Nelson, M., et al. *Efficacy and Safety of Maraviroc plus Optimized Background Therapy in Viremic, ART-experienced Patients Infected with CCR5-tropic HIV-1 in Europe, Australia, and North America: 24-Week Results, abstract 104aLB*. in *14th Conference on Retroviruses and Opportunistic Infections, San Francisco, CA*. 2007. http://www.retroconference.org/2007/home.htm.

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

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

161. Pert, C.B., et al., *Octapeptides deduced from the neuropeptide receptor-like pattern of antigen T4 in brain potently inhibit human immunodeficiency virus receptor binding and T-cell infectivity.* Proc Natl Acad Sci U S A, 1986. **83**(23): p. 9254-8.

162. Redwine, L.S., et al., *Peptide T blocks GP120/CCR5 chemokine receptor-mediated chemotaxis.* Clin Immunol, 1999. **93**(2): p. 124-31.

163. Polianova, M.T., et al., *Antiviral and immunological benefits in HIV patients receiving intranasal peptide T (DAPTA).* Peptides, 2003. **24**(7): p. 1093-8.

164. Ferrer, M. and S.C. Harrison, *Peptide ligands to human immunodeficiency virus type 1 gp120 identified from phage display libraries.* J Virol, 1999. **73**(7): p. 5795-802.

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

166. Biorn, A.C., et al., *Mode of action for linear peptide inhibitors of HIV-1 gp120 interactions.* Biochemistry, 2004. **43**(7): p. 1928-38.

167. Cocklin, S., et al., *Broad-spectrum anti-human immunodeficiency virus (HIV) potential of a peptide HIV type 1 entry inhibitor.* J Virol, 2007. **81**(7): p. 3645-8.

168. Gopi, H.N., et al., *Click chemistry on azidoproline: high-affinity dual antagonist for HIV-1 envelope glycoprotein gp120.* ChemMedChem, 2006. **1**(1): p. 54-7.

169. Morikawa, Y., et al., *Expression of HIV-1 gp120 and human soluble CD4 by recombinant baculoviruses and their interaction in vitro.* AIDS Res Hum Retroviruses, 1990. **6**(6): p. 765-73.

170. Murphy, C.I., et al., *Temporal expression of HIV-1 envelope proteins in baculovirus-infected insect cells: implications for glycosylation and CD4 binding.* Genet Anal Tech Appl, 1990. **7**(6): p. 160-71.

171. Nkosi, S.P., H. Huismans, and M.A. Papathanasopoulos, *Purification and partial characterization of R5, R5X4, and X4 HIV-1 subtype C envelope glycoproteins expressed in insect cells.* J Med Virol, 2005. **76**(4): p. 459-63.

172. Jones, D.H., et al., *Efficient purification and rigorous characterisation of a recombinant gp120 for HIV vaccine studies.* Vaccine, 1995. **13**(11): p. 991-999.

173. Gao, F., et al., *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, 1996. **70**(3): p. 1651-67.

174. *Entrez Nucleotide Database* <http://www.ncbi.nlm.nih.gov/sites/entrez?db=Nucleotide> [Accessed: Jan 2005 - June 2006].

175. Abacioglu, Y.H., et al., *Epitope mapping and topology of baculovirus-expressed HIV-1 gp160 determined with a panel of murine monoclonal antibodies.* AIDS Res Hum Retroviruses, 1994. **10**(4): p. 371-81.

176. Dickey, C., et al., *Murine monoclonal antibodies biologically active against the amino region of HIV-1 gp120: isolation and characterization.* DNA Cell Biol, 2000. **19**(4): p. 243-52.

177. Gomez-Roman, V.R., et al., *Phage-displayed mimotopes recognizing a biologically active anti-HIV-1 gp120 murine monoclonal antibody.* J Acquir Immune Defic Syndr, 2002. **31**(2): p. 147-53.

178. Dairou, J., C. Vever-Bizet, and D. Brault, *Interaction of sulfonated anionic porphyrins with HIV glycoprotein gp120: photodamages revealed by inhibition of antibody binding to V3 and C5 domains.* Antiviral Res, 2004. **61**(1): p. 37-47.

179. McDougal, J.S., et al., *Mechanisms of human immunodeficiency virus Type 1 (HIV-1) neutralization: irreversible inactivation of infectivity by anti-HIV-1 antibody.* J Virol, 1996. **70**(8): p. 5236-45.

180. Durda, P.J., et al., *Characterization of murine monoclonal antibodies to HIV-1 induced by synthetic peptides.* AIDS Res Hum Retroviruses, 1988. **4**(5): p. 331-42.

181. Gorny, M.K., et al., *Human anti-V2 monoclonal antibody that neutralizes primary but not laboratory isolates of human immunodeficiency virus type 1.* J Virol, 1994. **68**(12): p. 8312-20.

182. Hioe, C.E., et al., *Neutralization of HIV-1 primary isolates by polyclonal and monoclonal human antibodies.* Int Immunol, 1997. **9**(9): p. 1281-90.

183. Laal, S., et al., *Synergistic neutralization of human immunodeficiency virus type 1 by combinations of human monoclonal antibodies.* J Virol, 1994. **68**(6): p. 4001-8.

184. Zolla-Pazner, S., et al., *Serotyping of primary human immunodeficiency virus type 1 isolates from diverse geographic locations by flow cytometry.* J Virol, 1995. **69**(6): p. 3807-15.

185. von Brunn, A., et al., *Principal neutralizing domain of HIV-1 is highly immunogenic when expressed on the surface of hepatitis B core particles.* Vaccine, 1993. **11**(8): p. 817-24.

186. Von Brunn, A., et al., *The principal neutralizing determinant (V3) of HIV-1 induces HIV-1 neutralizing antibodies upon expression on HBcAg particles. In: Vaccines 93*. 1993, Cold Spring Harbor Press; Cold Spring Harbor, New York. p. 159-165.

187. Nyambi, P.N., et al., *Conserved and exposed epitopes on intact, native, primary human immunodeficiency virus type 1 virions of group M.* J Virol, 2000. **74**(15): p. 7096-107.

188. Moore, J.P., et al., *Antigenic variation in gp120s from molecular clones of HIV-1 LAI.* AIDS Res Hum Retroviruses, 1993. **9**(12): p. 1185-93.

189. Thali, M., et al., *Characterization of conserved human immunodeficiency virus type 1 gp120 neutralization epitopes exposed upon gp120-CD4 binding.* J Virol, 1993. **67**(7): p. 3978-88.

190. Buchacher, A., et al., *Generation of human monoclonal antibodies against HIV-1 proteins; electrofusion and Epstein-Barr virus transformation for peripheral blood lymphocyte immortalization.* AIDS Res Hum Retroviruses, 1994. **10**(4): p. 359-69.

191. Crawford, J.M., et al., *Characterization of primary isolate-like variants of simian-human immunodeficiency virus.* J Virol, 1999. **73**(12): p. 10199-207.

192. Etemad-Moghadam, B., et al., *Determinants of neutralization resistance in the envelope glycoproteins of a simian-human immunodeficiency virus passaged in vivo.* J Virol, 1999. **73**(10): p. 8873-9.

193. Mascola, J.R., et al., *Protection of Macaques against pathogenic simian/human immunodeficiency virus 89.6PD by passive transfer of neutralizing antibodies.* J Virol, 1999. **73**(5): p. 4009-18.

194. Trkola, A., et al., *Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1.* J Virol, 1996. **70**(2): p. 1100-8.

195. Cavacini, L.A., et al., *Human monoclonal antibodies to the V3 loop of HIV-1 gp120 mediate variable and distinct effects on binding and viral neutralization by a human monoclonal antibody to the CD4 binding site.* J Acquir Immune Defic Syndr, 1993. **6**(4): p. 353-8.

196. Posner, M.R., et al., *Neutralization of HIV-1 by F105, a human monoclonal antibody to the CD4 binding site of gp120.* J Acquir Immune Defic Syndr, 1993. **6**(1): p. 7-14.

197. Posner, M.R., H. Elboim, and D. Santos, *The construction and use of a human-mouse myeloma analogue suitable for the routine production of hybridomas secreting human monoclonal antibodies.* Hybridoma, 1987. **6**(6): p. 611-25.

198. Posner, M.R., et al., *An IgG human monoclonal antibody that reacts with HIV-1/GP120, inhibits virus binding to cells, and neutralizes infection.* J Immunol, 1991. **146**(12): p. 4325-32.

199. Sullivan, N., et al., *CD4-Induced conformational changes in the human immunodeficiency virus type 1 gp120 glycoprotein: consequences for virus entry and neutralization.* J Virol, 1998. **72**(6): p. 4694-703.

200. Wyatt, R., et al., *The antigenic structure of the HIV gp120 envelope glycoprotein.* Nature, 1998. **393**(6686): p. 705-11.

201. Wyatt, R., et al., *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, 1995. **69**(9): p. 5723-33.

202. Barbas, C.F., 3rd, et al., *Recombinant human Fab fragments neutralize human type 1 immunodeficiency virus in vitro.* Proc Natl Acad Sci U S A, 1992. **89**(19): p. 9339-43.

203. Burton, D.R., et al., *A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals.* Proc Natl Acad Sci U S A, 1991. **88**(22): p. 10134-7.

204. Burton, D.R., et al., *Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody.* Science, 1994. **266**(5187): p. 1024-7.

205. Roben, P., et al., *Recognition properties of a panel of human recombinant Fab fragments to the CD4 binding site of gp120 that show differing abilities to neutralize human immunodeficiency virus type 1.* J Virol, 1994. **68**(8): p. 4821-8.

206. *National Centre for Biotechnology Information* <http://www.ncbi.nlm.nih.gov/>, [Accessed: March 2006].

207. Nicholas, K.B., N.H.B. Jr., and D.W.I. Deerfield, *GeneDoc: Analysis and Visualization of Genetic Variation.* EMBNEW.NEWS 1997. **4**: p. 14.

208. *ExPASy Proteomics Server* <http://ca.expasy.org/>, [Acessed March 2006 - May 2006].

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

210. Coetzer, M., et al., *Genetic characteristics of the V3 region associated with CXCR4 usage in HIV-1 subtype C isolates.* Virology, 2006. **356**(1-2): p. 95-105.

211. Jensen, M.A., et al., *A reliable phenotype predictor for human immunodeficiency virus type 1 subtype C based on envelope V3 sequences.* J Virol, 2006. **80**(10): p. 4698-704.

212. Olshevsky, U., et al., *Identification of individual human immunodeficiency virus type 1 gpl20 amino acids important for CD4 receptor binding.* J Virol, 1990. **64**: p. 5701-7.

213. Adams, O., H. Schaal, and A. Scheid, *Natural variation in the amino acid sequence around the HIV type 1 glycoprotein 160 cleavage site and its effect on cleavability, subunit association, and membrane fusion.* AIDS Res Hum Retroviruses, 2000. **16**(13): p. 1235-45.

214. Fenouillet, E. and J.C. Gluckman, *Immunological analysis of human immunodeficiency virus type 1 envelope glycoprotein proteolytic cleavage.* Virology, 1992. **187**(2): p. 825-8.

215. Clevestig, P., et al., *CCR5 use by human immunodeficiency virus type 1 is associated closely with the gp120 V3 loop N-linked glycosylation site.* J Gen Virol, 2006. **87**(Pt 3): p. 607-12.

216. Scanlan, C.N., et al., *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, 2002. **76**(14): p. 7306-21.

217. Morikawa, Y., J.P. Moore, and I.M. Jones, *HIV-l envelope protein gp120 expression by secretion in E. coli: assessment of CD4 binding and use in epitope mapping.* Journal of Virological Methods, 1990. **29**: p. 105-14.

218. Lo, H.-R. and Y.-C. Chao, *Rapid Titer Determination of Baculovirus by Quantitative Real-Time Polymerase Chain Reaction.* Biotechnol. Prog., 2004. **20**: p. 354-60.

219. Murphy, C.I., et al., *Enhanced expression, secretion, and large-scale purification of recombinant HIV-1 gp120 in insect cell using the baculovirus egt and p67 signal peptides.* Protein Expr Purif, 1993. **4**(5): p. 349-57.

220. Wang, Y.H., A.H. Davies, and I.M. Jones, *Expression and purification of glutathione S-transferase-tagged HIV-1 gp120: no evidence of an interaction with CD26.* Virology, 1995. **208**(1): p. 142-6.

221. Zhang, C.W., et al., *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, 2001. **276**(43): p. 39577-85.

222. Srivastava, I.K., et al., *Purification and characterization of oligomeric envelope glycoprotein from a primary R5 subtype B human immunodeficiency virus.* J Virol, 2002. **76**(6): p. 2835-47.

223. Varadarajan, R., et al., *Characterization of gp120 and Its Single-Chain Derivatives, gp120-CD4D12 and gp120-M9: Implications for Targeting the CD4i Epitope in Human Immunodeficiency Virus Vaccine Design.* J Virol, 2005. **79**(3): p. 1713-23.

224. Jeffs, S.A., et al., *Expression and characterisation of recombinant oligomeric envelope glycoproteins derived from primary isolates of HIV-1.* Vaccine, 2004. **22**: p. 1032-46.

225. Ivey-Hoyle, M., et al., *Envelope glycoproteins from biologically diverse isolates of immunodeficiency viruses have widely different affinities for CD4.* Proc Natl Acad Sci U S A, 1991. **88**: p. 512-16.

226. Lopez, M., et al., *Microheterogeneity of the oligosaccharides carried by the recombinant bovine lactoferrin expressed in Mamestra brassicae cells.* Glycobiology, 1997. **7**(5): p. 635-51.

227. Voss, T., et al., *Expression of human interferon omega 1 in Sf9 cells. No evidence for complex-type N-linked glycosylation or sialylation.* Eur J Biochem, 1993. **217**(3): p. 913-9.

228. Kuroda, K., et al., *The oligosaccharides of influenza virus hemagglutinin expressed in insect cells by a baculovirus vector.* Virology, 1990. **174**(2): p. 418-29.

229. Yeh, J.C., et al., *Site-specific N-glycosylation and oligosaccharide structures of recombinant HIV-1 gp120 derived from a baculovirus expression system.* Biochemistry, 1993. **32**(41): p. 11087-99.

230. Bristow, G.W., et al., *Analysis of murine antibody responses to baculovirus-expressed human immunodeficiency virus type 1 envelope glycoproteins.* J Gen Virol, 1994. **75**: p. 2089-95.

231. Chen, H., et al., *Reintroduction of the 2G12 epitope in an HIV-1 clade C gp120.* Aids, 2005. **19**(8): p. 833-5.

232. Wilkinson, R.A., et al., *Structure of the Fab fragment of F105, a broadly reactive anti-human immunodeficiency virus (HIV) antibody that recognizes the CD4 binding site of HIV type 1 gp120.* J Virol, 2005. **79**(20): p. 13060-9.

233. Saphire, E.O., et al., *Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design.* Science, 2001. **293**(5532): p. 1155-9.

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

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

236. Cavacini, L.A., et al., *Interactions of human antibodies, epitope exposure, antibody binding and neutralization of primary isolate HIV-1 virions.* Aids, 2002. **16**(18): p. 2409-17.

237. Cavacini, L., et al., *Conformational changes in env oligomer induced by an antibody dependent on the V3 loop base.* Aids, 2003. **17**(5): p. 685-9.

238. Moore, J.P., et al., *Exploration of antigenic variation in gp120 from clades A through F of human immunodeficiency virus type 1 by using monoclonal antibodies.* J Virol, 1994. **68**(12): p. 8350-64.

239. Rizzuto, C. and J. Sodroski, *Fine definition of a conserved CCR5-binding region on the human immunodeficiency virus type 1 glycoprotein 120.* AIDS Res Hum Retroviruses, 2000. **16**(8): p. 741-9.

240. Rizzuto, C.D., et al., *A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding.* Science, 1998. **280**(5371): p. 1949-53.

241. Pantophlet, R., et al., *Analysis of the neutralization breadth of the anti-V3 antibody F425-B4e8 and re-assessment of its epitope fine specificity by scanning mutagenesis.* Virology, 2007. **364**: p. 441-53.

242. Bell, C.H., et al., *Structure of antibody F425-B4e8 in complex with a V3 peptide reveals a new binding mode for HIV-1 neutralization.* J Mol Biol, 2008. **375**(4): p. 969-78.

243. Nyambi, P.N., et al., *Mapping of epitopes exposed on intact human immunodeficiency virus type 1 (HIV-1) virions: a new strategy for studying the immunologic relatedness of HIV-1.* J Virol, 1998. **72**(11): p. 9384-91.

244. Sanders, R.W., et al., *The Mannose-Dependent Epitope for Neutralizing Antibody 2G12 on Human Immunodeficiency Virus Type 1 Glycoprotein gp120.* J. Virol., 2002. **76**(14): p. 7293-7305.

245. Calarese, D.A., et al., *Antibody domain exchange is an immunological solution to carbohydrate cluster recognition.* Science, 2003. **300**(5628): p. 2065-71.

246. Gray, E.S., et al., *N-linked glycan modifications in the gp120 of HIV-1 subtype C renders partial sensitivity to 2G12 antibody neutralization.* J. Virol., 2007: p. JVI.01106-07.

247. Gray, E.S., et al., *Insensitivity of paediatric HIV-1 subtype C viruses to broadly neutralising monoclonal antibodies raised against subtype B.* PLoS Med, 2006. **3**(7): p. e255.

248. Go, E.P., et al., *Glycosylation Site-Specific Analysis of HIV Envelope Proteins (JR-FL and CON-S) Reveals Major Differences in Glycosylation Site Occupancy,*

Glycoform Profiles, and Antigenic Epitopes' Accessibility. J Proteome Res, 2008. **7**(4): p. 1660-74.

249. Trkola, A., et al., *Cross-clade neutralization of primary isolates of human immunodeficiency virus type 1 by human monoclonal antibodies and tetrameric CD4-IgG.* J Virol, 1995. **69**(11): p. 6609-17.

250. York, J., et al., *Antibody binding and neutralization of primary and T-cell line-adapted isolates of human immunodeficiency virus type 1.* J Virol, 2001. **75**(6): p. 2741-52.

7. Appendices

Appendix 1: Standard Protocols and Recipes

A1. Bacterial Culture

A1.1. Solutions for Bacterial Culture

Luria Broth (LB) (1 litre)

10 g tryptone, 5 g yeast extract and 5 g Nacl were combined and the volume brought to one litre. The solution was then autoclaved and stored at room temperature until use.

Agar

3 g Agar was added per 200 ml of LB prior to autoclaving. Agar was then autoclaved and cooled slightly before it was used to pour plates.

Ampicillin or Chloramphenicol Stock (1000 x/ 100 mg/ml)

1 g ampicillin or chloramphenicol was dissolved in 10 ml 50% Ethanol

Transformation Buffer (100 mM $CaCl_2$, 10 mM PIPES-HCl, 15% Glycerol, pH 7)

1.4702 g $CaCl_2.2H_2O$, 0.3024 g PIPES and 15 ml glycerol were mixed and brought to a volume of 75 ml with dH_2O. This solution was brought to pH 7.0 with 10 M NaOH and made up to a final volume up to 100 ml with dH_2O. The transformation buffer was then autoclaved and stored at -20ºC.

A1.2. Preparation of competent *E. coli*

100 ml of Luria Broth (LB) was inoculated with either *E. coli* DH5α or *E. coli* BL21. For *E. coli* BL21 , chloramphenicol was included at a concentration of 0.1 mg/ml. Cultures were grown at 37°C overnight with vigourous shaking. The culture was then diluted 10 times in LB or LB with chloramphenicol and grown at 37°C with vigourous shaking for a further 2-3 hours until the bacteria were in the log phase of growth with an $A_{600} \sim 0.4$. Cells were pelleted at 2000 x g for 15 minutes at 4°C and the supernatant discarded. The pellet was carefully resuspended in 10 ml ice-cold transformation buffer per 50 ml culture and incubated for 20 minutes on ice. Cells were pelleted at 1500 x g for 15 minutes at 4°C and the supernatant discarded. The pellet was then carefully resuspended in 1ml ice-cold transformation buffer per 50 ml original culture volume and, working on ice, 75 µl aliquots of competent *E. coli* were made and stored at -70°C

A1.3. Transformation of competent *E. coli*

5 µl (approximately 50 ng) of each ligation reaction or 2 µl (approximately 50 ng) plasmid DNA was added to a 100 µl aliquot of competent *E.coli* and incubated on ice for 20-25 minutes. A mock transformation reaction containing no DNA was also set up. The cells were then heat shocked for 30 seconds at 42˚C and placed immediately on ice for 2 minutes. 250 µl SOC medium (Invitrogen; Carlsbad, CA) was added to each of the tubes, which were then placed in a shaking incubator at 37˚C for 1 hour. Each transformation reaction was then plated on 2 agar plates (100 µl and 250 µl) and incubated at 37˚C overnight.

E.coli DH5α were plated on plates containing 0.1 mg/ml ampicillin (Roche; Mannheim, Germany), while *E.coli* BL21 were plated on plates containing 0.1 mg/ml ampicillin and 0.1 mg/ml chloramphenicol (Sigma; St Louis, MO).

A2. Agarose Electrophoresis

A2.1 Solutions for Agarose Electrophoresis

0.5 M EDTA (pH 8.0)

93.05 g EDTA was dissolved in 400 ml dH_2O and brought to pH 8.0 with 10 M NaOH. It was then made up to 500 ml with dH_2O, autoclaved and stored at room temperature.

50 x TAE Buffer

242 g Tris base, 57.1 ml glacial acetic acid and 100 ml 0.5 M EDTA pH 8.0 were combined and the solution made up to 1 litre in dH_2O. The solution was made up to 1 x TAE by mixing 20 ml 50 x TAE with 980 ml dH_2O before use.

A3. SDS-PAGE

A3.1. Solutions for SDS-PAGE

4 x Running Gel Buffer (1.5 M Tris-HCl, pH 8.8)

36.3 g Tris was dissolved in 50 ml dH_2O, and the pH brought to 8.8 with HCl. The solution was then made up to a final volume of 200 ml with dH_2O and stored at 4°C in the dark.

4 x Stacking gel buffer (0.5 M Tris-HCl, pH 6.8)

3.0 g Tris was dissolved in 40 ml dH_2O, and the pH brought to 6.8 with HCl. The solution was then made up to a final volume of 50 ml with dH_2O and stored at 4°C in the dark

10% Sodium Dodecyl Sulphate (SDS)

10 g SDS was dissolved in 100 ml dH_2O

10% Ammonium Persulphate (APS)

0.1 g APS was dissolved in 1 ml dH_2O. This solution was made up fresh for each gel.

Running gel (For 4 gels):

12 ml Monomer solution (Sigma; St Louis, MO), 11.25 ml 4x running gel buffer, 21.15 ml dH_2O and 450 µl 10% SDS were combined. Just prior to pouring the gel, 225 µl 10% APS and 15 µl TEMED were added to the solution.

Stacking gel

2 ml Monomer solution (Sigma; St Louis, MO), 3.75 ml 4 x stacking gel buffer, 9 ml dH_2O and 150 µl 10% SDS were combined. Just prior to pouring the stacking gel, 75 µl 10% APS and 7.5 µl TEMED were added to the solution

2 x Treatment Buffer

2.5 ml 4 x stacking gel buffer, 4.0 ml glycerol, 2.0 mg bromophenol blue and 0.31 g Dithiothreitol were combined and the volume brought to 10 ml with dH_2O.

5 x Tank Buffer

15.14 g Tris, 72.726 g Glycine and 5 g SDS were combined and the volume brought to one liter with dH_2O. This solution was made up to 1 x tank buffer just before use.

Coomassie Staining Soultion

0.5 g Coomassie® Brilliant Blue R250 was combined with 800 ml methanol, and stirred until dissolved. 140 ml acetic acid was then added and the volume brought to 2 litres with dH_2O.

Destaining Solution 1 (40% Methanol, 7% Acetic Acid)

400 ml methanol and 70 ml acetic acid were combined and the volume brought to one liter with dH_2O.

Destaining Solution 2 (5% Methanol, 7% Acetic Acid)

50 ml methanol was combined with 70 ml acetic acid and the volume brought to 1 litre with dH_2O.

A3.2. Resolving SDS-PAGE gels

8% polyacrylamide gels were poured, and 35 µl of each sample loaded into the wells, together with 10 µl of Prestained Protein Marker, Broad Range (New England Biolabs; Ipswich, MA) and a positive control. Gels were run in duplicate for approximately 1.5 hours, at constant current of 12.5 mA, until the samples were stacked, followed by a constant current of 25-30 mA, until the dye front was approximately 0.5 cm from the bottom of the gel. One gel was then stained with Coomassie® Brilliant Blue R-250 overnight and destained the next day for two hours in Destain I, followed by Destain II until the background of the gel was clear.

A4. Western Blotting

A4.1. Solutions for Western Blotting

Transfer Buffer

100 ml methanol was combined with 100 ml 5 x tank buffer and the volume brought to 500 ml with dH_2O

10 x TBS

80 g NaCl, 2 g KCl and 30 g Tris were combined and the volume was brought to 800 ml with dH_2O. The pH was brought to 7.4 with HCl and the solution brought to a final volume of 1 litre. The solution was then autoclaved and diluted to 1 x TBS just before use. 1 ml Tween-20 was added to 1x TBS just before use to make T-TBS.

A4.2. Performing a Western Blot

After running SDS-PAGE, the duplicate SDS-PAGE gel was equilibrated in transfer buffer for 10 minutes, and then transferred to Hybond-C nitrocellulose membrane (Amersham Biosciences; Freiburg, Germany) on a Trans-Blot SD Semi-Dry Transfer Cell (Bio-Rad; Hercules, CA) for 1 hour and 15 minutes at constant current of 50 mA. Following protein transfer, samples were blocked for one hour shaking at room temperature or overnight at 4°C in 5% fat free milk powder in T-TBS . Three 5-minute washes in T-TBS followed before addition of the primary antibody.

After incubation with the primary antibody (as described in section 2.8.4), the membrane was washed three times for five minutes in T-TBS. The secondary antibody was then added at a concentration of 1:1000 diluted in 2 ml and laid flat on the membrane for 45 minutes. Depending on the primary antibody, the secondary antibody used was either anti-mouse in the case of the mouse monoclonal antibodies, or anti-Human in the case of plasma. After this incubation, the membrane was again washed three times for 5 minutes in T-TBS before protein detection.

Bound secondary antibody was detected on Western Blot membranes using enhanced chemiluminescence (ECL). A commercial substrate, SuperSignal West Pico Chemiluminescent Substrate (Pierce; Rockford, IL), was used according to the manufacturer's instructions. After the enzymatic reaction, membranes were exposed to Fuji Super Rx Medical X-Ray film (Africa Xray Industrial & Medical; La Rochelle, South Africa) for varying amounts of time, depending on the signal strength, before developing and fixing.

A5. Protein Concentration Determination

Protein was quantitated using a BCA Protein Assay Kit (Pierce; Rockford, IL). Briefly, a standard curve was set up in a 96-well plate (Nunc; Roskilde, Denmark) using the BSA standard provided in the kit. BSA was diluted in the same buffer as the protein to be quantitated to concentrations of 0 µg/ml, 2 µg/ml, 5 µg/ml, 10 µg/ml, 15 µg/ml, 20 µg/ml, 25 µg/ml and 50 µg/ml. 25 µl of standard (added in duplicate) or 10 x diluted protein (added in triplicate) pipetted into the plate, followed by 200 µl of working reagent made up according to the manufacturer's instructions. Plates were incubated at 56°C for up to an hour and the absorbance was read at 570 nm on a EL_X808 microplate reader (Bio-Tek; Winooski, VT). The protein concentration was then calculated according to the generated standard curve.

Appendix 2: Supplementary Figures

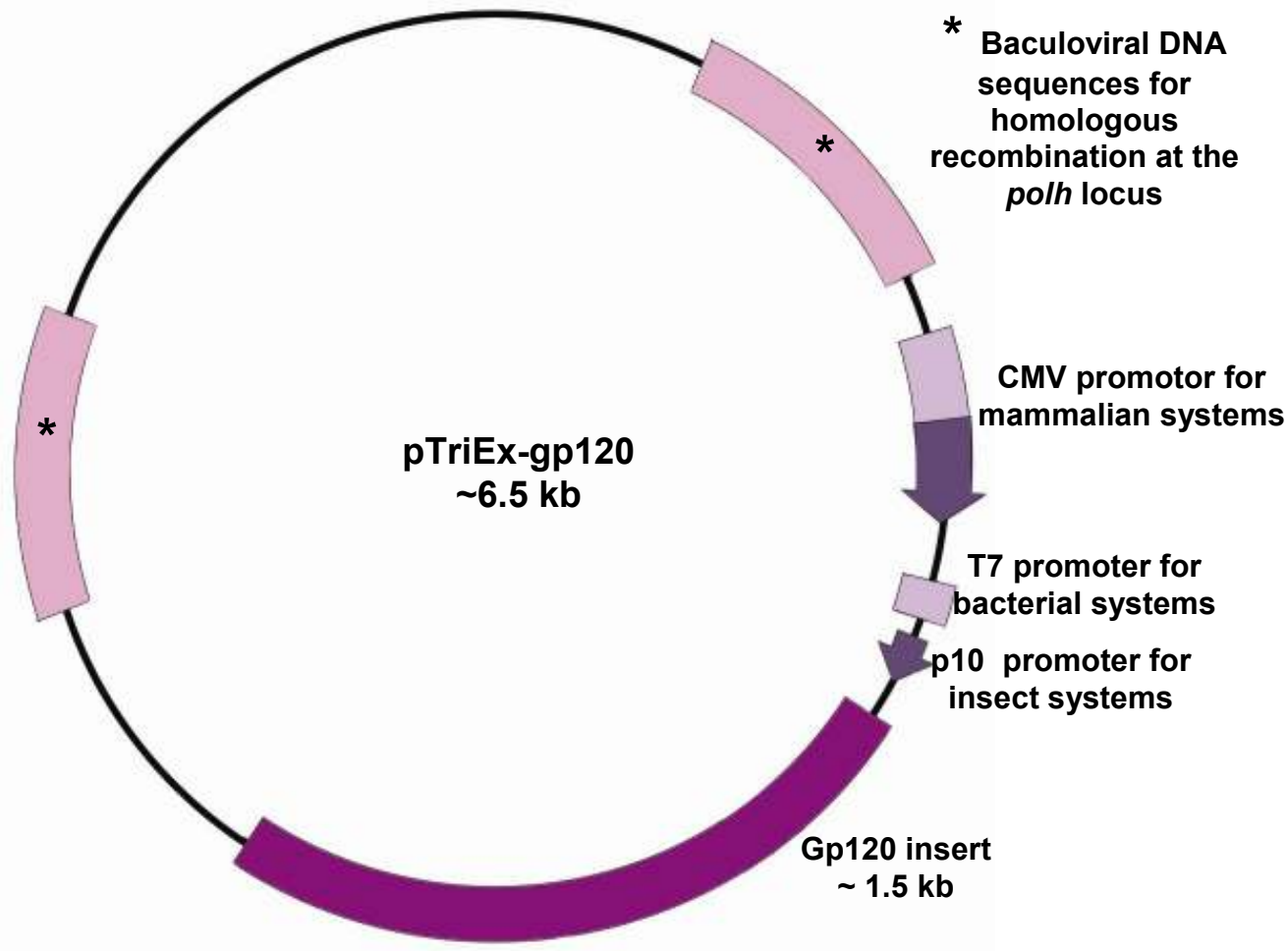


Figure S1: Schematic representation of pTriEx-gp120. The gp120-coding region was inserted into the multiple cloning site between the *Nco*I and *Xho*I digestion sites. The pTriEx-gp120 clones were approximately 6.5 kb in length. pTriEx contains promoters for bacterial, insect cell and mammalian protein expression. The promoter and cloning regions of pTriEx are flanked by segments of baculovirus genomic DNA (lef-2/ORF603 and ORF1629) This allows for homologous recombination with the baculoviral backbone at the *polh* locus to create TriEx recombinant viruses.

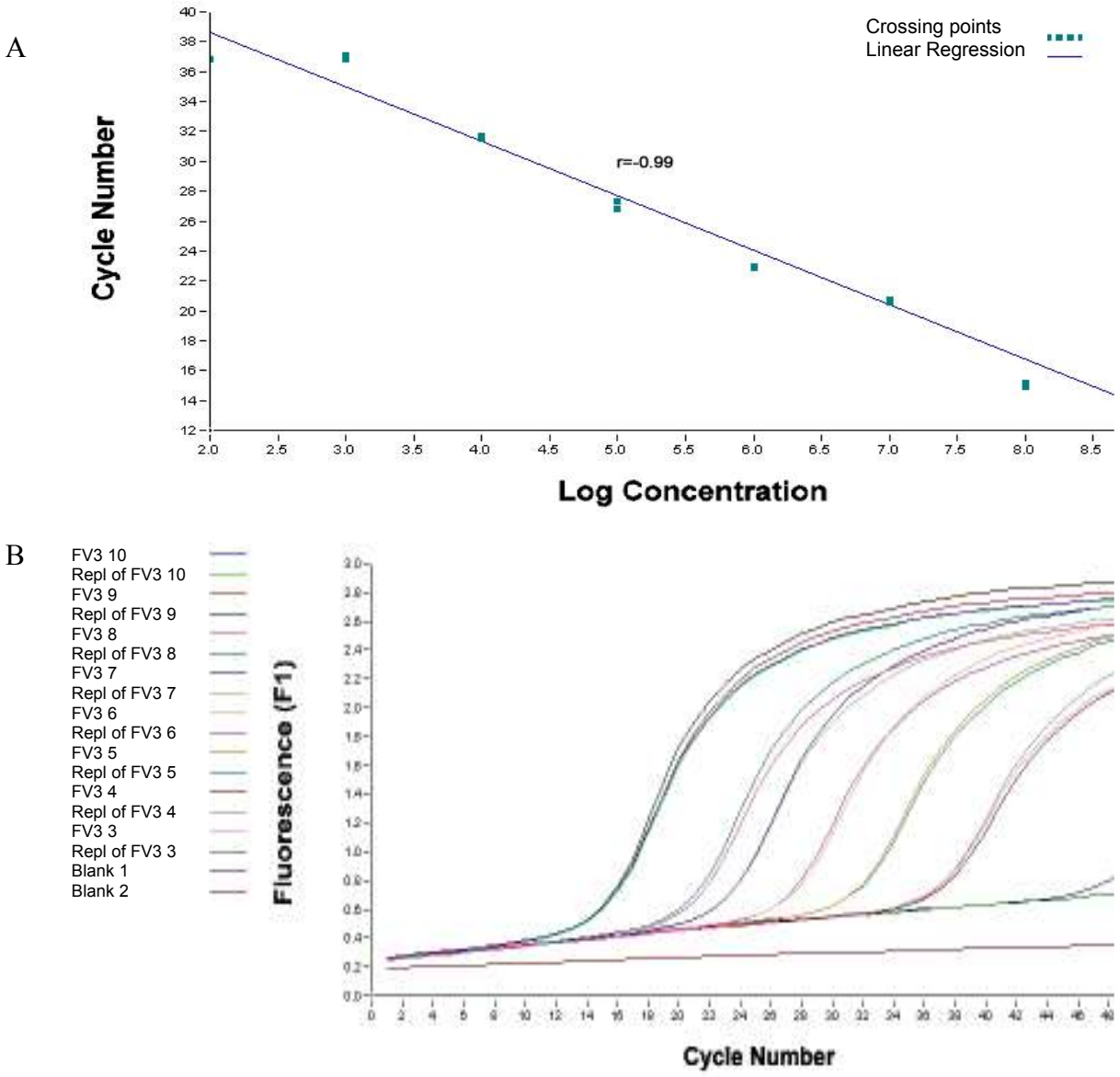
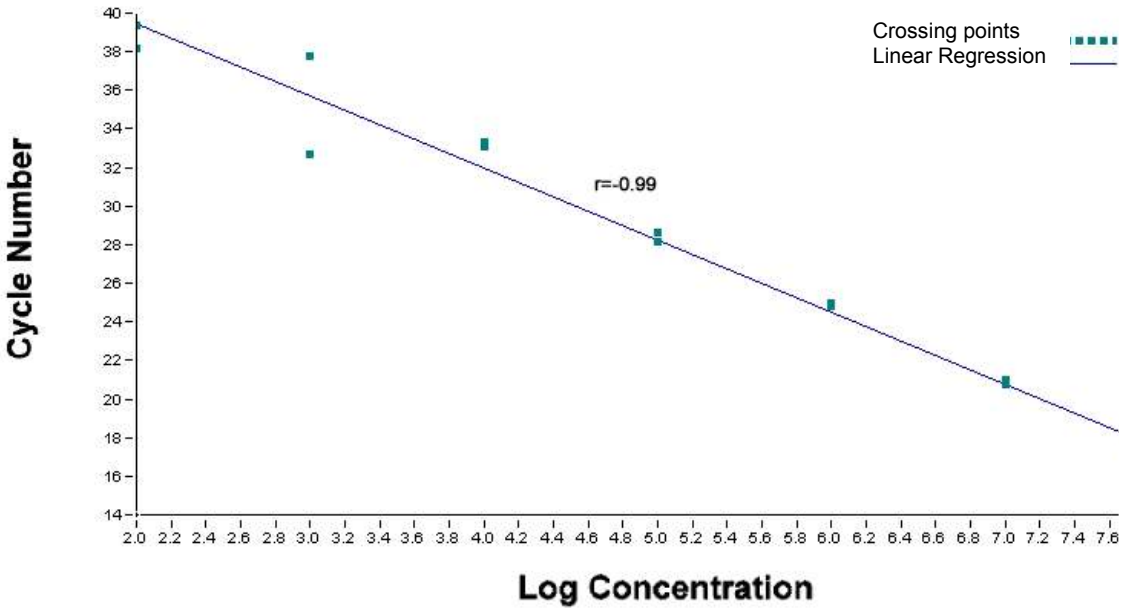


Figure S2: Standard Curve generation for FV3 by amplification of a 180 bp segment of gp120-encoding DNA. (A) Std curve generated by LightCycler® Software version 3.5.3. (B) Amplification curves for standards. Number to right of FV3 indicates the log number of copies per ml of the sample added to the reaction. i.e. For FV3 10, the concentration of the sample added to the PCR reaction was 1 x 10^{10} gp120 copies/ml

A



B

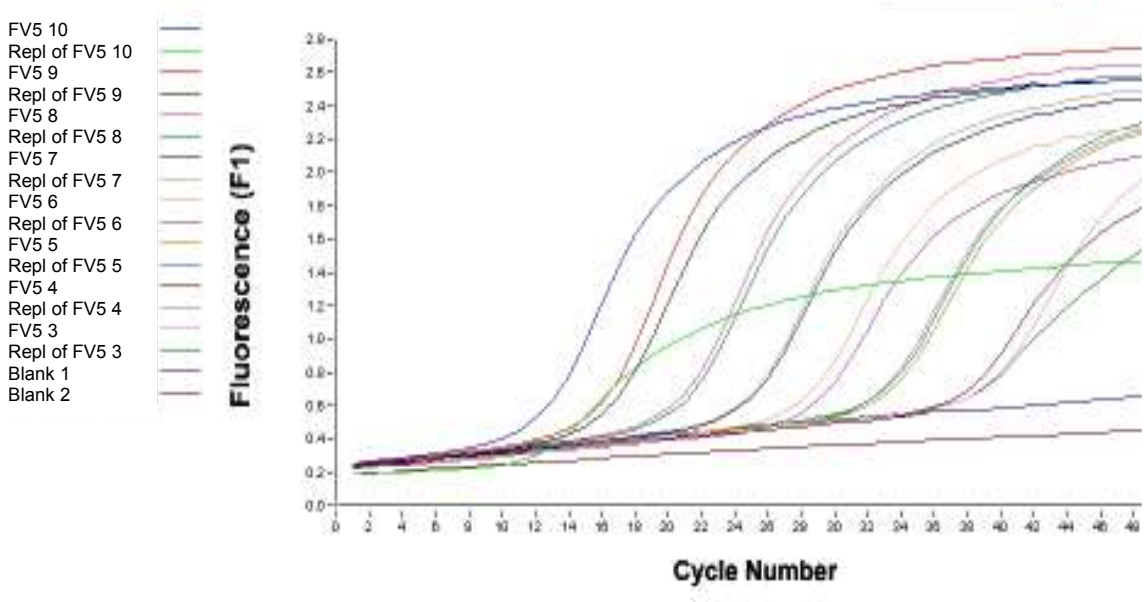


Figure S3: Standard Curve generation for FV5 by amplification of a 180 bp segment of gp120-encoding DNA. (A) Std curve generated by LightCycler® Software version 3.5.3. (B) Amplification curves of standards of known concentration. Number to right of FV5 in the key indicates the log number of copies per ml of the sample added to the reaction. i.e. For FV5 10, the concentration of the sample added to the PCR reaction was 1×10^{10} gp120 copies/ml

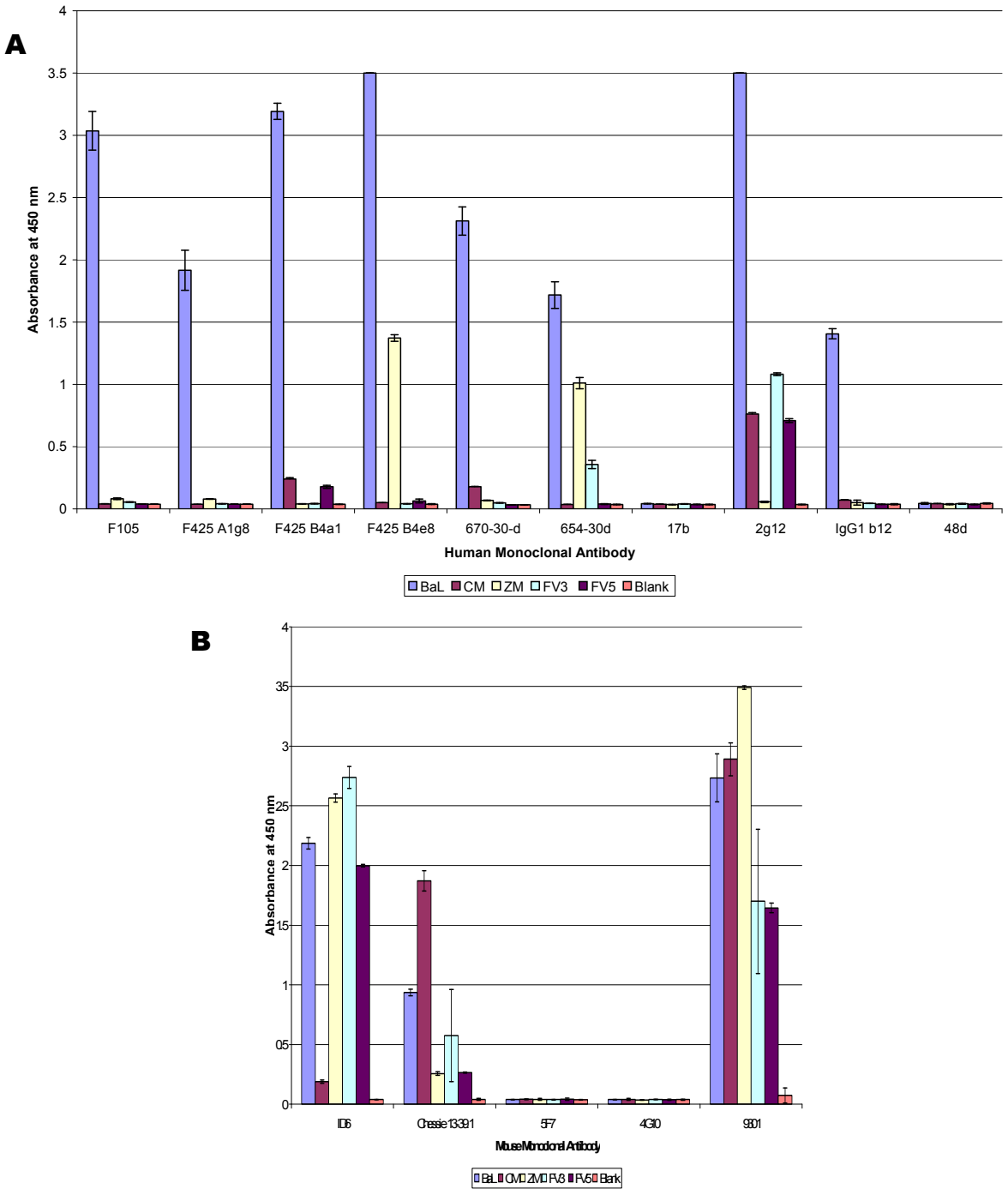
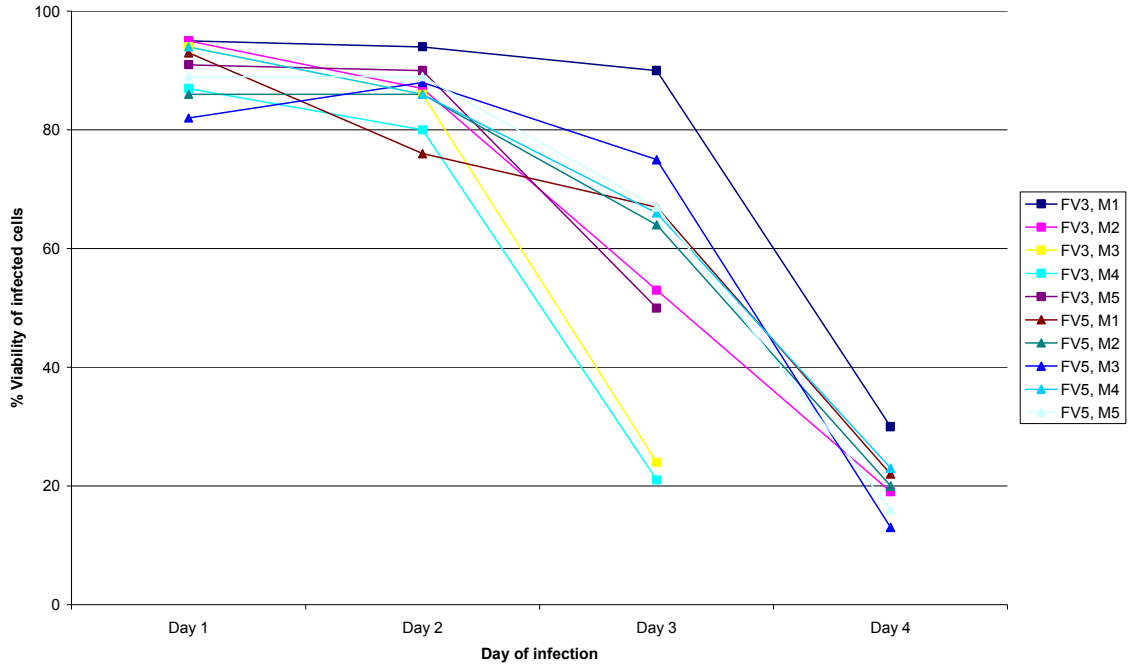


Figure S4: Graphs showing recognition of monoclonal antibodies by a panel of gp120s. Antibody recognition of gp120s was analysed by direct ELISA. Graphs show the mean A_{450} value of a triplicate experiment. Error bars indicate standard deviation. **(A) Human monoclonal antibody recognition (B) Mouse monoclonal antibody recognition.**



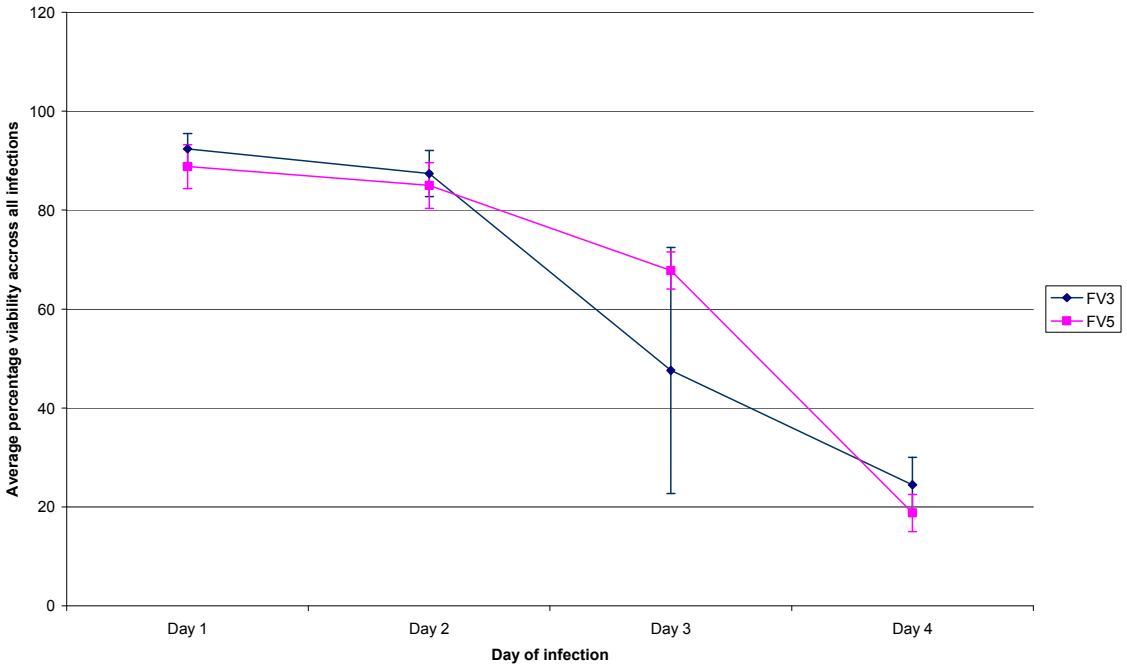


Figure S5: Diagrams showing percentage cell viability in insect cell suspension culture infections. (A) Cell Viability in individual infections, M: MOI in gp120-encoding copies/cell (B) All infection viabilities averaged for FV3 and FV5 each.

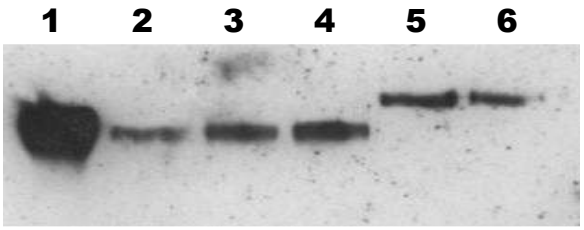


Figure S6: Comparison between insect cell-produced FV3 and mammalian cell-produced His-tagged FV3. Western Blot analysis. 1: BaL. 2,3,4: Insect cell- produced FV3. 5,6: Mammalin cell-produced His-tagged FV3.

Figure S7: Alignment of FV3 and FV5 with population-based amino acids of subtype C isolates used in other studies[171, 231]. Glycosylation sites important for formation of the 2G12 epitope are highlighted. Where the site is absent, the position it should be in is highlighted in green. Where sites are present, they are highlighted in pink.